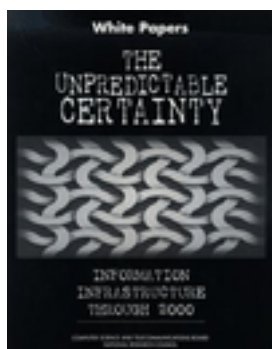


## The Unpredictable Certainty: White Papers



NII 2000 Steering Committee, Commission on Physical Sciences, Mathematics, and Applications, National Research Council

ISBN: 0-309-06036-2, 632 pages, 8.5 x 11, (1997)

**This free PDF was downloaded from:**  
<http://www.nap.edu/catalog/6062.html>

Visit the [National Academies Press](#) online, the authoritative source for all books from the [National Academy of Sciences](#), the [National Academy of Engineering](#), the [Institute of Medicine](#), and the [National Research Council](#):

- Download hundreds of free books in PDF
- Read thousands of books online, free
- Sign up to be notified when new books are published
- Purchase printed books
- Purchase PDFs
- Explore with our innovative research tools

Thank you for downloading this free PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to [comments@nap.edu](mailto:comments@nap.edu).

This free book plus thousands more books are available at <http://www.nap.edu>.

Copyright © National Academy of Sciences. Permission is granted for this material to be shared for noncommercial, educational purposes, provided that this notice appears on the reproduced materials, the Web address of the online, full authoritative version is retained, and copies are not altered. To disseminate otherwise or to republish requires written permission from the National Academies Press.

---

---

# The Unpredictable Certainty

**White Papers**  
**Information Infrastructure Through 2000**

NII 2000 Steering Committee  
Computer Science and Telecommunications Board  
Commission on Physical Sciences, Mathematics, and Applications  
National Research Council

National Academy Press

Washington, D.C. 1997

Copyright © National Academy of Sciences. All rights reserved.

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the steering committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

Support for this project was provided by the National Science Foundation under Grant No. IRI-9529473. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

International Standard Book Number 0-309-06036-2

Additional copies of this report are available from the Computer Science and Telecommunications Board, 2101 Constitution Avenue, N.W., Washington, D.C. 20418; [CSTB@NAS.EDU](mailto:CSTB@NAS.EDU) or <http://www2.nas.edu/cstbweb>.  
Copyright 1997 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

---

---

## NII 2000 Steering Committee

LEWIS M. BRANSCOMB, Harvard University, *Chair*  
CYNTHIA H. BRADDON, The McGraw-Hill Companies  
JAMES A. CHIDDIX, Time Warner Cable  
DAVID D. CLARK, Massachusetts Institute of Technology  
JOSEPH A. FLAHERTY, CBS Incorporated  
PAUL E. GREEN, JR., IBM T.J. Watson Research Center  
IRENE GREIF, Lotus Development Corporation  
RICHARD T. LIEBHABER, MCI Communications (retired)  
ROBERT W. LUCKY, Bell Communications Research  
LLOYD N. MORRISETT, John and Mary Markle Foundation  
DONALD W. SIMBORG, KnowMed Systems  
LESLIE L. VADASZ, Intel Corporation

### Staff

MARJORY S. BLUMENTHAL, Director

---

---

## Computer Science and Telecommunications Board

DAVID D. CLARK, Massachusetts Institute of Technology, *Chair*  
FRANCES E. ALLEN, IBM T.J. Watson Research Center  
JAMES CHIDDIX, Time Warner Cable  
JEFF DOZIER, University of California at Santa Barbara  
A.G. FRASER, AT&T Corporation  
SUSAN L. GRAHAM, University of California at Berkeley  
JAMES GRAY, Microsoft Corporation  
BARBARA J. GROSZ, Harvard University  
PATRICK HANRAHAN, Stanford University  
JUDITH HEMPEL, University of California at San Francisco  
DEBORAH A. JOSEPH, University of Wisconsin  
BUTLER W. LAMPSON, Microsoft Corporation  
EDWARD D. LAZOWSKA, University of Washington  
MICHAEL LESK, Bell Communications Research  
DAVID LIDDLE, Interval Research  
BARBARA H. LISKOV, Massachusetts Institute of Technology  
JOHN MAJOR, QUALCOMM Inc.  
DAVID G. MESSERSCHMITT, University of California at Berkeley  
DONALD NORMAN, Hewlett-Packard Company  
RAYMOND OZZIE, Rhythmix Corporation  
DONALD SIMBORG, KnowMed Systems Inc.  
LESLIE L. VADASZ, Intel Corporation  
MARJORY S. BLUMENTHAL, Director  
HERBERT S. LIN, Senior Staff Officer  
JERRY R. SHEEHAN, Program Officer  
ALAN S. INOUE, Program Officer  
JON EISENBERG, Program Officer  
JANET D. BRISCOE, Administrative Associate  
MARK BALKOVICH, Research Associate  
SYNOD P. BOYD, Project Assistant  
LISA L. SHUM, Project Assistant

---

---

## Commission on Physical Sciences, Mathematics, and Applications

ROBERT J. HERMANN, United Technologies Corporation, *Co-chair*  
W. CARL LINEBERGER, University of Colorado, *Co-chair*  
PETER M. BANKS, Environmental Research Institute of Michigan (ERIM)  
WILLIAM BROWDER, Princeton University  
LAWRENCE D. BROWN, University of Pennsylvania  
RONALD G. DOUGLAS, Texas A&M University  
JOHN E. ESTES, University of California at Santa Barbara  
MARTHA P. HAYNES, Cornell University  
L. LOUIS HEGEDUS, Elf Atochem North America, Inc.  
JOHN E. HOPCROFT, Cornell University  
CAROL M. JANTZEN, Westinghouse Savannah River Company  
PAUL G. KAMINSKI, Technovation, Inc.  
KENNETH H. KELLER, Council on Foreign Relations and the University of Minnesota  
KENNETH I. KELLERMANN, National Radio Astronomy Observatory  
MARGARET G. KIVELSON, University of California at Los Angeles  
DANIEL KLEPPNER, Massachusetts Institute of Technology  
JOHN KREICK, Sanders, A Lockheed Martin Company  
MARSHA I. LESTER, University of Pennsylvania  
NICHOLAS P. SAMIOS, Brookhaven National Laboratory  
CHANG-LIN TIEN, University of California at Berkeley  
NORMAN METZGER, Executive Director

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

## Preface

This book contains a key component of the NII 2000 project of the Computer Science and Telecommunications Board, a set of white papers that contributed to and complements the project's final report, *The Unpredictable Certainty: Information Infrastructure Through 2000*, which was published in the spring of 1996. That report was disseminated widely and was well received by its sponsors and a variety of audiences in government, industry, and academia. Constraints on staff time and availability delayed the publication of these white papers, which offer details on a number of issues and positions relating to the deployment of information infrastructure. The remainder of this preface is taken from the original preface of *The Unpredictable Certainty*. It provides more detail on the context in which the white papers were developed.

In October 1994, at the request of the Technology Policy Working Group (TPWG) of the Information Infrastructure Task Force, CSTB convened a steering committee to assess medium-term deployment of facilities and services to advance the nation's information infrastructure. The project was designated "NII 2000" by the steering committee, and its tasks were the following:

- To reach out to a broad range of industries with a stake in the future of U.S. information infrastructure—those industries expected to be major market drivers as well as those expected to be major service providers—to explore their expectations and motivations for technology deployment in the next 5 to 7 years;
- To infer from this exploration the extent to which there is a shared vision of the importance of common features of system architecture, such as interoperability or open system interfaces, and the alternative likelihood that major parts of the system will develop along proprietary, incompatible lines; and
- To conclude with suggestions to the U.S. government on public policy choices that might serve both the rapid, orderly, and successful development of information infrastructure and its satisfaction of important public interests.

To achieve these goals, the steering committee was asked by the TPWG to undertake a specific series of activities: convene a workshop of professionals and scholars to discuss and identify key issues related to technology deployment, call for white papers to gain further information on these issues, organize a forum to discuss the white papers and other key ideas, and write a synthesis report of its findings.

Following the workshop, the steering committee released a call for white papers on issues related to architecture and facilities, enabling technologies, recovery of costs, middleware technologies and capabilities, applications, equitable access and public service obligations, and research and development. The call was distributed through various media (the Internet, press advisories, direct mail, and so on) to producers of communications, computer, and software systems goods and services; Internet access and other network-based service providers; scholars specializing in relevant technical, economic, and public policy research and analysis; and project liaisons and other representatives of industries and sectors believed likely to become major users of advanced information infrastructure (such as the arts, banking and finance, education, health care, government agencies, libraries, manufacturing, and transportation). The white papers were



distributed to participants at the spring forum and to interested federal agencies. Their content, representing a broad spectrum of views from knowledgeable participants in the evolution of information infrastructure, was a major component in the development of the steering committee's report, which quotes from and refers specifically to several of them.

## Contents

The National Information Infrastructure and the Earth Sciences: Possibilities and Challenges <i>Mark R. Abbott (Oregon State University)</i>	1
Government Services Information Infrastructure Management <i>Robert J. Aiken and John S. Cavallini (U.S. Department of Energy)</i>	10
Cutting the Gordian Knot: Providing the American Public with Advanced Universal Access in a Fully Competitive Marketplace at the Lowest Possible Cost <i>Allan J. Arlow (Telecommunications Consultant, Annapolis, Md.)</i>	18
The Role of Cable Television in the NII <i>Wendell Bailey (National Cable Television Association) and Jim Chiddix (Time Warner Cable)</i>	26
Competing Definitions of "Openness" on the GII <i>Jonathan Band (Morrison and Foerster, Washington, D.C.)</i>	31
Communications for People on the Move: A Look into the Future <i>Richard C. Barth (Motorola Incorporated)</i>	38
Building the NII: Will the Shareholders Come? (And If They Don't, Will Anyone Really Care?) <i>Robert T. Blau (BellSouth Corporation)</i>	44
The Electronic Universe: Network Delivery of Data, Science, and Discovery <i>Gregory Bothun (University of Oregon), Jim Elias (US West Communications), Randolph G. Foldvik (US West Communications), and Oliver McBryan (University of Colorado)</i>	57
An SDTV Decoder with HDTV Capability: An All-Format ATV Decoder <i>Jill Boyce, John Henderson, and Larry Pearlstein (Hitachi America Ltd.)</i>	67
NII and Intelligent Transport Systems <i>Lewis M. Branscomb and Jim Keller (Harvard University)</i>	76
Post-NSFNET Statistics Collection <i>Hans-Werner Braun and Kimberly Claffy (San Diego Supercomputer Center)</i>	85
NII Road Map: Residential Broadband <i>Charles N. Brownstein (Cross-Industry Working Team, Corporation for National Research Initiatives)</i>	97
The NII in the Home: A Consumer Service <i>Vito Brugliera (Zenith Electronics), James A. Chiddix (Time Warner Cable), D. Joseph Donahue (Thomson Consumer Electronics), Joseph A. Flaherty (CBS Inc.), Richard R. Green (Cable Television Laboratories), James C. McKinney (ATSC), Richard E. Ottinger (PBS), and Rupert Stow (Rupert Stow Associates)</i>	101
Internetwork Infrastructure Requirements for Virtual Environments <i>Donald P. Brutzman, Michael R. Macedonia, and Michael J. Zyda (Naval Postgraduate School, Monterey, California)</i>	110
Electric Utilities and the NII: Issues and Opportunities <i>John S. Cavallini and Mary Anne Scott (U.S. Department of Energy) and Robert J. Aiken (U.S. Department of Energy/Lawrence Livermore National Laboratory)</i>	123
Interoperation, Open Interfaces, and Protocol Architecture <i>David D. Clark (Massachusetts Institute of Technology)</i>	133

Service Provider Interoperability and the National Information Infrastructure <i>Tim Clifford (DynCorp Advanced Technology Services)</i>	145
Funding the National Information Infrastructure: Advertising, Subscription, and Usage Charges <i>Robert W. Crandall (Brookings Institution)</i>	156
The NII in the Home <i>D. Joseph Donahue (Thomson Consumer Electronics)</i>	165
The Evolution of the Analog Set-Top Terminal to a Digital Interactive Home Communications Terminal <i>H. Allen Ecker and J. Graham Mobley (Scientific-Atlanta Inc.)</i>	168
Spread ALOHA Wireless Multiple Access: The Low-Cost Way for Ubiquitous, Tetherless Access to the Information Infrastructure <i>Dennis W. Elliott and Norman Abramson (ALOHA Networks Inc.)</i>	178
Plans for Ubiquitous Broadband Access to the National Information Infrastructure in the Ameritech Region <i>Joel S. Engel (Ameritech)</i>	185
How Do Traditional Legal, Commercial, Social, and Political Structures, When Confronted with a New Service, React and Interact? <i>Maria Farnon (Fletcher School of Law and Diplomacy, Tufts University)</i>	190
The Internet, the World Wide Web, and Open Information Services: How to Build the Global Information Infrastructure <i>Charles H. Ferguson (Vermeer Technologies Inc.)</i>	201
Organizing the Issues <i>Frances Dummer Fisher (University of Texas at Austin)</i>	205
The Argument for Universal Access to the Health Care Information Infrastructure: The Particular Needs of Rural Areas, the Poor, and the Underserved <i>Richard Friedman and Sean Thomas (University of Wisconsin)</i>	209
Toward a National Data Network: Architectural Issues and the Role of Government <i>David A. Garbin (MITRE Corporation)</i>	217
Statement on National Information Infrastructure Issues <i>Oscar Garcia (for the IEEE Computer Society)</i>	228
Proposal for an Evaluation of Health Care Applications on the NII <i>Joseph Gitlin (Johns Hopkins University)</i>	233
The Internet—A Model: Thoughts on the Five-Year Outlook <i>Ross Glatzer (Prodigy Services [retired])</i>	237
The Economics of Layered Networks <i>Jiong Gong and Padmanabhan Srinagesh (Bell Communications Research Inc.)</i>	241
The Fiber-Optic Challenge of Information Infrastructures <i>P.E. Green, Jr. (IBM T.J. Watson Research Center)</i>	248
Cable Television Technology Deployment <i>Richard R. Green (Cable Television Laboratories Inc.)</i>	256
Privacy, Access and Equity, Democracy, and Networked Interactive Media <i>Michael D. Greenbaum (Bell Atlantic) and David Ticoll (Alliance for Converging Technologies)</i>	271
As We May Work: An Approach Toward Collaboration on the NII <i>Marjorie Greene (First Washington Associates)</i>	280
The Use of the Social Security Number as the Basis for a National Citizen Identifier <i>W. Ed Hammond (Duke University Medical Center)</i>	286

Estimating the Costs of Telecommunications Regulation <i>Peter W. Huber (Manhattan Institute), Boban Mathew (Yale University), and John Thorne (Bell Atlantic)</i>	292
Residential PC Access: Issues with Bandwidth Availability <i>Kevin C. Kahn (Intel Corporation)</i>	304
The National Information Infrastructure: A High-Performance Computing and Communications Perspective <i>Randy H. Katz (University of California at Berkeley), William L. Scherlis (Carnegie Mellon University), and Stephen L. Squires (Advanced Research Projects Agency)</i>	315
Nomadic Computing and Communications <i>Leonard Kleinrock (University of California at Los Angeles)</i>	335
NII 2000: The Wireless Perspective <i>Mary Madigan (Personal Communications Industry Association)</i>	342
Small Manufacturing Enterprises and the National Information Infrastructure <i>Robert M. Mason, Chester Bowling, and Robert J. Niemi (Case Western Reserve University)</i>	351
Architecture for an Emergency Lane on the NII: Crisis Information Management <i>Lois Clark McCoy and Douglas Gillies (National Institute for Urban Search and Rescue) and John Harrald (NIUSR and George Washington University)</i>	364
Aspects of Integrity in the NII <i>John C. McDonald (MBX Inc.)</i>	374
What the NII Could Be: A User Perspective <i>David G. Messerschmitt (University of California at Berkeley)</i>	378
Role of the PC in Emerging Information Infrastructures <i>Avram Miller and Ogden Perry (Intel Corporation)</i>	388
NII Evolution—Technology Deployment Plans, Challenges, and Opportunities: AT&T Perspective <i>Mahal Mohan (AT&amp;T Corporation)</i>	397
Enabling Petabyte Computing <i>Reagan W. Moore (San Diego Supercomputer Center)</i>	405
Private Investment and Federal National Information Infrastructure Policy <i>Organization for the Protection and Advancement of Small Telephone Companies (OPASTCO)</i>	412
Thoughts on Security and the NII <i>Tom Perrine (San Diego Supercomputer Center)</i>	416
Trends in Deployments of New Telecommunications Services by Local Exchange Carriers in Support of an Advanced National Information Infrastructure <i>Stewart D. Personick (Bell Communications Research Inc.)</i>	422
The Future NII/GII: Views of Interexchange Carriers <i>Robert S. Powers (MCI Telecommunications Inc.), Tim Clifford (SPRINT, Government Systems Division), and James M. Smith (Competitive Telecommunications Association)</i>	434
Technology in the Local Network <i>J.C. Redmond, C.D. Decker, and W.G. Griffin (GTE Laboratories Inc.)</i>	447
Recognizing What the NII Is, What It Needs, and How to Get It <i>Robert F. Roche (Cellular Telecommunications Industry Association)</i>	462

Electronic Integrated Product Development as Enabled by a Global Information Environment: A Requirement for Success in the Twenty-first Century <i>Thomas C. Rochow, George E. Scarborough, and Frank David Utterback (McDonnell Douglas Corporation)</i>	469
Interoperability, Standards, and Security: Will the NII Be Based on Market Principles? <i>Quincy Rodgers (General Instrument Corporation)</i>	479
Technology and Cost Models for Connecting K-12 Schools to the National Information Infrastructure <i>Russell I. Rothstein and Lee McKnight (Massachusetts Institute of Technology)</i>	492
Geodata Interoperability: A Key NII Requirement <i>David Schell, Lance McKee, and Kurt Buehler (Open GIS Consortium)</i>	511
Electronic Commerce <i>Dan Schutzer (Citibank Corporation)</i>	521
Prospects and Prerequisites for Local Telecommunications Competition: Public Policy Issues for the NII <i>Gail Garfield Schwartz and Paul E. Cain (Teleport Communications Group)</i>	538
The Awakening 3.0: PCs, TSBs, or DTMF-TV—Which Telecomputer Architecture Is Right for the Next Generation's Public Network? <i>John W. Thompson, Jr. (GNOSTECH Incorporated)</i>	546
Effective Information Transfer for Health Care: Quality versus Quantity <i>Gio Wiederhold (Stanford University)</i>	553
Integrating Technology with Practice: A Technology-enhanced, Field-based Teacher Preparation Program <i>Ronald D. Zellner, Jon Denton, and Luana Zellner (Texas A&amp;M University)</i>	560
RegNet: An NPR Regulatory Reform Initiative Toward NII/GII Collaboratories <i>John P. Ziebarth (National Center for Supercomputing Applications), W. Neil Thompson (U.S. Nuclear Regulatory Commission), J.D. Nyhart, Kenneth Kaplan (Massachusetts Institute of Technology), Bill Ribarsky (Georgia Institute of Technology), Gio Wiederhold, Michael R. Genesereth (Stanford University), Kenneth Gilpatric (National Performance Review NetResults.RegNet and Administrative Conference of the United States [formerly]), Tim E. Roxey (National Performance Review RegNet. Industry, Baltimore Gas and Electric, and Council for Excellence in Government), William J. Olmstead (U.S. Nuclear Regulatory Commission), Ben Slone (Finite Matters Ltd.), Jim Acklin (Regulatory Information Alliance)</i>	576
Electronic Document Interchange and Distribution Based on the Portable Document Format, an Open Interchange Format <i>Stephen N. Zilles and Richard Cohn (Adobe Systems Incorporated)</i>	605

---

---

## **White Papers The Unpredictable Certainty Information Infrastructure Through 2000**

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

# 1

## The National Information Infrastructure and the Earth Sciences: Possibilities and Challenges

*Mark R. Abbott*

*Oregon State University*

As with most areas of science, computer networks (especially wide area networks) have become indispensable components of the infrastructure required to conduct environmental research. From transport of data to support for collaboration, to access to remote information processing resources, the network (which I will use as synonymous with the Internet) provides many essential services. In this paper I discuss the various technical challenges in the use of information networks for the Earth sciences. However, the technical issues, though important, are not the essential point. The network is, after all, only a collection of wires, switches, routers, and other pieces of hardware and software. The most serious issue is the content carried across these networks and how it engenders changes in the way Earth scientists relate to data, to each other, and to the public at large. These changes have impacts that are far more profound than access to bandwidth or new network protocols.

Most of the discussions of the national information infrastructure (NII) have focused on technical details (such as protocols) and implementation (e.g., provision of universal access), with little discussion of the impacts of an NII on the scientific process. Instead, discussions of the interactions between technology and human activities focus almost exclusively on the positive aspects of networks and social interactions. For example, networks have been extolled as tools for an expanding sense of community and participatory democracy. However, technology does not have only positive effects; the impacts are instead far more subtle and often more extensive than they first appear. They may not appear for decades. In this paper I neither extol nor condemn the impacts of computer networks on the conduct of science. Rather, it is essential that we become aware of these impacts, both positive and negative. I show that networks do far more than simply move bits; they fundamentally alter the way we think about science.

### EARTH SCIENCE AND NETWORKS

#### Data and Earth Science

Before exploring the role of networks in Earth science, I first briefly discuss the role of data in environmental research. As my background is in oceanography, my comments focus on the ocean sciences, but these observations are generally applicable to the Earth sciences.

Unlike experimental sciences such as chemistry or physics, most Earth sciences cannot conduct controlled experiments to test hypotheses. In some cases, though, manipulations of limited areas such as lakes or small forest plots can be done. Other branches of science that depend heavily on observations, such as astronomy, can observe many independent samples to draw general conclusions. For example, astronomers can measure the properties of dozens of blue dwarf stars. However, Earth science (particularly those fields that focus on large-scale or global-scale processes such as oceanography) must rely on many observations collected under a variety of conditions to develop ideas and models of broad applicability. There is only one Earth.



Earth science thus is driven strongly by developments in observing technology. For example, the availability of satellite remote sensing has transformed our view of upper ocean biology. The spring bloom in the North Atlantic, the sudden "flowering" of phytoplankton (the microscopic plants of the ocean) that occurs over a period of a few weeks, was thought to be primarily a local phenomenon. However, satellite imagery of ocean color (which is used to infer phytoplankton abundance) has shown that this event covers the entire North Atlantic over a period of a few weeks. Here, a new observing technique provided an improved understanding of what was thought to be a well-known process.

There are instances where new observing systems have transformed our understanding of the Earth. For over 40 years, the State of California has supported regular sampling of its coastal waters to understand the relationship between ocean circulation and fisheries production. A sampling grid was designed based on our understanding of ocean processes at the time. When satellite images of sea surface temperature (SST) and phytoplankton abundance became available in the early 1980s, they revealed a complex system of "filaments" that were oriented perpendicular to the coast and sometimes extended several hundred miles offshore. Further studies showed that these filaments are the dominant feature of circulation and productivity of the California Current, yet they were never detected in the 40-year record. The original sampling grid had been too widely spaced. This example shows as can sometimes lead us to design observing systems that miss critical processes.

The interaction between ideas and observations occasionally results in more subtle failures, which may be further obscured by computing systems. A notable example occurred during the 1982–1983 El Niño/Southern Oscillation (ENSO) event. ENSO events are characterized by a weakening of the trade winds in the tropical Pacific, which results in a warming of the eastern Pacific Ocean. This shift in ocean circulation has dramatic impacts on atmospheric circulation, such as severe droughts in Australia and the Pacific Northwest and floods in western South America and southern California. The 1982–1983 ENSO was the most dramatic event of this century, with ocean temperatures 5°–6°F warmer than normal off southern California. This physical event strongly influenced ocean biology as well. Lower than normal salmon runs in the Pacific Northwest are associated with this major shift in ocean circulation.

The National Oceanic and Atmospheric Administration (NOAA) produces regular maps of SST based on satellite, buoy, and ship observations. These SST maps can be used to detect ENSO warming events. Because of the enormous volume of satellite data, procedures to produce SST maps were automated. When SST values produced by the satellites were higher than a fixed amount above the long-term average SST for a region, the computer processing system would ignore them and would use the long-term average value instead (i.e., the processing system assumed that the satellite measurements were in error). As there was no human intervention in this automated system, the SST fields continued to show "normal" SST values in the eastern tropical Pacific in 1982. However, when a NOAA ship went to the area in late 1982 on a routine cruise, the ocean was found to be significantly warmer than had ever been observed. An alarm was raised, and the satellite data were reprocessed with a revised error detection algorithm. The enormous rise in SST over much of the eastern Pacific was revealed. The largest ENSO event of the century had been hidden for several months while it was confidently predicted that there would be no ENSO in 1982.

This episode reveals that the relationship between data and ideas has become more complex with the arrival of computers. The increasing volume and complexity of the data available for Earth science research have forced us to rely more heavily on automated procedures. Although this capability allows us to cope with the volume, it also relies on precise specification of various filters and models that we use to sort data in the computer. These filters may reflect our preconceived notions about what the data should actually look like. Although computers and networks apparently place more data into our hands more rapidly, the paradox is that there is increasing distance between the scientist and the actual physical process. This "hands-off" approach can lead to significant failures in the overall observing system.

As noted by Theodor Roszak<sup>1</sup>, raw data are of little value without an underlying framework. That is, ideas come before data. There must be a context for observations before they can make sense. A simple stream of temperature readings will not advance science unless their context is defined. Part of this framework includes the ability to repeat the measurements or experiment. Such repeatability strengthens the claim of the scientist that the process under study is a general phenomenon with broad applicability. This framework also includes a

historical context as well. Because of the strong observational nature of Earth science, it depends on earlier field programs to develop new theories and understanding.

With the availability of networks and computers, the task of obtaining, managing, and distributing data has become critical. The information system has become part of the observing system. The way we collect, store, and retrieve our data becomes another set of filters, just like the sampling strategy or measurement technique. Often technical problems obscure the underlying science. That is, the information science issues can dominate the Earth science issues. I explore these technical issues in the next section.

### Earth Science and Information Systems

The network and computational requirements for Earth science focus on the more obvious problems of bandwidth, accessibility, ease of use, and so on. I argue that although these issues are important, the profound shift in networks and computational systems has exacerbated the fundamental conflicts between information systems and the conduct of science while simultaneously obscuring these conflicts. Analogous to the analysis of television by Mark Crispin Miller<sup>2</sup>, information technology has both revealed these problems and hidden them within the medium itself.

### Technical Requirements

Earth science relies heavily on close interactions between many researchers from many disciplines. A single scientist cannot reserve an entire oceanographic research vessel for a cruise. Such expeditions require the work of many scientists. The study of problems such as the impacts of climate change on ocean productivity require an understanding of the physical dynamics of both the atmosphere and ocean, besides knowledge of ocean biology. Earth scientists must therefore develop effective mechanisms for sharing data.

Along with the need to share data and expertise among widely dispersed investigators, the characteristics of the data sets impose their own requirements. As Earth science moves toward an integrated, global perspective, the volumes of data have increased substantially. Although the dominant data sources continue to be Earth-observing satellites, field observations have also grown significantly. Sensors can now be deployed for longer time periods and can sample more rapidly. The number of variables that can be measured has increased as well. A decade ago, a researcher would come back from a cruise with a few kilobytes of data; today, a researcher will return with tens of gigabytes. However, these numbers are dwarfed by the data volumes collected by satellites or produced by numerical models. The Earth Observing System (EOS), which is planned by the National Aeronautics and Space Administration (NASA), will return over 1 terabyte per day of raw data.

The demands for near-real-time access to data have also appeared. Satellite-based communications to remote sampling sites have opened up the possibilities of having rapid access to observations, rather than waiting several months to recover environmental sensors. With more capable database servers, data can be loaded and made accessible over the network in hours to days after collection. This is in contrast to an earlier era when data were closely held by an individual investigator for months or even years. Although real-time access does open new areas for environmental monitoring and prediction, it does not necessarily address the need to accumulate the long, consistent, high-quality time series that are necessary for climate research. The pressures to meet everyday demands for data often distract scientists from the slower retrospective analyses of climate research.

As data become more accessible faster, public interest increases. As with the comet impact on Jupiter in 1994, public interest in science and the environment can often far exceed the anticipated demand. The EOS Data and Information System (EOSDIS) was originally designed to meet the needs of several thousand Earth science researchers. Now it is being tasked with meeting the undefined needs of the much larger general public<sup>3</sup>. This presents many technical challenges to an agency that has little experience dealing with a potentially enormous number of inexperienced users.

## Technical Challenges

Against this backdrop of new technical requirements, Earth science is facing a new set of technical challenges in addition to the continuing challenge of network bandwidth. The structure of the Internet is undergoing massive changes. With the departure of National Science Foundation funding for the network backbone and for the regional providers as well in the near future, there will be increasing emphasis on commercial customers. Interoperability and functionality of the network access points remain problematic. The possibility of balkanization of the network is real and not insignificant. The number of users has also expanded rapidly, and with the appearance of new applications such as the World Wide Web (WWW), the effective bandwidth has dropped dramatically.

As the science community becomes a smaller and smaller fraction of the network user community, network providers focus less on meeting scientific needs and more on meeting commercial needs. The search for bandwidth has become more intense. Telecommunication companies claim that new protocols (such as asynchronous transfer mode [ATM]) and new hardware (fiber optics) will usher in a new era of unlimited bandwidth. Some researchers claim that software "agents" will reduce the need for bandwidth by relying on intelligence at each node to eliminate the need for bulk transfers of data. Besides the technical hurdles to bringing such technologies to market, there are economic forces that work against such developments. In a process that is well known to freeway designers and transportation planners, bandwidth is always used to capacity when the direct costs of the bandwidth are not borne by the users. Funding for the network is hidden from the user so that any increase in personal use of network capacity is spread over every user. In reality, bandwidth is not free, though it is essentially free to the individual user. Even in the case of a totally commercial network, it is likely that the actual costs will be amortized over a broad customer base so that an individual user will have little incentive to use bandwidth efficiently. Although I pay a certain amount directly for the interstate highway system through my gasoline bills, the actual costs are hidden in many other fees and spread over a broad range of the population, many of whom may never even use the highway system.

With the rise of commercial Internet providers such as America Online and NetCom, will this situation change? Will users pay the true costs of using the network as opposed to paying only a marginal cost? I would argue that this is unlikely on several grounds. First, many users currently have access to virtually free Internet services through universities and other public institutions; it will be difficult to break this habit. Second, the government, through its emphasis on universal access, is unlikely to completely deregulate the system so that rural users (who truly are more expensive to service) will pay significantly more than urban users. Third, and perhaps more compelling, network bandwidth is no different from any other commodity. Every second that the network is not used, revenue is lost. Thus it is in the network providers' interest to establish a pricing structure that ensures that at least some revenue is generated all the time, even if it is at a loss. Some revenue is always greater than no revenue, and the losses can be made up elsewhere in the system. This is a well-established practice in the long-distance telephone industry as well as in the airline industry. Off-peak prices are not lower to shift traffic from peak periods to less congested periods; they are designed to encourage usage during off-peak periods, not reduce congestion during peak periods. They raise the "valleys" rather than lowering the "peaks."

The final threat to network bandwidth is the proliferation of new services. There is no doubt that as network bandwidth increases, new services become available. Early networks were suitable for electronic mail and other low-bandwidth activities. This was followed by file transfers and remote logins. The WWW dramatically increased the capabilities for data location and transfer. Just as the interstate highway system fosters the development of new industries (such as fast food franchises and overnight package delivery systems), so also has the Internet. As with highways, new services create new demand for bandwidth. Although considerable effort is being devoted within the NII to develop rational pricing strategies, it is more likely that the search for bandwidth will continue. It appears to be a law of networks that spare bandwidth leads to frivolous traffic.

As services and users proliferate, it has become more difficult for users to locate the data of interest. Search tools are extremely primitive, especially when compared with the tools available in any well-run library. Various "web crawlers" will locate far more irrelevant information than relevant information. For a user who does not know exactly where to find a specific data source, much time will be spent chasing down dead ends, or circling back to the beginning of the search. Although the relative chaos of the network allows anyone to easily

set up a data server, this chaos confounds all but the most determined user. There are no standard methods for defining the contents of a server and without "truth in advertising" rules, servers, that may superficially appear to be relevant may contain totally irrelevant information. The only users who have time to "surf" the network are those who have nothing else to do, as noted by Negroponte<sup>4</sup>. Once the appropriate server has been located, there is no consistent method for indexing and archiving data. Although data may be online, it often still requires a phone call to locate and access the relevant data sets.

With the increase in computer processor speeds and desktop mass storage, it has become increasingly important to match network speeds with these other components of the system. Following Amdahl's Rule where 1 bit of input/output requires one instruction cycle, this implies that a 1,000-MIPS (million instructions per second) processor will require a 1-gigabit-per-second network interface. Assuming that processor performance continues at a rate of 50 to 100 percent per year, we will have 1,000-MIPS machines in about 1 to 2 years. It is becoming commonplace for researchers to have 10-to 20-gigabyte disk storage systems on their desktop; in 2 to 3 years it will not be unusual for scientists to have 100 gigabytes of disk subsystems. Network speeds are increasing at a far slower rate, and the present Internet appears to many users to be running slower than it did 3 to 4 years ago.

This imbalance between processor speed, disk storage, and network throughput has accelerated a trend that began many years ago: the decentralization of information. In an earlier era, a researcher might copy only small subsets of data to his or her local machine because of limited local capacity. Most data were archived in central facilities. Now it is more efficient for the individual scientist to develop private "libraries" of data, even if they are rarely used. Although the decentralization process has some advantages, it does increase the "confusion" level. Where do other researchers go to acquire the most accurate version of the data? How do researchers maintain consistent data sets? On top of these issues, the benefits of the rapid decrease in price/performance are more easily acquired by these small facilities at the "fringes" of the network. Large, central facilities follow mainframe price/performance curves, and they are generally constrained by strict bureaucratic rules for operation and procurement. They are also chartered to meet the needs of every user and usually focus on long-term maintenance of data sets. In contrast, private holdings can evolve more rapidly and are not required to service every user or maintain every data set. They focus on their own particular research needs and interests, not on the needs of long-term data continuity or access by the broader community.

Since small, distributed systems appear to be more efficient and provide more "niche" services than do central systems, it has become harder to fund such centers adequately (thus further reducing their services.) This situation is similar to the "Commons" issue described by Hardin nearly 30 years ago<sup>5</sup>. For the centralized archive, each individual user realizes great personal benefit by establishing a private library while the cost is spread evenly over all of the users. It is therefore in the interest of the individual user to maximize individual benefit at the cost of the common facility. Not until the common facility collapses do we realize the total cost. The situation is similar in the area of network bandwidth.

The final two technical challenges facing Earth science and networks encroach into the social arena as well. The first is the area of copyrights and intellectual property. Although scientists are generally not paid for the use of their data sets or their algorithms, there is a strong set of unwritten rules governing use and acknowledgment of other scientists' data sets. With the network and its rapid dissemination of information and more emphasis on the development of Web browsers, this set of rules is being challenged. Data are being moved rapidly from system to system with little thought of asking permission or making an acknowledgment. Copying machines have created a similar problem, and the publishing industry has reacted vigorously. However, the rate at which a copying machine can replicate information is far slower than that for a computer network, and its reach is far smaller. As new data servers appear on the network, data are rapidly extracted and copied into the new systems. The user has no idea of what the original source of the data is nor any information concerning the data's integrity and quality.

Discussions over copyrights have become increasingly heated in the past few years, but the fundamental issues are quite simple. First, how do I as a scientist receive "payment" for my contributions? In this case, payment can be as simple as an acknowledgment. Second, how do I ensure that my contributions are not used incorrectly? For example, will my work be used out of context to support an argument that is false? With global, nearly instantaneous dissemination of information, it is difficult to prevent either nonpayment or improper use. After a few bad experiences, what incentive is there for a scientist to provide unfettered access?

The last technical challenge involves the allocation of resources. Network management is not for amateurs. The Internet requires experts for its operation. Web browsers are not easy to design, build, and maintain. Thus programming talent that was originally hired to engage in scientific analysis is spending a larger fraction of its time engaged in nonscientific activities. Although developments in the commercial field are simplifying these and associated tasks, they do cost money. With declining resources for science, one can ask whether this is a reasonable use of scarce resources. The pace of technological change is also increasing, so that the intellectual and technical infrastructure that was assembled 5 to 10 years ago is largely irrelevant. For example, it is becoming harder to hire FORTRAN programmers. Experts on ATM have not yet appeared.

The computer industry and Earth science researchers have generally focused on these challenges from a technological point of view. That is, bandwidth will increase through the provision of ATM services. Data location will be enhanced through new WWW services. Copyrights (if they are not considered absolutely irrelevant or even malevolent) can be preserved through special identification tags.

The technology optimists have decided that the problems of Earth science (and science in general) can be solved through the appropriate application of technology. That is, the fundamental problems of science will be addressed by "better, faster, cheaper" technical tools. On the science side of the issue, it appears that the scientific community has largely accepted this argument. As information systems approach commodity pricing, scientists acquire the new technology more rapidly in an attempt to remain competitive in an era of declining federal support. As put forth by Neil Postman<sup>6</sup>, this is the argument of "technology." That is, the technology has become an end in itself. The fundamental problem of science is understanding, not the more rapid movement of data. Although we have seen that the link between understanding and observations is perhaps more closely entwined in the Earth sciences, we must be aware of the implications of our information systems for how we conduct science. It is to this point that I now turn.

### THE HIDDEN IMPACTS

A recent book by Clifford Stoll<sup>7</sup> provided a critical examination of the information infrastructure and where we are headed as a society. Although he makes many valid points, Stoll does not provide an analysis as to how we arrived at this predicament. I draw on analyses of the media industries (especially television) by Postman and Miller that show some parallels between information systems and mass media. I do not argue that we should return to some pristine, pre-computer era. There is no doubt about the many positive aspects of information technology in the Earth sciences. However, it is worth examining all of its impacts, not just the positive ones.

### Information Regulation

Postman postulates that culture can be described as a set of "information control" processes<sup>8</sup>. That is, we have established mechanisms to separate important knowledge from unimportant knowledge (such as a school curriculum) and the sacred from the profane (such as religion). Even in the world of art and literature, we are constantly making judgments as to the value of a particular work of art. The judgments may change over time, but the process remains.

We are constantly inundated with information about our world, either from the natural world or by our creations. Somehow we must develop systems to winnow this information down to its essential elements. In the scientific world, we have established an elaborate set of rules and editorial procedures, most of which are not written down. For example, experiments that cannot be repeated are viewed as less valuable than those that can be. Ideas that cannot be traced to previous studies are viewed with more skepticism than those that can be traced through earlier research. Occasionally, a new idea will spring forth, but it, too, must go through a series of tests to be evaluated by the scientific community. This "editorial" process essentially adds value to the information.

The second point is that the scientific community believes that there is an objective reality that is amenable to observation and testing. Although clearly science is littered with ideas that rested on an individual's biases and processes that were missed because they did not fit our preconceived notions, we still believe that the

natural world can be measured and understood in a predictable manner. Assumptions and biases can at least be described and perhaps quantified. Scientific knowledge is more than a set of opinions.

Through the scientific process, researchers add value to data. Useful data are separated from useless data and interpreted within a framework of ideas. Data are therefore placed in a structure that in turn can be described. Raw information that is presented out of context, without any sense of its historical origins, is of little use. Thus Earth science is not limited by lack of observations; rather, the correct types of observations are often missing (e.g., long, calibrated time series of temperature).

The process of adding value has arisen over the last several centuries. Publishing in peer-reviewed journals is but one example of how valid data are separated from invalid data, and how observations are placed within a larger framework for interpretation. Although data reports are often published, they are perceived to be of less importance than journal articles. Even "data" produced by numerical models (which are the products of our assumptions) are viewed as less valuable than direct measurements.

### The Network and Science

There is no doubt that networks simplify many tasks for Earth science. However, there are many obvious problems, such as the separation of information from its underlying context, the difficulty in locating information of interest, and the lack of responsibility for the quality and value of a particular data set. Much as with television, it has become difficult to separate advertising from reality on the network. A recent discussion on the future of research universities in *Physics Today*<sup>9</sup> highlighted some troublesome issues associated with networks. Graduate students have become increasingly unwilling to use the library. If reference materials are not available online in digital form, then students deem them to be irrelevant. Although electronic searches can be helpful, it is clear that this attitude is misguided. Most scientific documents will never be placed online because of the associated costs. Second, digital storage is highly ephemeral and can never be considered a true archive. There will always be a machine and software between the reader and the data, and these tools are always becoming obsolete. Third, digital searching techniques follow rigid rules. Present search methods are quite sparse and limited. Thus far, no one has shown the ability to find material that truly matches what the reader wants although the search was incorrectly specified. Such serendipitous discoveries are common in libraries.

The network is becoming a place of advertising, with little true content and little personal responsibility. Home pages are proliferating that merely test what the network is capable of doing instead of using the network to accomplish a useful task. We have fixated on the technology, on the delivery system for information, rather than on the "understanding system"<sup>10</sup>. Speed, volume, and other aspects of the technology have become the goals of the system. Although these are useful, they do not necessarily solve the problems of collaboration, data access, and so on. In fact, they can distract us from the real problems. The issue of scientific collaboration may require changes in the promotion and tenure process that are far more difficult than a new software widget.

The emphasis on speed arises in part from the need to have short "return on investment." In such an environment, market forces work well in the development of flexible, responsive systems. For Earth science, this is a useful ability for some aspects of research. For example, development of new processing algorithms for satellite sensors clearly benefits in such an environment. However, this short-term focus is not sufficient. Long-term climate analysis, where the data must be collected for decades (if not centuries) and careful attention must be paid to calibration, will not show any return on investment in the short run. These activities will "lose" money for decades before one begins to see a return on the investment. In a sense, long time series are "common facilities," much like network bandwidth and central archives. They are the infrastructure of the science.

### The Network and Television

The early days of television were filled with predictions about increased access to cultural activities, improved "distance" learning, and increased understanding between people. The global village was predicted to be just around the corner. However, the reality is quite different. Television may fill our screens with

information, but the nature of the medium has not increased our understanding<sup>11</sup>. Science programming is useful as a public relations tool (and such contacts with the public should be encouraged), but the medium is not suitable for scientific research.

As discussed above, one of the key activities of science is the structuring of information to sort valid from invalid. But computer networks increasingly encourage disjointed, context-free searches for information. Our metaphors resemble those of television and advertising, rather than those of science and literature. Home pages encourage rapid browsing through captivating graphics; long pages of text are viewed as a hindrance.

In the early days of television, the message was in front of the viewer, as discussed by Miller<sup>12</sup>. It was clear what product was being advertised. Today, both the medium and its aesthetics are in front. The smooth graphics, the rapidly shifting imagery, the compelling soundtrack are the primary elements. The advertising and the product are in the background. Feelings about products rather than rational evaluations are transmitted to the viewer. This sense of the aesthetic has permeated other media, including film, and to some extent, newspapers. Presentation is more important than content.

The early days of computers were characterized by cumbersome, isolated, intimidating machines that were safely locked away in glass rooms. The computer was viewed as a tool whose role was clearly understood. Today, computers are ubiquitous. Most Earth scientists have at least one if not two computers in their offices, plus a computer at home. This demystification of computers has been accompanied by much emphasis on the interface. Computers are now friendly, not intimidating. Their design now focuses on smoothness, exuding an air of control and calm. As described in a biography of Steve Jobs<sup>13</sup>, design of the NeXT workstation focused almost exclusively on the appearance of the machine. Most of the advances in software technology have focused on aesthetics, not on doing new tasks. These new software tools require more technical capability (graphics, memory, and so on) to support their sophisticated form. This emphasis on form (both hardware and software) violates the Bauhaus principle of form following function. The computer industry appears to focus more on selling a concept of information processing as opposed to selling a tool.

Postman<sup>14</sup> has described print as emphasizing logic, sequence, history, and objectivity. In contrast, television emphasizes imagery, narrative, presentation, and quick response. The question is, Where do computer networks sit with regard to print versus television? There is no doubt that networks are beginning to resemble television more than print. The process of surfing and grazing on information as though it were just another commodity reduces the need for thoughtful argument and analysis. Networks encourage the exchange of attitudes, not ideas. The vast proliferation of data has become a veritable glut. Unlike television, anyone can be a broadcaster; but to rise above the background noise, one must advertise in a more compelling manner than one's competitors.

## CONCLUSIONS AND RECOMMENDATIONS

Networks will continue to play an important role in the conduct of Earth science. Their fundamental roles of data transport and access cannot be denied. However, there are other effects as well that are the result of a confluence of several streams. First, the next-generation networks will be driven by commercial needs, not by the needs of the research and education community. Second, the sharp decrease in price/performance of most computer hardware and the shortened product life cycles have required the science community to acquire new equipment at a more rapid pace. Third, expected decreases in federal funding for science have resulted in greater emphasis on competitiveness. This confluence has caused scientists to aim for rapid delivery of information over the network. Without the regulations and impedance of traditional paper publishing, scientists can now argue in near real time about the meaning of particular results. "Flame" wars over electronic mail are not far behind. The community now spends nearly as much time arguing about the technical aspects of the information delivery system as it does in carrying out scientific research.

Networks allow us to pull information out of context without consideration of the framework used to collect and interpret the data. Ever-increasing network speeds emphasize the delivery of volume before content. If all information is equally accessible and of apparently equal value, then all information is trivial. Science is at risk of becoming another "consumable" on the network where advertising and popularity are the main virtues.

Long, thoughtful analyses and small, unpopular data sets are often overwhelmed in such a system. Similar processes are at work in television; the metaphors of the TV world are rapidly appearing in the network world.

One can successfully argue that Earth science is currently limited by the lack of data (or at least the correct data), but an equally serious problem is the inability to synthesize large, complex data sets. This is a problem without a technological solution. While information systems can help, they will not overcome this hurdle. Delivering more data at a faster rate to the scientist will obscure this fundamental problem. Indeed, technology may give the appearance of solving the problem when in reality it exacerbates it. As stated by Jacob Bronowski,

This is the paradox of imagination in science, that it has for its aim the impoverishment of imagination. By that outrageous phrase, I mean that the highest flight of scientific imagination is to weed out the proliferation of new ideas. In science, the grand view is a miserly view, and a rich model of the universe is one which is as poor as possible in hypotheses.

Networks are useful. But as scientists, we must be aware of the fundamental changes that networks bring to the scientific process. If our students rely only on networks to locate data as opposed to making real-world observations, if they cannot use a library to search for historical information, if they are not accountable for information that appears on the network, if they cannot form reasoned, logical arguments, then we have done them a great disservice.

The balance between market forces with their emphasis on short-term returns for individuals and infrastructure forces with their emphasis on long-term returns for the common good must be maintained. There is a role for both the private sector and the public sector in this balance. At present, the balance appears to be tilted toward the short term, and somehow we must restore a dynamic equilibrium.

## NOTES

- [1] Roszak, Theodore. 1994. *The Cult of Information: A Neo-Luddite Treatise on High-Tech, Artificial Intelligence, and the True Art of Thinking*. University of California Press.
- [2] Miller, Mark Crispin. 1988. *Boxed In: The Culture of TV*. Northwestern University Press.
- [3] U.S. Government Accounting Office. 1995. "Earth Observing System: Concentration on Near-term EOSDIS Development May Jeopardize Long-term Success," Testimony before the House Subcommittee on Space and Aeronautics, March 16.
- [4] Negroponte, Nicholas. 1995. "000 000 111—Double Agents," *Wired*, March.
- [5] Hardin, Garrett. 1968. "The Tragedy of the Commons," *Science* 162:1243–1248.
- [6] Postman, Neil. 1992. *Technopoly: The Surrender of Culture to Technology*. Knopf, New York.
- [7] Stoll, Clifford. 1995. *Silicon Snake Oil: Second Thoughts on the Information Superhighway*. Doubleday, New York.
- [8] Postman, *Technopoly*, 1992.
- [9] Physics Today. 1995. "Roundtable: Whither Now Our Research Universities?" March, pp. 42–52.
- [10] Roszak, *The Cult of Information*, 1994.
- [11] Postman, *Technopoly*, 1992.
- [12] Miller, *Boxed In*, 1988.
- [13] Stross, Randall. 1993. *Steve Jobs and the NeXT Big Thing*. Atheneum, New York.
- [14] Postman, *Technopoly*, 1992.



2

## Government Services Information Infrastructure Management

*Robert J. Aiken and John S. Cavallini*  
*U.S. Department of Energy*

### ABSTRACT

The growth and deployment of the Government Services Information Infrastructure (GSII), and its relationship with the national information infrastructure (NII), will require the federal government to drastically change the way it acquires and deploys telecommunications for support of the government's businesses. The GSII is an enlightened attempt by the Clinton/Gore Administration to form a "virtual government" crossing agency boundaries to interact more closely with industry and with the public, as well as with state and local government, to greatly improve the delivery of government services. The GSII is that portion of the NII used to link government and its services, enable virtual agency concepts, protect privacy, and support emergency preparedness needs. The GSII will have to be an integral component of the NII in order to do so. The GSII and other private sector efforts will have a significant impact on the design, development, and deployment of the NII, even if only through the procurement of such services.

This paper concludes that the federal government must adopt new mechanisms and new paradigms for the management of the GSII, including improved acquisition and operation of GSII components in order to maximize taxpayer benefits, to optimize the delivery of government services, to ensure that the GSII is an integral component of the NII, and to adopt industry standards rather than setting them. Government requirements and applications, as well as the available technologies to address those requirements, will continue to evolve; therefore so must the government's use of technologies and services. The requirements from federal government services and the users of these services logically form affinity groups that more accurately and effectively define these common requirements, that drive the adoption and use of industry standards, and that provide a significant technology marketplace to capture the vision of the NII both now and in the future.

It is critically important that the federal government adopt a management scheme that improves its ability to work with U.S. industry to ride the coming Third Wave,<sup>1</sup> as opposed to being wiped out by it. A new management scheme is also needed to improve cooperation between the government and its partners (i.e., the private sector, academia, and state and local governments) and to improve the chances for a successful deployment of both the GSII and NII. This new management scheme must be built upon a modular and evolutionary approach for the GSII as well as for other large systems developments to greatly improve the successful use of information technology (IT) by the government, especially to provide service to the private sector and to the U.S. Citizenry.

---

NOTE: This paper was commissioned by the Government Information Technology Services (GITS) subcommittee of the Committee on Applications and Technology of the Information Infrastructure Task Force (IITF). It was drafted for the GITS by John S. Cavallini and Robert J. Aiken of the Department of Energy. It was reviewed and endorsed for submission to the NII 2000 Forum by the GITS members. Copyright © National Academy of Sciences. All rights reserved.

## BACKGROUND

Today, the private sector is developing and deploying information infrastructure. At the same time, the government is concurrently developing and deploying information infrastructure, mostly through contracts with the private sector. Under the lead of and at the encouragement of the Clinton/Gore Administration, the federal government has placed increased emphasis on the development and deployment of an NII as a strategic priority. This emphasis results from the understanding that properly leveraged information and information technology are among the nation's most critical economic resources, for manufacturing industries as well as for more modern services industries for economic security and for national security.

The Clinton/Gore Administration has made a commitment to work with business, labor, academia, public interest groups, Congress, and both state and local government to ensure the development of an NII that enables all Americans to access information and communicate with each other using combinations of voice, data, images, or video at anytime, anywhere.<sup>2</sup> This commitment was articulated very well by the National Performance Review (NPR) through its emphasis on using IT as a key element in creating a government that works better and costs less.<sup>3</sup> The President and Vice President recognize the need to use IT to improve Americans' quality of life and to reinvigorate the economy. To this end, they outlined a three-part agenda for spreading IT's benefits to the federal government: (1) strengthen leadership in IT, (2) implement electronic government, and (3) establish support mechanisms for electronic government. Thirteen major IT areas were identified for accomplishing the three-part agenda:

1. Provide clear, strong leadership to integrate IT into the business of government;
2. Implement nationwide, integrated electronic benefit transfer;
3. Develop integrated electronic access to government information and services;
4. Establish a national law enforcement/public safety network;
5. Provide intergovernmental tax filing, reporting, and payments processing;
6. Establish an international trade data system;
7. Create a national environmental data index;
8. Plan, demonstrate, and provide government-wide electronic mail;
9. Improve government's information infrastructure;
10. Develop systems and mechanisms to ensure privacy and security;
11. Improve methods of IT acquisition;
12. Provide incentives for innovation; and
13. Provide training and technical assistance in IT to federal employees.

Development of an NII is not an end goal in or of itself. Government requires an infrastructure to conduct its business more effectively and to deliver services to the American citizenry at lower cost to the taxpayers. A number of suitable national-scale applications or uses of the NII have been identified and documented by the IITF's Committee on Applications and Technology.<sup>4</sup> These uses of the NII, in addition to nationwide humanistic applications such as health care and education, include the fundamental businesses or enterprises of the federal government such as law enforcement, electronic commerce (including benefits), basic research, environment, health care, and national security, and as such represent a significant set of driving requirements for NII deployment.

In recognizing this fact, the NPR concluded that the government use of IT and development of information infrastructure should be improved and better coordinated in order to effectively address government business requirements. The NPR made approximately 60 recommendations for action in this regard, including the development of a plan for a GSII to electronically deliver government services and to integrate electronic access to government-provided information and services. The GSII is that portion of the NII used to link government and its services, enable virtual agency concepts, protect privacy, and support emergency preparedness needs. It was also recognized that better integration and coordination were required not only across federal government

agencies, but also across federal and state governments, and local and tribal governments, to the private sector and to the public at large.

To achieve these goals, the Vice President established the GITS working group in December 1993. The mission of the GITS working group is to promote the improvement of agency performance through the use of IT, accelerate the deployment of advanced networking technologies, and in conjunction with the Office of Management and Budget (OMB) and General Services Administration (GSA), establish procurement and implementation policies, practices, and directives designed to improve productivity and reduce costs. The working group's responsibilities include implementing the NPR's recommendations, coordinating federal agency and White House initiatives for evolving the NII, and enhancing cross-agency IT collaborations as well as government and industry IT activities.

### THE GSII: ANALYSIS

The GSII is that portion of the NII used to link government and its services, enable "virtual agency" concepts, protect privacy, and support emergency preparedness needs. The "virtual agency concept" allows services and functions to cross traditional organizational and geographic boundaries. For example, a U.S. citizen could, in this fashion, receive multiple services such as social security, food stamps, tax preparation assistance, information on the national park system, and so on through a single interaction with one agency.

The attributes of an effective GSII must include timely infusion of appropriate IT to meet requirements; coordination of infrastructure across government requirements and with the private sector; effectiveness, while also allowing for innovation and flexibility to meet specific needs; information surety to guard citizens' rights and privacy while also providing for national security needs; cost-effectiveness of IT acquisition and development while also encouraging a competitive technological environment; ubiquitous accessibility to all citizens, including those with disabilities; ease of use for locating government information or for service delivery; transparent (to the user) integration of the GSII portion of the NII with the rest of the Internet, NII, and the global information infrastructure; shared resource use across government entities; and the mechanisms, processes, and procedures to appropriately manage such a diverse infrastructure.

As opposed to the open system interconnect seven-layer model, information infrastructure is more often currently viewed as having three or four layers; a recent report of the National Research Council (*Realizing the Information Future*) proposes a standard open bearer service to promote, and perhaps maximize, interoperability across the NII.<sup>5</sup> This is a significant feature that the GSII must adopt and integrate as it evolves. Nevertheless, merely focusing on the NII layer concept and other technical models can lead one to neglect two of the most important components of the GSII and the NII. They are people, both using and providing services over the NII, and the coordination mechanisms needed to effectively manage all GSII activities and efforts.

Standards are often cited as the only way to achieve interoperability across the NII or GSII. However, standards can often be too much of a good thing.<sup>6</sup> The technology life cycle averages 12 to 18 months, with new hardware and software being introduced at a dizzying pace. The current standards processes have not been able to keep up. Competition among many vendors is crucial for reducing cost as well as ensuring a viable "technological gene pool" for both current and future requirements. Innovation requires new concepts. Flexibility is required to meet a wide and diverse set of special requirements and can enable the incremental evolution of the GSII and NII in a stepwise manner. Large systems development must be modular and better coordinated. Today's infrastructure has an enormous amount of different components composed of various technologies and standards. Therefore, a modular, seamless integration and evolution of the multicomponent GSII into the evolving NII will need to be based primarily on voluntary processes and proven interoperability solutions rather than on mandated standards. These and many more facts lead us to the conclusion that standards alone are not sufficient.

Furthermore, although the Internet protocol suite was originally developed as a military specification, it was not adopted as a federal procurement standard or an industry standard. Nevertheless, through versatility and capability to meet the research community's requirements, while supporting a productive range of interoperability, the Internet Transmission Control Protocol (TCP)/Internet Protocol (IP) became the most widely used

internetworking protocol. Therefore, common requirements and mutually beneficial, modular, evolutionary technical solution(s) will provide the best gauge of the range of interoperability that is ultimately needed.

Current federal coordination mechanisms are a result of the Brooks ADP Act of 1965 which was passed in an era characterized by mainframe computers when time-shared processing was the primary technique for managing scarce IT resources. Reflecting the centralized technology, the decisional and oversight responsibility for planning, acquiring, and managing IT was taken from the heads of the various executive agencies and vested in the administrator for general services. The theory was that investments in IT must be centrally managed and that only a dedicated cadre of IT professionals could perform the function.

Over the years, the original model eroded and technology became decentralized. Today, agencies plan and manage their own IT investments; the authority to contract for or to procure IT solutions is "delegated" to the agency heads. As a condition of the delegation, agencies are subject to an additional, centralized acquisition and management oversight. That oversight, however, is generally redundant in nature (i.e., oversight checkers outside an agency are checking internal oversight checkers) and neither set of checkers possesses the institutional expertise or the resources to adequately understand, let alone manage, the government's increasingly decentralized and diverse IT infrastructure. Ultimately, the centralized control model of the Brooks Act, which reduced both the responsibility and the authority of the heads of federal agencies, contributed to significant failures in IT management and seriously misspent resources.

The old centralized approach could achieve interoperability neither across the government infrastructure nor with the private sector, and it was even counterproductive in efforts to do so. In the future, an NII will undoubtedly incorporate multiple paradigms of internetworking, interoperability, and communications, thus making the task of coordination and interoperability even more difficult. Logically then, the question is how to promote interoperability. We propose starting with what the GSII will be used for—the business functions that can be considered GSII applications—and with those people who share common interests in both using and providing the GSII. The applications should determine what standards and technologies are required and will provide interoperability among their own constituency as well as with other groups, if properly coordinated.

It should be noted, however, that GSA has taken steps to try to improve the government's acquisition of IT and to recognize this new paradigm of distributed IT functionality and management. Two notable examples are the Time Out Program that tries to get faltering large-systems efforts back on track and the Government-wide Acquisitions Contracts Program that empowers lead agencies to conduct multiagency contracts.

### APPLICATIONS AND AFFINITY GROUPS

Government services and uses of the NII for law enforcement, benefits, and health care touch every community. Requirements for research, the environment, and national security go beyond the geographic requirements. The latter also require advances in the state-of-the-art, leading-edge IT capabilities. Hence, government applications provide significant financial leverage and incentive for NII deployment, as well as for new IT development. If one adds applications for education, generic electronic commerce, and energy delivery and management to this set of federal NII applications, the cost leveraging for deploying the NII becomes significant. For example, in energy, "[e]lectric utilities already serve over 95 percent of American homes (a percentage point above telephone companies)" with the "likely requirement that all these homes will need access to advanced telecommunications to manage energy" consumption.<sup>7</sup> It is unlikely that telecommunications for energy demand management will replace conventional access mechanisms, especially for the last mile to residences. However, the energy utilities telecommunications infrastructure can augment, leverage, and enhance the cable and telecommunications industry infrastructure and facilitate access to the NII, and hence the GSII, for almost all Americans, including those in remote rural areas.

In addition to the impact the energy utilities may have on the NII, FTS 2000 and its successor will have a more direct impact on both the NII and GSII, since it will be a major vehicle for supplying telecommunications and services to the federal government. Today's FTS 2000 "currently provides intercity telecommunications for 1.7 million federal government users."<sup>8</sup> Post-FTS 2000 is expected to go beyond the current FTS 2000 by delivering intercity telecommunications as well as value-added services such as electronic mail, key management,

teleconferencing, and more. This deployment will be an integral component of the GSII and will use industry-provided services almost exclusively, thereby leveraging the products of industry rather than competing with them. However, this reliance on and use of industry products and services will not lessen the impact that the federal government, via procurement vehicles, will have on the NII and GSII; therefore, the proper coordination of the deployment and management of the GSII with respect to the NII is even more critical.

The focus should be on achieving an effective, consultative, and collegial interagency process for managing the GSII. This will require creating a methodology for managing evolving IT policies, priorities, and programs that provides and coordinates the framework within which the mission, as well as the administrative, IT requirements, and activities of individual federal departments and agencies are conducted. The principal focus of the current top-level IT management is on what today are perceived as the common goals—mainly the administrative requirements for finance, personnel, facilities, and so on—of the overall enterprise called the federal government. Implementation of a new enterprise model, based on an understanding of GSII requirements driven by these other mission applications as well as those that represent government services to its customers and their non-government requirements, is essential to making the government work better and cost less.

The Federal Internetworking Requirements Panel (FIRP), created by the National Institute of Standards and Technology at the request of the Department of Energy's Office of Scientific Computing and the high-performance computing and communications community, issued a report that recommends increased responsibility for this shared infrastructure, such as a GSII, for the mission areas of federal agencies or their logical affinity groups in as compatible a way as practicable with the common vision of the federal government.<sup>9</sup> The term "affinity group" means government agencies, or functional interest groups therein, that share information electronically and have common IT requirements. The NPR emphasized the need for improved infrastructure for cross-agency groups. In addition, the FIRP report recommended the development of affinity groups to enhance cross-agency collaborations. Taken together, these factors make it reasonable to require that each agency explicitly ensure that GSII issues, including interoperability, are addressed not only within a given federal agency, but also with external affinity groups, industry, and the public, in which these agency mission or "enterprise" activities take place.

The federal research community is an excellent example of an affinity group. During the past decade, the federal research community has placed a very high priority on the application of advanced information technologies to enable advances in many science disciplines. As experimentation becomes too expensive, too unsafe, or too environmentally unsound, there is an increase in the importance and value of computational experiments, collaborative virtual reality, and distributed computing infrastructure technologies for remote access to one-of-a-kind facilities, shared data, and dispersed collaborations. Correspondingly, a sound and capable NII is needed, since the research community crosses many organizational boundaries, large geographical distances, and multiple capability needs. By virtue of its use of advanced capabilities, the research community or affinity group has mobilized to coordinate requirements and to cooperate in their solution.

Interagency, cooperative activities in the mid-1980s, prior to the start of the High Performance Computing and Communications Initiative (HPCCI), included studies to examine the need for common infrastructure. These studies resulted in the development of the National Research and Education Network component of the HPCCI,<sup>10</sup> which proposed, and subsequently implemented, Internet technologies to support the internetworking needs of the research programs of the federal agencies. In one notable case, the Energy Sciences Network (ESnet) extended these common affinity group requirements to include other administrative requirements by implementing a multiprotocol network, installing gateways for multiple e-mail systems, and so on. And although technically different, the ESnet, the National Aeronautics and Space Administration's Science Internet, the NSFNET, and the ARPANET all were able to interoperate within an acceptable range for the research community as separate, but integral, modules in a network of networks. These activities and studies drew upon the expertise and the involvement of many academic and industrial researchers as well as organizations external to the federal agencies, constituting a large affinity group and adding a large user base to the Internet and establishing its technologies as a viable and productive technology paradigm.

This success notwithstanding, the research community of each agency also needs to interact electronically with other parts of its own agency (e.g., using ESnet solutions for administration and Information resource management), which are not normally a part of the research affinity group, as well as with its affinity group

colleagues. These requirements of the research community did not always converge within the group (although convergence was generally discussed and studied); however, convergence (and ultimately interoperability) was even more problematic outside of the research community.

Affinity groups do, however, represent a very powerful method for identifying common requirements, coordinating infrastructure needs, and promoting and maximizing interoperability of applications and services across agency boundaries, extending these "functional" areas as virtual agencies. Affinity groups can also extend to industry and the private sector. Affinity groups could establish a common perspective for evaluating new technologies, eliminating unwarranted redundancies, interacting with various affinity subgroups or working groups, and sharing data, information, and knowledge about their enterprise or business area and how IT promotes effectiveness. By focusing on common requirements and solutions in this manner, affinity group activities can result in application-driven standardization for supporting important common functions, for setting priorities for new tasks, and for understanding the minimum capabilities needed to perform common business functions. They can also enhance the overall coordination for multiorganizational, distributed enterprises, as well as other attributes needed to maximize coordination for multiagency, distributed government services (i.e., the virtual agency) in the information future through the GSII.

### OPTIMIZING GSII AND NII COMPATIBILITY

Federal IT management reform is needed to correct the current problems with regard to managing IT in the federal government to achieve a management scheme that works better. A new management scheme should set up mechanisms to aid agencies in carrying out their responsibilities, to evaluate agency IT investments via performance measures of programmatic results and products, and to promote both compatibility and interoperability across recognized affinity groups. We propose the establishment of a high-level "leadership council" that brings together recognized leaders of larger or more critical mission-driven affinity groups along with government and private sector IT services providers to:

1. Promote cooperation among agencies by empowering lead agencies to conduct more multiagency procurements, by coordinating across affinity groups, and by seeking opportunities for consolidation and cooperation, where appropriate;
2. Set strategic direction and priorities for common infrastructure services and to identify lead or executive agencies, when appropriate, for procuring or providing common services;
3. Oversee a government-wide IT innovation fund;
4. Evaluate the work of agency activities through "independent" or external technology review panels or committees; and
5. Make policy recommendations to OMB to improve overall GSII effectiveness and to enhance coordination, such as changes to the Federal Advisory Committee Act to increase private sector involvement in GSII planning and decision making.

It should be noted that administrative functions for personnel, finance, procurements, facilities management, and so on logically combine the traditional federal information resource management (IRM) organizations together into an affinity group. One could conclude that this grouping results in a federal government affinity group based on electronic commerce. Individual agencies would still require policy and oversight function for IRM or IT management activities; however, it is not envisioned that multiple large centralized IRM organizations would or could promote GSII interoperability goals (e.g., large redundant organizations both in the GSA and in the agencies) or adequately serve the GSII goals and objectives well into the information future.

## GSII PLAN

Promoting interoperability, coordination, and information exchange across agencies and affinity groups is of paramount importance for achieving the goals of the NII and for creating an evolving and successful GSII. The best means of achieving this goal is through the use of the very same IT that will make the GSII successful—in other words, use of the technologies that we advocate for all agencies' businesses to further enhance the management and oversight of the GSII. As an example, a very important IT for furthering this goal is the World Wide Web (WWW), which is a product of both the federal and international high-energy physics community and also of the high-performance computing and communications research communities.

The GITS working group has endorsed the use of the WWW to create and maintain a "living" and "evolving" document accessible to all over the net.<sup>11</sup> The GITS working group has the responsibility for implementing the NPR IT recommendations, one of which is to develop a GSII plan. Understanding the changing nature and the wide variety of GSII components, elements, and layers, the GITS decided to create a collaborative, on-line document on the WWW.<sup>12</sup> The document consists of summarized reports of various affinity groups, agencies, panels, committees, and so on presenting the most current thinking with regard to GSII technology direction, issues, applications and management. It allows for interactive feedback and comment. All contributions will be referred to a GITS subgroup and/or be addressed by other expert groups.

On-line implementation plans, updates, and dialogue for the GSII, as well as its committees, activities, documents, and plans, will help to promote common understanding of issues and their status as well as to establish the foundation for the discussion of new ideas and/or requirements by government, industry, and the public on both a national and international basis.

## FEDERAL VERSUS NONFEDERAL ISSUES FOR THE GSII

Some issues that need to be resolved for the deployment of the NII result from the differences between policies, regulations, and practices of the federal government versus those in the private sector. It is timely that the Congress and the Administration are now committed to telecommunications reform, as this will help lay the foundation for resolving some of the GSII versus NII issues in the future.

Key issues that need to be addressed for successful GSII and NII integration include procurement reform, key and certificate management infrastructure, electronic signature, intellectual property, common carrier status and open access for the network and information service providers, standards setting, and cost recovery for shared infrastructures.

## RECOMMENDATIONS

First, federal IT management reform is needed to deal with Third Wave<sup>13</sup> (i.e., truly information age) organizations and business so that the federal government can, in fact, achieve its NPR goal of creating a government that works better and costs less.<sup>14</sup>

Second, the necessary features and mechanisms for achieving a successful GSII should, at a minimum, include:

1. Federal agency flexibility to meet mission requirements in a cost-effective manner;
2. Accountability based on technical success leading to programmatic outcomes and enforced through the budget process;
3. Support for easy-to-use mechanisms for interagency sharing of services and operations, including franchising, cross-servicing, and multiple agency contracts;
4. Provision of services/infrastructure as required by communities of interest (e.g., affinity groups for government business areas), by agencies, and by their customers;

5. Standards selected by affinity groups with technical assistance and possible facilitation by the lead agency or by the GITS working group;
6. An interagency fund for innovative IT programs;
7. A ready source of expert, objective advice for complex system acquisitions;
8. Central collection of information, as needed, to support government-wide priority setting, information sharing, cross-agency and cross-affinity-group coordination, and infrastructure investment optimization; and
9. A fast, equitable way of handling contract disputes.

Third, improvement in government and private sector cooperation (e.g., via extended affinity groups) is required to achieve a more timely acquisitions process; a more responsive standards process; open-access interoperability between NII and GSII services providers; and revision of the Federal Advisory Committee Act (FACA) to provide relief for some activities with regard to the GSII and NII to encourage rather than to discourage such cooperation. The FACA can often discourage government activities from utilizing academia, industry, and other private sector organizations in developing priorities and goals for designing and implementing the GSII and ensuring that the GSII is an integral component of the NII.

Lastly, continued dialogue on the direction of development and deployment of the GSII—especially relative to its superset, the NII—via the WWW implementation of the GSII Plan, is needed to ensure convergence of these two very important national resources and to achieve the optimum "range of interoperability" and the maximum benefit that one could expect from such a complex and diverse infrastructure.

## NOTES

1. Toffler, Alvin. 1980. *The Third Wave*. William Morrow & Company, New York, pp. 233, 262, and 404–415.
2. Information Infrastructure Task Force. 1993. *The National Information Infrastructure: Agenda for Action*. Information Infrastructure Task Force, Washington, D.C., September 15.
3. Office of the Vice President. *From Red Tape to Results: Creating a Government That Works Better and Costs Less*. U.S. Government Printing Office, Washington, D.C., September.
4. Committee on Applications and Technology (CAT), Information Infrastructure Task Force Committee. 1994a. *Putting the Information Infrastructure to Work*. Information Infrastructure Task Force, U.S. Department of Commerce, Washington, D.C. Also, Committee on Applications and Technology (CAT), Information Infrastructure Task Force. 1994b. *The Information Infrastructure: Reaching Society's Goals*. U.S. Government Printing Office, Washington, D.C., September.
5. Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C., May.
6. Aiken, R.J., and J.S. Cavallini. 1994. "Standards: Too Much of a Good Thing?," *Connexions—The Interoperability Report* 8(8) and *ACM StandardView* 2(2).
7. U.S. Department of Energy. 1993. *Positioning the Electric Utility to Build Information Infrastructure*. DOE/ER-0638, Department of Energy, Washington, D.C., November.
8. Acquisition Working Group. 1994. *Analysis of POST-FTS2000 Acquisition Alternatives*. Interagency Management Council, Washington, D.C., September.
9. Federal Internetworking Requirements Panel. 1994. *Report of the Federal Internetworking Requirements Panel*. National Institute of Standards and Technology, Washington, D.C., May 31.
10. Federal Coordinating Council for Science, Engineering, and Technology, Office of Science and Technology Policy. 1991. *Grand Challenges: High Performance Computing and Communications, The FY 92 U.S. Research and Development Program*. Committee on Physical, Mathematical, and Engineering Sciences, Office of Science and Technology Policy, Washington, D.C., February 5.
11. GITS Working Group. 1994. *Vision for Government Information Technology Services and the NII*. GITS Working Group, Washington, D.C., July.
12. See <http://www.er.doe.gov/production/osc/gsiiplan>.
13. Toffler, *The Third Wave*, 1980.
14. Gore, Albert. 1993. *Creating a Government That Works Better & Costs Less: Reengineering Through Information Technology*. U.S. Government Printing Office, Washington, D.C., September 1993.



### 3

## **Cutting the Gordian Knot: Providing the American Public with Advanced Universal Access in a Fully Competitive Marketplace at the Lowest Possible Cost**

*Allan J. Arlow*

*Telecommunications Consultant, Annapolis, Md.*

### **BACKGROUND**

#### **The Advanced Universal Access Dilemma**

A common vision is shared among both the business community and the federal political establishment as to what the broad characteristics of the national information infrastructure (NII) should be. Both the Democratic Administration<sup>1</sup> and the Republican Congress<sup>2</sup> share the view with the private sector that the NII must be built by private business in a competitive marketplace, rather than directed and regulated by government. Moreover, it should be a "network of networks," consisting of interconnected but competing systems, not just constructed but also operated in accordance with the competitive market model.

Yet neither the Congress nor the Administration is content with laissez-faire, purely market-driven deployment. A sense of national urgency, not only to ensure our ability to compete with other nations in the global economy, but also to help address major domestic concerns, has prompted other, not easily reconcilable goals. The first is to have the advanced infrastructure deployed as widely and rapidly as possible, with particular emphasis on early access by education and health care providers and users.<sup>3</sup> There is also the second, broader social concern about preventing the potential disparity between "information haves" and "information have-nots" for critical advanced access to the NII at reasonable and affordable rates. The current Senate proposal is to provide subsidies to designated carriers of last resort—a continuation of the existing narrowband mechanisms and an extension of those mechanisms into the realm of new and as yet unknown services.<sup>4</sup>

This public vision of the NII is not just internally inconsistent. It relies on a foreknowledge of the events that will define future technology deployment. Yet the actual composition of the anticipated advanced physical architecture and technology that will link schools, hospitals, government, businesses, and residences and will ultimately constitute the NII is unknown today. Although it is a given that advances in technology should benefit certain broad areas of our society, such as education and health care, there is no consensus as to what specific applications will be most valuable to the economy and the social fabric of the country.

#### **The Magnitude of Current Basic Service Subsidies and Proposed Broadband Service Subsidies Under Current Operating Arrangements**

The dimensions of the problem are well documented. The nationwide local exchange carrier (LEC) collection of revenues from customers and interconnecting carriers in 1992 exceeded \$91.5 billion. Subsidy level estimates from various sources range from \$4 billion to \$20 billion.<sup>5</sup> In a study of data from Tier 1 LECs which discussed only the averaging effects within individual LECs (excluding the complex subsidy flows among large and small LECs and interexchange carriers), costs for providing service to rural customers exceeded rural revenues by 35 percent—\$5 billion—or more than \$19.00 per line per month.<sup>6</sup>

A national policy to create a fully capable broadband network universally available in rural areas and provided by regulated carriers using traditional costing methods would require far greater subsidies: total loop and non-loop costs per line in rural areas would range from \$154 to \$168 per month per line, depending upon whether deployment completion was targeted for a 10- or 20-year period. Since rural telephony per line revenues average approximately \$54 per month, customer revenues or subsidies would have to bridge a \$100 per month gap. These estimates exclude the incremental cost of any customer premises equipment that would be necessary.<sup>7</sup> The regional bell companies and several of the largest cable multiple system operators are seeking or have entered into agreements with set-top box manufacturers, such as Digital and Scientific-Atlanta, to develop proprietary consumer premises equipment with an eye toward bringing the cost down to well below \$1,000 per unit.<sup>8</sup> Even if the nation chose to afford it, providing subsidies under a business-as-usual model would effectively prevent competitive entry and keep costs high. Since there is currently no two-way universal broadband network in place, there is no rationale for selecting and subsidizing one potential provider over another.

This paper does not propose to chart a hybrid middle ground between regulation and competition. It also neither evaluates existing and proposed narrowband transition plans nor offers any new transition mechanisms to end the current subsidy scheme for those universally available services. What this paper offers instead is a mechanism to provide universal two-way broadband service to high-cost areas in an industry- and technology-neutral way, through the use of a fully competitive marketplace, and at the lowest possible cost.

### The Need for a Separate Broadband Paradigm

As noted above, the current Senate bill, S. 652, calls for an evolving definition of universal service. The law's effect would be to pull new broadband services into the subsidized and regulated carrier-of-last-resort mandatory provisioning scheme as such technology gained popular acceptance.<sup>9</sup> As has already been discovered in narrowband communications markets supplied by multiple competing providers, the efficiencies of competition cannot be sustained if one or more parties' market behavior and obligations are being mandated by government. In an environment where there is a "network of networks" that both interconnect and compete with each other, intra-industry subsidies undermine and distort the marketplace while providing no direct end user benefit.

Although complaints about pricing and customer service led to the reregulation of cable television, cable companies are not monopoly suppliers of broadband services. A residential customer in search of broadband services has the technical capability to receive broadcast television, cable, multichannel multipoint distribution service, low-power satellite, and direct broadcast satellite signals. Portions of the entertainment and information that are received via such media are also available from newspapers, broadcast radio, on-line computer services, video cassettes, laser disks, computer diskettes, and CD-ROMs. The first principles of a new broadband paradigm rest on a free market foundation. Although robust narrowband local exchange competition will eventually arrive and enable the deregulation of that market, advanced infrastructure investment must go forward now in a "clean state" environment, in which government dictates neither the financial nor technical terms for the offering of new services. A distinct broadband deployment-friendly and regulation-free environment, possessing the following characteristics, is necessary:

- *LEC broadband investments that are neither subsidized nor regulated.* It appears to be a political certainty that some price and service regulation of local exchange service will continue far into the future either in rate of return or price cap form. Given this fact, shareholders, not telephone ratepayers, should take the risks and rewards of broadband deployment. Whether in the form of video dialtone or more vertically integrated arrangements, broadband assets and expenses should be unregulated, below-the-line allocations or should reside in separate subsidiaries or affiliates.
- *Unregulated cable system broadband and telephony offerings.*
- *Mandatory interconnection.* All providers of telecommunications services, whether or not they have been previously regulated, should be required to interconnect with each other,<sup>10</sup> but there should be no other day-to-day regulatory oversight or standards to which providers are obliged to adhere.

A great deal of capital is currently being invested with the hope of creating the technologies, products, and services that will ultimately prove victorious in the marketplace and constitute the NII. Any participation by government for the purpose of ensuring universal access to those advanced broadband services by the general public must, therefore, be *technology and service neutral* and must have the smallest possible impact on the marketplace.<sup>11</sup>

### **The Development of the Competitive Market for Broadband Services**

Telephone, cable, computer, software, and telecommunications equipment vendors have made massive commitments with each other to upgrade their products and services for a two-way broadband communications world.<sup>12</sup> At the same time, legal barriers have been struck down in several federal court cases,<sup>13</sup> and many state legislatures have either granted or are considering granting basic service price regulation instead of rate-of-return regulation, in return for lower prices, interconnection with competitors, and advanced infrastructure investments.<sup>14</sup>

As with other industries, it will be the combined effect of individual companies' strategies, and the public's reaction to the products and services offered pursuant to those strategies, that will determine how the total market for access to advanced two-way broadband services will actually develop.<sup>15</sup> Obviously, in an unregulated marketplace, deployment will take place at differing speeds throughout the country. Different suppliers, only some of whom have ever been under Federal Communications Commission (FCC) jurisdiction, may be supplying access to advanced services. Many businesses in major markets today already have multiple suppliers of two-way, high-speed narrowband access. However, there is one certainty: variations in public acceptance of particular means of access will cause higher or lower rates of market penetration and differing market shares among competing access suppliers; winners and losers will emerge, and along with them, certain technologies will become more ubiquitous and others will fall by the wayside.

It will be the marketplace winners in this initial deployment phase that will help define, *de facto*, the standard for which it means to have advanced access to the national information infrastructure and thus become "advanced access interfaces" (AAIs). The role of government in this environment should be to augment the forces of the marketplace, rather than interfere with them, so as to bring about the availability of advanced access on a universal basis. In order for such access to be universal, it will require legislation enabling the FCC, in partnership with the states, via a Joint Board, to administer a plan that provides the incentives for universal deployment at reasonable rates within a competitive market environment. The mechanisms of that plan, their operation, and their impact are described below.

## **ADVANCED UNIVERSAL ACCESS DEPLOYMENT PLAN**

### **Phase 1: FCC-State Joint Board "Advanced Access Interface (AAI) Recognition Rulemaking Proceeding"**

The marketplace success of various types of advanced network access interfaces will, as noted above, create a *de facto* definition of what it means to be an "information have" in American society. This definition would become *de jure* when penetration of potential AAIs reached a specific level, either set by Congress or delegated to the FCC or the Joint Board. The plan's purpose is to prevent the creation of a long-term "have not" community of sparsely populated or untenably high-cost service areas, while preserving the operation of the competitive marketplace.

Pending legislation uses an imprecise standard for the circumstances under which subsidies should be provided for new services.<sup>16</sup> A specific, easily recognized benchmark is far more desirable, because investment is augmented by reducing the risk that comes from regulatory uncertainty. The benchmark could be, for example, that 80 percent of institutions and businesses and 50 percent of residences in the top 25 metropolitan service areas have advanced access interfaces that have, as a minimum, the technical capability sufficient to simultaneously receive five channels and send one channel of National Television System Committee quality video (assuming that

this reflects the abilities of the most popular residential interfaces). It is important that access be viewed in terms of *capabilities* and not technologies, in order to ensure that any subsidies that assist deployment are technology neutral. For example, requiring a dedicated residential 440-MHz downlink and 6-MHz uplink is not meaningful if advanced access can also be in the form of a set-top box that couples digital cellular radio linked to massive regional servers with a half-meter dish of a co-venturing direct broadcast satellite system. What is most important is that a particular advanced access interface *has succeeded in the market*.

Once this plateau has been reached—i.e., when the level of market penetration has become sufficient to find that a core group of "information haves" has evolved—the implementation of the plan can begin. The first phase of the plan is an "Advanced Access Interface Recognition Rulemaking Proceeding." In such a proceeding, the AAIs with a defined minimum market share whose specifications have by this time been well documented, will be described with particularity and declared by the FCC-State Joint Board to be critical to the national interest. Companies, standards groups, associations, and others may submit their AAIs for recognition, because it will be these interfaces, and only these, whose deployment will be eligible for subsidies.<sup>17</sup>

In order to ensure the availability of access throughout the country via equipment that individual users may purchase, all of the accepted interfaces for which subsidies will be made available must, in return, be open and nonproprietary. Any proprietary rights in the AAI identified by the rulemaking proceeding become subject to strict public policy limitations; in that way, no provider of technology can hold hostage to its licensing fee or product purchase demands those customers for whom a highly competitive marketplace is unavailable.<sup>18</sup>

### Phase 2: Joint Board "Inverse Auction" Proceeding

Once the AAIs have been officially recognized, the Joint Board undertakes to implement universal access through a proceeding with the following elements:

- By public notice the Board divides the country into geographic market areas, much as the FCC has done for cellular and personal communication service (PCS) licensing, in which parties may compete for the right to be the subsidy recipient.
- If not previously set forth by enabling statute, national market penetration targets are adopted; for example, AAI NII access available to 100 percent of institutions by 2005 and to 90 percent of businesses and 80 percent of residences by 2015.
- If not previously set forth by the enabling statute, the proceeding would specify the service standards and administrative procedures that an applicant would be obligated to meet.
- In order to ensure the affordability of the service, the proceeding would also specify the "Top Rate" at which AAI NII access could be charged.<sup>19</sup>
- Rules for parties wishing to participate in the Inverse Auction would allow any party to apply for as many or as few markets as it wishes upon showing proof of general qualifications.
- A bid would consist of a response to the following question: What is the smallest dollar amount (the "Performance Payment") the bidder will accept to shoulder the supplier-of-last-resort responsibility for building the infrastructure to meet the AAI national market penetration target for the market in question?
- The party submitting the lowest bid in each market would be designated the Universal Access Supplier (UAS) for that market. The UAS would have many of the familiar service standards requirements now imposed upon common carriers (readiness to serve, time from request to time of installation, operator services, blocking rates, E-911, and so on), but they would be voluntarily assumed a part of the bidding process.<sup>20</sup>

As was widely noted at the recent auction of PCS spectrum, it is highly unlikely that strategy and game theory will play a major role in the bidding process. Some of these strategies may lead to later problems if not recognized early on. For example, an applicant who already has a significant market presence and believes that economies of scale are critical to long-term profitability may "lowball" a bid to discourage potential competitors from entering the market altogether. An applicant who has historically had protection from competition may also be tempted to lowball if it believes that it can bluff the government into providing a bailout or relaxing or

changing its rules with respect to cross-subsidies, pricing, or other requirements, as an unmet deadline approaches. As a consequence, it will be necessary to have penalties for failure to perform and, perhaps, performance bonds to protect the public from parties who underbid recklessly and then seek to abandon market obligations.<sup>21</sup>

### **The Post-Award Competitive Marketplace**

The award of UAS status would have little immediate impact on the marketplace. Unlike radio licenses or exclusive service territories, no scarce franchises are granted. All parties, UAS and non-UAS alike, would still be free to compete in deploying the infrastructure, selling to any customer on a competitive basis. Only the UAS would be required, upon request, to provide access to the requesting customer at no more than the Top Rate—an umbrella price for both its own and competitors' comparable access offerings—and annually report on its progress toward meeting the deployment target date. Neither federal nor state agencies would oversee the day-to-day construction or offering of AAIs, either by the UAS or other suppliers, until the UAS filed to receive its Performance Payment. This competitive marketplace coupled with the UAS Performance Payment incentives would encourage the most rapid deployment and meet the national deployment schedule targets.

Although the competitive marketplace debate has centered on visions of "last mile" competition among local exchange carriers, cable systems, and interexchange carriers, advances in technology, auctioning of spectrum, and the lessening of regulation will open the door to many potential participants, especially consortia using combinations of diverse technologies. Among possible providers:

- Direct broadcast satellite systems with broadband downstream capabilities may find synergies with cellular companies that will have tremendous excesses of capacity in rural markets. Using digital radio technologies, the cellular carriers may not only be able to handle full narrowband transmission demands but also will have sufficient reserve bandwidth to load-manage two-way broadband transmissions from business and residential customers as well.
- UHF (and even VHF) television spectrum is largely vacant or underutilized in most rural areas and is sufficient to enable an entrepreneur to provide low-cost broadband digital access over wide spaces of sparse population and, with low-power cellular technology, over more densely inhabited areas.
- Electric utilities, the "third wire," will be able to participate in the market much as telephone and cable companies; since electricity is likely to remain largely a regulated monopoly service, accounting or separate subsidiary requirements will keep the unregulated, competitive services free of improper cross-subsidy.

### **Distribution of the Performance Payment and Market Growth Beyond the Initial Deployment Period**

At the end of the deployment target period (or prior thereto if the UAS has met the goal ahead of schedule), the UAS would petition for its Performance Payment under a "UAS Completion Proceeding." The distribution mechanism could be administered at either the state or federal level:

- At the state level, the Public Service Commission (PSC) or other authorized body would review the UAS's affidavits of satisfaction of the access requirement and dispense the Performance Payment accordingly.
- If a federal proceeding is the policy route of choice, the FCC could review the UAS's affidavits of satisfaction of the access requirement. The PSC of the state in which the UAS market is located would be the designated intervenor in the Performance Payment approval process.<sup>22</sup>

In either case, the subsidy would be simple to administer and should provide incentives for the desired behavior. Drawing rights could be made available to each state PSC or to the FCC by the Treasury for the amount of subsidies maturing in markets within their jurisdiction during each year. Since the state PSC or the

FCC could withhold certification and thereby delay the Performance Payment, the UAS would have an incentive to make the strongest case and be responsive to customer complaints and government inquiries. However, in order to be sure that the subsidies are directed to the purposes for which they were originally bid, neither the state PSC nor the FCC should be permitted to use the proceeding as an opportunity to negotiate quid pro quo forgiveness of missed targets in return for lower prices or regulatory oversight of operations, and so on.

The resulting high market penetration of advanced access would probably also reflect a competitive marketplace. Ideally, it should render unnecessary both continuing subsidy of remaining and growth infrastructure investment and subsidies for reasonably priced access. Nevertheless, although these mechanisms will have enabled the initial deployment of advanced access, there may be some markets where there is a continuing need to assure the availability of service out into the future or to allow for circumstances where competition may not be adequate to ensure affordable rates. If such circumstances were to occur, the state PSC could petition the FCC to provide for subsequent rounds of UAS bidding for a designated market in time blocks, i.e., bidding to be the provider of AAI access at no more than the Top Rate to all requesting parties on a last-resort basis for a designated number of years.

### Funding the Performance Payment

There is already an abundance of proposals on how and from whom subsidies for narrowband universal service should be collected.<sup>23</sup> For those required to participate in providing the subsidy, the issue is not only the participation itself, but also the annual amount and open-ended duration of the subsidy. This plan addresses those issues and limits their ability to disrupt both the marketplace and the participants' efforts at business planning.

The first principle with respect to imposing any subsidy should be to spread the burden as broadly and fairly as possible across market participants. This is especially true in this instance because there will be a variety of communications technologies and varying levels of distribution of computing power (i.e., some systems of access may rely on large regional servers connecting to comparatively simple consumer premises equipment, while others may have set-top boxes whose capabilities rival those of today's workstations). Excise taxes on the value of shipments of computers and peripherals, telecommunications equipment, and satellite ground and mobile equipment, along with taxes on LEC, cellular, IXC, and CATV service revenues, will provide that base.

Ordinarily, subsidies distort markets because transfer payments flow from the pockets of companies designated efficient and profitable into the pockets of the designated inefficient and needy. This plan neither designates nor offers preferences to any potential provider.<sup>24</sup> Consequently, these proposed subsidies would flow only to parties providing access via interfaces that have already proven themselves successful in the marketplace. As a result, the distortion of transfer payments is minimized; the funds will flow back to the same industry participants who contributed to the fund and, potentially, in roughly similar proportion, preserving market efficiency as well as equity.

The second principle with respect to imposing a support burden is that it be predictably measured and finite. Under this plan, the national liability and maturity date for each and every market will be known well in advance of the time that any payment comes due, because the bidding and selection will occur years before the Performance Payment will be distributed. The process can be further fine-tuned by holding auctions on a staggered basis, using an accurate cross section of markets rather than the largest markets first. Such a procedure would enable the establishment of an approximate market price for each type of service area as the process went forward. The industry tax could then be set in a manner that would ensure proper levels of funding.

### Political Considerations

The collection of new taxes is always politically unpopular, but the fact that these funds would be collected from a discreet group and segregated and ultimately paid out to members of that same class of taxpayers, while simultaneously benefiting the general public in every congressional district, should alleviate much of the concern about government involvement. Since there is no day-to-day administration of a high-cost

service fund or other regulatory oversight of the potential recipients, there is no need for an independent entity to handle the funds in order to avoid a debate on the application of federalist principles. Opposition would also likely come from independent telephone companies, who have been both recipients of subsidies and protected from competition within their own service territories in the provision of narrowband services. Yet they will be positioned better than almost any other competitor to take on the role of UAS and will likely be the low bidders in their local markets; they are also likely to form partnership or consortia with other independent companies and bid successfully for the regional market.

Although this plan envisions a major role for state agencies, the overall operation of the proposal will preempt state jurisdiction over the day-to-day local activities, such as pricing, service standards, and so on, traditionally regulated at the state level. There is already, however, a recognition within many states, as noted above, that new services will be offered in a competitive, market-driven environment, where regulation does not need to be substituted for competition. With respect to the UAS itself, the state will be able, at its option, to participate in or control key aspects of the approval process. Finally, it should be noted that the plan, which will enable more rapid deployment of infrastructure, will by its operation not only improve the quality of life within the jurisdiction, but also increase commerce for both CPE and ancillary services, thereby increasing local and state sales and income tax receipts.

### CONCLUSION

The ideal of relying solely on unfettered competition among multiple suppliers of broadband access to reach a goal of affordable, universally available advanced services carries with risk of failure that policymakers within the United States are unwilling to take. At the same time there is a realization that neither regulation nor managed competition are viable long-term options. The proposals set forth in this paper offer one possible avenue to achieve advanced universal access in a fully competitive environment with the least economic and social costs and minimal regulation.

### NOTES

1. Information Infrastructure Task Force (IITF). 1993. *The National Information Infrastructure: Agenda for Action*. IITF, U.S. Department of Commerce, Washington, D.C., September 15.
2. S. 652, Telecommunications Competition and Deregulation Act of 1995, 104th Congress, 1st Session, Report No. 104-23, Sec. 4–5, p. 3.
3. *Ibid.* at Sec. 103(d), pp. 304–310.
4. *Ibid.*
5. Weinhaus, Carol, and the Telecommunications Industries Analysis Project Work Group (Weinhaus et al.). 1994. *Getting from Here to There: Transitions for Restructuring Subsidies*, p. 16.
6. Weinhaus et al. 1994. *Redefining Universal Service: The Cost of Mandating the Deployment of New Technology in Rural Areas*, pp. 8–9.
7. *Ibid.* at p. 12.
8. *Telecommunications Reports*. 1995. "Ameritech to Deploy Scientific-Atlanta Technology," 61(8), February 27, p. 7; and "RFP Aims to Lower Set-Top Box Manufacturing Costs," 61(9), March 6, p. 27.
9. S. 652, Report No. 104-23, Sec. 103(d), pp. 39–40.
10. This is arguably not new regulation. Indeed, if there were no prior history of telecommunications regulation, it might well be argued that new legislation to ensure such interconnection would not be necessary. Application of Sec. 3 of the Clayton Act might reasonably be interpreted to prevent telecommunications service suppliers from denying other service providers access to its customers. See *Datagate v. Hewlett-Packard*, 914 F.2d 864 (9th Cir. 1991); *XETA Inc. v. ATEX Inc.*, 852 F.2d 1280 (D.C. Cir. 1988); and *Digidyne Corp. v. Data General Corp.*, 734 F.2d 1336 (9th Cir. 1984).
11. This paper does not discuss programs to fund or provide broadband services to those individuals or institutions determined to be in need of special subsidies. Rather, it is limited to the issue of how best to make access to advanced broadband services universally available to the general public. Nevertheless, certain principles that further the overall goal of advanced universal access do apply: in a competitively neutral marketplace, public support, whether targeted recipients are poor individuals or distressed education, health, and public institutions, must provide the recipients

with the flexibility to obtain services from competing vendors rather than require last-resort carriers to offer services at noncompensatory rates. Armed with the ability to select services from among several vendors, the needy will make choices that will reflect the efficiencies of the marketplace.

12. A recent example is PacTel's \$150 million head-end and set-top box purchase agreement as part of an interactive video network. See *Wall Street Journal*. 1994. "Scientific-Atlanta Inc.—Concern Is Awarded Contract to Equip PacTel's Network," April 24, p. B4. Yet even this appears modest by current standards; cf. Ameritech's \$700 million in contract commitments to Scientific-Atlanta, note 8, supra. A great deal of publicity has surrounded announcements by regional Bell operating companies on the formation of programming consortia (Bell Atlanta, PacTel, and NYNEX; Ameritech, BellSouth, SBC Communications, and Walt Disney), as well as speculation with regard to US West's future intentions regarding its investment in Time Warner.

13. C&P Telephone Co. of Virginia et al. v. U.S. et al., 42 F.3d 181 (4th Cir. 1994); NYNEX v. U.S., 1994 WL 779761 (Me.); US West v. U.S., 48F.3d 1092 (9th Cir. 1994).

14. States where bills were introduced in 1995, and which currently appear to be under active consideration, include Colorado (HB 1335), Texas (HB 2128), and Tennessee (rival bills SB 891 and 827, HB 695 and 721). The Wyoming Telecommunications Act of 1995 (HB 176) was signed into law in March, and Georgia's Telecommunications and Competition Development Act of 1995 (SB 137) became law in April. In North Carolina, as of this writing, HB 161 awaits the governor's signature.

15. *Telecommunications Reports*. 1995. "Competitors and Allies: Cable TV Companies, Telcos, Broadcasters Jointly Will Create New Markets," 61(14), April 10, pp. 16–20.

16. S. 652, Report No. 104-23, Sec. 103(d), p. 40.

17. First in the market, especially in new technology, does not guarantee dominance, or even success (e.g., 8-track cassettes, Commodore computers, Atari and Magnavox video game systems, Bowmar calculators) even if the technology is superior (Betamax, digital audiotape). Commercial success is a reflection of market acceptance and manufacturing efficiency and yields a reduction in incremental risk per unit manufactured. If the public is required to subsidize the initial build out of long-term infrastructure, there must be a high level of confidence that the access interfaces installed under subsidy will have a long useful life and that individuals who purchase equipment and later relocate or need to access nonlocal information sources will not find their investments rendered worthless.

18. Any proprietor is free not to participate in the proceeding so that there is no state taking of intellectual property.

19. It would be reasonable to expect that the rate would be in the top quartile of rates charged nationally, i.e., still affordable by definition, but not a windfall rate when compared to those paid by the majority of customers nationwide.

20. In calculating its bid, the applicant would, of course, factor in such items as the costs of meeting the service standards obligations, the Top Rate, current and projected levels of competitive deployment in the market, demographics, terrain, and so on. The development of competitive markets will provide vast historical data on costs, price elasticity of demand, and the like. Obviously, within that price-demand elasticity curve, the higher the Top Rate, the lower the Performance Payment necessary to attract the successful bidder.

21. The UAS could be required to post a performance bond at least 5 years prior to the target date. If it advises the FCC that it is abandoning its UAS status, qualified but previously unsuccessful bidders in the market may then rebid for the right to be the UAS. The Treasury would collect on the bond to the extent that the second bid series produced a result that was higher than the original UAS's bid plus a penalty and administrative premium.

22. In multistate markets, PSCs would designate members to sit on a panel that would serve as the sole intervenor.

23. See Weinhaus et al., supra, *Getting from Here to There: Transitions for Restructuring Subsidies*, 1994, for a review of the most commonly discussed alternatives. Although based on recent prices paid for PCS spectrum, an auction of unused UHF and VHF spectrum could well provide all of the funds necessary to finance advanced universal access; for the purposes of this paper, it is assumed that the funds must be raised from market participants.

24. This is particularly important if, by operation of the program, a UAS receives what could be considered a windfall profit: if the market develops more rapidly and efficiently than anticipated and the performance targets are met with little or no investment that is less than fully compensatory, the UAS would receive the Performance Payment merely for having been "on call."



## 4

# The Role of Cable Television in the NII

*Wendell Bailey, National Cable Television Association*

*Jim Chiddix, Time Warner Cable*

There are over 11,000 cable television systems in the United States. They, along with the equipment vendors that have traditionally worked with them, represent a springboard to an extremely high capacity, nearly ubiquitous broadband two-way information infrastructure serving both the business and residential communities. This existing resource passes over 97 percent of the homes and businesses in the United States and is connected to nearly 65 percent of the residences. The cable industry has made significant improvements in the quality and reliability of its networks over the past few years and has begun to put in place the structures needed to transform its systems into highly flexible multiple service vehicles. Yet, though we have continually expressed a willingness to make the investment necessary to bring about this transformation of our facilities, there are significant impediments to that goal. The primary barrier to the realization of our full participation in this national information infrastructure is posed by excessive regulation of the cable industry. This federally imposed burden, which undermines our financial capabilities, is coupled with other levels of regulation and barriers to the provision of communication services posed by state regulators and local franchise authorities. When these restrictions are linked with such problems as denial of switch interconnection access and usurious pole attachments and duct rental contract provisions by potentially competitive local exchange carriers, it becomes clear that the most capable and flexible network available today can contribute its full resources to the achievement of a national information infrastructure/global information infrastructure (NII/GII) only if the government promulgates a rational set of regulations that make it possible for full and complete competition to flourish.

The basis for today's cable television networks is coaxial cable, a radio transmission medium capable of transporting a large number of separate radio carriers at different frequencies, each modulated with analog or digital information. It is common practice to use filters to segregate the high- and low-frequency portions of the spectrum to allow simultaneous transmission of information in both directions. This ability to transport coexisting carriers carrying different kinds of information (essentially separate "networks") provides an enormous amount of flexibility and capacity. Today's coaxial systems are, however, arranged architecturally in a broadcast topology, delivering the same spectrum of information-bearing radio carriers (commonly called "channels") to every customer in the community.

The advent of fiber-optic transmission technologies optimized for broadband use during the last decade allows a cost-effective upgrade of existing broadcast coaxial networks to a hybrid fiber coax (HFC) architecture, with fiber trunks providing transmission to and from small neighborhoods of a few hundred homes. This arrangement of fiber and coax segments the traditional coaxial-only transmission plant into many localized areas, each providing a localized assortment of information. When combined with the economics of high-speed digital switching technologies, this architecture allows the simultaneous delivery of multichannel television transmissions and switched video, voice, and data services to and from individual homes within these small serving areas. The upgrade of existing coaxial cable TV networks to a hybrid fiber coax architecture as described above costs less than \$150 per home passed and can be accomplished over the period of a few years.

The current design and economic considerations of a hybrid fiber coax cable network call for an initial passing of 500 homes (of which about 300 currently subscribe to cable), but it can be further segmented into smaller and smaller coaxial-serving areas. Its potential digital capacity is at least 1.5 gigabits per second downstream (to the home), and 500 megabits per second upstream. This is in addition to the transmission of 80 channels of broadcast National Television System Committee and high-definition television signals. Networks

configured this way can provide for extremely high-speed access to both symmetrical and asymmetrical data services, as well as the ability to evolve to even higher speeds and higher usage levels through further segmentation of the coaxial "last mile" of the network. Further, because analog and digital carriers using completely different modulation schemes and protocols can coexist, an evolutionary path is kept open for continuing advances in digital services and technologies, while, at the same time, allowing the pragmatic adoption of technology options to build cost-effective systems today.

In building a hybrid fiber coax plant, the cable television industry is providing a cost-effective, reliable, ubiquitous transmission system that can simultaneously support many separate networks, the sum of which makes the necessary upgrade investment a sustainable one for any company in our industry. Time Warner, for example, currently has plans to build at least four separate networks on the foundation of its fiber-upgraded plant. Cable Television Laboratories, an industry R&D consortium, is working with a group of broadband equipment manufacturers to develop an open standard that is called the Spectrum Management Application, to allow the coexistence of these multiple networks and the maximization of the efficiency with which the radio frequency spectrum within the transmission plant is used.

In Time Warner's upgraded plant, for example, there are plans for the coexistence, and separate operation, of networks that will continue the broadcast of scores of analog TV channels and will begin the delivery of high-quality digital telephone service, the provision of high-speed personal computer interconnection service, and access to a wide range of interactive video services. An additional network designed for interconnecting PCS radio microcell sites may be integrated into that residential network, or may be operated independently.

Time Warner's personal computer interconnection services will incorporate support of the Internet Protocol (IP), and they are in the process of working with a number of companies that are designing the necessary hardware and software systems needed to provide a service of this type. Many companies in the cable television industry envision initially offering access to a variety of online service providers, as well as e-mail and direct connection to the Internet at speeds currently unavailable to almost anyone, anywhere. We do not pretend to know how the Internet will evolve; there are those who claim that it will one day provide video, shopping services, and all the rest. It is far from that today and has many shortcomings, as recent security problems have demonstrated. But regardless of whether PC interconnection ultimately flows through a number of competing national online services or through the laissez-faire anarchy of the Internet, cable intends to offer a highly competitive avenue for local residential and business access to any viable service provider.

On a separate front, the industry is just beginning to understand interactive television services through projects like Time Warner's Full Service Network in Orlando, and while this and other trials are beginning to teach us a few hard-won lessons, it is far too early to set the standards for technologies that are still in their early innovative phase. We are, however incorporating standards wherever we can, particularly for content (MPEG, for example.) There will still be public policy questions to be dealt with in interactive TV, but they should wait until the technology and business mature at least to the point that such issues can be intelligently debated based on information and experience that we will gain from the trials that are under way.

An example of the harm that government regulation can do is evidenced by the so-called rate regulations that the Federal Communications Commission (FCC) enforces on the cable television industry. These regulations are over 600 pages in length and are so complicated that the commission has issued corrections and clarifications that total several hundred additional pages. While many people think that these rules are just about the subscription fees that operators may charge to a customer, they do not understand that these regulations also directly affect whether or not we can include the cost of a new piece of equipment or a mile of fiber-optic cable in the cost of "doing business" and adjust our rates to recover the expense. In fact, the top FCC official recently replied, in response to a question about "upgrade incentives," that upgrades are just a code word for rate hikes. Any premature attempts to set either mandated technological standards or regulations shaping business structure have the real potential to slow innovation in a field that may have great future value to American business and society. Such standards may be called for by industries that have a stake in the status quo and wish to limit or confine the directions that innovation may take, but these calls must be resisted if the country is to benefit fully from the fruits of the process of invention and exploration.

In the meantime, the PC networks that the industry is building will follow a much better understood path, since they will use an existing installed base of home computers and their operating software, and an existing communications model as a starting point. Although companies such as Time Warner will be one of many content providers, these will be networks that success in a competitive world will demand be kept open, and they will be as symmetrical as actual usage demands.

The cable industry views the competitive provision of wireless and wired telephone service as an immediate opportunity. The benefits of competition to the American public can only be realized, however, if four legislative and regulatory principles are observed:

- The elimination of historic state and local barriers to competition in telecommunications;
- The creation of requirements for interconnection, access, compensation, unbundling, collocation, pole and conduit sharing, and number portability and dialing parity by the incumbent telephone monopoly;
- The prevention of interference by local authority in the growth of competing telecommunications services; and
- The recognition that to enhance telephone competition, debilitating cable rate regulation must be reformed.

The industry strongly supports the concept of universal service and agrees that each telecommunications provider should carry its fair share of the burden to ensure that universal service remains a reality. However, universal service should not be maintained through subsidies flowing directly to particular telecommunications providers. Rather, such subsidies should flow directly to the consumers in need of such support, if the concept of truly competitive service is to be maintained in a multiple-provider environment.

In summary, the existence of a ubiquitous broadband cable television system in this country affords an almost unique opportunity to see the rapid realization of a series of extremely powerful digital networks. Some can offer competition to the existing telecommunications monopolies; some can interconnect computers and computer-based interactive television terminals in ways that can lead to an explosion of new and highly innovative services. In order to open these benefits to the American public, however, government will have to take an active role in lowering barriers posed by incumbent telephone companies and by state and local governments, as well as by current federal cable television regulation. Government will also have to exercise restraint in order to allow innovative and entrepreneurial forces to chart the way into a digital future without the imposition of mandated standards designed to protect a variety of existing interests.

The cable television industry has much to offer in helping this nation to realize the potential of the NII. The other network providers are seeking alliances with us and we with them. If the network of the future is to have a chance to be what its promoters believe it can be, government needs to provide guidance, not hindrance. If it can do that, we can do the rest.

#### ADDENDUM<sup>1</sup>

1. For your company or industry, over the next 10 years, please project your best estimate of scheduled construction of new broadband facilities to the total residential and small business customer base, in two-year increments.

Over the next 10 years, we project that the cable industry will upgrade its plant to hybrid fiber/coax architecture with relatively small node areas (approximately 500 homes passed on the average) to virtually the entire cabled universe. Since small businesses are for the most part mixed with residences, these same projections would apply to both bases. Currently, there are approximately 97 million residences in the country, 95 percent of which are passed by cable plant. (Sources: A.C. Nielsen and Paul Kagan.)

2. Over the same period, please project dates when each family of service will first become available, and its subsequent penetration of the total base, again in two-year increments.

The four major families of service which we foresee are multichannel broadcast signals; digital telephony; personal computer network services; and switched interactive multimedia services. Multichannel video services are already available, of course, to all homes currently passed by cable TV plant—96 percent of U.S. households. Cable telephony services are beginning to be offered in certain markets in 1995. PC modem services will become available in 1996 and interactive multimedia services are expected to become commercially viable in 1997.

Projected penetration of all U.S. households and estimated market share are shown in Table 1 (availability of these services would track the percentages in #1, from 1999 on):

TABLE 1 Projected Penetration of Advanced Cable Services in U.S. Households Percentage

Year	Cable Telephony	PC Network	Interactive Multimedia
1995	1	0	0
1997	2	4	0
1999	5	6	4
2001	10	8	10
2003	15	12	15
2005	20	18	25

3. Please outline the architecture(s) which will be used to build this broadband plant.

The architecture which will be used to build this plant will be hybrid fiber/coax architecture, as outlined earlier in our paper, with fiber nodes located in neighborhood with (on the average) 500 homes passed. Based on today's cable penetration, this would mean 300 cable subscribers per node, on average.

4. Please outline the peak switched digital bandwidth (in kbps or Mbps) available to an individual residential or small business user when you launch broadband service, and how that bandwidth can evolve to respond to increased peak traffic and to new, high capacity services (which may not now exist).

Peak digital bandwidth to and from the customer can be examined in the aggregate or on a service-by-service basis. In the aggregate, with hybrid fiber/coax architecture to 500 passings, we must make some penetration assumptions to have a meaningful answer. If we assume that cable maintains its current 60 percent penetration, this results in 300 customers per node. If we assume that 40 percent of those customers avail themselves of digital services of one kind or another, that results in 120 homes. If we further assume that peak usage of digital services is 33 percent (a conservatively high assumption), then the maximum number of simultaneous users being served by a given fiber node at any one time would be 40.

In cable systems as currently being upgraded, the spectrum from 50 MHz to 750 MHz is reserved for outgoing transmissions, but 50 MHz to 550 MHz is assumed to be reserved for broadcast analog services, with 200 MHz remaining for digital services. Assuming 356 QAM digital modulation, with an efficiency of 7 bits/Hz, this results in a total outgoing or "downstream" capacity of 1.4 gigabits/sec. This, when divided by 40 users at peak usage time, yields approximately 35 megabits/sec of outgoing bandwidth available per customer. Even given the expected volume of interactive video delivery, this number is ample for outgoing transmissions. It easily accommodates telephony, PC modem, and interactive multimedia applications.

Incoming bandwidth currently spans the spectrum from 5 to 40 MHz, a total of 35 MHz. Because of noise addition problems in this portion of the spectrum, modulation is probably limited to QPSK with an efficiency of about 2 bits/Hz, yielding 70 megabits/second. If this is divided by the 40 peak users to be fed from a fiber node, this yields 1.75 megabits/second available to each user. This is more than sufficient for telecommunications, including video telephony, and is sufficient as well for easily foreseen applications in PC network and interactive multimedia services.

Both the downstream bandwidth and upstream bandwidth can be increased in several ways. First, the number of customers per fiber node can be reduced on a neighborhood-by-neighborhood basis, based on usage, through the use of spare fibers (which are being installed) to subdivide nodes into smaller neighborhoods. In addition to spare fibers, wavelength division multiplexing can be used on a single fiber to the same end. It is easy to foresee average node sizes of 125 homes passed per node or less, resulting in at least a fourfold increase in the numbers cited above.

There is also the ability to dramatically increase the return spectrum. Time Warner's project in Orlando successfully makes use of 100 MHz of spectrum, from 900 MHz to 1 Ghz, in the return direction. This spectrum has fewer noise problems than the low frequencies cited above, so higher modulation efficiencies are possible. However, assuming the same kind of QPSK modulation used at the low frequencies, this would yield an additional 200

megabits/second of transmission capability, or 5 megabits/second per peak user. Again, this number can be multiplied through segmentation as outlined above.

It is possible to push these numbers significantly further. If very high speed, truly symmetrical capacity is required, frequencies above 1 Ghz can be used. The cut-off frequency of the coaxial cable employed is close to 2 Ghz, allowing for very significant expansion of capacity for high speed symmetrical services.

5. Please project the capital investment you or your industry plan to make on a per home-passed basis to install broadband infrastructure, and on a per subscriber basis to install specific services.

Our experience to date indicates that an investment of between \$125 and \$135 per home passed is required to upgrade existing coaxial cable television plan to the hybrid fiber/coax architecture referenced earlier.

Assuming a 15 percent penetration rate, we expect the incremental costs per customer moving into telephony to be no more than \$1,000 per customer. This investment is largely variable in nature, is made incrementally as telephony customers are added.

It is estimated that PC modem services will cost between \$400 and \$600 per customer, again, incrementally against only those customers taking the service. This covers the cost of the PC modem, as well as central routers, servers, gateways, and support systems.

It is estimated that interactive multimedia servers will cost between \$700 and \$800 per incremental subscriber, again accounting for terminal equipment in the home as well as switches, servers, and associated central investments.

6. Please respond to the concerns raised in Vice President Gore's letter (copy of letter attached) regarding the ability of users of your network to original content for delivery to any or all other users, versus the control of all content by the network operator.

The concerns outlined by Vice President Gore are largely addressed in our original paper. We expect to support several different coexisting networks on our broadband transmission system. These range from regulated common carrier-type symmetrical telecommunications services, like telephony, to highly experimental asymmetrical interactive entertainment services. In the middle ground will be a PC network, with great capacity. This network will be as symmetrical as it needs to be, given marketplace demand. As outlined above, we have the ability to expand network capacity in pursuit of the amount of symmetry that makes sense. However, premature installation of capacity and symmetry, in advance of demand, will be prohibitively expensive and, we believe, will not be supported by private investment.

7. Please specifically enumerate the actions which you or your industry believe that the federal government should take to encourage and accelerate the widespread availability of a competitive digital information infrastructure in this country.

We specifically address these points in our paper. To reiterate, they are:

- The elimination of historic state and local barriers to competition in telecommunications;
- The creation of requirements for interconnection, access, compensation, unbundling, collocation, pole and conduit sharing, and number portability and dialing parity by the incumbent telephony monopoly;
- The prevention of interference by local authority in the growth of competing telecommunications services; and
- The recognition that to enhance telephone competition, debilitating cable rate regulation must be reformed.

#### NOTE

1. All projections (unless noted) are the estimates of the authors and do not represent an official position of the National Cable Television Association or Time Warner Cable.

## 5

# Competing Definitions of "Openness" on the GII

*Jonathan Band*<sup>1</sup>

*Morrison and Foerster, Washington, D.C.*

The dozens of private-sector and governmental position papers concerning the emerging global information infrastructure (GII) all agree that it must be an "open" system. As the Clinton Administration recognized in its 1993 *Agenda for Action*, the information infrastructure

will be of maximum value to users if it is sufficiently "open" and interactive so that users can develop new services and applications or exchange information among themselves, without waiting for services to be offered by the firms that operate the NII [national information infrastructure].<sup>2</sup>

The position papers similarly agree that the desired openness could be achieved only through standardized points of interconnection (in technical parlance, interface specifications). In the words of the *Agenda for Action*:

To assure interoperability and openness of the many components of an efficient, high-capacity NII, standards for voice, video, data, and multi-media services must be developed. Those standards also must be compatible with the large installed base of communications technologies, and flexible and adaptable enough to meet user needs at affordable costs.<sup>3</sup>

Further, the position papers all agree that governments should not impose the standards; rather, the private sector should develop them. All concur that standards organizations will have a primary role in establishing the GII standards (de jure standards), but some acknowledge that many of the standards inevitably will be set by the market (de facto standards).

At this juncture, the position papers begin to diverge. The divergence arises over the issue of proprietary control of the standards: how much control, if any, should the inventor of the standard be able to exercise over the practice of the standard by others? The amount of control that can be exercised over a standard is inversely related to how "open" that standard really is. During the course of 1994, different firms and trade associations articulated disparate views on the issue of proprietary control of standards, reflecting their narrow commercial interests. During 1995, governments and international bodies began to assert positions as well. Although all proclaim fealty to the goal of openness, a given entity's definition of openness turns on the extent of proprietary control it would allow over GII standards. This paper examines the different definitions of openness that have emerged over the past 2 years.

### U.S. PRIVATE SECTOR DEFINITIONS OF OPENNESS

During 1994, many U.S. companies and trade associations began to express their opinions on "openness" on the information infrastructure. These opinions can be classified into four definitions ranging from the restrictive Microsoft definition to the expansive Sun Microsystems definition.

#### Microsoft

The Microsoft position starts from the assumption that hardware interfaces specifications are patentable and that software interface specifications can receive both copyright and patent protection. Microsoft further

believes that there is no need to create special rules for the GII; rather, the current laws applying to the software industry (which have served Microsoft extremely well) are adequate for the GII environment.

Microsoft will achieve "openness" by voluntarily licensing its application program interface to all third party application developers. Thus, anyone who wishes to develop a product that attaches to a Microsoft operating system will be able to do so. Microsoft, however, has not agree to license its interface specifications to firms seeking to develop operating systems that compete directly with Microsoft operating systems.<sup>4</sup>

Microsoft's narrow definition of openness—permitting attachment—flows from its business plan. Microsoft hopes to dominate the market for the operating system of the "set-top box"—the entry point of the information infrastructure into individual homes or businesses. By controlling the standard for the set-top box operating system, Microsoft will be able to exercise control over access to the entire infrastructure. Microsoft wants to encourage third party vendors to develop applications that will run on its operating system; the more applications, the more desirable the operating system becomes and the more likely that the market will adopt it as a de facto standard. At the same time, Microsoft wants to prevent the development of a competing set-top box operating system that is compatible with all the Microsoft-compatible applications.

### Computer Systems Policy Project

The Computer Systems Policy Project (CSPP), whose members include computer systems vendors such as Apple and IBM, shares many of the same intellectual property assumptions as Microsoft. Thus, it believes that hardware interfaces specifications are patentable and software interfaces specifications are both patentable and copyrightable. The CSPP differs from Microsoft in that it believes that special rules should apply in the information infrastructure environment. Specifically, the CSPP believes that the owner of an interface that is adopted as an infrastructure standard should be required to license it on reasonable and nondiscriminatory terms, not only to developers of attaching products but also to developers of competing products. That is, the interface specifications should be readily available to all vendors so that they could "build products that are compatible with both sides of the interface."<sup>5</sup> Further, the proprietor of an interface standard should be able to revise it only with timely notice or public process.

The CSPP position represents an underlying fear of Microsoft's market power. By requiring the licensing of the interface standards to the developers of competing as well as attaching products, CSPP hopes to prevent a Microsoft set-top box operating system monopoly even if the Microsoft interfaces emerge as the industry standard. Moreover, by permitting revision of standard interfaces only with timely notice, CSPP hopes to prevent the lead time advantages Microsoft's applications developers would otherwise have over third party developers. (These advantages are one of the subjects of the ongoing litigation over the Microsoft consent decree.)

The CSPP approach may work well enough for de jure standards set by a standards body. The standards body may, through negotiations, extract significant concessions from the proprietor. But what if the market, as opposed to a standards body, adopts Microsoft's set-top box operating system as a de facto standard? Pursuant to what authority will Microsoft be forced to license its interface specifications to competitors, or provide timely notice of upcoming revisions? Moreover, who will determine the "reasonableness" of the license fees demanded by Microsoft? Indeed, CSPP itself recognizes the shortcomings of its approach. It has conceded that in GII markets that are not competitive, the government may need to intervene "to ensure that critical interfaces are open."<sup>6</sup> Nonetheless, the CSPP continues to insist that the government "refrain from implementing compulsory licensing related to standards."<sup>7</sup>

### The American Committee for Interoperable Systems

The American Committee for Interoperable Systems (ACIS), whose members include Storage Technology, AT&T Global Information Solutions, Amdahl, and Broderbund Software, starts from a somewhat different intellectual property assumption than Microsoft and CSPP. While it agrees that hardware and software interface specifications are patentable, it doubts that many such specifications will meet the rigorous statutory

requirements for patentability. Moreover, unlike Microsoft and CSPP, ACIS believes that copyright cannot protect software interface specifications. It reaches this conclusion in reliance on the recent appellate court decisions in *Computer Associates v. Altai*<sup>8</sup> and *Lotus v. Borland*.<sup>9</sup> Further, ACIS believes that any incidental copying that occurs during software reverse engineering is lawful pursuant to *Sega v. Accolade*.<sup>10</sup>

ACIS, accordingly, believes that existing intellectual property law, as understood by ACIS, permits sufficient openness in the GII. Because few interface specifications would receive patent protection, and no software interface specifications would receive copyright protection, the firms that develop both de facto and de jure GII standards would rarely be able to exercise proprietary control over them. Additionally, the ability to reverse engineer the interfaces reduces the necessity of mandatory disclosure.

### Sun Microsystems

Sun, like ACIS, believes that copyright does not extend to software interface specifications. Nonetheless, Sun does not conclude that the current intellectual property laws provide sufficient openness on the GII. It fears the ability of the de facto standard inventor to obtain patents, to keep certain interfaces hidden, and to change the specifications without notice. Because of the importance of the information infrastructure to the world economy, Sun believes that the government should overcome these obstacles to openness by designating critical infrastructure interface specifications as "barrier-free." Other developers would have complete access to these specifications for nominal consideration.

### Incentives and Openness

Microsoft and CSPP argue that the ACIS and Sun positions—which at a minimum deny copyright protection for software interface specifications—would eliminate an important incentive for innovation. According to Microsoft,

Without the incentive offered by the ability to license intellectual property, the information infrastructure would not get built. R&D of the type needed to develop complex products like interactive television requires the investment of hundred of millions of dollars. Companies must be able to recoup those investments by licensing the rights to use the fruits of those investments. In addition, public domain standards give international competitors free ride on technology and intellectual property developed here in the U.S.

Similarly, CSPP states that "[d]evelopers of specifications for interfaces must be able to retain ownership of and benefit from the intellectual property that goes into the specifications, in order to maintain incentives to develop new technologies."<sup>12</sup>

Sun and ACIS reply to this criticism by drawing a distinction between interface specifications and interface implementations: "Interface specifications are pieces of paper; implementations are actual products or services."<sup>13</sup> Removing protection from an interface specification does not lead to removal of protection from a particular implementation of that specification. Indeed, proprietary implementations of nonproprietary specifications provide the bases for rigorous competition between providers: "This combination of nonproprietary interface specifications and proprietary implementations meets the imperative of balancing the requirement of providing incentives to developers of new technology with the societal need for interoperability along the information infrastructure."<sup>14</sup>



## GOVERNMENTAL AND INTERNATIONAL DEFINITIONS OF OPENNESS

Governments and international organizations began to adopt positions on GII openness in the weeks leading up to the G-7 Ministerial Conference on the Information Society at the end of February 1995. These positions are somewhat vaguer than those articulated by the U.S. private sector.

### EUROBIT-ITIC-JEIDA

In January 1995, information technology trade associations from Europe (EUROBIT), the United States (Information Technology Industry Council; ITIC) and Japan (JEIDA) met to create a joint position paper they hoped would influence the G-7 meeting the following month. The first section of the joint paper dealt with interoperability and openness:<sup>15</sup>

The key to interoperability is the development and implementation of open interfaces. An interface is open if its specifications are readily and nondiscriminatorily available to all vendors, service providers, and users, and if such specifications are revised only with timely notice and public process.

The joint paper stresses that those open interfaces should be the product of "private-sector-led voluntary consensus standards development processes."<sup>16</sup> It also draws a distinction between interface specifications and implementations: "Interface specifications provide the information and technical parameters for how systems, products and services communicate with each other, but should be limited to that information necessary to achieve interoperability, allowing suppliers to develop different implementations with distinguishing characteristics."<sup>17</sup> Nonetheless, the joint paper recognizes a role for proprietary technologies in GII standards: "When a proprietary technology is incorporated into a standard, the developer must voluntarily agree to license the technology on reasonable terms and conditions, demonstrably free of discrimination."<sup>18</sup>

The joint paper's views in general run parallel to those of CSPP: Intellectual property rights can reside in technology included in a GII standard, but the proprietor must license the technology on reasonable and nondiscriminatory terms and revise the specifications only with timely notice. The joint paper unfortunately also shares the infirmity of the CSPP paper of not explaining how the proprietor of a de facto standard will be required to make its interface specifications available to competitors. The joint paper, however, does not explicitly state that copyright protects software interface specifications. Thus, the references to proprietary technology could be interpreted as referring to patented hardware and software interface specifications. Indeed, the joint paper was probably intentionally left ambiguous on this point. While many ITIC and CSPP members historically have supported copyright protection for software interface specifications, many JEIDA members have opposed it. By discussing "proprietary technology" in the abstract, the joint paper could satisfy both constituencies.

### United States

In February 1995, the Clinton Administration's Information Infrastructure Task Force issued its *Agenda for Cooperation* amplifying on themes Vice President Gore had articulated in a speech in March 1994, to the International Telecommunications Union in Buenos Aires. One of the five core principles for the GII recognized by Vice President Gore and the *Agenda for Cooperation* is providing open access. The *Agenda for Cooperation* further recognizes that open access can be achieved only through interoperability and standardization:<sup>19</sup>

An essential technical element of the open access concept is interoperability, i.e., the ability to connect applications, services, and/or network components so that they can be used together to accomplish tasks. As the GII will be based on many different existing and emerging components at local, national, and global levels, it is imperative that these components be interoperable. The key to interoperability is global standards.

In listing recommended government action to achieve the goal of open access through global standards, the *Agenda for Cooperation* states that the U.S. should join with other countries to "[e]ncourage an open, voluntary standards-setting process that does not denigrate intellectual property rights...."<sup>20</sup> Like the EUROBIT-ITI-JEIDA formulation, the U.S. government appears to recognize intellectual property rights in standards adopted by standards bodies without taking a specific position on the copyrightability of software interface specifications. The discussion in the *Agenda for Cooperation* is so general, however, that it does not contain the protections included in both the joint paper and the CSPP paper: licensing on reasonable and nondiscriminatory terms, and revision with timely notice and public process.

Interestingly, a separate section of the *Agenda for Cooperation* appears to address this issue, as well as the problem of proprietary rights in de facto standards. When describing the need to create a flexible regulatory framework that fosters competition, the *Agenda for Cooperation* states that such a regulatory framework should clearly indicate:<sup>21</sup>

- The means by which new entrants can gain market access, e.g., ... licensing requirements ...;
- The nondiscriminatory terms and conditions of interconnection to an incumbent operator's network and of supplying information services over the network....

Here, the U.S. government appears to go beyond the joint paper's and CSPP's call for voluntary licensing; the quoted passage reflects an intent to mandate licensing by law. In other words, the *Agenda for Cooperation* takes the problem of proprietary rights in de facto standards seriously.

### European Union

The European Commission in September 1994, issued an "initial theme paper" for the G-7 meeting. The theme paper recognizes the importance of interoperability in permitting competition in the development of the infrastructure, which in turn will ensure that users receive the widest range of services at the lowest possible cost. To this end, "G-7 governments are invited to express their support for a consensual standardization process which is open, voluntary, and private sector-led."<sup>22</sup> The Commission notes that such a process "would help avoid two pitfalls. On the one hand, the unilateral imposition of mandatory standards by public authorities and, on the other, the emergence of de facto standards from monopolistic market positions."<sup>23</sup>

Although it is properly concerned about de facto standards and resultant monopolization, the Commission is unrealistic in its belief that standards bodies alone will solve the problem. Standards bodies tend to work slowly, and thus probably will not keep pace with the highly complex, rapidly evolving GII. Accordingly, de facto standards will emerge notwithstanding the best intentions of government and industry.

### G-7 Nations

At the conclusion of the G-7 Ministerial Conference, the G-7 parties issued a statement strongly supporting openness. The current regulatory framework must evolve to "put the user first."<sup>24</sup> The framework "must be designed to allow choice, high quality services and affordable prices."<sup>25</sup> This consumer welfare will flow from "dynamic competition,"<sup>26</sup> which in turn will result from "interconnectivity and interoperability."<sup>27</sup>

The G-7 parties specifically commit themselves to "[p]ursue the interconnectivity of networks and interoperability of services."<sup>28</sup> This goal will "be achieved through the promotion of a consensual standardization process which is market-led and which encourages open interfaces."<sup>29</sup> The G-7 partners recognize the need to accelerate the standardization process so that it can develop "timely and market-responsive standards."<sup>30</sup>

The statement does not explicitly discuss de facto standards. Nonetheless, it reflects a concern with the possible monopolization of the GII. Accordingly, the statement indicates that competition rules need to be interpreted and applied so as to encourage new entrants and promote global competition. Further, competition authorities must "shield[] against ... risks of abuse of market dominance."<sup>31</sup>

Moreover, the G-7 statement contains no reference to the incorporation of proprietary technology in standards. The silence on this topic appears to be a result of a compromise between the United States and the European Union.<sup>32</sup>

### CONCLUSION

The United States, the European Union, and the G-7 have all rejected the Microsoft definition of openness; in their view, openness requires that the standard interface specifications be available to all vendors, not only those developing attaching products. The United States and the G-7, at least implicitly, have recognized the shortcomings of the CSPP/EUROBIT-ITI-JEIDA approach; they understand that standards bodies alone cannot eliminate the danger posed by proprietary control of de facto standards. The G-7 solution appears to be vigilant antitrust enforcement. In contrast, the U.S. solution appears to be statutorily imposed compulsory licensing.

It is unlikely that Congress would require compulsory licensing for critical GII interface specifications. Even if it did, such licensing might serve as an effective remedy only if the government remained actively involved to arbitrate disputes concerning the reasonableness of license fees and the timeliness of disclosure. Further, active government participation might not succeed in preventing costly and time-consuming litigation over these issues. Vigilant antitrust enforcement has the same deficiency: It is cumbersome and resolves only one narrow dispute at a time.

For these reasons, the most effective approach may well be that of ACIS. If governments clearly affirmed the recent U.S. case law denying copyright protection for software interface specifications, much of the problem posed by de facto interface standards would vanish. To the extent problems remained on the margins—e.g., the occasional patent over an interface specification—compulsory licensing or antitrust enforcement could address them.

In any event, governments need to focus more attention on this issue immediately. The private sector has forged forward with the GII, already outpacing the standards bodies. As a practical matter, it will be much easier to solve the problem of de facto standards before the standards emerge on the GII. Once they do emerge, they will create vested interests willing to spend significant sums to resist any change to the status quo.

### NOTES

1. Jonathan Band is a partner in the Washington, D.C., office of Morrison & Foerster. Sections of this article appear in Band and Katoh. 1995. *Interfaces on Trial*. Westview Press, and the forthcoming Japanese language version of this book.
2. Information Infrastructure Task Force. *The National Information Infrastructure: Agenda for Action*. Information Infrastructure Task Force, Washington, D.C., September 15, p. 9.
3. Id.
4. See Myhrvold, Nathan P. 1994. "Interactive Video Systems," statement before the House Subcommittee on Telecommunications and Finance of the Committee on Energy and Commerce, February 1.
5. Computer Systems Policy Project. 1994. *Perspectives on the National Information Infrastructure: Ensuring Interoperability*. Computer Systems Policy Project, Washington, D.C., February.
6. Kerkeslager, Ellwood R. 1994. "Electronic Commerce and Interoperability in the National Information Infrastructure," statement before the House Subcommittee on Technology, Environment and Aviation of the Committee on Science, Space and Technology, May 26, p. 8.
7. Computer Systems Policy Project. 1995. *Perspectives on the Global Information Infrastructure*. Computer Systems Policy Project, Washington, D.C., February, p. 6.
8. 982 F.2d 693 (2nd Cir. 1992).
9. 49 F.3d 807 (1st Cir. 1995), petition for cert. filed (U.S. June 7, 1995).
10. 977 F.2d 1510 (9th Cir. 1992).
11. Myhrvold statement, p. 23.
12. Kerkeslager statement, p. 3.

13. See Rosing, Wayne. 1994. "Interactive Video Systems," statement before the House Subcommittee on Telecommunications and Finance of the Committee on Energy and Commerce, February 1, p. 2.
14. American Committee for Interoperable Systems. 1994. *Comment on International Aspects of the NII*, August, p. 4 n.2.
15. EUROBIT-ITI-JEIDA Paper (Jan. 1995), p. 5.
16. Id., p. 7.
17. Id., p. 5.
18. Id., p. 7.
19. Information Infrastructure Task Force. 1995. *Global Information Infrastructure: Agenda for Cooperation*. Information Infrastructure Task Force, Washington, D.C., February, pp. 14–15.
20. Id., p. 16.
21. Id., p. 17.
22. European Commission. 1994. *Initial Theme Paper*. September.
23. Id. Canada also seems to rely entirely on standards bodies to achieve the goal of interoperability. See *The Canadian Information Highway* (April 1994), pp. 13–14 and 23–24.
24. Documents Resulting from the G-7 Ministerial Conference on the Information Society, issued Feb. 26, 1995, *reprinted in Daily Executive Report* (Bureau for National Affairs, Washington, D.C.), Feb. 28, 1995, at M-5.
25. Id.
26. Id.
27. Id.
28. Id.
29. Id.
30. Id.
31. Id.
32. See Hudson, Richard L., and James Pressley. 1995. "G-7 Nations Make Gains in Facilitating Access to Information Superhighway," *Wall Street Journal*, February 27, p. A7A.

## 6

# Communications for People on the Move: A Look into the Future

*Richard C. Barth*  
*Motorola Incorporated*

### STATEMENT OF THE ISSUE

When Americans think of wireless communications systems, they often think only of cellular phones because of the explosive growth of this particular technology in many parts of the world. While that form of wireless communication is only one of many—paging, television broadcasting, police two-way radios, and many others come to mind—a closer look at wireless, and especially cellular technology and systems, is instructive regarding the growth overall of wireless communications, not just in the United States but around the world.

While seemingly ubiquitous in some business settings, the use of cellular phone and data communications systems is only in its infancy. Despite its dramatic growth, penetration rates for cellular technology-based systems in the United States are still running at less than 10 percent of American households. Compare this to penetration rates for another key component of the global information infrastructure, computers—PCs—which are estimated to be in 30 percent of American households. For many reasons the growth of cellular technology-based systems will likely continue to increase dramatically. This paper seeks to highlight the reasons for growth of wireless systems generally and cellular systems specifically. All of these changes represent a significant part of the evolving national information infrastructure (NII).

### BACKGROUND

There is no shortage of open issues in the NII debate. The categories of security and privacy, interoperability, spectrum availability, information access, ease of use, portability, ubiquity, network availability and manageability, applications development, multimedia, and network components provide just a partial list of the open NII issues. For our purposes here, the focus is on just two of these: portability and ubiquity. This focus is deliberate. Without satisfying these two requirements, the convenience, services, and applications that are visualized cannot be delivered, and the NII, rather than being a bold step forward, will in fact be a step backward.

After some 100 years of technological progress in communications, we live today in a world where voice communications are virtually ubiquitous. That means that today almost anyone can call—i.e., have a voice conversation—with almost anyone else anywhere at anytime. To do this with cellular and cordless technologies, the phones are locally or regionally wireless, and the wireless network that supports them is implemented by a parallel wired network that is highly complementary. There are some limitations in terms of access, costs, and competition, but recent private- and public-sector activities to implement personal communication services (PCS)—an extension of cellular technology—will go a very long way toward improving these limitations. Thus, holding private voice conversations will completely meet the anytime, anywhere standard of service.

In contrast to voice communications are visual or video-based services, which in many cases are still comparatively expensive and tightly controlled. Whereas anyone can make a phone call, only those designated can broadcast a television show or movie. Further, access to information and the role of computing in communications among people have been slower to develop. Only recently, with the advent of easy to use online services, are computers in the home, at schools, and in the workplace being used to supplement the papers, books, and access to learning services that remain pretty much as they were at the turn of the century—in news stands,

libraries, and schools. Radical changes in how people access data through computers are projected fairly confidently, at least in the developing countries.

The development of the technologies and networks that is best described by the NII is dramatically changing all of this, and by doing so is empowering all citizens with the conveniences and opportunities that will result from making all of the services personally accessible. With an NII, the world of "voice and imagery" are merging, along with more ready transfer of data, to meet the anytime, anywhere standard of service. Individuals will have full access not just to voice services as they do today, but also to image-based services and information services that are now only being imagined. This NII will have full mobility and connectivity that will be made possible by completing second-generation systems and bringing on the third-generation wireless systems that will become part of the NII.

Before getting into what this represents in terms of new functionality—efficiencies and services—it is appropriate to discuss why this vision could be at risk—that is, what could easily happen if vision and action don't match with the opportunity for portability that wireless technologies offer to the NII concept.

The promise of the NII lies in three synergistic forces—the availability of bandwidth brought on by developments in fiber and signaling, the availability of computing brought on by the microprocessor and the march of the semiconductor industry, and the emergence of competition and choice brought on by new telecom policies worldwide. The wireless component of these forces of technology is critical, especially next generation paging, cellular PCS, and dedicated systems used by public safety and critical industries.

Until recently, everything you could receive on your home wall-attached television, you could receive on your portable television, whether you chose to use it in another room, or on a camp out or while at a sporting event. That started to change with cable when the delivered wired bandwidth for television services was effectively increased by two orders of magnitude beyond that available in the radio frequency allocations for television. A similar shift occurred in computing over roughly the same time period. Early on, what you could do with a portable computer, or what we then called a portable computer, was pretty much what you could do with your office or home computer. That changed when local area networks (LANs) and computer networks came into being. With that transition, the portable computer became a comparative weakling to its LAN-based equivalent. These changes initially went unnoticed—after all, at least the new portable computer was portable, if a little out of touch, and who really needed 100 channels of television in any event?

Let us hold this perspective and move forward in time as the NII begins to deliver on its promise. People can talk face to face, and so groups can interact and decisions are made more quickly; families are united though they live miles apart; high-speed computing and information access are available in the home and office, and as a result people are more productive and better informed. Telecommuting becomes a reality, lowering energy consumption. But whereas in today's world most of the communications services that are available to a worker at a desk are available to a worker on the move, that is no longer necessarily true in the future—unless, that is, broadband wireless services are brought into line with broadband wired services.

This scenario prompts two questions: does it matter what is lost and what is gained, and, if it does, can it be done with the technology that is available and the other constraints that are likely to apply? The answer to both questions is yes.

### ANALYSIS AND FORECAST

Let us start with the first question, Does it matter? Broadly, we have already seen the high value people put on mobility. That value has generated vast new high-growth industries that not only have made the U.S. citizenry safer and more personally in touch, but also have made U.S. industry more efficient while driving substantial new export markets as well. But it is what happens in specific circumstances and industries that is perhaps more important. In other words, the applications must be carefully examined.

Many of the most interesting applications of wireless technology require the availability and dependability of private land-mobile communications—that is, the system dedicated to provide best-fit solutions to the communications needs and critical industries and protection of the public. These systems are a primary factor that has allowed the United States to establish and maintain its position as the world's leading producer of goods

and services. Private land-mobile radio is used by all segments of the industrial, business, public safety, public service, and land transportation mobile work force. Private land-mobile systems have become an indispensable part of the operation of this work force. The continued growth of this nation's commercial and public service activities demands additional communication capabilities. It is imperative that the industrial and public safety sectors have access to new imaging and decision processing/remote file access technologies. Even with the availability of some personal communication services offered by private and common carriers, public safety, public service, and industrial users will continue to satisfy their specialized communication requirements through private systems.

The private land mobile radio user community is a necessary ingredient in maintaining global competitiveness. Motivated by the constant need of the private sector to improve productivity and services, private users will invariably migrate to the specific communications solutions that provide the greatest advantage to their operations. An additional allocation of radio spectrum is essential if these users and their industries are to continue to flourish in increasingly competitive global markets.

### **Unique Communication Services Required**

Some of the unique services anticipated as being required to serve the critical day-to-day operational needs of critical industries and of public safety and public service organizations include the following.

#### **Crime Control**

- Mobile transmission of fingerprints, mug shots, warrants, and other images to and from law enforcement field personnel;
- Mobile transmission of maps, floor layouts, and architectural drawings for control of crime-in-progress operations;
- Tactical use of live mobile video for hostage, arrest, and surveillance operations;
- High-resolution graphics and electronic transfer of maps and other graphic information to police vehicles;
- Vehicle and personnel tracking systems;
- Locator service to address personnel security utilizing wearable devices containing wireless transmitters ("wireless dog tags"); and
- On-board information and security systems for mass transit vehicles.

#### **Energy Conservation and Management**

- Advanced distribution automation (remote monitoring, coordination, and operation of distribution and transmission components from centralized locations, including load management, advanced metering, and system control functions);
- Demand side management ("DSM") systems (e.g., managing the consumption of electric power and natural gas);
- Transmissions to monitor and record pipeline flow and pipeline pressure indicators; and
- Real-time monitoring, alerting, and control in situations involving handling of hazardous materials.

#### **Health Care and Fire/Emergency Management Systems**

- Remote monitoring of patients' vital signs in health care facilities to provide continuous patient monitoring and immediate response in the event of a patient crisis;

- Mobile transmission of maps, floor layouts, and architectural drawings to assist firefighters and other response personnel in the rescue of individuals involved in emergency situations;
- Transmission of visual signals and physician instructions in support of rescue operations;
- High-speed transmission of high-resolution medical imagery and data from paramedics to hospitals; and
- Automated inventory control.

### **Pollution Control**

- High-resolution graphics and electronic transfer of maps and other graphic information to mobile users;
- Management and remediation operations following spills or other crises;
- Real-time monitoring, alerting, and control in situations involving handling of hazardous materials; and
- Visual inspection of pipes and cables exposed during excavation projects.

### **Improving Industrial Productivity**

- Automatic transmission of messages advising of impending shortages of parts in a manufacturing environment;
- Vehicle and personnel tracking systems;
- Locator service to address personnel security utilizing wearable devices containing wireless transmitters ("wireless dog tags");
- Remote safety and security inspection of inaccessible locations;
- Automation of process and quality control functions;
- Transmission of scheduling and cost updates, job site inspection results, and performance assessments relating to construction projects; and
- Wireless face-to-face conferences between in-house production and sales personnel.

Many of these applications can be satisfied through the application of wireless technologies developed initially for the cellular market. There will also be a variety of special "niche" requirements that, by virtue of their highly specialized environment and exacting reliability requirements, will tend to be incompatible with consumer-oriented, carrier-provided PCS services that are evolving from the cellular technologies.

For example, a variety of advanced technology services will be required to ensure the safety and effective functioning of both underground and elevated transit and rapid rail transportation systems. In addition, there will be very specialized requirements for other critical industrial and public safety operations conducted in underground environments. Further, there will be a requirement for special broadband video and data systems designed to provide highly reliable communications networks in inherently dangerous settings. Private user emerging technology systems will fulfill a critical role in ensuring the safe and efficient functioning of maintenance crews and fuel and other service personnel working on highly congested flight lines.

### **Allocation of Spectrum**

Cellular and PCS spectrum allocations over the past several years have been critical to the introduction of some of these technologies. However, the expected rapid growth of wireless systems, based on user demand, requires that government policymakers assure a continued reallocation of spectrum to these needs. The recent spectrum allocation for PCS will not satisfy the need for spectrum for private emerging technologies. The regulatory scheme adopted for PCS makes it impractical, if not impossible, for private users to obtain and use their own PCS licenses for the new telecommunications technologies they need. Moreover, PCS carrier-licensees are inherently unlikely to offer the specialized solutions needed by public safety and critical industries.

Another factor that requires full analysis is the mobility of these new systems that are so critical to the evolution of the NII. While the overall agenda is for full mobility for the NII, its implementation fortunately



divides into two phases. The first requires immediate attention, and the second requires specific actions toward the turn of the century. The details are as follows.

### **Phase I for NII Mobility**

- *Spectrum for Second Generation Low Earth Orbit (LEO) Satellite Systems.* While the first generation of LEO technology is only just now being brought to market, it is not too early to plan for additional spectrum in anticipation of its success. We support the Federal Communications Commission (FCC) proposal to allocate an additional 70 MHz to allow for expansion of existing systems and the emergence of anticipated competitive systems.
- *Spectrum for Industrial and Public Safety Digital Systems with Broadband Capability.* It has always been a priority of the FCC to ensure that all needed spectrum for public safety and critical industry support is made available. As such, the tradition of support and forward-looking solutions for public safety and private industry is a long one that has been marked by the continued leadership of the United States. To prepare for the next series of needed changes, it is estimated that 75 MHz of spectrum is needed to deliver digital systems with broadband capability. These systems will not support continuous full motion video, but they will support selected slow scan video, image transmissions, file searches, building layouts, hazardous chemical mapping, and finger prints.

### **Phase II for NII Mobility**

Analog cellular, paging, and private systems provided the first generation. Digital systems that "remined" the existing spectrum, PCS, and the first phase of the NII mobility initiatives make up the second generation. Phase II for NII mobility makes up the third generation. Third-generation systems for private or public use and for data, paging, image, or voice provide similar functionality with flexible broadband capability, increased capacity, satellite system interconnectivity, and global roaming. These systems allow voice, but they also provide video. They support data, but they also support data at LAN rates. They deliver the full capability of NII to the mobile worker and the mobile person. Clearly, substantial spectrum will be necessary to support competing public systems, wireless cable access, and needed private systems with this capability. Efforts are just beginning to access how much spectrum may be needed and where that spectrum will be found in each of the world's regions and countries.

### **Growth of Wireless Systems**

Finally, let us focus on the key issue of the growth of wireless systems based on cellular technologies. As a percent of the world population, users of cellular technologies account for less than 1 percent. While growth rates for cellular systems have been running at 40 to 50 percent per year, that growth may well increase or at least continue for many years because of the large market opportunities that remain. By the end of this year alone, there will be over 70 million users of cellular technology worldwide.

Currently, the overall voice and telephony usage of telephony services is estimated at somewhat more than 4 trillion minutes, both wired and wireless. That will grow to nearly 8 trillion minutes by the year 2003. Correspondingly, the wireless component is now about 70 billion minutes at present, but that is expected to grow to 2 trillion minutes in 2003. So while it now amounts to a little over 1 percent of the total, predictions are that it will increase to 25 percent of the total. While that represents a tremendous opportunity, it is also a tremendous cause for concern if, as noted in the introduction of this paper, vision and action do not match this demand for spectrum and technology.

## RECOMMENDATIONS

### The Role of Government

The analysis presented above leads to the question, How can government help? First and foremost, government needs to accept what history has taught, that is, that mobility is essential. Wireless solutions need to be an explicit part of the NII agenda. An initial 165 MHz of spectrum if we leave in ITS, and 145 MHz if not, needs to be allocated for industrial and public safety services, IVHS and satellite services. Substantial additional spectrum will be required beyond that for third-generation systems. In fact the government itself has projected a need for approximately 250 MHz over the next 10 years for wireless terrestrial and satellite services. Government assistance needs to be focused on making spectrum available. "Re-mining" of existing broadcast television and broadcast auxiliary spectrum should be considered in light of the capability of Phase II systems to deliver both broadband data and video. Clearing the spectrum is not just a regulatory challenge. Solutions need to be developed to migrate existing services to either wireline or new spectrum.

The United States has led the world with its communications and computing visions in the past, and with mobility as part of the NII agenda it will do so again well into the next century.

### The Role of Standards

The network architecture of the NII must support its goal to facilitate true national mobility and connectivity. Open standards and interfaces should be adopted to assure a competitive supply of equipment to the operators and users of the NII. A level playing field and fair competition between independent device manufacturers will ensure affordable pricing and continued technological development. Naturally, the standards should address the need for privacy while allowing room for the innovative use of technology to provide access and security.

## SUMMARY

In conclusion, the NII features of portability and ubiquity are key to its success. But these aspects of the NII can be realized only if U.S. government regulators free up appropriate spectrum resources so that the private sector can develop these new markets.

7

## Building the NII: Will the Shareholders Come? (And if they Don't, Will Anyone Really Care?)

*Robert T. Blau*

*BellSouth Corporation*

More often than not, public policy debates concerning the national information infrastructure (NII) begin with a presumption that the proverbial "information superhighway" will be built regardless of what the government does. The only thing public policymakers really have to worry about, the reasoning goes, is to ensure that users, rival service vendors, and equipment vendors have affordable access to the nation's interoperable network of networks.

Many knowledgeable observers further assume that the telecommunications and cable television industries will move aggressively to upgrade their respective networks over the next 5 to 10 years. They presumably will do this to take advantage of new market opportunities spawned by interactive multimedia services, and to respond to competition. Still others, including several key federal officials, contend that the government will play a positive and constructive role in facilitating significant amounts of capital investment needed to extend broadband networks to households, businesses, schools, hospitals, and other public institutions throughout the country.<sup>1</sup>

This paper examines these expectations from the perspective of telephone company shareholders. Several key issues are addressed. How do returns on Bell company investment in local network facilities compare with returns on investment opportunities outside local telephone markets? Have shareholders been rewarded or penalized by Bell company decisions—given the prevailing regulatory environment—to upgrade their respective wireline telephone networks in recent years? On balance, do shareholder returns matter much to anyone other than the shareholder and telephone company managers, and if so, to whom, how, and why?

### BACKGROUND

#### Customer Expectations

In a recent Delphi survey, the consulting firm of Deloitte & Touche questioned 120 executives in the information, communications, and entertainment industries about how soon they expect various new communications products and services to arrive on the scene, and how rapidly markets for these products and

---

NOTE: Robert T. Blau, Ph.D., CFA, is executive director of policy analysis for BellSouth Corporation. He would like to thank Stephen Barreca, manager, infrastructure planning at BellSouth Telecommunications, for his valuable comments and assistance in developing aspects of this analysis that concern Bell company deployment of advanced network technologies. Views expressed in this paper are those of the author and do not necessarily reflect those of BellSouth or its subsidiaries.

services will develop.<sup>2</sup> With regard to the commercialization of interactive multimedia networks and devices, the panel came to the following conclusions:

- During the next 5 years, both local telephone companies and cable television operators will make significant progress in building new "infrastructure." A majority (54 percent) expect that advanced network technology will be available to 10 to 25 percent of telephone company residential customers, while a plurality of those surveyed thought 25 to 45 percent of all cable television customers would be served by upgraded networks by the end of the decade. Using Census Bureau population projections for the year 2000, this translates into a possible range of 10 to 25 million homes for telephone companies and 25 to 45 million homes for cable television operators.
- A majority (57 percent) of the executives surveyed believe that over a quarter of all schools, libraries, hospitals, and clinics will be connected to a fiber optic network by the end of the decade, and 23 percent believe that over 45 percent of these public institutions and agencies will have such a link.
- A majority (54 percent) of the participants believe that by 1998–2000 more than one-quarter of all U.S. households will have at least one computer with online capability, and 39 percent believe the penetration rate will be in the range of 25 to 45 percent of all households.

Interestingly, information industry executives surveyed by Deloitte & Touche also had a generally favorable view about the role of government in promoting investment in the NII. Approximately 62 percent thought the net effect of government actions by 1998–2000 will be positive (41 percent) or neutral (21 percent) "in terms of encouraging investment, fostering research and development, and promoting the rapid spread of advanced information, communications, and entertainment offerings."<sup>3</sup>

### Satisfying User Expectations

Are expectations for building the information superhighway realistic? In several key respects, the answers to that question will rest with those individuals who will be asked to put up the significant sums of capital needed to upgrade the NII, and particularly ubiquitous local telecommunications networks where the lion's share of these costs will be incurred.

If shareholders believe that risk-adjusted returns on investment in advanced network technologies will remain competitive with returns on alternative investment opportunities, then those technologies will be deployed and the new service features they make possible will be brought to the market in a timely manner. If, on the other hand, shareholders do not regard prospective returns on network investment to be high enough to compensate for risk incurred, then lesser amounts of discretionary capital spending will be committed to new network technologies. In that event, telephone company deployment of new technologies will slow, possibly to the point of lagging user expectations and needs. This could be especially problematic for developers and users of new multimedia service applications requiring substantially more bandwidth than is readily available over the public switched-telephone network (PSTN).

The balance of this paper analyzes relationships between shareholder returns, network investment, and the deployment of the types of advanced network technologies that will make up the NII. It begins with a discussion of recent trends in each and their implications for future investment. The paper concludes with a brief discussion of steps that telecommunications policymakers must take to create a more technology-friendly market environment.

## ANALYSIS

### Relationships Between Network Investment and Shareholder Returns

Figure 1 highlights relationships between regional Bell company (RBC) investment in regulated wireline networks and total shareholder returns between 1988 and 1994. Network investment is expressed as the percentage of Bell telephone company operating cash flow used to acquire wireline network plant and equipment. Unlike reported earnings, which are subject to the vagaries of different financial and regulatory accounting practices, operating cash flow provides an accurate measure of cash that a business generates after it pays its out-of-pocket operating expenses, taxes, and interest on its debt.<sup>4</sup>

Besides financing the acquisition of new plant and equipment for their regulated wireline networks, local Bell telephone companies principally use their operating cash flow in one of two ways: to pay dividends to their regional holding company shareholders, or to finance investment opportunities outside local networks. The ratio of wireline network investment to cash flow from local telephone operations, therefore, is a good comparative measure of how the individual companies—and their shareholders—view the relative attractiveness of using available cash flow to upgrade their local network facilities.<sup>5</sup> This is particularly true of the RBCs, since they currently finance nearly all capital expenditures (and pay all dividends) with internally generated funds (e.g., operating cash flow).

Figure 1 also depicts cumulative changes in total shareholder returns for each of the regional Bell companies between 1988 and 1994. Total shareholder returns include the percentage change in the price of an individual stock plus its dividend yield (i.e., dividends paid divided by the price of the stock at the time). For purposes of this analysis, cumulative shareholder returns are based on monthly returns and assume that all dividends paid on a particular stock are immediately reinvested in that stock.

Figure 1 highlights a definite inverse relationship between the ratio of network investment to operating cash flow from local telephone operations and total shareholder return. Differences in shareholder returns between the individual regional companies over the 1988–94 period also were quite substantial.

- If a shareholder had invested \$1,000 in a market weighted portfolio containing all seven RBC stocks on January 1, 1988, and reinvested all dividends paid, the portfolio would have increased in value to \$2,407 on December 31, 1994. This represents a gain of 141 percent, as compared with a cumulative return of 132 percent on the S&P 500. During this period, the seven RBCs reinvested 65.6 percent of their combined cash flow from their local telephone companies operations back into their regulated wireline networks.
- Between 1988 and 1994, three of the seven regional Bell companies—US West, BellSouth, and NYNEX—reinvested 71.3 percent of their local telephone companies' combined operating cash flow in wireline network plant and equipment. Had the same shareholder invested \$1,000 in these three stocks on January 1, 1988, this market weighted portfolio would have increased in value (assuming dividend reinvestment) to \$2,055, for a gain of 105 percent.
- During the same seven-year period, three other RBCs—Ameritech, Pacific Telesis, and Southwestern Bell—reinvested only 58.7 percent of their combined cash flow from local telephone operations in their respective regulated wireline networks. Had \$1,000 been invested in these three stocks on January 1, 1988, the value of this market weighted portfolio would have increased to \$3,019 by December 31, 1994, for a gain of 202 percent.

Given the size of the RBCs and the capital intensity of their local telephone operations, these differences in shareholder returns—and the emergence of an inverse relationship between capital spending on wireline network plant and equipment and shareholder returns—could have an important bearing on future investment in local network facilities. If the recent past is prologue, the level of discretionary capital spending on local wireline networks (i.e., capital expenditures over and above those required to maintain the quality of local telephone service at current levels) also will determine how rapidly broadband multimedia and other advanced service features are brought to the market.

### Relationships Between Capital Spending on Local Telephone Plant and Equipment and the Deployment of Advanced Network Technologies

Tables 1 and 2 highlight the degree to which differences in wireline network investment among the seven RBCs are reflected in how rapidly each of the regional companies upgraded their respective local wireline networks. The tables depict penetration rates and substitution ratios, respectively, for ten advanced network technologies that the Federal Communications Commission routinely tracks through its ARMIS Infrastructure Reports.<sup>6</sup> The RBCs and GTE file these statistics with the Commission annually, and they are currently available for 1989–93.

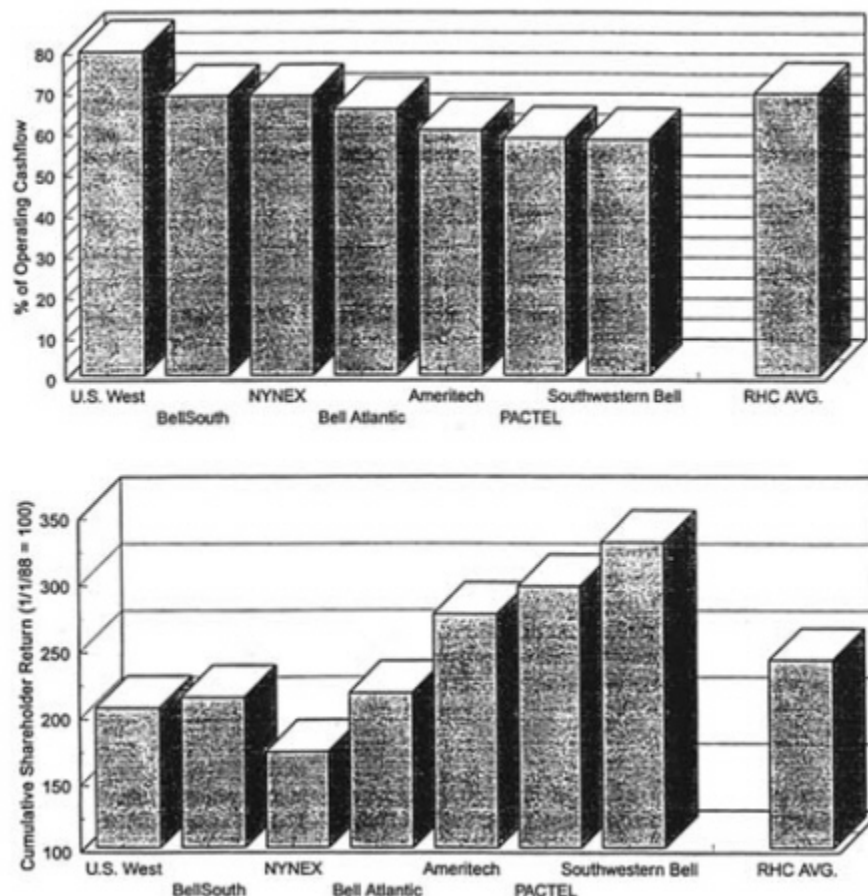


Figure 1 (Top) RHC Network Investment. Percent of Bell Telephone Company operating cash flow reinvested in local wireline networks between 1988 and 1994. (Bottom) RHC shareholder return index, 1988–94.

SOURCE: Compustat and One Source.

Penetration rates shown in Table 1 represent the percentage of a company's total number of switches, access lines, and communications channels equipped with a particular technology (e.g., ISDN, Signaling System 7, digital stored program control, fiber optics, etc.). Substitution ratios are based on the Fisher-Pry model commonly used to project ongoing increases in the penetration of new technologies expressed as a percentage of total usage.<sup>7</sup>

Technologies depicted in Table 2 also are categorized in three broad groupings designed to provide comparative measures of the following:

- *Digital connectivity*, which includes digital stored program control access lines, ISDN access lines, digital interoffice links, and fiber-equipped channels;
- *Deployment of fiber optic capacity*, which includes interoffice fiber links and fiber-equipped channels; and

- *Overall network modernization*, which reflects the deployment of all digital connectivity and fiber-optic technologies plus Signaling System 7 equipped access lines.

Substitution ratios for each of these groups were calculated by averaging the substitution ratios for the individual technologies that make up that group. These composite measures of technology deployment are presented along with the ratio of network investment to operating cash flow, and cumulative shareholder returns for the 1988–94 period.

**TABLE 1 Network Investment, Technology Penetration Rates, and Shareholder Returns**

Technology Deployment BOC Comparison (as of 12/31/93)	US West		NYNEX		BellSouth		Bell Atlantic		Ameritech		PACTEL		Southwestern	
	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank
Digital Stored Program Control Switches	74.8%	6	85.5%	3	86.0%	2	89.0%	1	84.8%	4	79.3%	5	73.4%	7
Digital SPC Access Lines	53.2%	5	74.4%	1	69.2%	3	69.8%	2	66.5%	4	53.0%	6	45.5%	7
SS7-317 Access Lines	72.4%	5	70.1%	6	94.2%	2	97.7%	1	79.8%	4	83.4%	3	64.2%	7
ISDN Capable Switches	11.5%	5	8.6%	6	19.3%	4	36.2%	1	26.3%	3	26.4%	2	6.3%	7
ISDN Access Line Capacity	29.0%	5	21.6%	6	39.5%	3	53.2%	1	46.0%	2	35.7%	4	11.2%	7
Fiber Sheath-Kilometers	6.7%	3	7.3%	2	6.6%	4	9.0%	1	6.2%	5	4.5%	7	5.5%	6
Fiber Links	75.7%	5	87.9%	2	88.5%	1	72.5%	6	83.6%	3	61.4%	7	76.1%	4
Digital Links	98.3%	2	88.1%	7	99.4%	1	95.8%	6	97.9%	4	96.9%	5	97.9%	3
Fiber Equipment Channels	2.9%	6	4.2%	3	11.4%	1	8.8%	2	4.1%	4	2.5%	7	4.0%	5
Fiber Working Channels	3.0%	6	3.9%	5	14.0%	1	9.2%	2	4.3%	4	1.8%	7	5.4%	3
Network Investment as % of Operating Cash Flow	79.4%	1	68.6%	2	68.5%	3	65.2%	4	60.1%	5	58.2%	6	57.8%	7
Cumulative Shareholder Return	205%	6	172%	7	213%	5	216%	4	275%	3	296%	2	329%	1

SOURCE: Federal Communications Commission, Compustat, and OneSource.

**TABLE 2 Network Investment, Technology Substitution Rates, and Shareholder Returns**

Advanced Network Technologies	US West		NYNEX		BellSouth		Bell Atlantic		Ameritech		PACTEL		Southwestern	
	Sub. Rate	Rank	Sub. Rate	Rank	Sub. Rate	Rank	Sub. Rate	Rank	Sub. Rate	Rank	Sub. Rate	Rank	Sub. Rate	Rank
Digital Stored Program Control Switches	0.30	6	0.30	5	0.27	7	0.38	1	0.33	3	0.32	4	0.37	2
Digital SPC Access Lines	0.23	6	0.27	5	0.28	4	0.35	2	0.37	1	0.29	3	0.17	7
SS7-317 Access Lines	1.18	7	2.68	2	1.24	6	1.95	3	1.65	4	1.37	5	0.15	1
ISDN Capable Switches	0.50	5	0.78	3	0.74	4	0.40	6	0.84	2	0.97	1	0.20	7
ISDN Access Line Capacity	0.22	7	1.05	3	0.87	4	0.30	6	1.19	1	1.11	2	0.51	5
Fiber Sheath-Kilometers	0.22	5	0.23	4	0.16	7	0.24	1	0.18	6	0.24	2	0.24	3
Fiber Links	0.86	2	0.57	4	1.42	1	0.39	6	0.55	3	0.21	7	0.45	5
Digital Links	0.72	2	0.29	7	1.09	1	0.59	4	0.64	3	0.37	6	0.39	5
Fiber Equip. Channels	0.31	#	0.23	#	0.38	1	0.28	2	0.05	#	0.38	#	0.38	#
Fiber Working Channels	0.36	#	0.29	#	0.35	1	0.33	2	0.10	#	0.36	#	0.39	#
Technology Composite	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank	Avg	Rank
Fiber	0.59	2	0.40	4	0.90	1	0.34	5	0.30	7	0.30	6	0.42	3
Digital Connectivity	0.37	6	0.46	4	0.66	1	0.38	5	0.56	2	0.54	3	0.36	7
Total (Network Modernization)	0.59	6	0.85	2	0.88	1	0.64	4	0.74	3	0.62	5	0.34	7
	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank	Percent	Rank
Network Investment as % of Operating Cash Flow	79.4%	1	68.6%	2	68.5%	3	65.2%	4	60.1%	5	58.2%	6	57.8%	7
Cumulative Shareholder Return	205%	6	172%	7	213%	5	216%	4	275%	3	296%	2	329%	1

SOURCE: Federal Communications Commission, Compustat, and OneSource.

Finally, Tables 1 and 2 rank order the seven RBCs by penetration and substitution rates for each of the individual technologies and the three composite technology groups described above, and by the ratio of network investment to operating cash flow and total shareholder returns.

As expected, those Bell companies that reinvested larger portions of cash flow from their local telephone operations in wireline network plant and equipment generally deployed advanced network technologies more rapidly. Figure 2 shows, however, that decisions by Bell company managers to accelerate the introduction of advanced network gear did not have a positive effect on shareholder returns. If anything, the opposite has been true. Between 1989 and 1993, for instance, BellSouth ranked first in overall network modernization, but only fifth in cumulative shareholder returns for the 1988–94 period.<sup>8</sup> Southwestern Bell, on the other hand, ranked last among the Bell companies in overall network modernization but first in cumulative shareholder returns.

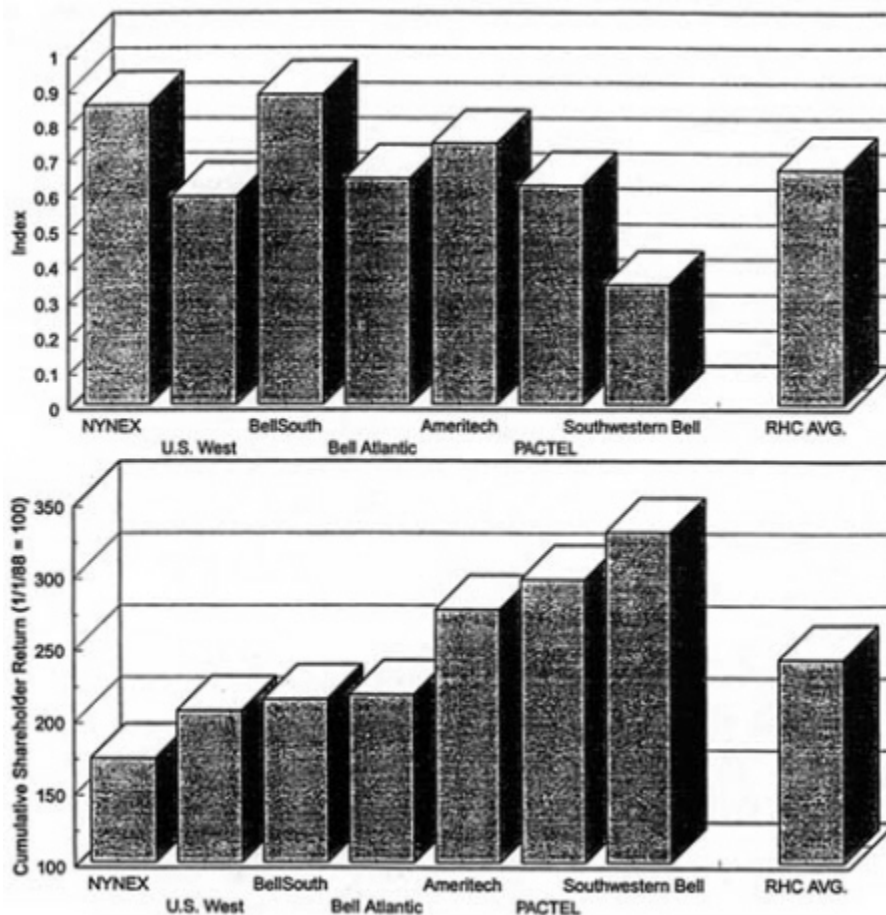


Figure 2 (Top) Network Modernization Index, 1988 to 1993. (Bottom) Cumulative Shareholder Index, 1988 to 1994. NOTE: Substitution rates measure how rapidly new network technologies replace old ones. Network Modernization Index is the average of substitution rates for the following technologies: Digital-SPC Access Lines, SS7-317 Access Lines, ISDN Access-Line Capacity, Fiber and Digital Links, and Fiber-Equipment Channels. SOURCE: FCC ARMIS Reports, Compustat, and One Source.

These relationships, of course, do not imply that shareholders have some inherent aversion to the Bell companies upgrading their portion of the nation's information infrastructure. What the data suggest, however, is that new or incremental revenue and earnings opportunities that investors expected to result from the deployment of wideband and broadband technologies have not been large enough, at least in recent years, to compensate for the capital cost and financial risk of installing these facilities sooner rather than later. This has been true even though the RBCs' local telephone operations accounted for roughly 86 percent of their total revenues between 1988 and 1994, 89 percent of their combined operating cash flow, 95 percent of their total earnings before extraordinary charges (e.g., write-offs of obsolete telephone plant and equipment), and 94 percent of their net income. (See Table 3.)



### Implications for Upgrading the NII

In view of the Bell telephone company contributions to the regional holding companies' overall earnings, why do shareholders seem at all concerned about using internally generated funds from local telephone operations to upgrade their wireline networks? To understand why this is so and what it means for future investment in advanced network technologies requires at least some appreciation of factors that investors consider in valuing common stocks. Besides current earnings and dividends, these factors commonly include earnings and dividend growth, and the degree of financial risk associated with owning a given company's shares.<sup>9</sup> And while the Bell telephone companies' current after-tax income might seem reasonable, if not ample today—either in absolute terms or as a percentage of the holding companies' net income—investor attitudes about their future earnings growth and business risk are decidedly less positive. This is due largely to an unsettled regulatory environment.

**TABLE 3 Earnings Growth Comparisons for the Regional Bell and S&P 500 Companies, 1988–94**

	1994	1993	1992	1991	1990	1989	1988	Avg. Ann. Growth (1988-94)
<b>REGIONAL BELL COMPANY AVERAGE</b>								
Total Revenue	12,623	12,031	11,762	11,514	11,382	10,983	10,600	2.95%
Operating Cash Flow	3,744	3,659	3,647	3,323	3,142	3,353	3,200	2.65%
Earnings Before Extraordinary Expenses	1,391	837	1,332	1,047	1,211	1,180	1,274	1.48%
Net Income	767	-250	814	808	1,211	1,187	1,274	-8.11%
<b>BELL TELEPHONE COMPANY AVERAGE</b>								
Total Revenue	10,683	10,389	10,087	9,894	9,794	9,606	9,457	2.05%
Operating Cash Flow	3,252	3,230	3,219	2,879	2,847	3,049	3,025	1.21%
Earnings Before Extraordinary Expenses	1,171	788	1,257	1,071	1,192	1,132	1,234	-8.7%
Net Income	591	-24	756	842	1,192	1,127	1,234	-11.55%
<b>UNREGULATED BELL BUSINESS AVERAGE</b>								
Total Revenue	1,940	1,642	1,675	1,620	1,588	1,377	1,143	9.22%
Operating Cash Flow	492	429	428	444	295	304	175	18.80%
Earnings Before Extraordinary Expenses	220	49	75	-24	19	48	40	32.86%
Net Income	176	-226	58	-34	19	60	40	28.01%
<b>S&amp;P 500 COMPANY AVERAGE</b>								
Total Revenue	7,718	7,015	6,846	6,689	6,701	6,209	5,713	5.14%
Operating Cash Flow	874	770	722	707	729	691	717	3.36%
Earnings Before Extraordinary Expenses	475	331	285	235	311	328	338	5.84%
Net Income	458	263	128	220	311	336	352	4.48%

SOURCE: Compustat.

Table 3 compares recent growth in average revenues, operating cash flow, earnings before extraordinary charges, and net income for the seven regional Bell holding companies, their local telephone companies, and their unregulated business operations. For comparative purposes, these same data are provided for the companies that currently make up the S&P 500, a broad market average. The data show that growth of the Bell telephone companies' after-tax earnings, whether before or after extraordinary charges, was negative and well below average earnings growth for the S&P 500 companies. By contrast, after-tax earnings from the regional Bell companies' unregulated businesses grew at an average annual rate of 28 percent over the 1988–94 period. This

explains why securities analysts now attribute as much as 36 percent of the regional holding companies' total market value to their unregulated, nonlocal telephone company operations.<sup>10</sup>

While lackluster earnings growth among the Bell telephone companies are attributable to several factors, including slow growth in the general economy between 1988 and 1994, two industry-specific developments stand out. One relates to competitive entry into the most lucrative segments of local telephone markets (e.g., exchange access and intra-LATA toll). Competition has limited growth in traffic carried over the Bell company networks while forcing the local telephone companies to reduce prices. The combination of modest traffic and access line growth and lower rates, in turn, has kept local telephone company revenues, earnings, and operating cash flow flat.

At the same time, legal and regulatory barriers to market entry like the Modified Final Judgement (MFJ) inter-LATA restriction, the video program prohibition, and until recently the MFJ information services restraint have walled off the Bell telephone companies from new revenue streams that could substantially hasten the recovery of capital invested in advanced network technologies. Investors understand this and are cautious about the Bell companies spending billions of dollars on network capacity that legal and regulatory barriers to market entry may prevent them from using.

Their concerns are compounded by expectations that deploying broadband technology in an increasingly competitive market environment will change network economics in ways that render telephone company earnings more volatile than they have been in the past, thereby increasing the risk of owning their stock.<sup>11</sup> Unlike today's copper based, narrow band telephone networks, tomorrow's wideband and broadband architectures will greatly increase bandwidth available to end users. Quantum increases in network capacity also will reduce the marginal or incremental cost of transporting communications traffic both in absolute terms and relative to the fixed cost of building and maintaining network capacity. Similarly, as telephone company operating leverage (i.e., the ratio of a firm's fixed costs to its total cost) increases, the addition, or loss, of traffic will have an increasingly pronounced impact on earnings since large portions of those gains or losses in marginal revenue will flow directly to the bottom line.

Prospects that local telephone companies could lose a significant portion of their local telephone revenues to rival vendors have been made more apparent in recent years by growth in demand for wireless communications services and the Internet. The WEFA Group recently forecast, for instance, that over the next 10 years, increases in network capacity available on commercial wireless communications systems (e.g., cellular telephone, personal communication services, and so on) will be large enough to accommodate not only natural growth in demand for mobile telephone services, but also nearly all narrowband voice and data traffic carried over wireline networks, as illustrated in Figure 3.<sup>12</sup> Should these forecasts pan out, wireless operators will no doubt attempt to leverage their "excess capacity" by integrating into traditional wireline telephone markets. If they do, price competition between wireless and wireline carriers will intensify, and wireless systems will capture voice and data traffic traditionally carried over wireline telephone networks.

The same could be said of the Internet. As Christopher Anderson of the *Economist* magazine recently observed.<sup>13</sup>

If the Internet does become a "data dialtone," regular dialtone might find itself out in the cold. It is already possible to make telephone calls on the Internet from specially equipped computers; the spread of multimedia PCS and faster Internet connections could make this commonplace. At the same time companies are turning an increasing amount of their telephone traffic into digital data and sending it through private data networks, saving up to half their telecom costs. Regulation permitting, this traffic could eventually move to the Internet. For the telephone companies, the only decision is whether they participate in the cannibalization of their revenues or watch it happen.

In light of prospects for increased competition, low earnings growth, and increased business risk, it is understandable why shareholders appear to have rewarded the Bell companies for minimizing discretionary capital spending on advanced network technologies in recent years. Overall, Bell company investment opportunities outside their regulated wireline networks have simply been more attractive in recent years primarily because these investments offer better earnings potential, often with less risk.

Though certainly of interest to investors and Bell company managers, the question remains whether inverse relationships between network investment and shareholder returns really matter much to anyone else. The answer, it turns out, will likely depend on growth in demand for access to the Internet, and the types of multimedia services that are just beginning to emerge on the World Wide Web.

### The Internet Community's Stake in Local Network Investment

According to the Internet Business Center, demand for access to the World Wide Web (WWW) increased at annual rates of 443,931 percent in 1993 and 1,713 percent in 1994, bringing the total number of WWW users in the United States to an estimated 4 million. The Internet Society further predicts that between June 1995 and January 2000 monthly traffic on the WWW (measured in bytes) will increase by a factor of 50,000! Although no one knows whether such forecasts will prove to be at all accurate, there is little question that the Internet will continue to expand at unprecedented rates for the foreseeable future.

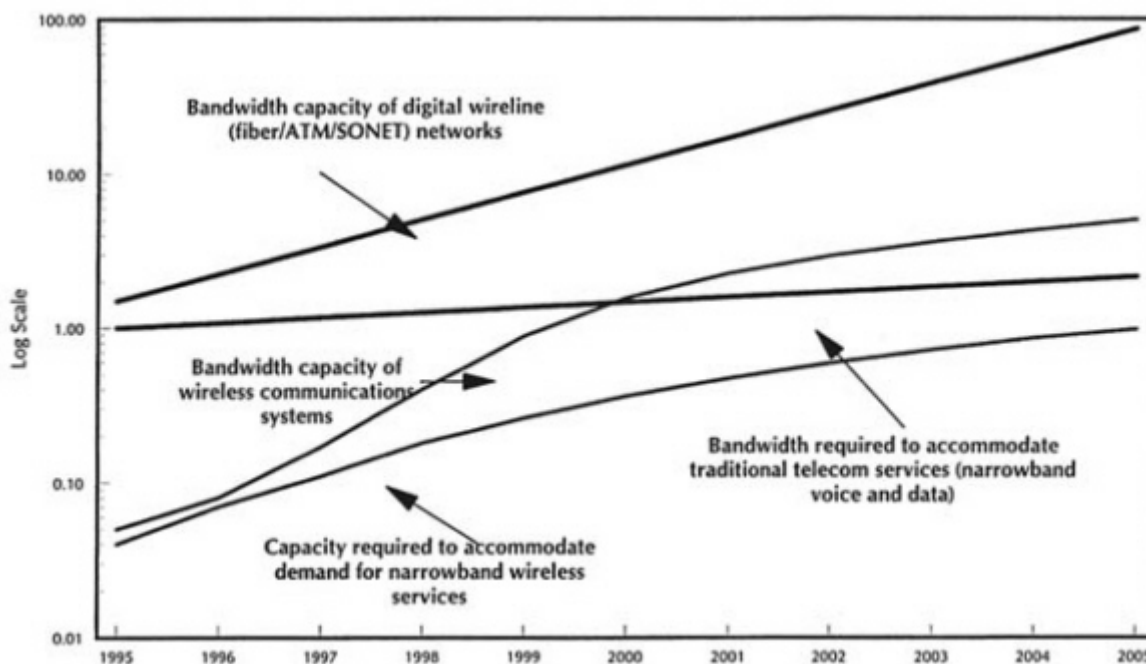


Figure 3 Bandwidth available on wireline and wireless networks, 1995 to 2005  
SOURCE: The WEFA Group.

Internet growth is of particular interest to local telephone companies for three key reasons. The first, as referenced above, has to do with prospects that significant amounts of voice and data traffic that traditionally have been carried over the PSTN could conceivably be routed over the Internet. Because use of the Internet today is effectively free, this shift would enable consumers to bypass the system of usage-based pricing arrangements and subsidies that the telecommunications industry has long relied on to recover its costs while keeping basic residential telephone rates at universally affordable levels (i.e., below cost).

At the same time, Internet traffic rides on network capacity leased from facilities-based telecommunications carriers. As such, growth in demand for services like the WWW also represents new market opportunities for local telephone companies. Routing Internet traffic should result in more intensive and more efficient use of embedded network facilities. If priced properly, it also could help recoup the cost of deploying

wider band technologies that Internet users will need if multimedia service applications available on the WWW are to continue to evolve.

Third, and perhaps most important, because multimedia applications on the WWW consume far more bandwidth than do electronic mail or other text-based services on the Internet, accommodating rapid growth in demand for access to the WWW could prompt local telephone companies to expand the capacity of their respective wireline networks significantly. The need for additional capacity should become all the more apparent as information resources available on the WWW continue to proliferate.

As new resources come online, demand for access to the WWW will increase along with its value to users as well as information service providers. Similarly, as the value of the WWW increases (e.g., by the square of the number of new users added to it during any given period), online sessions also should increase in duration (e.g., from current levels of 25 minutes per session versus an average of 5 minutes for local voice telephone call) for the simple reason that there will be more users and services to interact with. The combination of more businesses and residents spending more time online, accessing increasingly sophisticated multimedia services that require substantially larger amounts of bandwidth to transport, could press the limits of many local telephone networks within a relatively short period of time.

Near term, this demand for added capacity on the PSTN will be handled through the deployment of ISDN, asymmetrical digital subscriber loop (ADSL) equipment, and other advanced technologies that enable local telephone companies to increase bandwidth that can be made available to individual users over existing copper phone lines. Longer term, as demand for interactive full motion video applications develops, narrowband and wideband network technologies will need to give way to fiber optics or architectures that integrate fiber optics and coaxial cable capacity. Either way, the capital cost of accommodating WWW and other multimedia applications could be substantial.

Rendering a local telephone line ISDN capable, for instance, typically costs of \$100 and \$200 for lines already served by a digital switch, and between \$300 to \$500 if digital switch capacity has to be installed.<sup>14</sup> At year end 1993, Southwestern Bell had only 11.2 percent of its access lines equipped with ISDN, while 45.5 percent of its lines were served by digital switches. At these levels of penetration (and assuming ISDN installation costs fall at the midpoints of the latter two ranges), the total cost of making the company's remaining 88.8 percent of access lines ISDN ready would be roughly \$3.4 billion. As \$3.4 billion represents 42 percent of Southwestern Bell's total capital expenditures during the 1988–93 period, a commitment by the company to make ISDN widely available, say in the next 5 years, would constitute a major capital spending decision. Unless prospective earnings from its local telephone operations improve significantly, making such a commitment also would arguably be at odds with the company's fiduciary responsibility to maximize shareholder value.

### **PUBLIC POLICY RECOMMENDATIONS**

There is no question that the government will play a key role in balancing shareholders' interest in realizing reasonable returns on network investment, and consumer interests in seeing the PSTN upgraded. How public policymakers play their part also will determine when advanced technologies will be deployed, where, and on what scale. And though there are several steps they could take to ensure that the capabilities of the PSTN keep pace with user needs, three changes in telecommunications policies should prove especially helpful in this respect.

#### **Immediately Replace All Legal and Regulatory Barriers to Market Entry with User-Friendly Interconnection Requirements**

Given the less-than-favorable relationships between network investment and shareholder returns, coupled with substantial increases in bandwidth that the deployment of fiber optics, ISDN, ADSL, and other digital network technologies could bring about, it is imperative that federal and state officials immediately abandon any and all regulations that restrict how the PSTN is used. Notwithstanding the value of abundant bandwidth to the Internet community, other users, and growing numbers of information intensive industry groups that rely on

telecommunications networks to produce, distribute or add value to their wares, local telephone companies will not invest in advanced network technologies and capacity that they may be precluded from using by law or regulation.

Policymakers should recognize this fact and focus on developing a set of user-friendly network interconnection requirements that will encourage the development of a wider array of new service applications over which the capital cost of upgrading local telephone networks can be spread. Telecommunications legislation recently adopted by the U.S. Senate represents an important and much needed step in this direction. If enacted, the bill would allow the RBCs to provide inter-LATA long distance and video program services and to manufacture telecommunications equipment provided they comply with a 14-point checklist of interconnection requirements that will open up local telephone markets to more competition.

As Frank Governali, a leading telecommunications securities analyst with First Boston, put it:<sup>15</sup>

We'd love to see a bill passed. One, we think it would be good for the companies and the stocks to get this uncertain process behind us. Second, selfishly, we're tired of having to observe the inane activities of Washington so close-up. Although there is no chance that the current Senatebill is exactly the same as the one that may ultimately get [enacted into law], we think passage of a reasonably similar bill would be viewed positively by the market and cause many of the telecommunications companies' stocks to rise. By having a bill passed, investors could then observe the new operating environment and try to pick those companies that could do well in the new environment. Without a bill, the selection process becomes more difficult and the stocks more volatile, which is what we've seen over the past twelve months.

### **Complete the Transition from Cost Plus Rate-of-Return Regulation to Price Regulation**

In addition to replacing legal and regulatory barriers to network utilization with pro-competitive, user friendly interconnection requirements, federal and state officials need to complete the transition from rate-of-return regulation to a pure form of price regulation for noncompetitive telecommunications services. Moving to a pure form of price regulation also should have a favorable impact on efficient network investment in two important respects.

First, like telecommunications legislation, price regulation would eliminate a considerable amount of regulatory uncertainty that has discouraged network investment in recent years. Price regulation would accomplish this by eliminating any need for regulators to concern themselves with the types of capital expenditures regulated telephone companies are allowed to undertake, how local network costs are allocated between different services for rate-making purposes, or how rapidly new network plant and equipment is depreciated. Because these types of regulatory requirements are especially susceptible to being "gamed" by rival service providers for purely anticompetitive ends, their removal would make for a much more predictable and investment-friendly market environment.

Equally important, price regulation would give local telephone companies a much-needed opportunity to increase returns on network investment, while affording consumers much better protection from having to pay the cost of bad investment decisions by regulated carriers. Under price regulation, a local telephone company's shareholders would bear the brunt of ill-conceived or poorly timed investment in fiber-optic feeder plant or some other advanced technology, since price constraints would preclude the company from automatically passing these costs through to the ratepayer. At the same time, however, if a company's investment in advanced network technologies and capabilities succeeded in raising earnings, then those benefits would accrue to its shareholders, much as they would in any other business. Incentives to invest in new technologies would then improve.

Shifting investment risk from the local telephone companies' customers to their shareholders is especially important in today's market environment. This is primarily because consumer demand for many new multimedia services that telephone companies will rely on to help recoup the cost of deploying wideband and broadband network capacity will be far less predictable and, in all likelihood, far more volatile than demand for plain old

voice grade telephone service. Policymakers need to recognize that the emergence of a less certain and more volatile market environment will necessarily raise the risk of investing in new telecommunications technologies. And that unless regulatory pricing policies are adjusted in ways that compensate shareholders for bearing that added risk, capital expenditures on network facilities will slow, and network capabilities will likely lag behind user needs.

### **Allow Local Telephone Companies to Fashion New Business Models for Accommodating Internet Users**

Finally, if rapid growth in demand for Internet services continues to reshape the face of electronic communications, as many believe it will, local telephone companies and their regulators will need to experiment with new models for providing and pricing bandwidth that Internet users will no doubt want. What these models might entail is, at this point, unclear. Suffice it to say, however, that they will need to balance the telephone companies' need to recoup the cost of deploying advanced network technologies with the Internet community's desire to keep the Internet free of toll or other usage-based charges.

How these needs and interests ultimately get worked out will likely be accomplished through a considerable amount of trial and error. In some instances, regulators may need to give residents as well as business customers the latitude to pay part of the front end cost of installing advanced service features such as ISDN or ADSL in exchange for getting access to those features sooner and at lower monthly rates. In other areas, local telephone companies may want to provide customers free access to the Internet in exchange for agreeing to subscribe to a second telephone line for some specified amount of time.

In any case, new business models that telephone companies follow in responding to Internet user requirements may differ somewhat from ways local telephone service traditionally has been packaged and priced. Federal and state officials should recognize the need for flexibility in this area and allow any reasonable trials to go forward, preferably with minimal regulatory delay.

### **NOTES**

1. The estimated cost of deploying a ubiquitous broadband network in the United States ranges from \$1,000 to \$2,000 per business and residential access line depending on how much fiber and associated optoelectronics is deployed in telephone company feeder plant. See, for example, Egan, Bruce L. 1994. "Building Value Through Telecommunications: Regulatory Roadblocks on the Information Superhighway," *Telecommunications Policy* 18(8):580-583.
2. Of the 120 executives surveyed by Deloitte & Touche, 25 percent were with telecommunications companies; 24 percent were with broadcast or cable television firms; 21 percent were with consumer electronics or personal computer manufacturers; 14 percent were with publishing/advertising firms; and 16 percent were with entertainment companies. Fifty-five percent of the respondents were chairs, presidents, or chief executive officers, 30 percent were executive vice presidents or vice presidents, and the remaining 30 percent were senior managers. See Deloitte & Touche. 1995. *Interactive Multimedia Age II: Report on the 12-Question Update Survey*. Deloitte & Touche, May 1995.
3. *Ibid.*, pp. 2-3.
4. See Rappaport, Alfred. 1986. *Creating Shareholder Value: The New Standard for Business Performance*. Free Press, New York, pp. 19-45.
5. See Hackel, Kenneth S., and Joshua Livnat. 1992. *Cash Flow and Security Analysis*. Business One, Irwin, New York, pp. 138-214.
6. See Kraushauer, Jonathan M. 1995. *Infrastructure of the Local Operating Companies Aggregated to the Holding Company Level*. Federal Communications Commission, Washington, D.C., April.
7. See Fisher, John C., and Robert H. Pry. 1971. *Technology Forecasting and Social Change*. American Elsevier Publishing Company, New York.
8. Because there are time lags between capital expenditures on various technologies depicted in [Table 1](#) and their actual deployment, and additional lags between the deployment of those technologies and incremental sales/earnings growth that should result from their availability, it was appropriate to compare network investment/shareholder returns for

1988–94 with the Federal Communications Commission's technology deployment data, which is currently available for the 1989–93 period only.

9. The discounted dividend model holds that  $P = E(D)/(k - g)$  where

- $P$  = price of the stock;
- $E(D)$  = expected dividend to be paid in the next year (which is a function of earnings multiplied by the percentage of net income that the firm pays out as dividends);
- $k$  = the firm's cost of equity capital, which includes interest rates on risk-free Treasury bonds plus a risk premium that compensates investors for risk incurred by holding that company's stock; and
- $g$  = growth in dividends, which is a function of the firm's return on equity multiplied by the percentage of future net income that the firm is expected to retain and reinvest.

See Bodie, Zvi, Alex Kane, and Alan J. Marcus. 1989. *Investments*. Irwin, Homewood, Ill., pp. 473–480.

10. See Yanis, Steven R., and Thomas J. Lee. 1995. "The Regional Holding Companies Are More Than Plain Old Telephone Companies," *Telecommunications Services*, Oppenheimer & Company Inc., January 26, p. 8.

11. Increased earnings volatility adds to the risk of owning a share of stock, because it raises the probability that the company in question may not be able to pay future dividends or, for that matter, sustain day-to-day business operations. See Bodie, Zvi, Alex Kane, and Alan J. Marcus. 1989. *Investments*. Irwin, Homewood, Ill., pp. 130–143.

12. WEFA Group. 1995. *Economic Impact of Deregulating U.S. Communications Industries*. WEFA Group, Burlington, Mass., February, p. 29.

13. Anderson, Christopher. 1995. "The Internet Survey," *The Economist*, July 1, p. 18.

14. See Egan, Bruce L. 1994. "Building Value Through Telecommunications: Regulatory Roadblocks on the Information Superhighway," *Telecommunications Policy* 18(8):580–583.

15. See Governali, Frank. 1995. *Weekly Industry and Valuation Report*, First Boston, June 16.

## 8

# The Electronic Universe: Network Delivery of Data, Science, and Discovery

*Gregory Bothun, University of Oregon*

*Jim Elias, US West Communications*

*Randolph G. Foldvik, US West Communications*

*Oliver McBryan, University of Colorado*

### ABSTRACT

When technologists discuss the "information highway," their conversations usually revolve around high-end technologies such as "broadband" and "asynchronous transfer mode" (ATM). Unfortunately, this focus on the exclusive use of very high bandwidth solutions fails to reckon with the limitations confronting many learning institutions, and especially most of the K-12 environment, in gaining access to the "highway" or even exploring their options.

This paper describes a set of advanced networking tests that were conducted in Oregon and Colorado during 1994 and early 1995. High-speed ATM capabilities were effectively merged with readily available T1-based connectivity to deliver affordable information highway capabilities to K-12 educational institutions. Innovative collaborations between the university and K-12 level served to accelerate the introduction of the new technology to the K-12 level. A range of application capabilities were effectively delivered over the integrated ATM/T1-based infrastructure. Multimedia workstations were used for video to the desktop, desktop video teleconferencing, and access to multimedia servers via Mosaic and the World Wide Web.

The tests that were conducted, and the results and trends, are covered in detail in this paper. These findings are then used to suggest technology deployment models for the next 5 to 7 years.

### STATEMENT OF THE PROBLEM

The concept of "network everywhere" is a goal that is achievable through cooperation between education, industry, telecom carriers, and the government. But why is this important? Its importance lies in a paradigm shift in how the individual interacts with and views the world. Almost 300 years ago, the philosopher John Amos Comenius penned these words: "It is lamentable, utterly unjust and insulting that while all men are admitted to God's theatre, all are not given the chance of looking at everything."

This is a description of an inclusive philosophy of education that has never really been implemented. Instead, education has focused intensely upon the exclusive theme of "educating each according to his or her needs." Ironically, K-12 education is now in a position of reviving this old idea through the use of new technology: the high-speed network. High-speed network connectivity will provide the K-12 classroom with a vast array of resources in all disciplines.

Is access to high-speed networking a reasonable goal for the K-12 classroom? Unfortunately, the high media profile of fiber optics, high bandwidth applications, and advanced communications technologies has distracted many institutions from exploring other options for access. This is the case with most of the K-12 environment. There may be other ways for them to deal with some of the limitations confronting them. This is particularly true in most of the western states, where low population densities and correspondingly high deployment costs prevail. These challenges created an opportunity to explore the limits and possibilities of defining a high-speed ATM trial network for the educational community that could be integrated with the existing



geographically dispersed public network. This trial environment offered an opportunity to explore existing biases about fiber optic and high-speed networks, overcome the constraints of distance and dispersed population centers, and bridge the ever widening educational gap between technological "haves" and "have nots."

Education and lifelong learning are primary keys to societal success on both a personal and a national level. As our information-oriented society continues to move rapidly forward, tremendous pressures are exerted on schools at every level of learning. The rate of information acquisition in learning and scientific inquiry is astounding, making it extremely difficult for teachers and school districts to keep up. The problem is particularly difficult for both rural and urban schools with limited enrollments, diverse student populations, and constrained budgets. Many students in these types of schools are being taught advanced topics by minimally qualified teachers with outdated textbooks. These students are placed at a decided disadvantage if they advance to the college level and are forced to compete with students coming from schools with advanced curricula based on more current information. Worse yet, certain subjects may not even be taught in some schools due to lack of resources or expertise.

The trial activities described in this paper support the premise that these shortcomings can be addressed by means of a high-speed network to integrate communications transport at a variety of speeds and bandwidths and effectively partner university-level experts with teachers and students in the K-12 system. Such a network, both technological and human, can allow for a more efficient delivery of information and curricular material at both the university and K-12 levels.

Beginning in April 1994 and extending through March 1995, US West engaged in a set of technical trials of asynchronous transfer mode (ATM) technology in western Oregon and in Boulder, Colorado. Partnering with several leading universities and a number of other organizations, these trials explored issues surrounding the use of advanced networking technologies in combination with existing network services. Issues that were addressed included network platforms and architectures, along with an understanding of the types of applications that would use such an advanced mix of networking technology. Many planners assumed that high-end niche applications associated with such things as supercomputers would predominate. However, we soon realized that trial participants were placing a strong emphasis on the extension of advanced technologies to secondary schools. Even as that trend began to emerge, some felt that while extension of advanced technologies to secondary schools might be technically feasible, economic and social factors would make such applications unworkable.

Innovative work conducted separately by the University of Oregon and the University of Colorado proved the skeptics wrong. Advanced capabilities were in fact extended to a number of secondary schools in both Oregon and Colorado with encouraging results. This paper discusses the experiments that were performed, the trends observed, and subsequent plans made for follow-on activities. Insights gained are used to project a baseline of technologies that could be deployed over the next 5 to 7 years. We believe that these results should influence the deployment of advanced technology to secondary classrooms and could serve as a model for effective future cooperation between universities and K-12 schools.

## BACKGROUND

US West announced a multiphase ATM strategy in October 1993.<sup>1</sup> Key elements of this strategy included small, scalable ATM switches flexibly and economically deployed in a distributed architecture, as was done in the western Oregon and Boulder, Colorado, trials. The results were positive, and US West subsequently announced availability of an ATM-based Cell Relay Service offering in its 14-state region in January 1995.<sup>2</sup>

Experimentation in Oregon was conducted in conjunction with the Oregon Joint Graduate Schools of Engineering as part of "Project NERO" (Network for Education and Research in Oregon).<sup>3,4</sup> Five widely dispersed graduate-level engineering schools (Oregon State University, Oregon Graduate Institute, University of Oregon, Portland State University, and Oregon Health Sciences University) and several state office buildings were linked together via a network of ATM switches and associated OC3c and DS3 lines in several major cities in western Oregon. In addition, connectivity was also extended to a teacher training workshop at school district headquarters in Springfield, Oregon. Experimentation in Boulder, Colorado,<sup>5</sup> was conducted in conjunction with

the University of Colorado<sup>6</sup> and included three government laboratories, four industrial laboratories, a public library, four public schools, and even a private residence.

## ANALYSIS OF PRIMARY TRENDS

### LANs at All Locations

All major trial participants had an extensive imbedded base of legacy LANs. For most large organizations, the LAN environment has expanded rapidly over the past 10 years. For example, less than 10 years ago the campus of the University of Oregon in Eugene had only one Ethernet LAN serving about 50 host computers. Today the campus has more than 80 LANs spread across 70 buildings supporting 6,000 host computers. A complex infrastructure of LANs, bridges, and routers has evolved over the intervening years to support these networks, which have become an integral part of the University of Oregon environment.

The level of sophistication of these campus LANs continued to increase as the trial progressed. Typical campus configurations eventually included a variety of ATM switches providing gateways to remote locations and connectivity between central routers. Direct ATM connectivity was extended to workstations and hosts to allow them to communicate directly over ATM using native IP per RFC 1577<sup>7</sup> as well as to distributed ATM hubs supporting 10 BASE-T and 100 Mbps Ethernet for devices using LAN emulation.<sup>8</sup>

Based on this high level of sophistication, one might make the mistaken assumption that LANs are therefore too sophisticated for school districts to deploy. However, we found that the trial school district and library locations had numerous LANs already in place. For example, the School District Headquarters in Springfield, Oregon, has an Ethernet LAN supporting a number of high-end workstations. A Macintosh lab in a school in Albany, Oregon, has a mix of Local Talk and Ethernet LANs supporting 65 Macintosh devices. The four participating middle schools and high schools in Boulder, Colorado, all have Ethernet LANs supporting a mix of devices. Even the Boulder Public Library has an Ethernet LAN with a sophisticated set of routers and servers supporting local and remote access to online card catalogues and banks of CD-ROMs. While the existence of LANs may not be typical for all K-12 schools, their presence at our trial participant locations is an important indicator of future direction. Local area networks are quickly becoming a plug-and-play item that routinely comes with off-the-shelf workstations and can be installed by the local school administrator, the local library staff, the home user, or (in the case of the Issaquah school district near Seattle) by the students.

### IP Networking and Mix-and-Match Telecom

In addition to the universal presence of local area networks, all trial participants accessed the network via routers and TCP/IP. Typical high-end configurations utilized a router to access the public wide area network via either a DS3-based (45 Mbps) or an OC3c-based (155 Mbps) ATM user network interface (UNI). The router often worked in conjunction with local ATM switches as previously discussed.

Existing distributed Internet infrastructures in both Oregon and Colorado are typically based on TCP/IP routers that are interconnected with T1 lines running at 1.5 Mbps. The university trial customers used ATM to construct new Internet infrastructures still based upon TCP/IP routers, but with 155 Mbps rather than 1.5 Mbps interconnectivity. Connectivity at 155 Mbps was then extended to a central campus hub where it was distributed in turn to a few individual desktop computers. By using ATM with routers and TCP/IP, the trial participants were able to gain a 100-fold increase in raw speed and aggregate carrying capacity without changing their basic protocol structure. One might assume that high-end transport options would preclude the participation of secondary schools; however, this did not prove to be the case. The use of IP-based networking provided the ability to cost effectively bring secondary schools into the networked environment.

Routers emerged as the universal anchors to enable the new ATM transport option without affecting existing applications and infrastructures. Routers support most types of network connectivity, providing the ability to transparently mix and match various types of networking options in different parts of the network. With

the central dominance of IP-based routers, telecommunications transport became a transparent commodity to be mixed and matched based on cost-performance analyses for each individual location.

Using this extreme level of flexibility, the trial participants proceeded to deploy ATM in their backbone networks and their central locations where high aggregations of traffic were important, while extending trial connectivity to outlying schools and libraries via less expensive T1 or frame relay links. All of the secondary schools were in fact brought into the new environment via lower-cost T1-based telecommunications links in the existing copper-based public switched network. The universal use of IP-based routers to support all types of telecommunications transport allowed the creation of sophisticated "internets" employing the complete spectrum of telecommunications options from low speed to high speed. This is an extremely important trend to consider as we move forward into the next 5 to 7 years.

### **Multimedia Workstations**

All trial participants made extensive use of multimedia workstations. These workstations provided the ability to receive audio/video broadcasts at the desktop, to engage in desktop video teleconferencing, and to access multimedia servers. Initial testing was primarily with high-end SGI or Sun workstations with multicast routing capabilities built into their UNIX kernels. One could easily assume that using high-end workstations would preclude the participation of secondary schools. However, this was not the case. In some situations the expedient option of loaning high-end workstations to the individual schools was adopted so that proof-of-concept work could proceed. As the trial proceeded, other less expensive terminals became available, such as X Window terminals and AV-capable Macintoshes. The use of multimedia workstations from the secondary schools to the university environments was an extremely important trend observed during the trial, along with the increasing availability of affordable multimedia-capable workstations.

### **Broadcast Video to the Desktop**

The broadcast of educational video directly to the desktop was accomplished via M-bone (multicast backbone) capabilities.<sup>9,10,11</sup> M-bone originated from an effort to multicast audio and video from Internet Engineering Task Force (IETF) meetings.<sup>12</sup> By late 1993 M-bone was in use by several hundred researchers located at approximately 400 member sites worldwide. The M-bone is a virtual multicast<sup>13,14,15</sup> network that is layered on top of portions of the physical Internet. The network is composed of islands that can directly support IP multicast (such as Ethernet LANs), linked by virtual point-to-point links called "tunnels." More recently, several vendors have supplied native IP multicast routing protocols. Multicast routing information was distributed using PIM (Protocol Independent Multicast).<sup>16,17</sup> Desktop video delivery was initially accomplished via the M-bone tool "nv," with later migration to a newer video tool called "vic" that supported not only nv but also other video encodings and standards such as H.261 and JPEG.

M-bone has been routinely used for some time to broadcast various events to individuals' desks via the Internet. Examples include "NASA Select," the NASA in-house cable channel broadcast during space shuttle missions, and "Radio Free VAT," a community radio station. "Internet Talk Radio" is a variation on this technique to conserve network bandwidth by making prerecorded programs available as data files to be retrieved via ftp and then played back on the local workstation or other appropriate devices as desired. This can have a powerful impact on the inclusion of current, nearly up-to-the-minute information in curriculum development. Programming as varied as talks by Larry King and the "Geek of the Week" are made available via this technique.<sup>18</sup>

These broadcast services have been significantly expanded recently via the Internet Multicasting Service (IMS), which broadcasts live multicast channels on the Internet such as satellite-based programming like Monitor Radio and World Radio Network, network-based programming such as CBC News and SoundBytes, and live links into the National Press Club, the U.S. Congress, and the Kennedy Center for the Performing Arts.<sup>19</sup>

The Oregon universities used these capabilities to broadcast computer science colloquia out over the Internet. They found that bandwidth as low as 56 kbps was adequate for broadcast of the image of an instructor, while 500 to 750 kbps was required for proper transmission of text images with hand-pointing motions. They also extended experimentation to the actual teaching of an engineering statistics class to several registered students as an alternative to attendance, using specially equipped video classrooms.

Broadcast of various events via M-bone proved to be very popular at the four secondary schools in Colorado. Every time the space shuttle flew they tapped into the 24-hour-a-day M-bone broadcasts of the live space shuttle activities and monitored the shuttle flight with great interest. A window was brought up on the workstation screen on which the students could watch live pictures of the space shuttle as the mission proceeded. In November 1994 the students monitored the "Prince Charles Video Event" from Los Angeles via M-bone. In addition, movies precompressed via MPEG were broadcast out over the network. In December, an MPEG-compressed portion of a Star Trek movie was successfully broadcast out at 30 frames/s via schools' T1 connectivity and was then decompressed at the receiving workstations. Based on this success, it has been proposed that MPEG-compressed university courses could be delivered to desktop workstations at secondary schools with minimal T1-based access capabilities. Experiences in Colorado showed that only 128 kbps was generally needed for typical M-bone broadcasts, while full-frame 30 frame/s VHS quality video functioned very well via MPEG compression at T1 speeds. One of the more creative uses of M-bone broadcast capabilities was an "end of Boulder trial party" broadcast out over the Internet from a residence directly connected to an ATM switch. A 3-week-old baby at the party was given the distinction of being the "youngest person to ever do an Internet broadcast."

### Desktop Video Teleconferencing

Closely related to the delivery of video to the desktop is the use of desktop workstations for video teleconferencing. Desktop video teleconferencing emerged as a strong trend in both the Oregon and the Boulder trials. M-bone proved itself to be extremely useful and versatile for desktop teleconferencing and collaboration, as did InPerson software from SGI and CU-SeeMe software for Macintosh computers from Cornell University. Other similar software packages are also becoming more readily available on the market. In general, this type of software supports simultaneous conferencing with as many as four or five other people, along with the ability to share application and white-board windows.

These capabilities were used extensively and successfully in Colorado on a daily basis for four- or five-way conferences. T1-based connectivity was adequate for this type of conferencing, although more efficient compression by the associated software would be helpful. Dr. McBryan used the InPerson software to teach classes at the four Boulder secondary schools. He networked with the four public schools, each of which had an SGI Indy workstation in a classroom. The students then clustered around the Indy workstation while Dr. McBryan taught the class from his home. He taught classes on how to write World Wide Web (WWW) home pages so that the schools could come online with their own servers. All four schools subsequently came online with their own servers based on what they had learned in these online classes. This was an effective demonstration of cooperative work between universities and secondary schools via the network. The capability was so successful that the schools wanted to obtain a large-screen projection capability to make it easier for students to participate and to see the screen. It is significant to note again that this was done entirely with T1-based connectivity on the existing copper-based public infrastructure as well as fiber optics.

In October 1994 a similar experiment was conducted in Oregon at a demonstration of the "21st Century Classroom" for the ITEC Expo's "Computer Office and Automation Show" in the Portland Convention Center. A group of students from Grant High School in Portland participated in a live audio/video exchange with faculty in the Physics Department of the University of Oregon. During this exchange the students were shown how to use a software package called Mosaic and were then able to "sit in" on a physics lecture delivered at the University of Oregon. The students had full two-way audio/video capabilities. They could see the lecturer and the other students and could ask questions during the lecture.

Other Colorado applications included the ability for a student ill at home to interact with teachers and other students. Similar capabilities were used at the college level for off-hours consultation between instructor and students, with video consultation as an enhanced alternative to the current electronic mail procedure. All of these applications have direct value in delivering more timely and customized learning to students in dispersed locations.

### **Multimedia Servers, Mosaic, and the World Wide Web**

The final trend is closely related to the use of desktop workstations for receipt of teleconferencing and video broadcasts. Desktop access to multimedia servers was an extremely important area of emphasis in both trials and was successfully extended to the secondary schools. The National Center for Supercomputing Applications (NCSA) at the University of Illinois developed an environment called "NCSA Mosaic"<sup>20,21</sup> that enables wide-area network-based information discovery and retrieval using a blend of distributed hypermedia, hierarchical organization, and search functionality. NCSA has developed Mosaic client software for X Window Systems, Macintosh, and Microsoft Windows. NCSA Mosaic's communications support is provided courtesy of the CERN World Wide Web (WWW) project's common client library.<sup>22</sup> No longer isolated within the academic community, the WWW has emerged as one of the Internet's most popular technologies. Commercial companies now advertise the availability of their "home pages," associated software is readily available, and commercial bookstores are stocked with numerous basic how-to books on Mosaic and the WWW.

As used in the Oregon and Boulder trials, the WWW servers were accessed via the Internet and controlled a set of self-paced courses that contained a mixture of video, audio, and text. A student would take a particular course via a multimedia workstation, with the course consisting of multiple windows on the screen, including full motion film clips with audio, along with associated interactive text. The multimedia courseware is delivered over the WWW and is received and viewed using a WWW reader such as Netscape or Mosaic.

### **Examples of Technology Uses**

The combination of IP networking capabilities, multimedia workstations, video to the desktop, video teleconferencing, and multimedia servers was used in creative and effective ways in both trials. During the summer of 1994 a teacher training workshop was held for Springfield (Oregon) School District science teachers on the use of multimedia X-terminals as network information retrieval machines. As previously mentioned, T1 connectivity was extended to the Springfield School District headquarters as part of the ATM trial in Oregon. Two X-terminals were located there, connected to the university hosts and used by the teachers to practice information access as well as the preparation of their own multimedia courseware.

During the 1994 fall and 1995 winter term at the University of Oregon, three physics classes were done entirely using Mosaic. The students in those classes were able to retrieve the class notes offline and hence were no longer compelled to incessantly take notes during the lecture. As judged by improved performance on exams, this situation seems to have led to better student comprehension of the material. The Mosaic browser allowed the students to personally annotate each lecture page and thus "take notes on the notes." The use of class newsgroups and e-mail allowed the students to work collaboratively on some class assignments as well as provide more useful feedback to the instructor than from the traditional lecture format. Normal lectures were still given, but the lecture material itself was in the form of Mosaic pages projected to the class on an LCD projector. In the winter 1995 term, in a course on alternative energy, students working together in self-organized teams successfully constructed WWW pages of their own in response to a homework assignment to create a virtual energy company and advertise it to the world. This effort and the course can be accessed at <http://zebu.uoregon.edu/phys162.html>.

In mid-November 1994, Dr. Bothun went to the Cerro-Tololo Interamerican Observatory in Chile to make observations with the telescopes located there. Even though the telescope and classroom were separated by 10,000 miles, successful interactive classroom sessions did occur and the class was able to follow along with the observations as the data were reduced and then placed on the WWW in quasi-real time. These data can be found

at <http://zebu.uoregon.edu/fp.html>. The delivery of scientific data, in this case digital imaging from a telescope, to the K-12 classroom through the Mosaic interface is perhaps the most powerful educational application on today's Internet. As part of the NERO project, the University of Oregon is hoping to get its Pine Mountain Observatory (located at an elevation of 7,000 feet, 26 miles SE of Bend, Oregon) on a T1 link to the Internet for purposes of delivering access to a scientific instrument and data acquisition directly to a networked K-12 classroom. Another example of this kind of activity is provided by the recent space shuttle mission. Dr. Bothun was part of the scientific experiment on board the *Endeavor* and kept several WWW pages updated about mission progress and the scientific data being obtained. Digital optical photos of the targets that were imaged with the ultraviolet telescope were made available to the Internet audience so that interested followers of the mission could get some idea of the kind of object currently being imaged by the shuttle astronauts. This effort can be seen at <http://zebu.uoregon.edu/uit.html>.

As previously mentioned, a class on how to create WWW home pages was taught via teleconference to four Boulder, Colorado, secondary schools. Subsequent to receiving this remote training, each school set up its own WWW server on the Internet. The students then proceeded to set up their own WWW home pages, learned how to scan pictures into the computer, and then made their pictures available online via the Internet. The students used the network heavily as part of school projects. As part of an oceanography project, they used WWW "search engines" to research material and make personal contacts with researchers around the country. Individual students made personal contacts with professionals at Scripps Institute in La Jolla, California, and with researchers at Sea World. They also successfully accessed supercomputer centers at the National Center for Atmospheric Research and at NOAA. In addition, a teacher at one of the high schools developed a chemistry class for remote delivery to students at middle schools. In conjunction with this class, some of the high school students mentored middle school students via the network. It is precisely this kind of new, dynamic interaction between experts, information resources, and students at various levels that precipitates the kind of quantum leap the experience of learning can undergo.

Finally, a number of demonstrations were held in conjunction with efforts associated with the Council of Great City Schools, a nonprofit organization representing 50 of the nation's largest urban public school systems. These demonstrations included deployment of a technical model for a high-performance classroom at the Council's Annual Conference of Technology Directors in Denver, Colorado, in May 1994, and the design and deployment of a "21st Century Classroom" for a computer office and automation show at the Portland Convention Center in October 1994.

### ANALYSIS AND FORECAST: A LOOK TO THE FUTURE

A number of follow-on projects are either proposed or under way as a result of these successful experimental activities. These follow-on activities, when taken in conjunction with the various experiments, are excellent examples to look to for help in forecasting networking trends for educational institutions over the next 5 to 7 years.

The schools in Boulder, Colorado, used their workstations and associated connectivity in an impressive and successful way. The teachers and students adapted very well to the advanced technology, perhaps even more successfully than anticipated. In 1993 the voters approved a bond issue to allocate funds for provision of a T1 line to every school in the Boulder Valley School District by early 1996. Within a year the schools will be running their own 51-school network, routinely doing the things that were on the cutting edge during the trial period. The ability of schools to actually know and experience the value of these networked services facilitates solicitation of taxpayer support for efforts of this type. When the community sees direct value, then the resources and synergies required to build these networks (both technological and human) emerge more easily.

The schools have a talented pool of teachers and students who will continue with the work begun during the trial. Dr. McBryan has continued to work with the schools and the public library and is currently testing delivery of highly compressed, 30 frame/s, full-frame, VHS-quality video to the schools. The work conducted in Boulder could serve as a model of cooperation between secondary schools and local universities, extending university expertise into the secondary schools and "bootstrapping" them into the information age.

In October 1994 the University of Colorado was awarded a Department of Commerce National Telecommunications and Information Administration grant to develop the Boulder Community Network (BCN). BCN collects information resources from all available sources in Boulder County and makes them available to the whole population through the World Wide Web. Many residents have access through work or school, including schools at the elementary and secondary levels. To widen access, information kiosks have been deployed at strategic locations throughout the county, including libraries and public buildings. The Senior Citizens Center and several similar institutions were provided with both computer platforms and network access. Over 50 community organizations now supply and update information on a regular basis and participate in BCN management and development. Commercial information resources are also available on BCN, including a countywide restaurant menu guide and shopping information. BCN may be accessed at <http://bcn.boulder.co.us/>.

Dr. McBryan was one of the founding members of BCN, is principal investigator on the NTIA grant, and has been BCN Technical Director from the outset. As a result the BCN information resources were rapidly made available to the T1-connected schools, resulting in interesting curriculum developments. As examples, Spanish classes are translating some BCN home pages into Spanish and a history class plans an online archive of tapings of early Boulder residents, both available to all residents via BCN. The marriage of community network and school Internet connectivity represents a new paradigm for outreach in education.

Looking forward into the future and taking a 5- to 7-year view, the Boulder participants see a number of key trends developing. In today's environment, students must leave their classrooms and go to a computer lab. In the future the computers must be located directly in the classrooms. CD-ROMs will be installed on every desktop, allowing individualized instruction programs to be developed with each student having his or her own programs and subjects. Even today schools can acquire a Performa 630 with a CD-ROM for \$1,200 or less. Video broadcast and video on demand will mix freely with CD-ROM capabilities. All three media and their supporting network topologies will bring high-quality educational and current affairs materials into the classroom.

In October 1994 the University of Oregon received one of 100 NTIA grants from the U.S. Department of Commerce. The university in partnership with 15 other Lane County agencies will use the \$525,000 grant to create the Lane Education Network. This high-speed computer network will connect the participating agencies' information systems in a seamless, fully accessible community network. The grant will finance demonstration projects, electronic classrooms, and distance-learning, off-campus classes. Agencies in the partnership include three local school districts. In one of those school districts, voters recently passed a \$37.5 million bond levy, \$3 million of which is designated for instructional technology.

In conjunction with these associated activities in Oregon a cooperative effort is currently being proposed called "The On-line Universe: Network Delivery of Multimedia Resources in Science to the K-12 Classroom and the General Public." Building on the successful experiments outlined above, the project will focus in several areas. Multimedia educational materials will be developed in the areas of astronomy and the environmental sciences. These materials will be delivered via Internet connectivity at the T1 level between the University of Oregon, the Oregon Museum of Science and Industry (OMSI), and individual science classrooms at Grant High School in Portland and at Springfield and Thurston High Schools in Springfield, Oregon. All materials will be organized and accessible through the Mosaic interface. In addition, M-bone connectivity will be used at these sites for desktop conferencing and live interactions between students, university professors, K-12 teachers, and museum educational staff. The existing multimedia Internet classroom on the University of Oregon campus will be used to develop a new classroom at OMSI to serve as a training area for K-12 science teachers in the Eugene and Portland areas.

Along the same lines but on a larger geographic scale, the Council of Great City Schools is proposing an "Urban Learning Network" to link urban high schools in 14 states to deliver useful information and practical educational services to teachers and students in the real-world environments of urban community high schools. The proposed list of services and applications to be provided should sound familiar at this point: video on demand, interactive multimedia courses, distance instruction or teleconferencing, electronic field trips, and fully equipped smart classrooms.

These various activities show that electronic networks offer the possibility of forming learning communities of educators and learners who can share and develop courseware. The process can be highly interactive, and feedback from the students can be integrated into improving the content. Most importantly, a

network of this type facilitates educational outreach activities between higher education, K-12 school systems, libraries, and science museums. These new, network-based partnerships will allow any expert's resources to be delivered to a larger audience with the aim of improving education through the delivery of real data in an interface that promotes learning as a discovery process. Such capabilities will place vast amounts of information at the fingertips of educators, students, and ordinary citizens.

### SUMMARY AND RECOMMENDATIONS

This trial and our collective observations support the expectation that traditional barriers between student, teacher, and professional researcher will erode and be replaced by a learning community of individuals who are collectively interested in a particular subject. Herein lies the paradigm shift. The keeper of the knowledge will no longer be the individual K-12 teacher or the professor. Rather, students will now be able to access and interact with a diverse and distributed knowledge base. Professional researchers can make their real data available via interactive networks to this learning community, and science can be taught as a discovery process rather than a collection of facts "known" about the physical world. This approach duplicates what the professional scientist does and can go a long way toward improving science education.

These technological breakthroughs and educational links between universities and K-12 schools underscore the value of education at every level. The electronic superhighway will increasingly allow experts to share their research and foster an excitement for learning among interested students, and in this process they can also help K-12 teachers navigate quickly through the vast and often times confusing information highway. Thus, through this partnership, the "looking at everything" ideal becomes increasingly real with seemingly endless possibilities for exploration, dialogue, and learning.

Through the activities of this trial two unexpected insights were gained relative to the communications technologies: (1) the power and pervasiveness of TCP/IP as a common denominator for educational networking was dramatically underscored; and (2) the scalability, robustness, and aggregation strengths of ATM switching make it an ideal fabric for a public switched network to support educational and developing multimedia applications. Also, the fact that the IP networking of education communities is already in place and expandable, and that applications were deliverable at T1 bandwidths (allowing for the use of existing copper-based infrastructures), means that K-12 and learning communities can begin accelerated access to the information highway today, migrating to other exclusively fiber networks over time. Although DS3 and other higher speed network capabilities are often desirable, it is currently neither feasible nor financially possible to extend these capabilities universally to all schools. Until the value of these networks is comprehended, "affordability" remains a circular argument. The network demonstrated in this trial comes within the reach of anyone when properly scaled, and its incredible value can be experienced by learners and educators.

Society's challenge will be to provide schools and universities with the resources and support they need to have full and equal electronic access to information in this new era. Such a partnership extends also to business and industry, whose success ultimately depends on qualified personnel entering the work force. This trial highlighted the possibilities that will emerge when a learning community discovers the value that can be derived from the effective application of technology and communications networking. Communities of interest exist everywhere in human society. Each community must individually shape the vision and define the value that these networks can deliver. There is no "one size fits all solution." The trial demonstrated that the technologies are very flexible, with many of them already in place to deliver the "value" the users want. Value has a direct relationship to "affordability." Obtaining funding for these technological advancements and communications networks will require creativity from the communities that build them.

### NOTES

1. Wallace, B. 1993. "US WEST Plots ATM Course with Planned Expansion," *Network World*, October 18, pp. 28 and 31.



2. Greene, T. 1995. "US WEST Places Its ATM Services Cards on the Table," *Network World*, February 6, p. 23.
3. Owen, S. 1993. "Network for Engineering and Research in Oregon," proposal submitted to Education Division of National Aeronautics and Space Administration, September 23.
4. Foldvik, R., D. Meyer, and D. Taylor. 1995. "ATM Network Experimentation in the State of Oregon," *Proceedings of the IEEE International Conference on Communications*, June.
5. Foldvik, R., and O. McBryan. 1995. "Experiences with ATM: The Boulder Perspective," *Proceedings of the IEEE International Conference on Communications*, June.
6. Research supported in part by NSF Grand Challenges Application Group grant ASC-9217394 and by NASA HPCC Group grant NAG5-2218.
7. Laubach, M. 1994. "IETF Draft RFC: Classical IP and ARP Over ATM," RFC 1577, January.
8. "LAN Emulation Over ATM: Draft Specification," ATM Forum Contribution 94-0035.
9. Macedonia, M., and D. Brutzman. 1993. "MBONE, the Multicast Backbone," Naval Postgraduate School, December 17, [taurus.cs.nps.navy.mil:pub/mbmg/mbone.hottopic.ps](http://taurus.cs.nps.navy.mil/pub/mbmg/mbone.hottopic.ps).
10. Macedonia, M., and D. Brutzman. 1994. "MBONE Provides Audio and Video Across the Internet," *Computer* 27(4):30-36.
11. Casner, S. 1993. "Frequently Asked Questions (FAQ) on the Multicast Backbone (MBONE)," May 6, available via anonymous ftp from [venera.isi.edu/mbone/faq.txt](http://venera.isi.edu/mbone/faq.txt).
12. Casner, S., and S. Deering. 1992. "First IETF Internet Audiocast," *ACM SIGCOMM Computer Communication Review*, July, pp. 92-97.
13. Deering, S. 1989. "Host Extensions for IP Multicasting," RFC 1112, August.
14. Moy, J. 1993. "Multicast Extensions to OSPF," IETF draft, July.
15. Deering, S. 1988. "Multicast Routing in Internetworks and Extended LANs," *Proceedings of the ACM SIGCOMM 1988*.
16. Deering, S., D. Estrin, D. Farinacci, V. Jacobson, C. Liu, and L. Wei. 1995. "Protocol Independent Multicast (PIM): Motivation and Architecture," Internet draft, draft-ietf-idmr-pim-arch-00.txt, November.
17. Deering, S., D. Estrin, D. Farinacci, V. Jacobson, C. Liu, and L. Wei. 1995. "Protocol Independent Multicast (PIM): Protocol Specification," Internet draft, draft-ietf-idmr-pim-spec-01.ps, January 11.
18. *Release 1.0*. 1994. "The Internet Multicasting Service: Ted Turner, Watch Out!," 94(2):10.
19. Refer to <http://town.hall.org/radio/live.html>.
20. Andreessen, M. 1993. "NCSA Mosaic Technical Summary," May 8, [marca@ncsa.uiuc.edu](mailto:marca@ncsa.uiuc.edu).
21. Valauskas, E. 1993. "One-Stop Internet Shopping: NCSA Mosaic on the Macintosh," *ONLINE*, September, pp. 99-101.
22. Powell, J. 1994. "Adventures with the World Wide Webs, Creating a Hypertext Library Information System," *DATABASE*, February, pp. 59-66.

## 9

# An SDTV Decoder with HDTV Capability: An All-Format ATV Decoder

*Jill Boyce, John Henderson, and Larry Pearlstein  
Hitachi America Ltd.*

This paper describes techniques for implementing a video decoder that can decode MPEG-2 high-definition (HD) bit streams at a significantly lower cost than that for previously described high-definition video decoders. The subjective quality of the pictures produced by this "HD-capable" decoder is roughly comparable to current DBS delivered standard-definition (SD) digital television pictures. The HD-capable decoder can decode SD bit streams with precisely the same results as a conventional standard-definition decoder. The MPEG term Main Profile at Main Level (MP@ML) is also used to refer to standard-definition video in the sequel.

The decoder makes use of a pre-parser circuit that examines the incoming bit stream in a bit-serial fashion and selectively discards coded symbols that are not important for reconstruction of pictures at reduced resolution. This pre-parsing process is performed so that the required channel buffer size and bandwidth are both significantly reduced. The pre-parser also allows the syntax parser (SP) and variable-length decoder (VLD) circuitry to be designed for lower performance levels.

The HD-capable decoder "downsamples" decoded picture data before storage in the frame memory, thereby permitting reduction of the memory size. This downsampling can be performed adaptively on a field or frame basis to maximize picture quality. Experiments have been carried out using different methods for downsampling with varying results. The combination of the pre-parser and picture downsampling enables the use of the same amount of memory as used in standard definition video decoders.

The decoder selects a subset of the 64 DCT coefficients of each block for processing and treats the remaining coefficients as having the value zero. This leads to simplified inverse quantization (IQ) and inverse discrete cosine transform (IDCT) circuits. A novel IDCT is described whereby the one-dimensional 8-point IDCT used for decoding standard definition pictures is used as the basis for performing a reduced complexity IDCT when processing high-definition bit streams.

A decoder employing the above techniques has been simulated using "C" with HDTV bit streams, and the results are described. Normal HDTV encoding practices were used in these experiments. The bit streams were decoded according to the concepts described herein, including pre-parsing, the effects of reduced memory sizes, simplified IDCT processing, and the various associated filtering steps. The pre-parser and resampling result in a certain amount of prediction "drift" in the decoder that depends on a number of factors, some of which are under the control of the encoder. Those who have viewed the resulting images agree that the decoding discussed in this paper produces images that meet performances expectations of SDTV quality.

The HD-capable video decoder, as simulated, can be expected to be implementable at a cost only marginally higher than that of a standard definition video decoder. The techniques described here could be applied to produce HD-capable decoders at many different price/performance points. By producing a range of consumer products that can all decode HDTV bit streams, a migration path to full HDTV is preserved while allowing a flexible mix of video formats to be transmitted at the initiation of digital television service.

There is an ongoing policy debate about SDTV and HDTV standards, about a broadcast mix of both formats, and about how a full range of digital television might evolve from a beginning that includes either SDTV or HDTV or both. This paper offers technical input to that debate, specifically regarding consumer receivers that could decode both SDTV and HDTV digital signals at a cost only marginally higher than that of SDTV alone.

There are at least two areas of policy debate in which these issues are relevant:

1. *What is the right mix of HDTV and SDTV as the digital service evolves over time?* There are a variety of introduction scenarios for digital television, ranging from HDTV only, to SDTV only, to various mixes of the two. To preserve the HDTV broadcast option no matter how digital television service is introduced, SDTV receivers must be able to decode the HDTV signal. It is assumed here that SDTV receivers with such HDTV-decoding capability are both practical and cost effective. It is thus entirely practical to preclude SDTV-only receivers. Therefore, the introduction of SDTV would not prevent later introduction of HDTV because fully capable digital receivers would already be in use.
2. *How quickly can National Television System Committee (NTSC) broadcasting be discontinued?* The receiver design approach described herein can be applied to low-cost set-top boxes that permit NTSC receivers to be used to view digital television broadcasts. The existence of such decoders at low cost is implicit in any scenario that terminates NTSC broadcast.

### **COST AND COMPLEXITY OF FULL-RESOLUTION HDTV DECODER COMPONENTS**

The single most expensive element of a video decoder is the picture storage memory. A fully compliant video decoder for U.S. HDTV will require a minimum of 9 MBytes of RAM for picture storage. An HDTV decoder will also require at least 1 MByte of RAM for channel buffer memory to provide temporary storage of the compressed bit stream. It can be expected that practical HDTV video decoders will employ 12 to 16 MBytes of specialty DRAM, which will probably cost at least \$300 to \$400 for the next few years and may be expected to cost more than \$100 for the foreseeable future.

The IDCT section performs a large number of arithmetic computations at a high rate and represents a significant portion of the decoder chip area. The inverse quantizer (IQ) performs a smaller number of computations at a high rate, but it may also represent significant complexity.

The SP and VLD logic may also represent a significant portion of the decoder chip area. At the speeds and data rates specified for U.S. HDTV, multiple SP/VLD logic units operating in parallel may be required in a full HDTV decoder.

### **COST REDUCTIONS OF HDTV DECODER**

This section describes several techniques that can be applied to reduce the cost of an HD-capable decoder. The following decoder subunits are considered: picture storage memory, pre-parser and channel buffer, SP and VLD, inverse quantizer and inverse discrete cosine transform, and motion compensated prediction. The discussion refers to [Figure 1](#), which is a block diagram of a conventional SDTV decoder; and [Figure 2](#), which is a block diagram of an HD-capable decoder. The blocks, which appear in [Figure 2](#) but not in [Figure 1](#), have been shaded to highlight the differences between an HD-capable decoder and a conventional SD decoder.

#### **Picture-Storage Memory**

As described in Ng (1993), the amount of picture-storage memory needed in a decoder can be reduced by downsampling (i.e., subsampling horizontally and vertically) each picture within the decoding loop. Note in [Figure 2](#) that residual or intradata downsampling takes place after the IDCT block and prediction downsampling is done following half-pel interpolation blocks. The upsample operation shown in [Figure 2](#) serves to restore the sampling lattice to its original scale, thus allowing the motion vectors to be applied at their original resolution. Although this view is functionally accurate, in actual hardware implementations the residual/intra downsampling operation would be merged with the IDCT operation, and the prediction downsample operation would be merged with the upsample and half-pel interpolation. In an efficient implementation the upsample—half-pel interpolation—downsample operation is implemented by appropriately weighting each of the reference samples extracted from the (reduced resolution) anchor frame buffers to form reduced resolution prediction references.

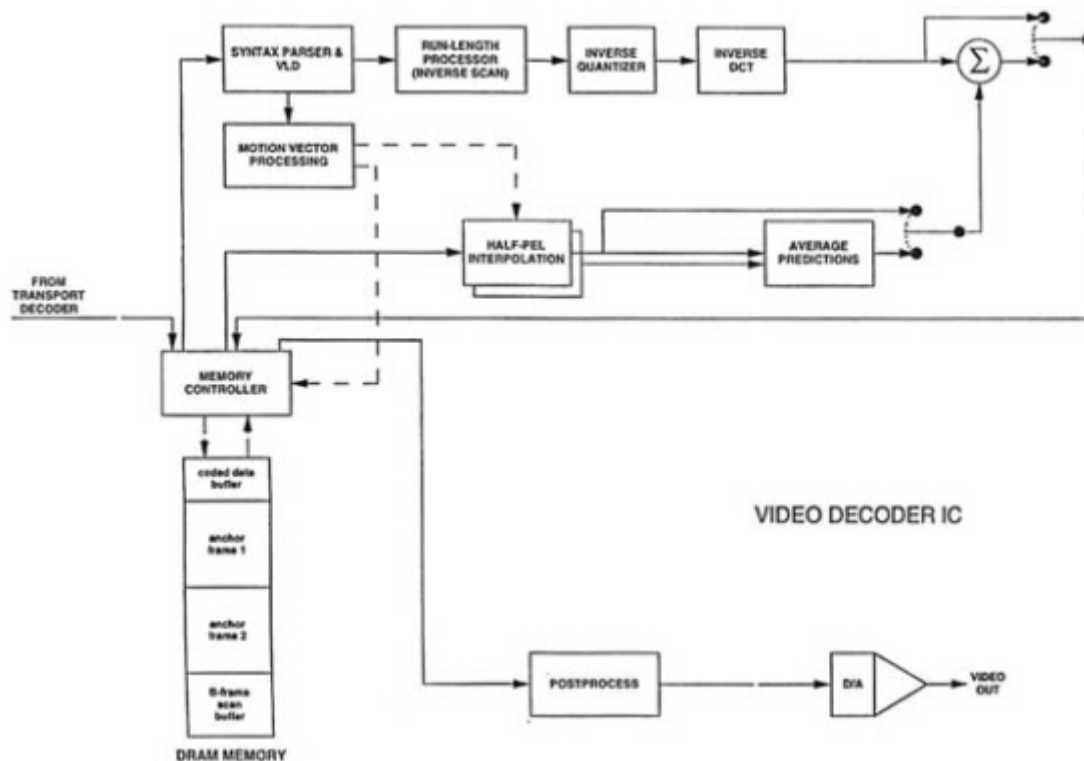


Figure 1 Block diagram of a conventional SDTV video decoder.

The weights used in this operation depend on the full-precision motion vectors extracted from the coded bit stream.

Experiments have shown that it is important that the prediction downsampling process is near the inverse of the upsampling process, since even small differences are made noticeable after many generations of predictions (i.e., after an unusually long GOP that also contained many P-frames). There are two simple methods:

- Downsample without filtering (subsample), and upsample using bilinear interpolation; and
- Downsample by averaging and upsample without filtering (sample and hold).

For both of these methods the concatenated upsample-downsample operation is identity when motion vectors are zero. Both methods have been shown to provide reasonable image quality.

For the residual/intra downsampling process it is possible to use frequency domain filtering in lieu of spatial filtering to control aliasing. Frequency domain filtering is naturally accomplished by "zeroing" the DCT coefficients that correspond to high spatial frequencies. Note that the prediction filtering may introduce a spatial shift—this can be accommodated by introducing a matching shift in the residual/intra downsampling process, or by appropriately biasing the motion vectors before use.

When processing interlaced pictures, the question arises as to whether upsampling and downsampling should be done on a field basis or on a frame basis. Field-based processing preserves the greatest degree of temporal resolution, whereas frame-based processing potentially preserves the greatest degree of spatial resolution. A brute-force approach would be to choose a single mode (either field or frame) for all downsampling.

A more elaborate scheme involves deciding whether to upsample or downsample each macroblock on a field basis or frame basis, depending on the amount of local motion and the high-frequency content. Field based processing is most appropriate when there is not much high-frequency content and/or a great deal of motion. Frame-based processing is most appropriate when there is significant high-frequency content and/or little motion.

One especially simple way of making this decision is to follow the choice made by the encoder for each macroblock in the area of field or frame DCT and/or field- or frame-motion compensation, since the same criteria

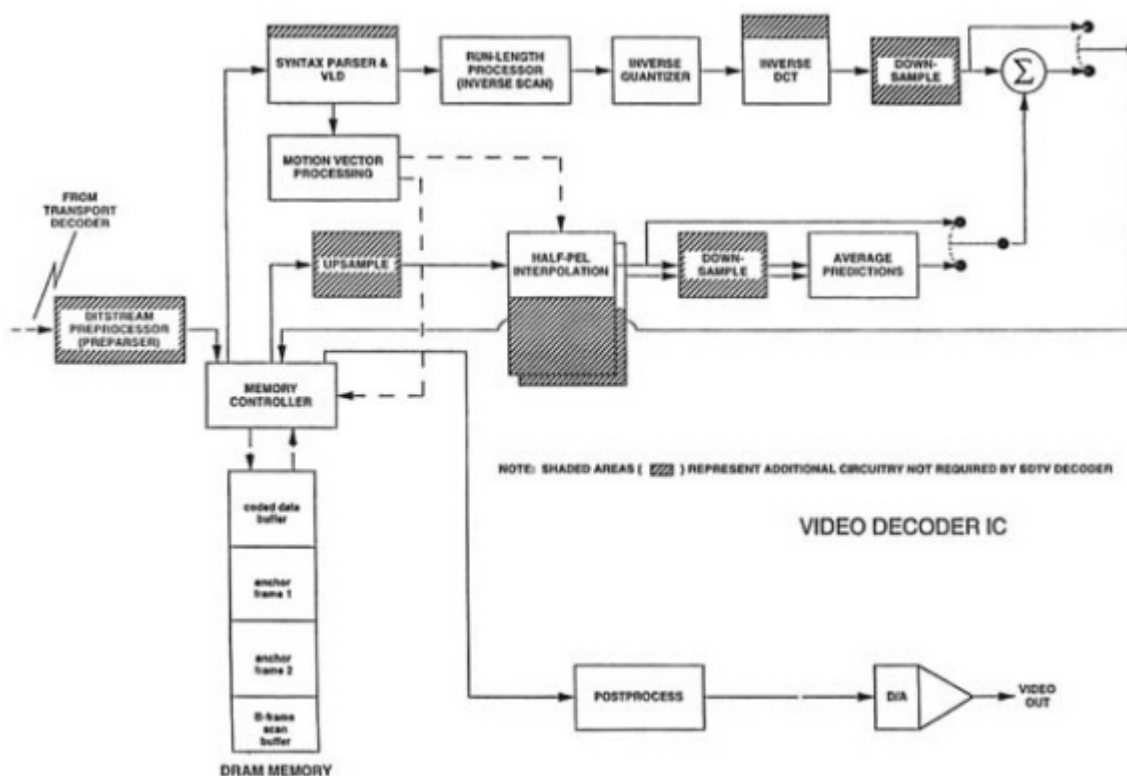


Figure 2 Block diagram of a low-cost HDTV video decoder.

may apply to both types of decisions. Although field conversion is not optimal in areas of great detail, such as horizontal lines, simulations show that if a single mode is used, field is probably the better choice.

In MPEG parlance, SDTV corresponds to Main Level, which is limited to  $720 \times 480$  pixels at 60 Hz, for a total of 345,600 pixels. U.S. ATV allows pictures as large as  $1920 \times 1080$  pixels. Sequences received in this format can be conveniently downsampled by a factor of 3 horizontally and a factor of 2 vertically to yield a maximum resolution of  $640 \times 540$ , a total of 345,600 pixels. Thus the memory provided for SDTV would be adequate for the reduced-resolution HD decoder as well. It would be possible to use the same techniques with a smaller amount of downsampling for less memory savings.

In a cost-effective video decoder, the channel buffer and picture-storage buffers are typically combined into a single memory subsystem. The amount of storage available for the channel buffer is the difference between the memory size and the amount of memory needed for picture storage. Table 1 shows the amount of picture-storage memory required to decode the two high-definition formats with downsampling. The last column shows the amount of free memory when a single 16-Mbit memory unit is used for all of the decoder storage requirements. This is important since cost-effective SDTV decoders use an integrated 15-Mbit memory architecture. The memory not needed for picture storage can be used for buffering the compressed video bit stream.

TABLE 1 Reduced Resolution Decoder Memory Usage

Active Horizontal	Active Vertical	H Scale Factor	V Scale Factor	Downsampled Horizontal	Down-sampled Vertical	\$ Frames Stored	Down-sampled Memory Required	Free Memory With 16 MBITS DRAM
1920	1,080	3	2	640	540	3	12,441,600	4,335,616
1280	720	2	2	640	360	2	5,529,600	11,247,616
720	480	—	—	—	—	3	12,441,600	4,336,616

NOTE: Shaded row reflects SDTV format for reference.

As indicated in Table 1, the  $1920 \times 1080$  format is downsampled by 3 horizontally and 2 vertically. This results in efficient use of memory (exactly the same storage requirements as MP@ML) and leaves a reasonable amount of free memory for use as a channel buffer.

The natural approach for the  $1280 \times 720$  format would be to downsample by 2 vertically and horizontally. This leaves sufficient free memory that the downconverter would never need to consider channel buffer fullness when deciding which data to discard.

After decoding of a given macroblock, it might be immediately downsampled for storage or retained in a small buffer that contains several scan lines of full-resolution video to allow for filtering before downsampling. The exact method of upsampling and downsampling is discussed below; it can greatly affect image quality, since even small differences are made noticeable after many generations of predictions.<sup>1</sup> The upsampling and downsampling functions are additional costs beyond that for an SD decoder.

The general concept of reducing memory storage requirements for a lower-cost HDTV decoder is known in the literature. This paper adds pre-parsing and new techniques for performing downsampling and upsampling.

### Pre-parser and Channel Buffer

A fully compliant HDTV decoder requires at least 8 Mbits of high-speed RAM, with peak output bandwidth of 140 MBytes/sec for the channel buffer. With the use of a pre-parser to discard some of the incoming data before buffering, the output bandwidth can be reduced to a peak of 23 MBytes/sec and the size of the channel buffer can be reduced to 1.8 to 4.3 Mbits. (The lower number is required for MP@ML and the higher number is the amount left over in the SDTV 16-Mbit memory after a  $1080 \times 1920$  image is downsampled by 3 horizontally and 2 vertically, including the required 3 frames of storage.)

The pre-parser examines the incoming bit stream and discards less important coding elements, specifically high-frequency DCT coefficients. It may perform this data selection while the DCT coefficients are still in the run-length/amplitude domain (i.e., while still variable-length encoded). The pre-parser thus serves two functions:

- It discards data to allow a smaller channel buffer to be used without overflow and to allow reduction of the channel-buffer bandwidth requirements.
- It discards run-length/amplitude symbols, which allows for simpler real-time SP and VLD units.

The pre-parser only discards full MPEG code words, creating a compliant but reduced data rate and reduced-quality bit stream. The picture degradation caused by the pre-parsing operation is generally minimal when downsampled for display at reduced resolution. The goal of the pre-parser is to reduce peak requirements in later functions rather than to significantly reduce average data rates. The overall reduction of the data rate through the pre-parser is generally small; for example, 18 Mbps may be reduced to approximately 12 to 14 Mbps.

The channel buffer in a fully HDTV decoder must have high-output bandwidth because it must output a full macroblock's data in the time it takes to process a macroblock. The pre-parser limits the maximum number of bits per macroblock to reduce the worst-case channel buffer output requirement. The peak number of bits allowed per macroblock in U.S. HDTV is 4608; this requires an output bandwidth of 140 MBytes/sec even though the average number of bits per macroblock is only 74. The pre-parser retains no more than 768 bits for each coded macroblock, thereby lowering the maximum output bandwidth to 23 MBytes/sec, the same as for MP@ML.

The pre-parser also removes high-frequency information (i.e., it does not retain any non-zero DCT coefficients outside of a predetermined low-frequency region). Pre-parsing could remove coefficients after a pre-specified coefficient position in the coded scan pattern, or it could remove only those coefficients that will not be retained for use in the IDCT. This reduces the total number of bits to be stored in the channel buffer.

In addition to discarding data to limit bits per coded macroblock and high-frequency coefficients, the pre-parser also alters its behavior based on the channel buffer fullness. The pre-parser keeps a model of buffer occupancy and removes coefficients as needed to ensure that the decreased size channel buffer will never

overflow. As this buffer increases in occupancy, the pre-processor becomes more aggressive about the amount of high-frequency DCT coefficient information to be discarded.

This decoder management of its own buffer is a key difference between the enhanced SDTV decoder and a "normal" SDTV decoder. In a "normal" encoder/decoder combination, the encoder limits the peak data rate to match the specifications of the decoder buffer; it is the responsibility of the encoder to assure that the decoder buffer does not overflow. In the enhanced SDTV decoder outlined in this paper, the decoder can accept bit streams intended for a much larger buffer (i.e., an HDTV bit stream) and can perform its own triage on the incoming bit stream to maintain correct buffer occupancy.<sup>2</sup>

This pre-parser is an additional cost over a stand-alone SD decoder, but the cost and complexity are low since it can run at the relatively low average incoming-bit rate. The pre-parser is significantly less complex than a full-rate SP and VLD because of its slower speed requirement and because it parses but does not have to actually decode values from all of the variable length codes.<sup>3</sup>

### Syntax Parser and Variable-Length Decoder

The computational requirements for the SP and VLD units of the downconverter are substantially reduced by implementing a simplified bit-serial pre-parser as described above. The pre-parser limits the maximum number of bits per macroblock. It also operates to limit the number of DCT coefficients in a block by discarding coefficients after a certain number, thus reducing the speed requirements of the SP and VLD units.

At the speeds and data rates specified for U.S. HDTV, multiple SP/VLD logic units operating in parallel may be required. The pre-parser limits the processing speed requirements for the HD downconverter to SDTV levels. Thus the only additional requirement on the SP/VLD block for decoding HDTV is the need for slightly larger registers for storing the larger picture sizes and other related information, as shown in [Figure 2](#).

### Inverse Quantization and Inverse Discrete Cosine Transform

Reduced complexity inverse quantizer (IQ) and inverse discrete cosine transform (IDCT) units could be designed by forcing some predetermined set of high frequency coefficients to zero. MPEG MP@ML allows for pixel rates of up to 10.4 million per second. U.S. ATV allows pixel rates of up to 62.2 million per second. It is therefore possible to use SDTV-level IQ circuitry for HDTV decoding by ignoring all but the 10 or 11 most critical coefficients. Some of the ignored coefficients (the  $8 \times 8$  coefficients other than the 10 or 11 critical coefficients) will probably have already been discarded by the pre-parser. However, the pre-parser is not required to discard all of the coefficients to be ignored. The pre-parser may discard coefficients according to coded scan pattern order, which will not, in general, result in deleting all of the coefficients that should be ignored by later processing stages.

Processing only 11 of 64 coefficients reduces the IQ computational requirement and significantly decreases the complexity of the IDCT. The complexity of the IDCT can be further reduced by combining the zeroing of coefficients with the picture downsampling described above.

IDCT circuitry for performing  $8 \times 8$  IDCT is required for decoding SD bit streams. A common architecture for computing the two-dimensional IDCT is to use an engine that is capable of a fast, one-dimensional, 8-point IDCT. If the SC IDCT engine were used when decoding HD bit streams, it could perform about three 8-point IDCTs in the time of an HDTV block. Thus the SD IDCT can be used to compute the IDCT of the first three columns of coefficients. The remaining columns of coefficients would be treated as zero and thus require IDCT resources.

A special-purpose IDCT engine would be implemented to do the row IDCTs. It would be especially simple since five of the eight coefficients would always be zero, and only two or three output points would have to be computed for each transform. Note that only four rows would have to be transformed if no additional filtering were to be performed prior to downsampling.

For blocks in progressive frames, or that use field IDCTs, coefficients might be selected according to the following pattern (retained coefficients are represented by "x"):

```
XXXX00000  
XXXX00000  
XXXX00000  
XX0000000  
000000000  
000000000  
000000000  
000000000
```

For blocks that use frame DCTs on an interlaced picture, we might discard coefficients with the following pattern:

```
XXXX00000  
XXXX00000  
000000000  
000000000  
000000000  
000000000  
XX0000000  
XXXX00000
```

This pattern of retained coefficients maintains temporal resolution represented by differences between the two fields of the frame in moving images.

### Motion Compensated Prediction (MCP)

Assume that the anchor pictures<sup>4</sup> have been downsampled, as described above. The data bandwidth to the motion compensation circuitry is thus reduced by the same factor as the storage requirement. As described above, motion compensation is accomplished by appropriately interpolating the reduced resolution picture-reference data according to the values of the motion vectors. The weights of this interpolation operation are chosen to correspond to the concatenation of an anti-imaging upsampling filter, bilinear half-pel interpolation operation (depending on the motion vectors), and optional downsampling filter.

### Complexity Comparisons

A block diagram of the HD-capable video decoder is shown in [Figure 2](#). This can be compared with [Figure 1](#), "SDTV Video Decoder Block Diagram," to identify the additional processing required over an SD decoder. Complexity comparisons between a full-resolution HD decoder, SD decoder, prior art HD downconverter,<sup>5</sup> and the HD-capable decoder described in this paper are shown in [Table 2](#). The total costs of the HD downconverter/SD decoder are not significantly greater than the cost of the SD decoder alone.

### PREDICTION DRIFT

In MPEG video coding a significant portion of the coding gain is achieved by having the decoder construct a prediction of the current frame based on previously transmitted frames. In the most common case, the prediction process is initialized by periodic transmission of all intra-coded (I-frames). Predicted frames (P-frames) are coded with respect to the most recently transmitted I- or P-frames. Bidirectionally predicted frames



TABLE 2 Complexities of New HD-capable Decoder and Prior Decoders

Function	ATSC Standard HDTV	Prior Art HD Downconverter	New HD-capable Decoder	MP@ML (SDTV)
Pre-parser			< 10,000 gates 19 Mbits/sec (partial decode)	
Channel buffer size	8 Mbits	8 Mbits	1.8 to 4.3 Mbits	1.8 Mbits
bandwidth	140 MBytes/sec BW	140 MBytes/sec BW	23 MBytes/sec	23 MBytes/sec
Total off-chip memory requirements	96 Mbits specialty DRAM	16 Mbits DRAM + 8 Mbits specialty DRAM	16 Mbits DRAM	16 Mbits DRAM
SP/VLD	93M coefficients/sec	93 M coefficients/sec	15.5 M coefficients/sec	15.5 M coefficients/sec
IDCT	1.5 M blocks/sec	1.5 M simple blocks/sec (HD) + 240 K full 8 × 8 blocks/sec (SD)	1.5 M half complexity simple blocks/sec (HD) + 240 K full 8 × 8 blocks/sec (SD)	240 K full 8 × 8 blocks/sec (SD)
Upsample/Downsample	—	1,000 to 2,000 gates?	1,000 to 2,000 gates?	—
Decodes HDTV	Yes	Yes	Yes	No
Decodes SDTV	Yes	No	Yes	Yes

(B-frames) are coded with respect to the two most recently transmitted frames of types I or P.

Let the first P-frame following some particular I-frame be labeled P1. Recall that the decoder described above downsamples the decoded frames before storage. Thus, when P1 is to be decoded, the stored I-frame used for constructing the prediction differs from the corresponding full-resolution prediction maintained in the decoder. The version of P1 produced by the HD-capable decoder will thus be degraded by the use of an imperfect prediction reference, as well as by the pre-parsing and downsampling directly applied to P1. The next decoded P-frame suffers from two generations of this distortion. In this way the decoder prediction "drifts" away from the prediction maintained by the encoder, as P-frames are successively predicted from one another. Note that the coding of B-frames are successively predicted from one another. Note that the coding of B-frames does not contribute to this drift, since B-frames are never used as the basis for predictions.

Prediction drift can cause visible distortion that changes cyclically at the rate of recurrence of I-frames. The effect of prediction drift can be reduced by reducing the number of P-frames between I-frames. This can be done by increasing the ratio of B-frames to P-frames, decreasing the number of frames between I-frames, or both. As a practical matter, however, special encoding practices are neither needed nor recommended. Experiments have shown that reasonable HD video-encoding practices lead to acceptable quality from the HD-capable decoder described here.

### SIMULATION RESULTS

Test images were chosen from material judged to be challenging for both HDTV and SDTV encoding and decoding. Progressive sequences were in the 1280 × 720 format; interlaced sequences were 1920 × 1024 and 1920 × 1035 (we did not have access to 1920 × 1080 material). The images contained significant motion and included large areas of complex detail.

Some of the bit streams used for testing were encoded by the authors using their MPEG-2 MP@HL software; others were provided by Thomson Consumer Electronics. Decoding at HDTV was done using the authors' MPEG-2 MP@HL software. The HD-capable decoder algorithms described above were simulated in "C" and tested with the HDTV bit streams.

The simulation included accurate modeling of all the processes described here. The bit stream was parsed; channel buffer and picture memory sizes, IDCT processing, and IDCT coefficient selection were all in accord with these explanations; upsampling and downsampling were applied to use the original motion vectors.

The resulting HDTV and downconverted images were examined and compared. Although the downconverted images were of discernibly lower quality than the decoded full-HDTV images, observers agreed that the downconversion process met performance expectations of "SDTV quality."

### CONCLUSIONS

This paper describes an enhanced SDTV receiver that can decode an HDTV bit stream. The enhancements needed to add HD decoding capability to an SD decoder are modest, even by consumer electronics standards. If all receivers included the capabilities described here, an introduction of SDTV would not preclude later introduction of HDTV because fully capable digital receivers would already be in use. The techniques described in this paper also permit design of low-cost, set-top boxes that would permit reception of the new digital signals for display on NTSC sets. The existence of such boxes at low cost is essential to the eventual termination of NTSC service.

### REFERENCES

- Lee, D.H., et al. Goldstar, "HDTV Video Decoder Which Can Be Implemented with Low Complexity," Proceedings of the 1994 IEEE International Conference on Consumer Electronics, TUAM 1.3, pp. 6-7.
- Ng, S., Thomson Consumer Electronics, "Lower Resolution HDTV Receivers," U.S. Patent 5,262,854, November 16, 1993.

### NOTES

1. The "predictions" mentioned here are the P-frames within the GOP sequence. The downsampling and preparsing processes alter the image data somewhat, so that small errors may accumulate if unusually long GOP sequences contain many P-frames. B-frames do not cause this kind of small error accumulation, and so good practice would be to increase the proportion of B-frames in long GOP sequences or to use GOPs of modest length. Receiver processing techniques can also reduce any visible effects, although they are probably unnecessary.
2. In this paragraph, the buffering operation and its associated memory are treated as distinct from the picture-storage memory. This distinction is useful for tutorial purposes, even though the two functions may actually share the same physical, 16-Mbit memory module.
3. Note that the macroblock type, coded block pattern, and run-length information must be decoded.
4. Anchor pictures are I- and P-frames in the MPEG GOP sequence. The downsampling that has been applied to them by the decoding techniques described here means that the motion vectors computed by the encoder can no longer be directly applied.
5. The term "downconverter" as used here applies to hardware that reduces the full HDTV image data to form an SDTV-resolution picture. The appropriately enhanced SDTV decoder described here inherently includes such an HDTV "downconverter."

## 10

# NII and Intelligent Transport Systems

*Lewis M. Branscomb and Jim Keller*

*Harvard University*

Just as the highway metaphor has driven the vision of advanced communications infrastructure, so also are advances in communications pushing the vision of our transportation systems future. How do these two systems interrelate? What policy issues does this relationship raise?

Over the past half-century, the U.S. highway system has advanced regional and national economic development by enhancing access to markets for goods, services, and people. It has also provided direct quality-of-life benefits by providing easier access to both work and leisure. Now the traditional model for surface transportation is reaching its limit. In many areas systems are at or beyond capacity. Building new or bigger roads is not the answer because of space and budget constraints and environmental concerns. Instead, the focus of transportation experts is now on promoting more efficient use of existing capacity. The central theme of these efforts is more efficient integration of existing transportation components through the use of information technology.

The Department of Transportation's Intelligent Transportation Systems (ITS) Program is the focal point for coordinating the development of a national ITS system. The current program was funded in 1991 under the Intermodal Surface Transportation Efficiency Act of 1991. In the federal ITS program user services are broken down into 7 areas consisting of 29 applications (Table 1). Anticipated benefits of ITS include reduced travel time and pollution and increased traveler safety and convenience. Although there are clear public-interest benefits to ITS services, the public-sector role in ITS development will in many respects be indirect. Consistent with activity in other components of the national information infrastructure (NII), development will be a complex mix of public and private interactions.

The federal role in the development of ITS will be primarily that of an enabler, just as it is for the NII. ITS systems involving fixed facilities will initially develop locally and regionally, whereas commercial products for use in vehicles and in trip planning will be sold nationally and internationally. The Department of Transportation will need to take a leadership role in a variety of system architecture and standards development, deployment, and coordination issues to ensure that all of these systems will come together smoothly into a coherent national system. Standards will be needed to address interoperability at different layers in the ITS architecture and issues of data compatibility across systems. Data collected locally must integrate smoothly into national systems, requiring broad agreement on data definitions and coding. Local system developers will have to agree on what information is important, what it will be called, and how it will be recorded.

The standardization issues are complex relative to the traditional telecommunications environment, as they span a broader array of technologies and systems. At the same time, the environment for standardization is relatively weak. Telecom standards evolved with a common platform and a stable (indeed regulated) competitive environment, but ITS will consist of heterogeneous systems and a relatively independent set of players. As with other areas of computing and communications, players will face a conflicting set of incentives to seek standardization, including incentives to differentiate (for competitive advantage and to lock in market share) and to seek interoperability (to expand the market for services and to lower the costs for customers to migrate to their system).

TABLE 1 ITS User Services

User Service Area	Application
Travel and Transportation Management	En-route driver information
	Route guidance
	Traveler services information
	Traffic control
	Incident management
Travel Demand Management	Emissions testing and mitigation
	Demand management and operations
	Pre-trip travel information
Public Transportation Operations	Ride matching and reservation
	Public transportation management
	En-route transit information
	Personalized public transit
Electronic Payment	Public travel security
	Electronic payment services
Commercial Vehicle Operations	Commercial vehicle electronic clearance
	Automated roadside safety inspection
	On-board safety monitoring
	Commercial vehicle administrative processes
	Hazardous material incident response
	Freight mobility
	Emergency notification and personal security
Emergency Management	Emergency vehicle management
	Emergency notification and personal security
Advanced Vehicle Control and Safety Systems	Longitudinal collision avoidance
	Lateral collision avoidance
	Intersection collision avoidance
	Vision enhancement for crash avoidance
	Safety readiness
	Pre-crash restraint deployment
	Automated vehicle operation

NOTE: User services as defined by the National ITS Program Plan, ITS America, March 1995.

Like standards development, many other aspects of ITS development and deployment will involve new and complex coordination issues. The technologies for many ITS applications have already been developed, but a variety of nontechnical issues will determine how and when these applications become widely available. Liability, intellectual property, security, privacy, and data ownership issues will all influence the commitment of private firms to deploying ITS services and the interest of users in adopting them.

Those in government and in industry responsible for planning and coordinating ITS developments recognize that some of the communications systems, such as those supporting traffic flow control and addressing emergency situations, will require real-time, quick response capability with high reliability. Such systems will probably have to be dedicated to ITS applications. At the other extreme, trip planning and many other offline applications can surely be supported by the nation's general purpose data networks—the NII. It is unclear, however, where the boundaries lie and what combination of architecture, commercial strategies, public services, and regulatory constraints will serve to define this relationship. In short, is the ITS a domain-specific application of the NII? Or is ITS a special environment whose information systems support is specialized and only loosely coupled, through data sharing, with the NII?

### INTELLIGENT TRANSPORTATION SYSTEMS

The mission of the ITS program is to improve the safety, efficiency, and capacity of the country's surface transportation system through the use of information technology. The ITS program is coordinated within DOT,

bringing together each of the major program areas within the department. The systems integration and real-time information requirements of ITS include not only managing the flow of real-time information to and from individual vehicles, but also the seamless coupling of different modes of transportation. In addition, many commercial applications for consumer motorists are anticipated. These include traffic management, vehicle tracking, electronic toll collection, augmentation of driver perception, automated emergency intervention, real-time traffic and travel information, trip planning, and eventually automation of at least parts of the driving process, such as collision avoidance.

Eventually the ITS will call for a high-integrity, real-time network system. This system will take inputs from highway sensors, from vehicle global positioning system (GPS) systems, and from other information gathering systems. This information will be continuously compiled in a system of databases with a dynamic model of the local, regional, and national highway system, and will be used to provide real-time information on optimum routes, based on such factors as least time and fuel economy. Although there is a higher degree of homogeneity among ITS applications and data requirements relative to the NII, the real-time requirements, security, and scale of some parts of the ITS make it one of the most challenging NII applications.

### THE GOVERNMENT ROLE

Like the NII initiative of which it is a component, ITS is at the forefront of changes in how the federal government will relate to the states and the private sector. In the post-Cold War economy, agency initiatives are being driven by a new set of requirements. These include an active role for industry in project design, selection, and execution; reliance on private investment; mechanisms to ensure commercial adoption; complex management structures; and a federal role in consensus building.<sup>1</sup> ITS fits well into this model. The ITS program, and in particular the definition of ITS technical requirements, have been developed with active participation from ITS America, a broad-based industry consortium, and with the understanding that ITS will rely heavily on private investment and state and local deployment.

The ITS program was established by Congress through the Intelligent Vehicle Highway Systems (IVHS) Act,<sup>2</sup> part of the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991. This legislation authorizes the secretary of transportation to conduct a program to research, develop, operationally test, and promote implementation of ITS systems.<sup>3</sup> Though the secretary is authorized in these areas, it is clearly indicated in the legislation that this role is intended to be cooperative and facilitatory. The secretary is directed to seek transfer of federally owned or patented technology to the states and the private sector. The secretary is also directed to consult with the heads of the Commerce Department, the Environmental Protection Agency, the National Science Foundation, and other agencies, as well as to maximize the role the private sector, universities, and state and local governments in all aspects of the program.<sup>4</sup>

The management of ITS is conducted by the Joint Program Office (JPO) in the secretary of transportation's office. The JPO manages ITS activities in all areas of the department, as well as working actively to coordinate and build consensus among ITS stakeholders. As directed by ISTEA, DOT has produced a planning document, the "IVHS Strategic Plan—Report to Congress" (December 1992), outlining the program activities, roles, and responsibilities. DOT is also funding the development of a national ITS architecture. This development is being conducted by private consortia under the direction of DOT and is still in process. Currently, the architecture is loosely defined, identifying ITS system elements and estimating the communication and information requirements of each. It does not specify the technical characteristics of component interfaces that will be required for interoperability.

To examine federal policy in the development of ITS, it is perhaps easiest to break down the areas for federal activity. As ITS is broad in scope, this covers quite a bit of ground and will vary between different ITS application areas. For example, in the case of emergency fleet management, technology development will occur under a more traditional procurement model. Many traveler information services may be provided by commercial information services, and onboard capabilities will require the participation of auto manufacturers and wireless telecommunications companies. This participation could, in principle, be achieved either through market incentives or by mandating compliance.

Opportunities for federal activity in the development of ITS can be broken down into the following areas:

- *Promoting the development and testing of ITS technologies and applications.* This is an area in which DOT has been actively engaged. A number of demonstration projects are now under way, most addressing multiple ITS applications. These projects should go beyond traditional R&D to include a diffusion component aimed at assisting future implementors. The projects should focus not only on feasibility, but also on measuring benefits and cost-effectiveness.
- *Fostering an environment that will engage state and local agencies and private firms in the deployment of ITS services.* The availability of ITS services depends largely on the efforts of state and local agencies and private firms to implement them. Some ITS services will come about on their own, based on their commercial viability. In other cases, regulatory mandates or public-sector procurement will be needed to stimulate product development and availability. DOT must clearly articulate its ITS goals to allow systems developers time to anticipate these requirements.
- *Promoting the development and adoption of standards and a national ITS architecture that will ensure that local and regional systems coalesce into a coherent national system and allow integration of U.S. components into international systems.* Facilitating the deployment of an interoperable set of ITS implementations represents many challenges. ITS will likely be deployed as a set of semiautonomous, local implementations and will involve stakeholders in many branches of the computing, communications, and transportation industries. For these systems to coalesce smoothly into a national system, broad consensus will need to be achieved early on among product and service developers and local implementors. It will also require a robust architecture that can expand in both scope and scale to accommodate unanticipated service requirements.

In pursuit of compatibility between local and regional systems, DOT has initiated the National ITS Architecture Development Program. This program was established in September 1993 and its requirements driven, based on the 29 ITS user services (see [Table 1](#)). The architecture seeks to define the systems components and component interactions for a national ITS. Given the breadth of interoperability issues related to ITS, such a high level of coordination will be necessary, but it is not without risk. Key challenges facing planners will be ensuring acceptance and conformance and designing a system that will be able to support unanticipated requirements and applications. To ensure flexibility, the architecture should be based on information requirements and interfaces, not specific technologies or the particular way in which the service is provided. For example, when placing a phone call, a user is typically indifferent to whether the call goes over copper, fiber, or any other medium, as long as basic performance standards are met.

Fulfilling these architectural guidelines will require political as well as technological prowess. It will be difficult to achieve consensus in the diverse environment of ITS developers and implementors. The most important means will be to keep the process open and participatory. This is the approach DOT has taken so far, and private-sector participation has been strong. Another opportunity to achieve conformance will be the procurement lever. To some degree, DOT will also be able to tie conformance to federal funding. However, not all local and regional implementations will receive federal funds, and while it is possible to mandate conformance as a requirement for general DOT highway funding, these funds may be shrinking and yield less influence.

Such a top-down approach is interesting to consider in contrast to other infrastructure development efforts. Part of the success of the Internet, for example, has been its ability to evolve from the middle out, to be able to run over unanticipated types of networking technology and to support unanticipated applications. The ITS application areas are highly defined at present, but technology planning is historically an inexact science, and much will rest on how willing users are to pay.

To the extent that the Highway Trust Fund constitutes a major source of federal investment that can be tied to ITS objectives, the ITS is different from the larger information infrastructure of which it is a part—the national information infrastructure. However, the states have a much bigger role in ITS than in the NII, where state roles are largely confined to telecommunications regulations—authority that will likely be substantially curtailed in the future. So whereas the NII requires more sensitive relationships between the federal government and private industry, the ITS requires a three-way collaboration between a federal department with significant

financial resources; states with the major responsibility for both investment and operations; and private industry, without which neither level of government can realize its objectives.

- *Identifying and supporting those ITS application areas that provide clear benefit to the public but would not otherwise be deployed by the private sector.* ITS represents a broad array of application areas. Some of these are consumer-oriented conveniences, while others are public goods that will extend transportation and nontransportation benefits to the public at large. Federal ITS investment should be made on an application-by-application basis, based on the anticipated service benefits and the potential for private investment (with and without federal support).
- *Clarifying areas of law that may inhibit the adoption of ITS services, including product and service liability, data ownership, privacy, and security.* The move toward ITS services brings up a number of yet-to-be-defined areas of law. Some of these issues are specific to transportation, and others are more broadly relevant in the emerging information society. One area is liability. As transportation systems become more complex and new systems, technologies, and organizations influence the movement of people and materials, liability concerns begin to touch a larger array of players. It is currently unclear to what extent information service and systems providers will be responsible for failures ranging from inconvenience to catastrophe. Do public-sector service providers risk liability? These ambiguities are a potential deterrent to the development and availability of ITS products.

ITS has also produced concerns about the abuse of personal privacy. The information gathering potential of ITS is tremendous. ITS systems may be able to identify where individuals (or at least their vehicles) are and where they have been. Who owns this information, and who will have access to it? Will there be limits on its use? Will service providers be able to sell this information to marketers?

The DOT role in ITS will differ fundamentally from the earlier DOT role in developing the federal highway system, and it anticipates challenges federal policymakers will face in other sectors in the future. Other areas of federal activity, such as environmental management, housing, health care, education, and social services, are becoming more information intensive. ITS offers a proving ground for federal efforts to coordinate the development of intelligent infrastructure. A critical factor in managing this effort will be to maximize the extent to which systems can be leveraged across sectors. If ITS can be leveraged in ways that will decrease the marginal cost of developing infrastructure for other areas, it can help to jump-start these efforts and offset or reduce ITS costs.

### ITS INFRASTRUCTURE REQUIREMENTS

Most ITS applications have an inherent communications component, and in the minds of many, ITS brings a vision of dedicated communications networks. However, in looking at the infrastructural elements, the communications element, while pervasive, is generally not beyond the scope of anticipated NII capabilities. Instead, it appears that ITS communications requirements may be met largely through general-purpose infrastructure. Special-purpose infrastructure needed to support ITS can be broken down into five general areas:

- *Dedicated networks.* This category is intended to identify those ITS applications that may not be supported by general-purpose NII infrastructure. The communications requirements of ITS applications can be grouped into three categories. The first is autonomous stand-alone systems that will not be part of a larger network, for example, intersection collision avoidance systems, which will communicate between approaching vehicles and intersection-based sensors. The next category is applications that will likely be supported by commercially available communications services, for example, pre-trip or en-route travel information that will be a low-bandwidth, "bursty" application and that will likely be served by otherwise available circuit or packet-based wireless or wire-line services. The third includes applications that will require the support of a dedicated, special-purpose network. The dedicated network category will include only this last set of applications.

- *Roadside element deployment.* Sensors and broadcast devices that will be required along roadways to sense or communicate with vehicles. Broadcast devices may signal vehicles concerning road conditions, speed limit, or other factors. Roadside elements may also communicate with central servers managing traffic or freight information, but this will not necessarily require a special-purpose network.
- *Information/database systems.* These systems will be the heart of ITS. Shared access will be provided to these systems to manage information storage and retrieval and processing of ITS information.
- *Service points.* These are site-specific locations at which ITS services will be administered. Examples include electronic payment, weigh in-motion, and emissions testing sites.
- *Onboard vehicle components.* Displays, sensors, and communication devices will be installed in vehicles for the purpose of supporting ITS. This includes both autonomous systems (route guidance using GPS and onboard CD-ROM maps) and communications-dependent systems (collision avoidance).

Table 2 maps the 29 ITS applications against these five elements of required infrastructure. This chart is not perfect, as it does not recognize the changing requirements within application areas between early and advanced applications. It also does not recognize that some applications may be enhanced by access to data gathered by the traffic control application. Despite these limitations, Table 2 does point out that a significant component of ITS services may be provided through general-purpose communications infrastructure. Only two applications are identified as having a dedicated network requirement—traffic control and automated vehicle operation. In particular, the primary need for a dedicated ITS network will be for automated vehicle operation, perhaps the furthest out on the time horizon of the ITS applications. Traffic control is also identified as requiring a dedicated network because of the large proliferation of sensors that will be required in some areas. Of course, these dedicated networks may be either owned or leased by transportation agencies.

Table 2 offers a simplified but practical overview of the systems that will be required to support ITS. The Information Infrastructure Project in Harvard's Science, Technology, and Public Policy Program is currently exploring this area in its project on intelligent transportation systems and the national information infrastructure. The project will lead to a workshop in July 1995 that will investigate opportunities for linkages between these initiatives, including infrastructure requirements.

The conclusion supported by Table 2 that most ITS services will not require a dedicated network is based on anticipated rather than available NII capabilities. For many applications there are a variety of potential technological solutions, including public ATM, frame relay, or SMDS networks and the Internet. The Internet is a desirable solution from cost and ubiquity perspectives but is currently lacking in speed, reliability, and security. It is now a "best-efforts" network, with no guarantee of delivery. The Internet technical community recognizes the need to move beyond best efforts service to provide a variety of quality-of-service levels consistent with the needs of different groups of users and applications.<sup>5</sup> Efforts are under way to explore the development of quality-of-service parameters, including a reserved bandwidth service. Similarly, a variety of systems of encryption are now being developed and tested to ensure secure communications.

In terms of communications services available today, mobile communications are virtually ubiquitous and are becoming more so. Mobile phones are virtually free today, and it would be reasonable to expect that they will soon be bundled with most new cars. Exploring the provision of basic ITS services through mobile cellular technologies may offer a low-cost means of developing and testing their marketability.

The trend so far in the development of ITS services has been dedicated private network solutions. Cost comparisons made by these early implementors have heavily favored the use of owned versus leased infrastructure. However, these estimates may not provide a balanced perspective. Analysis of owning versus outsourcing needs to consider customized rates available for multiyear contracts, as opposed to full tariffs, and anticipate the availability of advanced virtual network services. The cost differential is also being driven by the lower cost of capital faced by public agencies. One means of leveling the playing field in this area is industrial bonds, which would allow private corporations to float tax-free bonds to raise money for public sector projects.



TABLE 2 ITS User Services Infrastructure Requirements

User Service Area and Applications	Infrastructural Elements				Onboard Vehicle Components
	Dedicated Network	Roadside Element Deployment	Information/ Database	Service Points	
<b>Travel and Transportation Management</b>					
En route driver information		x	x		
Route guidance		x	x		x
Traveler services information			x		
Traffic control*	x	x	x		
Incident management		x	x		
Emissions testing and mitigation				x	x
<b>Travel Demand Management</b>					
Demand management and operations		x	x		
Pre-trip travel information		x	x		
Ride matching and reservation			x		
<b>Public Transportation Operations</b>					
Public transportation management			x		x
En-route transit information		x	x		x
Personalized public transit			x		
Public travel security			x		
<b>Electronic Payment</b>					
Electronic payment services				x	
<b>Commercial Vehicle Operations</b>					
Commercial vehicle electronic clearance			x	x	
Automated roadside safety inspection			x	x	
Onboard safety monitoring					x
Commercial vehicle administrative processes			x		
Hazardous material incident response		x	x		x
Freight mobility			x		
<b>Emergency Management</b>					
Emergency notification and personal security			x		
Emergency vehicle management			x		
<b>Advanced Vehicle Control and Safety Systems</b>					
Longitudinal collision avoidance					x
Lateral collision avoidance					x
Intersection collision avoidance		x			x
Vision enhancement for crash avoidance					x
Safety readiness					x
Pre-crash restraint deployment					x
Automated vehicle operation	x		x		x

\*Traffic control can be considered a core application that provides information available to enhance other ITS applications.

It appears that there is considerable room for a larger open network role in provision of ITS services. Realization of this role will require active dialogue between the telecommunications community and the ITS community. A shared infrastructure approach can offer benefits beyond reduced cost. From the perspective of a national ITS program, expanding the role of public communications service providers can provide a more stable environment for the deployment of ITS services. The case for owning versus outsourcing will of course vary from project to project, depending on the proximity of the site to existing infrastructure and its value as right of way.

The debate between DOT officials and communications service providers has often painted a black-and-white picture of the ownership issue. There are possible win-win solutions that recognize the value of roadside right-of-way and explore options in trading access for bandwidth and sharing revenue from excess capacity.

In most areas, ITS networks have been procured based on a predefined architecture. Alternatively, states could purchase services rather than networks. This would allow communications service providers to determine what network architecture could most cost effectively meet the service requirements in the context of their overall business. This might also place expertise in ITS communications where it would more easily migrate to serve other areas. If the larger role in ITS network planning is undertaken by public-sector officials, the resulting expertise and learning curve benefits will remain local. If this responsibility falls more heavily on the private sector, other regions may benefit from the providers' expertise. It may also allow services to roll out consistent with underlying economics, rather than arbitrary politically defined geographic areas.

Assuring the maximum benefit from shared infrastructure opportunities will require an active role on the part of DOT and other ITS constituents in articulating ITS requirements and in participating in NII and Internet development activities. ITS has been largely absent in reports coming out of the Clinton administration's Information Infrastructure Task Force (ITTF), the focal point for federal coordination on NII activities. The IITF offers a high-level platform for DOT to articulate its ITS vision and seek integration with both federal and private-sector NII development activities.

Referring back to Table 2, the information component is a dominant infrastructural element. People, vehicles, and freight will all move across regions and ITS deployment areas and will require the integration of ITS information across these regions. Standardization of data structures will be a critical element in the success of ITS and needs to be addressed.

## CONCLUSION

Interest in advanced infrastructure is at an unprecedented level. The Internet and the NII have been embraced by both parties and are virtually unavoidable in the media. This offers fertile ground for the advancement of ITS on both technical and programmatic fronts. To date, ITS activity has occurred almost exclusively in the transportation domain. The conclusion that much if not most of the network data communications needs of the ITS can be met by the NII, if it develops into fully accessible services that can interoperate with all the specialized ITS subsystems, suggests an important dependency of ITS on NII outcomes. Yet there is no formal mechanism, other than participation in the ITS America discussions, through which federal interests in the NII can be brought together with state interests. Such a mechanism needs to be put in place.

A task force launched on April 25, 1995, by Governors Richard Celeste and Dick Thornburgh is responding to a request by Jack Gibbons, the director of the Office of Science and Technology Policy, to examine how federal-state relationships might be strengthened or restructured in the area of science and technology. Sponsored by the Carnegie Commission on Science, Technology and Government, the National Governors' Association, and the National Council of State Legislators, the task force will be looking at a number of specific cases, including the intelligent transportation system. Out of this work, or perhaps out of the President's NII Task Force, a mechanism needs to emerge for engaging state governments in the architecture and policies for the NII. If this linkage is made through the NII Task Force, it follows that the Department of Transportation, together with other departments and agencies concerned with the nation's transportation systems, needs to participate actively in the work of the task force, to ensure that interoperability, security, and other features of the NII are appropriate for the national transportation system's needs.

## NOTES

1. Branscomb, Lewis M. "New Policies, Old Bottles," *Business and the Contemporary World*.
2. In the fall of 1994 the IVHS Program was officially renamed the Intelligent Transportation Systems Program in recognition of the program's multimodal scope.
3. Intermodal Surface Transportation Efficiency Act of 1991, Part B, Section 6051.

4. Intermodal Surface Transportation Efficiency Act of 1991, Part B, Section 6053.
5. Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future*. National Academy Press, Washington, D.C., May, p. 6.

## 11

# Post-NSFNET Statistics Collection

*Hans-Werner Braun and Kimberly Claffy*  
*San Diego Supercomputer Center*

### ABSTRACT

As the NSFNET backbone service migrates into a commercial environment, so will access to the only set of publicly available statistics for a large national U.S. backbone. The transition to the new NSFNET program, with commercially operated services providing both regional service as well as cross-service provider switching points, or network access points (NAPs), will render statistics collection a much more difficult endeavor. In this paper we discuss issues and complexities of statistics collection at recently deployed global network access points such as the U.S. federal NAPs.

### BACKGROUND

The U.S. National Science Foundation (NSF) is making a transition away from supporting Internet backbone services for the general research and education community. One component of these services, the NSFNET backbone, was dismantled as of the end of April 1995. To facilitate a smooth transition to a multiprovider environment and hopefully forestall the potential for network partitioning in the process, the NSF is sponsoring several official network access points (NAPs) and providing regional service providers with incentives to connect to all three points<sup>[1,2]</sup>

NAPs are envisioned to provide a neutral, acceptable use policy (AUP)-free meeting point for network service providers to exchange routing and traffic. The three NSF-sponsored priority NAPs are

- Sprint NAP, in Pennsauken, NJ;
- Pacific Bell NAP, in Oakland/Palo Alto, CA; and
- Ameritech NAP, in Chicago, IL.

NSF also sponsors a nonpriority NAP in Washington, D.C., operated by Metropolitan Fiber Systems.

The Sprint NAP was operational as of November 1994, but the other NAPs were not yet ready until mid-1995, mainly because Sprint was the only NAP that did not try to start off with switched asynchronous transfer mode (ATM) technology, but rather began with a fiber distributed data interface (FDDI) implementation. In addition, NSF is sponsoring a very-high-speed backbone service (vBNS), based on ATM technology, to support meritorious research requiring high bandwidth network resources. The vBNS represents a testbed for the emerging broadband Internet service infrastructure in which all parts of the network will be experimented with: switches, protocols, software, etc., as well as applications. It will be a unique resource for network and application researchers nationwide to explore performance issues with the new technologies (e.g., how host systems and interfaces interact with ATM components of the wide area network) <sup>[1,2]</sup>.

### BOX 1 MERIT NOTIFICATION OF THE CESSATION OF NSFNET STATISTICS COLLECTION

NSFNET performance statistics have been collected, processed, stored, and reported by the Merit Network since 1988, in the early stages of the NSFNET project. During December 1994, the numbers contained in Merit's statistical reports began to decrease, as NSFNET traffic began to migrate to the new NSF network architecture.

In the new architecture, traffic is exchanged at interconnection points called network access points (NAPs). Each NAP provides a neutral interconnection point for U.S.-based and international network service providers. Once the new architecture is in place, Merit will be unable to collect the data needed to continue these traffic-based reports. The reports will be discontinued by spring 1995.

SOURCE: [NIC.MERIT.EDU/nsfnet/statistics/READ.ME](http://NIC.MERIT.EDU/nsfnet/statistics/READ.ME), January 9, 1995.

As the NSFNET era comes to a close, we will no longer be able to rely on what was the only set of publicly available statistics for a large national U.S. backbone (Box 1). The transition to the new NSFNET program, with commercially operated services providing both regional service as well as cross-service provider network switching points (NSPs), will render statistics collection a much more difficult task. There are several dimensions of the problem, each with a cost-benefit tradeoff. We examine them in turn.

## DIMENSION OF THE PROBLEM

### Contractual and Logistical Issues

In the cooperative agreement with the NAPs, the National Science Foundation has made a fairly vague request for statistics reporting. As with the NSFNET program, NSF was not in a position to specify in detail what statistics the network manager should collect since NSF did not know as much about the technology as the providers themselves did. The situation is similar with other emerging network service providers, whose understanding of the technology and what statistics collection is possible is likely to exceed that of NSF. The NAPs and NSPs, however, at least in early 1995, were having enough of a challenge in getting and keeping their infrastructure operational; statistics have not been a top priority. Nor do the NAPs really have a good sense of what to collect, as all of the new technology involved is quite new to them as well.

A concern is that the NAPs will likely wait for more specific requirements from NSF, while NSF waits for them to develop models on their own. Scheduled meetings of community interest groups (e.g., NANOG, IEPG, FEPG, EOWG, Farnet)<sup>3</sup> that might develop statistics standards have hardly enough time for more critical items on the agenda, e.g., switch testing and instability of routing. The issue is not whether traffic analysis would help, even with equipment and routing problems, but that traffic analysis is perceived as a secondary issue, and there is no real mechanism (or spare time) for collaborative development of an acceptable model.

**Cost-benefit tradeoff: fulfill the deliverables of the cooperative agreement with NSF, at the least cost in terms of time and effort taken away from more critical engineering and customer service activities.**

### Academic and Fiscal Issues

Many emerging Internet services are offered by companies whose primary business thus far has been telecommunications rather than Internet protocol (IP) connectivity. The NAP providers (as well as the vBNS provider) are good examples. Traditionally phone companies, they find themselves accustomed to having reasonable tools to model telephony workload and performance (e.g., Erlang distributions). Unfortunately, the literature in Internet traffic characterization, both in the analytical and performance measurement domains, indicates that wide-area networking technology has advanced at a far faster rate than has the analytical and theoretical understanding of Internet traffic behavior.

The slower and more containable realms of years ago were amenable to characterization with closed-form mathematical expressions, which allowed reasonably accurate prediction of performance metrics such as queue lengths and network delays. But traditional mathematical modeling techniques, e.g., queuing theory, have met with little success in today's Internet environments.

For example, the assumption of Poisson arrivals was acceptable for the purposes of characterizing small LANs years ago. As a theory of network behavior, however, the tenacity of the use of Poisson arrivals, whether in terms of packet arrivals within a connection, connection arrivals within an aggregated stream of traffic, or packet arrivals across multiple connections, has been quite remarkable in the face of its egregious inconsistency with any collected data<sup>3,4</sup>. Leland et al.<sup>5</sup> and Paxson and Floyd<sup>6</sup> investigate alternatives to Poisson modeling, specifically the use of self-similarity (fractal) mathematics to model IP traffic.

There is still no clear consensus on how statistics can support research in IP traffic modeling, and there is skepticism within the community regarding the utility of empirical studies that rely on collecting real data from the Internet; i.e., some claim that since the environment is changing so quickly, any collected data are only of historical interest within weeks. There are those whose research is better served by tractable mathematical models than by empirical data that represent at most only one stage in network traffic evolution.

A further contributing factor to the lag of Internet traffic modeling behind that of telephony traffic is the early financial structure of the Internet. A few U.S. government agencies assumed the financial burden of building and maintaining the transit network infrastructure, leaving little need to trace network usage for the purposes of cost allocation. As a result Internet customers did not have much leverage with their service provider regarding the quality of service.

Many of the studies for modeling telephony traffic came largely out of Bell Labs, which had several advantages: no competition to force slim profit margins, and therefore the financial resources to devote to research, and a strong incentive to fund research that could ensure the integrity of the networks for which they charge. The result is a situation today where telephone company tables of "acceptable blocking probability" (e.g., inability to get a dialtone when you pick up the phone) reveal standards that are significantly higher than our typical expectations of the Internet.

We do not have the same situation for the developing Internet marketplace. Instead we have dozens of Internet providers, many on shoestring budgets in low margin competition, who view statistics collection as a luxury that has never proven its utility in Internet operations. How will statistics really help keep the NAP alive and robust since traffic seems to change as fast as they could analyze it anyway?

We are not implying that the monopoly provider paradigm is better, only observing aspects of the situation that got us where we are today: we have *no way to predict, verify, or in some cases even measure* Internet service quality in real time.

There is some hope that some of the larger telecommunication companies entering the marketplace will eventually devote more attention to this area. The pressure to do so may not occur until the system breaks, at which point billed customers will demand, and be willing to pay for, better guarantees and data integrity.

Cost-benefit tradeoff: undertake enough network research to secure a better understanding of the product the NAPs sell, without draining operations of resources to keep the network alive. Failing that, fund enough research to be able to show that the NAPs are being good community members, contributing to the "advancement of Internet technology and research" with respect to understanding traffic behavior, at the least cost in terms of time and effort taken away from more critical engineering and customer service activities.

### Technical Issues

With deployment of the vBNS and NAPs, the situation grows even more disturbing. The national information infrastructure continues to drive funding into hardware, pipes, and multimedia-capable tools, with very little attention to any kind of underlying infrastructural sanity checks.

And until now, the primary obstacles to accessing traffic data in order to investigate such models have been political, legal (privacy), logistic, or proprietary.

With the transition to ATM and high-speed switches, it will *no longer even be technically feasible* to access IP layer data in order to do traffic flow profiling, certainly not at switches within commercial ATM clouds. The NAPs were chartered as *layer 2 entities*, that is, to provide a service at the link layer without regard for higher layers. Because most of the NSFNET statistics reflected information at and above layer 3 (i.e., the IP layer), the NAPs cannot use the NSFNET statistics collection architecture<sup>7</sup> as a model upon which to base their own operational collection. Many newer layer 2 switches (e.g., DEC gigaswitch, ATM switches) have little if any capability for performing layer 3 statistics collection, or even looking at traffic in the manner allowed on a broadcast medium (e.g., FDDI, Ethernet), where a dedicated machine can collect statistics without interfering with switching. Statistics collection functionality in newer switches takes resources directly away from forwarding of frames/cells, driving customers toward switches from competing vendors who sacrifice such functionality in exchange for speed.

**Cost-benefit tradeoff: (minimalist approach) Collect at least cost what is necessary to switching performance.**

### Ethical Issues

Privacy has always been a serious issue in network traffic analysis, and the gravity of the situation only increases at NAPs. Regardless of whether a NAP is a layer 2 entity or a broadcast medium, Internet traffic is a private matter among Internet clients. Most NAP providers have agreements with their customers not to reveal information about individual customer traffic. Collecting and using more than basic aggregate traffic counts will require precise agreements with customers regarding what may be collected and how it will be used. Behaving out of line with customers' expectations or ethical standards, even for the most noble of research intentions, does not bode well for the longevity of a service provider.

An extreme position is to not look at network header data (which incidentally is very different from user data, which we do not propose examining) at all because it violates privacy. Analogous to unplugging one's machine from the Ethernet in order to make it secure, this approach is effective in the short term but has some undesirable side effects. We need to find ways to minimize exposure rather than surrendering the ability to understand network behavior. It seems that no one has determined an "optimal" operating point in terms of what to collect, along any of the dimensions we discuss. Indeed, the optimal choice often depends on the service provider and changes with time and new technologies.

We acknowledge the difficulty for the NAPs, as well as any Internet service provider, to deal with statistics collection at an already very turbulent period of Internet evolution. However, it is at just such a time, marked ironically with the cessation of the NSFNET statistics, that a baseline architectural model for statistics collection is most critical, so that customers can trust the performance and integrity of the services they procure from their network service providers, and so that service providers do not tie their hands behind their backs in terms of being able to preserve robustness, or forfend demise, of their own clouds.

**Cost-benefit tradeoff: accurately characterize the workload on the network so that NAPs and NSPs can optimize (read, maintain) their networks, but at the least cost to these privacy ethics we hold so dear.**

### UTILITY OF ANALYSIS: EXAMPLES

There is no centralized control over all the providers in the Internet. The providers do not always coordinate their efforts with each other, and quite often are in competition with each other. Despite all the diversity among the providers, the Internet-wide IP connectivity is realized via Internet-wide distributed routing, which involves multiple providers, and thus implies a certain degree of cooperation and coordination. Therefore, there is a need to balance the providers' goals and objectives against the public interest of Internet-wide connectivity and subscribers' choices. Further work is needed to understand how to reach the balance.<sup>8</sup>

Y. Rekhter

As the underlying network becomes commoditized, many in the community begin to focus on higher layer issues, such as the facilities, information, and opportunities for collaboration available on the network. In this context, facilities include supercomputers and workstations, information includes World Wide Web and gopher servers, and opportunities for collaboration include e-mail and multiuser domains (MUDs). One could think of these three categories as corresponding to three types of communication: machine-to-machine, people-to-machine, and people-to-people.

Within each category, multiple dimensions emerge:

- *Context*: static topical, such as cognitive neuroscience research or geographically topical, such as local news;
- *Temporal*: from permanent resources to dynamically created communication resources used for brief periods, e.g., distributed classroom lecture or seminar; and
- *Geographic distribution*: which may require transparency at times and boundary visibility at another.

As we build this higher layer infrastructure taking the lower layers increasingly for granted, the need for statistics collection and analysis is not diminished. On the contrary, it is even more critical to maintain close track of traffic growth and behavior in order to secure the integrity of the network. In this section we highlight several examples of the benefits of traffic characterization to the higher layers that end users care about.

### Long-Range Internet Growth Tracking

Few studies on national backbone traffic characteristics exist<sup>9</sup>, limiting our insight into the nature of wide area Internet traffic. We must rely on WAN traffic characterization studies that focus on a single or a few attachment points to transit networks to investigate shorter-term aspects of certain kinds of Internet traffic, e.g., TCP<sup>10</sup>, TCP and UDP<sup>11</sup>, and *dns*<sup>12</sup>.

The authors have devoted much attention in the last 2 years to investigating the usefulness, relevance, and practicality of a wide variety of operationally collected statistics for wide-area backbone networks. In particular, we have undertaken several studies on the extent to which the statistics that the NSFNET project has collected over the life of the NSFNET backbone are useful for a variety of workload characterization efforts. As the NSFNET service agreement ends, leaving the R&E portion of Internet connectivity in the hands of the commercial marketplace, we are without a common set of statistics, accessible to the entire community, that can allow even a rough gauge of Internet growth. We consider it important to establish a community-wide effort to support the aggregation of such network statistics data from multiple service providers, such as that being developed in the IETF's opstat group<sup>13</sup>. We view a consensus on some baseline statistics collection architecture as critical to Internet long-term stability.

### Service Guarantees and Resource Reservation

Another ramification of the transition of the R&E portion of Internet connectivity from the NSFNET service into a commercial marketplace is the need for a mechanism to compare the quality of service providers. Rather than procured via collaborative undertakings between the federal government and academia and industry, services in today's internetworking environments will be market commodities. Effective provision of those commodities will require the ability to describe Internet workload using metrics that will enable customers and service providers to agree on a definition of a given grade of service. Furthermore, metrics for describing the quality of connectivity will be important to market efficiency since they will allow customers to compare the quality of service providers when making procurement decisions.

In addition, users will want to be able to reserve bandwidth resources, which will require that the network provider have an understanding of the current traffic behavior in order to efficiently allocate reservations without leaving unused reserved bandwidth unnecessarily idle.



However, the research community has yet to determine such metrics, and the issue requires immediate attention. A precursor to developing common metrics of traffic workload is a greater understanding of network phenomena and characteristics. Insight into workload requires tools for effective visualization, such as representing the flow of traffic between service providers, or across national boundaries. Without such tools for monitoring traffic behavior, it is difficult for an Internet service provider to do capacity planning, much less service guarantees.

In addition to our studies of operational statistics, we have also undertaken several studies that collect more comprehensive Internet traffic flow statistics, and we have developed a methodology for describing those flows in terms of their impact on an aggregate Internet workload. We have developed a methodology for profiling Internet traffic flows which draws on previous flow models<sup>14</sup> and have developed a variety of tools to analyze traffic based on this methodology. We think that NAP and other network service providers would gain great insight and engineering advantage by using similar tools to assess and track their own workloads.

### Accounting

The ability to specify or reserve the services one needs from the network will in turn require mechanisms for accounting and pricing (else there is no incentive not to reserve all one can or not to use the highest priority). Many fear pricing will stifle the open, vibrant nature of the Internet community; we suggest that it may rather motivate the constructive exploration of more efficient network implementation of innovative networking applications over the Internet. Understanding the possibilities for resource accounting will also aid commercial network providers in the process of setting cost functions for network services.

### Web Traffic and Caching

The World Wide Web is another domain in which operational collection and analysis of statistics are vital to support of services. Similar to our NSFNET analysis work, we have explored the utility of operationally collected Web statistics, generally in the form of *http* logs. We analyzed two days of queries to the popular Mosaic server at NCSA to assess the geographic distribution of transaction requests. The wide geographic diversity of query sources and the popularity of a relatively small portion of the Web server file set present a strong case for deployment of geographically distributed caching mechanisms to improve server and network efficiency.

We analyzed the impact of caching the results of queries within the geographic zone from which the request was sourced, in terms of reduction of transactions with and bandwidth volume from the main server<sup>15</sup>. We found that a cache document timeout even as low as 1,024 seconds (about 17 minutes) during the two days that we analyzed would have saved between 40 percent and 70 percent of the bytes transferred from the central server. We investigated a range of timeouts for flushing documents from the cache, outlining the trade-off between bandwidth savings and memory/cache management costs. Further exploration is needed of the implications of this tradeoff in the face of possible future usage-based pricing of backbone services that may connect several cache sites.

Other issues that caching inevitably poses include how to redirect queries initially destined for a central server to a preferred cache site. The preference of a cache site may be a function of not only geographic proximity, but also current load on nearby servers or network links. Such refinements in the Web architecture will be essential to the stability of the network as the Web continues to grow, and operational geographic analysis of queries to archive and library servers will be fundamental to its effective evolution.

For very heavily accessed servers, one must evaluate the relative benefit of establishing mirror sites, which could provide easier access but at the cost of extra (and distributed) maintenance of equipment and software. However, arbitrarily scattered mirror sites will not be sufficient. The Internet's sustained explosive growth calls for an architected solution to the problem of scalable wide area information dissemination. While increasing network bandwidths help, the rapidly growing populace will continue to outstrip network and server

capacity as people attempt to access widely popular pools of data throughout the network. The need for more efficient bandwidth and server utilization transcends any single protocol such as *ftp*, *http*, or whatever protocol next becomes popular.

We have proposed to develop and prototype wide area information provisioning mechanisms that support both caching and replication, using the NSF supercomputer centers as "root" caches. The goal is to facilitate the evolution of U.S. information provisioning with an efficient national architecture for handling highly popular information. A nationally sanctioned and sponsored hierarchical caching and replicating architecture would be ideally aligned with NSF's mission, serving the community by offering a basic support structure and setting an example that would encourage other service providers to maintain such operations. Analysis of Web traffic patterns is critical to effective cache and mirror placement, and ongoing measurement of how these tools affect, or are affected by, Web traffic behavior is an integral part of making them an effective Internet resource.

### Regulation

Although in previous years we have investigated telecommunications regulation issues in both the federal and state arenas, we prefer instead to invest, and see others invest, in research that might diminish the need for regulatory attention to the Internet industry. For example, Bohn et al.<sup>16</sup> have proposed a policy for IP traffic precedence that can enable a graceful transition into a more competitively supplied Internet market that might reduce the need for federal interest in regulation. Their paper discusses how taking advantage of existing IP functionality to use multiple levels of service precedence can begin to address disparities between the requirements of conventional and newer and more highly demanding applications.

Our approach thus far been based on the belief that at least so far, the less regulation of the Internet—and in fact the more progress in removing regulatory barriers for existing telecommunications companies so they can participate more effectively in the Internet market—the better. However, regulation may be necessary in order to foster a healthy competitive network environment where consumers can make educated decisions regarding which network service providers to patronize. A key requirement in analyzing the relative performance of and customer satisfaction with network service providers is public availability of statistics that measure their capacity, reliability, security, integrity, and performance.

### RECOMMENDATIONS

In this section we highlight several examples of recommended statistics collection objects and tools.

Reaching consensus on the definition of a community-wide statistics collection architecture will require cooperation between the private and public sectors. We hope that key federally sponsored networks such as the NSF very high speed backbone network service (vBNS)<sup>17</sup> and the federally sponsored NAPs can serve as a role model in providing an initial set of statistics to the community and interacting with the research community to refine metrics as research reveals the relative utility of various ones.

### Existing Work

The last community-wide document of which we are aware was Stockman's RFC 1404<sup>18</sup>, "A Model for Common Operational Statistics." Since that time the Internet environment has changed considerably, as have the underlying technologies for many service providers such as the NAPs. As a result these specific metrics are not wholly applicable to every service provider, but they serve as a valuable starting point. We emphasize that the exact metrics used are not a critical decision at this point, since refinements are inevitable as we benefit from experience with engineering the technologies; what is essential is that we start with some baseline and create a community facility for access and development in the future.

From Stockman<sup>19</sup>:

The metrics used in evaluating network traffic could be classified into (at least) four major categories:

- Utilization metrics (input and output packets and bytes, peak metrics, per protocol and per application volume metrics)
- Performance metrics (round trip time [rtt] on different protocol layers, collision count on a bus network, ICMP Source Quench message count, count of packets dropped)
- Availability metrics (longer term) (line, route, or application availability)
- Stability metrics (short-term fluctuations that degrade performance) (line status transitions, fast route changes (flapping), routes per interface in the tables, next hop count stability, short-term ICMP anomalous behavior).

Some of these objects are part of standard simple network management protocol (SNMP) MIBs; others, of private MIBs. Others are not possible to retrieve at all due to technical limitations; i.e., measurement of a short term problematic network situation only exacerbates it or takes longer to perform than the problem persists. For example, counts of packets and bytes, for non-unicast and unicast, for both input and output are fairly standard SNMP variables. Less standard but still often supported in private MIBs are counts of packet discards, congestion events, interface resets, or other errors. Technically difficult variables to collect, due to the high resolution polling required, include queue lengths and route changes. Although such variables would be useful for many research topics in Internet traffic characterization, operationally collected statistics will likely not be able to support them. For example, one characteristic of network workload is "burstiness," which reflects variance in traffic rate. Network behavioral patterns of burstiness are important for defining, evaluating, and verifying service specifications, but there is not yet agreement in the Internet community on the best metrics to define burstiness. Several researchers<sup>20</sup> have explored the failure of Poisson models to adequately characterize the burstiness of both local and wide-area Internet traffic. This task relies critically on accurate packet arrival time stamps, and thus on tools adequate for packet tracing of the arrivals of packets at high rates with accurate (microsecond) time granularities. Vendors may still find incentive in providing products that can perform such statistics collection, for customers that need fine-grained examination of workloads.

The minimal set of metrics recommended for IP providers in Stockman<sup>21</sup> were packets and bytes in and out (unicast and non-unicast) of each interface, discards in and out of each interface, interface status, IP forwards per node, IP input discards per node, and system uptime. All of the recommended metrics were available in the Internet standard MIB. The suggested polling frequency was 60 seconds for unicast packet and byte counters, and an unspecified multiple of 60 seconds for the others. Stockman also suggested aggregation periods for presenting the data by interval: over 24-hour, 1-month, and 1-year periods, aggregate by 15 minutes, 1 hour, and 1 day, respectively. Aggregation includes calculating and storing the average and maximum values for each period.

### Switched Environments

In switched environments, where there is no IP layer information, the above statistics are not completely applicable. Without demand from their customer base, many switch vendors have put collection of statistics at a second priority since it tends to detract from forwarding performance anyway. Some ATM switches can collect per-VC (virtual circuit) statistics such as those described in the ATM MIB<sup>22</sup>.

One alternative for providers that support IP over an internal ATM infrastructure is to collect the IP statistics described above, and in fact use objects such as host-host matrices to plan what number and quality of ATM virtual circuits might be necessary to support the IP traffic workload. For switched FDDI environments, the provider could collect statistics on each LAN coming into the switch, and collapse it during analysis to determine possible compound effects within the switch, in addition segmenting traffic by interface, customer, and perhaps by protocol/application. If there is no access to network layer information, such as at NSF NAPs or certain ATM switches, the network service provider will still have an interest in these statistics, since sorting the

resulting arrays would give the NSP an indication of what fraction of traffic comes from what number of users, which may be critical for planning switched or permanent virtual circuit configuration, and by the same token for accounting and capacity planning purposes. However, converting virtual circuits to end customers, most likely IP customers for the near future, requires maintaining mappings to higher layers.

### Dedicated Studies of IP Workloads

Even with a solid set of operational statistics there are times when one wants dedicated collection to gain greater insight into short-term dynamics of workloads. For example, there are limitations of the operationally collected statistics for the NSFNET for describing flows in terms of their impact on an aggregate Internet workload<sup>23</sup>. We have developed tools for supporting operational flow assessment, to gain insight into both individual traffic signatures as well as heavy aggregations of end users. We have tested our methodology using packet header traces from a variety of Internet locations, yielding insight far beyond a simple aggregated utilization assessment, into details of the composition of traffic aggregation, e.g., what components of the traffic dominate in terms of flow frequency, durations, or volumes. We have shown that shifts in traffic signatures as a result of evolving technologies, e.g., toward multimedia applications, will require a different approach in network architectures and operational procedures. In particular, the much higher demands of some of these new applications will interfere with the ability of the network to aggregate the thousands of simultaneous but relatively short and low-volume flows that we observe in current environments.

The methodology<sup>24</sup> defines a flow based on actual traffic activity from, to, or between entities, rather than using the explicit setup and teardown mechanisms of transport protocols such as TCP. Our flow metrics fall into two categories: metrics of individual flows and metrics of the aggregate traffic flow. Metrics of individual flows include flow type, packet and byte volume, and duration. Metrics of the aggregate flow, or workload characteristics seen from the network perspective, include counts of the number of active, new, and timed out flows per time interval; flow interarrival and arrival processes; and flow locality metrics. Understanding how individual flows and the aggregate flow profile influence each other is essential to securing Internet stability, and requires ongoing flow assessment to track changes in Internet workload in a given environment.

Because flow assessment requires comprehensive and detailed statistics collection, we recognize that the NAPs and other service providers may not be able to afford to continuously monitor flow characteristics on an operational basis. Nonetheless we imagine that NAP operators will find it useful to undertake traffic flow assessment at least periodically to obtain a more accurate picture of the workload their infrastructure must support. The methodology and tools that implement it<sup>4</sup> will be increasingly applicable, even on a continuous basis, for NAP tasks such as ATM circuit management, usage-based accounting, routing table management, establishing benchmarks by which to shop for equipment from vendors, and load balancing in future Internet components.

The methodology can form a complementary component to other existing operational statistics collection, yielding insights into larger issues of Internet evolution, e.g., how environments of different aggregation can cope with contention for resources by an ever-changing composition and volume of flows. Internet traffic cross-section and flow characteristics are a moving target, and we intend that our methodology serve as a tool for those who wish to track and keep pace with its trajectory. For example, as video and audio flows and even single streams combining voice and audio become more popular, Internet service providers will need to parametrize them to determine how many such end user streams they will be able to support and how many more resources each new such stream would require. Multicast flows will also likely constitute an increasingly significant component of Internet traffic, and applying our methodology to multicast flows would be an important step toward coping with their impact on the infrastructure.

### Community Exchange

A key requirement in analyzing the relative performance of and customer satisfaction with network service providers is public availability of statistics that measure their capacity, reliability, security, integrity, and performance. One vehicle conducive to the development of an effective marketplace is a client/server-based architecture for providing access to statistics of various NSPs. Each NSP would support a server that provides query support for collected statistics for its clients or potential customers. MCI provides this functionality for statistics collected on the NSF-sponsored vBNS network. Community-wide participation in such an open forum would foster a healthy competitive network environment where consumers could make educated decisions regarding which network service providers to patronize.

### CONCLUSION

When we get through we won't be done.

—Steve Wolff, then director, DNCRI, NSF on the NSFNET transition (one among many since 1983) at 1994 Farnet meeting

We conclude by mentioning an issue of recent popular interest as well as symbolic of the larger problem: Internet security and prevention of criminal behavior. Ironically, workload and performance characterization issues are inextricably intertwined with security and privacy. Much of the talk about the Internet's inherent insecurity due to the inability to track traffic at the required granularity is misleading. It is not an *inability* to examine traffic operationally that has prevented it thus far, whether for security or workload characterization (and the same tools could do both), but rather its priority relative to the rest of the community research agenda.

As a result, the gap has grown large between the unambiguous results of confined experiments that target isolated environments, and the largely unknown characteristics of the extensive Internet infrastructure that is heading toward global ubiquity.

Empirical investigation and improved methodology for doing so can improve current operational statistics collection architectures, allowing service providers to prepare for more demanding use of the infrastructure and allowing network analysts to develop more accurate Internet models. In short, we can contribute to a greater understanding of real computer networks of pervasive scale by reducing the gaps among (1) what network service providers need; (2) what statistics service providers can provide; and (3) what in-depth network analysis requires.

We encourage discussion as soon as possible within the community on developing a policy on statistics collection/exchange/posting of available NAP/Internet service provider statistics, with supporting tools to allow greater understanding of customer requirements and service models, equitable cost allocation models for Internet services, verification that a given level of service was actually rendered, and evolution toward a level of Internet performance that matches or surpasses that of most telecommunication systems today.

### AUTHOR INFORMATION

Hans-Werner Braun is a principal scientist at the San Diego Supercomputer Center. Current research interests include network performance and traffic characterization, and working with NSF on NREN engineering issues. He also participates in activities fostering the evolution of the national and international networking agenda. San Diego Supercomputer Center, P.O. Box 85608, San Diego, CA 92186-9784; email address: [hwb@sdsc.edu](mailto:hwb@sdsc.edu).

Kimberly Claffy received her doctoral degree from the Department of Computer Science and Engineering at the University of California, San Diego in June 1994, and is currently an associate staff scientist at the San Diego Supercomputer Center. Her research focuses on establishing and improving the efficacy of traffic and performance characterization methodologies on wide-area communication networks, in particular to cope with the

changing traffic workload, financial structure, and underlying technologies of the Internet. San Diego Supercomputer Center, P.O. Box 85608, San Diego, CA 92186-9784; email address: [kc@sdsc.edu](mailto:kc@sdsc.edu).

## REFERENCES

- [1] Braun, H.-W., C.E. Catlett, and K. Claffy. 1995. "<http://vedana.sdsc.edu/>," tech. rep., SDSC and NCSA, March. San Diego Supercomputer Center.
- [2] Braun, H.-W., C.E. Catlett, and K. Claffy. 1995. "National Laboratory for Applied Network Research," tech. rep., SDSC and NCSA, March. San Diego Supercomputer Center.
- [3] Caceres, R., P. Danzig, S. Jamin, and D. Mitzel. 1991. "Characteristics of Wide-area TCP/IP Conversations," Proceedings of ACM SIGCOMM '91, pp. 101–112, September.
- [4] Paxson, V. 1994. "Empirically-Derived Analytic Models of Wide Area TCP Connections," IEEE/ACM Transactions on Networking 2(4), August.
- [5] Leland, W., M. Taqqu, W. Willinger, and D. Wilson. 1994. "On the Self-similar Nature of Ethernet Traffic (extended version)," IEEE/ACM Transactions on Networking, February.
- [6] Paxson, V., and S. Floyd. 1994. "Wide-area Traffic: The Failure of Poisson Modeling," Proceedings of ACM SIGCOMM '94, February.
- [7] Claffy, K., H.-W. Braun, and G.C. Polyzos. 1993. "Long-term Traffic Aspects of the NSFNET," Proceedings of INET '93, pp. CBA—1:10, August.
- [8] Rekhter, Y. 1995. "Routing in a Multi-provider Internet," Internet Request for Comments Series RFC 1787, April.
- [9] See Claffy, K., H.-W. Braun, and G.C. Polyzos. 1993. "Long-term Traffic Aspects of the NSFNET," Proceedings of INET '93, pp. CBA—1:10, August; Davis, M. 1988. "Analysis and Optimization of Computer Network Routing," unpublished Master's thesis, University of Delaware; Heimlich, S. 1988. "Traffic Characterization of the NSFNET National Backbone," Proceedings of the 1990 Winter USENIX Conference, December; Claffy, K., G.C. Polyzos, and H.-W. Braun. 1993. "Traffic Characteristics of the T1 NSFNET Backbone," Proceedings of IEEE Infocom 93, pp. 885–892, March; Claffy, K. 1994. "Internet Workload Characterization," Ph.D. thesis, University of California, San Diego, June; and Claffy, K., H.-W. Braun, and G.C. Polyzos. 1994. "Tracking Long-term Growth of the NSFNET Backbone," Communications of the ACM, August, pp. 34–45.
- [10] Caceres, R., P. Danzig, S. Jamin, and D. Mitzel. 1991. "Characteristics of Wide-area TCP/IP Conversations," Proceedings of ACM SIGCOMM '91, pp. 101–112, September.
- [11] Schmidt, A., and R. Campbell. 1993. "Internet Protocol Traffic Analysis with Applications for ATM Switch Design," Computer Communications Review, April, pp. 39–52.
- [12] Danzig, P.B., K. Obraczka, and A. Kumar. 1992. "An Analysis of Wide-area Name Server Traffic," Proceedings of ACM SIGCOMM '92, August.
- [13] Wolff, H., and the IETF Opstat Working Group. 1995. "The Opstat Client-server Model for Statistics Retrieval," April, draft-ietf-opstat-client-server-03.txt.
- [14] See Caceres, R., P. Danzig, S. Jamin, and D. Mitzel. 1991. "Characteristics of Wide-area TCP/IP Conversations," Proceedings of ACM SIGCOMM '91, pp. 101–112, September; and Jain, R., and S.A. Routhier. 1986. "Packet Trains—Measurement and a New Model for Computer Network Traffic," IEEE Journal on Selected Areas in Communications, Vol. 4, September, pp. 986–995.
- [15] Braun, H.-W., and K. Claffy. 1994. "Web Traffic Characterization: An Assessment of the Impact of Caching Documents from NCSA's Web Server," Second International World Wide Web Conference, October.
- [16] Bohn, R., H.-W. Braun, K. Claffy, and S. Wolff. 1994. "Mitigating the Coming Internet Crunch: Multiple Service Levels via Precedence," Journal of High Speed Networks, forthcoming. Available by anonymous ftp from <ftp.sdsc.edu/pub/sdsc/anr/papers/> and <http://www.sdsc.edu/0/SDSC/Research/ANR/kc/precedence/precedence.html>.
- [17] Braun, H.-W., C.E. Catlett, and K. Claffy. 1995. "<http://vedana.sdsc.edu/>," tech. rep., SDSC and NCSA, March. San Diego Supercomputer Center.
- [18] Stockman, B. 1993. "A Model for Common Operational Statistics," RFC 1404, January.
- [19] Stockman, B. 1993. "A Model for Common Operational Statistics," RFC 1404, January.
- [20] See Willinger, W. 1994. "Self-similarity in High-speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements," Statistical Science; Garrett, M.W., and W. Willinger. 1994. "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," Proceedings of ACM SIGCOMM '94, September; and Paxson, V., and S. Floyd. 1994. "Wide-area Traffic: The Failure of Poisson Modeling," Proceedings of ACM SIGCOMM '94, February.
- [21] Stockman, B. 1993. "A Model for Common Operational Statistics," RFC 1404, January.

- [22] Ahmed, M., and K. Tesink. 1994. "Definitions of Managed Objects for ATM Management, Version 7.0," Internet Request for Comments Series RFC 1695, August.
- [23] Claffy, K., G.C. Polyzos, and H.-W. Braun. 1995. "Internet Traffic Flow Profiling," IEEE Journal on Selected Areas in Communications, forthcoming.
- [24] Claffy, K., G.C. Polyzos, and H.-W. Braun. 1995. "Internet Traffic Flow Profiling," IEEE Journal on Selected Areas in Communications, forthcoming.

### NOTES

1. See <http://www.merit.edu> for more information on the NSFNET transition.
2. Specifically, NSF-sponsored regional providers, i.e., those that received funding from the NSF throughout the life of the NSFNET, will only continue to receive funding if they connect to all three NSF NAPs. Even so, the funding will gradually diminish within 4 years, an interval of time in which regional providers will have to become fully self-sustaining within the marketplace.
3. North American Network Operators Group, International Engineering Planning Group, Federal Engineering and Planning Group, Engineering and Operations Working Group, Federal Association of Regional Networks.
4. The software we used is available at <ftp://ftp.sdsc.edu/pub/sdsc/anr/software/Flows.tar.Z>.

## 12

# NII Road Map: Residential Broadband

*Charles N. Brownstein*

*Cross-Industry Working Team*

*Corporation for National Research Initiatives*

The Cross-Industry Working Team (XIWT) is in the process of examining how today's national information infrastructure (NII) can evolve into the future NII envisioned in the XIWT white paper, "An Architectural Framework for the National Information Infrastructure." This paper presents the methodology the XIWT is employing in this effort and some of the results obtained in residential broadband. In this context residential broadband refers to communications-based services using large amounts of access bandwidth to provide NII services to residential users.

### METHODOLOGY

To understand NII evolution, it is instructive to examine the various overlapping and interdependent technology and business segments making up the NII. Examples of these segments include residential broadband, satellite-based communications, intelligent transportation system, personal communications services, enterprise networking, home automation, public telephony, electronic commerce, and the Internet. Each of these segments will undergo important evolutionary changes during the next decade. By piecing together the evolutionary directions of these segments, a vision of the overall NII evolution will emerge.

The initial portion of this study has not been to determine how the NII should evolve, but rather to understand the current technical and business directions of the various NII segments. No a priori assumption was made that these segments would evolve in a consistent fashion or that the result would meet the Clinton administration's or industry's vision for the NII. Indeed, although certain segments of the NII are overlapping and interdependent, others are evolving with apparently little interaction. One of the goals of this study is to identify where inconsistencies exist between the evolutionary directions of the NII segments and to identify the degree to which the combined segments miss the vision put forward for the future NII. Based on the results of this study, specific recommendations will be developed for reshaping evolutionary directions toward the target vision.

To understand how these segments are likely to evolve, industry experts from the relevant segments of the NII were invited to a series of NII evolution workshop meetings of the XIWT. These industry experts were asked to address three questions: (1) How would you characterize today's NII with regard to this industry segment? (2) What changes do you forecast for this industry segment over the next 3 years? and (3) What changes do you forecast will occur over the next 6 years?

The 3- and 6-year time horizons were found to be convenient milestones. Most industry planners have well-articulated views of how their industry will change during the next 3 years based on existing exploratory efforts and announced business plans. Changes likely to occur in 6 years are harder to predict; still, this is within the planning horizon of most companies.

These industry experts were asked to present their most likely scenarios—the one they would bet on occurring, not a fanciful view of what could be possible given dedicated industry and government efforts. These forecasts are predicated on whatever regulatory and policy changes these industry experts believe will probably occur during this planning period.



## RESIDENTIAL BROADBAND EVOLUTION

Many services envisioned for the NII will demand significant amounts of communications bandwidth. To extend these services to the general public, broadband access networks to private residences and apartments will play a critical role in the NII. Residential broadband (RBB) networks are already widely deployed for entertainment video distribution services. Future residential broadband networks will provide increased bandwidth and two-way interactive capabilities supporting a wide variety of applications.

To characterize today's RBB networks and to understand how they are evolving to provide the capabilities needed in the future NII, the XIWT invited industry experts representing CATV companies, local exchange carriers, RBB equipment manufacturers, and satellite communications service providers to discuss current and future RBB networks.

The following is a summary of the views of these industry experts. It does not necessarily represent the views or positions of the XIWT or its member companies.

### RESIDENTIAL BROADBAND TODAY

#### Access Architecture

Today's residential broadband (RBB) is composed of over-the-air broadcast networks, CATV networks, microwave access networks, and direct reception from home satellite antennas. With the exception of emerging satellite-based delivery systems, today's RBB access networks are based on 6-Mhz analog channels. In a recent study of CATV networks conducted by CableLabs, typical downstream capacities were as follows:

- 22 percent have less than 30 channels;
- 64 percent have 30 to 53 channels; and
- 14 percent have 54 channels.

Although the amplifier housings employed in current CATV networks are designed to accommodate a return path amplifier (i.e., they are two-way ready), most of today's CATV systems have unactivated return channels. Roughly 20 percent of today's CATV systems use some fiber-optic links to bypass long amplifier chains in the trunk portion of the network. Currently a mix of 300-, 400-, 450-, and 550-MHz amplifiers is used. Service is typically provided to residences and apartments, with relatively few business locations connected to CATV networks. There is usually only a single CATV operator in a given service area, with nascent competition from microwave and direct broadcast satellite service providers. TVRO (television receive only) background antennas that are 1 to 2 meters in diameter are used by a small fraction of residential customers.

Services available over today's RBB networks typically consist of the following core set:

- Basic video;
- Subscription pay;
- Pay-per-view;
- Special events; and
- Shopping channels.

In addition, the following emerging services have been deployed on a limited basis:

- Near video on demand;
- Electronic video guides;
- Low-speed and high-speed data access;
- Digital video services via high-power satellites; and
- High-speed, downlink-only data via high-power satellites.

### **Residential Broadband: 3-Year View of Access Architecture**

Three years from now most CATV networks are expected to be 750-MHz systems providing 77 analog channels in the 54- to 550-MHz band and hundreds of digital channels in the 550- to 750-MHz band. Most CATV networks in that time frame will have fully activated return paths using the 5- to 40-MHz band. These systems will have a maximum chain of four to five amplifiers providing improved quality and reliability. There will be a clustering of cable systems in many metropolitan areas to provide connectivity to business and residential customers throughout these metropolitan areas.

Local exchange carriers will provide video dialtone (VDT) service in many metropolitan areas using either hybrid fiber-coax access or fiber-to-the-curb architectures. These VDT networks will include a level-1 gateway for selecting the video information programmer. The video information programmer will provide a level-2 gateway for selection of the specific video channel. Most VDT networks will include three to five level-2 providers, including the LEC itself.

High-power direct broadcast satellites allowing reception from small antennas (those less than two feet in diameter) will be widely available for supporting video and data services to residences. Terrestrial radio frequency access networks using cellular principles in the 30- to 40-GHz band will be deployed in a number of areas.

By 1998, many of today's emerging RBB services will be considered core services of most RBB networks. The following core RBB services are expected to be available:

- Basic video;
- Subscription pay;
- Pay-per-view;
- Special events;
- Shopping channels;
- Near video on demand;
- Electronic video guides;
- Low-speed and high-speed data access;
- Digital video services via high-power satellites; and
- High-speed, downlink-only data via high-power satellites.

In addition, the following emerging services are expected to be deployed on a limited basis in 1998:

- Telephony for second and third lines;
- Video telephony;
- Entrepreneurs providing niche information services (e.g., travel services);
- Interactive games;
- Greater variety of shopping services; and
- More education applications.

### **Residential Broadband: 6-Year View of Architecture**

From 1998 to 2001, fiber will migrate closer to residences, allowing available bandwidth to be shared among fewer customers and providing increased upstream bandwidth.

Despite current experimental deployments of 1-GHz cable systems, CATV networks in 2001 are expected to be predominantly composed of 750-MHz fiber serving area systems. Fiber node sizes will be reduced to 125 to 250 homes, with a maximum of 1 or 2 amplifier cascades with a move toward entirely passive networks. To provide additional upstream capacity, the analog portion of the spectrum could be reduced to 88 to 400 MHz and

the return path expanded to 5 to 70 MHz. Satellites with onboard ATM switching will be deployed supporting two-way bandwidth on demand.

By 2001, the following are expected to be core services of RBB networks:

- Basic video;
- Subscription pay;
- Pay-per-view;
- Special events;
- Shopping channels;
- Community bulletin boards;
- Electronic program guides;
- Near video on demand;
- High- and low-speed data services;
- Video shopping malls;
- Competitive access systems;
- Video conferencing;
- Overwhelming merchandising including electronic coupons;
- Telemedicine to the home; and
- Do-it-yourself guides (e.g., auto repair).

### SUMMARY

This paper presents the methodology being used by the XIWT in its NII evolution study and summarizes information collected on RBB evolution. The XIWT continues to collect information on industry evolution plans in other NII segments. This information will be included in a forthcoming white paper on NII evolution.

## 13

### The NII in the Home: A Consumer Service

*Vito Brugliera, Zenith Electronics*

*James A. Chiddix, Time Warner Cable*

*D. Joseph Donahue, Thomson Consumer Electronics*

*Joseph A. Flaherty, CBS Inc.*

*Richard R. Green, Cable Television Laboratories*

*James C. McKinney, ATSC*

*Richard E. Ottinger, PBS*

*Rupert Stow, Rupert Stow Associates*

#### INTRODUCTION

The present infrastructure for terrestrial broadcast and cable delivery to the home is analog based and is thus incompatible and noninteroperable with the concept of the national information infrastructure (NII), which is based on digital signal processing and transmission. However, the potential economic and technical advantages of digital transmission are so appealing that the transition to digital delivery techniques is inevitable.

The central problems associated with the transition to digital broadcast transmission to the home are consumer based. First, unlike all other services to the home, broadcasting is a mature industry with near universal access to the American home, and with some 170 million television receivers installed. Each household has on average 1.9 sets. Thus, even if the total present annual output of 28 million televisions was converted to digital receivers, 9 years would elapse before the analog service could be abandoned and digital service take its place. If universal access is a defining condition of the NII, then the present broadcast and cable-TV services together constitute the national information infrastructure.

Second, the terrestrial broadcast service that the consumer now enjoys is free, and the public expects it to remain so. Cable TV is not free, but 63 percent of television households (TVHH) are willing to pay \$300 per annum to receive over 40 additional channels of programming in addition to the delivery by cable of all local broadcast stations. In fact about 60 percent of all TVHHs view broadcast programs via cable, often because the quality of the broadcast signal delivered in metropolitan areas is deficient.

It is therefore clear that the consumer will be willing to make an investment in digital receiving equipment only if a service is provided that offers uniquely attractive features and a superior quality of picture and sound.

This paper, then, explores the means by which the NII can help to provide a significantly improved service to the present broadcast and cable universe, ideally with no increase in cost to the consumer. New services, functions, and means of delivery will be considered, but only in the context of providing them for all consumers at a price that they are willing and able to pay, and at a cost that is economically viable for both the program and service providers.

The primary service provided by broadcast and cable is passive entertainment with a wide range of program choice. The issue to be addressed is whether and how the NII can be adapted to continue this service while improving both its quality and diversity.

#### BACKGROUND

The broadcast and cable industries today share three major features: universal access, an entertainment service, and localism.

### Universal Access

The existing infrastructure serves almost all citizens. Ninety-eight percent of all households have televisions, including more than 15 million households rated as below the poverty level of income. Ninety-six percent have telephone service, and 63 percent subscribe to cable-TV service, which is available to 95 percent of TVHHs. In those areas where broadcast television reception is poor, the delivery of the broadcast signal via cable is an important complementary function of value to the consumer. In addition, cable TV provides over 40 cable networks on the average system, together with some distant and out-of-market broadcast stations. Thus the home consumer has a wide choice of programs.

The telephone system provides fully interactive voice communication and, with the addition of a fax machine, enables interactive communication of text and graphics. For other interactive services, such as home shopping programs delivered by cable TV, the consumer can use the telephone to place orders for the products selected.

For all these consumer services, the broadcast, cable, and telephone media have essentially met the goal of universal access. A historical account of the growth of the penetration of these media into the home is presented in Figure 1. The number of TVHHs increases at a rate of about 1 million per year, generally keeping pace with the growth in the number of households.

The measure of the "television of abundance" is seen in Table 1, where the services listed are the average of those provided by each of the top 60 television markets, a grouping that encompasses 73 percent of all TVHHs. In this average market there are eight local television stations.

Cable-TV systems provide 42 basic cable networks catering to a diversity of special interests, commonly called niche markets. These include typically a 24-hour national news service, local news, sports, talk shows, music video, documentaries, history, natural history, the arts, cooking, home improvement instruction, congressional proceedings, weather, and home shopping.

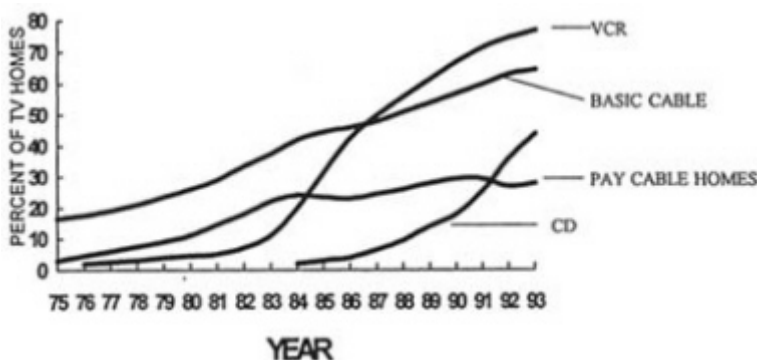


Figure 1 Penetration of consumer devices and services—video cassette recorders, compact disk players, basic cable, and pay cable—into U.S. households, 1975 to 1993.

TABLE 1 The Television of Abundance (Top 60 Markets)

Service	Average Number Provided by Each Top Market	
Current		
Broadcast	8	Local stations
Cable	42	Basic cable networks
	10	Pay-cable networks
	8	Pay-per-view channels
	5	Out-of-market broadcast stations
	3	Public access channels
Direct broadcast satellite	150	Channels
Total	226	Channels
Coming	40	Near-video-on-demand channels
TOTAL	266	Channels

In addition to carrying the eight local broadcast stations, five distant out-of-market broadcast stations are carried. Public access channels are also available to anyone wishing to address the audience on any subject they please.

The cable system offers also 10 pay-cable networks, to which the consumer may subscribe for a fee of approximately \$100 per annum per network. These pay-cable networks offer recent feature films, in addition to the production and origination of over 60 movies per year, at a production cost of \$3 million to \$5 million each. The pay-TV networks also acquire the transmission rights to other programs produced elsewhere.

Eight pay-per-view channels are offered, giving access to recent feature films, sporting events, and music spectaculars, for which a charge of about \$5 is levied for each program selected.

Direct broadcast satellite (DBS) service is now becoming generally available, offering some 150 channels, and while the market penetration is less than 1 percent of TVHHs, it is growing steadily. The programs offered are mainly those offered by pay-TV together with many pay-per-view channels. DBS is available and is being installed by a broad cross section of consumers, including many who have cable service, and among many of the 37 percent of TVHHs that do not or cannot subscribe to cable services, including the 5 percent of homes that are not passed by cable.

The existing infrastructure thus offers a total of 226 channels for the consumer's selection. When this infrastructure is converted to digital transmission and the use of digital compression, this number may be increased severalfold, and the so-called "500-channel universe" will have arrived, with perhaps hundreds of channels available for near-video-on-demand, or NVOD.

For the practical purposes of the home consumer, the services of terrestrial broadcast, cable, and the telephone together constitute a national information infrastructure with almost universal access.

### An Entertainment Service

The successful and deep market penetration of terrestrial broadcasting and cable stems in part from their satisfying the consumer's primary desire for passive entertainment, and not interactivity. In addition to the mass-market entertainment, niche-market entertainment is provided by cable-TV networks.

The second category of programming delivered by broadcast and cable is national and local news and information. The consumer's desire for this service is demonstrated by the fact that a broadcaster's local news service is the single most profitable type of programming. Indeed, television news is perceived by the consumer as the most trusted source of news, according to surveys conducted over the last decade.

Though the home television receiver is on for an average of 7 hours per day, it is in the evening prime-time hours that the audience is greatest, when 71 percent of the TVHHs, or 58 million people, are viewing the screen. The vast majority of this audience watches passive entertainment and sports.

The share of this active audience viewing each of the main delivery services is shown in Figure 2. The market demonstrates that entertainment remains the most desired service and that broadcasting and cable meet this need.

While the digital revolution may offer new services that enhance the entertainment value of programs delivered to the home, entertainment will remain the key element of such services if they are to succeed in the marketplace.

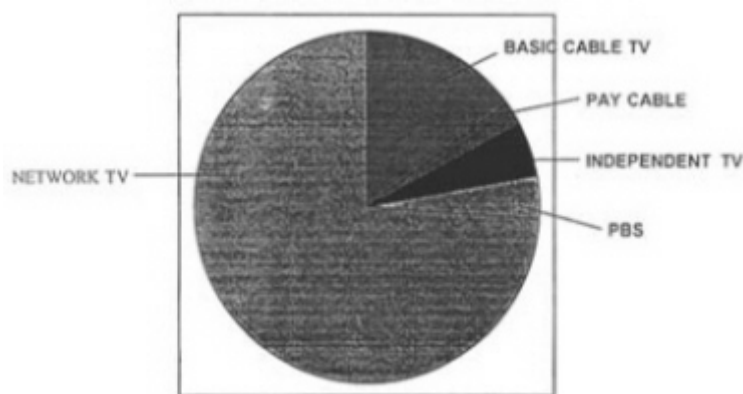


Figure 2 Share of viewing audience during prime time for each of the main categories of program providers, 1994.

### Localism

The ability of broadcasting and cable to deliver local news, public affairs, information, and sports has proven to be critical to the profitability of local broadcast stations. These services, supported by local commercial advertising, have proved vital in cementing the social fabric of a community. In times of crisis or local disaster, local broadcasters and cable TV provide a vital service to the populace.

In planning the development of the NII as a universal service, it is therefore essential that the infrastructure be compatible with the delivery of local content for each market and community.

### ANALYSIS AND PROJECTION

The plethora of new services and functions postulated for the home consumer, all enabled by applications of digital technology and their integration into the new infrastructure, should be examined in the context of three factors: cost to the consumer and the desires of the consumer, technical and economic feasibility, and regulatory issues.

#### Cost to and Desires of the Consumer

Whatever new services and functions the NII can deliver to the consumer, his willingness to acquire them will be a balance between desire and cost. The current market trials of new services are intended to demonstrate both consumer acceptance and willingness to pay.

In the existing infrastructure serving the home consumer, commercial television broadcasting is unique in that it is a "free" service, essentially paid for by advertising. The real cost to the consumer is the hidden cost he or she must pay for advertised products, a price that covers the additional cost of advertising.

By contrast, basic cable service costs the consumer about \$300 per annum, a price that 63 percent of TVHHs are willing to pay for the better reception of broadcast signals and for the additional programs provided by the cable networks and the local cable systems.

The declining rate of growth of subscribers to basic cable TV (see [Figure 1](#)) may well be influenced by the price barrier of \$300 per annum, with one out of every three households where cable-TV service is available declining to subscribe. It is unlikely to be due to any lack of variety or the entertainment value of the programs offered.

Pay-cable or "premium" channels are available to the consumer for an additional fee. Generally free of commercial advertising, each pay-cable channel costs about \$100 per annum. Despite the convenience and variety offered, this additional cost appears to be limiting the growth of this market. [Figure 1](#) demonstrates a low rate of penetration and growth for pay-cable networks. Over the last several years, the penetration of pay-cable has stabilized at 26 percent of TVHHs, or 41 percent of basic cable subscribers, having fallen from 57 percent over the last 10 years.

While the number of pay-cable networks offered has more than doubled in the past 10 years, the average number of pay-cable networks subscribed to by the consumer has risen only from 1.5 to 2.0.

From the above, it would appear that, for the types of programming and service now offered the consumer, the willingness or ability to pay is reaching its limit. Any further outlays by the consumer must therefore be based on more attractive programming or innovative services, perhaps made possible and economically viable by the NII. However, the NII only facilitates the delivery of information to and from the home, while new and attractive programs and services must still be made by the production and creative community.

For any new services to succeed in the home market, the cost to the consumer will be of paramount importance. If the new services are attractive enough, the consumer may redirect current disposable income used for other services to acquire the new services.

Figure 3 shows the growth in consumer spending on the major entertainment services. Overall, these have increased by 90 percent over the last 10 years in constant dollars. The main use of the consumer's discretionary funds (71 percent) has been for basic cable and home video services, while the growth in pay-cable, the cinema, and pay-per-view has been minimal. In addition to the obvious value of basic cable to the consumer, the appeal of home video or video rental is interesting, because it meets a consumer demand for a direct choice of entertainment programming at a specific time. This suggests that if pay-cable, pay-per-view, and NVOD can be offered at a price competitive with the video store (about \$2 per program), much of the video rental spending could be diverted to these services. Further, this diversion of funds can be effective only if a comparable inventory of programs—perhaps 40 at any one time—is made available.

Given these important conditions, an NVOD or pay-per-view service could enjoy spectacular growth at the expense of home video. For the program or content provider, significant economies in the number of cassettes manufactured and distributed are possible, because only a relatively small number will be required for distribution to the consumer. However, the consumer will still have to bear the cost of the cable system's right to transmit the program, together with the operational costs of storing and delivering the program.

In summary, any considerations of a different infrastructure for the delivery of entertainment to the home should be judged on the basis of the cost to the consumer, however great the convenience of using the service may be.

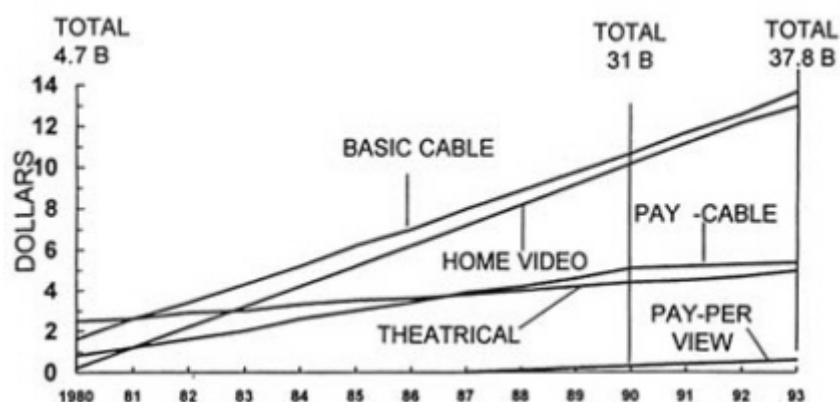


Figure 3 Consumer spending on major entertainment services, 1980 to 1993.  
SOURCE: P. Kagan Associates.

### Technical and Economic Feasibility

It is a common mantra that "people do not watch and listen to technology; they watch and listen to programs." While this is a self-evident truth, all systems for watching and listening to programs must pass through the critical funnel of technology. Experience has shown that the consumer is willing to pay for those applications of technology that offer superior performance and quality. The classic example is the market success of the compact disk. The music may be the same as that recorded on vinyl or audiocassette, but the quality of the sound is markedly improved, and CD penetration rates have been phenomenal.

We discuss below the economic feasibility of some of the proposed services that the NII might deliver in terms of the cost to the consumer and the cost of delivering the services.

### Digital HDTV

In pursuit of improved visual quality, the television industry has, since its inception, continuously strived to achieve better quality in the transmitted signal. If the desired gain in quality has not reached the consumer, it is because of the fundamental constraints of an aged analog transmission standard.

The transition to digital transmission offers the prospect of higher-quality service to the home. In fact, digital transmission permits the selection of a quality of transmission appropriate to the requirements of the signal



being transmitted and its intended use. To this end, a basic digital transmission format has been developed, containing a hierarchy of related and scalable transmission formats, each with a particular data bit rate appropriate to the requirements of the signal to be transmitted and to the bandwidth of the system under consideration.

The top member of this hierarchy is the high-definition television format developed by the Grand Alliance, a consortium of seven leading organizations under the aegis of the FCC's Advisory Committee on Advanced Television Service, or ACATS. This standard has been adopted by the Advanced Television Systems Committee, or ATSC.

The original objective of Japanese research in 1970 was to develop a signal origination and transmission system that would provide the full quality of the cinema experience in the living room, complete with CD quality sound. When, in 1987, the FCC formed the ACATS, its mandate was to study and evaluate proposed systems, and to recommend a proposed terrestrial transmission standard for the delivery to the home of a greatly improved and advanced television service, intended to replace the present standard, all "in the public interest."

The development work and the construction of a prototype system have now been completed by the Grand Alliance, and, following final tests of the complete system, the ACATS will be in a position to formally recommend a standard to the FCC by the end of 1995. It is already known that the system works, and works well.

This first all-digital transmission system has been designed to be compatible and interoperable with the NII concept, as will be the lower orders of the hierarchy that may be used for the transmission of low-definition television, graphics, data, and text.

The development work and the testing have enjoyed the active participation of CableLabs, a cable industry research organization representing all the major cable system companies in the United States and in Canada. The cable industry has researched the means by which HDTV can be delivered to cable-TV subscribers, and, in technological terms, is ready to implement such a service.

Although representing a revolutionary advance in television technology, HDTV is not a new consumer service but rather a greatly improved and advanced television service, serving the same constituency of TVHHs and offering the same or similar types of programs as those transmitted today. However, with five times the amount of information on the display, with a color rendition at least the equal of film, with a wide-screen display similar to that of the cinema, and with multichannel digital sound, the sense of presence given the viewer has been likened to "a window on the world."

While this paper is concerned only with the home consumer's interests, it may be remarked that the development of HDTV is a vital enabling application for a wide range of other services, including education, distance learning, medicine, printing and publishing, and indeed the whole field of video telecommunications. All these applications are compatible with, and complementary to, the intended use of the NII. Further, educational and medical services can be of great value to the consumer at home.

Returning to the interests of the home consumer, audience research on the demand for HDTV is sparse, but a recent poll of 500 randomly selected adults is useful. When asked to rate in order of preference the value of 21 new and proposed digital-based services to the home, ranking each on a scale of 1 to 10, 51 percent placed HDTV in the top three rankings, followed by NVOD with 49 percent. All other video services received a much lower ranking.

It thus appears that, once again, the consumer cares greatly about quality and foresees a market for HDTV in the home. By what means HDTV can be delivered most effectively and economically to the home, and at an acceptable price to the consumer, is discussed below.

### Interactive Services

The perceived consumer demand for interactive services will clearly influence the future infrastructure serving the home. Apart from the present telephone service that employs symmetric information flow, nearly all other proposed services will require a much smaller information flow from the consumer than that delivered to the home. Such asymmetry of data flow is important to the economic design of the interactive service.

In the multichannel universe now developing, an interactive device is needed to enable the consumer to navigate with ease and speed among the options presented. While individual viewers have different and diverse sets of interests in programs, they all share a common characteristic. It is established that people, when faced with a large number of options, cannot make a choice directly but instead reduce the number intellectually to between seven and nine. They are then able to make a selection by choosing among this smaller number.

Television audience analysis bears this out and shows that the average viewer selects a small group of program sources, typically four broadcast networks, perhaps a local independent broadcast station, and two niche market cable networks that respond to particular interests. In the future it will be important for the consumer to record and store favorite types of programs in the electronic channel navigator. Following this interaction, the navigator device may display only those upcoming programs that match the consumer's tastes.

Electronic games require interactive responses at a low data rate. When games are played online against the program provider or other remote players, the infrastructure must provide for this communication.

Home shopping services on cable require consumer responses made upstream. Ideally, these responses may take the form of individual and specialized questions, formatted as an upstream data flow on the cable system, rather than through the use of a telephone.

Provision must be made also for demanding various and specific sets of textual data, whether it be particular stock prices, sports statistics, particular news items, or additional data relating to a program that the consumer is watching. It should be possible to record with the program provider particular types of information to be delivered on a regular basis and stored in the consumer's home for later retrieval. In all these cases, only a low bandwidth upstream data flow is required. Upstream commands for VOD and NVOD programs and for information about the programs should be accommodated by the cable system.

Overall, cable and DBS are well placed to support asymmetric interactive applications, and unlike the telephone system, they do not have to be switched.

The Clinton administration would like to see an NII that would allow individuals to be producers as well as consumers of information. Certainly such a facility would be valuable for the consumer who wishes to invest in distance learning or to access remote medical services or electronic publishing services.

However, the great majority of consumers are interested in receiving a broad selection of mostly passive entertainment programs with the highest possible quality and at the lowest possible cost. Even with new services such as the Internet, which offers a remarkable range of information services at a rapid rate of growth, it is reported that the home consumer frequently uses it as an entertainment medium, spending much time surfing the World Wide Web for amusement rather than edification.

### **The Consumer's Equipment in the Home**

Some 30 percent of TVHs now have a computer that is used for nonbusiness purposes. Telecommunications, broadcasting, cable, and the computer are converging toward the use of a common digital technology. As television becomes a purely digital medium, it will also become a digital data stream. At the top end of the spectrum hierarchy there will be digital HDTV, and at the bottom end, a telephone twisted pair occupying the smallest bandwidth.

In the home, however, the convergence will not be complete. The large television display will remain the appropriate device for the viewing of pictures and for entertainment however the signal is delivered to the home. The high-resolution computer, on the other hand, is designed for reading text and graphics, and for their creation and manipulation, with the viewer seated close to the smaller screen. Given their different functions and features, the two displays will remain separate devices, although they will, for some applications, have to be interconnected.

Given the several digital data delivery media entering the home—including cellular, broadcast, cable, telephone, and DBS—practicality, economy, and convenience demand a single set-top box for the decryption, decoding, and processing of the signals for display. Furthermore, the increasing need for interactive upstream communication suggests that, if it should ever achieve universal penetration, the computer, with its associated telephone modem, is the most efficient location for the multisignal converter functions. The convenience of the

armchair remote control facing the television display will remain for the viewer of entertainment programs, but some of the switching functions it originates would be performed in the computer-based signal converter.

In this concept, the leads from the terrestrial antenna, the satellite dish, and the cable feed would all terminate at the computer box, where appropriate circuits would process the signals for display at, in many cases, multiple viewing locations in the home, each requiring different signals for display. This aspect of convergence will minimize the cost of equipment to be acquired by the consumer.

Given the rising rate of technological advance, it is inevitable that, in the future, incoming signals will be processed by a continuing stream of new or improved circuits and storage media. It is unrealistic to expect that the consumer will be able to replace the television set with the frequency necessary to accommodate the latest advances in digital technology. The display and cabinet in the TV set constitute 75 percent of the total cost of the equipment. While the display must be replaced once to accept the 16:9 aspect ratio of picture to be provided by digital broadcast and cable services, it then should remain in service for several years.

The changes in signal processing equipment that will occur may be accommodated by some form of set-top box, equipped with modular components and circuit cards and provided by the program carrier. The consumer does not have to buy them, but in effect pays for them through a monthly subscription to the service, as is now the case with the cable services.

This situation will obtain only during the introductory period of service, but as volume builds, the most popular services will be received by television sets or computers with integrated signal processing facilities.

### **Delivery Media**

Prima facie, in the interests of overall economy, it would seem sensible to minimize the number of separate data delivery media entering the consumer's home, but the monopolistic dangers of a single carrier delivering all signals to the home are considerable, absent a strong regulatory overview. On balance, the merits of market competition for delivery of information to the home are compelling.

Such competition, now developing between DBS and cable TV, where the programs delivered are similar, can lead only to lower costs for the consumer. What temporary advantage DBS may gain in quality of picture and sound is balanced by the local broadcasting availability, the number of programs offered by cable, and its interactive capability.

Terrestrial broadcast service has the important advantage of being free of charge and has the vital attribute of localism, but it cannot match the variety of programs offered by cable.

Developments planned by the telephone companies can lead to direct competition with basic and pay-cable services. However, the phone companies have recently shown a reluctance to quickly enter the television delivery business.

A cellular video transmission system may also serve as a carrier of programs nominally broadcast by local television stations. The interactive capabilities of such systems would enable upstream commands to switch between program sources at a local switching node. In addition to providing a full range of all broadcast and cable feeds, the cellular system would enable the transmission of data, graphics, and compressed video to other subscribers, and enable access to many nonentertainment services conceived for the NII. In this scenario, the current local broadcaster would continue originating all local programming but would no longer endure the cost of terrestrial broadcasting.

Some elements of a national infrastructure will remain common. The program creators and their products are sought by all. Some of them will remain independent, selling their programs on an open market to the highest bidder. The delivery media may acquire some program creators and thus gain from the vertical integration of their business. This has occurred in television broadcasting and cable but has not impaired the competitive market, because the consumer can always select another program carrier for personal entertainment needs.

It is asserted that, in the interests of the consumer, an open and competitive marketplace for the delivery of programs to the home is essential. This is best assured by the existence of multiple content carriers.

### Regulatory Issues

As proposed above, we believe that the future infrastructure for communication to the home will develop in the consumer's best interests if competing information carriers vie for this business. Regulatory intervention should have the sole purpose of ensuring that such competition exists. Only where no competing carrier exists is rate regulation required. Where a new service to the consumer is uniquely proposed by a single organization, the regulatory authorities should act to encourage the entry of a competing entity offering the same service.

It is further arguable that regulation should work to inhibit or even prohibit any single entity from owning or controlling both program production and program delivery operations, if such control results in the consumer being denied access to any available program source. Consumer access should be a prime consideration.

### RECOMMENDATIONS

We believe that the services required most by the consumer include entertainment and local and national news. The consumer wants the highest technical quality of picture and sound and a wide range of choices of programming and expects these services to be available at minimal cost.

The infrastructure necessary to meet these needs should have the following features:

1. There should be multiple competing delivery systems entering the home, where it is economically feasible. A governmental role in ensuring a level playing field may be appropriate.
2. The delivery systems should be able to carry all services now provided by broadcast, cable, DBS, and computer data networks.
3. The delivery systems should be able to carry a range of digital data rates, including that necessary for the FCC's initiative for HDTV service.
4. The delivery systems, as an element of the NII, following a period of innovation and standards development, should be interoperable with other elements of the NII, enabling access to nonentertainment and information services.
5. The delivery systems should permit interactive operation by the consumer, through the use of a simple interface.
6. The cost of delivering all services to the consumer should be such that universal access is practicable and affordable. In particular, the delivery of current terrestrially broadcast programs should continue to be free to the consumer.
7. The existing cable infrastructure, which is broadly available to almost all television households, should continue to offer a cost-effective path to a broadband component of the NII, which builds on its existing broadband architecture.
8. With the exception of Recommendation 1 above, no governmental role is foreseen as desirable or necessary.

## 14

# Internetwork Infrastructure Requirements for Virtual Environments

*Donald P. Brutzman, Michael R. Macedonia, and Michael J. Zyda*  
*Naval Postgraduate School, Monterey, California*

### ABSTRACT

Virtual environments constitute a broad multidisciplinary research area that includes all aspects of computer science, virtual reality, virtual worlds, teleoperation, and telepresence. We examine the various network elements required to scale up virtual environments to arbitrarily large sizes, connecting thousands of interactive players and all kinds of information objects. Four key communications components for virtual environments are found within the Internet protocol (IP) suite: lightweight messages, network pointers, heavyweight objects, and real-time streams. We examine both software and hardware shortfalls for internetworked virtual environments, making specific research conclusions and recommendations. Since large-scale networked virtual environments are intended to include all possible types of content and interaction, they are expected to enable new classes of sophisticated applications in the emerging national information infrastructure (NII).

### Conclusions

- Virtual environments are an all-inclusive superset of NII 2000.
- Any workable solution must address scaling up to arbitrarily large sizes.
- Lower- and middle-network layer problems are basically solved.
- Four key communications components are used in virtual environments.
- Virtual Reality Modeling Language (VRML) will take the World Wide Web (WWW) into three dimensions.

### Recommendations

- Upper application layers relating to entity interactions need funded research.
- A Virtual Environment Network Testbed will enable research and testing.
- Behaviors must be added to VRML for interaction with 3D graphics objects.
- Virtual reality transport protocol (VRTP) will soon be needed.
- Compelling applications, not protocols, will drive progress.

### OVERVIEW

Virtual environments (VEs) and virtual reality applications are characterized by human operators interacting with dynamic world models of increasing sophistication and complexity (Zyda et al., 1993) (Durlach and Mavor, 1995). Current research in large-scale virtual environments can link hundreds of people and artificial agents with interactive three-dimensional (3D) graphics, massive terrain databases, and global hypermedia and scientific datasets. Related work on teleoperation of robots and devices in remote or hazardous

locations further extends the capabilities of human-machine interaction in synthetic computer-generated environments. The variety of desired connections between people, artificial entities, and information can be summarized by the slogan "connecting everything to everything." The scope of virtual environment development is so broad that it can be seen as an inclusive superset of all other global information infrastructure applications. As the diversity and detail of virtual environments increase without bound, network requirements become the primary bottleneck.

The most noticeable characteristic of virtual environments is interactive 3D graphics, which are ordinarily concerned with coordinating a handful of input devices while placing realistic renderings at fast frame rates on a single screen. Networking permits connecting virtual worlds with realistic distributed models and diverse inputs/outputs on a truly global scale. Graphics and virtual world designers interested in large-scale interactions can now consider the worldwide Internet as a direct extension of their computer. We show that a variety of networking techniques can be combined with traditional interactive 3D graphics to collectively provide almost unlimited connectivity. In particular, the following services are essential for virtual world communications: reliable point-to-point communications, interaction protocols such as the IEEE standard distributed interactive simulation (DIS) protocol, WWW connectivity, and multicast communications.

## EXISTING INFRASTRUCTURE TECHNOLOGIES

### Layered Models

The integration of networks with large-scale virtual environments occurs by invoking underlying network functions from within applications. Figure 1 shows how the seven layers of the well-known open systems interconnection (OSI) standard network model generally correspond to the effective layers of the IP standard. Functional characteristic definitions of the IP layers follow in Box 1.

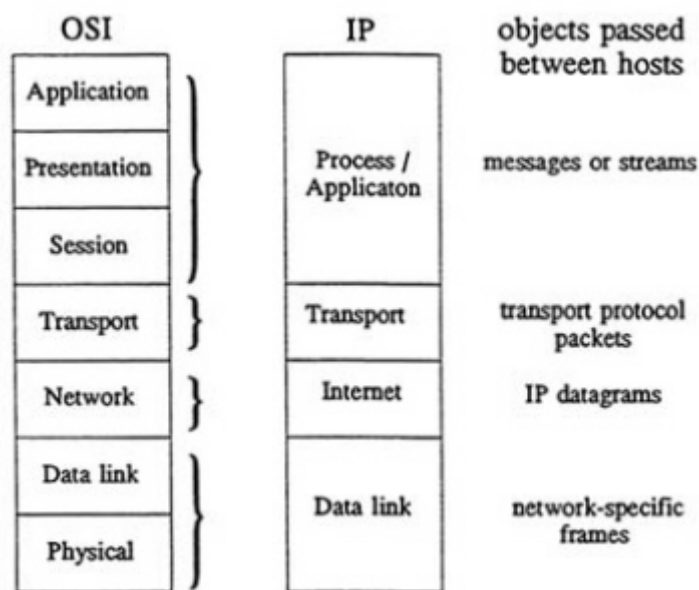


Figure Correspondence between OSI and IP protocol layer models, and objects passed between corresponding layers on separate hosts.

These diagrams and definitions are merely an overview but help illustrate the logical relationship and relative expense of different network interactions. In general, network operations consume proportionately more processor cycles at the higher layers. Minimizing this computational burden is important for minimizing latency and maintaining virtual world responsiveness.

### BOX 1 SUMMARY OF INTERNET PROTOCOL (IP) SUITE LAYER FUNCTIONALITY

- *Process/Application Layer.* Applications invoke TCP/IP services, sending and receiving messages or streams with other hosts. Delivery can be intermittent or continuous.
- *Transport Layer.* Provides host-host packetized communication between applications, using either reliable delivery connection-oriented TCP or unreliable delivery connectionless UDP. Exchanges packets end to end with other hosts.
- *Internet/Network Layer.* Encapsulates packets with an IP datagram that contains routing information; receives or ignores incoming datagrams as appropriate from other hosts. Checks datagram validity, handles network error and control messages.
- *Data Link/Physical Layer.* Includes physical media signaling and lowest level hardware functions; exchanges network-specific data frames with other devices. Includes capability to screen multicast packets by port number at the hardware level.

Methods chosen for transfer of information must use either reliable connection-oriented transport control protocol (TCP) or nonguaranteed delivery connectionless user datagram protocol (UDP). Each of these complementary protocols is part of the transport layer. One of the two protocols is used as appropriate for the criticality, timeliness, and cost of imposing reliable delivery on the particular stream being distributed. Understanding the precise characteristics of TCP, UDP, and other protocols helps the virtual world designer understand the strengths and weaknesses of each network tool employed. Since internetworking considerations affect all components in a large virtual environment, additional study of network protocols and applications is highly recommended for virtual world designers. Suggested references include Internet NIC (1994), Stallings (1994), and Comer (1991).

#### Internet Protocol

Although the variety of protocols associated with internetworking is very diverse, there are some unifying concepts. Foremost is "IP on everything," or the principle that every protocol coexists compatibly within the Internet Protocol suite. The global reach and collective momentum of IP-related protocols make their use essential and also make incompatible exceptions relatively uninteresting. IP and IP next generation (IPng) protocols include a variety of electrical, radio-frequency, and optical physical media.

Examination of protocol layers helps clarify current network issues. The lowest layers are reasonably stable with a huge installed base of Ethernet and fiber distributed data interface (FDDI) systems, augmented by the rapid development of wireless and broadband integrated services digital network (ISDN) solutions (such as asynchronous transfer mode [ATM]). Compatibility with the IP suite is assumed. The middle transport-related layers are a busy research and development area. Addition of real-time reliability, quality of service, and other capabilities can all be made to work. Middle-layer transport considerations are being resolved by a variety of working protocols and the competition of intellectual market forces. From the perspective of the year 2000, lower- and middle-layer problems are essentially solved.

#### Distributed Interactive Simulation

The DIS protocol is an IEEE standard for logical communication among entities in distributed simulations (IEEE, 1993). Although initial development was driven by the needs of military users, the protocol formally specifies the communication of physical interactions by any type of physical entity and is adaptable for general use. Information is exchanged via protocol data units (PDUs), which are defined for a large number of interaction types.

The principal PDU type is the Entity State PDU. This PDU encapsulates the position and posture of a given entity at a given time, along with linear and angular velocities and accelerations. Special components of an entity (such as the orientation of moving parts) can also be included in the PDU as articulated parameters. A full

set of identifying characteristics uniquely specifies the originating entity. A variety of dead reckoning algorithms permits computationally efficient projection of entity posture by listening hosts. Dozens of additional PDU types are defined for simulation management, sensor or weapon interaction, signals, radio communications, collision detection, and logistics support.

Of particular interest to virtual world designers is an open format Message PDU. Message PDUs enable user-defined extensions to the DIS standard. Such flexibility coupled with the efficiency of Internet-wide multicast delivery permits extension of the object-oriented message-passing paradigm to a distributed system of essentially unlimited scale. It is reasonable to expect that free-format DIS Message PDUs might also provide remote distributed connectivity resembling that of "tuples" to any information site on the Internet, further extended by use of network pointer mechanisms that already exist for the World Wide Web. This is a promising area for future work.

### **World Wide Web**

The World Wide Web (WWW or Web) project has been defined as a "wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents" (Hughes, 1994). Fundamentally the Web combines a name space consisting of any information store available on the Internet with a broad set of retrieval clients and servers, all of which can be connected by easily defined HyperText Markup Language (html) hypermedia links. This globally accessible combination of media, client programs, servers, and hyperlinks can be conveniently used by humans or autonomous entities. The Web has fundamentally shifted the nature of information storage, access, and retrieval (Berners-Lee et al., 1994). Current Web capabilities are easily used despite rapid growth and change. Directions for future research related to the Web are discussed in (Foley and Pitkow, 1994). Nevertheless, despite tremendous variety and originality, Web-based interactions are essentially client-server: A user can push on a Web resource and get a response, but a Web application can't independently push back at the user.

### **Multicast**

IP multicasting is the transmission of IP datagrams to an unlimited number of multicast-capable hosts that are connected by multicast-capable routers. Multicast groups are specified by unique IP Class D addresses, which are identified by  $1110_2$  in the high-order bits and correspond to Internet addresses 224.0.0.0 through 239.255.255.255. Hosts choose to join or leave multicast groups and subsequently inform routers of their membership status. Of great significance is the fact that individual hosts can control which multicast groups they monitor by reconfiguring their network interface hardware at the data link layer. Since datagrams from unsubscribed groups are ignored at the hardware interface, host computers can solely monitor and process packets from groups of interest, remaining unburdened by other network traffic (Comer, 1991; Deering, 1989).

Multicasting has existed for several years on local area networks such as Ethernet and FDDI. However, with IP multicast addressing at the network layer, group communication can be established across the Internet. Since multicast streams are typically connectionless UDP datagrams, there is no guaranteed delivery and lost packets stay lost. This best-effort unreliable delivery behavior is actually desirable when streams are high bandwidth and frequently recurring, in order to minimize network congestion and packet collisions. Example multicast streams include video, graphics, audio, and DIS. The ability of a single multicast packet to connect with every host on a local area network is good since it minimizes the overall bandwidth needed for large-scale communication. Note, however, that the same multicast packet is ordinarily prevented from crossing network boundaries such as routers. If a multicast stream that can touch every workstation were able to jump from network to network without restriction, topological loops might cause the entire Internet to become saturated by such streams. Routing controls are necessary to prevent such a disaster and are provided by the recommended multicast standard (Deering, 1989) and other experimental standards. Collectively the resulting internetwork of communicating multicast networks is called the Multicast Backbone (MBone) (Macedonia and Brutzman, 1994).



Improved real-time delivery schemes are also being evaluated using the real-time transport protocol (RTP), which is eventually expected to work independently of TCP and UDP (Schulzrinne and Casner, 1993). Other real-time protocols are also under development. The end result available today is that even with a time-critical application such as an audio tool, participants normally perceive conversations as if they are in ordinary real time. This behavior is possible because there is actually a small buffering delay to synchronize and resequence the arriving voice packets. Research efforts on real-time protocols and numerous related issues are ongoing, since every bottleneck conquered results in a new bottleneck revealed.

The MBone community must manage the MBone topology and the scheduling of multicast sessions to minimize congestion. Currently over 1,800 subnets are connected worldwide, with a corresponding host count equivalent to the size of the Internet in 1990. Topology changes for new nodes are added by consensus: A new site announces itself to the MBone mail list, and the nearest potential providers decide who can establish the most logical connection path to minimize regional Internet loading. Scheduling MBone events is handled similarly. Special programs are announced in advance on an electronic mail list and a forms-fed schedule home page. Advance announcements usually prevent overloaded scheduling of Internet-wide events and alert potential participants. Cooperation is key. Newcomers are often surprised to learn that no single person or authority is "in charge" of either topology changes or event scheduling.

### SOFTWARE INFRASTRUCTURE NEEDS

We believe that the "grand challenges" of computing today are not large static gridded simulations such as computational fluid dynamics or finite element modeling. We also believe that traditional supercomputers are not the most powerful or significant platforms. Adding hardware and dollars to incrementally improve existing expensive computer designs is a well-understood exercise. What is more challenging and potentially more rewarding is the interconnection of all computers in ways that support global interaction of people and processes. In this respect, the Internet is the ultimate supercomputer, the Web is the ultimate database, and any networked equipment in the world is a potential input/output device. Large-scale virtual environments attempt to simultaneously connect many of these computing resources in order to recreate the functionality of the real world in meaningful ways. Network software is the key to solving virtual environment grand challenges.

#### Four Key Communication Methods

Large-scale virtual world internetworking is possible through the application of appropriate network protocols. Both bandwidth and latency must be carefully considered. Distribution of virtual world components using point-to-point sockets can be used for tight coupling and real-time response of physics-based models. The DIS protocol enables efficient live interaction between multiple entities in multiple virtual worlds. The coordinated use of hypermedia servers and embedded Web browsers allows virtual worlds global input/output access to pertinent archived images, papers, datasets, software, sound clips, text, or any other computer-storable media. Multicast protocols permit moderately large real-time bandwidths to be efficiently shared by an unconstrained number of hosts. Applications developed for multicast permit open distribution of graphics, video, audio, DIS, and other streams worldwide in real time. Together these example components provide the functionality of lightweight messages, network pointers, heavyweight objects, and real-time streams (Box 2). Integrating these network tools in virtual worlds produces realistic, interactive, and interconnected 3D graphics that can be simultaneously available anywhere (Brutzman, 1994a,b; Macedonia, 1995; Macedonia et al., 1995).

#### Application Layer Interactivity

It is application layer networking that needs the greatest attention in preparing for the information infrastructure of the year 2000. DIS combined with multicast transport provides solutions for many

### BOX 2 FOUR KEY COMMUNICATIONS COMPONENTS USED IN VIRTUAL ENVIRONMENTS

- *Lightweight Interactions.* Messages composed of state, event, and control information as used in DIS Entity State PDUs. Implemented using multicast. Complete message semantics is included in a single packet encapsulation without fragmentation. Lightweight interactions are received completely or not at all.
- *Network Pointers.* Lightweight network resource references, multicast to receiving groups. Can be cached so that repeated queries are answered by group members instead of servers. Pointers do not contain a complete object as lightweight interactions do, instead containing only a reference to an object.
- *Heavyweight Objects.* Large data objects requiring reliable connection-oriented transmission. Typically provided as a WWW query response to a network pointer request.
- *Real-time Streams.* Live video, audio, DIS, 3D graphics images, or other continuous stream traffic that requires real-time delivery, sequencing, and synchronization. Implemented using multicast channels.

SOURCE: Macedonia (1995).

application-to-application communications requirements. Nevertheless DIS is insufficiently broad and not adaptable enough to meet general virtual environment requirements. To date, most of the money spent on networked virtual environments has been by, for, and about the military. Most of the remaining work has been in (poorly) networked games. Neither is reality. There is a real danger that specialized high-end military applications and chaotic low-end game "hacks" will dominate entity interaction models. Such a situation might well prevent networked virtual environments from enjoying the sustainable and compatible exponential growth needed to keep pace with other cornerstones of the information infrastructure.

#### Next-generation DIS

We believe that a successor to DIS is needed that is simpler, open, extensible, and dynamically modifiable. DIS has proven capabilities in dealing with position and posture dead reckoning updates, physically based modeling, hostile entity interactions, and variable latency over wide-area networks. DIS also has several difficulties: awkward extensibility, requiring nontrivial computations to decipher bit patterns, and being a very "big" standard. DIS protocol development continues through a large and active standards community. However, the urgent military requirements driving the DIS standard remain narrower than general virtual environment networking requirements.

A common theme that runs through all network protocol development is that realistic testing and evaluation are essential, because the initial performance of distributed applications never matches expectations or theory. A next-generation DIS research project ought to develop a "dial-a-protocol" capability, permitting dynamic modifications to the DIS specification to be transmitted to all hosts during an exercise. Such a dynamically adjustable protocol is a necessity for interactively testing and evaluating both the global and local efficiency of distributed entity interactions.

#### Other Interaction Models

Many other techniques for entity interaction are being investigated, although not always in relation to virtual environments. Intelligent agent interactions are an active area of research being driven by artificial intelligence and user interface communities. Rule-based agents typically communicate via a message-passing paradigm that is a natural extension of object-oriented programming methods. Common Gateway Interface (cgi) scripts function similarly, usually using hypertext transfer protocol (http) (Berners-Lee et al., 1994) query extensions as inputs. Ongoing research by the Linda project uses "tuples" as the communications unit for logical entity interaction, with particular emphasis on scaling up (Gelernter, 1992). MUDs (multiuser dungeons) and MOOs (MUDs object-oriented) provide a powerful server architecture and text-based interaction paradigm that is

well suited to support a variety of virtual environment scenarios (Curtis and Nichols, 1994). Passing scripts and interpretable source code over the network for automatic client use has been widely demonstrated for the multiplatform tool control language (Tcl) (Ousterhout, 1994). Recently the Java language has provoked interest over the possibility of simple and secure passing of precompiled program object files for multiplatform execution (Sun, 1995).

### Virtual Reality Modeling Language

The Web is being extended to three spatial dimensions thanks to virtual reality modeling language (VRML), a specification based on Silicon Graphics Inc. Open Inventor scene description language (Wernicke, 1994). Key contributions of the VRML 1.0 standard are a core set of object-oriented graphics constructs augmented by hypermedia links, all suitable for scene generation by browsers on PCs, Macintoshes, and Unix workstations. The current interaction model for VRML browsers is client-server, similar to most other Web browsers. Specification development has been effectively coordinated by mail list, enabling consensus by a large, active, and open membership (Pesce and Behlendorf, 1994; Pesca and Behlendorf, 1994–1995; and Bell et al., 1994).

Discussion has already begun on incorporating interaction, coordination, and entity behaviors into VRML 2.0. A great number of issues are involved. We expect that in order to scale to arbitrarily large sizes, peer-to-peer interactions will be possible in addition to client-server query-response. Although behaviors are not yet formally specified, the following possible view of behaviors extends the syntax of existing "engine" functionality in Open Inventor (Figure 2). Two key points in this representation follow. First, engine outputs only operate on the virtual scene graph, and so behaviors do not have any explicit control over the host machine (unlike CGI scripts). Second, behaviors are engine drivers, while engines are scene graph interfaces. This means that a wide variety of behavior mechanisms might stimulate engine inputs, including Open Inventor sensors and calculators, scripted actions, message passing, command line parameters, or DIS. Thus it appears that forthcoming VRML behaviors might simultaneously provide simplicity, security, scalability, generality, and open extensions. Finally, we expect that as the demanding bandwidth and latency requirements of virtual environments begin to be exercised by VRML, the client-server design assumptions of the HyperText Transfer Protocol (http) will no longer be valid. A Virtual Reality Transfer Protocol (vrtp) will be needed once we better understand how to practically deal with the new dynamic requirements of diverse interentity virtual environment communications.

### Vertical Interoperability

A striking trend in public domain and commercial software tools for DIS, Mbone, and the Web is that they can seamlessly operate on a variety of software architectures. The hardware side of vertical interoperability for virtual environments is simple: access to IP/Internet and the ability to render real-time 3D graphics. The software side is that information content and even applications can be found that run equivalently under PC, Macintosh, and a wide variety of Unix architectures. One important goal for any virtual environment is that human users, artificial entities, information streams, and content sources can interoperate over a range that includes highest-performance machines to least-common-denominator machines. Here are some success metrics for vertical interoperability: "Will it run on my supercomputer?" *Yes*. "Will it run on my Unix workstation?" *Yes*. "Will it also run on my Macintosh or PC?" *Yes*. This approach has been shown to be a practical (and even preferable) software requirement. Vertical interoperability is typically supported by open nonproprietary specifications developed by standardization groups such as the Internet Engineering Task Force (IETF).

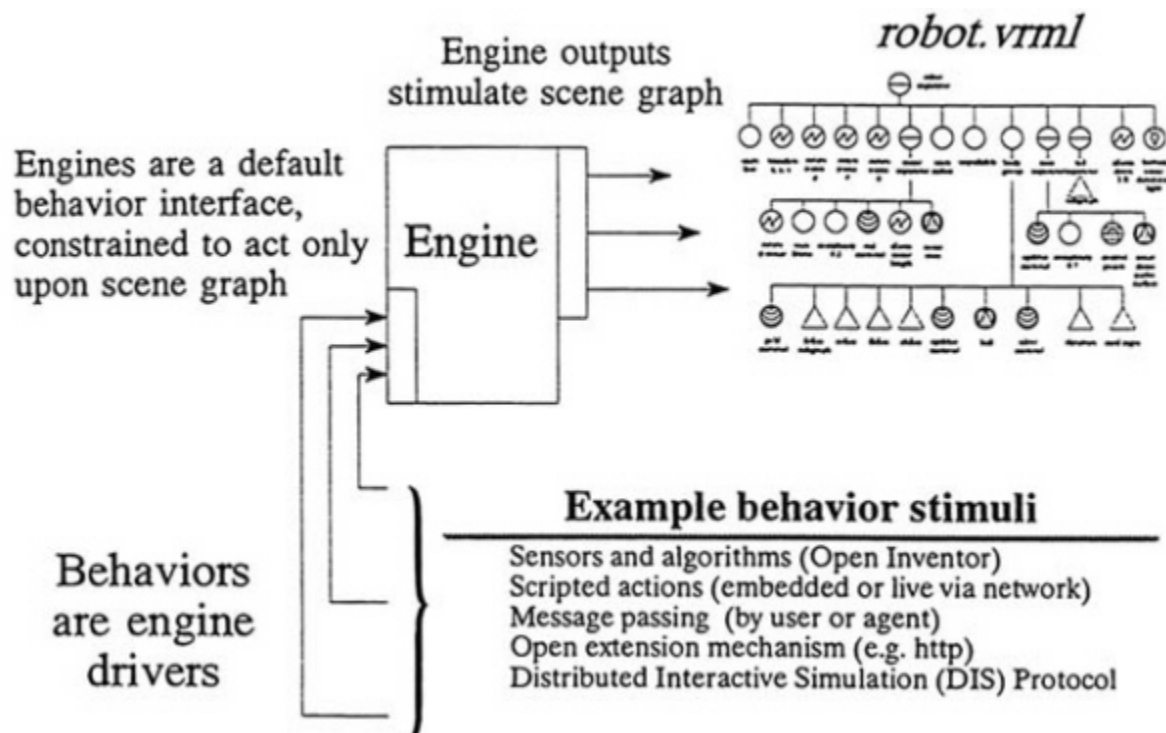


Figure 2 Proposed behavior interaction model for VRML.

## HARDWARE INFRASTRUCTURE NEEDS

### Research Testbed

The National Research Council report on virtual reality (Durlach and Mavor, 1995) made few recommendations for funding virtual environment hardware research due to active commercial development in most critical technologies. We agree with that assessment. However, the report also has a notable hardware-related recommendation regarding networks:

**RECOMMENDATION:** The committee recommends that the federal government provide funding for a program (to be conducted with industry and academia in collaboration) aimed at developing network standards that support the requirements for implementing distributed VEs [virtual environments] on a large scale. Furthermore, we recommended funding of an open VE network that can be used by researchers, at a reasonable cost, to experiment with various VE network software developments and applications. (Durlach and Mavor, 1995, p. 83)

The cost of high-speed network connections has precluded most academic institutions from conducting basic research in high-performance network applications. Those sites with high-performance connections are rarely free from the reliability requirements of day-to-day network operations. A national VE Network Testbed for academia and industry is proposed as a feasible collaboration mechanism. If rapid progress is expected before 2000, it is clearly necessary to decouple experimental network research from campus electronic mail and other essential services. The International Wide-Area Year (I-WAY) project is a proposed experimental national network that is applications-driven and ATM-based (I-WAY 95). It will connect a number of high-performance computing centers and supercomputers together. I-WAY may well serve as a first step in the direction of a national testbed, but additional efforts will be needed to connect institutions with lesser research budgets. Finally, it must be noted that design progress and market competition are bringing the startup costs of high-speed local

area networks (e.g., FDDI, ATM) within reach of institutional budgets. At most schools, it is the off-campus links to the Internet that need upgrading and funding for sustained use.

### **Other Problems**

In order to achieve broad vertical integration, it is recommended that proprietary and vendor-specific hardware be avoided. Video teleconferencing (VTC) systems are an example of a market fragmented by competing proprietary specifications. Broad interoperability and Internet compatibility are essential. Closed solutions are dead ends. In the area of new network services such as asynchronous transfer mode (ATM) and integrated services digital network (ISDN), some disturbing trends are commonplace. Supposedly standardized protocol implementations often do not work as advertised, particularly when run between hardware from different vendors. Effective throughput is often far less than maximum bit rate. Latency performance is highly touted and rarely tested. Working applications are difficult to find. Network operating costs are often hidden or ignored. Application developers are advised to plan and budget for lengthy delays and costly troubleshooting when working with these new services.

### **APPLICATIONS**

We believe that working applications—not theories and not hype—will drive progress. In this section we present feasible applications that are exciting possibilities or existing works in progress. Many new projects are possible and likely to occur by the year 2000 if virtual environment requirements are adequately supported in the information infrastructure.

#### **Sports: Live 3D Stadium with Instrumented Players**

Imagine that all of the ballplayers in a sports stadium wear a small device that senses location (through the Global Positioning System or local electrical field sensing) and transmits DIS packets over a wireless network. Similar sensors are embedded in gloves, balls, bats, and even shoes. A computer server in the stadium feeds telemetry inputs into a physically based, articulated human model that extrapolates individual body and limb motions. The server also maintains a scene database for the stadium complete with textured images of the edifice, current weather, and representative pictures of fans in the stands. Meanwhile, Internet users have browsers that can navigate and view the stadium from any perspective. Users can also tune to multicast channels providing updated player positions and postures along with live audio and video. Statistics, background information, and multimedia home pages are available for each player. Online fan clubs and electronic mail lists let fans trade opinions and even send messages to the players. Thus any number of remote fans might supplement traditional television coverage with a live interactive computer-generated view. Perhaps the most surprising aspect of this scenario is that all component software and hardware technologies exist today.

#### **Military: 100,000-Player Problem**

"Exploiting Reality with Multicast Groups" describes groundbreaking research on increasing the number of active entities within a virtual environment by several orders of magnitude (Macedonia, 1995; Macedonia et al., 1995). Multicast addressing and the DIS protocol are used to logically partition network traffic according to spatial, temporal, and functionally related entity classes. "Exploiting Reality" further explains virtual environment network concepts and includes experimental results. This work has fundamentally changed the distributed simulation community, showing that very large numbers of live and simulated networked players in real-world exercises are feasible.

### **Science: Virtual Worlds as Experimental Laboratories for Robots and People**

In separate work, we have shown how an underwater virtual world can comprehensively model all salient functional characteristics of the real world for an autonomous underwater vehicle (AUV) in real time. This virtual world is designed from the perspective of the robot, enabling realistic AUV evaluation and testing in the laboratory. Real-time 3D computer graphics are our window into that virtual world. Visualization of robot interactions within a virtual world permits sophisticated analyses of robot performance that are otherwise unavailable. Sonar visualization permits researchers to accurately "look over the robot's shoulder" or even "see through the robot's eyes" to intuitively understand sensor-environment interactions. Theoretical derivation of six-degrees-of-freedom hydrodynamics equations has provided a general physics-based model capable of replicating a highly nonlinear (yet experimentally verifiable) response in real time. Distribution of underwater virtual world components enables scalability and rapid response. Networking allows remote access, demonstrated via MBone audio and video collaboration with researchers at distant locations. Integrating the World Wide Web allows rapid access to resources distributed across the Internet. Ongoing work consists primarily of scaling up the types of interactions, datasets, and live streams that can be coordinated within the virtual world (Brutzman, 1994a,b).

### **Interaction: Multiple CAVEs using ATM and VRML**

A CAVE is a type of walk-in synthetic environment that replaces the four walls of a room with rear-projection screens, all driven by real-time 3D computer graphics (Cruz-Neira et al., 1993). These devices can accommodate 10 to 15 people comfortably and render high-resolution 3D stereo graphics at 15-Hz update rates. The principal costs of a CAVE are in high-performance graphics hardware. We wish to demonstrate affordable linked CAVEs for remote group interaction. The basic idea is to send graphics streams from a master CAVE through a high-speed, low-latency ATM link to a less expensive slave CAVE that contains only rear-projection screens. Automatic generation of VRML scene graphs and simultaneous replication of state information over standard multicast links will permit both CAVEs and networked computers to interactively view results generated in real time by a supercomputer. Our initial application domain is a gridded virtual environment model of the oceanographic and biological characteristics of Chesapeake Bay. To better incorporate networked sensors and agents into this virtual world, we are also investigating extensions to IP using underwater acoustics (Reimers and Brutzman, 1995). As a final component, we are helping establish an ambitious regional education and research network that connects scientists, students from kindergartens through universities, libraries, and the general public. Vertically integrated Web and MBone applications and a common theme of live networked environmental science are expected to provide many possible virtual world connections (Brutzman, 1995a,b).

### **CONCLUSIONS**

- Virtual environments are an all-inclusive superset of NII 2000.
- Any workable solution must address scaling up to arbitrarily large sizes.
- Lower- and middle-network layer problems are basically solved.
- Four key communications components are used in virtual environments:
  - Lightweight entity interactions (e.g., DIS PDUs);
  - Network pointers (e.g., URLs), usually passed within a lightweight entity interaction that includes identifying context;
  - Heavyweight objects (e.g., terrain databases or large textured scene graphs) which require reliable connections for accurate retrieval; and
  - Real-time streams (e.g., audio and video), usually via MBone.

- The Virtual Reality Modeling Language (VRML) will take the Web into three dimensions and make powerful applications available across platforms that were previously restricted to expensive graphics workstations.

### RECOMMENDATIONS

- It is the upper network application layers relating to entity interactions that need funded research work.
- A Virtual Environment Network Testbed is needed to enable research schools to cheaply connect and test experimental services without jeopardizing production networks.
- Behaviors are needed in VRML that allow 3D graphics objects to be driven by
  - Internal algorithms (e.g., Open Inventor sensors),
  - Scripted actions (embedded or live via network streams),
  - Message passing (by user or agent, e.g., MUDs/MOOs),
  - Open extension mechanism (e.g., http attributes), and
  - DIS (or superior) entity interaction protocol.
- Virtual Reality Transport Protocol (VRTP) will soon be needed:
  - Bandwidth, latency, and efficiency requirements will change relative to the design assumptions of current http servers;
  - Lightweight objects, heavyweight objects, multicast real-time streams;
  - Small sets of servers will combine to serve very large client/peer base.
- Compelling applications, not protocols, will drive progress. some examples:
  - Sports*: live 3D stadium with instrumented players; user chooses view;
  - Military*: 100,000-player problem;
  - Science*: virtual worlds as experimental laboratories for robots, people; and
  - Interaction*: affordable linked CAVEs for remote group interaction.

### PROJECTIONS

If one considers the evolving nature of the global information infrastructure, it is clear that there is no shortage of basic information. Quite the opposite is true. Merely by reading the *New York Times* daily, any individual can have more information about the world than was available to any world leader throughout most of human history! Multiply that single information stream by the millions of other information sources becoming openly available on the Internet, and it is clear that we do not lack *content*. Mountains of content have become accessible. What is needed now is *context*, a way to interactively locate, retrieve, and display the related pieces of information and knowledge that a user needs in a timely manner.

Within two lifetimes we have seen several paradigm shifts in the ways that people record and exchange information. Handwriting gave way to typing, and then typing to word processing. It was only a short while afterwards that preparing text with graphic images was easily accessible, enabling individuals to perform desktop publishing. Currently people can use 3D real-time interactive graphics simulations and dynamic "documents" with multimedia hooks to record and communicate information. Furthermore such documents can be directly distributed on demand to anyone connected to the Internet. In virtual environments we see a further paradigm shift becoming possible. The long-term potential of virtual environments is to serve as an archive and interaction medium, combining massive and dissimilar data sets and data streams of every conceivable type. Virtual environments will then enable comprehensive and consistent interaction by humans, robots, and software agents within those massive data sets, data streams, and models that recreate reality. Virtual environments can provide meaningful context to the mountains of content that currently exist in isolation without roads, links, or order.

What about scaling up? Fortunately there already exists a model for these growing mountains of information content: the real world. Virtual worlds can address the context issue by providing information links similar to those that exist in our understanding of the real world. When our virtual constructs cumulatively

approach realistic levels of depth and sophistication, our understanding of the real world will deepen correspondingly. In support of this goal, we have shown how the structure and scope of virtual environment relationships can be dynamically extended using feasible network communications methods. This efficient distribution of information will let any remote user or entity in a virtual environment participate and interact in increasingly meaningful ways.

Open access to any type of live or archived information resource is becoming available for everyday use by individuals, programs, collaborative groups, and even robots. Virtual environments are a natural way to provide order and context to these massive amounts of information. Worldwide collaboration works, for both people and machines. Finally, the network is more than *a* computer, and even more than *your* computer. The Internet becomes *our* computer as we learn how to share resources, collaborate, and interact on a global scale.

## REFERENCES

- Bell, Gavin, Anthony Parisi, and Mark Pesce, "The Virtual Reality Modeling Language (VRML) Version 1.0 Specification," draft, <http://www.eit.com/vrml/vrmlspec.html>, November 3, 1994.
- Berners-Lee, Tim, Luotonen Cailliau, Ari Nielsen, Henrik Frystyk, and Arthur Secret, "The World-Wide Web," *Communications of the ACM*, vol. 37, no. 8, August 1994, pp. 76–82.
- Brutzman, Donald P., "A Virtual World for an Autonomous Underwater Vehicle," *Visual Proceedings, Association for Computing Machinery (ACM) Special Interest Group on Computer Graphics (SIGGRAPH) 94*, Orlando, Florida, July 24–29, 1994a, pp. 204–205.
- Brutzman, Donald P., *A Virtual World for an Autonomous Underwater Vehicle*, Ph.D. Dissertation, Naval Postgraduate School, Monterey, California, December 1994b.
- Brutzman, Donald P., "Remote Collaboration with Monterey Bay Educators," *Visual Proceedings, Association for Computing Machinery (ACM) Special Interest Group on Computer Graphics (SIGGRAPH) 95*, Los Angeles, California, August 7–11, 1995.
- Brutzman, Donald P., "Networked Ocean Science Research and Education, Monterey Bay, California," *Proceedings of International Networking (INET) 95 Conference*, Internet Society, Honolulu, Hawaii, June 27–30, 1995. Available at <ftp://taurus.cs.nps.navy.mil/pub/i3la/i3laisoc.html>.
- Comer, Douglas E., *Internetworking with TCP/IP, Volume I: Principles, Protocols and Architecture*, second edition, Prentice Hall, Englewood Cliffs, New Jersey, 1991.
- Cruz-Neira, Carolina, Jason Leigh, Michael Papka, Craig Barnes, Steven M. Cohen, Sumit Das, Roger Engelmann, Randy Hudson, Trina Roy, Lewis Siegel, Christina Vasilakis, Thomas A. DeFanti, and Daniel J. Sandin, "Scientists in Wonderland: A Report on Visualization Applications in the CAVE Virtual Reality Environment," *IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, San Jose, California, October 25–26, 1993, pp. 59–66 and CP-3.
- Curtis, Pavel, and David A. Nichols, "MUDs Grow Up: Social Virtual Reality in the Real World," Xerox Palo Alto Research Center, Palo Alto, California, 1994. Available at <ftp://ftp.parc.xerox.com/pub/MOO/papers/MUDsGrowUp.ps>.
- Deering, Steve, "Host Extensions for IP Multicasting," Request for Comments (RFC) 1112, <ftp://ds.internic.net/rfc/rfc1112.txt>, August 1989.
- Durlach, Nathaniel I., and Anne S. Mavor, eds., *Virtual Reality: Scientific and Technological Challenges*, National Research Council, National Academy Press, Washington, D.C., 1995.
- Foley, Jim, and James Pitkow, eds., *Research Priorities for the World-Wide Web*, National Science Foundation (NSF) Information, Robotics and Intelligent Systems Division Workshop, Arlington, Virginia, October 31, 1994. Available at <http://www.cc.gatech.edu/gvu/nsf-ws/report/Report.html>.
- Gelernter, David, *Mirror World—Or the Day Software Puts the Universe in a Shoebox ... How It Will Happen and What it Will Mean*, Oxford University Press, New York, 1992.
- Hughes, Kevin, "Entering the World-Wide Web (WWW): A Guide to Cyberspace," Enterprise Integration Technology Inc., May 1994. Available at <http://www.eit.com/web/www/guide/>.
- IEEE Standard for Information Technology—Protocols for Distributed Interactive Simulation (DIS) Applications*, version 2.0, Institute for Simulation and Training report IST-CR-93-15, University of Central Florida, Orlando, Florida, May 28, 1993.
- Internet Network Information Center (NIC), *Request for Comments (RFC) archive*, <ftp://ds.internic.net>, 1994.
- International Wide-Area Year (I-WAY) project, 1995, information available at <http://www.iway.org>.



- Macedonia, Michael R., and Donald P. Brutzman, "MBone Provides Audio and Video Across the Internet," *IEEE COMPUTER*, April 1994, pp. 30–36. Available at <ftp://taurus.cs.nps.navy.mil/pub/i31a/mbone.html>.
- Macedonia, Michael R., *A Network Software Architecture for Large Scale Virtual Environments*, Ph.D. Dissertation, Naval Postgraduate School, Monterey, California, June 1995.
- Macedonia, Michael R., Michael J. Zyda, David R. Pratt, Donald P. Brutzman, and Paul T. Barham, "Exploiting Reality with Multicast Groups: A Network Architecture for Large-Scale Virtual Environments," *IEEE Computer Graphics and Applications*, 1995, to appear.
- Ousterhout, John K., *Tcl and the Tk Toolkit*, Addison-Wesley, Reading, Massachusetts, 1994.
- Pesce, Mark, and Brian Behlendorf, moderators, "Virtual Reality Modeling Language (VRML)," working group home page, <http://www.wired.com.vrml>, 1994–1995.
- Reimers, Stephen, and Donald P. Brutzman, "Internet Protocol over Seawater: Towards Interoperable Underwater Networks," *Unmanned Untethered Submersibles Technology 95*, Northeastern University, Nahant, Massachusetts, September 25–27, 1995, to appear.
- Schulzrinne, Henning, and Stephen Casner, "RTP: A Transport Protocol for Real-Time Applications," Audio-Video Transport Working Group, Internet Engineering Task Force, working draft, Oct. 20, 1993, available as <ftp://nic.ddn.mil/internet-drafts/draft-ietf-avt-rtp-04.ps>.
- Stallings, William, *Data and Computer Communications*, fourth edition, Macmillan, New York, 1994.
- Sun Microsystems Corporation, Java language home page, 1995, <http://java.sun.com/>.
- Wernicke, Josie, *The Inventor Mentor: Programming Object-Oriented 3D Graphics with Open Inventor™*, Release 2, Addison-Wesley Publishing, Reading, Massachusetts, 1994.
- Zyda, Michael J., David R. Pratt, John S. Falby, Paul T. Barham, Chuck Lombardo, and Kirsten M. Kelleher, "The Software Required for the Computer Generation of Virtual Environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 2, no. 2, MIT Press, Cambridge, Massachusetts, Spring 1993, pp. 130–140.

### AUTHOR INFORMATION

Don Brutzman is a computer scientist working in the Interdisciplinary Academic Group at the Naval Postgraduate School. His research interests include underwater robotics, real-time 3D computer graphics, artificial intelligence, and high-performance networking. He is a member of the Institute of Electrical and Electronic Engineers (IEEE), the Association for Computing Machinery (ACM) Special Interest Group on Computer Graphics (SIGGRAPH), the American Association for Artificial Intelligence (AAAI), the Marine Technology Society (MTS), and the Internet Society (ISOC).

Mike Macedonia is an active duty Army officer and Ph.D. candidate at the Naval Postgraduate School. He received a M.S. degree in telecommunications from the University of Pittsburgh and a B.S. degree from the U.S. Military Academy. His research interests include multicast data networks, real-time computer graphics, and large-scale virtual environments. His leadership on the Joint Electronic Warfare Center and CENTCOM staffs was instrumental in successfully deploying and integrating advanced computers, networks, and telecommunications systems during Operations Desert Shield and Desert Storm.

Mike Zyda is professor of computer science at the Naval Postgraduate School. His research interests include computer graphics, virtual world systems, and visual simulation systems. He is executive editor of the journal *PRESENCE: Teleoperators and Virtual Environments*, published by MIT Press. His recent accomplishments include organizing and serving as general chair for the 1995 ACM SIGGRAPH Symposium on Interactive 3D Graphics.

## 15

# Electric Utilities and the NII: Issues and Opportunities

*John S. Cavallini and Mary Anne Scott*

*U.S. Department of Energy*

*Robert J. Aiken*

*U.S. Department of Energy/Lawrence Livermore National Laboratory*

The electric utility industry is immersed in a changing environment, and deregulation is a key pressure driving this change. Careful strategic planning is needed to determine what the future business of the electric utilities will be, and there are concerns about the utilities' competitive status. Embedded in this restructuring evolution and coupled to the growing need to provide more in supply-side and demand-side management is the opportunity for the electric utility industry to become a significant player in the deployment of the national information infrastructure (NII). As an Electric Power Research Institute (EPRI) study concluded,

Energy production and delivery will be tightly coupled with telecommunications and information services for the foreseeable future. In order to control access to the customer and prevent erosion of their customer bases, utilities will be driven to become more aggressive in deploying both supply-side information technologies for improved operation of their generation, transmission, and distribution facilities; and demand-side Energy Information Services (EIS). Those information services will enable utilities to provide higher quality services at lower cost with lower environmental impact, and to give their rate payers better control over their power usage<sup>1</sup>.

The entry of the electric utility industry into the telecommunications arena, which is driven by the need to provide energy information services for the generation, delivery, and utilization of electric power, can affect competition in this market and contribute to the goal of universal service. However, significant policy implications must be addressed.

### AN INDUSTRY IN TRANSITION

After two decades in which competition in the energy market increased somewhat, the Energy Act of 1992 provided legislative authority to the Federal Energy Regulatory Commission (FERC) to mandate open wholesale transmission access. As a result, there is an even greater imperative for the electric utility industry to reexamine its business strategy. Utilities must determine how to remain competitive as this historically vertically integrated monopoly becomes restructured into focused organizations that will compete in the marketplace for energy supply and delivery services. The overall issue is how to provide high-quality energy cost effectively, while reducing the need for the capital investment of new generating capacity, cutting back on the depletion of nonrenewable fuels, and reducing emissions. Coupled with these elements is the need to focus on customer service and satisfaction. On an individual utility basis, the issue is deciding what part of the market to focus on, how to maintain and build the appropriate customer base, and how to maintain and increase market share. The utility choice of business lines includes generation, transmission, distribution, and value-added services, and these businesses may or may not be segregated into separate subsidiaries.

In a strongly regulated environment, industry has little incentive to be innovative—there is a vested interest in proven technology. The prevailing philosophy is that it's okay to be Number Two and that there is no need to lead the pack. However, the competitive, nonregulated marketplace encourages innovation, and industries adopt strategies to manage rather than avoid risks in their use of new technologies. The electric utility industry is

moving toward the latter and as such will be forced to consider the opportunities offered through telecommunications and information services to a much greater extent. As is discussed below, this is fortuitous for the deployment of the NII. It supports the vision that it is possible to create an information infrastructure and the tools to support smart energy decisions by all consumers and lessen the U.S. balance of payments and dependence on foreign energy sources.

Many electric utilities have extensive telecommunications facilities that they use for conducting their business—maintenance and operation of their generation and distribution systems to ensure reliability and quality of service to their customer base. But some utilities rely on services provided by telecommunications providers. Important assets that utilities bring to the table that could be used in providing telecommunications services for themselves or others include extensive rights-of-ways extending to businesses and residences, a ubiquitous presence in residential and business locations, and extensive system facilities (such as conduits, poles, transmission towers, and substation sites) that could be used in telecommunications networks.

The FERC recently announced a Notice of Proposed Rulemaking (NOPR) that outlines the mandate for the electric utilities to open their transmission facilities to all wholesale buyers and sellers of electric energy<sup>2</sup>. Included in the NOPR is a notice (RM95-9-00) of a technical conference on "real-time information networks" (RINs) that would give all transmission users simultaneous access to the same information under industry-wide standards. Thus it is clear that within a relatively short time frame, the need for this information system and telecommunications service will directly affect the 137 utilities required to open up their transmission facilities.

### **The Organization of the Electric Utility Industry**

There are three main categories of electric utilities today—the investor-owned utility (IOU), the municipal utility, and the rural cooperative. A fourth, often-discussed utility is the registered holding company (RHC), a subcategory of investor-owned utilities. These are the multistate, investor-owned holding companies that must be registered with the U.S. Securities & Exchange Commission (SEC) under the Public Utility Holding Company Act (PUHCA). Reform of this act is being considered.

When the electric utility industry was born late in the nineteenth century, it appeared first in the form of IOUs with the corresponding corporate charters shaping its existence. As electric service spread and came to be viewed as a necessity rather than a luxury, public discontent grew with the selective nature of the coverage provided by the IOUs. By 1900, many cities and some counties had created a municipal utility as a unit of the local government, with the charter to provide service to all its constituency. Financing was primarily through tax proceeds and future system revenues. Later, during the New Deal, the rural cooperative was born as a customer-owned, not-for-profit membership corporation. This progression of organizations was driven by requirements of customers and (to some extent) strategic objectives of the industry. This progression continues today, albeit in a somewhat different form.

About 95 percent of the U.S. population is served by the electric utilities<sup>3</sup>. Of these customers, 76.4 percent are served by IOUs, 13.7 percent by municipal utilities, and the remaining 9.9 percent by rural cooperatives<sup>4</sup>. Few can choose their supplier of electricity—in most cases it has to be the local provider. However, the possibility for options in choosing a provider offered through what is referred to as "retail wheeling" could make more competitive choices available<sup>5</sup>.

### **Utility Authority with Respect to Telecommunications**

Each utility organizational type functions differently within regulatory environments that vary according to jurisdictional boundaries. As discussed extensively in a report examining these issues, all have legal authority with respect to telecommunications to build infrastructure and to deliver at least some telecommunications-based services related to the delivery and consumption of energy<sup>6</sup>. It is legally sound, though possibly contentious, for a utility to build the facilities needed to communicate with its customers and possible suppliers and to develop the information services needed for effective energy management. However, the reaction of regulators and

competitors could be negative should the utility attempt to exploit the excess capacity created by using this infrastructure. The utility has at least three strong arguments to counter any challenges and support its position:

First, since utilities have clear rights to bring such facilities and enterprises into being, they have considerable leverage with regulators and competitors alike to achieve new arrangements that acknowledge the utilities' rights and interests and build upon them.

Second, since utilities have undeniable rights to read their own electric meters through telecommunications pursuant to their existing authorities, they may also have First Amendment Protections...to send any information they wish over their wires....

Finally, the nation's need to finance construction of the NII and make it universal gives electric utilities a compelling reason to gain regulatory favor for the use of and profit from their excess telecommunications capabilities<sup>6</sup>.

Of these arguments, the final is the most compelling, because it provides leverage that can further the goal of universal access for the NII, especially in rural areas. The cable industry and telecommunications providers are less likely to provide the necessary infrastructure in rural areas because of the low-density customer base.

### **Benefits of Energy Information Services**

There are several characteristics of energy supply, distribution, and consumption that can be and have been exploited to realize efficiencies and energy savings. The use of telecommunications-based information services could potentially enhance efficiency and savings substantially. From the industry's perspective there are many benefits to be accrued from the application of both supply- and demand-side information services. From the supply side, real-time information permits a shift to less conservative operations. Smoothing of peak power loads is possible, and spinning reserves can be reduced. Increased information can lead to reduced energy theft. Load balancing between utilities is feasible. Meters can be read remotely, and remote control of service is possible. Most importantly, power can be used more efficiently, reducing the need for new generating facilities.

For example, linking electronic sensors and automated control systems along transmission and distribution lines offers the possibility for applications such as the optimization and brokerage of dispatch and emissions credits, automatic meter reading, and remote service control. Similar deployment could be made in a commercial building with transmission to the utility of detailed information on end-use energy consumption, along with temperature and other relevant measurements. The analyses of such data could produce models that could be applied to improve utility load forecasting, improve building energy models, improve analysis of building energy efficiency, and diagnose electrical system problems in the building.

Energy consumption varies with the consumer as well as the time of day. Residential, commercial, and industrial sectors have differing requirements. The use of real-time pricing (RTP), which provides a direct relationship between the cost of electricity and its demand, can provide an incentive to encourage customers to conserve power when demand (and hence cost) is highest. This choice gives the customer a reasonable amount of discretionary control. The use of direct load management involves direct utility control of customer appliances to reduce consumption during periods of peak demand. Target appliances include those used for electrical heating, air conditioning, and hot-water heating. The demand-side management approach with the most stringent time response requirements is rapidly curtailable loads (RCL). Load must be shed within 20 seconds or less. Only large commercial and industrial customers currently have interruptible power tariffs. Customer applications are generally referred to as demand-side management (DSM). A more comprehensive term, energy information services (EIS), is used when DSM is coupled with remotely provided information services.

Although each of these approaches (as well as others) is currently being used, the benefits derived can be enhanced with the increased application of the evolving information and telecommunications technologies.

## ISSUES AND APPROACHES

The pressures of increased competition and rapidly evolving technology converge forcefully enough to mandate that utilities develop robust business strategies. There are also compelling arguments to justify the consideration of public policy issues related to electric utility involvement of the electric utility in an effective deployment of NII. Those arguments are tied to economics, technology, and regulatory issues. Utilities, legislators, and the public are struggling with the notion that utilities could provide the "last-mile" energy service capability as well as more general NII access to the consumer, either competitively or together with the cable and telecommunications services sectors. Should the electric utilities provide the last-mile access to support their energy services requirements, only about 5 percent of the fiber-optic network capacity would be needed to satisfy that requirement. The excess capacity could be leased or sold to other information service providers, subject, of course, to appropriate regulatory provisions. All these elements are important to the vision for the NII wherein deployment is accomplished economically and equitably in an open, competitive marketplace and with universal service assured.

### Economic Issues

Since electricity is fundamental in some way to every product and every service in the United States, it is not unexpected that the cost of electricity should have an economic impact. Consistent cuts in real electric power costs during the 1950s and 1960s were a factor in economic growth. The following decade saw a reversal in that trend, with real prices for electric power increasing. Energy information services offer the means to help reduce total energy costs. Summarizing the benefits discussed above, total energy costs can be reduced through (1) improved operating efficiency, safety, and productivity of electric utilities; (2) optimized power consumption, reduced energy costs, and improved energy efficiency for customers; and (3) deferral of capital investments in new generation, transmission, and distribution facilities, and minimization of environmental impacts. In addition, there are at least two other significant effects with positive economic implications: (1) the creation of new energy-related businesses and jobs, and (2) new growth opportunities for utilities as well as for other sectors of the economy. Finally, telecommunications facilities and services provided by the electric utility that did not exist before would be available to support the emerging information infrastructure. These benefits can be illustrated quantitatively by economic research analyses, by proposed utility projects and their projections, and by utility projects in operation.

In a plan filed by Energy, *Least-Cost Integrated Resource Plan—A Three-Year Plan*, with several public utility commissions in its service area, that utility proposed a Customer-Controlled Load Management (CCLM) project that would eventually serve over 400,000 customers. Its projections showed a cost/benefit ratio of 1 to 1.57. That is, the projected 20-year electricity benefits (or avoided energy supply costs) would be \$1,845 per household, while the cost of deploying and maintaining broadband telecommunications infrastructure would be \$1,172 for each household<sup>7</sup>. The local regulators in the city of New Orleans, included in the plan, challenged Energy on the basis of technical and regulatory uncertainties. The utility's intentions for the use of excess capacity and accounting for it were never clearly stated. Critics were concerned that the utility stockholders would exploit the windfall excess bandwidth for profitable, unregulated telecommunications. This concern must be dealt with for any future such proposals.

There are also indirect economic benefits of electric utility contributions to information infrastructure and services deployment. Several recent studies have considered and quantified the ties between telecommunications and competitiveness. A selection is highlighted here.

1. By instituting policy reforms to stimulate infrastructure investments in broadband networks, an additional \$321 billion in GNP growth would be possible over a 16-year period beginning in 1992.<sup>8</sup>
2. Jobs can be created through investments in telecommunications infrastructure. Should California increase its telecommunications infrastructure investment by \$500 million to \$2.3 billion per year over an 18-year

period, such investments would produce 161,000 jobs, \$9.9 billion in additional personal income, and \$1.2 billion in additional state and local tax revenues.<sup>9</sup>

3. Productivity is enhanced through access to telecommunications infrastructure. It has been calculated that between 1963 and 1991, telecommunications have saved the U.S. economy \$102.9 billion in labor and capital expenditures<sup>10</sup>.
4. Export activities are also supported by telecommunications infrastructure investments. One study demonstrated that between 1977 and 1982, increased competitiveness induced by improvements in telecommunications infrastructure led to an increase in U.S. exports of over \$50 billion<sup>11</sup>.

A recent study conducted by DRI/McGraw Hill under conservative assumptions and determined that the deployment of energy information services and the full use of the telecommunications infrastructure supporting those services could make a significant contribution to the U.S. economy. Specifically, it would do the following<sup>12</sup> :

- Improve U.S. productivity by reducing total real energy costs by \$78 billion between 1995 and 2010;
- Increase national employment by an average of 63,000 jobs per year;
- Produce a cumulative \$276 billion increase in GNP between 1995 and 2010;
- Achieve a cumulative increase of \$25 billion in U.S. exports because of improved business productivity;
- Reduce the federal deficit by 2010 by a cumulative \$127 billion; and
- Increase real personal income by a total of \$173 billion, or \$280 per household, through both energy cost savings and improved economic activity.

As noted above, most utilities have telecommunications infrastructure that they use for their internal business needs. Some have begun to develop and test energy information services. Few, however, have entered into providing telephone or cable TV services. An exception is the municipally owned electric utility in Glasgow, Kentucky, a community with a population of 13,000. In 1988, the Glasgow Electric Power Board installed a broadband network designed to support the standard communications needs of the utility, to provide innovative new energy information services, and to offer cable TV services. The program has been quite successful. An interesting observation made by this utility is that a customer might not be interested in having his or her water heater controlled to gain a credit of \$3 to \$4 a month on the electric bill but may well be interested in having the water heater controlled in exchange for reception of a premium channel such as HBO<sup>13</sup>. This observation offers support for the concept of having a single gateway into the home for all information services. The customer's perception of value is just as important as the real value.

This competition in cable TV in Glasgow has reduced prices. Before the broadband network installation, the local cable provider, TeleScripps, offered standard service in the area for \$18.50. After the utility offered its service at \$13.50, TeleScripps immediately dropped its rate to \$5.95 and increased the basic service from 21 to 48 channels. Despite this aggressive competition, the utility has reached a market share of about 50 percent. A more significant point, however, is that the local citizens in Glasgow have reduced their cable TV bills, and those savings have stayed in the local economy and are supporting local development<sup>13</sup>.

Statistics are not available on the energy savings results, since to this point those have been limited. Glasgow purchases all its electricity from TVA, and at the time the network was installed TVA did not offer a time-differentiated wholesale rate structure. It was not possible to pass the cost savings from load shift to customers. However, when wholesale wheeling becomes available, Glasgow is clearly positioned to use that to its advantage. It is also well positioned to provide its population access to the NII.

### Recent Strategic Alliances

In recent months there have been a number of alliances announced that are aimed at teaming players in the areas of energy services with those who provide telecommunications services and develop sensors. Retail

competition for large customers is likely in the near future, analogous to the situation wherein many large facilities have more than one major telecommunications service provider selected on the basis of the best possible price and for ensuring reliability. The time line for extending retail choice to residences and other industrial and commercial sectors is less certain except for isolated pockets that demonstrate certain advanced technology applications. In any case, it is clear that the alliances mentioned below, as well as others, are moving in anticipation of this eventuality by virtue of changes in their business and technical models.

In January 1995, AT&T announced the formation of a multicompany team "to develop and deliver a cost-effective, fully integrated, two-way customer communications system for the utility industry"<sup>14</sup>. Its development partner in this activity is Public Service Electric & Gas (PSG&E). Members of the team are American Meter Company, Anderson Consulting, General Electric, Honeywell, and Intellon. The short-term objective is to provide remote meter reading, power outage detection, real-time load management, and warning of meter tampering. The long-term objective will focus more on increasing customer control of their energy use. A notable observation about this alliance is that the utility has managed to share the risks<sup>15</sup>. Although AT&T is heading the project, the niche expertise of several other companies is key to the success of the endeavor.

In mid-April 1995, TECO Energy and IBM announced a pilot agreement "to demonstrate and evaluate an advanced 'smart' home energy management and communications system that will enable residential electric utility customers to better control and track their energy consumption, right down to their appliances"<sup>16</sup>. In addition, the system can serve as a gateway to the home for a variety of communications by serving as the interface for providers of voice, video, and data services—it can be the access point for the emerging information infrastructure. The system is configured to connect energy measuring and monitoring devices to a personal computer in the home and to a control processor attached outside the home. This processor acts as a central controller for a local area network via the existing in-house electrical wiring. With this configuration, residential customers can measure their energy use and costs 24 hours a day, and utilities can administer new, more effective energy management programs.

Also in January 1995, CableUCS was announced as a consortium of four of the nation's top cable operators—Comcast Corporation, Continental Cablevision, Cox Communications, and Tele-Communications Inc.<sup>17</sup>. CableUCS was formed to foster, build, and manage strategic relationships between cable operators and utilities—gas and water, as well as electric—and to promote the development of equipment and systems that will use the two-way telecommunications capabilities being deployed by the cable companies.

Other notable strategic alliances are (1) Energy with First Pacific Networks, (2) SoCal Edison with Cox Cable, and (3) PG&E with Microsoft and TCI. In addition to the energy services mentioned specifically above, others addressed by these alliances include energy theft detection, customer surveys, real-time pricing, power quality monitoring, and distribution system automation.

### Factors Influencing Strategic Alliances

As discussed above, the electric utilities have a choice—deploy and operate their own telecommunications infrastructure, or use infrastructure provided by other telecommunications service providers. A study by Anderson Consulting stated, "In most cases, the benefits and risks of using a third-party provider's network outweigh the benefits and risks of a utility owning and operating its own network"<sup>18</sup>. This is especially true with respect to the utilities' Supervisory Control and Data Acquisition (SCADA), which requires real time and high reliability.

Although most of the management of the electric grid is through use of the utilities' own telecommunications infrastructure, there are some experimental projects that use cable, wireless, and other services for providing last-mile access for energy services management. Possible synergism between the utilities and the cable and telecommunications providers could be exploited such that the utilities could take advantage of the last-mile infrastructure already in place to address the energy application currently deployed. However, the future will require enhanced capabilities, as noted in the Anderson Consulting study: "For the present, narrowband alternatives such as radio-based infrastructures are adequate to deliver many of the communications-enabled

services being considered. Nevertheless, trends in other industries are moving toward customer interfaces through televisions and personal computer; these interfaces will require broadband<sup>18</sup>.

To use existing cable infrastructure for two-way communication, the cable providers will be required to retrofit their infrastructure to handle high-bandwidth, two-way interactive traffic. In addition, since most of their current infrastructure is residential, they will have to partner with other providers or extend their reach to the business and industrial sectors. The cable companies reach about 62 percent of residences<sup>19</sup>. The local telecommunications providers, local-exchange carriers or regional Bell operating companies (RBOCs), have a much larger footprint with respect to residential customer base, 94 percent—one percentage point less than the electric utilities<sup>13</sup>. However, even though the telecommunications industry has made strides in the amount of information provided over conventional twisted pair lines, a more economical approach for the utilities is probably providing fiber directly to their customers. This could be funded by the sale of excess capacity as mentioned above or through energy savings earned through energy management services by the utility or the customer.

### Technologies

The current technologies for energy information services are quite diverse. They range from proprietary SCADA and energy services protocols to systems based on open protocols, such as Open Systems Interconnection (OSI) and the Transmission Control Protocol/Internet Protocol (TCP/IP).

Most residential solutions rely on the X10 protocol, implemented with simple command controls using frequency modulated power line carrier communications. CEBus, a protocol proposed by the Electrical Industry Association, and the LONWORKS protocols developed by a joint effort of the Echelon Corporation, Motorola, and Toshiba, are two new entries in the residential energy management arena<sup>20</sup>. These two solutions can be implemented over a variety of media that include twisted pair lines, infrared, fiber, coax, and radio frequency. Yet most of these solutions have significant drawbacks for addressing the requirements of a general local area network configured to support advanced information and energy appliances.

Advanced energy distribution and management systems are being tested with infrastructure deployed by utilities, through Internet services, and with infrastructure formed by combining the resources of existing utility, telecommunications, and cable providers<sup>13,21</sup>. There are also some instances of more advanced systems. PG&E is using the TCP/IP protocol suite for managing its energy distribution system. In addition, EPRI provides an energy information services network that is multiprotocol, using both OSI and TCP/IP based services such as WAIS and Mosaic.

In the future, technical requirements will require compatibility with existing as well as emerging technologies such as asynchronous transfer mode (ATM), multimedia, field data transfer, security, wireless quality of services/resource management, TCP/IP, and more. Trials are under way for enhanced monitoring and predictive maintenance of energy distribution and generation systems. This application requires a large number of addressable nodes with the capability for gathering and transmitting observation data to a centralized location for analysis. Access to such energy-related information, including accounting and billing information, will be required by users. Such access will be over both the Internet and any alternate route provided as part of the utility/user infrastructure for real-time energy demand management and control. To meet this requirement, utilities must operate a multiprotocol system, and the end user will need an inexpensive multiprotocol gateway/interface/desktop box. Real-time protocol support will be needed for such applications as time-of-day pricing or real-time pricing and both multicast protocols and the infrastructure to handle the reverse aggregation of responses in real time. Scalable network management tools are also required to handle the plethora of devices employed to support both the energy related and information related infrastructure. These same tools must also allow for the end-to-end maintenance and operation of the energy system over the various media that may be used. Finally, secure mechanisms and protocols are needed before any real energy related business will be conducted over an open packet switched network such as the Internet. Remote control of end-user appliances can be done only in a secure environment, and users will likely demand reasonable control over the other data and information that pass through their information access gateway. Technologies are beginning to emerge that can handle these requirements, and so considerable growth in this area can be forecast.



A real need exists that is not being adequately addressed. A general purpose gateway and a control model implementation should be configured for residences so that consumers can control both their energy and their information appliances from a single source—a PC or a unit attached to their television. In addition this concept should permit users to control what data and information pass beyond the walls of their premises. To effect such a solution, the utilities must define the data and information exchange paradigm, that is, the electronic data interchange (EDI) between the utility and the customer. The beginnings of this activity are the focus of part of the recent NOPR issued by FERC and discussed above.

The decision of utilities to adopt a narrow set of technical standards for the purposes of achieving interoperability or economies of scale may at first seem to be a sound choice. But over the long term this decision may put them at a disadvantage<sup>22</sup>. The NII and the Internet, as well as any energy system infrastructure, will continue to be a heterogeneous mix of media and protocols. Hence, interoperability will be possible through the use of open access and standards for gateways and interfaces, not through an end-to-end homogeneity based on a single protocol.

### RECOMMENDATIONS AND ISSUES TO BE RESOLVED

The question could be asked, Is there a compelling application for the NII that is serving as a driving force for a rapid deployment? And are the applications that could be suggested as an answer to this question contributing to the competitiveness of the United States? Why not put more emphasis on providing energy information services, since this is a ready NII application? Why not partner electric utilities with other service providers? There may not be sufficient revenues to deploy the NII without participation by the electric utilities<sup>12</sup>.

#### Technology Needs

In early 1994, the National Research Council was directed to examine the status of the High Performance Computing and Communications Initiative (HPCCI), the main vehicle for public research in information technologies. Two of the recommendations from the report of the committee that conducted that review are relevant here<sup>23</sup>.

- Recommendation 7. "Develop a research program to address the research challenges underlying our ability to build very large, reliable, high-performance, distributed information systems based on the existing HPCCI foundation" (p. 10).

The three areas identified by the committee where new emphasis is critical to supporting the research needs associated with the information infrastructure are "scalability, physical distribution and the problems it raises, and interoperative applications."

- Recommendation 8. "Ensure that research programs focusing on the National Challenges contribute to the development of information infrastructure technologies as well as to the development of new applications and paradigms" (p. 10).

It is clear from the discussion of technologies that these recommendations are relevant. For example, the development of an open nonproprietary premises gateway may prove to be the single most important advancement for both the energy services and the NII information services. It will enable new and innovative supporting hardware, software, systems, and services that will not only drive the energy supply and consumption applications arena but also feed the further development of the NII.

The electric utilities will need to deploy telecommunications infrastructure to support the energy management services, but they should also provide generic NII and Internet access. The technologies involved must be capable of interoperating with and using the services of the other service providers.

Appropriate security mechanisms are required for the delivery of adaptive energy services and for interutility as well as utility-to-consumer business activities. An issue that needs to be addressed is the need for consumer privacy. Industry, commercial concerns, and homeowners will all want to control who collects and who has access to the information collected regarding their use of electricity.

### Regulatory and Legislative Issues

Deregulation in progress or being considered in both the electric utility and the telecommunications arenas will affect how the utilities will or will not be involved in the NII. The 103rd Congress considered telecommunications reform legislation but failed to pass the proposed bills. The 104th Congress is likewise considering such reform. The question of how that should be structured should be framed with the ultimate interests of the public as the strongest driver. Is the "two-wire" model the appropriate scenario, and is it truly competitive? Although it is agreed that everyone should be able to compete, the problem lies with the details of how and when that can be accomplished.

Another factor not often discussed is the cost of deploying the NII. If the cost for deploying to one residence is about \$1,000, deployment to 100 million homes would cost \$100 billion—clearly not pocket change. The consequences of the two-wire model effectively double this cost. What markets will be there to recover such costs? Can the United States afford to spend twice as much as is needed?

There is general agreement between both political parties that the electric utilities should be allowed to enter the marketplace the same as any other "entity." As such the utilities could go head-to-head with the other service providers. But is this the best approach for all parties? The Anderson study<sup>18</sup> has a theme of dissuading the utilities from building the networks themselves—the two-way broadband communications that are projected to create benefits of \$700 million a year. Instead, Anderson suggests that the electric utilities should partner with the cable TV companies and invest in upgrading the conventional video networks. This may make sense if the utilities are to own equity in the resulting network. This would allow the utilities to finance the investment through ratable contributions since the justification is energy management applications.

The electric utilities, however, also have the opportunity to partner with the local telephone companies. These companies are struggling under the burden of regulation that prevents them from entering two competitive information services—video and long distance. Should another entity such as the electric utility build and manage the infrastructure, the telephone companies would be in a more advantageous position for competition.

Regardless of how the legislation proceeds, the electric utilities should be positioning themselves for the future. Someone who has considerable experience in dealing with telecommunications issues has this to say about what utilities should be doing today<sup>24</sup>:

- Start building "common infrastructure"—switched broadband telecommunications networks that the utility will use itself to deliver energy information services to consumers.
- Work with regulators, where applicable, to ensure that they accept such facilities in the utility's electric rate bases. This should be premised on the utility's undertaking to deliver energy information services via its infrastructure to all its electric customers in a negotiated time frame.
- Lease "excess capacity" to cable operators, telephone companies, and others on a nondiscriminatory basis—subject to regulation-sanctioned protections of incumbents to guarantee their continuity of service and recovery of investment.
- Voluntarily avoid competing with service providers to encourage amicable differentiation of market segments and to safeguard an open market in services.

### REFERENCES

- [1] "Business Opportunities and Risks for Electric Utilities in the National Information Infrastructure," EPRI TR-104539, Electric Power Research Institute, October 1994.
- [2] "The Energy Daily," Vol. 23, No. 61, p. 1, Thursday, March 30, 1995.
- [3] *Statistical Abstract of the United States—1993*, pp. 561, 563, 589, 728.

- [4] *Pocketbook of Electric Utility Industry Statistics*, Edison Electric Institute, p. 119.
- [5] "Orders Instituting Rulemaking and Investigation of the Commission's Proposed Policy Governing Restructuring of California's Electric Services Industry and Reforming Regulation," California Public Utilities Commission, R.94-04-031 and I.94-04-032, April 20, 1994.
- [6] "Positioning the Electric Utility to Build Information Infrastructure," DOE/ER-0638, November 1994.
- [7] *The PowerView System: Two-Way Communications for Electric Utility Applications*, Edison Electric Institute, 1994.
- [8] Cohen, Robert. 1992. *The Impact of Broadband Communications on the U.S. Economy and on Competitiveness*. Economic Strategy Institute, p. 1.
- [9] Cronin, Francis, et al., *Telecommunications Network Modernization and the California Economy*. DRI/McGraw Hill, May 1993.
- [10] Cronin, Francis, et al., *Pennsylvania Telecommunications Infrastructure Study*, Vol. V, pp. XIII-28, 29, DRI/McGraw Hill, May 1993.
- [11] Cronin, Francis, et al., "Telecommunications Technology, Sectoral Prices, and International Competitiveness," *Telecommunications Policy*, September/October 1992.
- [12] *The Role of Electric Companies in the Information Economy*, Mesa Research, Redondo Beach, Calif., August 1994.
- [13] *Glasgow's Fully Interactive Communications & Control System*, Glasgow Electric Plant Board, Glasgow, Ky.
- [14] "News from AT&T," press release, January 23, 1995.
- [15] "Surfing the Energy Channel," *Frontlines*, Bruce W. Radford, ed., May 15, 1995.
- [16] Press release from TECO Energy and IBM, April 1995; contact Mike Mahoney, TECO; and John Boudreaux, IBM.
- [17] Press release, January 23, 1995; contact Steve Becker, Cox Communications; Doug Durran, Continental Cablevision; Joe Waz, Comcast Corporation; and Lela Cocoros, Tele-Communications Inc.
- [18] *The Role of Broadband Communications in the Utility of the Future*, Anderson Consulting, San Francisco, Calif., 1994.
- [19] Kagan, Paul, *Marketing New Media*, December 14, 1992, p. 4.
- [20] "Supply and Demand of Electric Power and the NII," *The Information Infrastructure: Reaching Society's Goals, Report of the Information Infrastructure Task Force Committee on Applications and Technology*, NIST Special Publication 868, U.S. Government Printing Office, September 1994.
- [21] *San Diego: City of the Future, The Role of Telecommunications*, International Center for Communications, San Diego State University, March 1994.
- [22] Aiken, R.J., and J.S. Cavallini, "Standards: Too Much of a Good Thing?," *Connexions: The Interoperability Report* 8(8), and *ACM StandardView* 2(2), 1994.
- [23] Computer Science and Telecommunications Board, National Research Council, *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure*. National Academy Press, Washington, D.C., 1995.
- [24] Private communication, Steven R. Rivkin, Washington, D.C.

## 16

# Interoperation, Open Interfaces, and Protocol Architecture

*David D. Clark*

*Massachusetts Institute of Technology*

### ABSTRACT

This paper provides a framework for defining and understanding the terms *interoperation* and *open*; this framework is then used to develop operational criteria for assessing different approaches to achieving interoperation. This framework is related to the report released by the NRC entitled *Realizing the Information Future* (RTIF<sup>1</sup>). To illustrate the utility of the framework, two other interoperation proposals, one developed by the Cross-Industry Working Team (XIWT<sup>2</sup>) and one by the Computer Systems Policy Project (CSPP<sup>3</sup>), are compared. Finally, as an illustration of the framework, it is mapped onto a particular set of network protocols, the Internet protocol suite.

### INTRODUCTION

Superficially, interoperation is simply the ability of a set of computing elements to interact successfully when connected in a specified way. From the consumer's perspective, this operational test is sufficient. However, at a deeper level, it is useful to ask *why* interoperation was achieved in a particular situation, because the *why* question will shed some light on when and where this interoperation can be expected to prevail. Interoperation can be achieved in a number of ways, with different implications.

Interoperation is based on the use of common standards and agreements. One could imagine a computer application being specified in such a way that all the needed standards are described as part of the application itself. However, in most cases a more modular approach is taken, in which the application is defined to depend on certain externally defined services. These services, called *protocols* in network parlance, are often organized in layers, which help structure the dependencies that exist in achieving interoperation. Thus, in the protocols that are part of the Internet protocol suite, an application such as mail (which has its own specifications and standards) depends on the transport control protocol (TCP), which in turn depends on the network Internet protocol (IP). These protocols, organized into layers of building blocks, provide an infrastructure that makes it easier to achieve interoperation in practice. This sort of dependency in protocols is often illustrated by drawing the protocols in a stack, as in [Figure 1](#).

However, if one examines the various layers of protocols, one usually finds that there is a layer below which common standards need not be used to achieve interoperation. Again, using the Internet suite as an example, mail can be exchanged between two machines even though one is attached to an Ethernet and one to a token ring LAN.

Why are common standards needed at one layer but not at another? Certain protocols are designed with the specific purpose of bridging differences at the lower layers, so that common agreements are not required there. Instead, the layer provides the definitions that permit *translation* to occur between a range of services or technologies used below. Thus, in somewhat abstract terms, at and above such a layer common standards contribute to interoperation, while below the layer translation is used. Such a layer is called a "spanning layer" in this paper. As a practical matter, real interoperation is achieved by the definition and use of effective spanning layers. But there are many different ways that a spanning layer can be crafted.

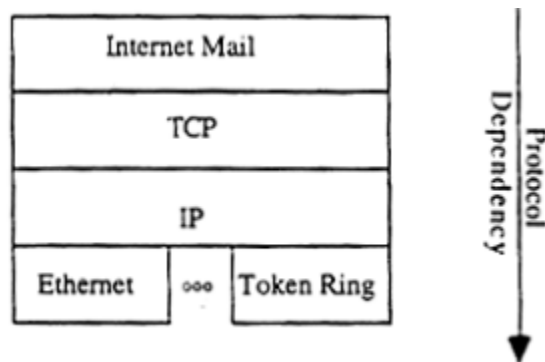


Figure 1 A simple protocol stack, organized as layers.

### THE INTERNET PROTOCOLS AS AN EXAMPLE

Expanding the Internet example above may make these points clearer. The Internet protocol suite contains a protocol that plays the role of a spanning layer, IP. Consider how IP supports the forwarding of simple text mail. The format of mail is defined by the standard RFC-822,<sup>1</sup> the required common agreement at the application layer. This protocol in turn depends for its delivery on the Internet's TCP. And finally, TCP uses the services of IP, which provides a uniform interface to whatever network technology is involved.

How does the IP spanning layer achieve its purpose? It defines a basic set of services, which were carefully designed so that they could be constructed from a wide range of underlying network technologies. Software, as a part of the Internet layer, translates what each of these lower-layer technologies offers into the common service of the Internet layer. The claim is that as long as two computers running RFC-822 mail are connected to networks over which Internet is defined to run, RFC-822 mail can interoperate.

This example illustrates the role of the spanning layer as the foundation on which interoperation sits. To determine whether two implementations of RFC-822 mail can interoperate, it is not necessary to look at the implementation of Internet over any specific network technologies. The details of how IP is implemented over Ethernet or over ATM are not relevant to the interoperation of mail. Instead, one looks at the extent, in practice, to which IP has succeeded in spanning a number of network technologies. And the practical conclusion, as reflected in the marketplace, is that IP defines a successful spanning layer. The functions and semantics of the IP layer are well defined, as shown by the fact that many companies have become successful by selling routers, the devices that implement the translation required by IP.

### A PROPOSAL FOR A SPANNING LAYER

As a part of its proposed open data network (ODN) architecture for the national information infrastructure (NII), the report *Realizing the Information Future* (RTIF<sup>1</sup>) proposes a specific spanning layer, a module it calls the ODN bearer service. This is illustrated on p. 53 of the report, in the "hourglass picture," which is reproduced here in Figure 2. It illustrates a collection of applications at the top (presumably desirous of interoperation), and at the bottom a collection of network technologies, which support the interoperation.

In the middle of this picture, at the narrow point in the hourglass, is the ODN bearer service. This layer is the key to the approach that RTIF takes to interoperation. The bearer service provides a set of capabilities sufficient to support the range of applications illustrated above it. It implements these capabilities by building on the more basic capabilities of the various network technologies below. The bearer service would thus span the broad range of network technologies illustrated below it, hide the detailed differences among these various technologies, and present a uniform service interface to the applications above. The bearer service is thus an example of a spanning layer, with specific features and capabilities.

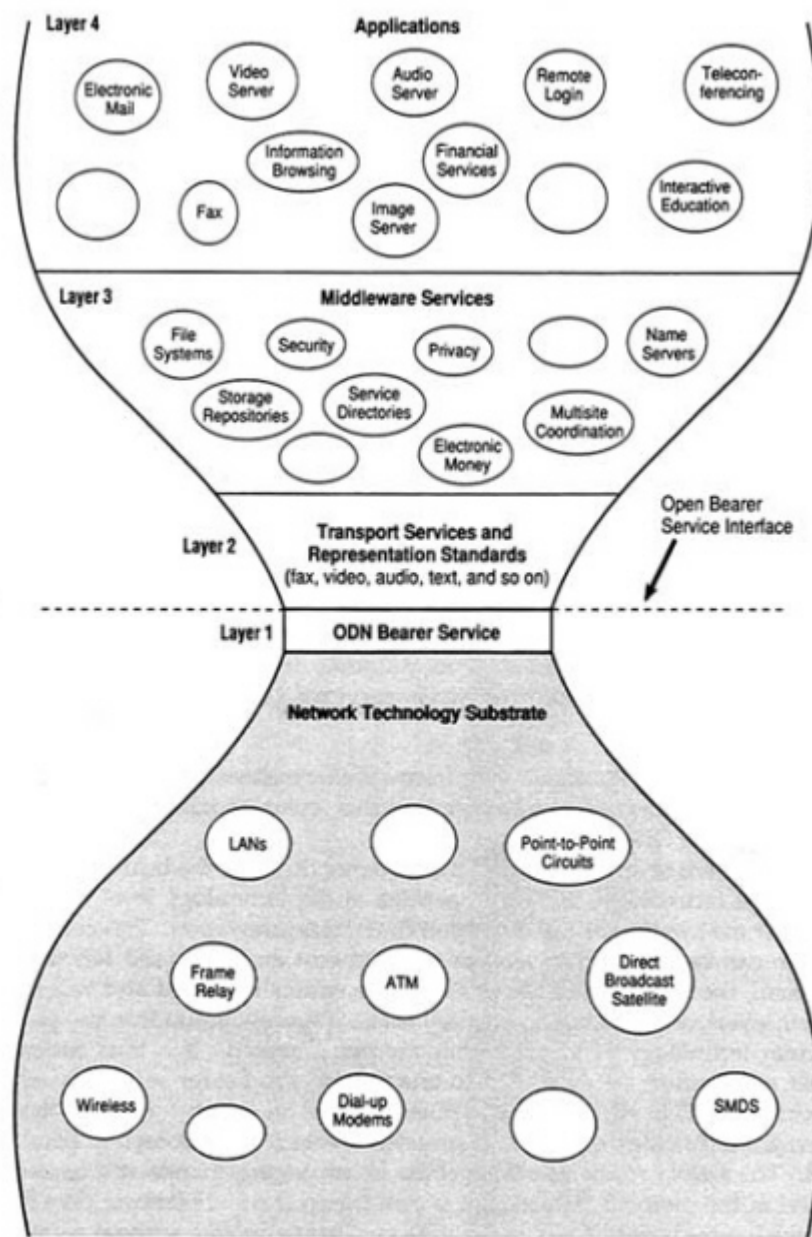


Figure 2 A four-layer model for the ODN (reprinted from<sup>1</sup>, p. 53).

The ODN bearer service is shown as the narrow point in the hourglass to illustrate that there must be a narrowing of the range of alternatives at that layer. There can be a wide range of applications supported over the bearer service and of technologies utilized under it, but the bearer service itself must represent the point at which there is a single definition of the provided capabilities. The power of the scheme is that programmers for all the applications write code that depends only on this single set of capabilities and thus are indifferent to the actual technology used below. If instead there were competing alternatives in the definition of the bearer service, the application programmers would have to cope with this variation, which in the limit is no more useful than having applications deal directly with the individual network technologies. In computer science terms, if there are  $N$  applications and  $M$  technologies, the bearer service reduces the complexity of the resulting situation from  $N \times M$  to  $N + M$ . But for this to succeed, the bearer service must be a point of agreement on a single set of capabilities.

### OTHER EXAMPLES OF A SPANNING LAYER

The bearer service is not the only way that one could achieve interoperation. There are other approaches in use today, with different objectives as to the range of spanned technologies and the range of supported applications. Some specific examples may make the alternatives clear.

- *National Television System Committee (NTSC) video delivery*: All the systems that are part of delivering NTSC—over the air broadcast, cable, VCRs, satellite, and so on—interconnect in an effective manner to deliver a video signal from various sources to the consumer. But NTSC video delivery is the only application supported. This circumstance has a different set of objectives for the range of technology that it spans below and the range of applications that it supports above. Compared to the RTIF hourglass, this would look more like a funnel sitting on its wide end.
- *ATM (asynchronous transfer mode)*: Some of the developers of ATM, which is one of the network technologies in the bottom of the RTIF hourglass, have expressed the ambition that an eventually ubiquitous ATM would come to serve for all networking needs. Were this to happen, it would not then be necessary to define a separate bearer service above it, because the service provided by ATM would become the universal basis for application development. While this vision does not seem to be working out in practice, it illustrates an approach to interoperation at a lower point than the RTIF bearer service. This, if drawn, might look like a funnel sitting on its narrow end. The virtue of this approach is that a single technology solution might be able to offer a sophisticated range of services, and thus support an even wider range of applications than, for example, the Internet approach. So the funnel, while narrow at the bottom, might be wide at the top.

All of these schemes have in common that there is some "narrow point" in their structure, with a single definition of service capability. Whether it is a definition tied to one application, or to one network technology, this narrow point in the picture corresponds to a spanning layer and is what forms the foundation of the approach to interoperation.

### COMPARING APPROACHES TO INTEROPERATION

The various examples given above suggest a fundamental way to evaluate different approaches to interoperation, which is to assess what range of technology they can span "below" and what range of applications they can support "above." The essential claim of the RTIF bearer service is that the NII needs an hourglass rather than either of the two possible funnel pictures as its key spanning layer. An approach that spans a broad range of networks is needed because there has never been, and in practice never will be, a single network technology sufficiently powerful and general to meet everyone's needs. The ATM solution, if fully embraced, would require replacing all the legacy networks, such as Ethernet, token ring, the telephone digital hierarchy, and so on. And even if this victory over heterogeneity occurred, we must expect and plan for new innovations, which will in turn replace ATM. For example, there is an ongoing research effort attempting to define a next-generation network based on wavelength division multiplexing (WDM) over fiber. This evolution will always happen; it is a necessary consequence of innovation, cost reduction, and so on.

At the same time, the NII must support a broad range of applications, not just one, because there is wide disagreement and uncertainty as to what the successful applications of the NII will finally be. One school believes that the NII will be dominated by video on demand. Others think it will be more Internet-like, with a range of interactive and transaction-oriented applications, and a wide range of service providers on a global infrastructure. Not knowing for sure what the range of applications is, it seems foolish to propose an application-specific approach to interoperation. So RTIF proposed a middle ground of interoperation, which permits both a range of applications above and a range of technologies below.

The Internet provides a concrete example of this tradeoff in span below and range of common function above. While the Internet protocol is the primary spanning layer of the suite, it is not the only one. For example, the Internet mail transfer protocol SMTP (RFC-821) is carefully written to be independent of TCP or any other

transport protocol. Thus Internet mail has "built in" an alternative spanning layer. Any reliable byte stream can be used to deliver Internet mail. And indeed, there are viable products, called mail gateway, that support mail forwarding between networks with different transport protocols.

Using the mail protocol as a spanning component provides a very wide range of spanning options. This wider range of technology and protocol options translates into a wider range of real world deployment; IP-level Internet access reaches about 86 countries today, while Internet mail reaches about 168.<sup>2</sup> The price for this is that the only application supported is mail. There is no access to remote login, file transfer, the World Wide Web, and so on. So interoperation at the mail level is, again, a funnel sitting on its wide end.

### SPANNING LAYERS—ONE OR MANY?

An obvious question at this point is whether there must be a single point of narrowing in the hourglass or, instead, there can be multiple points of common agreement. In fact, as these examples suggest, the situation that prevails in the real world is more complex than the single hourglass might imply, and one finds multiple points of narrowing at the same time. But the idea of a narrow point is always present in any general approach to interoperation. The hourglass of the RTIF should thus be thought of as both an abstract and a concrete approach to interoperation. It illustrates the key idea of a spanning layer and then makes a specific assertion as to where that point ought to be.

### WHAT IS OPENNESS?

Since successful interoperation is based on the common use of protocols and standards, a key issue is whether the necessary standards are available openly, so that different implementors can use them as a basis for implementation. Real interoperation is thus closely related to the issue of openness. There are two aspects to defining *open*. One is the question of just how public, free, and available a specification needs to be in order to be called open. The issues here are control of intellectual property rights and the control of how interfaces may be used. For a good discussion of competing views on this point, see the paper by Band<sup>4</sup> in this volume.

The other aspect of openness is its relationship to the layering that was described above. If our goal is *open interoperation*, what specifications need to be open? All, or perhaps only some? The key to sorting this out is to realize that, from the consumer perspective, what matters is the interoperation of applications—the software packages that actually perform functions directly useful to users. Since applications are what customers care about, the constructive definition of openness starts with applications. An application supports open interoperation if the definition of the application and each of the layers below it are open (by whatever definition of open prevails), down to an acceptable spanning layer. It is a simple test.

Is Internet text mail an open application? Mail is defined by RFC-822, which is an open standard.<sup>3</sup> It depends on SMTP, the Internet mail transfer protocol, which is open, and this depends on TCP, which is open; this in turn runs on IP, which is open. Since IP is a spanning layer, Internet text mail provides open interoperation.

It is usually not necessary to look below the spanning layer to see if all those network technology standards are open, exactly because that is the benefit of a spanning layer. Instead, one looks to the market to see if a spanning layer is successful. Even if some of the technologies below the spanning layer were proprietary, applications still would provide open interoperation, so long as the translation boxes that implement the spanning layer have dealt with the issues of attaching to that technology.

Even if all the definitions at the application layer are open, the application does not support open interoperation if any of the layers it depends on between it and a spanning layer are not open. An application that is openly defined, but that runs on top of a proprietary service layer that runs on TCP and IP, does not support open interoperation. This situation is uncommon but illustrates how the test for open interoperation might fail.



### SYMMETRY—A FINAL ASPECT OF INTEROPERATION

A final aspect of interoperation is the issue of whether, when two (or more) entities interact, they do so as equals or peers, or instead in some asymmetric way, for example as client and server, or as provider and consumer. Protocols are sometimes designed in such a way that not all the participants in the protocol can play all the roles. For example, some of the early remote file access protocols were designed so that the functions that implemented the file server and the file user were separately specified. This made it easy to produce implementations that omitted the server code, so that the resulting software could only play the role of the user. It was then possible to sell the server code for additional money.

In general, symmetry, like openness, tends to reflect business issues. However, some asymmetries are justified on the basis that they usefully simplify one end of an interaction. For example, some of the tasks having to do with network operation (such as resolution of faults or traffic routing) are often designed so that the network carries most of the burden and the end-node has a simplified interface. The other framework documents discussed below offer additional insights on the implications of asymmetric interfaces.

### WORKING WITH HETEROGENEITY OR PAPERING OVER DIFFERENCES?

There are two different views or philosophies about a spanning layer and its manifestation, which is a translation of some sort. One view is that heterogeneity is inevitable, and thus translation is inevitable. A properly designed spanning layer is thus a key to successful technology integration. The other view is that a proper technology design would include the goal of interoperation as a core function, and so the need for a higher level translation capability is a papering over of engineering failures at a lower level.

In different situations, both views may be true. The phone system today represents an example of an integrated approach to spanning, in which there is a wide range of underlying technologies, from copper wire to the digital hierarchy, and (in the future) ATM. However, interoperation among all of these has been a goal from the beginning and is integrated into the technology as a basic capability.

In contrast, Internet interoperation has not been engineered into most of the technologies over which it runs. The Internet has sometimes been called a "hostile overlay," because it runs above (and invisible to) the underlying technology. This sometimes leads to operational issues, since the management tools below the Internet level cannot be used to manage the Internet. However, it is the key to the growth and evolution of the Internet, since the Internet is not restricted to operating over only those technologies that were designed for that purpose.

In fact, this distinction is probably key to understanding how new services will first emerge and then mature in the NII of the future. Given that spanning layers can be defined to operate on top of other spanning layers (for example, the Internet mail spanning layer on top of the IP spanning layer), it is easy to bring a new spanning layer as well as a new application into existence by building on one of the existing interfaces. If the service proves mature, there will be a migration of the service "into" the infrastructure, so that it becomes more visible to the infrastructure providers and can be better managed and supported.

This is what is happening with Internet today. Internet, which started as an overlay on top of point-to-point telephone circuits, is now becoming of commercial interest to the providers as a supported service. A key question for the future is how the Internet can better integrate itself into the underlying technology, so that it can become better managed and operated, without losing the fundamental flexibility that has made it succeed. The central change will be in the area of network management: issues such as accounting, fault detection and recovery, usage measurement, and so on. These issues have not been emphasized in many of the discussions to this point and deserve separate consideration in their own right.

## COMPARISON TO OTHER MODELS

RTIF took a stand as to where the spanning point ought to be: a layer just above the network technology options, which the report called the ODN bearer service. Other reports have attempted to define interoperability in a somewhat different way, by cataloging *critical interfaces*.

The Computer Systems Policy Project (CSPP<sup>3</sup>) has proposed four critical interfaces; the Cross-Industry Working Team (XIWT<sup>2</sup>), seven. They are summarized in [Table 1](#), where they are grouped to best show their commonalities. These interfaces are proposed as the key to open interoperability. How do these relate to the RTIF model and the framework developed here? Although the approach to interoperability expressed in these two models may seem rather different from the definition proposed above, based on the identification of a spanning layer, these models can be mapped easily onto this framework and express a similar set of objectives.

TABLE 1 Interfaces of the CSPP and XIWT Models

CSPP	XIWT
Network-Network	Network-Network
Appliance-Network	Appliance-Network
	Resource-Network
Application-Application	Appliance-Appliance
	Resource-Resource
	Resource-Application
Appliance-Application	System Control Point-Network

### The Network-Network Interface

A good place to start with the CSPP/XIWT models is the network-network interface. This interface connects different network technologies and thus implies a spanning layer. However, it could be implemented in several ways as discussed above.

- One approach would be technology specific, such as the network-network interface (NNI) of the ATM Forum.
- The Internet or the RTIF ODN model would employ a low-level but technology-independent spanning layer.
- Another approach would be a higher-level but application-independent layer such as a client-server invocation paradigm.
- Finally, a network-network interface could be constructed at the application layer—for example, a mail gateway.

Just calling for an open network-network interface does not define which of the above is intended. The XIWT report suggests that there may be several forms of the network-network interface, while the CSPP report suggests that their concept is a fairly low-level interface.

### The Application-Application Interface

Next, the CSPP report identifies an interface called the application-application interface, which it describes as the protocols that one NII application uses to communicate with another application. The XIWT has a more detailed structure, with three versions of this interface as listed in [Table 1](#). To unravel what this interface

is, one must start with some known network-network interface, since that characterizes the spanning layer and thus the foundation of the common standards. We then apply the test for open interoperation offered above. Two applications have an open interface for interoperation if their interface specifications, and all the layers below them, are open, down to a suitable spanning layer. That spanning layer is the network-network interface.

These options are clearly illustrated in the CSPP report. In one option,<sup>4</sup> there is a messaging interface between the networks, which is a high-level (application-specific) example of a network-network interface. Another option,<sup>5</sup> instead of having a messaging interface between the networks, has a lower-level network-network interface, together with the other standards necessary to interwork the applications over that interface. The coexistence of these two forms is the exact analogy of the Internet example above, where mail can interwork either by a messaging interface (RFC-822 mail gateway) or by a full protocol stack of 822 over SMTP over TCP over IP, which would be the lower-level network-network interface.

### **The Appliance-Network Interface**

The appliance-network interface defines the way an appliance (host, set-top box, or whatever) attaches to the network. This interface, although it may not be so obvious, is another manifestation of the spanning layer. An Internet example may help illustrate this. There are low-level technology interfaces that connect appliances to networks, such as the Ethernet or the token ring standards, or ISDN. But the key to Internet interoperation is that the applications in the appliance are insulated from which technology is being used by the Internet layer. Thus, the Internet layer not only spans divergent network technologies but also "spans" the host interface to those technologies.

The above discussion suggests that the appliance-network interface could be similar to the network-network interface discussed above. In particular, the capabilities of the spanning layer must be supported across both the appliance-network and the network-network interface. To first order, in the Internet protocol these two are the same. In practice, there are differences in detail, because it is desirable to simplify the appliance-network interface as much as possible. For example, the Internet does not define appliances (hosts) as participating in the full routing protocol, which is an essential part of the Internet network-network interface, but instead provides a more simple appliance interface to routing. The ATM Forum is defining both the user-network interface (for users, or appliances) and the network-network interface, for hooking networks together. They have many similarities, but they differ in the details of what commands can be issued across the interface to set up connections, request services, and so on. The ODN model, which does not develop the network-network interface in much detail, defines them only to the extent that both must support the same end-to-end capabilities, both based on the bearer service.

### **The Appliance-Application Interface**

The CSPP identifies the appliance-application interface, which is the API for the infrastructure (e.g., the operating system on the computer). It is concerned with software portability as much as interoperability. Thus, RTIF does not emphasize this interface. But the interface makes a good illustration of the utility of a spanning layer to insulate higher levels from issues and options at the lower levels.

Applications today can be ported between a Unix, a Windows system, and a Macintosh, perhaps with some difficulty, but with some success. Even though the operating systems have different interfaces, there is *some* alignment at the abstract level among the services of the three. This is not totally true, and some applications do not port, but the point is that if an application builder accepts an abstract service definition (an interface to a spanning layer) to the operating system, then that application is much more likely to avoid getting trapped onto a concrete interface that is proprietary or has license problems. That freedom is the equivalent of what the bearer service gives us. The lower layer, whether the operating system or the network technology, can evolve. For example, if the well-known token ring patent had been too much of a problem, the community could have abandoned the specific technology, without any changes at the higher levels. In exchange, the application

builder must accept that the details of the appliance technology may be hidden to some extent beyond this abstract interface, and someone (the application builder or a third party) must write "glue code" at the bottom of the application to port the application to each operating system. This code is the equivalent of what the router vendors write in their devices to map the IP layer onto specific network technologies, and the software above the "network device driver" code, which performs the same mapping in the host.

So this interface, even though it is somewhat outside the framework proposed for interoperation, can illustrate why an abstract service definition (a spanning layer) is powerful as a point of convergence.

### Other Interfaces

In addition to the interfaces discussed above, the XIWT report proposes an interface to a service control point, which is outside the scope of this discussion. It also breaks apart the previously discussed interfaces into variants that reflect the different roles of an information resource provider and a user of the network. There are three variants of the application-application interface: the appliance-appliance interface, the resource-resource interface, and the resource-appliance interface, and there are two variants of the appliance-network interface: the appliance-network and the resource-network interfaces. These variations in the interfaces capture the idea, presented above, that the various participants in an interaction may not play as equals, but rather in some asymmetric way.

The Internet philosophy is that any user of the network is also to some extent a provider, and the distinction between user and provider is not sharp but a continuous gradation. Thus, the Internet, in its IP layer, makes no discrimination between the two. At the technology level, a provider might like to attach to the Internet in ways that are different from those of a user, perhaps with higher bandwidth, or with more redundancy, and so on. These differences are important, but the Internet protocols treat this as happening below the spanning layer.

One might take a different tack and provide a (resource) provider attachment paradigm that is different from an appliance (user) attachment in its relation to the spanning layer itself. One would have to see the technical arguments in favor of this discrimination to judge the benefit. In some cases, the purpose of this distinction has been not technical but financial—the provider, if he must attach using a different interface, may be thus identified and separately billed, presumably at a higher rate.

As noted above, the appliance-network interface and the network-network interface have a lot in common, in that they must both support the same end-to-end capabilities. In the Internet case, the network-network interface is *symmetric*, which means that the networks meet as equals. The routing protocols do not force one network to be controlled by another, nor do the management protocols. The appliance-network interface is intentionally asymmetric; for example, as discussed above, the appliance cannot participate in the routing protocols but must use a surrogate protocol.

The purpose of this asymmetry was simplicity, but one should also note that asymmetry, especially in control functions, mandates the shifting of functions to one side of the interface, which affects both the technical and the business arrangements of the network. For example, in the telephone system, there is a network-network interface of sorts between a PBX and the public telephone system. However, this interface is not symmetric and does not let the PBX attach to the public system as a full peer. This is, no doubt, not a mere oversight.

There are many other examples of interfaces that might or might not be symmetric. The network-network interface sits at a point where symmetry or lack thereof in the interface will dictate to a large extent the relative power and business position of the various providers. This is also true of application-application interfaces. A critical point of assessment for approaches to interoperation is the symmetry, or lack thereof, in the interface.

### SUMMARY

This paper proposes a framework for understanding and assessing interoperation and openness.

We offer a careful definition of interoperation, beyond the operational test of "does it work?" Two implementations of an application can interoperate if (1) they are based on common definitions and standards at

the application layer, (2) they use supporting services in a consistent manner, (3) they are based on a common spanning layer as a foundation for these services, and (4) the range of underlying network technologies to which the applications are actually attached is within the range that the spanning layer can reach.

There are two important aspects to this definition. First, it is stated in terms of interoperating applications. If the question of interoperation has been framed in terms of a different entity, such as a computer, there is no well-formed answer. Computers may interoperate, or not, depending on what application is running. Second, the spanning layer is defined as the "foundation" on which the definition rests. This word was chosen because one does not have to look below it when testing for interoperation. If the spanning layer is well defined and is known to work over the network technology in question, then the fact that the application is defined in terms of this spanning layer is sufficient.

A spanning layer is characterized by three key parameters, which define the sort of interoperation it supports:

- The span of infrastructure options over which the interoperation occurs;
- The range of higher-level applications the spanning layer supports; and
- The degree of symmetry in the services of the spanning layer and its interfaces, and in the higher layers up to the application definitions.

Different approaches to interoperation can be evaluated by assessing these characteristics; this paper attempts to map the RTIF model and the XIWT and CSPP interfaces in this way. In general, there is no conflict between these models, but the RTIF framework<sup>1</sup> is more specific on certain points, in particular as to the exact modularity of its bearer service, which provides the basis in the RTIF framework of the network-network, the appliance-network, and (if it separately exists) the resource-network interface. The XIWT<sup>2</sup> and CSPP<sup>3</sup> reports are more specific, on the other hand, in calling for a distinct and full definition of the network-network interface, while RTIF discusses only those aspects that relate to the end-to-end capabilities of the bearer service.

Using this framework, RTIF observes that openness is a characteristic of interoperation. That is, if all the protocols that accomplish a particular sort of interoperation are open, from the application down to the relevant spanning layer, then one is justified in calling that form of interoperation open. Again, the definition is in terms of an application, which, from the user's perspective, is the relevant objective of an interoperation test.

RTIF also uses the Internet architecture as an example of a concrete set of protocols that embody an RTIF-style bearer service approach to interoperation: Internet defines a largely symmetric bearer service, the Internet protocol itself, which has been shown in the marketplace to span a wide range of network technologies and support a wide range of applications. The network-network interface in Internet is defined by IP itself, together with the routing, management, and other control protocols. It is completely symmetric. The appliance-network interface is defined by IP, together with ICMP and the other host-related control protocols. There is no separate resource-network interface.

### ACKNOWLEDGMENT

A number of people have read and commented on this paper. Neil Ransom, from the XIWT, and Tom Gannon, from the CSPP, provided valuable feedback on my comparisons. Marjory Blumenthal and Louise Arnheim, both of the Computer Science and Telecommunications Board of the NRC, offered extensive comments that greatly improved the paper.

### REFERENCES

- [1] Computer Science and Telecommunications Board (CSTB), National Research Council, *Realizing the Information Future: The Internet and Beyond*, National Academy Press, Washington, D.C., 1994.
- [2] Cross-Industry Working Team (XIWT), Corporation for National Research Initiatives, *An Architectural Framework for the National Information Infrastructure*, CNRI, Reston, Va., 1994.

- [3] Computer Systems Policy Project (CSPP), *Perspectives on the National Information Infrastructure: Ensuring Interoperability*, Computer Systems Policy Project, Washington, D.C., 1994.
- [4] Band, Jonathan, "Competing Definitions of 'Openness' on the NII," in this volume.
- [5] Object Management Group (OMG), *Object Management Architecture Guide*, Object Management Group, Framingham, Mass., 1990.

## APPENDIX I MORE DETAIL ON THE BEARER SERVICE: WHAT IS IT EXACTLY? WHERE IS IT?

The general argument in favor of an hourglass is not enough to support the specific details of the RTIF bearer service approach. Other spanning layers could be defined that, while still "in the middle," are at a higher level in the protocol stack.

One approach would be to standardize a protocol by which a user program can call a server across the network, such as the client-server invocation protocol proposed by the Object Management Group (OMG<sup>5</sup>). Interoperation through a standard client-server invocation paradigm would in fact permit a wide span below, because it could run not only over a range of network technologies but also over a range of protocol suites. It could, for example, operate over TCP/IP or vendor specific protocols such as Novell's IPX or IBM's System Network Architecture (SNA).

Another approach would be to standardize an interface directly above the transport layer in the protocol stack, corresponding to TCP in the case of the Internet suite. IBM has recently proposed such a thing, a spanning layer that spans TCP and SNA. Such a layer would permit an application from either the Internet suite or SNA to run over any combination of the two protocols.<sup>6</sup> Spanning both TCP and SNA covers a wide range of underlying technologies and might prove powerful.

What is the limitation of these schemes? First, these potential spanning layers are separated from the actual network technology by more protocol modules (more layers of software) than the ODN bearer service is. These intervening modules make it less likely that a higher-level spanning layer can interact with the network technology to change the quality of service delivered. Various applications need different performance from the underlying network. Some services, like real-time video, require delivery within a predictable delay. Others, like e-mail, may require only the most minimal of best-effort delivery. (This variation in offered performance is called quality of service, or QOS, in the network community.) Most existing protocol suites do not provide a useful way, from high in the protocol "stack," to reach down across the layers and control the QOS. This argues for an interoperation layer at a low point in the protocol stack, close to the network technology itself.

For example, in the case of the SNA/TCP spanning layer, the capabilities of the service are constrained by the particular protocols out of which the service was built, TCP or SNA, which offer a reliable delivery protocol with no ability to bound delivery delays. So the SNA/TCP spanning service, although it has a broad span over network technology, cannot make use of other QOS in the lower layer and thus cannot support applications such as real-time audio and video.

Second, one of the features of a spanning layer must (almost always) be an address space or numbering plan.<sup>7</sup> One cannot interoperate with something one cannot name or address. So any protocol layer designed to span a range of lower-level technologies must implement an address space at this layer. Experience suggests that providing a single global address space at a low level is most effective.

There are a number of examples of different address or name spaces in use today. The Internet mail protocol was designed as a spanning layer and thus has its own address space, the mailbox names. The syntax for mailbox names can get quite complex and can express a range of functions such as source routing. However, since the address space is not as well defined or engineered as the IP address space and is not as well supported by management tools, problems with mailbox names and mail forwarding when using mail as a spanning layer are well known.

Another name space of interest is the naming of information objects in the World Wide Web, the so-called universal resource locators, or URLs. The WWW transfer protocol, HTTP, is currently defined to run on top of TCP. However, it provides its own name space and thus might be able to function as an application-specific spanning layer. This extension would require some reengineering of the URL name space, an area of current research.

In contrast to IP or the Internet mail protocol, if one were to ask if TCP (the transport protocol that sits above IP) could be a spanning layer, the answer is no. It is defined to depend on IP and cannot be run on top of another internetwork layer protocol. TCP is part of the Internet interoperation story (it defines a popular common service), but it is not a spanning layer.

## NOTES

1. The Internet standards are specified in a series of documents called RFCs, or Requests for Comments, which are available on the Internet. Contact [rfc-info@isi.edu](mailto:rfc-info@isi.edu).
2. These numbers are from the international connectivity data collected by Larry Landweber and provided by the Internet Society as a part of its Web information collection. See <http://info.isoc.org/home.html>.
3. All Internet standards are published for use in any way without any constraints or license. This perhaps represents the most open form of "open."
4. CSSP [3], p. 16, example 3A.
5. CSSP [3], p. 17, example 3E.
6. Note that this proposal does not support the interworking of an SNA application with an Internet application. By our definition of interworking, that would require a spanning layer specific to each pair of applications to be interconnected. The goal here is more modest: to take existing applications from either protocol suite and support them over a new spanning layer that allows them to sit above both TCP and SNA.
7. There are examples where a spanning layer tries to avoid a common address space by supporting multiple address spaces or by translating among several. These approaches are often complex.

## 17

# Service Provider Interoperability and the National Information Infrastructure

*Tim Clifford*

*DynCorp Advanced Technology Services*

### ABSTRACT

The vision of a national information infrastructure (NII) introduces many scenarios that imply the use and assumption of flexible, interoperable telecommunications services. The existing telecommunications infrastructure offers interoperability between local exchange carriers (LECs) and interexchange carriers (IXCs) for traditional circuit switched and dedicated services and among peers at agreed exchange points in the case of the Internet. The anticipated evolution of commercial telecommunications services does not guarantee that competing service providers will interoperate.

Current trends compromise the basis for interoperability among the providers of the telecommunications infrastructure. The merging of LECs and IXCs toward joint competition suggests a model similar to today's IXC environment, with its extremely limited interoperability. The commercialization of the Internet offers a second example where the Internet policy of free exchange of data tends to become compromised by the growing call for settlements between carriers. Early experience with advanced services such as ATM and frame relay, likely key enablers of the NII, also suggests a reluctance on the part of competing carriers to interoperate. For example, the long-term availability of X.25 public data networks has not resulted in interoperability among competing service providers.

The recommendations suggest several possible approaches to ensure the continuation and expansion of interoperability among telecommunications service providers. A first approach examines past models of interoperability, (e.g., for voice service or the Internet) and examines the suitability of these models to an NII. As technology-related issues will come into play, prototyping activities could be established to define interfaces, address performance, and consider management between service providers. The Network Access Points (NAPs) established by the National Science Foundation may also be considered as a government-sponsored vehicle for interoperability. Finally, it is advisable to continue the analysis of policy and regulatory reform currently under way to ensure continued interoperability among service providers in the NII.

### STATEMENT OF THE PROBLEM

The vision of a national information infrastructure (NII) introduces many scenarios that imply the use of a flexible, interoperable telecommunications infrastructure. The existing telecommunications infrastructure offers interoperability between LECs and IXCs for traditional circuit switched and dedicated services and among peers at agreed-upon exchange points in the case of the Internet. The evolution of commercial telecommunications services does not guarantee that competing service providers will interoperate. In fact current trends compromise interoperability among future providers of telecommunications services, thereby creating the potential for many islands where exchange of information beyond island boundaries becomes difficult, less responsive, and uneconomical.



## BACKGROUND

The following sections offer a brief background of the broad and diversified telecommunications market. The first section summarizes the current service providers (with a focus on local and interexchange carriers) and uses the Internet as an example of telecommunications services offered outside of the traditional carrier envelope. The second section describes the generic services currently available and focuses on the interoperability relationships in today's environment necessary for user-to-user communication. Interoperability today typically occurs at the lowest common denominators of service; the third section therefore provides an example of future services available to the NII. Interoperability among service providers in the realm of advanced services appears to be a significant challenge.

### Telecommunications Service Providers

The telecommunications infrastructure is composed of an extensive set of physical communications facilities owned by a relatively small set of telecommunications companies. The LECs and the IXC provide most of the bulk of the actual physical facilities that deliver telecommunications services. However, a broader definition of infrastructure is offered here that includes a large set of service providers that leverage the physical infrastructure to deliver service. This latter group includes the carriers themselves but also involves many organizations that build services based on specific markets or communities of interest. The Internet is the most prominent example of this latter group, where several thousand networks make up the Internet and only a small subset of the LECs and IXCs actually provide Internet services. In addition, new elements of infrastructure have emerged in the areas of cable infrastructure and wireless media. Although many of the cable and wireless providers have become tied to traditional carriers, their presence implies a strong potential for telecommunications evolution. With the exception of the Internet, the existing infrastructure offers limited interoperability, principally between a federally mandated tier of service providers and at the lowest common levels of information transport.

Local service implies that a service provider is principally restricted to offering connectivity within local access transport areas (LATAs). The local service area continues to be dominated by the regional Bell operating companies (RBOCs). However, many local service providers include traditional service providers and an emerging class of competitive access providers (CAPs). The traditional service providers offer RBOC-like services in independent areas of the country and include a broad spectrum, from large providers such as GTE and Sprint Local (both over 100 years old) to small providers such as Splitrock Telephone Co. of South Dakota. Traditional services offered by the RBOCs and other local providers have increasingly featured voice services, private lines, and a gradual influx of local data services. CAPs, such as Metropolitan Fiber Systems and Teleport Communication Group, have focused on the provision of dedicated access between high-density customer locations (e.g., government and business offices) and the IXCs. The CAP focus can be attributed to the strong business opportunity for local access and the large investment required to reach private residences. CAPs have also gradually increased their portfolio of services offered to include data services. The most prominent upcoming change in the local service area is the likely removal of restrictions on the RBOCs for the delivery of services beyond the local area.

The IXCs, or long-distance carriers, provide telecommunications services between the LATAs. The divestiture of the RBOCs by AT&T in 1984 created the opportunity for new long-distance service providers and has resulted in the creation of several prominent competitors of AT&T, principally MCI and Sprint, and a strong cadre of growing service providers, such as Willtel and LCI. The resulting competition has brought rapid technology evolution, new services, and lower costs to long-distance services. The long-distance infrastructure has also become highly robust and diverse, currently including multiple high-capacity networks with independent control structures. For example, a typical span in a major carrier network may consist of several active fiber pairs, each carrying in excess of 1 gigabit per second (Gbps) of capacity. The IXCs make up multiple cross-country (east to west and north to south) routes that, taken together, represent hundreds of Gbps of capacity. The IXCs support many telecommunications services, dominated by voice (especially for the top three providers) and private line. In addition, the IXCs offer various public services for data or video-based services.

During this period of competition in the long-distance market, the local carriers have played a key role as the common exchange point among the services provided by the IXC's. However, in the local area, where the traditional telecommunications monopoly has remained in effect, technology, services, and prices have tended to lag behind the long-distance market, often significantly. Although the local carriers enabled the emergence of multiple long-distance carriers through a common and expensive local infrastructure, the capability gaps and higher prices in the local area have become considerable. The IXC's typically claim that approximately 40 percent of their total costs result from local carrier charges. The local carriers have moved to close these gaps as the promise of unrestricted competition has become feasible. The IXC's have also begun to posture for the entry of local providers through arrangements to build local infrastructures that will lessen the local monopoly of service.

The Internet is one example of telecommunications services that offer a value-added capability on top of the carrier infrastructure. It has become prominent over the last several years, principally due to an unabated growth rate. As a simple definition, the Internet is a collection of networks bound together using a common technology. The Internet originated with Department of Defense research on networking and sharing of computing resources. Primarily through government subsidies for research and education support, the Internet grew significantly through the 1980s and began to attract commercial attention due to its growth and strategic government and private investment. Currently the Internet has become a major source of commercial enterprise and continues to grow rapidly. An explosion of commercial Internet service providers has paralleled the growth of the Internet and offers wide area connectivity among the expanding population of Internet users and component networks. The Internet has also become a prime model for many visions of the NII, because of its ability to interconnect various technologies, people, and sources of information.

Overall, the telecommunications infrastructure, taken as an entity without regard for regulatory or business concerns, offers a multifaceted and ubiquitous platform for the delivery of services consistent with the NII. Current market and regulatory factors have placed restrictions on the delivery of services by local providers. As a result, local services tend to be less capable and more expensive than those provided by IXC's. This status quo has begun to evolve rapidly as local providers prepare for a relaxation of their restrictions. The IXC's continue to enhance their abilities to deliver many services at ever-lower prices. Numerous service providers, best represented by the Internet community, have capitalized on this infrastructure to provide effective information services.

### Telecommunications Services

It is also worthwhile to briefly review the services delivered by the service providers. The telecommunications infrastructure supports three types of connection that enable a wide variety of services:

- Circuit switched services that allow dynamic establishment of a connection between users and support voice, video, and modem-based traffic;
- Private line services where a carrier establishes a full period connection between users at a predefined capacity; and
- Statistically multiplexed services that offer the ability to carry traffic dynamically as required by the user. Public data networks, the Internet, and emerging frame relay and ATM services employ statistical multiplexing.

Especially important to this paper are the relationships between the providers of these services that enable interoperability from user to user. Although transparent to a user, virtually all telecommunications services in fact require an exchange of information between multiple service providers. Today a subtle but complex web of relationships addresses the information transfer, operational support, and business mechanisms necessary to make telecommunications a reality. Interoperability among carriers has typically settled to the least common denominators of service such as switched voice or private line services. The demand of the NII, for more robust, featured services, poses a more difficult and a new challenge of service interoperability to the growing cadre of service providers.

Telephone-based voice services are by far the dominant mode of telecommunications. This dominance can be expressed in many ways, including percentage of infrastructure consumed, ubiquity of service, and generation of revenue. Voice services have evolved over the past 100 years to where significant complexities in the delivery of service have become transparent to the typical user. The voice service infrastructure also supports a large proportion of all data-oriented services through its support of modem-based communications. In general, voice services involve two handoffs between carriers: A local carrier passes a call to the IXC, and the IXC passes the call to the destination local carrier.

Telecommunications carriers also provide private line services, which allow users to connect separate sites with basic transport services. A private line represents an extension of the carrier infrastructure to an individual user; the carriers devote a dedicated time slot of their network to that customer. A user may attach a large variety of customer premises equipment (CPE) to the private line to support virtually any type of service. Large users of telecommunications use private line services as the cost of dedicated services tends to be high relative to that for switched services such as voice (note that while customers only pay for voice services when they are in use, private lines incur flat rate charges independent of usage). Paradoxically to the higher relative costs, private lines represent the least possible value added by the service provider. However, private line services provide the underpinning of virtually all services. For example, the Internet has long been based on the connection of devices that implement the Internet protocols (i.e., IP routers) with private lines.

In fact, service providers distinguish their services from customer use of private lines by the movement of value-added equipment from the customer premises to the service provider premise. For example, in the case of Internet services, many large enterprises have deployed Internet technology within the enterprise as IP routers interconnected by dedicated services. In effect, these enterprise networks are analogous to a private internet (i.e., many enterprise networks interconnected). Conversely, an Internet service provider deploys Internet technology within its infrastructure and offers service through dedicated connections from individual user locations to the backbone network. Similarly, enterprise networks based on local intelligent equipment such as smart multiplexers, X.25 equipment, or frame-relay technology continue to proliferate.

Public data services originated with the same government development activities as the Internet but moved directly to the support of commercial services. The emergence of public data networks based on the X.25 protocol suite through the 1980s created a global infrastructure for data transport. However, unlike the Internet, X.25 networks tended to focus on services for individual users over a virtual private network, rather than between independent networks. These virtual private networks (a concept later extended to voice services) allowed many users to share the backbone of a single service provider. These X.25 networks have recently found new life as the access vehicle of choice for consumer information services, such as America Online, owing to their deployed dial-up facilities. X.25 networks also offer a working model for carrier network interconnection in the form of the X.75 gateway, the technical designation for interconnection between X.25 networks provided by different public-service providers. However, while X.75 offers the technical mechanism for interoperability, competing service providers have been reluctant to deploy X.75 connections. Public data services have also reemerged in the form of frame-relay and ATM services.

Similar to the emergence of public services for data communications, carriers have initiated the delivery of public videoconferencing services over the past several years. A video-based service takes the form of providing telecommunications service between public or private video rooms. The video rooms, essentially conference rooms modified to support video teleconferences, are interconnected via carrier facilities. The carriers have offered reservation and gateway services to allow users to schedule facilities ahead of time and ensure that the video systems at both the source and destination ends will interoperate. The ability of a service to offer interoperability between disparate types of video equipment has become an important feature of the overall service. In a limited fashion, the video teleconference capabilities of the carriers support interoperability, although primarily in the context of specific user implementations rather than generic service features.

The NII can be expected to increasingly take advantage of a wide range of service capabilities. Today's common model for the NII, the Internet, tends to focus on data exchange modalities. Conventional telecommunications services offer limited models for interoperability among competitive carriers or for advanced services. We can expect that the NII will increasingly embrace multimedia services that will require a robust

service infrastructure and place demands on service provider interoperability that exceed the scope of existing interoperability models.

### Emerging Infrastructure and Services

The strong growth and competition in the telecommunications industry have accelerated the introduction of advanced technologies to the existing infrastructure. The result is a dynamic set of services that offer greatly enhanced services and new capabilities to the NII. These capabilities will begin to change the manner of delivery of information services, likely with significant improvement in price and functionality. However, the early definition of these services has tended to follow existing models for carrier interoperability (i.e., via private line or eventually local carrier interconnection to the IXC). Rather than attempt a comprehensive review of emerging technologies, the following section discusses the potential of asynchronous transfer mode (ATM), a good model for the introduction of new telecommunications services enabled through advanced technology.

### Advanced Technologies

Since its definition in the mid-1980s, ATM has represented something of a holy grail to the telecommunications community. ATM, as defined, supports all communications services and thereby represents a merging of the telecommunications and computing portions of the information technology industry. The formation of the ATM Forum in 1991, now including over 700 members from all segments of industry and government, sparked the development of ATM specifications to support early ATM implementation. The emergence of early ATM capabilities in late 1992, along with ATM's steady growth over the past 30 months, has created considerable excitement as the harbinger of a new era of networking consistent with the NII. Over this time ATM has taken a preeminent position on the leading edge of the telecommunications industry. ATM simultaneously offers service providers more effective operation and users greater flexibility. Service providers, from large public carriers to the telecommunications groups of private organizations, have embraced the ATM concept as a means to streamline internal operations and improve the delivery of user services. For example, ATM will enable carriers to significantly reduce overall provisioned bandwidth and allow instantaneous provisioning of service through the statistical multiplexing capability of an ATM backbone. Although fiber-optic networks have greatly reduced the overall cost of network capacity, carrier investments in network facilities still represent billions of dollars in investment and operating expense. The ability of an ATM approach to reduce costs and improve operations constitutes a significant market advantage. In addition to enhanced internal factors, ATM will enable service providers to deliver advanced services at a fraction of current prices and create new markets for services.

The vision of an ATM-based network also offers significant advantages to the user community. Existing telecommunications services mirror the carrier networks, where multiple independent networks share a common time division substrate to deliver separate services. For example, the voice, Internet, and video services mentioned here exist as separate networks within the carrier facilities, requiring separate capacity, operational support, and management. Users tend to have a relatively small set of options for service, which are rigidly defined. ATM offers a dramatically different view. A customer may send any type of information—voice, video, data, image, or multiple types at once—at transfer rates from 1 to 1 billion bits per second. ATM also promises independent quality of service guarantees for each application, thereby allowing users to select services within their own context of budget, urgency, and quality, all on a dynamic basis. Despite the great promise of ATM, true interoperability of ATM services offered by competing carriers—that is, addressing technological, operational, and business concerns—appears several years away.

## Emerging Services

The evolution of information technology has significantly enhanced the capabilities and range of information services. Numerous reports and articles have already described emerging information services, but three trends become important to future service provider interoperability:

- Information services as an entity are growing. Providers such as Prodigy, Compuserve, and America Online are showing phenomenal growth. The Internet also continues to grow rapidly, especially the phenomenon of the World Wide Web. The access and availability of critical information can become a differentiating feature of a service provider.
- Many telecommunications carriers have described intentions to provide information content. Though the concept of carriers providing information has been slow to develop, and somewhat dominated by applications such as video movies on demand, it seems likely that more useful and broader cases will arise. Several carriers have developed prototype capabilities for the support of applications such as collaborative research, distance learning, and telemedicine.
- Telecommunications service providers will try to move away from competition founded in the lowest price for a commodity service and differentiate their transport through information content.

The continued evolution of information services and their combination with telecommunications will strongly influence the NII. The influence will be seen in the changing nature of competition between telecommunications carriers, information providers, and new hybrids of the two. We can expect that commercialization of information services will affect interoperability as a market driven by information content may evolve.

## ANALYSIS AND FORECAST

Interoperability among telecommunications carriers is a key attribute of the infrastructure. For example, voice or switched services and private line and Internet services all support interconnection between users connected to different carrier services. Meanwhile, private networks and emerging services tend to struggle for interoperability beyond specific provider boundaries. The current models for interoperability among service providers offer a reasonable baseline today. However, key trends such as regulatory changes, market forces, and technology development are emerging that affect the continued assumption of interoperability.

### Current Models for Interconnection

There are three significant models for service provider interoperability that may come into play within the future NII. Since divestiture, local carriers have set up agreements with IXCs for the exchange of services. These agreements have focused on two types of service: switched service supporting voice; and private lines that transport many types of information, with little value added from the carriers aside from provision of the connection. Although these arrangements are technically sophisticated, the most important feature involves the business mechanisms that track the exchange of information and convert the volume of information to financial settlements. Meanwhile, the data community has set up interoperability models consistent with data exchange. The X.75 protocol was developed to address the exchange of information between national public data networks that employed the X.25 protocol. The Internet is thriving and is synonymous with the concept of interconnected networks. It presents a powerful model for the NII, often used interchangeably with future NII concepts. However, today's Internet does not support explicit business mechanisms for financial settlement but has begun to address settlements between service providers and is likely to evolve to some appropriate model. The existing models for interconnection offer limited guidelines for future service provider interoperability.

### **LEC-IXC Interconnect**

The most important instance of service provider interconnect, in terms of the volume of traffic handled, supports the connection of switched traffic, such as voice, and private lines between LECs and the IXCs. This relationship emerged with the divestiture of AT&T in 1984, the establishment of many local carriers, and the gradual introduction of competition in the IXC area. The interconnects are mandated by the agreements surrounding the divestiture providing for equal access to local facilities to any IXCs prepared to provide service.

The LEC-to-IXC interconnect employs sophisticated technology oriented to the establishment of connections on a dynamic basis between users that often span three service providers. Both LECs and IXCs establish multiple connections between their networks and other carriers—for example, Bell Atlantic supports interconnects to Sprint, AT&T, and MCI—and the interconnect discipline has thus been replicated many times. Each of the carriers has invested heavily in improving the signaling that controls these interconnects. Improved signaling both increases performance (e.g., in the sense of call set-up time) and enhances feature availability. In addition to the technology inherent in these interconnects, the establishment of business mechanisms that measure traffic exchange and enable financial settlements represents a major contribution to the interconnect model. The volume of both the traffic (tens of billions of call minutes per month) and the financial exchange (billions of dollars per year) makes the current LEC-IXC a powerful model for interconnect.

The LEC-IXC model does not necessarily represent the best possible approach. The interconnect is mandated through regulatory means rather than through the dictates of an open market. The absence of any substantial interconnect scheme between competing service providers suggests that the current scheme may not extend well to competitive-service providers. For example, interconnection between Sprint, MCI, and AT&T is mostly nonexistent; it is unclear whether typical LEC to IXC agreements would translate to two IXCs.

The technologies associated with this interconnect have not been shown to adapt well to other types of services. The IXC-LEC emphasis on the lowest common denominators of telecommunications, switched voice services and private lines, does not extend well to an NII vision that takes advantage of various advanced telecommunications services. Recent efforts have begun to address the interconnection of advanced services such as ATM and frame relay between LEC and IXC. However, operational interconnects have only reached planning stages, with issues like service guarantees and appropriate financial settlements still poorly understood.

### **Public Data Network Interconnection: X.75**

A second form of carrier interconnect in place since the early 1980s and widely deployed is the gateway between public data (X.25-based) networks, referred to as X.75. This technology applies to gateways that may exist between public X.25 data networks, often those operated by national telecommunications carriers (i.e., France Telecom; in the United States, the primary X.25 carriers are MCI and Sprint). The X.75 interconnect offers an historical model for large data service provider connection in a commercial environment. Though limited in functionality and becoming obsolete, the X.75/X.25 model is important because it supports the data environment, involves financial settlements between carriers (including international entities), and bears some similarity to the emerging ATM environment (i.e., it supports connection-oriented service).

X.75 operates by setting up virtual connections from user to user as a spliced connection of multiple virtual connections. In this sense X.75 represents the opposite of the Internet IP protocol, which uses a connectionless model for service delivery. X.75 also thereby tracks current efforts toward the development of ATM services, which is also connection oriented and a potential model for future services. The X.75 gateways also include mechanisms for the tracking of data exchange and settlement procedures between networks. As an interconnect between public networks, the X.75 model has the added advantage of defining interconnects between international entities. Many countries still operate a single, monopolized telecommunications carrier and therefore a single public data network. Notably, the evolution of public data networks has tended to downplay the interoperability offered through X.75 connections. The expansion of global services by primary carriers has created direct competition between service providers and resulted in reduced desire for interoperability.

## Internet

The Internet offers an interesting model for interoperability. From its inception, virtually anyone could connect to the Internet, given compliance with the set of procedures and formats dictated by an evolving set of Internet standards. By the late 1980s an architecture emerged that included a backbone transit network, the NSFNET (which supported connectivity between regional service providers) and an acceptable use policy (which restricted traffic to research and education). The NSFNET exchanged traffic with other major government backbone networks, through Federal Internet Exchanges (FIXs) on the East (FIX-east) and West (FIX-west) coasts. The FIXs provide a technical and policy-based interchange point, but they do not include mechanisms typically found in business environments for collection of traffic data and billing.

The growth of the Internet has precipitated commercialization. At first, small service providers and then gradually larger service providers (including Sprint and then MCI) developed services that employ Internet technologies and offer connectivity to the Internet (a combination usually referred to as Internet service). The commercial service providers created a Commercial Internet Exchange, the CIX, to allow the exchange of commercial, policy-free traffic. The shift to commercialization, with the government as a major catalyst, is prompting the gradual withdrawal of government influence. Through the NSF's solicitation 93-52, the NSF has decommissioned the NSFNET as the transit backbone for the domestic Internet and promulgated a commercial infrastructure that includes three primary Network Access Points (similar to FIXs) and commercial network service providers (NSPs). Both the NAPs and the CIX are provided by third-party vendors (the facilities for each are actually provided by carriers—Sprint, Ameritech, PacBell, and MFS) and support the exchange of traffic.

At this time there are limited barriers or policy for connection at a NAP or a CIX. The NSF has set a policy that any entity funded by the NSF must employ an NSP that is connected to all three primary NAPs. In this way the NSF has ensured that the research and education of the Internet will be fully connected, at least for the 2 to 3 years planned for the NSF funding. Within each access point members must establish separate bilateral agreements to facilitate interconnection. In some cases, service providers have included place holders for settlement agreements in these bilaterals; however, no agreements have been set at this time.

## Attributes of Service Provider Interconnection

The preceding profiles of major interconnection methods suggest several key attributes for service provider interconnection.

- *Technology.* The interconnection method must employ technologies capable of supporting the types and volume of information exchanged. The technologies must provide scalability to ensure that future growth can be supported.
- *Flexibility.* The interconnection should support flexibility—for example, through support of multiple forms of information, including voice, video, and data.
- *Replicable.* The interconnection should enable replication and preferably employ agreed-upon standards that require consensus among a broad part of the service provider community.
- *Stability.* The introduction of new members to an interconnection point should not affect the existing service providers.
- *Operational support.* The interconnection point should be supported by solid operational agreements between connecting members to address failures in the connection.
- *Management.* The interconnect point should enable oversight of activities at the interconnect point to include measurement of traffic exchange, service quality, and security.
- *Financial.* The interconnect point should define methods for settlement of traffic between service providers.

### Key Trends Affecting Carrier Interoperability

Today's telecommunications environment exists through a web of interoperability agreements among numerous service providers. In general, these agreements exist to facilitate the fundamental exchange of information. In virtually all cases the agreements exist between noncompetitive entities where cooperation benefits both parties. However, many commercial and regulatory trends have begun that threaten this basis for service provider interconnection. The expected relaxation of constraints on LECs for the provision of long-distance services will create a new competitive relationship among LECs and IXC. Both groups have rushed to form strategic relationships and extend capabilities to prepare for this new era. In addition, as telecommunications services have become more of a commodity, service providers have looked for ways to differentiate their services through technological improvements as well as in combination with information services. The Internet is undergoing a parallel change where the trend toward commercialization will create competitive relationships among previously complementary entities.

Current trends in the telecommunications regulatory structure suggest strongly that local carriers will begin to extend their role beyond local transport services to include wide area connectivity. The LECs have already won the right to provide long-distance services for cellular-based customers. Local carriers will also continue their expansion into other nontraditional service avenues, especially in the area of information service. Although the pace of this evolution of the LEC role is unclear, it is likely that the local carriers will tend to dominate the local infrastructure for many years to come. Despite major IXC investments, the extensive, in-place LEC connectivity to every residence will be extremely difficult to duplicate. Early in this evolution, it therefore appears likely that the LECs will dominate locally and have multiple options for long-distance exchange.

The IXCs have also begun extensive efforts to move into alternate forms of service, including Internet services, local service, and cellular/wireless-based services. The IXCs have recognized the large portion of costs apportioned to local service, along with the growing threat of LEC competition in the long-distance arena, and have begun to focus on the extension of facilities to local areas. For example, a major IXC announced a partnership with three major cable companies and a CAP in 1994. These activities recognize the likelihood that the IXCs will soon directly compete with the LECs for customer-to-customer service. The evolving roles of LECs and IXCs suggest the possibility of a paradigm shift in agreements for traffic exchange between carriers. For example, today the flow of financial settlements travels in one direction, from the IXCs to the LECs; future environments where today's LECs and IXCs become competitors suggest a two-way flow.

Both LECs and IXCs have also recognized that simple transport services will become more like commodity services. As a result, many carriers have begun to expand their services to include advanced services and information capabilities. Advanced services, such as ATM, enable the service providers to create value with their offerings and differentiate themselves from other providers based on the features and capabilities of the services offered. The addition of information services on top of the transport presents a further opportunity for carriers to differentiate. A likely evolution of information services and telecommunications will bind the two together. The purchase of information services will therefore imply the selection of a specific carrier.

The commercialization of the Internet presents a parallel scenario for interoperability. As in the traditional carrier networks, interoperability is a defining feature of the Internet. However, the open exchange of information that created the Internet has occurred without the influence of business considerations. As the Internet has grown, its potential value to commercial ventures has become apparent. These trends have resulted in an Internet dominated by commercial service providers (note that Sprint's SprintLink service recently became the largest global supporter of the Internet, surpassing the NSFNET). As commercial ventures, Internet services still represent a small fraction of carrier revenues. Therefore, today Internet service providers exchange information more or less freely through agreed-upon exchange points without traffic measurement or settlements. However, as revenues grow, one can reasonably expect that competitive influences will drive the Internet toward more concise, business-oriented operation.

Interoperability tends to contradict service provider efforts to move away from commoditization. As the carrier networks tend to interoperate, they share features and functionality and offer access to the same sources of information, thereby reducing the differentiation associated with information-based services. Interoperability between emerging services, such as ATM, has also evolved slowly. Although this is partially due to the



immaturity of the technologies, the desire for stable customers also contributes to inertia over service interoperability. This is desirable from an NII perspective, but the creation of interoperable networks tends to counter marketplace tendencies among competitors.

### Summary

Expansion of the roles and capabilities of LECs and IXC's will eliminate the current stratification between local and long-distance services. It will also erase the natural requirement for interoperability between LECs and IXC's. Financial settlements for the exchange of services will follow market-based formulas dictated by individual service providers. The larger carriers may be able to leverage their greater reach of services to force unbalanced settlements. Although the evolution of these trends is uncertain, the coexistence of parallel, competing service providers suggests an unstable environment that may affect end-to-end connectivity and the support of emerging NII applications.

The commercialized Internet will begin to mirror the evolving carrier environment over time. The major IXC's have already captured prominent positions in the commercialization of the Internet. As Internet services reach traffic/revenue levels comparable to those of mainstream carrier services, service providers will move toward settlement scenarios that will once again favor the larger providers. The traditional carriers own the facilities and a growing share of the expertise that make a growing, commercial, unsubsidized Internet feasible. However, numerous nontraditional service providers now make up a substantial portion of the Internet. Though these providers depend on the traditional carriers for physical connectivity, they have established strong foundations based on Internet-level services and a growing ability to delivery information services. Segmentation of the Internet appears highly unlikely, but the effect on connectivity is at least uncertain.

The possibility of limited interoperability among service providers, including those offering Internet services, should have a profound effect on the emerging NII. In principle, many of the applications contemplated in NII literature would still be feasible; but numerous obstacles would almost certainly emerge. For example, the concept of telemedicine—the delivery of medical services over the NII—implicitly assumes a seamless telecommunications fabric. Otherwise, medical resources that reside on one service provider's network could not reach a patient on another network. Although this scenario is consistent with the desire for service providers to differentiate, it clearly threatens the realization of telemedicine services.

### RECOMMENDATIONS

The recommendations presented here suggest several possible approaches to ensure the continuation and expansion of interoperability among telecommunications service providers. Overall, further analysis needs to be conducted on the nature of interoperability for the NII and the suitability of the existing models for this requirement. Today it seems that interoperability among competing service providers has been assumed as a natural attribute of a telecommunications infrastructure. On the contrary—there are numerous examples where today's competing service providers do not interoperate. Furthermore, the evolution of the telecommunications infrastructure in terms of technology, economics, and service relationships will tend to eliminate existing areas of interoperability and provide disincentives for further development of interoperable services.

The regulatory changes under way, one of which will relax restrictions on LECs, represent a major influence on service provider interoperability. Explicit consideration of the effect of interoperability of these events should be undertaken. In addition, serious consideration should be applied to additional regulatory actions necessary to ensure continued interoperability among service providers in the NII.

The Internet represents a national and global treasure of information. As the Internet continues toward commercialization, the effects on interoperability among service providers become less clear. The growth of the Internet has brought traditional service providers into the role of Internet service delivery. Although the traditional carriers play a key role in the scaling of the Internet because of their ownership of the physical infrastructure, they also bring the traditional service models and notions of interoperability. The future of the

Internet is widely debated, but specific attention should be given to interoperability as the Internet migrates to a fully commercial enterprise.

Numerous technological factors will affect the evolution of the NII and the interoperability among service providers. Prototype models for carrier interconnection for various services, technologies, and special user communities should be established to determine feasibility and preferred approaches for deployment in the future NII. The NAPs offer an excellent platform for these prototypes, as they represent by definition a means for interoperability among service providers.

## 18

# Funding the National Information Infrastructure: Advertising, Subscription, and Usage Charges

*Robert W. Crandall*

*Brookings Institution*

The stunning changes in information and communications technologies are focusing attention on the gap between the potential services available from new telecommunications networks and the much more limited services now available from our existing communications infrastructure. While this focus is useful in galvanizing political support for action, it may divert attention from the magnitude of the economic task involved in building the new information superhighway. Investors will be committing enormous resources to risky new technologies with uncertain information on future service revenues, the cost of these services, and the prospect of further technical breakthroughs that could render new facilities obsolete in a short time.

In this environment, policymakers must be careful not to foreclose the opportunity to develop new facilities, new services, and new sources of economic support for both. The federal and state governments should be encouraged to allow new services to develop over a variety of distribution systems and to permit vendors of these services to select the funding mechanisms for such systems or services. Indeed, it is precisely because we have already begun to open our telecommunications markets to competition and to reduce government regulation of them that we are now likely to develop the new national information infrastructure more rapidly than most countries and that other countries are moving to emulate us.<sup>1</sup>

In this paper, I review the recent historical data on the evolution of U.S. communications and media industries with a particular focus on the sources of revenue that have propelled this growth and the effects of restrictive government regulation. In addition, I provide a brief review of the potential costs of the new technologies and the uncertainties involved in implementing them. My conclusions are as follows:

- Advertiser-supported media have shown the most rapid growth since 1980. Even subscriber-supported media, such as cable television or the print media, have relied heavily on advertising for their growth.
- Most of the competing technologies for the national information infrastructure will require large capital expenditures, perhaps as much as \$125 billion, in a risky environment.
- Regulatory restraints on new media, designed to protect incumbents, have inevitably suppressed the growth of the communications sector and have often proved unnecessary or even counterproductive.
- In the current risky environment, it would be imprudent to exclude any new services or sources of funds that could be used to defray the costs of building the national information infrastructure (NII).

---

NOTE: Robert Crandall is a senior fellow at the Brookings Institution. The views expressed herein are the author's and not necessarily those of the Brookings Institution, its trustees, or other staff members.

The Coalition for Advertising-Supported Information and Entertainment, which sponsored this paper, is a joint effort of the Association of National Advertisers, the American Association of Advertising Agencies, and numerous leading advertisers and advertising agencies across the country. Copyright © National Academy of Sciences. All rights reserved.

## THE GROWTH OF COMMUNICATIONS MEDIA

I define "communications media" as all of those delivery systems through which information and entertainment are transmitted to and among businesses and consumers. These include traditional print media (newspapers, magazines, journals, etc.), broadcasting, cable television, satellites, and traditional and newer wireless forms of telephone service. Historically, these various media occupied somewhat independent markets in the United States. The print media provided information and advertising aimed at specialized national markets or at decidedly local markets. Radio broadcasting provided entertainment services for which it had little direct competition other than records, but it competed with print media in disseminating local and national news. When television broadcasting emerged, it began to offer entertainment that competed with motion picture theaters, but it too became an important distribution medium for national and local news. Cable television developed as a simple form of television-broadcasting retransmission, but—when allowed to develop without regulatory controls—rapidly became a medium for distributing a wide range of entertainment and information services. More recently, direct satellite transmissions have begun to compete with cable television and broadcasting in the distribution of national entertainment and information services.

Virtually all of the above media have developed with at least some form of advertiser support. In the case of cable television, it was the FCC's decision to allow cable operators to import distant advertiser-supported signals that provided the impetus for cable systems to expand into large urban markets and to increase their channel capacity. These expansion decisions, in turn, facilitated the development of new basic cable networks that were also supported by advertising.

The telephone industry, on the other hand, has traditionally been supported by direct customer access and usage payments. The only source of advertiser support for traditional telephone carriers has been from the publication of Yellow Pages directories. This reliance upon direct customer support was dictated by the nature of the service: two-way switched voice communications. It would have been difficult and disruptive to have advertisements interspersed with these conversations. However, with the development of computer connections through modems or local-area networks, the delivery of information and entertainment services over telephone circuits became possible. As with other communications services (other than voice telephony), customers may now be offered a choice: pay for these services directly, accept advertisements in them, or accept some combination of the two. For example, electronic Yellow Pages may be advertiser supported, require a subscriber usage charge, or both. Similar options may be available for electronic want ads, home-shopping services, or even electronically delivered financial services.

The role of advertising revenues in the growth of the various media is evident in [Table 1](#). Since 1970, there has been substantial growth in all media that offer advertiser-supported content. Video markets, in particular, have grown substantially, but even the print media have enjoyed an expansion of total revenues from advertising and direct circulation fees. In recent years, multichannel video media, such as cable television and direct broadcast satellite (DBS), have grown more rapidly than traditional radio-television broadcasting. In the next few years, as additional DBS services and even terrestrial multichannel wireless systems are built, video service revenues are likely to continue to grow, supported by both advertising and direct consumer outlays.

For more than a decade, regulation restricted the growth of pay or subscription television services in order to protect off-the-air broadcasters. This, in turn, limited cable television to the retransmission of advertiser-supported off-the-air television broadcasting. This repressive regulation was slowly eliminated by the courts, the FCC, and the Congress in the period from 1977 to 1984, spurring cable to grow very rapidly and to expand its channel capacity.

Even though cable subscription revenues grew very rapidly, however, broadcasting revenues continued to expand, in part because cable retransmissions made marginal UHF stations more viable, a result predicted by academic studies of the medium.<sup>2</sup> Thus, cable subscription growth induced further growth of advertiser-supported broadcasting. But even though cable operators would be able to survive and even expand by relying solely on direct subscriber revenues, they also found that they could now tap advertising markets to grow even more rapidly. Direct cable television advertising expenditures began to grow very rapidly in the 1980s, increasing from virtually nothing in 1980 to nearly 20 percent of total cable revenues in 1993.

TABLE 1 The Growth of Communications Media Revenues Since 1970

Year	TV Homes (million)	TV Advertising Revenue (billion \$)	Cable TV Homes (million)	Cable TV Subscriber Revenues (billion \$)	Cable TV Advertising Revenues (billion \$)	TV/Cable Revenues per TV Household (\$)	Newspaper & Periodical Circulation Revenues (billion \$)	Newspaper & Periodicals Advertising Revenues (billion \$)	Telephone Industry Total Revenues (billion \$)	Telephone Industry Directory Advertising Revenues (billion \$)
1970	58.5	3.6	5.1	0.3	—	62	3.0 <sup>a</sup>	7.0	19.4	0.7
1975	68.5	5.3	9.8	0.8	—	77	5.3 <sup>a</sup>	9.7	33.4	1.1
1980	76.3	11.4	17.5	2.4	—	149	9.2 <sup>a</sup>	18.1	59.8	2.3
1985	84.9	20.3	36.7	8.8	0.8	249	13.2	30.9	92.3	2.3
1990	90.8	26.8	51.7	17.2	2.5	323	17.9	40.1	114.6	2.6
1993	93.9	28.7	57.2	22.2	4.0	348	20.3	41.2	128.7	2.9

<sup>a</sup> Estimated by the author from Census data.

SOURCES: Paul Kagan Associates; Statistical Abstract of the United States; FCC, *Statistics of Communications Common Carriers*.

TABLE 2 Real (Inflation-Adjusted) Growth in Revenues for Video, Print, and Telephone Sectors, 1970–93 (1982–84 Dollars)

Year	Television Broadcasting and Cable Television	Newspapers and Periodicals	Telephone Companies
1970	10.0	25.8	50.0
1980	16.7	33.1	72.6
1993	38.0	42.6	89.1
Average Annual Growth, 1970–80	5.1%	2.5%	3.7%
Average Annual Growth, 1980–93	6.3%	1.9%	1.6%

NOTE: All data deflated by the Bureau of Labor Statistics Consumer Price Index.

The telephone industry, supported almost exclusively by subscriber payments for conveying voice and data messages, has grown much more slowly. Telephone companies have not moved aggressively into the provision of content, in part because of regulatory policies that have excluded them from some of the most important of these markets. As a result, their growth has been much slower than that of the print and video media. Indeed, in real (inflation-adjusted) terms, telephone revenues increased at a rate of only 1.6 percent per year from 1980 through 1993. (See Table 2.) During the same period, broadcasting and cable revenues rose at an annual real rate of 6.3 percent per year, and newspaper and magazine revenues rose at a real rate of 1.9 percent per year.

Obviously, telephone companies recognize that their growth has slowed dramatically and have therefore moved aggressively to seek relief from regulation that has kept them from offering new services. In this pursuit, the regional Bell operating companies (RBOCs) have won court reversals of the restriction on their provision of information services that was part of the decree breaking up AT&T.<sup>3</sup> In addition, invoking the First Amendment, several RBOCs have recently mounted successful First Amendment challenges to the legislative ban on telephone company provision of cable television service in its own region.<sup>4</sup>

### NEW COMMUNICATIONS TECHNOLOGIES

In the next few years, cable television operators, telephone companies, and satellite operators will be facing off in the competition to build new broadband distribution networks. These new networks, utilizing a mixture of wireless and fiber technologies, may offer a variety of switched and distributive broadband services. At this juncture no one knows which technologies will ultimately succeed, nor which services will propel these new technologies.<sup>5</sup> This is a risky environment for private investors and policymakers alike given the magnitude of the necessary investments.

Cable television operators now have one-way "tree-and-branch" broadband networks composed of fiber optics and coaxial cables. These networks are typically unable to offer switched services to most subscribers, although such services are possible with major investments in switching equipment and changes in network architecture. To upgrade cable networks simply to provide switched voice (telephone) services would probably cost \$300 to \$500 per subscriber.<sup>6</sup> To provide full interactive switched video might cost as much as \$800 per subscriber depending upon the amount of fiber optics already in the cable system's feeder plant.

Telephone companies are now exploring technologies for expanding the bandwidth of their networks so that they can offer video dialtone or full switched video service. Many companies are pursuing a fiber/coaxial cable architecture that is likely to cost \$800 to \$1,000 per home passed.<sup>7</sup> A full fiber-to-the-curb architecture might be more expensive but provide greater flexibility in terms of bandwidth available for one-way and interactive switched video applications. Simply upgrading the existing copper wire distribution plant with asynchronous digital subscriber loop (ADSL) technology<sup>8</sup> is much more expensive per subscriber, but allows the telephone company to target its investment at only those households that wish to subscribe to its video services.

A recent entrant into the video distribution market is high-powered satellite delivery to the home. This direct broadcast satellite (DBS) service has been launched by Hughes Aircraft and Hubbard Broadcasting and

will soon be emulated by two other consortia, including one that now operates with a lower-powered C-band service. Each of these enterprises requires about \$750 million to \$1 billion in initial investment plus possible start-up losses.<sup>9</sup>

Other wireless services are attempting to compete with cable systems for one-way, distributive video services, but none has yet been shown to be a viable launching pad for the interactive broadband network of the NII variety. It is possible, however, that terrestrial wireless services can be adapted for this purpose, particularly if the FCC shifts to using market mechanisms for all spectrum allocations. Multichannel multipoint distribution services (MMDS) or cellular cable systems could have sufficient channel capacity, particularly with digital signal compression, to offer competition to wired systems. Finally, it is possible that a system of orbiting satellites, such as that being developed by Motorola for narrowband communications, could also provide switched broadband services.

This brief review of the technical possibilities for building the NII is not intended to be comprehensive. Rather, it is included to show that the number of competing technologies is growing and that the ultimate winner or winners is (are) not known or even likely knowable at this time. Some telephone companies are reexamining their earlier commitments to a fiber/coaxial cable technology.<sup>10</sup> At least one large cable television/telephone company merger has failed, and other consortia may be foundering. DBS systems may have arrested some of the momentum for building interactive terrestrial broadband networks because the latter would probably have to rely heavily on traditional one-way video programming at first. However one interprets the recent technological and economic developments in this sector, it is quite clear that building versions of the NII is not becoming less risky.

Even at costs as low as \$800 per home, the cost of extending the NII to all residential households is at least \$80 billion. To extend it to all business and residential customers would require at least \$120 billion.<sup>11</sup> This latter estimate is almost exactly equal to the book value of the entire net plant of all regional Bell operating companies and about 80 percent of the book value of the net plant for all telephone companies in the United States.<sup>12</sup> Thus, if the NII is to be built by established telephone companies with technology now under development, it would probably require a near doubling of these companies' assets. Since no one is seriously advocating government support that would cover even a small fraction of this added investment, private firms must have the ability to obtain large cash flows from a variety of sources to fund such a massive expansion of our communications infrastructure.

The sobering fact about any such large investment in the NII is that it will likely be overcome by different or improved technology soon after it is built. Perhaps such networks can be continuously adapted to these technical improvements, but it is also possible that a completely different technology—perhaps based on orbiting satellites, for example—could render an \$80 billion to \$120 billion terrestrial investment obsolete within a decade. Under these circumstances, investors should be permitted to explore all possible sources of revenues from the marketplace if we expect them to commit such large amounts of capital to so risky an enterprise.

### RECENT TRENDS IN COMMUNICATIONS-SECTOR CAPITAL FORMATION

The explosion in new technology might be expected to generate a boom in investment in the various communications sectors. The advances in fiber optics, electronic switching, asynchronous switching, digital compression, and consumer electronics are providing all market participants with the opportunity to deliver a multitude of new services if they invest in new equipment.

The evidence on capital formation in the communications sector only partially reflects this rosy assessment. Telephone companies responded to the new opportunities by investing in new switching systems and fiber-optics distribution systems through the early 1980s. Much of this investment was spurred by competitive entry into long-distance markets and even some urban local-exchange markets and by the necessity for local carriers to provide equal access to these new competitive carriers in their switching systems. Thereafter, however, telephone company investment slowed markedly. (See [Figure 1](#).) In the past 5 years, investment in this sector of the economy has been insufficient to maintain the net value of its capital stock.<sup>13</sup>

By contrast, the radio-television sector continued to expand its net capital stock through 1993. Fueled by cable deregulation in 1977–79, the net capital stock in this sector nearly doubled between 1979 and 1989—a

period in which telephone plant increased by less than 20 percent. (See Figure 2.) This growth reflects the substantial expansion in the number of homes wired by cable in the 1980s as well as the investments in increasing channel capacity during these years.

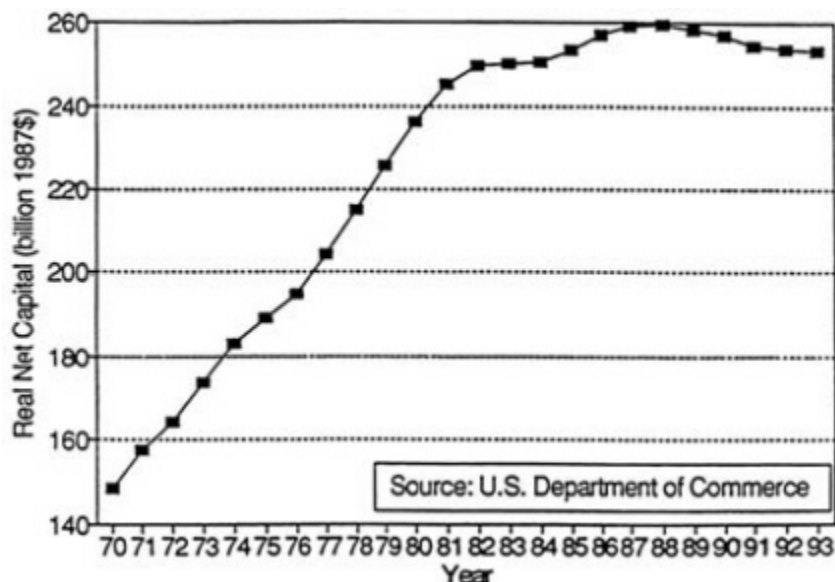


Figure 1 Real net capital stock—U.S. telephone industry, 1970 to 1993.

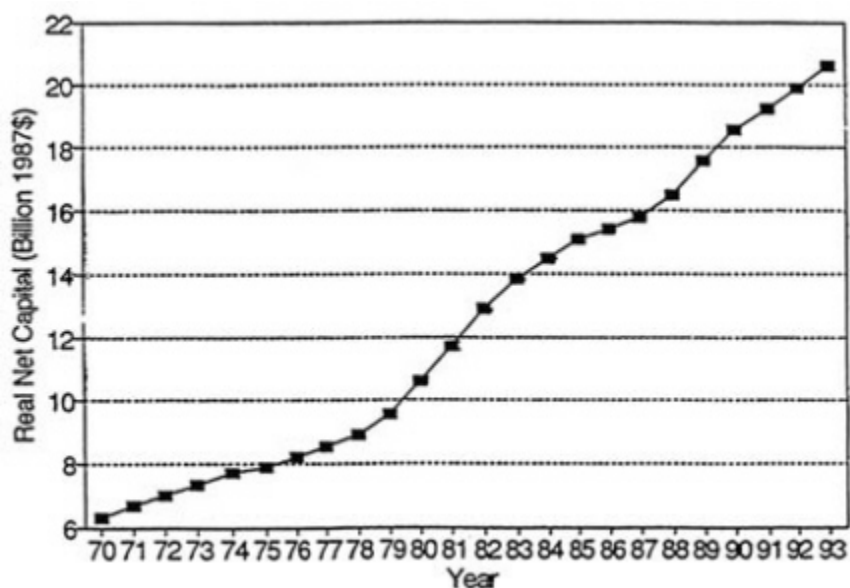


Figure 2 Real net capital stock—television and radio industry, 1970 to 1993.

A lesson to be learned from these divergent trends in capital formation is that regulatory constraints that restrain competition and new entry, whether in the name of promoting "universal service" or "fairness," create a substantial risk of slowing investment and the spread of new technology. Cable television investment accelerated after the severe regulatory restrictions on cable service, motivated by the desire to protect off-the-air broadcasting, were lifted in the late 1970s. The removal of these restrictions caused cable revenues to rise rapidly, fed by both advertising and subscriber fees, while traditional broadcast advertising revenues continued to rise. Telephone investment rose rapidly in the 1970s after entry restrictions on long-distance competitors were lifted. Once restrictions were placed on telephone company offerings of information and cable television services, however, investment in traditional telephony slowed dramatically. Were new investments in the NII to be constrained by the desire to cross-subsidize new universal-service offerings or by a concern that advertiser-supported services



might supplant other services, the prospects for rapid deployment of risky new technology could be severely impaired.

### **ADVERTISING VERSUS SUBSCRIBER FEES—THE IMPLICATIONS FOR ECONOMIC WELFARE**

Economists have studied the economic welfare implications of using advertising or direct subscriber payments to support radio and television programming. The results of this research suggest that a mix of the two support mechanisms is likely to be the optimal source of revenues.<sup>14</sup> The reasoning is fairly straightforward.

Advertisers are likely to be motivated by the desire to maximize their reach among a target population at the lowest possible cost consistent with persuading consumers of the attractiveness of their products. For this reason, they are likely to eschew the support of programming (or other software) that appeals to a narrow set of potential consumers. Thus, advertiser-supported services are likely to be less finely differentiated than are subscriber-supported services under most circumstances. If some consumers are not willing to pay much for highly differentiated services, their economic welfare will be greater with advertiser-supported services. It is for this reason, presumably, that more than one-third of all households now passed by cable television do not currently subscribe despite the fact that cable offers much more choice than the available array of broadcast stations in even the largest markets.

On the other hand, some consumers may desire more choices than those that can or would be supported totally by advertising. Subscriber-supported services can reflect the intensity of consumer preferences, not simply their willingness to use a service—the criterion that is important to advertisers. It is for this reason that pay-cable and pay-per-view services have proliferated in the current multichannel video environment. Only a few hundred thousand subscribers, each paying a substantial one-time user fee, may be required to support two or three hours of a nationally distributed service or event. Foreign soccer games, obscure older feature films, or live concerts may be offered for a one-time charge, while more popular events may be advertiser supported.

A third possibility for supporting services on the current or future information infrastructure is a mixture of user charges and advertiser support for the same service. Pay-per-view services could still carry advertising billboards or even interspersed advertising messages. For example, videocassettes and motion picture theaters are now supported by both subscriber charges and advertising. In addition, the same programming could be offered simultaneously without advertising but with a subscriber fee and with advertising and no subscriber fee. This combination of support mechanisms could lead to greater diversity and a larger array of available services.

It is important, therefore, that policies for developing the NII should not directly or inadvertently exclude new service options that might contribute importantly to the cash flows necessary to amortize the large investments required to build the infrastructure. As long as there are no prohibitions on advertiser support or direct subscriber payments in a competitive marketplace, there is little reason for policymakers to attempt to prejudice the choice of support mechanisms for the current information infrastructure or for the NII. It can always be argued that virtually any information or entertainment service has attributes of a public good. One consumer's use of such a service may not reduce the amount of it that is available for other consumers; therefore, no consumer should be excluded from consuming it through a price that is above zero. Advertising could help provide a solution to this problem since it can be employed to generate free broadcasting or print services with low subscriber fees, for example. However, without a combination of advertising and direct subscriber fees, there is no mechanism for consumers to indicate the intensity (or lack of intensity) of their preferences.<sup>15</sup>

### **REGULATING THE NII**

It is impossible for anyone to understand fully the potential of the changes that are now occurring in telecommunications technology. Even if the technical choices for network design could be forecast for the next 10 or 20 years, no one could possibly predict how these technologies would be used. The potential changes in the delivery of financial, shopping, entertainment, medical, and government services—to name a few—that are likely

to occur are so vast that they cannot possibly be forecast with any precision. But the ideal design of new networks will be highly influenced by both technology and the evolution of and demand for these new services. Therefore, investments in new networks will necessarily take place on a trial-and-error basis.

In this environment, it would be counterproductive for government policymakers to attempt to guide investments or to require carriers to offer services to segments of the population at nonremunerative rates in the name of "universal service." Such demands can only be made on carriers that are protected from competition in other markets because intra-firm subsidies require that the firm earn supra-competitive returns somewhere. If certain (early) NII services are selected as the domain for universal-service requirements, the protection afforded the carrier(s) in other markets may inhibit entry and the introduction of new technology elsewhere. This was precisely the outcome of the FCC's protection of television broadcasters from the development of cable in the 1960s and 1970s and of the federal/state protection of telephone companies from competition in customer-premises equipment, long-distance services, and wireless services.

The calls for preferential access to the NII from a variety of worthy public institutions are understandable, but they should be resisted if they require carriers or other service providers to cross-subsidize them from other service revenues. If there is a legitimate case for subsidizing some of these institutions' access to the NII, it would be far better to provide these subsidies directly from public tax revenues. Otherwise, the built-in cross-subsidies will serve as the rationale for barring new entrants with even newer services in the years to come. If there is one lesson to be learned from past exercises in regulatory protection in this sector, it is that such protection delays new services, such as cable television, cellular telephony, personal communication service, video-on-demand, or direct broadcast satellite services. Free entry to new facilities, new services, and all sources of revenues will provide the fastest route to the NII.

#### NOTES

1. For example, Japan is now urgently considering new approaches for allowing competitive entry into its telecommunications markets, including the possible emulation of our approach to breaking up AT&T. The European Community, which featured entrenched Postal, Telegraph, and Telephone authority (PTT) control of its national telephone monopolies, is now moving to liberalize its telecommunications sector by January 1998. Australia has already admitted a second telephone carrier, Optus, and New Zealand has not only privatized its national carrier and opened its market to competition but has also completely deregulated telecommunications. Finally, Canada has recently admitted entry into its long-distance sector and is considering the liberalization of local telephone service and video delivery services.
2. R.E. Park was the first to show that cable television would increase the viability of UHF broadcast stations. See R.E. Park, "Cable Television, UHF Broadcasting, and FCC Regulatory Policy," *Journal of Law and Economics*, (1972), Vol. 15, pp. 207-31.
3. This restriction was one of the line-of-business restraints on the RBOCs that were built into the antitrust decree that was entered in 1982 to break up AT&T. The information-services restriction was removed in 1991. (See Michael K. Kellogg, John Thorne, and Peter W. Huber, *Federal Telecommunications Law*. Boston: Little, Brown and Company, 1992, Section 6.4, for a detailed history of this restriction.)
4. The successful challengers have been U S West, Bell Atlantic, Southwestern Bell (now SBC), NYNEX, Southern New England Telephone, GTE, BellSouth, and Ameritech. These cases have been brought in various U.S. District Courts and, in two cases, U.S. Courts of Appeals. The Supreme Court has yet to rule on the issue.
5. An excellent example of this uncertainty was provided by Bell Atlantic in the past month. It has withdrawn its Section 214 applications for the hybrid fiber/coaxial cable and ADSL technologies, announcing that it now wishes to pursue the fiber-to-the-curb technology and various wireless technologies, such as MMDS.
6. See David P. Reed, *Putting It All Together: The Cost Structure of Personal Communications Services*. FCC, OPP Working Paper #28, November 1992, p. 35, and "Will the Broadband Network Ring Your Phone," *Telephony*, December 6, 1993, p. 34.
7. This estimate is based on recent filings by Pacific Bell before the Federal Communications Commission (see, for example, Robert G. Harris, "Testimony in Support of Pacific Bell's 214 Application to the Federal Communications Commission," December 14, 1994). Slightly higher estimates are to be found in David P. Reed, *Residential Fiber Optic Networks: An Engineering and Economic Analysis*. Boston: Artech House, 1992.

8. This technology allows telephone companies to deliver video services over the paired copper wires that are now used to connect subscribers to the companies' plant.

9. See "Digital TV: Advantage, Hughes," *Business Week*, March 13, 1995, pp. 67–68. Similar estimates may be found in Leland L. Johnson and Deborah R. Castleman, *Direct Broadcast Satellites: A Competitive Alternative to Cable Television?* RAND, 1991.

10. Bell Atlantic has recently withdrawn its Section 214 applications for hybrid fiber/coaxial cable technology but remains committed to building a broadband network using other technologies, such as fiber to the curb.

11. This estimate is based on a national total of about 150 million access lines. (United States Telephone Association, *Statistics of the Local Exchange Carriers, 1993*.)

12. FCC, *Statistics of Communications Common Carriers, 1993/94* edition, p. 38.

13. It should be noted that part of this slowdown is attributable to the liberalization of customer premises equipment. Since the late 1970s, customers have been able to purchase their own telephone handsets, key telephone systems, PABXs, modems, answering machines, and fax machines. As a result, perhaps 20 to 25 percent of all telephone-related capital equipment is owned by telephone customers. See R.W. Crandall, *After the Breakup: U.S. Telecommunications in a More Competitive Era*. Washington, D.C.: The Brookings Institution, 1991.

14. For an excellent review of this literature, see B.M. Owen and S.S. Wildman, *Video Economics*. Cambridge, Mass.: Harvard University Press, 1992, ch. 4.

15. For an excellent compendium of studies of the public-goods problem, see T. Cowen (ed.), *The Theory of Market Failure*. Fairfax, Va.: George Mason University Press, 1988.

## 19

### The NII in the Home

*D. Joseph Donahue*

*Thomson Consumer Electronics*

A successful national information infrastructure (NII) strategy must include a clear vision, plus an action program directed at bringing NII usage in the home to the pervasive level at which we use telephones and television today.

Fortunately, there is an approach that allows the unique strengths of two large industry groups to be applied to the introduction of extensive NII capabilities into a full cross section of American homes. These two industries are broadly defined as the computer and television industries. Neither alone can provide the full range of services in response to the consumer's interests and desires. Full utilization of the strengths of both industries will yield a win-win strategy that could greatly accelerate the introduction and acceptance of diverse NII services in consumer households.

The powerful capabilities of PCs, combined with online services and the Internet, have already provided the initial interest and stimulus for the concepts of a comprehensive NII. The importance of the continued growth and acceptance of these capabilities cannot be overemphasized. One need only look at the sales of computers and software—or the use of the Internet—to feel the ever-increasing utilization in commerce and in the home. And there is no end in sight for the dynamic expansion of the capability of these products and services.

The television industry—with 99 percent household penetration—can also make profound contributions to the acceptance and growth of NII in the home. Until recently, the television industry was solely based on analog technology, which has many limitations when viewed from today's vantage point. The current movement of the television industry to digital systems will allow television to diversify and expand its capabilities in a manner analogous to that of the products and services of the computer industry.

Almost all of the new television systems are based on MPEG 2, which uses flexible transport packets. New digital television signals are thus no longer simply television signals—they are MPEG 2 packetized bit streams that can be used to deliver any desired mix of video, audio, and data. Interactive services over television systems are now also a reality. For example, a Thomson Consumer Electronics-Sun Microsystems partnership recently announced an interactive operating system, "OpenTV," designed to work over digital systems with interactivity interfaced through the TV remote control or the PC.

The strong consumer interest in television will allow interactivity to be introduced in a nonthreatening manner to the broad segments of society not currently disposed to using a PC. As a result, all members of society will be able to learn to use interactivity with ever-increasing levels of sophistication. Digital television, which will initially be purchased for its entertainment value, can be a key vehicle that can be used to attract consumers and help finance the installation of the digital pathways to digital hardware in the home. Digital HDTV delivery over any media will provide a 19.4-Mbit/sec service to homes. In traditional one-way services such as terrestrial broadcast and direct broadcast satellite, the telephone line is used for the return path. This arrangement is standard in Thomson's Digital Satellite System (DSS) now used to deliver Direct-Tv and USSB signals nationwide.<sup>1</sup>

The dual evolution and penetration of computer and digital television services and hardware into the home represents a major win-win victory for all parties concerned. Both industries will use many of the advances of the other. PCs are adding audio and video. Television-receiving products will become digital with microprocessors

and interactivity. Both will stimulate the construction and acceptance of improved interactive networks to the home.

Consumer interest will determine whether interactivity is first introduced into homes through the PC or the digital television. Over time, homes will contain a powerful PC capability in a den or work location and a sophisticated digital television entertainment center in the family room. Many of the communication networks and software programs will serve both home setups. Full use of the attractive services and products of both industry groups will greatly accelerate the development and use of NII in the home.

Maximum implementation of digital television and interoperability across all media require certain federal government and industry actions.

- What is key is the completion of the Federal Communications Commission (FCC) Advisory Committee on Advanced Television Service (ACATS) process, with the FCC adoption of the Grand Alliance (GA) HDTV standard as early as possible—late 1995 or early 1996—as recommended by the May 1994 NIST Workshop on Advanced Digital Video in the NII and the recent report by the NII Technology Policy Working Group (TPWG). FCC action must include allocation of new 6 MHz transition channels to all broadcasters. This ACATS-FCC action is moving along toward completion.
- Establishment of new infrastructure network rules for the previously separate industries of local telephone, long-distance telephone, cable, broadcast, and so on. for all NII type services is vital. Maximum network development and investment must await a clear set of regulations.
- The more difficult government-industry challenge is the establishment of an open interoperable infrastructure within the digital video world. Closed proprietary systems and the potential for many different video systems and interfaces will retard consumer acceptance. Confusion over systems, standards, and interfaces, as in the past, will cause consumers to delay acquisitions. One key ingredient that would help is to provide consumers with the option of buying all home hardware at retail from competitive suppliers. Consumer decisions plus competition will help to establish open and interoperable systems and products.

A concern expressed in some quarters is industry's commitment to commercialize systems and products. Commercial commitments will not be a significant problem if the obstacles cited here can be dealt with. As an example, the digital television actions of Thomson Consumer Electronics are tabulated below. Every reasonable effort in the areas of standards, product development, and promotion is being supported, to accelerate the conversion of the home entertainment center into an exciting new interoperable digital center with uses far beyond those common today.

Thomson's digital video activities include the following:

- Key participant in development of MPEG 1 and MPEG 2 standards.
- Leading participant and early proponent of the use MPEG flexible packets for the U.S. HDTV standard. Charter member of earlier Advanced Digital HDTV Consortium and of the recent Grand Alliance HDTV development team.
- Developer, manufacturer, and marketer of the RCA Digital Satellite System (DSS), the first high unit volume digital video system ever implemented.
- In cooperation with Sun, announced an interactive digital operating system, "OpenTV," that is extremely economical and that can be interfaced through a TV remote control or a PC.
- Announced MPEG 2 encoder capability for SDTV (1995) and HDTV (1997).
- Announced capability to produce set-top receivers with full microprocessor capability.
- In cooperation with Hitachi, demonstrated and announced commercial plans (1996) for a digital D-VHS VCR for the home recording of the DSS signals.
- In cooperation with Toshiba and others, announced standard, manufacturing, and commercial plan (1996) for a digital videodisc (DVD) player.
- Chaired worldwide working group that reached consensus on a digital recording standard (DVC) for the GA HDTV system.
- Announced plans for manufacture and sale of digital television receivers with interactivity (1997).

- Announced plan for digital HDTV receivers with interactivity (1997).
- Announced program for a DVD-ROM product (1997).

**NOTE**

1. As an aside, Thomson announced the shipment of the one-millionth DSS home unit in less than ten months from a standing start. No other consumer electronics or other major product—color TV, VCR, CD, and so on—has ever been accepted at a rate even approaching this level. Consumers are prepared to accept the new digital television systems and hardware.

## 20

# The Evolution of the Analog Set-Top Terminal to a Digital Interactive Home Communications Terminal

*H. Allen Ecker and J. Graham Mobley  
Scientific-Atlanta Inc.*

### ABSTRACT

This paper addresses the evolution of the cable home terminal from past to present and presents a most likely scenario for its future evolution. Initially, a simple converter allowed subscribers to tune more channels than the initial VHF Channel 2 through Channel 13. Next, conditional access and addressability allowed programmers and network operators to offer subscription-paid premium programming. Today, advanced analog home communications terminals (HCTs) allow downloading of software to provide electronic program guides, virtual channels, and menu-driven navigators. Evolution will continue from advanced analog HCTs to a fully interactive digital HCT.

The enabling factors that allowed the cable industry to grow to become the primary entertainment delivery system to the home are the following:

1. The availability of satellite delivery of content to headends;
2. The availability of a broadband plant allowing many channels; and
3. The availability of set-top terminals that have grown in functionality to truly become HCTs with reverse path capability.

Currently, cable systems provide broadband delivery to the home that allows broadcasting of many channels. In the broadcast mode, all programs pass each HCT so that each and all HCTs can receive and display the programs. In most instances, premium programming is scrambled before transmission so that only authorized customers can descramble premium channels. The authorization can be handled through individual addressing of each HCT.

To evolve to an interactive digital terminal that allows client/server type applications, four key ingredients are required.

The first is the requirement for a migration strategy to digital that uses broadband hybrid fiber coaxial systems. This strategy allows the digital signals to coexist with analog broadcast channels from which most revenue is derived today.

The second ingredient is the development of high-density integrated circuits to reduce the cost of the complex digital and analog signal processing inherent in any interactive terminal. Circuit density will increase from 0.8 micron to 0.5 micron to 0.35 micron geometries over time.

Third is the development of a multimedia operating system (OS) and user interface (UI) that allows the user to navigate and interact with the wide variety of content applications made available. This OS/UI must also be programmer friendly so that many different applications providers can develop applications that will run on home terminals.

Finally, the system must be designed for full interoperability at critical interfaces and must operate efficiently within the regulatory environment. Specifically, the interactive home terminal must contain a network interface module (NIM), which would be provided by the network operator but could have a consumer digital entertainment terminal (DET) that could be moved from network to network and that has an open architecture operating system and an open application program interface (API).

## HISTORICAL BACKGROUND

Community antenna television, or CATV as it is now called, has its roots extending as far back as the late 1940s. It is widely accepted that the first CATV system was constructed in Mahanoy City, Pennsylvania, by an appliance store owner who wanted to enhance the sale of television sets by making programming available to customers that could not get reception of off-air signals since the community was located in a valley in the Appalachian Mountains. A large antenna was erected on top of a utility pole atop New Boston Mountain and was used to feed wires and repeater amplifiers. The system was "able to provide sufficient reception of television programs so as to not only sell television sets but to obtain subscribers to his cable connection"<sup>1</sup>. From this early attempt to provide cable service, additional systems were constructed with the intent to provide clear reception of locally generated signals to customers unable to receive off-air signals. By the mid-1960s there were over 1,300 cable systems feeding approximately 1.5 million customers.

Architecturally, these early cable systems were merely distribution systems that received and amplified off-air signals without processing and reassigning redistribution frequency. However, the 1970s brought new technological developments. The launching of geostationary satellites, coupled with the allowance by the FCC to make these satellites available for domestic communications, brought about the growth of satellite delivery of programming directly to cable headends. For the first time, television programmers and programming distributors could use satellites to broadcast their programming with complete U.S. coverage. CATV systems could receive and distribute these signals via cable to customers with little incremental investment. Television broadcast stations such as WTBS in Atlanta and WGN in Chicago made use of this technology to create superstations with full continental U.S. coverage of their signals. In addition, FM satellite receivers, large satellite receiving antennas, and video processing equipment provided cable headends with the necessary electronic equipment to receive and distribute these programs over selected channels in the cable system.

## HOW THE ANALOG CABLE SET TOP EVOLVED

For the first time, there were more channels available over the CATV system than the normal Channel 2–13 VHF television could tune. This availability of greater channel capacity led to the development of the cable tuner or "set-top converter." These early units, developed in the late 1970s, were capable of tuning up to 36 channels. They did little more than convert the tuned cable channel to a standard Channel 3 or Channel 4 output frequency so that a standard VHF television set could receive the cable signals. In effect, these units were the cable analog of the UHF converter prevalent in the late 1950s. Recognizing the vast market made available by the satellite and cable distribution system in place, content providers such as Home Box Office, Cinemax, Showtime, The Movie Channel, and The Disney Channel as well as ESPN, USA Network, and others sought to use the existing infrastructure to provide "cable only" delivery of premium and "niche" programming content to a growing audience hungry for content variety.

This programming was meant to service only those who purchased it and therefore required means for selectively enabling only those viewers who paid for these special services. To fulfill the requirement, frequency traps were installed in cable drops to nonsubscribing customers. This technique had several drawbacks:

1. Traps were used to prevent reception, proving to be a costly proposition since the expense was distributed over nonsubscribing customers.
2. Customers who changed their viewing tastes required a service call by the cable operator to either remove or add traps.
3. If the cable operator added a new premium service, new traps were needed for each subscriber *not* desiring the new service.

Scrambling and addressability removed these road blocks and made premium programming a viable business. In the 1980s, scramblers began replacing frequency trap technology. Premium programming content was scrambled before being redistributed over the cable system. Newer set-top converters were provided that



contained addressable descramblers. These units had unique preloaded serial numbers so that individual converters could be addressed remotely from the cable headend and either enabled or disabled for descrambling of particular premium programming content or tiers of programs.

### **More Channels and Better Signal Quality Using Analog Fiber**

In an attempt to increase channel capacity, improve penetration, and provide better quality of service, cable systems in the 1990s substituted analog fiber technology for the coaxial RF trunk and distribution systems. Analog fiber provided a number of advantages:

1. It reduced the number of amplifiers needed to span large distances, thereby increasing reliability and signal quality.
2. It allowed cable systems to penetrate deeper into outlying areas without a sacrifice in signal quality.
3. It reduced the powering requirements.
4. Analog fiber was selected instead of digital fiber because it made conversion from fiber back to analog coaxial distribution inexpensive and simple.

The addition of analog fiber also changed the architecture of a typical cable distribution plant. In the process of providing better signal quality deeper into the distribution system, the topology slowly converter from a "tree and branch" layout, in which feeder lines split from a main trunk, to a "star" topology where clusters of subscribers were connected to a fiber node that was fed directly from the cable headend. This architecture, known as fiber to the serving area (FSA) or hybrid fiber coax (HFC), reduced the number of subscribers being fed from a single node and increased the number of nodes fed directly from the cable headend. This evolution has enhanced a cable system's ability to convert from a broadcast topology, where bandwidth is assigned to a program, to an interactive topology, where bandwidth is assigned specifically to connect a content provider and a single customer.

Hybrid fiber coax also provides more bandwidth for reverse path signaling. In the Time Warner Full Service Network in Orlando, reverse path amplifiers are used to connect HCTs to the fiber node in the spectrum between 750 MHz and 1 GHz. Reverse path time division multiple access (TDMA) signals are converted from RF to analog, fiber at the nodes and piped back to the headend over fiber. Other systems operate in the same way except that the band from 5 to 30 MHz is used between the fiber nodes and the home HCT. The "star" topology of HFC provides more reverse path bandwidth per home because of the lower number of homes connected to a node.

### **Cable Becomes a System**

Cable in the 1990s has evolved into a complete communications system that provides broadband delivery to the home of many channels, both basic unscrambled as well as premium scrambled channels. Through the use of computer-controlled system managers coupled with a conditional access system, individual set-top terminals can be quickly and efficiently electronically addressed to allow for the descrambling of particular premium channels.

To aid in program selection, advanced analog set-top terminals today feature downloaded fully integrated program guides and a system navigator that informs subscribers of the vast programming available. Also, using the 150 kbps average bandwidth available in the vertical blanking interval of a standard video channel, these units incorporate "virtual channels" that can be used to send sports updates, financial services, news, and other types of digital services. These units also can be equipped with a standard read-only-memory (ROM)-based operating system or with an addressably renewable operating system (AROS) that allows the operator to download new software to the HCT via the system manager to provide new user interfaces, unique electronic program guide on-screen formats, bitmapped graphics, and multilingual on-screen displays.

Also, reverse path signaling has made impulse pay-per-view (IPPV) and near-video-on-demand (NVOD) a practicality. With an integrated IPPV module, subscribers can view events at their leisure. Also, NVOD capability offers subscribers advantages similar to those of VCR tape rental by defining events with staggered start times on multiple channels. Advanced HCTs have the capability to emulate a VCR's pause, fast-forward, and rewind features by making use of the staggered simulcast nature of NVOD channels and the built-in intelligent software resident in the HCT to keep track of the appropriate channel needed to view a desired program segment.

### **Digital Communications with Analog TV Signals: An Added Dimension**

While current cable systems use analog signals for video and audio, advancements in digital technology now allow cable systems to add a digital video layer to increase channel capacity with little or no increased distribution bandwidth.

Two technological breakthroughs in digital processing are clearing the way for digital video and audio program content. The first is the adoption of standards for digitizing, compressing, and decompressing video programming. In 1992, the Moving Picture Experts Group (MPEG) of the ISO set out to develop a set of standards for digitizing and compressing an analog video signal. This international standards group laid the groundwork for standardizing the algorithms, syntax, and transport format necessary to allow interoperability among different suppliers of video compression and decompression systems. The attitude at the outset was that if digital television was to flourish, equipment built by different vendors must be compatible and interchangeable. The international standards adopted by the MPEG committee allow freedom to be creative in the encoding process within the parameters of the defined "encoding language" while maintaining compatibility with standard MPEG decoders. The analogy in the computer programming world is to say that software programmers can approach a programming problem in many different ways; however, if their program is written in C, for example, any C compiler can compile and execute the program.

The second development was a means for delivering the digital signals to the customer. Schemes for modulating and demodulating an RF carrier using either quadrature amplitude modulation (QAM) or vestigial sideband modulation (VSB) have been developed. These approaches are compatible with standard analog cable systems and can deliver data information rates up to 36 Mbps in a single 6-MHz channel. The combination of digital compression of video and audio and digital transmission over cable can increase the number of video services in a single 6-MHz channel by a factor of approximately 5 to 10 depending on programming content and picture resolution.

Not only does digital processing increase capacity, but it also allows greater flexibility in the types of services that can be provided. No longer will there be the restriction that the information provided be a video signal. Digital computer games, digital music, and other multimedia applications will all be likely candidates for the new broadband digital cable.

The cable set-top terminal needed to fully utilize these new services will evolve to a fully interactive home communications terminal (HCT) that allows client/server functionality.

### **Migration Strategy for Broadband Cable Systems of the Future**

So how do we evolve from the cable system of today that broadcasts analog video services to the home to the fully interactive digital system of the future? This evolution will most likely occur in four phases.

The first phase in the migration would employ advanced analog home communications terminals (HCTs) that are compatible with existing hybrid fiber coax distribution plants. These terminals, available today, not only can tune up to 750 MHz but also have both forward- and reverse-path digital communications capability for downloading digital applications, providing on-screen display of program guides and hand-held remote navigators for ordering pay-per-view and NVOD.

The second phase most probably will use digital broadcast HCTs that can not only tune and display the standard analog video but also receive and decode digital video services. These digital services would be overlaid on the HFC cable plant using unused spectrum above the standard analog video channels. This strategy would allow the digital services to coexist with analog broadcast channels from which most of today's revenue is derived. System upgrades and even "new builds" will need to provide today's analog channels to existing set-top terminals and customers who are content in viewing basic and premium scrambled broadcast services. A hybrid fiber coax system transmits the digital information on analog RF carriers designed to be frequency multiplexed into a normal broadband channel. The bandwidth required is compatible with a normal analog video channel so that the digital signals can be assigned a vacant space in the spectrum or, because of enhanced ruggedness, can be placed at the upper end of the spectrum outside the guaranteed performance range of a distribution plant.

The third phase in the migration would provide the analog/digital HCTs with full interactive capability. This phase would require an incremental upgrade in the HCT to provide more memory and enhanced reverse path signaling and full interactivity. However, the major upgrades would evolve over time as the plant architecture migrates from a "tree and branch" to a "star" configuration and more cable headends become interconnected. This change would be coupled with the growth of content-provider digital applications stored on file servers, market demand for specialized applications, and the success of true "video dialtone" as compared with NVOD.

The final phase, illustrated in [Figure 1](#), is a fully integrated broadband network that provides full connectivity to the home for POTS services, broadband digital modems for personal computers (PCs), analog and digital video services, digital applications, and utility monitoring. Either a fully integrated HCT or a split between the functionality of an HCT and a customer interface unit (CIU) could be used. This last phase could happen as slowly or as rapidly as needed depending on market demand for interactive services.

The phases outlined make sense because they allow for the coexistence of analog only, digital broadcast, and full interactive services and HCTs on the same system and do not require flash cuts to implement, nor do they sacrifice analog video revenues. Thus, the revenues track with the cost of upgrades.

#### **COST DRIVERS FOR THE FULLY INTERACTIVE HOME COMMUNICATIONS TERMINAL**

This migration will be paced by key cost drivers in the HCT, the distribution plant, and the headend. HCT cost appears to be the controlling factor because each digital user must have a digital HCT. Also, developing an economic model that allows enough incremental revenues to support needed upgrades will be key.

The cost of the fully interactive digital HCT depends on the availability of a cost-effective set of application-specific integrated circuits (ASICs) for digital and analog signal processing and digital memory inherent in the interactive digital terminal. Currently most if not all the necessary ASICs have been designed to perform the critical processing functions. These functions include digital demodulation of 64/256 QAM digital signals, adaptive delay equalization of the digital signals, error correction, program demultiplexing and conditional access, and finally MPEG video and audio decompression. Also, microprocessors and graphics accelerators are available for a platform for the terminal operating system (OS) and user interface (UI). Several trial systems such as the Time Warner Full Service Network in Orlando, Florida, and the US West fully interactive trial in Omaha, Nebraska, are using HCTs that include these ASICs. In 1995 other systems using these chips will be deployed by BellSouth, Ameritech, Southern New England Telephone, and PACTEL.

However, these chips are currently designed using today's 0.8-micron technology, and the number needed to build a fully interactive HCT is about seven to nine chips containing about 975,000 gates and 1,500 pinouts. To reduce cost, the goal is to design the required functionality into one or two chips. Projections are that by mid-1996 the number of ASICs could be reduced to four using 0.5-micron technology, and by 1998 the goal of two chips could be realized using 0.35-micron technology. Reducing the number of chips would dramatically reduce the number of pinouts, thereby decreasing cost and increasing liability.

The second key cost driver is the cost per subscriber for the headend and distribution plant. In an analog broadcast or digital broadcast scenario, including NVOD, the cost per subscriber for the headend and distribution plant can be normalized over the number of subscribers. Therefore, increasing penetration reduces operating cost

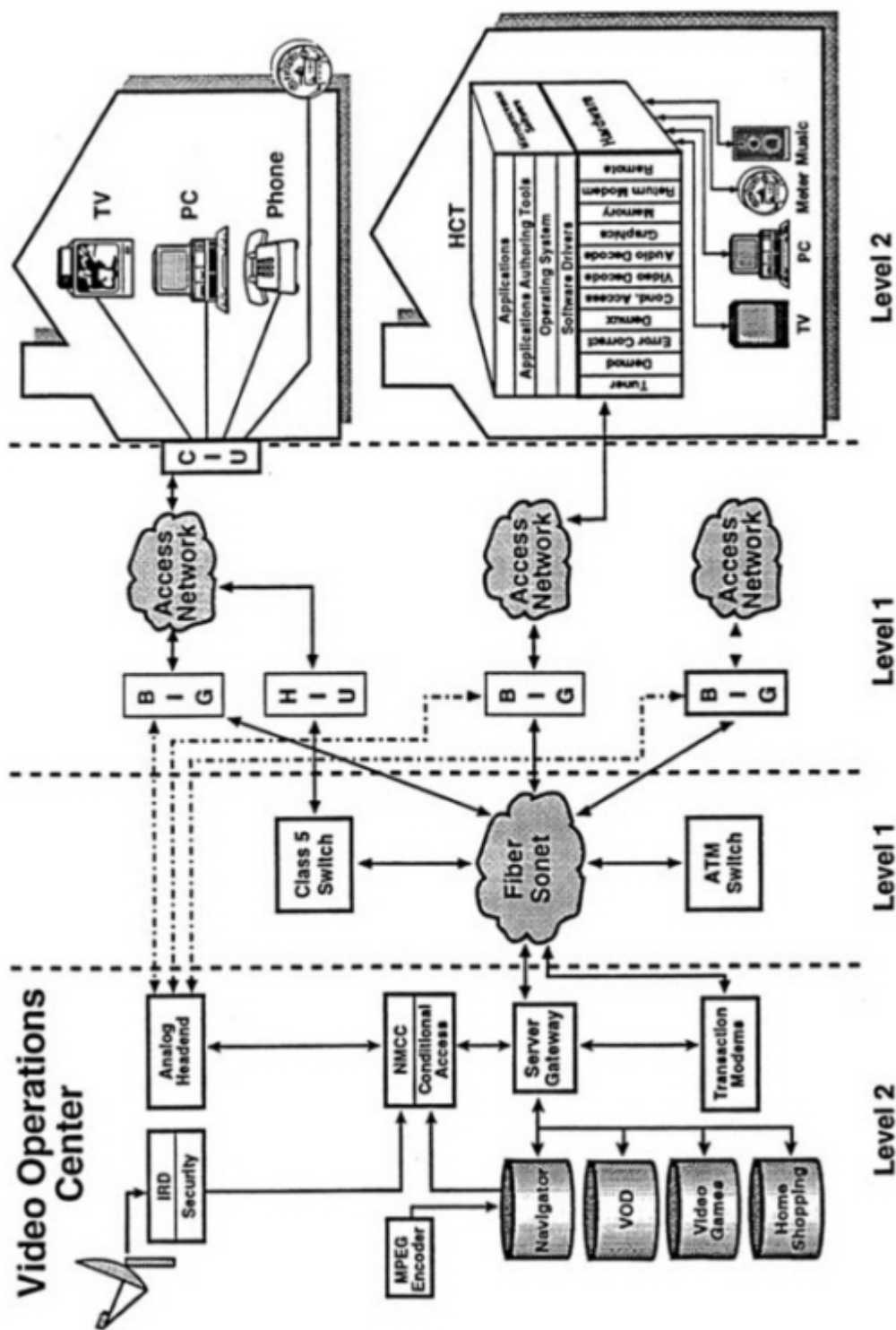


Figure 1 A fully integrated broadband network providing full connectivity to the home.

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

per subscriber. However, in a fully interactive network, interactive server/client bandwidth must grow proportionally with the number of subscribers, so that plant and headend cost increases with penetration to meet the anticipated heavier demand for simultaneous interactive sessions between client and server. An acceptable value for subscriber costs for the fully interactive system is heavily dependent on the economic model for incremental revenues that might be expected from the additional services available to a customer. To put it another way: Just how much is a customer willing to pay and for what types of services? To answer this question, a number of fully interactive trials are being conducted for consumer acceptance as well as technical evaluation.

Current projections for 1995–96 revenues that might be expected with various services using advanced analog HCTs, digital broadcast mode HCTs, and digital interactive HCTs are shown in [Table 1](#).

TABLE 1 Projected Revenues from Services over Various HCTs, 1995–96 (billions of dollars)

	Advanced Analog HCT	Digital Broadcast HCT	Digital Interactive HCT
Broadcast video	28	28	28
Pay per view/NVOD	8	14	16
Video on demand	—	—	4
Other services	4	8	28
TOTAL	40	50	76

SOURCE: Paul Kagan, 1994, Projections for 1995/96 Cable Revenues.

These projections indicate that a fully interactive HCT could provide almost a 100 percent increase in monthly revenues over an advanced analog unit. However, a large portion of the projected increase is in the vague category defined as "other services" that include home shopping and other services whose revenue potential is not proven.

Currently, a reasonable compromise is to provide advanced analog HCTs with NVOD capability for the near term, but add additional capacity using digital broadcast, and offer digital interactive services on a customer demand basis with digital interactive HCTs.

#### REQUIREMENT FOR INTEROPERABILITY AND INDUSTRY STANDARDS

The promise of emerging digital interactive services cannot reach critical mass without the elements of *interoperability*, *portability*, and *open standards* at critical interface points in the system. Because of the complexity of the systems involved, these elements are best resolved through industry groups representing major providers of equipment as opposed to government standards that could overlook critical issues.

The following are one set of definitions for the above terms:

- *Interoperability* is the ability to easily substitute or mix similar products from different vendors. Examples are television sets, VCRs, fax machines, and almost any electronic product sold at retail.
- *Equipment portability* is defined as consumers' ability to move their owned equipment across a town, state, or country and still be able to use the equipment as before in a new location. Examples are TVs, VCRs, radios, cellular telephones, direct-TV, and almost any public telecommunications equipment. Currently, CATV set-top converters do not fit this category.
- *Open standards* are standards that provide complete descriptions of interfaces at critical points in the system with reasonable, nondiscriminatory license fees if any intellectual property is involved. Examples are NTSC, PAL, DOS, MPEG, telecommunications interface protocols, and many others.

Interoperability, consumer equipment portability, and open standards are critical to the success of a fully interactive system for at least three reasons.

First, the risk is lowered for all participants. Equipment manufacturers are able to invest in the development of products that are guaranteed to be compatible with all networks and the suppliers of content on those networks.

Second, it assures the lowest cost hardware and content. Competition between equipment manufacturers to produce in large volume and therefore achieve economies of scale is enhanced because these products can serve the entire market base as opposed to many smaller proprietary networks. Also, content providers will be inclined to participate in content and applications development because these applications will be compatible with all consumer terminals.

Third, it will jump start the creation of many new interactive applications by removing the chicken-and-egg phenomenon. Once standards are agreed upon, content providers can begin working on applications software knowing that network architectures and consumer terminals will be available in the near future. Likewise, once attractive applications are available, consumers can acquire home terminals and subscribe to the network.

A major step toward the goals of achieving interoperability, portability, and industry standards has been accomplished by the MPEG committee in the adoption of standards for compressing, transporting, and decompressing video signals. As a result, chip manufacturers have developed a set of decompression chips that are functionally interchangeable, and equipment manufacturers now have a video standard to work from. Because of MPEG's success, a digital audiovisual council called DAVIC was formed in April 1994 to help resolve remaining compatibility issues. This committee currently has over 120 member companies worldwide. The charter for DAVIC is to assist in the success of the emerging digital video/audio applications and services by providing timely agreed-upon specifications for interfaces and protocols. DAVIC is working on a timetable that called for a strawman baseline document in December 1994, experiments and tests to occur during the spring and summer of 1995, and publication of final recommendations by December 1995.

While the goals of DAVIC are ambitious to say the least, much of the intent could be realized by recognizing that two critical interface points exist in the fully interactive network.

The first interface where standards are important is the connection between broadband integrated gateways at a cable headend and the data networks used to achieve full connectivity between cable systems and between cable systems and telephone companies. An interface standard must be adopted that allows SONET, FDDI, Ethernet, ATM, and other protocols to provide inputs to the access part of the network easily. This approach would provide the connectivity necessary to access remote content providers either over satellite, or from remote video file servers on the metropolitan network or connection through a trunk from other areas. Interactive applications providers with the capability to communicate over the Internet or get applications resident outside the local cable network can also be sources of content.

The second critical interface connects the broadband network to the home. If interoperability and portability are to be realized along with network integrity and flexibility, the home terminal must split into a part that *must* be provided by the network operator and a part that is not necessarily a part of the network and could evolve to becoming consumer electronics sold at retail.

A first part of the terminal that could be called the network interface module (NIM) must be provided by the network operator for several reasons:

- The technical characteristics of networks will evolve over time. This evolution will include network-unique modulation of digital streams that will start out using 64 QAM but migrate to 256 QAM or even VSB modulation, upgrades in network-specific signaling techniques for both forward and reverse path, and a reallocation of bandwidth used to transmit the digital channels.
- The HCT will need to interface with other types of delivery media including 28 GHz wireless, 2 GHz wireless, or even a satellite network.
- Networks will have different security and conditional access systems.
- Software needed to run the HCT and communicate with network control will be network specific.

The NIM would be provided by the network provider as a standard sized plug-in module with standardized I/O on the customer side and specialized interface to the network. It would contain the tuner, demodulators, conditional access system, and reverse path signaling common to a specific network and would

allow the many different network architectures, signaling structures, and connection management systems, many of which are already in place today, to communicate with a part of the HCT that one day might be purchased at retail. To add further flexibility and network integrity, the addressability and conditional access system (AC&E) could be installed on a "smart card" that plugs into the NIM.

This HCT concept can be explained by Figure 2, which shows the major functional blocks contained in an interactive digital HCT. These include the following:

- A network interface module (NIM) as described above to provide connectivity, conditional access, and control between the network and the HCT;
- A platform for executing multimedia applications, including an operating system (OS) and a published application program interface (API) and a user interface (UI) to process downloaded applications;
- A digital entertainment terminal (DET) to receive and process both MPEG compressed digital video and audio, as well as analog video for presentation to a video monitor or TV set (it would also contain all the necessary home entertainment interfaces for TVs, stereos, PCs, and other consumer electronics); and
- Software that is network-specific to interface with the DET and the NIM for constructing reverse path messages and interpreting addressable instructions sent downstream to the terminal.

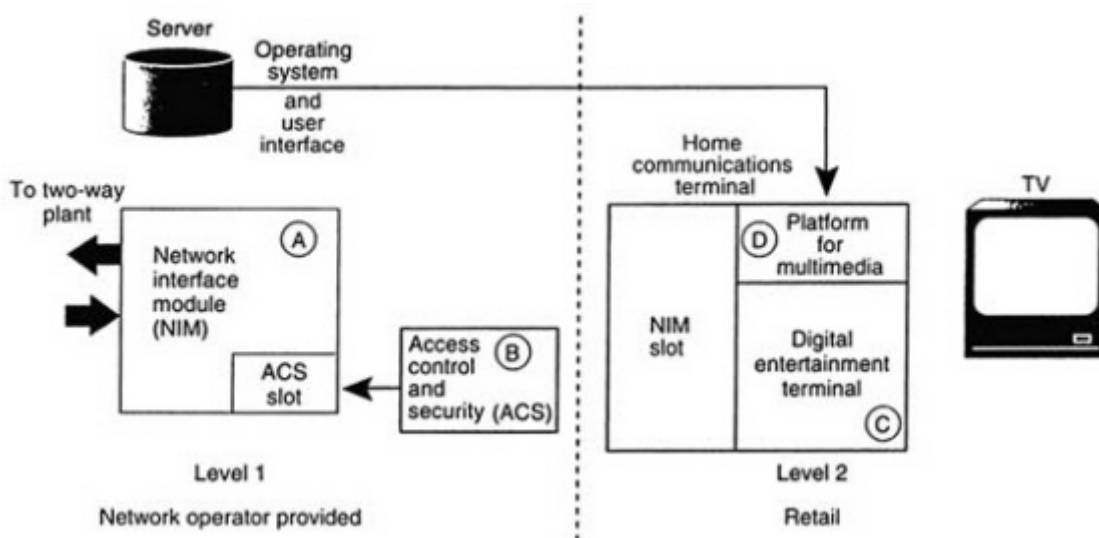


Figure 2 Proposed two-part broadband digital home communications terminal architecture to resolve interoperability and retail sales issues.

### CONCLUDING REMARKS

The following statements are supported by the previous sections:

- Today's advanced analog HCTs include digital communications that provide controlled access, downloadable digital applications (e.g., electronic program guides and virtual channels), and a significant revenue-producing service with analog video and NVOD.
- Digital interactive HCTs have processing power equivalent to PCs and the additional ability to communicate over broadband networks and provide multimedia applications.
- Digital HCTs are already available for trials in digital networks. Their wide-scale deployment will be made possible by using highly integrated silicon to reduce cost of the HCT.
- For an economic transition from analog to digital that does not preclude revenue-producing analog video, digital HCTs will also incorporate analog as well as digital capability.

- Standards will allow the market to develop more rapidly and will help drive the cost of the fully interactive HCT down faster.
- NIMS and DETS allow networks to be deployed that provide interoperability and portability for the network provider and the customer without inhibiting technology and service growth.
- HFC provides the plant configuration that allows a smooth transition to the future and can be the platform for a fully interactive network that includes video, audio, telephony, and data without sacrificing analog services and associated revenues.

#### **REFERENCE**

- [1] Goodale, James C., 1989, "All About Cable," Law Journal Seminars-Press, Chapter 1, p. 6.



## 21

# Spread ALOHA Wireless Multiple Access: The Low-Cost Way for Ubiquitous, Tetherless Access to the Information Infrastructure

*Dennis W. Elliott and Norman Abramson*  
*ALOHA Networks Inc.*

### STATEMENT OF THE PROBLEM

The ability to access the national information infrastructure (NII) on a tetherless, broadband basis is often discussed. But today, the capability to interchange information with and through the network on a tetherless basis is limited and expensive. The promises of new tetherless access approaches for data, such as the current personal communication system (PCS) implementation approaches, have been only that—promises. But a new wireless multiple-access approach will change this situation in a revolutionary way. This new wireless multiple-access approach is called Spread ALOHA and is being developed under U.S. Small Business Innovative Research grants from the National Science Foundation and from the Advanced Research Projects Agency of the Department of Defense.

Envision a robust nationwide packet radio data network in place with millions of users having tetherless, broadband user communications devices allowing easy, low-cost automatic interface into the public and private networks of the national information infrastructure—anywhere, anytime. A user with a portable PC, a personal digital assistant (PDA), or another device containing a Spread ALOHA PCMCIA card or embedded chip would have instant access to a network of choice.

This paper describes the market need for a wireless multiple-access approach that offers robust, wireless multiple access to the NII at an affordable price. A strategic plan for implementation of Spread ALOHA architecture having an increase of two orders of magnitude in capability over existing wireless data multiple access approaches is discussed.

### BACKGROUND

#### The Market: What People Want and Need

As the NII evolves, users will increasingly want and demand ready access to acquire information, send information, or communicate with each other. Some of this information may be manually or automatically generated or requested by a user or by an application or stored in a database for manual or automatic request.

Access to the NII will either be by wire (telephone, cable TV, etc.) or wireless. The wireless case for millions of users is the focus of this paper.

Wired data access to the NII and predecessors with capabilities adequate for certain applications and with distinct evolutionary paths or alternatives exists today. But wired data access has a fundamental limitation. It is not always where the users are, and it does not travel easily with the users. One has to "plug into another jack" whenever one moves. The wire is a tether. Thus, wireless data access can prove a boon to those users who want and need information access wherever they are. The ability to operate on a tetherless basis generates new power to use information for almost everyone. (Those who have experienced a good, tetherless computing or data access situation can testify to this.)

But wireless data access capability to interchange information with and through the network on a tetherless basis is limited and expensive today. To date, new wideband tetherless access approaches for data appear to be "vaporware"—many have tried and none have succeeded. There are fundamental technology limitations in wireless multiple access that have led to economic limitations of these dreams.

### **The Market: Quantization**

Characterizing a market in which one introduces a product of at least two orders of magnitude more capability than currently exists is difficult at best. The following discussion forms a probable baseline that would be the lower bound for the market addressed by Spread ALOHA architecture.

The addressed market for nomadic, tetherless computer networking employing Spread ALOHA architecture consists of a large fraction of the users of portable computing devices, such as portable PCs and PDAs. The reality is, however, that with the introduction of a breakthrough technology such as Spread ALOHA, a new emphasis on new devices and applications comes into play, which tends to stimulate and transform the market. ALOHA Networks expects that the PC/PDA market will be only the base for the nomadic, tetherless computing network market. New applications using a PCMCIA card (beyond PC and PDA applications) and embedded Spread ALOHA wireless technology will develop as well.

Even recent forecasts for growth of the mobile data market can now be revised upward:

The mobile data market in the United States will increase from 500,000 users in 1994 to 9.9 million users in 2000—with a compound annual growth of 64 percent, according to a new report from BIS Strategic Decisions.... The "mobile professional" sector—professional employees who spend 20 percent or more of their time away from their desks—represent a potential user population of 31.9 million by 2000....<sup>1</sup>

The rebirth of the PDA market will hinge on re-positioning the devices as communications-oriented PC companions.... BIS predicts portable PC users will increasingly communicate using mobile data services. We forecast that 2.8 percent of portable PCs that are wirelessly-enabled (300,000 units in 1994) will grow to 16 percent of the installed base, or 2.6 million units, in 1998. The increased usage of mobile data will be a direct result of improvements in availability, functionality, and pricing for services. We also expect that the mobile data market will grow at a compound rate of 80 percent through 1998. Although at most 12 percent of PDAs are currently wirelessly enabled, that percentage will grow to 75 percent by 1998 [1.7 million units from authors' graph]. The current low percentage rate reflects the death of low-cost mobile data alternatives. Once users have more cost-effective options to choose from, the number of wirelessly-enabled PDAs will climb.<sup>2</sup>

From the above forecast made without knowledge of the breakthrough Spread ALOHA technology, it can be assumed that the market for user units (as opposed to infrastructure) will likely be about 4.3 million units in 1998 and will approach 10 million units by 2000.

### Existing and Near-Term Marketplace Alternatives

There are several alternative wireless data service approaches available or nearly available today. Broadly speaking, the services can be broken into three general categories: one-way, packet-based services (primarily paging); two-way, packet-based services; and circuit-based services. Packet-based services chop the information that is to be sent into data packets, attach addresses identifying the transmitter and the recipient if needed, and send the packet over a channel that is shared by multiple users. Circuit-based services allocate a specific transmission channel to the end user when a transmission is requested, and that user holds that circuit until the transmission is completed.

The most common form of circuit-transmitted data utilizes cellular modems to provide mobile data communication over the existing analog cellular infrastructure. By incorporating sophisticated error correction protocols, these modems attempt to compensate for the relatively low line quality and allow portable computer or fax machines to communicate on a wireless basis. This approach benefits from the existing broad coverage of analog cellular and, except for the continuing problem of broken connections, is effective for large file transfers. But the quality of service has been generally low, and the price to the user has been high because of long set-up times, low data rates, and high cellular airtime pricing.

Wide area applications such as the ARDIS and RAM Mobile Data services are struggling. (According to a *New York Times* article,<sup>3</sup> ARDIS reportedly has 38,000 users in 10,700 towns and cities and RAM Mobile Data has 17,000 users in 7,000 towns and cities). These services, which are employed where customer data must be captured regionally or nationally, do not live up to their initial promise because they are narrowband and therefore fundamentally expensive and limited in data rate.

Cellular digital packet data (CDPD) services by AT&T McCaw, GTE, and others, which use the analog cellular infrastructure in the United States, are moving toward implementation in individual cellular carriers' regions. CDPD transmits packet data in the limited voice analog cellular infrastructure. Though it employs a packet structure good for many data applications, there are two limitations: (1) access to the infrastructure requires much of the cumbersome multiple access required for voice uses and (2) data rates are limited to what fits into the analog voice channels of that infrastructure, which are less than ideal.

Metricom, Geotek, and Nextel frequency hop/time division and time division, multiple-access—based services and new two-way paging services, such as MTEL's Destineer, are also being deployed. Metricom offers a narrowband approach of uncertain throughput today, and its peer-to-peer architecture limits the data volumes its network can handle—not a desirable trait for ubiquity. None of these systems can provide tetherless data access that is wideband, robust, user unlimited, and inexpensive.

### The Technology

Though broadcast transmission (the transmission of significant amounts of data from one to many) is a well-understood problem, the transmission of significant amounts of data from many to one (i.e., the network node) can be quite difficult. The new wideband Spread ALOHA wireless multiple-access technology, contained in forthcoming chips and boards for manufacturers of nomadic, tetherless computing products, allows robust, "bursty," tetherless access to the NII at an extremely low cost. The implementation of Spread ALOHA is simpler and far less expensive than any of the various PCS or wide area data network approaches defined to date.

Spread ALOHA is an advanced wireless multiple access technology that can provide the capabilities required for digital networks with large numbers of remote terminals. Spread ALOHA combines the proven simplicity and operational flexibility of conventional narrowband ALOHA wireless multiple access with the high bandwidth and high throughput of spread spectrum. (Conventional ALOHA multiple access was developed at the University of Hawaii and is employed in RAM Mobile Data's network, in the Inmarsat maritime satellite system, and in many VSAT networks, and it is the underlying basis for Ethernet.)

Spread ALOHA compares favorably to time division multiple-access (TDMA) and frequency division multiple-access (FDMA) approaches because its capacity is data limited, not channel limited as are TDMA and

TABLE 1 Comparison of Spread ALOHA to Two Other Access Technologies

	Conventional ALOHA	Spread Spectrum CDMA	Spread ALOHA
Bandwidth	Low	High	High
Number of users	Unlimited	Low	Unlimited
Complexity	Low	Very high	Low
Flexibility	High	Low	High

FDMA. That is, capacity in the Spread ALOHA architecture is limited only by the amount of data transmitted in the network rather than by the number of users who access the network. Table 1 compares Spread ALOHA to the other data limited approaches, conventional ALOHA and Spread Spectrum code division multiple access(CDMA). It illustrates why Spread ALOHA is the most efficient multiple-access technique for large numbers of users.

The number of users in either conventional ALOHA or Spread ALOHA is limited only by the total data rate of the channel. Conventional ALOHA is low bandwidth and thus has a low data rate because of the practical requirements of maintaining a constant pulse energy as the data rate increases. In CDMA the *practical limit* on the number of users is cell constrained by the requirement to implement a separate receiver at the hub station for each active user. In the IS-95 CDMA standard for cellular voice, the maximum number of users per cell is less than 40.

Spread ALOHA can be viewed as a version of CDMA that uses a single, common code for all remote transmitters in the multiple-access channel. In a Spread ALOHA channel, different users are separated by a random timing mechanism as in a conventional ALOHA channel rather than by different codes. Since only a single code is used in a Spread ALOHA channel, only a single receiver is required in the Spread ALOHA hub station, rather than a separate receiver for each remote terminal as is required in CDMA. In addition, because of the elimination of multiple codes, many of the most complicated features required in a CDMA receiver can be removed. The elimination of unnecessary system complexity makes possible a degree of system flexibility, which can be important in today's rapidly evolving wireless markets.

For example, a Spread ALOHA hub station need only be capable of synchronizing to received signals, all of which use the same code, a much simpler problem than that faced by a CDMA hub, where the codes received are all different. In a Spread ALOHA hub station, packet interference can be eliminated by a cancellation process made practical by the fact that the interference terms generated by all packets are identical. And in a Spread ALOHA channel it is possible to select a spreading code that has only half as much interference as codes used in a CDMA channel.

Spread ALOHA can provide a throughput 100 times greater than any conventional ALOHA network now in operation and is much easier to implement than current wideband multiple-access techniques, such as CDMA. No other technology can provide robust data networking with large numbers of users. Spread ALOHA combines the proven simplicity and operational flexibility of a conventional ALOHA multiple-access channel with the high bandwidth and high throughput of a spread spectrum channel.

## ANALYSIS AND FORECAST

### General

Spread ALOHA will make it possible to build a nationwide broadband packet-radio data network allowing easy, low-cost, automatic interface into the public and private networks of the national information infrastructure. A user with a portable PC, a PDA, or another device containing a PCMCIA card or embedded chip would have instant access to the network of choice. Smaller campus, metropolitan, or regional networks could be addressed initially as a way of beginning what could ultimately be a national network.

Spread ALOHA technology also holds great promise for data/voice PCS applications. Because of its lower cost, Spread ALOHA offers a potential for ubiquity that does not exist with other approaches for supporting data applications within PCS. However, the PCS market cannot be easily approached without an

established standard for Spread ALOHA PCS applications for voice and data. The establishment of a standard is expected to take some time, but ALOHA Networks, in cooperation with others, plans to begin these efforts in 1995, looking toward the second phase of PCS standards and equipment in 1998–99. At the bottom line, however, if a nationwide data network can be implemented before the acceptance of a standard in this area, Spread ALOHA will be positioned as a de facto standard. ALOHA Networks believes that its Spread ALOHA technology will ultimately form the basis for the most viable PCS data standard.

### **Economic Advantages**

The economics of using the Spread ALOHA technology are compelling. Spread ALOHA technology allows for a simple implementation and has no user population limits. This simple implementation can lead to low-cost user units. In addition, the cost of implementing the hub or microcell facilities is lower for a large number of users than any alternative.

ARDIS, RAM Mobile Data, and GTE's CDPD services appear to price their services at about \$0.50–\$1.00 per kilobyte.<sup>4</sup> This is essentially a user cost of \$0.50 per second of use! AT&T McCaw Cellular has announced CDPD service prices ranging from \$0.08 to \$0.16 per kilobyte.<sup>5</sup> The typical average monthly bill for these services has been estimated to range from \$25 to \$200.

ALOHA Networks estimates that Spread ALOHA multiple access can substantially increase network capacity as well as individual "burst" transmission rates without significant added cost over other alternatives. The existing and planned data networks tend to be constrained to an operating rate of about 20 to 50 kilobits per second. With large network volumes, price per kilobyte could be reduced by one to two orders of magnitude. As a corollary, Spread ALOHA could allow a pricing of *\$0.50 to \$1.00 per 100 kilobytes* if one assumes demand is stimulated by such a substantial price decrease.

ALOHA Networks anticipates that by 2000 users of a high proportion of notebook PCs, PDAs, and other embedded microprocessors in portable platforms will expect to be able to communicate with remote points, public networks such as the Internet, private networks, or with the user's office. In fact, with a low-cost tetherless approach, the ALOHA Networks' Spread ALOHA technology could significantly stimulate the market for notebook PCs, PDAs, and other devices not yet conceived! With the anticipated growth of nomadic computing, this would translate into 10 million to 100 million user units (either PCMCIA cards or embedded) in 2000.

### **System Issues**

The nomadic, tetherless computing network is envisioned as a nationwide system, accessible from almost anywhere in the United States. This network could either be integrated with other networks or interfaced to other networks at various nodes. The hub or microcell stations would be spaced according to propagation characteristics in every area, similar to PCS or perhaps ARDIS or RAM Mobile Data. (In fact, the existing infrastructure of these networks could be employed for this.)

### **Frequency Spectrum**

Spread ALOHA can operate in almost any frequency band. However, the system envisioned here would operate in a given, yet to be determined frequency band. The possibilities are (1) allocation of a new frequency band for this service, (2) use of the existing ISM bands, (3) use of the existing ESMR bands, or (4) use of the existing and future PCS bands. Since the radio frequency transmission is spread spectrum, the selected band must, of course, be suitable for such uses. The frequency band approach, which will allow the fastest implementation but yet allow the expected ubiquitous growth, should be explored.

## **Network Infrastructure**

The network infrastructure for a nationwide nomadic, tetherless computing network would entail microcell sites with the Spread ALOHA hubs. These hub sites would be trunked together using existing telecommunications company and interexchange carrier facilities. At selected hub sites, the network would be interfaced with the Internet and other Internet-like networks.

To establish ubiquitous coverage, from 2,500 to 12,000 hub sites will have to be established, depending upon the selected frequency band and coverage patterns. A Spread ALOHA installed infrastructure of this size is estimated to cost between \$100 million and \$500 million. The exact cost depends on what the coverage pattern is, whether the data network overlays another network, and whether a voice network is implemented simultaneously. Such a network can be implemented on a phased basis, covering the highest user population areas first.

## **Remote User Terminal Communications Devices**

ALOHA Networks envisions remote user terminal communications devices that are small and inexpensive. Assuming large quantities of devices, ALOHA Networks estimates the cost of the Spread ALOHA chip set or chip for remote user communications cards in a microcellular system to be substantially below \$100 in the 1998 time frame, with the normal "Moore's Law" cost reductions beyond 1998. User software in the terminal device would employ "standard" user software such as the General Magic or other user operating software products. Ultimately, ALOHA Networks envisions these devices as being embedded in many different computing appliances, with the communications and microprocessor elements not particularly discernible to the user.

## **Strategic Relationships**

To implement such a concept, the existing and potential wireless infrastructure owner/operators must be involved in the evolution of the system together with wireless networking equipment manufacturers. These parties must have a reasonably common objective.

## **RECOMMENDATIONS**

### **Forum for Development of Wireless Infrastructure**

Establish a forum for those private- and public-sector entities involved in infrastructure for wireless data. Encourage analysis of the Spread ALOHA architecture and the establishment of strategic relationships among the parties, assuming that the effectiveness of that architecture is demonstrated. In conjunction with existing private-sector infrastructure providers, develop an implementation approach to overlay a Spread ALOHA architecture on existing wireless networks and determine the most appropriate frequency allocation. Coordinate and make appropriate filings with the Federal Communications Commission for the selected frequency use. This infrastructure implementation will be the critical factor in realizing such a nomadic, tetherless computing network.

### **PCS Data Standards**

Encourage standardization proceedings to be initiated for a Spread ALOHA wireless air interface for PCS data and voice/data applications. The lowest-cost user PCMCIA card approach would require a hub infrastructure similar in coverage and spacing to voice PCS. Though it is recognized that the initial PCS implementations will be oriented toward telephony, the second implementation should be more attentive to data, thus offering a good opportunity for a broadband approach—Spread ALOHA.

### **Integrate Tetherless Approach with NII Planning**

The concept of tetherless data access to the NII should be integrated into other NII studies and planning. Most such studies and planning envision a person sitting at a desk. The tetherless concept is an important aspect of making full use of the NII in life beyond relatively static libraries, schools, and offices.

### **ADDITIONAL RESOURCES**

Abramson, Norman (editor). 1993. *Multiple Access Communications: Foundations for Emerging Technologies*, IEEE Press, New York.  
Abramson, Norman. 1994. "Multiple Access in Wireless Digital Networks," invited paper, *Proceedings of the IEEE*, September.

### **NOTES**

1. *Mobile Data Report*. 1995. "BIS Estimates U.S. Market to Reach 9.9 Million Users in 2000," Vol. 7, No. 8, April 24.
2. Nelson, Paul, and Dan Merriman, BIS Strategic Decisions. 1994. "Wirelessly Enabling Portable Computers: A Major Growth Opportunity," *The Red Herring*, September/October, pp. 64–65.
3. Flynn, Laurie. 1994. "The Executive Computer: 3 Ways to Be Unplugged Right Now," *New York Times*, December 4.
4. Leibowitz, Dennis, Eric Buck, Timothy Weller, and John Whittier. 1995. *The Wireless Communications Industry*. Donaldson, Lufkin & Jenrette, New York, Winter 1994–95, p. 34.
5. *Mobile Data Report*. 1995. "McCaw Prices CDPD as Low as 8 Cents/K to Cover 75 Percent of its Markets in 1995," Vol. 7, No. 8, April 24.

## 22

# Plans for Ubiquitous Broadband Access to the National Information Infrastructure in the Ameritech Region

*Joel S. Engel*  
*Ameritech*

### STATEMENT OF THE PROBLEM

The national information infrastructure (NII) can be visualized as consisting of four components: (1) access transport from the premises of the end user to an access node; (2) access nodes on an intracity, intercity, and international network of networks; (3) the internodal network of networks, including the internetwork gateways; and (4) information repositories. The access nodes, network, and information repositories—the last three components—are common facilities; their costs are predominantly proportional to usage and can be shared among many users. In contrast, the access transport—the first component—is a dedicated facility connecting the premises of a single user; its cost is predominantly fixed, independent of usage, and borne by that single user.

As a result, the access transport—popularly referred to as the "last mile," although, in fact, it is typically a few miles in length—presents the greatest challenge to the provision of a ubiquitous, truly "national," wideband NII. Because the access nodes and information repositories can be provided modularly and grow with usage, they are already being put in place, by small entrepreneurs as well as by large service providers. Since the intracity, intercity, and international traffic can be multiplexed into the traffic stream on the existing telecommunications network, that portion of the "information superhighway" already exists. The major investment hurdle, then, is the deployment of ubiquitous broadband access to individual user premises.

At first glance, it might appear that cable television systems constitute such broadband access, but closer analysis reveals that this is not so. Cable systems are predominantly one-way downstream, some with minor upstream capability, usually relying on polling to avoid interference among upstream signals. More constraining, they are broadcast systems, with 50 or so video channels being simultaneously viewed by several thousand users. The same cost hurdle described above exists to upgrade these systems to provide dedicated broadband access to individual users.

Currently available technologies for access transport span a wide range of capabilities and costs. Since the costs cannot be shared among users, at the low end of the range they are shared among usages, particularly with voice telephony. Modems provide dial-up capability over standard telephone lines at speeds typically up to 9,600 bits per second, with less common capability at 14,400 and 28,800 bits per second. Integrated services digital network (ISDN) is rapidly becoming widely available, providing dial-up capability at 56, 64, and 128 kilobits per second. Various local carriers, including the local exchange carriers, are offering frame relay service at 56 kilobits per second and 1.5 megabits per second, and switched multimegabit digital service at 1.5 and 45 megabits per second, with plans for migration to asynchronous transfer mode (ATM) at OC-1 (45 megabits per second) and higher SONET rates. However, all of these require a dedicated leased digital line between the user's premises and the carrier's switch of a capacity equal to the peak bit rate used. As a result, these higher-speed services are cost effective only for large user locations that generate sufficient traffic to fill them. The challenge, again, is to provide cost-effective, high-speed access to individual homes and small business locations.



## DEPLOYMENT PLANS FOR THE AMERITECH REGION

Ameritech has committed to providing broadband access in its region, using a hybrid optical fiber and coaxial cable architecture that can support a wide range of applications, including video services similar to current cable television, expanded digital multicast services, interactive multimedia services, and high-speed data services. By providing a platform for a wide range of applications, a cost-effective solution is achieved (economic data are provided in a later section).

Construction is expected to begin in 1995 in the Chicago, Cleveland, Columbus, Detroit, Indianapolis, and Milwaukee metropolitan areas, ramping up to an installation rate of 1 million lines per year by the end of 1995 and continuing at that rate to deploy 6 million lines throughout the region by the end of the decade. In December 1994, the Federal Communications Commission granted approval for Ameritech to construct the first 1.256 million lines, distributed among the metropolitan areas as follows:

- Chicago—501,000,
- Cleveland—137,000,
- Columbus—125,000,
- Detroit—232,000,
- Indianapolis—115,000, and
- Milwaukee—146,000.

## TECHNICAL ARCHITECTURE

The system architecture employs a hybrid transport network of optical fiber and coaxial cable. Signals are delivered over Ameritech's ATM network to video serving offices, each serving 100,000 to 150,000 customer locations. The signals are then distributed on optical fiber to individual nodes, each serving a total of 500 customer locations, not all of whom may actually subscribe to the service. From each node, the signals are distributed on four parallel coaxial cable systems, each serving 125 customer locations. With this architecture, the coaxial cable network is less than 2,000 feet in length and contains, at most, three amplifiers to any customer location.

The signal on both the optical fiber and the coaxial cable is a broadband analog video signal. The initial deployment will have a bandwidth of 750 megahertz, with capability for upgrade to 1 gigahertz when the reliability of such electronics becomes proven, yielding 110 channels of standard 6 megahertz video bandwidth. The allocation of these 110 channels to various applications is flexible and will be adjusted to satisfy user needs. Based on current estimates, approximately 70 of the channels will carry analog video signals for applications similar to current cable television, including basic and premium channels and pay-per-view. The remaining, approximately 40, of the channels will be digitized using 256 quadrature amplitude modulation, yielding a usable bit rate of over 36 megabits per second on each channel. Approximately 30 of these digitized channels will be used for multicast services, with multiple users viewing each transmitted program. Approximately 10 of the digitized channels will be used for switched interactive services, for which each user requires a dedicated digital circuit for the duration of the session.

On the digitized channels, the video signals will be compressed using the MPEG-2 compression standard. Depending on the particular application, each such signal will require a fraction of the 36-megabit-per-second or greater capacity. The signals will be multiplexed at the video serving offices and demultiplexed by the customer premises equipment, using the MPEG-2 transport layer protocol.

In addition to the downstream capacity, the system will have an upstream capability provided by up to 20 channels, each of 1.5-megabit-per-second capacity. Depending on local conditions of noise and interference, it is expected that at least 15 of these will be usable on each coaxial cable system serving 125 customer locations.

The system is intended to be a platform for a wide range of applications. Accordingly, the customer premises equipment may be a set-top box for use with a television set or an adjunct to a personal computer.

## SYSTEM CAPACITY

As described above, each branch of the distribution network will consist of an optical fiber to a node serving 500 customer locations, with four coaxial cable buses from the node, each serving 125 customer locations. Each branch will have a downstream capacity of approximately 360 megabits per second for switched interactive services.

The system is designed for economical growth in capacity as the market expands and traffic increases. Initially, each fiber branch will share a linear laser in the video serving office with two other fibers, through a passive optical splitter, so that the 360-megabits-per-second downstream capacity will serve 1,500 customer locations, not all of whom may subscribe to the service. When the traffic grows to exceed this capacity, the fibers can be driven by independent lasers so that the 360 megabits per second downstream capacity will serve 500 customer locations. When the traffic requires it, up to four fibers can feed each node, one for each coaxial cable bus, so that the 360 megabit per second downstream capacity can serve 125 customer locations.

The downstream capacity will be assigned to the users on a per-session basis, depending on the particular application. There is some uncertainty about the bit-rate required for each application. Human factors experiments and customer trials appear to indicate that live programs other than athletic events, compressed in real time and displayed on full-size television screens, can be delivered at 3 megabits per second. Material that is prerecorded and stored in compressed digital form can be delivered at a lower bit-rate, since the compression is not performed in real time. During the compression process, the results can be observed, and the parameters of the compression algorithm can be optimized to the particular material. Similarly, video material that accompanies text and occupies only a portion of a computer screen requires less resolution and can be delivered at a lower bit-rate. Not all interactive applications will involve video, and those that do will generally involve material that is stored in compressed digital form and displayed in a window occupying a portion of the screen, so that an average of 3 megabits per second per session is probably a conservative estimate. At that average bit-rate, the 360-megabits-per-second downstream capacity could serve 120 simultaneous sessions.

Further, although the initial system will assign a fixed bit-rate for the entire duration of the session, equal to the bit-rate required for the most demanding segment, the capability will exist for statistically multiplexing the signals and using a bit-rate based on instantaneous requirement. This capability will be employed if necessary. In that event, an average bit-rate per session of 3 megabits per second would be quite conservative.

It is estimated that, during the peak usage period of the day, 15 percent of the subscribers will be using the service at the same time. That would consume the entire 120-session capacity when 800 of the 1,500 customer locations became subscribers, equal to 53 percent market penetration. At that point, the next step in capacity growth, sharing the capacity among 500 customer locations, would be implemented. If the peak usage per subscriber turned out to be higher than the 15 percent estimated, then the capacity growth would be implemented at a lower number of users. This would not affect the economic viability of the system, since it is the total amount of usage that will generate both the need and the revenue to support the increased capacity.

At the limit of the current architecture, the 120-session capacity will be shared by the 125 customer locations on each coaxial cable bus. That would support 96 percent simultaneous usage at 100 percent market penetration. If that turned out to be insufficient, because of multiple simultaneous users per customer location, or higher bit-rate applications, statistical multiplexing could be employed.

In addition to the downstream capability, each coaxial cable will support at least 15 usable upstream channels at 1.5 megabits per second each, and these will be multiplexed onto the fiber from the node to the video serving office. Therefore, unlike the downstream capacity, this upstream capacity will be shared by 125 customer locations from the start. These upstream channels must utilize a time division multiple access protocol, which does not allow for 100 percent "fill"; nevertheless, traffic studies indicate that a single 1.5-megabit-per-second upstream channel could easily support all of the digital video requirements, including video on demand, of the 125 customer locations, leaving the remaining 14 or more for the more interactive multimedia and data applications.

There are two parameters of the interactive services that determine the required upstream capacity: (1) the frequency (and size) of the upstream messages generated by the user and (2) the required latency, or speed of response to the message. Analyses of the types of applications that are anticipated indicate that the second

parameter is controlling. If enough upstream capacity is provided to assure that the transport delay contributes no more than 100 milliseconds, it does not matter how frequently the user generates upstream commands.

Of course, if the users were to generate large files for transmission in the upstream direction, rather than the inquiry and command messages typical of interactive services, this analysis would not apply. But, by definition, such users would be generating sufficient traffic to use one of Ameritech's other, more symmetric, data offerings cost effectively, and would not be using the hybrid fiber/coaxial cable system.

At a 100-millisecond latency, each upstream channel could support approximately 20 simultaneous sessions, for a total of at least 280 simultaneous sessions on the 14 or more upstream channels. This is greater than the 125 simultaneous sessions supportable by the downstream capacity that will only be generated by 96 percent simultaneous usage at 100 percent market penetration.

### ECONOMIC ANALYSIS

At this time, the FCC has approved construction of the first 1.256 million lines, and the economics for that initial phase, which is extracted from Ameritech's application for construction approval, is presented below.

Table 1 presents the economic data for the first 10 years. It is important to note that these data are strictly for the transport network; they do not include any costs or revenues for the provision of "content."

Construction of the first 1.256 million lines is planned for completion early in the third year. Both market penetration and usage per customer are expected to grow throughout the period, as shown by the revenue forecast. These revenues are for the transport of all types of content, including broadcast television, video on demand, interactive multimedia services, and high-speed data access. The total costs each year consist of three components: (1) the costs of constructing additional lines, which ends in the third year; (2) the costs of adding equipment for additional customers and additional usage per customer; and (3) the costs of providing customer service.

The cash flow each year, revenues minus costs, is discounted to Year 1 equivalent values and cumulated, and is presented in Table 1 as cumulative discounted cash flow (CDCF). As shown, the CDCF turns positive in the eighth year. In actuality, as additional lines are constructed in each of the six major metropolitan areas, building on the initial base, economies of scale are expected to make the economics even more favorable.

TABLE 1 Economic Data for Initial Phase of Ameritech's Deployment of Broadband Access Transport Infrastructure

Year	Customer locations passed (millions)	Revenues (\$millions)	Costs (\$millions)	Discounted cumulative cash flow (\$millions)
1	0.5	0.963	131.2	(130.2)
2	1.058	47.9	162.5	(233.8)
3	1.256	84.2	85.8	(235.2)
4	1.256	115.5	65.8	(198.4)
5	1.256	137.4	75.4	(157.0)
6	1.256	154.0	71.1	(106.9)
7	1.256	162.9	65.1	(53.5)
8	1.256	176.3	66.4	0.8
9	1.256	191.1	68.3	55.6
10	1.256	200.2	48.1	117.0

### CONCLUSIONS

The greatest challenge to the realization of a ubiquitous wideband national information infrastructure, an NII that is truly "national," is the provision of economical broadband access transport to individual residences and small businesses—popularly referred to as the "last mile." Access transport to large business locations can be based on usage, because such locations, as well as the nodes and internodal network and the information repositories, are all shared facilities and are used (and paid for) by multiple users. By contrast, access transport

to individual customer locations must be sized for the peak requirements of the most demanding application, even though that capacity is unused most of the time, and its cost must be borne by that individual customer.

Although these facilities cannot be shared by multiple customers, they can be shared among multiple services to the individual customer. Ameritech has committed to providing broadband access in its region, utilizing a hybrid optical fiber and coaxial cable architecture, supporting a wide range of applications. These include video services similar to current cable television, expanded digital multicast services, and interactive multimedia services, as well as high speed data services. By providing a platform for a wide range of services, a cost-effective solution is achieved.

Ameritech plans to begin construction in 1995, ramping up to a rate of 1 million lines per year by year end and continuing at that rate to deploy 6 million lines throughout the region by the end of the decade. The FCC has granted approval for construction of the first 1.256 million lines, in the Chicago, Cleveland, Columbus, Detroit, Indianapolis, and Milwaukee metropolitan areas.

## 23

# How Do Traditional Legal, Commercial, Social, and Political Structures, When Confronted with a New Service, React and Interact?

*Maria Farnon*

*Fletcher School of Law and Diplomacy*

*Tufts University*

### STATEMENT OF THE PROBLEM

It is somewhat ironic that the Internet, which was originally pioneered for national security reasons in the United States, has evolved into an international network. In fact, it is the open, international character of the Internet that is largely responsible for its growth. Its explosive expansion has been exported from North America to Europe and Asia, in particular. At the same time, the Internet poses serious challenges to many traditional structures at many levels: regulatory, commercial, legal, and cultural.

Many governments have enthusiastically embraced the introduction of the Internet without understanding its subversive character. As new services like the Internet travel to countries where government intervention in the market is strong, and where the telecommunications industry remains a state-owned monopoly, these countries will seek to define the Internet in traditional terms. The determination to construct a national information infrastructure (NII) by imposing standards, reserving service transmission for a monopoly provider, and controlling content reflects a complete misunderstanding of the Internet model.

This paper evaluates the development of the international Internet as a case study of how the evolution and diffusion of a new service are affected by preexisting factors. In other words, how do traditional legal, commercial, social, and political structures, when confronted with a new service, react and interact? Even if the Internet represents an entirely new paradigm of services technology, it seems clear that domestic governments and industries will seek to apply a familiar framework, and thus exert the same control and derive the same benefits that the current telecommunications system allows. After reviewing the historical structure of the telecommunications industry, this paper evaluates the introduction of networking in several geographic regions to determine how a new service affects the traditional telecommunications model. The question then becomes, What can be done to preserve the originality of the Internet as a decentralized, scalable network on a global basis?

### BACKGROUND

#### **The Old Paradigm: Government Ownership and Monopoly Service**

Telecommunications operators (TOs) have traditionally been structured as state-owned monopolies for several reasons:

1. *Natural Monopoly*. Telephone service has long been cited as an example of a "natural monopoly," characterized by significant ex ante investment (initial fixed costs) and increasing returns to scale. These factors imply that the size of the firm must be large to capture economies of scale and to recoup the initial investment.

2. *Public Good.* Government ownership is explained by the existence of "public goods," where the marginal cost of providing this good to an additional user is zero, and where it is difficult or impossible to limit consumption of the good. Thus, a commercial firm has no incentive to produce this type of product. An example of a "pure" public good is national defense. Telephone service, once the network is in place, might also be considered a public good.
3. *Incomplete Markets.* Incomplete markets occur whenever markets fail to produce a good or service, even where the cost is less than what people are willing to pay. Telephone service to rural areas or even to low-volume residential users could be considered an incomplete market, because firms may not generate enough business to recoup their investment. Thus, governments have justified their monopoly of telephone services as a welfare obligation, with the ultimate goal of providing access to every citizen, or "universal service."
4. *National Security.* Keeping the communications infrastructure out of foreign ownership and in the control of the government was long considered vital for protecting the national security, especially in times of crisis or war.

Though it is overly simplistic to suggest that state ownership is inherently less efficient than private ownership, because of the large size of state-owned enterprises (SOEs), they tend to attract stakeholders who depend on the actions of the firm for benefits.<sup>1</sup> The TO, for example, is often the largest employer in a country. Stakeholders include employees, suppliers, and the relevant government bureaucracies. Because the benefits they derive are significant, stakeholders have an incentive to organize in order to pressure the firm, and through this organization gain their power. Even though shareholders—namely the public—are the legal owners of the firm, they remain too dispersed to organize effectively. The stakeholders thus assume de facto control of residual rights. Governments are willing to maintain inefficient SOEs for fear of antagonizing powerful stakeholders, who are also political constituents.

TOs in many countries across Europe, Asia, Africa, and Latin America have clearly attracted strong groups of stakeholders, including labor unions, equipment suppliers, and the government agencies that direct the TO. The employees are concerned foremost with retaining their employment, while the domestic equipment suppliers often depend on the national TO for the bulk of their orders. The third stakeholder, the government, also has an interest in preserving the TO under state ownership. By setting high tariffs for long-distance use, and subsidizing shorter domestic calls, politicians garner public support. Furthermore, a profitable TO can be used to cross-subsidize other operations, such as the postal service. In essence, the policy of high tariffs for phone usage constitutes an indirect tax.<sup>2</sup>

### A Changed Consensus

The question is why the consensus that supported the "natural monopoly" and state ownership of telephone services began to dissolve in the early 1990s. There are essentially three motivating factors:

1. Introduction of digital technology;
2. Privatization of SOEs around the world; and
3. Globalization of business.

The most fundamental driver for this trend is the revolution in digital communications technology. A government monopoly was easy to justify for a service that was fairly undifferentiated (i.e., basic phone calls) and for which demand was relatively inelastic. Furthermore, the provision of universal service was a legitimate goal of governments, given that private industry was unwilling to assume the cost of laying lines to every household. However, since the 1970s, telephone penetration no longer constitutes a legitimate measure of the effectiveness of the telephone company. With the introduction of enhanced services, including voice mail, toll-free calling, and the establishment of data networks, the diversity and scope of services have become much more important indicators of progress.

The extraordinary growth of private value-added networks (VANs), especially within multinational organizations, demonstrates the failure of the TO system to adjust to technological innovation. The VAN market is characterized by high product differentiation and elastic demand, meaning narrow groups of users who demand highly customized services. It seems clear that in markets where the product is highly differentiated and demand is more elastic, government monopolies tend to be less flexible in meeting the demands of the market because of the constraints imposed by stakeholders.<sup>3</sup> Telecommunications has assumed a vital role in the competitive advantage of firms, to the point where pricing and the availability of advanced services affect location decisions. This presents a fundamental conflict with the traditional goals of the TO, determined by the stakeholders, of maximizing revenues through high tariffs, and of controlling entry to the network and to the equipment procurement market. Telephone rates set by TOs are often not cost based, but created according to political pressures for cross-subsidization. Although some constituencies benefit from these subsidies, including residential and rural callers, they raise the costs of doing business and thus harm overall economic efficiency.

In addition to the onset of digital technology, fiscal and competitive pressures have convinced many governments of the need to privatize the telecommunications industry. Telecommunications is now seen as an advanced factor that will determine a nation's competitive position in the global marketplace. However, governments have encountered difficulty attempting to deprive TO employees of their civil servant status, or more drastically, of their jobs. The political pressures inherent in the privatization process have often resulted in unhappy compromises that are at odds with the new services. For example, the ability of the European TOs to retain their monopoly of voice services even after 1998 will provoke tensions, especially when the Internet is used to carry voice.

Finally, changes in the telecommunications industry have mandated the globalization of companies within the industry. These changes have taken place at the regulatory level, as well as at the levels of the customer, the competition, and the industry as a whole. The consumer who purchases voice, data, and video services is growing more and more homogeneous across national borders. At one time, consumers in the developed world demanded more sophisticated products, whereas consumers in developing nations strove simply to acquire basic telephone service. However, reliable communication is now recognized as providing the means to successful competition across the world, and new technology has given consumers the opportunity to leapfrog outdated communications systems to achieve a certain measure of equality. Thus, the cellular phone is no longer simply a status symbol for the wealthy but is instead a vital business tool from London to Moscow. The globalization of many industries provides for synergistic development: as financial services, for example, become standardized worldwide, the technology that supports these services must likewise go global.

The U.S., British, and Japanese companies that underwent liberalization in the 1980s, and are thus free from political constraints that bind other TOs, have been much more aggressive in their international strategies. These firms are seeking to provide one-stop shopping for all telecommunications services by creating worldwide, seamless networks. At the same time, these companies have all realized that no single firm possesses the resources to offer global services alone. The result has been a host of alliances, including AT&T-Unisource (the TOs of Holland, Switzerland, and Sweden), Eunetcom (France Telecom and Deutsche Telecom), and the British Telecom—MCI joint venture.

Together, the revolution of digital technology, the privatization of SOEs across industrialized as well as developing countries, and the internationalization of business have totally undermined the traditional TO structure. The internationalization of the Internet can only hasten this trend. The Internet, as a decentralized, uncontrolled service, stands in direct contradiction to the centralized, government-controlled model of the old TO system. The old system includes not only the state enterprises themselves, but also the entire system of standards setting (including supranational institutions like the International Telecommunications Union), a pricing structure distorted by political considerations, a national body of law that regulates areas including intellectual property and pornography, and even the national security of certain countries. Yet in responding to the challenge of the Internet model, countries often rely on traditional methods of control that are rendered ineffective by the new services. In Singapore, for example, Internet users are encouraged to report seditious or pornographic material to the Singapore Broadcasting Authority. In the United States, Senator Jim Exon sponsored the Communications Decency Act in response to pornographic images available on the Internet. In Europe, the European Union's (EU) Council of Ministers has agreed on a directive that limits the transmission of information about individuals,

especially to countries that do not offer adequate protection of privacy. All of these efforts will be impossible to enforce without literally pulling the plug on users.

The problem is that many societies remain wedded to traditional structures that no longer match the new communications services. The EU has set 1998 as the deadline for the liberalization of the telecommunications market. The state-owned TOs will face mandated competition in every area but voice communications. However, new services like the Internet can integrate voice, data, and images, rendering old distinctions between basic telephony and value-added services meaningless. Perpetuating the voice monopoly of the TOs will significantly hinder the advance of countries that fail to recognize the advancing convergence of technologies and services. Though governments should continue to ensure that the majority of the population receives telephone access, the Internet demonstrates the foolishness of reserving a small slice of communications services for a monopoly provider.

### ANALYSIS

The Internet has grown far beyond the original vision of a network connecting research institutions and universities in the United States. More than 70 countries now have full TCP/IP connectivity, linking over 30 million users, and about 150 countries have at least e-mail service. About 1 million users are joining the Internet every month worldwide.<sup>4</sup> Governments across the world have identified international networking as critical to their national competitiveness, although few seem to actually have considered the subversive role of the Internet in undermining traditional political, commercial, and social structures.

#### Political Issues

As Subcomandante Marcos, the leader of the Zapatista rebellion in the Chiapas region of Mexico, explained, "What governments should really fear is a communications expert."<sup>5</sup> Though traditional communications outlets such as radio and television are centralized and often controlled by the government, the Internet remains decentralized, diffuse, and nearly impossible to police. Thus, communiqués from the Zapatistas are transmitted throughout the world via the Internet. Likewise, during the 1991 coup in Moscow, reformers trapped inside the Russian Parliament sent out bulletins on the Internet, which the Voice of America picked up and then transmitted back to Russia over the radio. Obviously, the most repressive societies that restrict access to the necessary hardware—including North Korea and Iraq—will not be toppled by the information revolution. It is regimes like China and Singapore that recognize advanced telecommunications as a vital component of economic development, yet seek to maintain control over new services, that will face the most serious dilemmas.

#### Legal Issues

The totally unregulated nature of the Internet has allowed the diffusion of content that violates not only political controls but also legal structures. Some of the most public controversies have surrounded the diffusion of pornographic images on the Internet. In the United States, Senator Jim Exon sponsored the Communications Decency Act, which would set fines and jail terms for "lewd, lascivious, filthy, or indecent" material on the Internet or other networks. In New Zealand, the primary link to the Internet is threatened by a new antipornography law.<sup>6</sup> U.S. service providers have responded by increasingly policing their own users. For example, America Online recently cut off subscribers using the network to transmit or receive child pornography images and eliminated the online fan club of the music group Hole.

Other pressing legal issues include the protection of intellectual property rights and of privacy on the Internet, as well as censorship more generally. Essentially, the Internet's tradition of open access and its user-driven content challenge the traditional methods employed by societies—including copyright laws and limits on distribution—to restrict access to and use of controversial or protected information.



### Standards Setting

Standards setting has long been of critical importance in the telecommunications industry, even more so as internationalization has progressed. The interface of various TOs, equipment suppliers, and users requires some degree of cooperation and uniformity. Standards for basic telephony services have traditionally been set by national governments through their ownership of the monopoly TO, and through cooperative agreements brokered by international organizations such as the International Telecommunications Union (ITU). However, the inability of these organizations to keep pace with the advent of new services has allowed users and other nontraditional groups to subvert the top-down, "perfection-oriented" model of standards setting.

The success of the Internet centers on its structure as an eminently flexible, open architecture, which has allowed its evolution to be determined largely by the users themselves. Bodies like the ITU have grown increasingly irrelevant with the introduction of new services such as the Internet, and have seen their turf eroded by new organizations that do not necessarily have official government sanction. For example, the Massachusetts Institute of Technology (MIT) and the European Laboratory for Particle Physics (CERN) announced in July 1994 that the two institutions will work together to further develop and standardize the World Wide Web (WWW).<sup>7</sup> Likewise, countries such as Denmark and Singapore, which are aiming to become the networking hubs for their respective regions, are seizing the initiative to establish the regional standard. As Peter Lorentz Nielsen of the Danish Ministry of Research commented, "We can't wait for Brussels to act."<sup>8</sup>

The problem of compatibility obviously arises when systems such as UNIX and EDI meet, but users who become dependent on networking—from students to home shoppers to small and large businesses—will increasingly demand interoperability. The Internet offers the possibility of entirely user-driven services and technologies. Thus, despite the recent pronouncements of the Group of Seven (G7), governments may ultimately have little to say about the Internet or other new services. The debate over integrating the tangle of supranational standards institutions (ITU, CCIT, ITSC, and so on) may therefore be moot as well. The internationalization of data networks from the Fidonet to the Internet demonstrates above all the importance of open, flexible architectures that can be adapted to regional demands and constraints. Centralized standards setting makes as little sense as centralized network control.

## REGIONAL STUDIES

### Asia

Although the Internet is growing at a phenomenal rate in Asia, three critical components are lacking: legal and social structures, and software interfaces. Furthermore, dictatorships (China) or merely repressive regimes (Singapore) will undertake significant political risks by introducing networks that have international reach and that are impossible to perfectly monitor. In addition, the dominance of English on the international Internet provokes charges of "linguistic imperialism" that might "weaken indigenous languages and colonize the minds of the Asians."<sup>9</sup> Nevertheless, within east Asia only Laos and Burma currently remain without any form of Internet access.

### Hong Kong

Although Hong Kong's Supernet is the oldest and main provider of access lines on the territory, eight other access providers have appeared since late 1994. In a classic example of the threat the Internet poses for traditional TOs, the Hong Kong Telecommunications Authority raided and closed down seven commercial access providers to the Internet in March 1995 for operating without a license. The companies claim that providing access to the Internet falls outside of current regulations.<sup>10</sup>

## Singapore

Singapore leads Asia in building digital networks to support an information infrastructure, and is clearly aiming to serve as the hub for the rest of the region. Despite Singapore's rapid advance, the government is attempting to police the Internet by encouraging users to report any criminal or antisocial behavior to other users as well as to the Singapore Broadcasting Authority, a government agency. As George Yeo, the minister for information and the arts in Singapore, admitted, "SBA cannot police on its own without the support and cooperation of the members of the Singapore cyberspace community."<sup>11</sup> Nevertheless, the Telecommunications Authority of Singapore plans to issue licenses for two more access providers for the Internet in addition to Singapore Telecom, which is currently the sole provider of Internet access through Singnet.

## China

China's leaders have quickly realized that the explosive economic development of the country can only be sustained by a modern telecommunications infrastructure. As a result, the Chinese government has been forced to introduce limited competition and to welcome foreign investment. This is especially significant for a country ruled by an authoritarian regime, where communications have long been defined as a national security issue. In June 1994, China broke up the telephone monopoly, which resulted in the formation of several new companies. Liantong Communications was formed in July 1994 by the cities of Shanghai, Beijing, and Tianjin as the country's second operator. The People's Liberation Army established Ji Tong Communications as another new carrier. Despite the urgent need for capital, however, foreigners have so far been banned from actually owning and operating networks, although they have been allowed to invest in the production of telecommunications equipment. Foreign companies are thus forbidden to take any equity stake in telephone networks.

China has one of the lowest telephone penetration rates in the world, with only three lines available for every 100 people.<sup>12</sup> The government plans to lay about 15 million lines per year until it reaches 100 million lines by the end of the century. However, China's leaders have not restricted their vision to basic voice communications, but also plan to leapfrog to the latest technologies through the so-called "Golden Projects." The Golden Projects include three separate initiatives:

1. *Golden Bridge*: the backbone for data, voice, and image transmission;
2. *Golden Card*: a national credit card transaction system; and
3. *Golden Gate*: a paperless trading system to support offices, financial institutions, and foreign trade.

To accomplish the Golden Projects, the Chinese government has called for overall planning, joint construction, unified standards, and a combination of public and special information networks.<sup>13</sup> China will concentrate its investment in three main areas: digital switching equipment, fiber optic cable networks, and expanding the mobile phone networks.<sup>14</sup> The Chinese Ministry of Electronics Industry assigned the Golden Bridge project to Ji Tong Communications company, which will construct a nationwide information network using satellites and fiber optic cables. Although many major cities have already built local data networks, the Golden Bridge will ultimately link 500 cities and more than 12,000 information sources, including large enterprises, institutions, and government offices. Ji Tong has entered an alliance with the China National Space Industry Administration, as well as with IBM, to support construction. Ji Tong has established the Shanghai Jinqiao Network Engineering Company—together with Shanghai Pushi Electronics Company and two university and scientific institutions—to begin building the Golden Bridge in Shanghai. The Jinqiao Network Engineering Company will be responsible for designing, building, and maintaining a system that is to link the State Council, provincial capitals, more than 400 cities, and over 1,000 large enterprises throughout the country.<sup>15</sup>

However, the reality of the communications culture in China remains in conflict with the stated goals of the Golden Bridge. At present, information centers for the central government's 55 ministries store 75 percent of all official information in China but are not interconnected and keep almost all of their information secret. According to a report in *Economic Information Daily* in November 1994, only about 6 percent of all official

information is allowed to publicly circulate. Partly as a result of the continuing perception of information as proprietary, only 300 of the country's 600 data banks are in fact operational, and fewer than 60 of these offer public access.<sup>16</sup> Thus, most of the software and equipment that China has introduced from abroad to support the information infrastructure remains useless. *Economic Information Daily* pointed out that 90 percent of the 100,000 electronic mailboxes imported from the United States in 1992 remain idle. Essentially, the premise of the Golden Bridge remains questionable within a regime whose attitude toward information is ambivalent at best, and where institutional jealousies regarding data prevent its diffusion even among government offices. There are signs, however, that this is slowly changing. For example, China's central bank released current money supply figures in 1994 for the first time since the communist takeover.<sup>17</sup>

### Europe

At the moment, about 1 million computers in Western Europe are connected to the Internet. However, the development of international networks in Europe has been hampered by the perpetuation of the state-owned, monopoly TO structure. As a result, movement toward a European Information Infrastructure has been mostly controlled by the state telecommunications companies (telcos), and thus fragmented along national lines. For example, while Britain offers the most liberal telecommunications market, where even cable television operators are allowed to offer voice services, not only does France Telecom continue to maintain a total monopoly over voice communications, but the French government has also failed to propose any framework for privatization. In Italy and Spain the basic telephone system itself requires significant investment.

In addition, concerns about U.S. cultural dominance, which have pervaded EU legislation on media and film, have entered the discussion on the Internet. France is especially resistant to opening the country to international networks that might threaten local culture and language.<sup>18</sup> (Ironically, Internet traffic is growing at a monthly rate of 15 percent in France.<sup>19</sup>) EU Commissioner Edith Cresson of France has proposed a 0.5 percent telecommunications tax to subsidize European content for online services. Jacques Santer named the erosion of cultural diversity as the number one risk of the technological revolution in the opening speech of the G7 Conference on the Information Society.<sup>20</sup>

Finally, European governments disagree among themselves as to what information belongs online—Scandinavia's tradition of open government contrasts with the more secretive British structure.<sup>21</sup> Partial elements of an information infrastructure already exist in Europe in the form of national academic and commercial networks, but Europe still lacks guidelines for coordinated regional development. Denmark is determined to establish itself as the network leader in Europe, with the goal of bringing all government offices, hospitals, schools, businesses, and research institutions online by the year 2000 in the "InfoSociety 2000" project. As the first mover in Europe, Denmark also hopes to create the standards for the whole continent. According to the Danish proposal, all government agencies would go online first, creating a large body of users to drive costs down before introducing the network to the general population. Yet Denmark has not addressed the monopoly position of the state-owned TO, TeleDenmark, which was only partially privatized in 1994. Furthermore, InfoSociety 2000 reveals a statist approach to telecommunications development that is seriously misplaced in a world where the private sector continually pushes innovation and creates its own standards.

The concern that Europe will be left behind as the United States roars ahead on the information superhighway prompted the creation of an EU task force led by Martin Bangemann, the German Minister of Economic Affairs. The task force consists of 20 politicians and the heads of major European communications and computer firms. The task force reported their findings to the European Council of Ministers in May 1994, beginning their statement with an affirmation of "market mechanisms as the motive power to carry us into the Information Age. It means developing a common regulatory approach to bring forth a competitive Europe-wide market for information services. It does not mean more public money, financial assistance, subsidies, or protectionism."<sup>22</sup> Nevertheless, there remains a wide gap between these words and the reality of telecommunications development in Europe. In contrast with the haphazard and anarchic evolution of the Internet in the United States, the process in Europe aims to be much more deliberative and top-down. As Horst Nasko, Vice Chairman at Siemens Nixdorf, stated "In the U.S. they seem to shoot first and aim later. ... What we are

trying to do is aim first and shoot when we are sure about what to do."<sup>23</sup> This sentiment obviously contradicts the decentralized character of the Internet that has been so critical to its widespread acceptance.

### **Africa**

An underdeveloped telecommunications structure continues to hamper Africa's access to all levels of service. Overall, charges remain much higher and line quality much worse than elsewhere. Besides the lack of infrastructure, the legal model of most national TOs forbids third-party traffic, making the commercial provision of online services impossible. Nevertheless, led by nongovernmental organizations (NGOs) such as the South African Development Community (SADC), as well as by commercial entities, some progress has occurred. The SADC's Early Warning Unit gathers advance data on the food supply situation from each of the 10 member countries as well as from international agencies to provide advance notice of impending food shortages.<sup>24</sup> The SADC employs the Magonet node through Harare, and converts to Fidonet nodes as they become available to communicate with various institutions via e-mail. Rascal Datacom is currently working with Barclays Bank to establish an Overseas Banking Retail Network (OBRN), which will provide e-mail service to 11 African countries, the Middle East, the Caribbean, and the United Kingdom. The initial phase of the projects is centered in Nairobi and is based on a combined LAN and X.25 solution. Ethernet LANs provide connectivity between users in major sites, whereas users across broader regions will apply more flexible solutions appropriate to local user requirements and available services and technologies.<sup>25</sup> By employing solutions that match local needs and infrastructure, both the OBRN project and the SADC Early Warning system demonstrate that online connectivity is possible and highly useful even in regions that have missed the traditional route to telecommunications development.

Another route to establishing connectivity in less developed areas is the galvanizing force of a "local leader." For example, an individual in South Africa has connected the country's research and academic Internet backbone to similar institutions in Botswana, Lesotho, Mauritius, Mozambique, Namibia, Swaziland, Zambia, and Zimbabwe.

### **Latin America**

As in Africa, local telecommunications infrastructure remains an obstacle to high-speed international links. Although international operators like MCI and AT&T offer service, coverage is not total, forcing users to employ the local TO. The "last-mile problem" becomes especially acute for communications outside of metropolitan areas. Many multinationals in Latin America have thus opted to create private VANs. Eastman Kodak, for example, relies on a private X.25 data network that links to its Rochester, New York-based computing center and to other host mainframes in Mexico and Brazil. Kodak opted for the X.25 network because of its ability to operate on bad wiring.<sup>26</sup> As companies in Africa and Latin America create their own networking solutions, this sometimes allows the local population to piggyback on private carriers. For example, investment by Banco Pacifico has encouraged the diffusion of Internet connectivity in Ecuador.<sup>27</sup> However, tight control by the government on international traffic can limit this option. Thus, the Internet remains fairly underdeveloped in Latin America.

## **FORECAST AND RECOMMENDATIONS**

### **The International Internet versus Local Networks**

Many multinational corporations have already established proprietary value-added networks (VANs) to carry vital data among far-flung subsidiaries, suppliers, and customers. Likewise, industry groups in various countries have created their own networks to promote information sharing. An example of the latter is Electronic Data Interchange (EDI) in Europe. EDI was invented by the retailing and automobile industries as a "business

process and not a technology."<sup>28</sup> The essential characteristics of the Internet as a global, scalable, decentralized network that delivers information at cost stands in contrast to the highly centralized, planned, and focused nature of these local business networks.<sup>29</sup> Nevertheless, both services are similar in that they are user- rather than vendor-driven.

In terms of standards, the Internet assigns arbitrary addresses, while EDI employs an X.400 messaging system, in which addresses are logical and hierarchically structured. Searches on the Internet require a good deal of time and luck, but searches on local networks are usually constructed to be both dependable and repeatable. Furthermore, the services offered on business-oriented VANs are carefully planned to meet specifically defined information needs. Mike Holderness, writing in *The Guardian*, called these local VANs the "Outernet" and views the emerging information infrastructure as a battle between the Internet and the Outernet models.

The Internet has succeeded precisely because it is not tightly controlled to deliver specific bits, but instead offers whatever the users themselves see fit to make available. By definition, this structure cannot adopt a standard to offer 100 percent reliability. In contrast, the Outernet perpetuates the hierarchical, "one-to-many" communication model of traditional media.<sup>30</sup> Thus, the question is whether the Internet can succeed as a model for an international network, given the concerns of national governments over issues of control, or whether the Outernet model will succeed instead.

### Leaders versus Followers

Besides the less developed countries (LDCs), even the member nations of the EU are worried that the United States is beating them in the race to construct an information superhighway. South African Deputy President Thabo Mbeki reminded the Group of Seven (G7) meeting in March 1995 that Manhattan offers more telephone lines than all of sub-Saharan Africa. Politicians warn that the digital revolution will divide societies and nations between the information-rich and information-poor. This pervasive sense of some countries gaining a lead in the information revolution is somewhat misplaced. Though the dearth of actual hardware obviously limits the ability of a country to connect to international networks, once connectivity is established, even the economically poor immediately become information-rich.

The question remains, If the global information infrastructure will be constructed by blending national information infrastructures, what of the countries that lack the resources to develop this infrastructure? The extent of network services has been directly related to overall economic development.

In many developing countries, connectivity is available only through a Fidonet link to a few PCs. In contrast, for most of the industrialized world, building a national backbone is now a top priority. For many of these countries—including Australia, Chile, Japan, South Africa, and Switzerland—this has resulted from the creation of research and academic networks.<sup>31</sup> Nevertheless, the Fidonet, which is based on a simple protocol, has provided even the poorest countries with a networking tool to store messages and check for transmission errors. Furthermore, the Fidonet enables users to evade the local TO by allowing messages to be polled from a site outside of the country.

As mentioned in the "Regional Studies" section above, often a country will act as a local leader within a region, such as Singapore in Southeast Asia, Costa Rica in Central America, or South Africa in southern Africa. As demand and experience increase, the connections may evolve from Fidonet to UUCP to IP.<sup>32</sup>

### Obstacles to the International Internet

Despite the clear intention of the industrialized world to foster the building of national backbones, and the gradual diffusion of connectivity in many developing countries, the traditional TO structure, as well as the resulting legal and commercial models this structure fosters, remains a serious obstacle to a truly international Internet. Although technical difficulties can be overcome with resources from institutions such as the World Bank, NGOs, and governments themselves, the traditional mindset of control over the communications infrastructure and services is more difficult to displace. As discussed in the "Background" section of this paper,

governments have long justified their ownership of the telecommunications operator on the basis of national security reasons and also derive significant political and revenue benefits from this ownership. Although this structure has been seriously undermined in the United States, the European Union, and parts of Asia, it remains strong elsewhere.

### CONCLUSION

Ideally, an international network like the Internet should provide a protocol that is easily adapted to a wide variety of infrastructure development stages, and offer services that can be tailored to respect the cultural, legal, and regulatory norms of every country. However, the model that the Internet has provided demonstrates that an international network will, by definition, still act to undermine many traditional structures that have evolved around the old TO system. Rather than seeking to impose old standards of behavior and control on the Internet, governments can best encourage the development of national information infrastructures by eliminating the inherent conflicts that exist between the new services and the domestic organization of telecommunications. This means introducing competition into all levels of service and allowing the market to drive pricing and standards.

### NOTES

1. This discussion is drawn specifically from a lecture by Professor Aaron Tornell of Harvard University, 1 March 1994, and more generally from his course, "Privatization" (Economics 1470, Spring 1994). I am grateful to Professor Tornell for providing the methodological framework for analyzing the structure of an SOE.
2. Dutch, Raymond M. 1991. *Privatizing the Economy, Telecommunications Policy in Comparative Perspective*. University of Michigan Press, Ann Arbor, Mich., p. 141.
3. Dutch, Raymond M. 1991. *Privatizing the Economy, Telecommunications Policy in Comparative Perspective*. University of Michigan Press, Ann Arbor, Mich., pp. 144–145.
4. *OECD Observer*, February 1995; and Goodman, S.E., L.I. Press, S.R. Ruth, and A.M. Rutkowski. 1994. "The Global Diffusion of the Internet," *Communications of the ACM*, Vol. 8, August, p. 27.
5. Watson, Russell, John Barry, Christopher Dickey, and Tim Padgett. 1995. "When Words Are the Best Weapon," *Newsweek*, February 27, p. 36.
6. Lever, Rob. 1995. "Battle Lines Drawn in Debate on Smut in Cyberspace," *Agence France Presse*, March 29.
7. Isa, Margaret. 1994. "MIT, European Lab to Work on World Computer Network," *Boston Globe*, July 8, p. 62.
8. Neuffer, Elizabeth. 1995. "Europe's Tiny Cyberpower," *Boston Globe*, March 27, p. 1.
9. Dixit, Kunda. 1994. "Communication: Traffic Jams on Superhighway," *Inter Press Service*, August 8.
10. Liden, Jon. 1995. "Internet: Asia's Latest Boom Industry," *International Herald Tribune*, March 8.
11. Richardson, Michael. 1995. "Singapore to Police Information Superhighway," *International Herald Tribune*, March 9.
12. Ramsay, Laura. 1995. "China Sees Great Leap Forward in Telecommunications Ability," *Financial Post*, April 1, p. 30.
13. Yu'an, Zhang. 1994. "China: Country Embracing Information Revolution," *China Daily*, October 7.
14. Murray, Geoffrey. 1995. "China Sets Priorities for Massive Telecom Investment," *Japan Economic Newswire*, March 31.
15. Wei, Li. 1994. "China: Company's Creation Marks 1st Step Towards Network," *Shanghai Star*, October 18.
16. Chandra, Rajiv. 1994. "China-Communications: Dial 'S' for Secrecy," *Inter Press Service*, November 28.
17. Murray, Geoffrey. 1995. "China Sets Priorities for Massive Telecom Investment," *Japan Economic Newswire*, March 31.
18. Neuffer, Elizabeth. 1995. "Europe's Tiny Cyberpower," *Boston Globe*, March 27, p. 1.
19. *Monthly Report on Europe*. 1995. "European Union: On-line Data Service to Shed Some Light on the EU," Vol. 127, March 23.

20. *RAPID*. 1995. "Opening Address by Jacques Santer at the G7 Conference on the Information Society," February 24.
21. Neuffer, Elizabeth. 1995. "Europe's Tiny Cyberpower," *Boston Globe*, March 27, p. 1.
22. Brasier, Mary. 1994. "European Business: Race for Knowledge on Super-highway," *The Daily Telegraph*, July 25.
23. Tate, Paul. 1995. "Putting Their Heads Together—Europe's Top Execs Take on Info Highway Challenge," *Information Week*, April 11, p. 26.
24. *African Economic Digest*. 1995. "Africa and the Information Superhighway," January 30.
25. *African Economic Digest*. 1995. "Africa and the Information Superhighway," January 30.
26. Lamonica, Martin. 1995. "Latin American Telecom Remains in State of Flux," *Network World*, March 13, p. 37.
27. Goodman, S.E., L.I. Press, S.R. Ruth, and A.M. Rutkowski. 1994. "The Global Diffusion of the Internet," *Communications of the ACM*, Vol. 8 (August), p. 27.
28. *Computing*. 1995. "EDI-Market Growth," January 26.
29. Holderness, Mike. 1995. "Networks: When Two Cultures Collide," *The Guardian*, February 2, p. 4.
30. Holderness, Mike. 1995. "Networks: When Two Cultures Collide," *The Guardian*, February 2, p. 4.
31. Goodman, S.E., L.I. Press, S.R. Ruth, and A.M. Rutkowski. 1994. "The Global Diffusion of the Internet," *Communications of the ACM*, Vol. 8 (August), p. 27.
32. Goodman, S.E., L.I. Press, S.R. Ruth, and A.M. Rutkowski. 1994. "The Global Diffusion of the Internet," *Communications of the ACM*, Vol. 8 (August), p. 27.

## 24

# The Internet, the World Wide Web, and Open Information Services: How to Build the Global Information Infrastructure

*Charles H. Ferguson*  
*Vermeer Technologies Inc.*

### SUMMARY

The rise of the Internet represents the most fundamental transformation in information technology since the development of personal computers. We are witnessing the emergence of an open, distributed, global information infrastructure based on the Internet, World Wide Web servers, and Mosaic. At the same time, inexpensive servers, fast networks, client-server technology, and visual software tools usable by nonprogrammers are transforming the strategic use of information by organizations. Taken together, these developments offer an opportunity to revolutionize information services and electronic commerce, generating both an enormous business opportunity and a chance to vastly improve access to information for everyone.

Over the next 5 years, the Internet software industry will construct the architectures and products that will be the core of information infrastructure for the next several decades. The promise of these technologies to enable a global information infrastructure can hardly be exaggerated. Decisions made during this critical period will have a profound effect on the information economy of the next century. But there is a serious risk that avoidable mistakes and/or entrenched economic interests will cause the opportunity to be lost or much reduced.

This paper therefore discusses the principles that should drive technology development and adoption in the Internet market, especially for the World Wide Web. Our goal is to promote the development of an open architecture Internet/Web software industry, and to support the deployment of the Internet/Web software industry, and to support the deployment of the most open, easy to use, and productive information infrastructure possible. We believe that a simple set of principles, if adhered to by vendors and buyers, can maximize the openness, interoperability, and growth of both the Internet-based infrastructure and the industry providing it.

Vermeer Technologies intends to succeed as a firm by contributing to the development of this open architecture, Web-based software industry. Vermeer is developing open, standards-based, client-server visual tools for collaborative World Wide Web service development. These visual tools will enable end-users to use the Internet to inexpensively develop and operate powerful World Wide Web information services, which currently require complex programming.

### THE OPPORTUNITY TO BUILD A GLOBAL INFORMATION INFRASTRUCTURE

The global Internet has rapidly evolved into the basis of an open, nonproprietary, global information infrastructure. The Internet now contains nearly 25,000 networks and 30 million users and is growing at a rate



of 10 percent per month. Within the next year, Internet connectivity will be bundled as a standard function with nearly all newly sold personal computers. The Internet architecture already provides excellent nonproprietary standards for basic connectivity, electronic mail, bulletin boards, and, perhaps most critically of all, the World Wide Web architecture for distributed information services. The Internet Engineering Task Force (IETF) and the newly formed MIT-CERN World Wide Web Consortium (W3C) will continue to evolve these basic standards (e.g., by adding security) and we support these efforts completely.

At the same time, there is an explosion of commercial investment in development of Internet software such as Web servers, Web browsers, and Internet access products. In general, we support the development of the Internet software industry because there is a real need to bring state-of-the-art, commercial software technology to the Internet. In this way the Internet and the Web, originally developed by and for researchers, can and should be made fully accessible to the entire world of end-users. We feel that this is an extremely important goal, nearly as critical as developing the Internet itself: information infrastructure should be easily usable by everyone, on any computer, and should not be available only to programmers and researchers. We do not require that everyone learn typesetting and printing to write a book; we should not restrict electronic publishing to those who can write computer programs.

The independent software industry has already developed a large set of technologies that address these problems in other markets and that could be of huge benefit to an Internet-based information infrastructure. State-of-the-art commercial technologies applicable to the Internet include visual tools and WYSIWYG techniques that enable end-users to develop applications that previously required programming; client-server architectures; online help systems; platform-independent software engineering techniques; and systematic quality assurance and testing methodologies. Adobe, Quark, Powersoft, the Macintosh GUI, and even Microsoft have used these techniques to make software easier to use. If these techniques were applied to Internet software, the result could be a huge improvement in everyone's ability to use, communicate, publish, and find information. However, commercial efforts must respect the openness, interoperability, and architectural decentralization that have made the Internet successful in the first place.

### **THE WORLD WIDE WEB AND THE REVOLUTION IN INFORMATION SERVICES**

With the possible exception of basic electronic mail, the World Wide Web (WWW) is the most vital and revolutionary component of Internet-based information infrastructure. The WWW architecture provides a remarkable opportunity to construct an open, distributed, interoperable, and universally accessible information services industry. The Web, started about 5 years ago, now contains tens of thousands of servers and is growing at a rate of 20 percent per month. It is now being used not only to publish information over the Internet but also to provide internal information services within organizations.

In combination with TCP/IP, the Internet, and SMTP-based e-mail integration services, the Web will enable the development of a new information services sector combining the universal access and directory services of the telephone system with the benefits of desktop publishing. If we develop this industry properly, and continue to honor the openness of the Web architecture, the result will be an explosion of information access and a huge new global industry.

The importance of the Web, of its open architecture, and of enabling everyone to use it can hardly be overstated. The World Wide Web offers, for the first time, the opportunity to liberate computer users, publishers, and information providers from the grip of the conventional online services industry. This \$14 billion industry, which includes such firms as America Online and Bloomberg, is strikingly similar to the mainframe computer industry. It once represented progress but has long since become technologically obsolete. It maintains its profitability only by charging extremely high royalties and by holding proprietary control over closed systems. Some current online services vendors continue to retard progress to maintain their financial viability.

There is consequently a real risk that entrenched incumbents in the online services industry will try to suppress the Web or to turn it into simply another collection of proprietary, closed, noninteroperable architectures. There is a similar risk that other companies, such as vendors of commercial Web servers,

browsers, tools, or system software, might attempt to do likewise to establish a new generation of closed, proprietary systems.

Such a return to the world of centralized, proprietary systems would be a disaster, but it need not take place. If developed properly by the emerging Internet software industry, the Web offers huge advantages relative to conventional, centralized online services and would generate gigantic revenues because the Web enables many applications unreachable by the current industry. Web-based services are enabling the free publication of huge quantities of information, real-time access to individual and workgroup data, the rise of large-scale internal corporate information services, the use of online services for educational and library applications, and the growth of information services and electronic commerce as a strategic component of all business processes.

Conventional services cannot do any of these things. They cannot make use of local, distributed, and/or real-time information stored on personal systems or workgroup servers; their capacity is severely limited; they are not interoperable with each other; they cannot be used for internal information services; they cannot be managed by those who create their content; they cannot integrate with database systems and legacy applications used in operating businesses; they are expensive and uneconomical for many services, including most free services; they cannot be linked with each other; and they cannot be viewed using a single, common graphical program such as Mosaic. In contrast, the World Wide Web offers the potential for millions of electronic publications, information services, authors, and publishers to evolve in a layered, open, interoperable industry with support from navigation and directory services.

### THE CURRENT SITUATION AND SOME PRINCIPLES FOR FUTURE DEVELOPMENT

The Internet, the Web, and Mosaic have already laid an excellent foundation for the development of standardized, open, distributed information services. However, two major problems remain. The first is that this foundation will come under attack from vendors interested in slowing progress or exerting control via closed systems. The second problem is that the Internet, and especially the Web, remain much too hard to use.

The first problem—attacks on the openness of the Web—must be dealt with simply by industry vigilance. Internet software vendors should adhere to, and customers should insist on, several basic principles in this industry. These include complete support for current and future nonproprietary IETF and W3C standards; open architectures and vendor-independent APIs; cross-platform and multiplatform products; and complete independence of each product from the other products and services provided by the same vendor. Thus buyers should resist attempts by vendors to link viewers, servers, tools, operating systems, specific information services, and/or Internet access provision with each other. Every product and architecture should be open and forced to compete independently.

The second problem—the fact that Web services are still overly difficult to create and use—requires further work. At present, tools are extremely hard to use, do not provide WYSIWYG capability, do not manage the complexity of hypertext-based services well, and do nothing to eliminate the need for custom programming. Most interesting user interactions with Web servers—sending or requesting electronic mail, performing text searches, accessing databases, creating or filling out forms—require custom programming on the Web server.

The importance of this problem is frequently underestimated because of the computer science origins of the Internet community. However, a few simple facts can illustrate this point. First, information services and/or Web servers can remain hard to develop only when there are few of them. There are still at most 100,000 Web servers in use, most of them deployed in the last 6 months. But this year, 2 million to 4 million Intel-based servers will be shipped, and the server market is growing at a rate of 100 percent per year. If in the long run 10 percent of all servers and 5 percent of all personal computers run Web servers, then the Web server installed base will soon be in the millions. If the average Web server holds content from 5 people, then it will be necessary or at least desirable to enable 5 million to 10 million people to develop Web services. There are not enough programmers to do that.

Even more importantly, the people who understand what the services should look like are the professionals close to the application, not the programmers currently required to code it. Furthermore, the development and maintenance of information services should be a seamlessly collaborative client-server activity.

It should be possible to develop, debug, and edit services over the Internet, using PC-based graphical tools, in collaboration with other remotely located or mobile developers.

We have seen situations like this before. Before development of spreadsheets, accountants performed spreadsheet computations by asking MIS to write a COBOL program. With the advent of PC-based spreadsheets, professionals could perform such computations themselves far more effectively and could share them with coworkers. Desktop publishing, presentation graphics, and visual application development tools such as PowerBuilder had similar effects. We believe that modern graphical tools will do the same for the construction of Web-based online information services.

### VERMEER TECHNOLOGIES AND ITS MISSION

Vermeer Technologies Inc. is a venture-capital funded independent software firm founded in 1994 by Charles Ferguson and Randy Forgaard. Vermeer Technologies intends to become an industry leader in Internet software by contributing to the construction of an open, standards-based information infrastructure available to everyone. In particular, we plan to make it easy for anyone to develop a Web-based information service, either for internal use within their organization or for publication on the Internet.

Accordingly, Vermeer is developing open, standards-based, client-server visual tools for collaborative World Wide Web service development. These visual tools will enable end-users and professionals (collaborating across the Internet) to inexpensively develop and operate powerful World Wide Web information services, without the need for programming. (These services currently require complex custom programming.) Nonprogrammers will be able to develop services for the first time, and professional developers will gain highly leveraged productivity tools. Our architecture also supports many usage models ranging from individual self-publishing to collaborative remote authoring for large commercial Web hosting services. Our architecture is platform independent and will be available on all major client and server computer platforms, on all operating systems, and for all standard-conforming commercial Web servers. Our vendor-independent, open APIs will enable us to construct partnerships with other industry leaders in complementary areas such as text indexing, electronic payment systems, high-performance Web servers, and other functionalities to be developed in the future.

Vermeer intends to rigorously support IETF and W3C standards and is a member of the World Wide Web Consortium. Vermeer's architecture relies on and supports all current standards and is designed to accommodate future standards as they are finalized. Vermeer is an entirely independent firm and has no entrenched interests derived from existing or proprietary products or businesses. Vermeer is therefore completely free to focus on the construction of the most open, easy to use, interoperable products possible.

## 25

# Organizing the Issues

*Francis Dummer Fisher*

*University of Texas at Austin*

### SUMMARY

The following should be kept in mind as the rich submissions to the NII 2000 project are synthesized:

1. *Division of markets.* The local telecommunications markets should be differentiated; the residential broadband market should be distinguished from the downtown business market and from the rural market. The downtown market is likely to take care of itself, and the rural market has unique problems. In view of the importance of universal coverage, the residential market should be emphasized.
2. *"Competition" and the closed character of cable TV.* It should be clearly recognized that cable TV, since it controls both network and content, has a strong economic interest in excluding other content suppliers from unrestricted access to the cable TV broadband network. Even were there to be more than one vertically integrated broadband network reaching homes, openness would not be assured without either regulation or antitrust litigation. Confusion between "competition" and "openness" should be avoided.
3. *State and local regulation.* Federal regulation was emphasized in the NII 2000 project white papers and discussion. Yet federal power to force openness may be limited, as suggested by the claim of those supplying video that the First Amendment confers rights to select whatever video messages the network owner wishes. It may ultimately be the power of local governments to place conditions on use of their rights-of-way that must be invoked to achieve universal and open use of broadband networks.
4. *A broadband connection to the Internet.* A connection to the Internet for anyone who is served by a local broadband network would achieve many of the goals of the national information infrastructure (NII) initiative.

### THE RESIDENTIAL HOME MARKET

The phrases "local exchange" and "local telecommunications market" can gain from greater precision. It is particularly important to distinguish between the market in the downtown business area and the residential market. The residential market can be defined as those homes and businesses that are passed by cable TV but that are not located in the downtown market. In the downtown market, traffic is great and distances small, and the relative cost of entry to serve the market is low. Many different networks and services will be provided. Hence, there is no problem of "openness." In the residential market, traffic is much lower in relation to the area covered. The problems in the residential market will be to assure universal open service by at least one provider and to arrange that the network will be open.

One reason for distinguishing the residential market from the rural market (customers not served by cable) is that the technology for providing broadband service in the former seems much clearer at this time. Indeed, for the residential market, there was in the NII 2000 project activities substantial agreement on the probable technology: a digital service over largely fiber networks with the ability to increase two-way bandwidth as demanded. Special problems in the rural market, besides providing the technology of choice, include difficulties in providing bandwidth "symmetry" and the likely need for subsidies to assure universal service in sparsely populated areas.

Since many of the most important policy issues that must be faced in constructing the NII are presented in assuring universal broadband communications to the home, it will help to address these issues in the specific context of the residential market.

### "COMPETITION" AND THE CLOSED CHARACTER OF CABLE TV

The wire broadband network that exists today in the residential market is that of cable TV. It is especially important, therefore, in considering how it might evolve as part of the NII to recognize the strong economic interest of cable TV in continuing to operate a closed network. The problem in designing the NII is only partially technical; it is equally economic. And unless we attend to it, economics may dictate technology.

"Openness" is easy to define. A network is open if it permits users, at their choice, to be connected to competing sources of information and permits providers easy access to users. In a truly open network, users and providers cannot be distinguished, although those connected to the network can be distinguished by the amount of bandwidth they require. The phone network is an example of an open network.

For cable TV, an open network would permit a customer to connect to packaged TV programs offered by firms that compete with the network owner and the packages it offers.

Cable TV is a closed, vertically integrated system. The existing cable TV system is well described in the white papers in this volume. What is important here is that the same company owns the network and sells, or arranges for the sale of, the content moving over the network. As contrasted with common carriage or an "open system," suppliers of content over cable TV do not compete directly for the business of customers, striking with customers whatever deal the market demands. The business arrangements are, rather, first between content producer and the cable company and then, for those products that cable TV decides to offer, between the cable TV company and those connected to its network.

It could be argued that the profits of cable TV flowing from its monopoly character are necessary to raise the money with which to upgrade the home network, but that harsh argument has not been advanced in the papers.

Perhaps understandably, the papers do not make clear the closed character of vertically integrated networks in the absence of regulation. Bailey and Chiddix state that "while companies such as Time Warner will be one of many content providers, these PC networks that the industry is building will be networks that success in a competitive world will demand be kept open." Rodgers contends that "where the network provider faces competition" it has "an incentive to make interconnection as easy as possible." These comments ignore the profit maximization behavior of competing virtually integrated companies. Powers et al. postulate that where there is unbundling and resale on a fair basis, providers of services can compete, but the additional statement that "effective competition can bring about those results" is not supported.

In Brugliera et al., one paragraph stands out:

It is further arguable that regulation should work to inhibit or even prohibit any single entity from owning or controlling both program production and program delivery operations, if such control results in the customer being denied access to any available program source. Consumer access should be a [sic] prime consideration.

Yet the general scenario presented in that paper is that of the "500 channel" technology that satisfies what is described as "the consumer's primary desire for passive entertainment, and not interactivity." One wonders how the writers would decide, if forced to choose for their homes today, between the telephone or the TV as a single permitted technology. Would not interactivity and choice be reasons for their likely preference for the phone? Why should those considerations diminish as bandwidth grows?

Of course, if one of the network competitors offers an open system, as does the phone company for transporting voice and data, cable TV as a competitor will probably offer equally open transport as to those services. According to Bailey and Chiddix, "Regardless of whether PC interconnection ultimately flows through a number of competing national on-line services or through the laissez-faire anarchy of the Internet, cable intends to offer a highly competitive avenue for local residential and business access to any viable service provider." Personick states, "All RBOCs have expressed a commitment to deploy broadband access services as quickly as

the market demand, technology cost trends, and regulatory/legal environment permit." These comments seem to apply to voice and data, and not to video. Phone companies entering the video market claim that they, like cable TV companies, have the right to use their video network to capture customers for those particular content services offered by them in their video-supplying role.

The interest of cable TV companies in closed systems is reflected not only in the fact that cable TV companies trade for much more than the capital costs of the network investments. It is shown also by the reluctance to be clear about "openness" as even an eventual goal for video networks. In Austin, Texas, where the city is offering access to its right-of-way (poles and conduits owned by the municipal electric utility) to any firm willing to operate an open network, the local cable company, far from applying for the resources that it could use in its investment program, has sought to block the city-sponsored effort to achieve openness.

Competition does not imply openness. Even if competing networks did develop in the home market, it does not follow that customers would have the option of being connected to any information source of choice, or that suppliers of content would be able to access the network, if the competing companies were vertically integrated. For it would probably be in the interest of both such "competing" video providers not to open their networks to other content providers. For example, each network might offer an alarm service, but a third alarm service might well be denied access over either network.

Huber et al. consider the possibility of overbuilding local networks and estimate that the savings would "swamp" the unnecessary costs of overbuilding. Their paper provides little support for this estimate, and its comparison between monopoly and oligopoly probably applies to the monthly charges for those content services offered. It does not follow that a competing deliverer of content could even gain access to either network.

If one cable network could handle all the traffic in the residential market and were open to all, the competition between services envisaged by Huber et al. could still take place, just as it can take place by firms using common poles or conduits and without the cost of overbuilding. This illustrates that it is important to be precise in specifying exactly what it is that is in competition: the services of local carriers (alarm services, voice mail, etc.), the content provided, or merely two largely identical fiber cables.

In sum, the economic interests of cable TV are at present in opposition to achieving the goal of the NII in an open network; to achieve the desired openness on these important broadband networks will require public action.

## FEDERAL, STATE, AND LOCAL REGULATION

### Interconnection and Regulation

Some of the papers in this volume intimate that the regulation required to achieve openness is fairly minimal. Arlow recommends that all providers of telecommunication services should be required to interconnect with each other, "but there should be no other day-to-day regulatory oversight or standards to which providers are obliged to adhere." Mohan urges the public sector to "remove impediments to the deployment and interconnection of competitive local exchange and access capabilities." These authors, and Huber et al., exemplify the tendency to treat deregulation and interconnection as feasibly consistent. In fact, however, mandated interconnection, without regulation of interconnection charges, would be meaningless, and with such price regulation closely resembles common carriage regulation, the heart of traditional government intervention in wired telecommunications.

### Role for States and Localities

In both the papers and the discussion of them at the forum, there was a strong tendency to identify government action as the action of the federal government. But it may well be that by applying the First Amendment to video content, courts have substantially reduced the federal power to prescribe the kind of openness for video that pervades the voice and data system of the telephone network and, to date, of the Internet.

As Ithiel deSola Pool presciently forecast, it may ultimately be the power of local government to control its right-of-way that provides the governmental lever to open up the cable video networks.

#### **A MANDATE FOR BROADBAND CONNECTION TO THE INTERNET**

In this political climate, it could be that openness and switched access to competing content suppliers might be more readily achieved through connection to the Internet than by what might sound like a backward invocation of common carriage regulation of video.

In view of the unsettled constitutional basis for a federal mandate of openness in broadband networks, it could be that a simple statement by the National Research Council urging localities to make broadband connection to the Internet a condition of renewing municipal cable franchises could be an important step in achieving the goals of the NII.

## 26

# The Argument for Universal Access to the Health Care Information Infrastructure: The Particular Needs of Rural Areas, the Poor, and the Underserved

*Richard Friedman and Sean Thomas*

*University of Wisconsin*

### STATEMENT OF THE PROBLEM

The national information infrastructure (NII) offers an opportunity for the poor, persons living in rural areas, and the otherwise underserved to obtain a wide range of educational and medical services to which they currently have only limited access. These benefits can occur only if these groups gain access to the NII. Universal access to the telecommunications infrastructure has been a stated goal for many decades. As a result of a combination of funding transfers, funded and unfunded mandates, and regulations, the telecommunications industry has achieved a 94 percent overall telephone access rate. Unfortunately, the very groups that may benefit most from access to the NII are the very ones that have had the lowest penetration rate in the existing telecommunications infrastructure.

Upgrading the current system to the higher bandwidths that will be required for the interactive capabilities required in telemedicine will be expensive. Unfortunately, current technologies are making it possible for competitive access providers to bypass the local access providers who are maintaining the existing telecommunications infrastructure. They can preferentially provide services at a discount to the lowest-cost clients. The local access providers are left with the higher-cost users such as the poor and rural population. They are increasingly finding this an unreasonable burden, and many existing carriers are finding it uneconomical to upgrade services to these populations.

These are the very populations that may benefit most from the NII. The impact on their education and health care will be dramatic and the benefits to society as a whole will be extensive. It will therefore be necessary for society (i.e., the federal government) to make sure that these populations have access to the NII. This will require a new funding mechanism that no longer relies on internal industry fund transfers. It will require additional government subsidies, new taxes on competitive access providers, or legislative mandates.

We argue that the best course would be a system of value-added taxes on all services that use the NII. The funds from such taxes could be used to underwrite the development and maintenance of the NII.

### BACKGROUND

Rural education and medical services are in the midst of dramatic changes. Shrinking rural populations are making community schools and hospitals hard to justify. While many states have legislated new curricula that attempt to improve the competency of students, rural schools generally lack the funds to attract the specialized teachers necessary for courses in these topics. In many rural areas there is a shortage of physicians. Local hospitals are being forced to close because they are noncompetitive. Medicare and Medicaid provide decreased physician and hospital reimbursement in rural areas. Many poorer patients in rural areas rely on Medicare and Medicaid funding for medical services, yet these funds no longer cover the costs of providing even these basic services. Many residents of rural areas must travel long distances to regional health care centers. Rural



preventive medical care is compromised because of the greater distances and limited access to services.

The economic competitiveness of rural areas is being compromised because of their inability to access high-speed transmission lines for electronic mail, telecommunications, video conferencing, and access to large centralized data bases used by libraries, in inventory control, and in updating of government records. Persons living in rural areas must pay toll charges when contacting government or regional health care offices, whereas residents in larger cities have no such expenses.

The case has been made that the NII will make it possible for rural, poor, and underserved populations to access high-quality educational and health care resources. These resources have previously been denied to these populations due to a combination of factors that include their inability to access these resources, the costs of these resources, and the limited number of high-quality providers of these services who were willing to relocate at sites convenient to these populations. If these services can be provided via the NII at a level of quality comparable to that provided to other populations, then the main impediment to these groups' obtaining this benefit will be their ability to access the NII.

Distance education holds the promise of bringing specialized teaching programs to poor and rural schools. Programs in foreign languages, science, and the arts, which are not available due to inadequate funding or class size, could be made available by video linkages. Participation in government programs and the provision of an "informed" electorate could result from improved access to computer-based educational resources. Rural constituents could sign up for government programs, renew licenses, submit applications for benefits, and obtain many other services, thus avoiding the time and expense now involved in travel to government offices. Health care consultation by experts at distant regional health care centers might be provided at local clinics. Preventive medicine programs, health education, access to health resources, and rapid clinical consultations are all possible consequences of access to the NII.

The provision of high-quality educational and health care services over the NII is currently being evaluated. Preliminary studies have shown that high-quality services are possible. The Mayo Clinic maintains video conferencing links between three sites in Florida, Arizona, and Minnesota. It recently conducted a telemedicine program linking its clinic in Rochester, Minnesota, with the Pine Ridge Indian Reservation in South Dakota. The Medical College of Georgia has several dozen sites on its telemedicine network. This year it has begun trials with direct links to patients' homes so that individuals recuperating from heart surgery can be linked to their doctors for follow-up. The Navy runs a teleradiology program estimated to save \$14 million a year. Studies in Georgia have shown that using telemedicine can increase the number of patients admitted to rural hospitals. Rural hospitals typically charge up to \$500 less per day than hospitals in large cities. Telemedicine offers a way to retain and retrain medical professionals and help keep rural hospitals open.<sup>1</sup>

Norway funds the largest telemedicine program in the world. In that country telemedicine is part of the government-run health care system. Practitioners regularly conduct consultations in a variety of specialties between remote hospitals and main academic centers. One ENT specialist suggested that the number of cases he receives from general practitioners has decreased by 50 percent since the introduction of telemedicine consultations.<sup>2</sup>

The NII is already being used to provide a wide range of medical information via bulletin boards, newsgroups, and electronic mailing lists via the Internet and commercial services such as Compuserve, Prodigy, and America Online. These services provide a great deal of patient-oriented medical information. The World Wide Web (WWW), a multimedia hypertext service, is the fastest growing component of the Internet. It contains a plethora of medical services, including newsgroups and bulletin boards as well as diagnostic, educational, and preventive services. Most major state and federal health care groups provide online access to their resources via the WWW. Many medical schools and university hospitals already provide patient information, course material, up-to-date medical literature, and even consultative services via the NII.

Many clinical services are available via the NII. The problem remains one of access. The cost of the computer equipment required to interface with the NII is falling dramatically and will eventually be only a small fraction of the expense of providing quality educational and health care services by conventional modalities. The main expense will be the heavy investment necessary to bring the telecommunications infrastructure to the homes and communities of all individuals, particularly those who are poor or live in rural or inner city locations. Given

their locations in either decaying, densely populated inner city areas or widely distributed and sparsely populated rural locations, it is extremely expensive to provide access.

The relatively low income of these individuals makes it unlikely that commercial providers of the NII can profit from providing this access. The commercial incentives for providing NII access to these groups may not exist. However, the benefit to the nation of providing quality education and health care to these individuals will be great. Because economic benefit will accrue to the nation as a whole rather than to the companies building and maintaining the NII, the usual capitalistic incentives that would make providing such services worthwhile are not appropriate. The government must either subsidize the provision of such services or somehow mandate that the providers of such services include these populations as a condition of any license to provide services to the remainder of the population.

In the past, mandates for rural electrification or universal telephone access took care of this problem. At the time these services were originally provided, it was decided that such universal access was in the public good. The idea of universal telephone service has been a foundation of the information policy of the United States since the Communication Act of 1934 was passed, creating the Federal Communications Commission

for the purpose of regulating interstate and foreign commerce in communications by wire and radio so as to make available, so far as possible, to all the people of the United States a rapid, efficient, nationwide, and worldwide wire and radio communications service with adequate facilities at reasonable charges, for the purpose of national defense, for the purpose of promoting safety of life and property through the use of wire and radio communications, and for the purpose of securing a more effective execution of this policy by centralizing authority.

The current telephone penetration rate for U.S. households is approximately 94 percent.<sup>3</sup> In addition, approximately 1.3 percent have a phone available or nearby. Roughly 4.5 percent of Americans (4.4 million households and 11.6 million individuals) have no telephone available.<sup>4</sup>

With respect to at-risk populations, the elderly actually do better than young parents with children with respect to telephone access. Access to telephone service for retired persons at all income levels is at the national average or better. Among the elderly, only those persons receiving Supplemental Security Income have a lower than average penetration of telephones (79.7 percent to 84.9 percent).<sup>5</sup> The disparity in access to telephone service is most pronounced for people of all ages with low incomes. It is noted that 31 percent of all families receiving food stamps have no telephone.<sup>6</sup> For families completely dependent on public assistance the percentage rises to 43.5.

Only 2.2 percent of homeowners have no telephone, but 21.7 percent of persons in public housing have no telephone and 40.2 percent of those living in hotel rooms or boarding houses have no telephone. Women and children are particularly vulnerable. Households headed by women with children who are living at or below the poverty line have a telephone penetration rate of only 50 percent. Race and ethnic background appear to confound the impact of income on telephone access. The percentage of white households without telephones is 8 to 10 percentage points lower than black and Hispanic households.

Persons living in rural communities outside metropolitan areas lack telephones in 9.9 percent of cases. In rural New Mexico only 88 percent of all households have phone service. In the Four Corners areas of Colorado, New Mexico, Arizona, and Utah, where 36,000 Navajo households are scattered about on 27,000 square miles, 21,300 households (63 percent) have no telephones. In that region installation fees can range as high as \$5,000 for a single telephone.<sup>7</sup>

It therefore appears that factors that affect telephone penetration are low income, household headed by a woman with children, race and ethnicity, and rural location. Yet these are the very groups that would appear to benefit from the recent advances in telecommunications.

## ANALYSIS

Today the public mandate for universal access is being removed or bypassed, and in some rural areas, universal access to telephone or electrical utilities does not in effect exist. To obtain these services the user must

pay for linkages to trunk lines that may be at a considerable distance from the home. The cost of such a link can be expensive. Consequently, many rural areas do not have cable or cellular telephone services. The equipment for new high-bandwidth connections necessary to provide state-of-the-art NII services may be prohibitively expensive for many rural and/or inner-city users.

However, these are the groups most likely to benefit from such services. The improvement in their education and health care will in turn be of great benefit to society in general. Once NII access is provided, the provision of such educational and health care services to these groups will represent only a modest additional marginal expense. It therefore is in the interest of the government and the general population that the NII is made available to these populations and further that it is provided as a governmental service or as a government mandate to the providers of such services.

The current method for providing near universal access to telephone services relies on funding transfers among carriers, transfers from other service providers, transfers between customers, and government contributions. The primary transfers have been from business and toll service users to basic residential subscribers and from urban to rural exchanges. Before the AT&T divestiture, these transfers were accomplished largely within a single corporate umbrella. Since then the systems has become more complex, with a system of rules governing transfers among carriers based on political compromise rather than economic logic. Currently there are three major fund transfer mechanisms: (1) funds transferred from toll services to local access providers, (2) transfer from low-cost local exchanges to high-cost local exchanges (i.e., from urban to rural exchanges), and (3) differential allocation of local service costs through rates charged to different customer classes (i.e., higher rates for business than for residential users). There is also a system of subsidization for social support services such as lifeline programs for low income customers, 911 emergency services, and services for the hearing impaired, all of which are covered through local telephone charges.<sup>8</sup>

With the arrival of increased competition and the exponential growth of providers of equipment, toll services, business services, etc., the rules are being altered. The difference between core services and access lines is blurring. As fiber and other broadband media continue to move into local neighborhoods and business districts and as the already existing cable networks begin to offer switched services, the points of interconnection among these paradigms are changing. The growth of alternate access providers has made it possible for large toll customers to directly connect to toll service providers without paying the usage-based within-system transfers that were traditionally collected through local access carriers. As local exchange competition develops, it will be more difficult for local exchange carriers to support high-cost areas or provide the default capacity relied on for system backup by competitive network providers. As cable television networks and wireless personal communication networks develop, they will make it more difficult for local access providers to continue to provide social services below cost. They will also threaten the ability of local exchange carriers to provide service to high-cost customers (inner city, rural, poor, etc.).<sup>9</sup>

The recent advances in communication technologies have raised the concern as to what is a reasonable basic telecommunications service. Is it plain old telephone service (POTS), fiber optics, microwave, satellite uplinks, etc.? Is there a single basic service modality? Some newer communications modalities such as cellular telephone access may be less expensive to provide to rural areas than are telephone lines, but the equipment to receive these services may be more expensive (i.e., cellular telephones). Will the telephone companies be required to provide only access, or access and equipment? What will be the minimum bandwidth that these companies will be required to supply?

If one agrees that there should be universal telecommunications access to the NII, then one must next decide the minimum bandwidth that should be allocated. Will it be for audio communication only, as previously defined, or must it now include video and computer communication capabilities? If it includes video, will it be single frame or action capabilities? Will it be five frames/s or the full 30 frames/s necessary for true action video? Will it be only black and white, or should it include color? For computer communication capabilities what will be the necessary speed? This is a particularly difficult challenge because we must select a bandwidth that is broad enough to supply a range of services (audio, freeze-frame video, action video, data interchange, etc.) yet economically realistic.

In deciding the minimum level of services to be provided we must also realize that the method of connecting to the NII will not be uniform. Some areas already have an extensive communications infrastructure

in place, and therefore it may be more economical to upgrade that existing infrastructure than to replace it wholesale. In other areas where no such infrastructure exists, consideration might be given to providing a newer technology. In some areas wireless communication may be more reasonable, while in others a cable system will be appropriate. After dealing with a relatively stable technology (twisted pair telephone line) for many decades, we are now faced with a rapidly changing panoply of technologies (digital telephone lines, fiber optic cables, microwave towers, satellite linkages, etc.) with variable communication bandwidth. In addition, the bandwidth of each modality is changing rapidly. Existing twisted pair analog telephone linkages can be upgraded to digital lines (ISDN), thus increasing their bandwidth. Using compressed video techniques, we can send video images over traditional telephone lines today in a manner not possible only six months ago.

It would appear that the minimal service configuration should be determined not by the physical type of linkage but rather by its bandwidth capability (i.e., its ability to attain a level of service rather than the physical configuration of that service).

Another major consideration must be the connecting focal point of that communication. Traditionally the communication linkage was to a physical location (home or office). Universal communication meant that all Americans had access to a telephone in their living space. This too is changing. We are now able to provide cellular communication linkages to individuals rather than to physical structures. Shortly we will be able to provide a seamless linkage across the entire nation so that one can move from cellular cell to cell without needing to register in the new area. The minimum standard for telecommunication access may involve service not to a physical household or building but rather to an individual.

In the health care field one could argue that fully interactive video capabilities are necessary if we are to be able to provide remote delivery access. Certainly, to provide remote consultation such video capabilities are necessary. However, if one is to provide remote teleradiology services, a higher bandwidth is necessary; to supply only patient information access and preventive services information, then perhaps a lower bandwidth would be appropriate.

Finally, if we are to provide universal service then we must define not only the bandwidth of that service but also its capabilities. In the past, universal access included access to the physical telephone. Does universal access now involve access to video equipment, computers, high-resolution screens, etc.? We now can use stethoscopes, otoscopes, and ophthalmoscopes at remote sites via telemedicine facilities. Will we include a remote stethoscope as the minimum telecommunications configuration in every household where a person with hypertension resides? In the end it would save lives and might be cost effective.

The major problem remains, Who is to maintain the NII? In the past these costs were either mandated to the carriers, who added them to the general cost of the service, or were explicitly included as an item on the bills of users of the system. They were essentially mandated "unfunded" government benefits, which are now unpopular and insupportable. We must therefore determine another way to support these services. This will become particularly important as access to the NII becomes a "portable" access that is not tied to a physical location.

Given the increasingly "portable" nature of this service, it will become more difficult to add the cost to a basic local access service charge. The concept of the local access company may gradually disappear as wireless systems spread. The only reasonable way to finance such a system is either via a tax on all telecommunications services or by government funding of the basic access service.

There are many ways the government can sustain the idea of universal access. One would be to mandate that the local access provider in a particular area provide service to all households and businesses or individuals living in that area at the same basic rate. This would continue the practice of cross subsidization within a local area and might make some local provider areas unprofitable for any carriers. One could set up a bidding system for local access provider areas, with the highest bidder getting the franchise for the area. These funds would then be available to subsidize services in local access areas that were not cost effective and did not attract any bidders. This would also result in cross-subsidization since companies successfully bidding for the more sought-after franchises would pay a higher fee and therefore find it necessary to charge higher tariffs to their customers. The major problem with this proposal is that the whole idea of a local access area is changing. The physical access location may be an individual. Would the area then become all the persons living in that area, working in the area, and visiting the area? Wireless communication negates the importance of the local access area. In addition,

microwave and satellite linkages make it possible for commercial access providers to bypass the local access provider. Such a mandate might be unenforceable.

A second method would be to follow that utilized in Japan, where the government supports local telecommunications services to all households. With this method the government could decide a basic level of services to all households or individuals and determine the most cost efficient method for providing it, be it telephone line, microwave tower, cellular telephone, etc. The government would then contract with commercial providers to provide these services. Such contracts could be "bid" to the lowest-cost provider who would agree to supply all households or individuals in an area for a set fee.<sup>10</sup> Such an undertaking would be very expensive and would have to be funded from tax revenue. This would, once again, result in a major cross subsidization from higher taxed groups to other groups. The chances of legislating a tax to pay for the free provision of services that most of us already pay for would be remote.

A third approach would be to use value-added service surcharges. These would be calculated as a fixed percentage of gross revenues that potentially could be collected from all businesses selling value-added telecommunication services. Value-added services include the services of all service providers interconnecting with the public switched network, except local loop services provided to homes, businesses, or eventually individuals by state-certified common carriers with provider-of-last-resort obligations.<sup>11</sup> The value-added tax would be the easiest to monitor. Funds from this tax could be used to subsidize basic services to all household or individuals.

The government has the most to gain in promoting universal telecommunications services. If, as expected, it improves our educational system, provides continuing medical education services, enhances health care, and brings these services to the poor, underprivileged, and rural populations, then the government (which currently supplies many expensive remedial series to these groups) would benefit the most. A better educated, healthier, and better informed electorate would greatly benefit the federal government. These benefits would be to the collective good and might not necessarily accrue to the companies supplying these telecommunication services. The government must therefore be the supplier of last resort.

## RECOMMENDATIONS

### Basic Levels of Service

The level of service must include a bandwidth specification based on type of service. We would argue that two-way, real-time video communication is the minimum level of service that should be accepted. Current twisted pair telephone communication will not realize this level of service, but combined digital circuits, fiber optic cables, and microwave and satellite modalities certainly can achieve this standard. There need be no single modality used for the NII. Depending on the area and/or population, a combination of these modalities might be appropriate.

### Physical Access Unit

We believe that the access unit must gradually change from the household to the individual. We believe mobile communication will continue to grow to the point that we will have individual telephone numbers or IP numbers of telecommunication access numbers. An actual geographic access location will cease to be important.

### Access Modality

We believe that there will be a gradual movement from the telephone as the standard of access to the NII into a broader audio and video interface. As a part of universal access, each home, business, and individual will have both audio and video capabilities as well as some minimal computer component. It is still too early to clearly define the actual instrument that will provide this interface, but it will surely be more complex than the

telephone. We must, however, begin to define the basic capabilities of this unit because it will dictate the level of service available.

### Method of Funding Universal Access

We favor a value-added tax on the gross revenues of all providers of telecommunications services. This would include not only competitive access providers but also groups that use the NII to provide data, entertainment, news, financial information, etc. Such a tax is the easiest to calculate and enforce. It should raise enough money to fund the basic telecommunications infrastructure. The government would fund the entire NII with funds from this tax. Once the bandwidth and physical access unit were agreed upon, the government would solicit bids for the supplying of services to a particular area. The suppliers would agree to supply services to all individuals or households in the designated area. As the physical access unit gradually changed from the household or business to the individual, the actual geographic area might be replaced by population groupings.

We believe that the government must provide an NII, that it cannot be achieved through a nonregulated environment. The benefits are potentially too great to allow any segment of the population to be "displaced" because of limited commercial cost-benefit analysis. We must consider the total benefit of universal access, and this can be done only with government intervention. Continued government involvement can result in efficiencies of scale, uniform standards, and universal access.

### REFERENCES

- Beamon, Clarice. 1990. "Telecommunications: A Vital Link for Rural Business," OPASTCO Roundtable.
- Belinfante, A. 1991. "Monitoring Report: Telephone Penetration and Household Family Characteristics," No. CC, Docket No. 80-286. Federal Communications Commission, Washington, D.C.
- Bumble, W.A., and G.J. Sidak. 1993. *Toward Competition in Local Telephone Markets*. MIT Press, Cambridge, Mass.
- Dordick, H.S. 1990. "The Origins of Universal Service," *Telecommunication Policy* 14(3):223-38.
- Dordick, H.S., and M.D. Fife. 1991. "Universal Service in Post-divertiture USA," *Telecommunications Policy* 15(2):119-28.
- Gallottini, Giovanna T. 1991. "Infrastructure: The Rural Difference," *Telecommunications Engineering and Management* 95(1):48-50.
- Hudson, Heather E. 1984. *When Telephones Reach the Village: The Role of Telecommunications in Rural Development*. Ablex, Norwood, New Jersey.
- Mueller, M.L. 1993. "Universal Telephone Service in Telephone History: A Reconstruction," *Telecommunications Policy* 17(July):352-69.

### NOTES

1. See Linder, A. 1994 "Global Telemedicine and the Future of Medical Science," *Healthcare Informatics*, November, pp. 63-66; and McGee, R., and E.G. Tangalos. 1994. "Delivery of Health Care to the Underserved: Potential Contributions of Telecommunications Technology," *Mayo Clinic Proceedings*, Vol. 69, pp. 1131-1136.
2. Linder, A. 1994. "Global Telemedicine and the Future of Medical Science," *Healthcare Informatics* November, pp. 62-66.
3. Schement, J.R. 1994. "Beyond Universal Service: Characteristics of Americans Without Telephones, 1980-1993," Communications Policy Working Paper #1, Benton Foundation, Washington, D.C.
4. Belinfante, A. 1989. "Telephone Penetration and Household Family Characteristics," No. CC, Docket No. 87-339. Federal Communications Commission, Washington, D.C.
5. Schement, J.R. 1994. "Beyond Universal Service: Characteristics of Americans without Telephones, 1980-1993," Communications Policy Working Paper #1, Benton Foundation.
6. Belinfante, A. 1989. "Telephone Penetration and Household Family Characteristics," No. CC, Docket No. 87-339. Federal Communications Commission, Washington, D.C.

7. "How to Flourish in an All Digital World," *Wilt Letter*, Vol. 1, No. 4, December 21, 1993.
8. Hudson, Heather E. 1994. "Universal Service: The Rural Challenge Changing Requirements and Policy Options," Working Paper #2, Benton Foundation, Washington, D.C.
9. Egan, B.L., and S. Wildman. 1994. "Funding the Public Telecommunications Infrastructure," Working Paper #5, Benton Foundation, Washington, D.C.
10. Egan, B.L., and S. Wildman. 1994. "Funding the Public Telecommunications Infrastructure," Working Paper #5, Benton Foundation, Washington, D.C.
11. Egan, B.L., and S. Wildman. 1994. "Funding the Public Telecommunications Infrastructure," Working Paper #5, Benton Foundation, Washington, D.C.

## 27

# Toward a National Data Network: Architectural Issues and the Role of Government

*David A. Garbin*

*MITRE Corporation*

The last two decades have seen several revolutions occur in the telecommunications field, encompassing both the underlying technologies used to construct networks and the services offered to customers over these networks. In this paper, we follow two threads on their converging paths: the emergence and evolution of packet switching as the dominant technology for data communications and the central management of computer-controlled switches as a mechanism to create virtual private networks (VPNs) out of a national infrastructure. By 1990, the second thread culminated in the dominance of VPNs for the voice communications requirements of large customers. Now, both threads can combine to create a similar phenomenon for data communications requirements.

These events are playing out against a background of explosive growth in requirements for data communications. Growing public interest in the Internet and the availability of user-friendly access tools is causing a doubling of Internet traffic every 12 months. Federal government programs focusing on service to the citizen and more efficient operations within government are driving federal agency requirements higher and higher. Finally, there is a national initiative to bring the benefits of reliable, inexpensive data communications to public institutions as a whole through the creation of a national information infrastructure (NII). The federal government role in bringing the NII into being is unclear at the present time; current proposals call for the private sector to play the major role in actually building infrastructure. However, it has been postulated that the federal government could use its vast purchasing power to facilitate the development of an open data network, a key building block of the NII.<sup>1</sup>

It is well known that telecommunications networks exhibit economy-of-scale effects; unit costs decrease as absolute volume increases. This paper explores the economics of single-agency networks, government-wide networks, and networks with the span of the proposed NII. Specifically, it explores the benefits of a government-wide network based on shared public switches and transmission facilities. Such a network would yield the best unit prices for the government and create a solid infrastructure base for the NII at no additional cost.

### BASIC CONCEPTS

Before we begin, a basic review of the key concepts underlying the issues explored in this paper is warranted to avoid any confusion over terms and definitions. These concepts are best discussed in terms of contrasts:

- Voice versus data communications,
- Circuit versus packet switching, and
- Shared versus dedicated networks.

In a simple sense, data communications are between computers, and voice communications are between people. However, that basic fact results in different characteristics for data traffic as opposed to voice traffic.



Data traffic is "bursty" in nature; it occurs in periods of intense activity followed by (sometimes long) periods of silence. If delivery of the data cannot be immediate, they may be stored in the network and delivered at a later time (how much later depends on the application). Data traffic is not tolerant of errors; any error passed to an application will cause problematic behavior of some kind. The above characteristics lead to the requirement for protocols to package the data, transmit the package through intermediate points, check for and correct errors, and deliver the package to the distant end. By contrast, voice communications are continuous in nature and have a real-time delivery requirement. A varying delay for a voice signal results in totally incomprehensible speech at the receiving end. However, voice signals are robust and tolerant of errors. Speech itself is so redundant that words remain comprehensible even at error rates of 1 in 100 in the digital bit stream. It is clearly more important for voice applications to deliver the signal on time than to deliver it with 100 percent accuracy.

These different requirements have led to different switching techniques for voice and data communications. Circuit switching sets up a path with a guaranteed bandwidth from end to end for each voice call. While this would also work for data communications, it would be inefficient since the reserved bandwidth would be unused most of the time. A technique known generically as packet switching was developed to effectively share transmission resources among many data communications users. When we refer to data networks in this paper, we are talking about networks employing some form of packet switching.

The final concept we need to explore is the concept of shared versus dedicated networks. In the early days of telephony, all customers used (shared) the public switched network for voice communications. The backbone of the public switched network was composed of large circuit switches located on the telephone companies' premises. By the 1960s, manufacturers began producing economical switches designed for use on customers' premises to lower costs for large users. While these were primarily for local service, it was soon discovered that large customers could connect these premises switches with private lines and create private networks for long distance voice communications. Public switched service rates at the time were high enough that these private networks were economical for large corporations (and the federal government).

## BACKGROUND

As packet switched data networks came into being in the 1970s, the private network alternative was the only one available to customers. There was no public packet switched data network, nor was there a large demand for one. Private data networks grew alongside the private voice networks, with packet switches on customer premises and private lines connecting the switches. Computer processing technology limited the capacity of packet switches to that required by a single large customer; little penalty was paid in not locating switches on carrier premises where they could be shared by many customers.

In the 1980s, two forces converged to spell the end of the private voice network. Divestiture created a very competitive interexchange market, and computer-controlled switch technology evolved to the point where the partitioning of a large network in software became feasible. In this case, the large network was the public switched network of each interexchange carrier that was serving the general population. Over time, competition drove the unit prices being offered to a wide range of customers down to levels consistent with the large total volume. Volume ceased to be a discriminator for price beyond a level of one-tenth of the federal government-wide traffic. This service, which now dominates the voice communications market, is called virtual private network (VPN) service.

The current approach to data communications for large customers is still the private network. There are two shortcomings to this approach: economies of scale beyond a single user are never obtained, and the proliferation of switches on user premises does not further the development of a national infrastructure. Technology such as asynchronous transfer mode (ATM) switches and high-capacity routers is emerging that makes carrier-premises switching feasible. At the same time, initiatives in government and in the research and education communities are generating a large future demand for data communications. The consolidated demand of the federal government could create an infrastructure that pushes unit costs well up on the economy-of-scale curve. However, this will only be the case if a shared network is used to satisfy these requirements. We call this network, based on standard interfaces and protocols, the National Data Network (NDN). This paper presents the

architecture and economics of such a network as it is used to satisfy the requirements of a single federal agency, the federal government as a whole, and the general population as the transport element of the NII. The implications of this architecture for universal access to the information highway will be apparent.

### ASSUMPTIONS

Any study of telecommunications alternatives must make assumptions about the technological and regulatory environment that will be present during the projected time period. In this case, the time horizon is from the present over the next 5 years. We assume that router and ATM switch technology that is now emerging will be ready for widespread deployment in operational networks during this period. Although the regulatory requirements that mandate the current restrictions relating to local exchange carriers (LECs), interexchange carriers (IXCs), and local access and transport areas (LATAs) are likely to be modified by the end of the period, the current structure is assumed for this study. Even if LATAs as a formal entity were removed, they would continue to be useful to represent statistical areas for traffic purposes.

All switching will be performed at LEC and IXC locations, beginning at the LEC central office, or wire center, serving the customer. Access to the wire center could be through a dedicated local loop or through a shared channel on an integrated services digital network (ISDN) local loop. Local loop costs between the customer's location and the wire center are not part of this study. Routers at wire centers or at tandem offices within a LATA will be provided by the LEC. LATA and regional switches that are used for inter-LATA transport will be provided by the IXC. Note that, in this scenario, both LECs and IXCs must implement shared subnetworks and cooperate in achieving the level of sharing needed for end-to-end economies to be realized. In most cases, this is not the case in today's networks. Each carrier generally uses the other only for dedicated circuits between its switches. The realization of an end-to-end shared network architecture is critical for the formation of a National Data Network.

A final assumption that does not affect the economics of this study, but that is essential to the viability of a shared network, is the successful addressing of security issues. Most present networks use encryption only on interswitch links; the traffic in the switches is in the clear and is protected by physical security. Since this would be an issue if the switch were on carrier premises, more complex security systems that encrypt traffic at the source before the network protocols are added are required to maintain security. Fortunately, such systems are being designed and deployed since they also provide a much higher level of security than the present system.

### STUDY METHODOLOGY

Three physical network models were formulated and evaluated for cost-effectiveness over a range of traffic levels. These models were crafted to represent the generic data communications requirements of a single federal agency, the federal government as a whole, and the general public as a user of the NII. The models differed primarily in the number and distribution of wire centers served by the network. The agency model served 1,350 wire centers; the distribution of wire centers served was derived from projected Treasury locations during the time frame of the study. The government model expanded the coverage to 4,400 wire centers. These corresponded to the wire centers currently served by FTS2000. For the NII model, all the wire centers in the continental United States (21,300) were served.

A network was designed for each model and traffic level. Transmission and switching capacity were sized to meet the throughput requirements for the traffic. Each network was then priced using monthly prices for transmission and amortized prices for switching equipment. A percentage factor was applied for network management based on experience with agency networks. The total costs and unit costs for each model and traffic level were then computed and analyzed. The following sections provide additional details on the network architecture and the traffic and cost models used in the analysis.

### Network Architecture

A four-level hierarchy was used for the network architecture; the locations of the switches were based on the current LEC and IXC structure that exists today. The equipment used for switching was based on the current technology of high-end routers at the lower levels of the hierarchy and ATM switches at the backbone level. As ATM technology evolves, the current routers will likely be replaced by ATM edge switches. For the purposes of this study, the capacity and cost factors of these two technologies would be similar.

At the top of the hierarchy is a backbone consisting of fourteen switches and connecting circuits. The country was divided into seven regions corresponding to the current regional Bell operating company (RBOC) boundaries. Two switches were placed in each region for redundancy. The topology of the interconnecting circuits was based on the interswitch traffic, with the requirement that each switch be connected through at least two paths. The backbone subnetwork is shown in [Figure 1](#).

Within each region, one switch is placed in each LATA in the region. This switch serves as the "point of presence" in the LATA for the backbone network. Each LATA switch is connected to one of the regional backbone switches. [Figure 2](#) illustrates this connectivity for the Northeast region, showing the two backbone switches and fourteen LATA switches.

Within each LATA, traffic from the wire centers is concentrated at tandem routers before being sent to the LATA switch. These tandem routers are located at LEC central offices, usually those serving a large number of exchanges. The number of tandem locations is somewhat dependent on the model in use, but a typical configuration is shown in [Figure 3](#) for the Maine LATA (seven tandem switches are shown).



**FIGURE 1 National Data Network backbone.**

Figure 1 National Data Network backbone.

The final level in the hierarchy consists of routers in the LEC wire centers serving actual users. Access to these routers occurs over the customer's local loop and can take various forms:

- Dedicated local loop connected directly to data terminating equipment,
- ISDN local loop using D or B channel packet mode access, and
- Analog local loop used in dial-up mode.

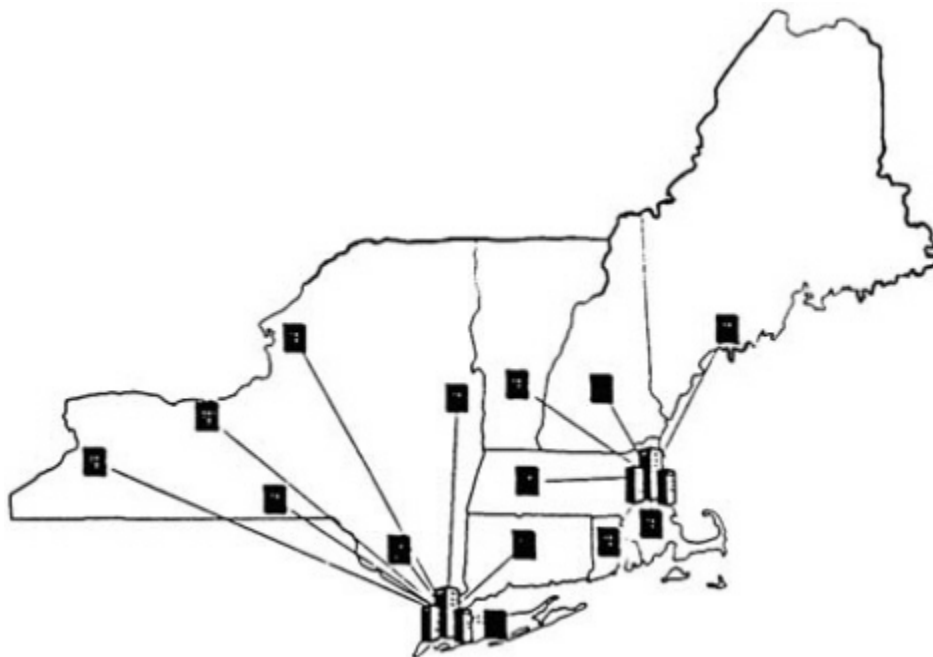


Figure 2 NDN LATA subnetwork (Northeast Region).

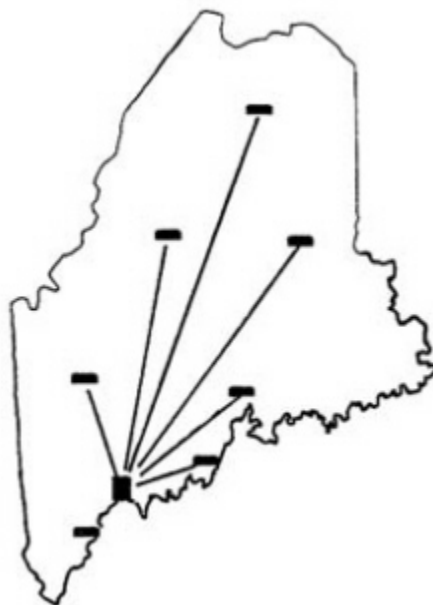


Figure 3 NDN tandem subnetwork (Maine LATA).

As stated above, the number of wire centers served is dependent on the model being evaluated.

### Traffic Model

A single, scalable traffic model was used to evaluate all three physical network models (agency model, government model, NII model). In the future, data applications will encompass all facets of government operation, not only data center operations. Consequently, the current LATA-to-LATA voice traffic distribution of the government as a whole was used as the basis for the traffic model. This reflects the level of government presence and communities of interest within the country. The resulting traffic matrix represents a generic approach to characterizing a national traffic distribution.

Within a LATA, the traffic was assigned to wire centers based on the number of locations served (or, in the NII model, the number of exchanges served). The base unit used to characterize traffic was terabytes per day (1 terabyte = 1 million megabytes). As a calibration point and a way to put the traffic units into perspective, Table 1 gives the approximate traffic volumes for existing and proposed networks.

TABLE 1 Traffic Volumes for Typical Networks

Network	Volume (terabytes/day)
Current FTS2000 Packet Switched Service	0.0075
Treasury TCS (projected)	0.1–0.4
March 1994 NSF Backbone	0.6
All FTS2000 Agencies	1.4–5.6
DOD Operational Networks	&223C;6.0
NII	>20.0

The current FTS2000 packet switched service (PSS) carries only a small percentage of the civilian agency data communications traffic; most of the traffic is carried over private networks using circuits procured under FTS2000 dedicated transmission service (DTS). The Department of the Treasury is procuring such a private network at this time and has estimated its traffic requirements over the life of the new contract. Analysis of current DTS bandwidth utilization by agency indicates that Treasury represents about 7 percent of the total FTS2000 agency requirement for bandwidth. This includes Department of Defense (DOD) circuits on FTS2000 but does not include the large number of DOD mission critical packet switched and IP router networks. The volume of traffic on these networks may be as large as the FTS2000 agency estimate.

As a point of comparison, the March 1994 Internet traffic traversing the NSFNET backbone is given.<sup>2</sup> This traffic doubled in the past year and shows every indication of increasing growth rate. Of particular interest would be the amount of regional Internet traffic that could use the infrastructure generated by the NDN for more economical access and transport service.

The traffic volume generated by a mature NII cannot be estimated; truly, the sky is the limit if any of the future applications being contemplated grabs the imagination of the general public.

This study used volume ranges appropriate to the physical model under consideration. The agency model was evaluated at volumes ranging from 0.006 to 0.5 terabytes per day. The government was modeled as representing 14 typical agencies and was evaluated at volumes ranging from 2 to 8 terabytes per day. Note that although the government model had 14 times the traffic of the typical agency, it utilized only 3.3 times as many wire centers. The NII model extended the reach of the network to 5 times as many wire centers; it was modeled as carrying the traffic of 8 government-size networks (16 to 64 terabytes per day).

### Circuit Cost Model

The cost of circuits used in this study was based on the current, maximally discounted tariffs for interoffice channels (i.e., channels between carrier premises). Local channels were not used since all switches in the study were located on carrier premises. Rates for OC-3 and OC-12 were projected as mature rates following the current economy-of-scale trends. Carrier projections for these rates support this view. The five line speeds used for circuits were as follows:

- 64 kbps,
- 1.5 Mbps (T-1),
- 45 Mbps (T-3),
- 155 Mbps (OC-3), and
- 620 Mbps (OC-12).

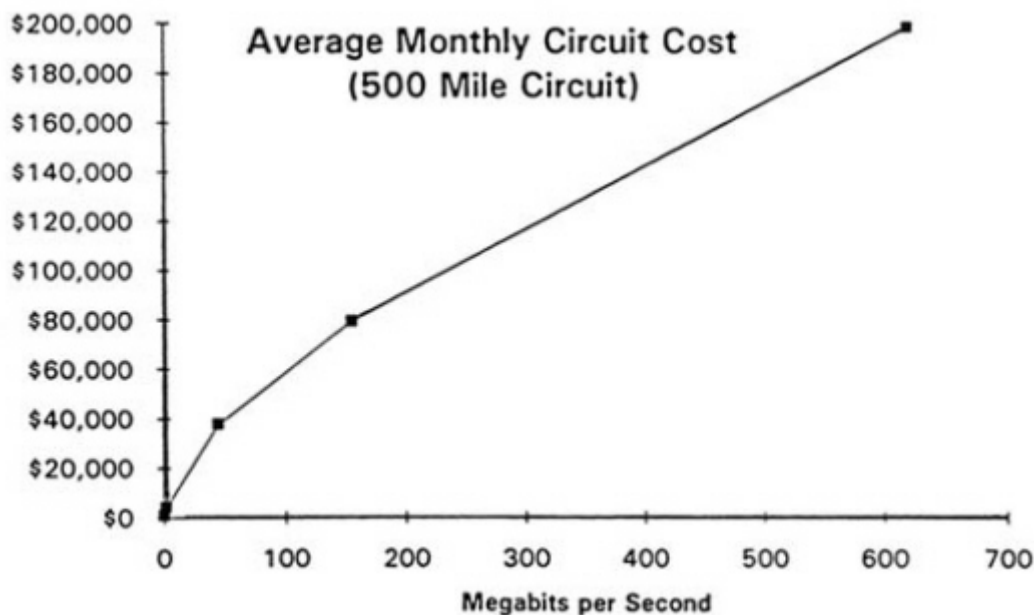


Figure 4 Circuit cost model example.

Figure 4 shows the monthly cost of a 500-mile circuit at different line speeds, illustrating the economies of the higher-speed circuits.

#### Equipment Cost Model

The wire center, tandem switch, and LATA switch cost models were based on high-end router technology (e.g., Cisco 7000). Serial interface processors were used for link terminations. ATM interface processors were assumed for T-3 links and above. ATM concentrator switches in the future should exhibit cost behavior similar to that of the router configurations. The backbone switch cost model was based on high-end, wide-area ATM switch technology (e.g., AT&T GCN-2000).

The one-time cost of equipment was amortized over a 5-year period to obtain an equivalent monthly cost that could be added to the monthly transmission costs. Before amortization, the capital cost of the equipment was increased by 20 percent to account for installation costs. Finally, a monthly cost of maintenance was added at a rate of 9 percent of the capital cost per year. These factors correspond to standard industry experience for these functions.

#### Management Cost Model

Network management costs were estimated at 25 percent of the equipment and transmission costs, based on current experience with agency networks. Implementing management cost estimates as a percentage assumes that network management will show the same economy-of-scale effects as the other cost elements. In fact, large networks will probably realize greater economies in network management than in any other area. These costs are driven mostly by personnel costs, which are relatively insensitive to traffic volume and only marginally related to number of locations.

### RESULTS

The results of the analysis are presented here in two formats. The first graph for each physical model shows the variation of monthly cost with volume. The second shows the variation of unit cost with volume. Unit costs are presented in cents/kilosegment (1 kilosegment = 64,000 characters).

### Agency Model

Figure 5a shows the variation of monthly cost versus volume for the agency traffic model. The curve demonstrates the classic economy-of-scale shape, although the effect is more discernible in the unit cost curve presented in Figure 5b. The unit costs predicted by the model at the lowest traffic levels are consistent with the current unit costs for FTS2000 packet switched service, which is operating at these traffic levels. It can readily be seen that large single agencies can achieve economies with private networks at their current volumes of 0.1 terabytes (TB) per day.

### Government Model

The monthly and unit cost curves for the government model are presented in Figures 6a and 6b. The combined costs of multiple single-agency networks comprising the same volumes are also shown. Significant cost savings are achievable with a government-wide network versus the multiple-agency networks. Unit costs at government-wide volumes (2 to 8 TB/day) are approximately one-third the unit costs realized using the volumes of even the largest single agencies (0.1 to 0.4 TB/day). The economies achievable for the smaller agencies would be much greater. A portion of the reason for the substantial economies realized is the more efficient use of facilities from the local wire centers up to the LATA switches. With multiple-agency networks, a large number of inefficient, low-speed circuits exist in parallel at the same local wire center. With a shared network, the traffic on these circuits can be bundled over more efficient, higher-speed circuits. The situation is made worse if the multiple agency networks use switches on customer premises rather than in wire centers.

### NII Model

The relative economies in moving from government-size networks to networks on the scale of the NII show a similar pattern (Figures 7a and 7b). The savings are not as great as in the previous example since extra costs are involved in extending the reach of the NII into all wire centers. Nevertheless, if the traffic increase is assumed to be on a par with the increased coverage (8 times the traffic with 5 times the wire centers covered), then the economies are still there and the enormous benefits of full coverage are realized.

The single NII network produces a 37 percent unit cost savings over the 8 multiple networks comprising the same volumes. Note that large increases in traffic from the wire centers already serving federal government traffic could be handled at little additional cost. This would be the case in most urban and suburban areas.

### Investment Costs

The cost figures presented above represent the costs as equivalent monthly costs, including the amortized cost of equipment and installation. It is instructive to break the equipment and installation costs out separately since these costs represent capital investment. In particular, Table 2 presents the additional investment required to carry increased traffic, given that a government-wide network carrying 4 TB per day of traffic has already been constructed. The investment in equipment required to build a network of that size is approximately \$160 million. While substantial, this investment is commensurate with the estimated investment made to provide FTS2000 services to the federal government in 1988. These investment costs would be recovered through service charges to the government.

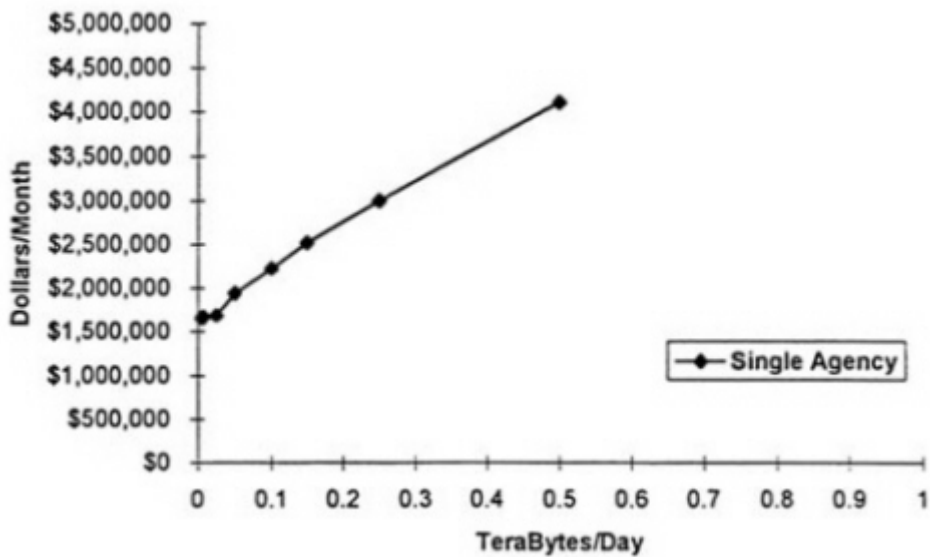


Figure 5a Monthly cost—agency model.

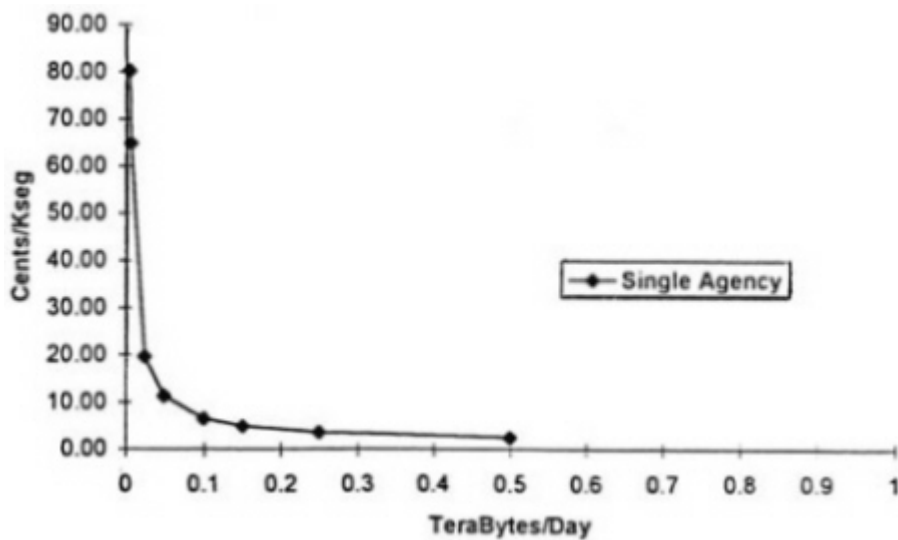


Figure 5b Unit cost—agency model.

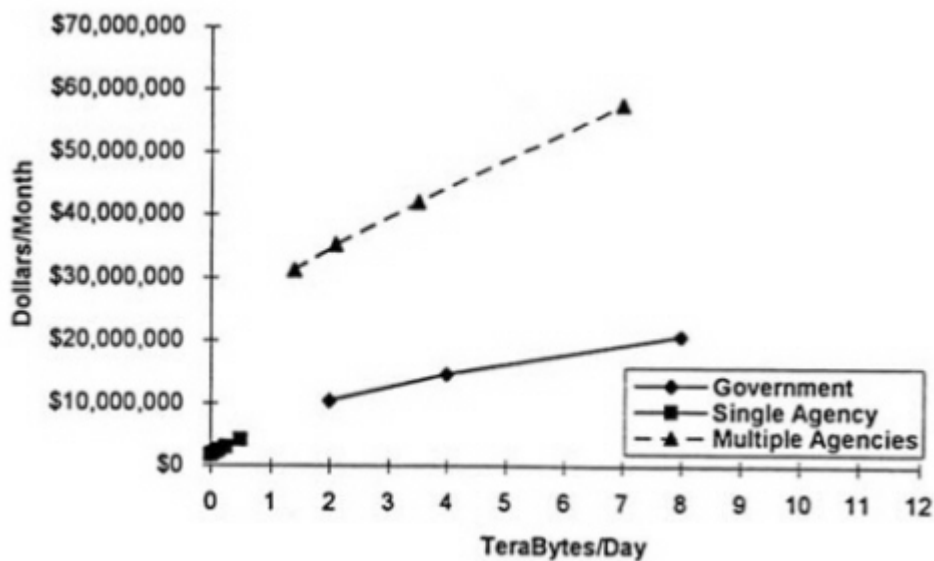


Figure 6a Monthly cost—government model.



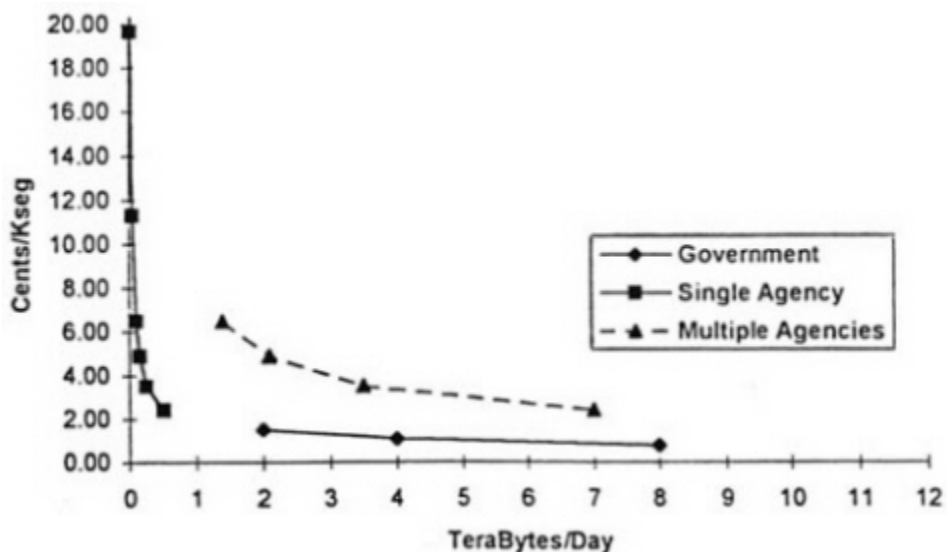


Figure 6b Unit cost—government model.

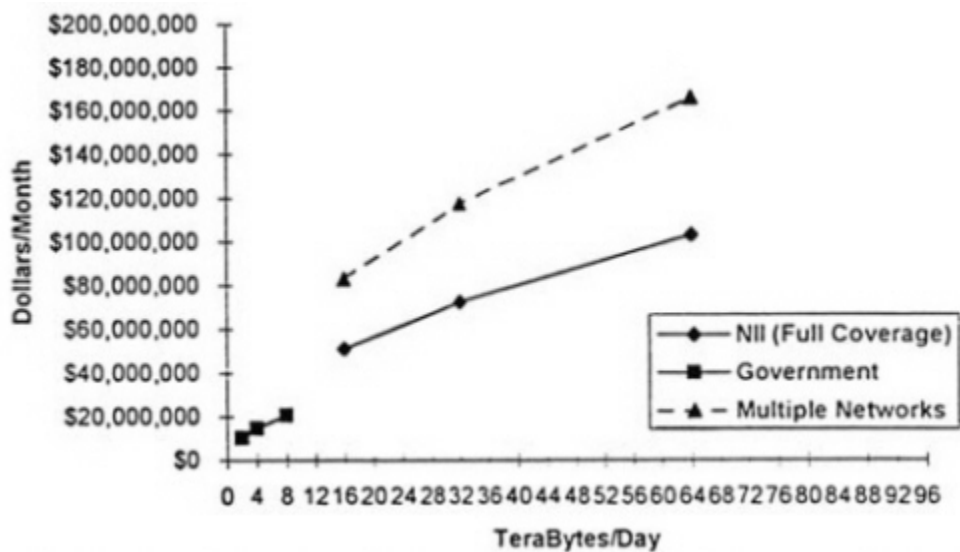


Figure 7a Monthly cost—NII model.

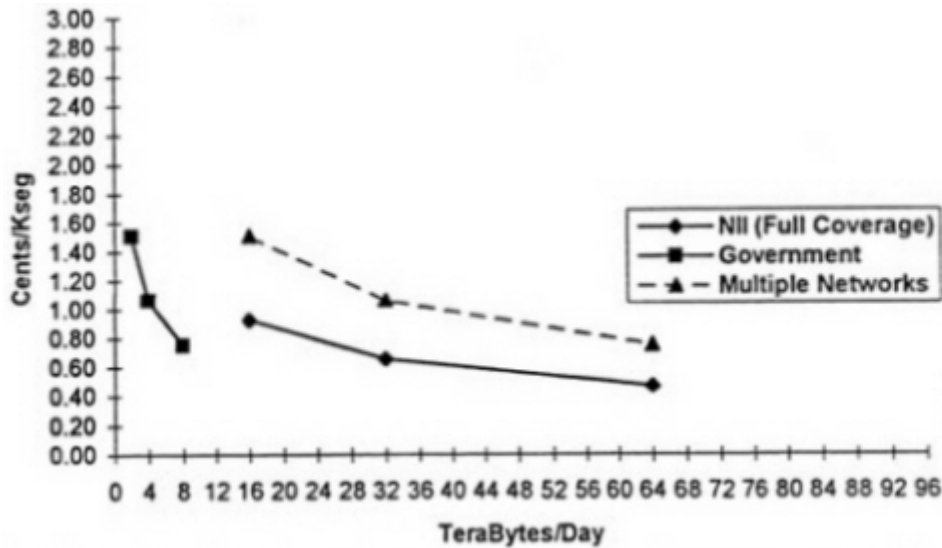


Figure 7b Unit cost—NII model.

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

TABLE 2 Additional Investment Needed to Carry Additional Traffic (millions of dollars)

	+4 TB/day	+12 TB/day	+28 TB/day	+60 TB/day
Government Model	34	67	123	199
NII Model	250	378	518	669

As Table 2 shows, additional traffic can be carried through the same wire centers that serve the government with little additional investment. Additional capital is needed to expand toward a fuller NII infrastructure that would require 5 times as many wire centers to be covered as are covered in the government network. However, the government network would still provide a significant jumping off point for the complete network. For example, an NII network serving all wire centers at 8 TB/day would require an investment of \$410 million in equipment (\$160M for the first 4 TB/day through the government wire centers plus \$250M for the additional 4 TB/day through the remaining wire centers). The government network would have already caused 40 percent of that investment to be made.

The largest portion of the investment and monthly costs is in the access areas of the network, the portion that is normally provided by LECs. This reinforces the point made above in this paper that the shared network concept must be extended all the way to the user. It also points out the need for uniform standards for interfaces and switching in all regions (a minimum requirement for any open data network).

### CONCLUSIONS

Three major conclusions can be drawn from the analysis presented above:

- The infrastructure costs of a National Data Network show a marked economy-of-scale effect at the volumes represented by the federal government data communications traffic.
- Significant economy-of-scale benefits can be achieved by aggregating agency requirements onto a common network.
- The infrastructure created to support federal government requirements can significantly reduce the cost of extending service to larger communities in the public interest.

The savings resulting from the NDN approach are substantial enough to justify the complexities of an aggregated procurement (coordination of requirements, security, standards). Such a procurement would have to be carefully structured to harness the competitive forces necessary to motivate both local and interexchange carriers to pass on the cost savings shown above through lower prices. The end result would be a quantum step forward for the government and the country on the road to the information technology future.

### NOTES

1. Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C.
2. Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C.

## 28

# Statement on National Information Infrastructure Issues

*Oscar Garcia for the IEEE Computer Society*

### STATEMENT OF THE PROBLEM

A key challenge facing the information technology community is to integrate human-computer interaction (HCI) research with that of broadband communication in an economically effective manner serving the broadest possible community. Over the next 5 to 7 years, we expect this issue to become one of the central economic and scientific drivers of the entertainment industry, as well as a key influence on how science and education are conducted in technologically advanced nations. Our organization is particularly concerned with the latter arena. We want to ensure that scientists, students, and other serious information seekers are able to exploit the new technology and, in the case of a large segment of our membership, also contribute to it. One of the many nontechnical barriers is the lack of a forum for dissemination of information about the national information infrastructure (NII), and the IEEE Computer Society, with its technical groups and wide-ranging publications, can be a significant facilitator of such a forum. The deployment should include a strong leveraging of the educational potential, given the more likely entertainment and business aspects. Few universities, high schools, and elementary schools are prepared to profit from the NII deployment, and even fewer understand what changes in their educational modus operandi are likely to take place if they participate. These issues are not frequently included in public policy studies of technology because the responsible curricula designers and implementors are often absent. One certain aspect is that usability via good interfaces must be taken seriously, given the broad spectrum of users.

### AREAS AND EXAMPLES

Five key areas of HCI research activity are summarized below, together with related key developments, enabling technologies, and capabilities:

1. *Human-computer dialogue.* Natural language, voice, and other modalities (pen, gestures, haptic, etc.) can be combined to produce dialogue techniques for HCI. These techniques can be robust enough to tolerate acoustic noise, disfluencies, and so on. Relevant topics here include speech recognition and synthesis; natural language processing and understanding;<sup>1</sup> lip-reading technologies in noisy environments;<sup>2</sup> and search, navigation,<sup>3</sup> and retrieval from cyberspace using multimedia.<sup>4</sup> Speaker identification can be combined with secure passwords for use in control of computer access, and language identification may support improved multilingual interfaces.<sup>5</sup> Understanding of natural language and speech also concerns semantics and prosody (for speech, obviously) and probably the use of semantic nets and semantic encyclopedias now under construction. Technologies of understanding beyond symbolic matching or statistical techniques allow searches invoked by description of the searched object in a less restrictive spoken or written form. Roe and Wilpon<sup>6</sup> present a thorough analysis of the potential uses of speech recognition and synthesis, including its use in aiding physically disabled people. Speech could play a very significant role if used in the new asynchronous transfer mode (ATM) technologies, which are designed, precisely, as a compromise between voice communications and video

transmission. The use of multiuser dialogue is likely to be extremely useful if there is usable software available, such as in the case of scientific research and engineering design.<sup>7</sup>

2. *Machine sensors for HCI and general I/O to facilitate telepresence and teleoperation.* Current computers are "sensory deprived," representing a barrier to both human-computer interaction and machine learning. Such sensors as microphones in single and array configurations, infrared and other means of scanning and computing distances, optical sensors at lightwave range including charge-coupled device (CCD) cameras of small size, haptic interfaces, and alternatives to click-and-point devices should be studied. Fusing sensor inputs to the computer with intelligent or learned action-response behavior would create a more realistic approach to machine learning and complex inferencing techniques, involving symbolic, fuzzy, and probabilistic approaches. This area has been researched with different objectives, but seldom with that of trying to improve the human-computer interface. Standardizing environments (e.g., via a human-computer interaction workbench; HCI-WB) can improve measurements. Such an experimental environment is also useful in the study of human behavior in real and virtual modalities related to the NII, and provides comparisons in human subject variabilities between real and virtual environment behavior, navigation, and orientation. The potential for research in the fusion of the modalities is enormous.<sup>8</sup> The challenge of this research area is to fuse multiple sensor inputs to the computer in a cohesive and well-coordinated manner. One such example would be the integration of a CCD camera input with a haptic experiment using force feedback and synthesized video output. Another helpful experiment could involve mechanisms for the localization of sound in virtual environments<sup>9</sup> using the HCI-WB.
3. *Large storage (archival and nonarchival), database, and indexing technologies,* including multiresolution and compression for different modalities. Video and audio technologies will require large compression factors and mechanisms for rapid encoding and decoding and are difficult to index and access for retrieval, and even then, mass storage database techniques will be required. This area is also indirectly related to the speech and video synthesis technologies, since high-resolution synthesis approaches imply efficient encoding, possibly at different resolution levels. Similarly, virtual environment research requires efficient storage and compression technologies for input and output. There are good reasons to believe, for example, that high-quality audio can be encoded at rates of 2,000 bps using dynamic adaptation of perceptual criteria in coding and articulatory modeling of the speech signal. Therefore, encoding research should include both generation and perceptual factors.<sup>10</sup> Additionally, multimedia databases require techniques for providing temporal modeling and delivery capabilities. A novel interface, called "query scripts," between the client and the database system adds temporal presentation modeling capabilities to queries. Query scripts capture multimedia objects and their temporal relationships for presentation. Thus, query scripts extend the database system's ability to select and define a set of objects for retrieval and delivery by providing a priori knowledge about client requests. This information allows the database system to schedule optimal access plans for delivering multimedia content objects to clients. One more example of an area of concern related to the overall throughput capability of the NII is the Earth Observing System (EOS) of NASA. This system is coupled with a data information system (DIS) in a composite EOSDIS, which is expected, when operational in 1998, to require transport of one terabyte per day of unprocessed data and possibly an order of magnitude more when processed, roughly equivalent to the total daily transport capacity of the current Internet. The question is, Will the NII provide the capacity for even a fraction of such volumes of data?
4. *Virtual environments and their use in networking and wireless communication (tethered and untethered) networked environments*<sup>11</sup> will have an impact on the NII. Virtual environments relate to telepresence and telecommuting, as well as to personal communication services for digital voice. The technologies for telepresence and telecommuting involve a mixture of multimedia and networking. Wireless communication technology also includes techniques such as geopositioning measures, local indoor infrared sensors for location, communications technologies at low, medium, and high bandwidth, and so on. The technical challenges of wireless messaging are well known.<sup>12</sup> In particular, the proposed use of ATM LANs will integrate virtual environment research at different sites with communication research.<sup>13</sup> The concept of virtual environments is taken here in a broad sense, including both head-mounted and enclosed CAVE-like environments,<sup>14</sup> telepresence, and their human factor considerations for the real-time and residual long-term psychological effects of immersion. Strong encouragement for a research emphasis on the human-computer interface is provided by the National Research Council's Committee on Virtual Reality Research and Development, whose final report<sup>15</sup> makes specific

recommendations for federal national research and development investments. Virtual environments are also likely to be used for education and experimentation over distances when there are sufficient educational and psychological developmental technologies and related network communication. For example, realistic virtual environments could be used for technology education in areas that involve high cost and risk, such as in welding training.<sup>16</sup>

5. *Applications of software engineering and CASE* to the R&D of complex software systems and browsers to be used in HCI.<sup>17</sup> Many modules of the software and interfaces for the different modalities might be developed in a compact and reusable manner, taking advantage of existing and newly developed software techniques. It has been found that 50 to 90 percent of all lines of code in industrial and military software are dedicated to human-computer interfaces.<sup>18</sup> In this sense, we include usability studies in the scope of software engineering measurements for interfaces.<sup>19</sup> A special interest is anticipated in experimentation with the facilitation of interactive multimedia educational software development, particularly related to science and engineering topics. Software financial investments in the NII applications would be affected by their ability to be easily accessible to the broad community of NII users. Software engineering for the NII is likely to have a flavor quite different from what has been done in the past at research institutes such as the Software Engineering Institute, strongly based on Ada environments.

### **INTERACTION AMONG TECHNICAL AND NONTECHNICAL (LEGAL/REGULATORY, ECONOMIC, AND SOCIAL) FACTORS**

There are legal concerns with regard to the balance between security and freedom of communications. In particular, a thorny issue to be discussed is the degree of responsibility, if any, that carriers have for transmitting illegal material or for the theft or penetration that may take place when security is breached. There are new socially explosive issue (pornography, copyright issues, etc.) that need to be addressed in the context of networks and information systems. They are related to the financial viability of the human-computer interaction on a large scale by big populations and have a tremendous impact on the publishing industry. A new type of "NII electronic forensics" needs to be established, and it must have a strong technical basis to stand legal scrutiny. This is an area that only highly secret intelligence agencies have dealt with and that universities have incorporated only sporadically in their research areas. It is a delicate area of concern for the public, since it is often related to security and privacy.

### **CONTINGENCIES AND UNCERTAINTIES**

The entertainment industry is most likely to dominate the field. It is most likely (but uncertain) that only a few educational institutions will be able to afford the expenditures associated with supplying educational services to their constituencies. It is not clear how the telephone and publishing industries will react and what their investments will be, but much of it will depend on intellectual property rights protection and the availability of sources of materials. How users will react to this can only be gleaned from some experiments such as the "electronic village" at Virginia Polytechnic Institute's Department of Computer Science. The Digital Libraries Initiative of NSF, ARPA, and NASA needs to continue and be more widely coordinated in a national forum accessible to all.

### **USERS**

Classes of users to be served include the following:

1. *The public.* The public will have access to the media. In general the spectrum of "public users" has a broad range of sophistication. A distribution of know-how would have a large number of naive users (mostly browsers and e-mail users) and then a small number of highly sophisticated users. Age is not a factor in the

experiential know-how. A study of this population should provide a sociotechnical profile, which would be useful in this study. These are people not associated with an educational, government, or industrial institution but rather "home users." They may be seeking information from educational, government, or industrial sources but on an irregular basis.

2. *Those associated with educational institutions*, such as students, teachers, administrators, and librarians.
3. *Those public servants who interface with the public and are in charge of dissemination of government information*; also, individuals able to provide services to those citizens who must be licensed, tallied by a census, taxed, certified for licenses and renewals, and so on. This population could be categorized into federal, state, and local government public servants.
4. *Industrial users*. This category has a large subcategory of entertainment, and possibly "edutainment." These are the salespeople on the network or electronic commerce providers of sources, technology, and products and include, of course, video-on-demand providers. There are many subcategories here.

Disadvantaged persons or those in geographical areas remote to broadband access will be the most difficult to serve, partly because of their technical access problems and partly because, in general, they will most likely be at the low end of user sophistication. They will also be those who are likely to benefit the most from having access to resources that would otherwise be unreachable.

### MARKET RAMP-UP

The market will have to provide "substance" or content. The cost of providing is high. How to provide substantive content, create a cottage industry of providers, allow those potential providers the opportunity to access and sell in a free market, and draw lines of responsibility and legality are but some of the issues that will determine the speed of the ramp-up. Interactivity is expensive, as is any two-way communication, but the bandwidth does not have to be symmetric in both channels. This is an area where technology could have an impact if we understand the human-computer aspects of interactive "dialogue" in a broad sense. Openness should mean possible accessibility to all the users who fall within the service potential of a provider on an equal basis, but should be restrictive, of course, on the basis of registration for cases where financial transactions are to take place. The determination of viable means to charge for services is a techno-economic factor that is of fundamental importance for early resolution and fast ramp-up. The scalability may also be viewed from the point of view of the user's sophistication and needs. Our "help" menus are insufficient and too slow to solve the problems of specialized use for nonspecialized but proper users of the facilities. New approaches to diagnosis of the user's difficulty are a part of the "HCI problem" and are required for fast progress by the public user and even by the moderately sophisticated industrial or government user.

### REFERENCES

1. Cole, R., O.N. Garcia, et al. 1995. "The Challenge of Spoken Language Systems: Research Directions for the Nineties," *IEEE Transactions on Speech and Audio Processing*, January.
2. Garcia, O.N., with A.J. Goldschen and E. Petajan. 1994. "Continuous Optical Automatic Speech Recognition by Lipreading," *Proceedings of the Twenty-Eight Annual Asilomar Conference on Signals, Systems, and Computers*, October 31–November 2, Pacific Grove, Calif.
3. Shank, Gary. 1993. "Abductive Multiloguing: The Semiotic Dynamics of Navigating the Net," *Electronic Journal on Virtual Culture* 1(1).
4. Vin, Harrick M., et al. 1991. "Hierarchical Conferencing Architectures for Inter-Group Multimedia Collaboration," *Proceedings of the ACM Conference on Organizational Computing Systems*, Atlanta, Ga., November.
5. Wilpon, J., L. Rabiner, C.-H. Lee, and E. Goldman, 1990. "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38, November, pp. 1870–1878.

6. Roe, D.V., and J. Wilpon (eds.). 1994. *Voice Communication Between Humans and Machines*. National Academy Press, Washington, D.C.
7. Anupan, V., and C.L. Bajaj. 1994. "Shastra: Multimedia Collaborative Design Environment," *IEEE Multimedia*, Summer, pp. 39–49.
8. Koons, David B., C.J. Saparrel, and K.R. Thorisson. 1993. "Integrating Simultaneous Inputs from Speech, Gaze, and Hand Gestures," in *Intelligent Multi-Media Interfaces*, M. Mayberry (ed.). AAAI Press/MIT Press, Cambridge, Mass., Chapter 11, pp. 257–276.
9. Gilkey, R.H., and T.R. Anderson. 1995. "The Accuracy of Absolute Speech Localization Judgements for Speech Stimuli," submitted to the *Journal for Vestibular Research*.
10. Flanagan, J.L. 1994. "Speech Communication: An Overview," in *Voice Communication Between Humans and Machines*, D.V. Roe and J. Wilpon (eds.). National Academy Press, Washington, D.C.
11. Kobb, B.Z. 1993. "Personal Wireless," *IEEE Spectrum*, June, p. 25.
12. Rattay, K. 1994. "Wireless Messaging," *AT&T Technical Journal*, May/June.
13. Vetter, R.J., and D.H.C. Du. 1995. "Issues and Challenges in ATM Networks," *Communications of the ACM*, special issue dedicated to ATM networks, February.
14. DeFanti, T.A., C. Cruz-Neira, and D. Sandin, 1993. "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of CAVE," *Computer Graphics Proceedings, Annual Conference Series*, pp. 135–142.
15. Durlach, N.I., and A.S. Mavor. 1995. *Virtual Reality: Scientific and Technological Challenges*. National Academy Press, Washington, D.C.
16. Wu, Chuansong. 1992. "Microcomputer-based Welder Training Simulator," *Computers in Industry*. Elsevier Science Publishers, pp. 321–325.
17. Andreesen, M. 1993. "NCSA Mosaic Technical Summary," National Center for Supercomputing Applications, Software Development Group, University of Illinois, Urbana, Ill.
18. Myers, Brad A., and Mary Beth Rosson. 1992. "Survey on User Interface Programming," *Proceedings SIGCHI'92: Human Factors in Computing Systems*. Monterey, Calif., May 3–7, p. 195.
19. Hix, D., and H.R. Hartson. 1993. *Developing User Interfaces: Ensuring Usability through Product and Process*, Wiley.

## 29

# Proposal for an Evaluation of Health Care Applications on the NII

*Joseph Gitlin*  
*Johns Hopkins University*

### STATEMENT OF THE PROBLEM AND ISSUES

Telemedicine is being advocated as a process of delivering health care to all segments of the population with the potential of reducing the cost of care while maintaining its quality. However, little is known about the efficacy and cost-effectiveness of the technology in routine diagnostic and therapeutic practice. A well-designed evaluation based on stringent criteria is needed to determine the merits of telemedicine in utilizing the national information infrastructure (NII) in the environment of health care reform. Within this context, some of the more important concerns related to the realization of telemedicine are the following:

- The lack of "bandwidth on demand" to provide data rates when they are needed at affordable costs to the health care community;
- The lack of availability of high-density, low-cost, digital storage and related software for efficient access by authorized users; and
- The lack of standards and interfaces for both health care data acquisition and for the effective use of such information. This is particularly applicable to integration with heterogeneous legacy systems used by a wide variety of health care providers.

### PROJECTIONS REGARDING THE PROBLEM AND RELATED ISSUES

Assuming the telecommunication "tariff" issue under the jurisdiction of federal, state, and local authorities can be resolved, it is anticipated that the technical barriers to bandwidth on demand will be overcome in the next 5 to 7 years. The efforts under way to develop a reliable storage and retrieval system that is suitable for medical images and other health care data should be realized before the year 2000. Though improved interfaces and system integration techniques are expected to be available shortly, the accommodation of heterogeneous legacy systems may be delayed by economic and cultural factors for several years.

### STATUS OF KEY DEVELOPMENTS

Recent developments in information technology and the recognition of the need for reform provide a unique opportunity for health care decisionmakers to capitalize on the availability of the NII. If, for example, medical imaging advances are to be available to all patients regardless of situation or geography, the storage, transmission, and retrieval of large volumes of data must be accommodated in all areas of the country. Also, access by secondary users to clinical information for teaching, research, and management (within appropriate security and privacy restrictions) requires that the information be readily available to medical students, research investigators, and health care policymakers.



Key developments in telecommunications essential to meeting the demand generated by health care activities include the availability of high-speed communication networks in sparsely populated areas. The economic and cultural aspects of the telemedicine applications can then be identified, demonstrated, and evaluated. Technology related to high-density storage and retrieval of medical images and related patient data is immature. Much work remains to be done to achieve reliable and cost-effective systems that will support patient care, medical research and education, and health care administration. The proliferation of personal computers and the increase in computer literacy are major factors in user acceptance of telemedicine and related technology.

### INTERACTION BETWEEN TECHNICAL AND NONTECHNICAL FACTORS

The regulatory authority of federal, state, and local agencies to set tariffs has resulted in barriers to the economic reality of telemedicine. If the quality of health care is to be maintained in a cost-effective manner using the NII, clinical data must be transmitted promptly within cost constraints.

Other legal and regulatory issues that must be addressed include the privacy of patients and health care providers, and the security of data against unauthorized access. Many questions need to be answered regarding the "ownership" of medical information and the responsibility for retention of medical records. The differing state medical licensure requirements must be rationalized to permit access, when needed, to specialists across state boundaries, and malpractice regulations need to be modified to eliminate unnecessary medical procedures that are performed solely to reduce the threat of litigation. In the area of administration, the adoption of a uniform claims data set would substantially reduce current processing activities related to reimbursement.

Though health care costs in the United States amount to approximately 15 percent of the gross domestic product, health care information requirements alone cannot support the development and deployment of the NII. However, health care is an important contributor to the information community and is one of many large economic segments that must be included in the utilization of the NII. If the cost of the NII is shared among a large number of major segments of the economy, the application of telecommunications will be facilitated. Further advances in storage and retrieval technology are largely dependent on government agencies and sectors of the economy other than health care. The special requirements of health care can then be met by modifying the basic developments designed to meet other needs.

The current trend of health care reform emphasizes the restructuring of the delivery system toward managed care corporations. The driving force behind this restructuring is the recognized need for cost containment. Today, the decisions to adopt new technology for use in health care are predominantly made by corporate managers rather than by individual practitioners. Since health care is a labor intensive activity, in this new climate technology that increases efficiency is more favorably received.

Increased access to quality care by patients regardless of situation or geography is the primary justification for telemedicine and for health care reform. To some extent, this implies "patient acceptance" of the NII and related technology; however, health care provider acceptance is pivotal to adoption of telemedicine in practice. The acceptance of new technology requires many cultural and procedural changes by physicians, nurses, and allied health care workers. These changes have already occurred in health care financial activities such as billing and reimbursement and in medical research but are lagging in patient care delivery functions.

### CONTINGENCIES AND UNCERTAINTIES

The investment in and deployment of new technologies applicable to health care are partly dependent on the success of health care reform. "Many of the political imperatives driving telemedicine derive from the anticipated use of managed care incentives to provide accessible low-cost health care to all Americans."<sup>1</sup> It is expected that the information infrastructure will be deployed because of impetus by government agencies and industries other than health care. The implementation of the NII for the delivery of health care is dependent on the cost-effectiveness of the technology as perceived by the decision makers within health care reform.

Other contingencies and uncertainties related to investment and deployment of the NII and its use in health care include the following:

- Availability of appropriate network connections throughout the country;
- Willingness by health care providers to share new technologies;
- Resolution of security and privacy issues;
- Cost-effective solutions to integration of legacy systems;
- Resolution of telecommunications tariff issues, particularly bandwidth on demand;
- Adoption of a uniform claims data set;
- Acceptance by physicians of "compressed" data, especially in medical imaging; and
- Completion of the comprehensive computer-based patient record.

### **KEY APPLICATIONS, ENABLING TECHNOLOGIES, AND CAPABILITIES**

Several recent technological developments make it possible for health care to take advantage of many of the capabilities offered by the NII. The following are among the more important developments available to health care providers:

- A range of workstations accommodates the spectrum of needs of health care providers, particularly the high-resolution and luminance requirements of radiologists, as well as the needs of other specialists and primary care physicians.
- Several standards have been developed for health care data interchange. These include the Digital Imaging Communication in Medicine (DICOM), HL7, and MEDIX P1157. However, it is necessary to identify and develop other standards that will facilitate further use of the NII by health care providers.
- Preliminary results of the large-scale effort to develop a comprehensive computer-based patient record are available, and there is momentum to complete the task.
- Various technical approaches have made "electronic signature" available. However, some legal questions remain to be answered before broad acceptance can be achieved.
- Recent advances in "speech recognition" technology are most important to health care provider input to medical records. This is especially applicable to medical imaging, where interpretation of examinations is basic to the specialty.
- The use of compression algorithms to decrease data volume has proved cost-effective and reduces transmission time. However, there is concern about the loss of pertinent information when "destructive" compression is applied. This is especially true of medical images, where radiologists require all of the original data for detecting subtle abnormalities.

### **CLASSES OF USERS**

In health care, there are several types of users to be considered when access to the NII is planned. Among those users who are relatively easy to accommodate and who may adapt quickly to the new technology are the following:

- Computer-literate health care providers, researchers, educators, and students in academic settings;
- Computer-literate health care providers and other personnel in managed care settings; and
- Government and insurance agencies concerned with reimbursement.

There are also a large number of potential health care users of the NII that will require substantial training and education, as well as appropriate hardware and software to be capable of using the infrastructure:

- Older health care providers in small groups and solo practices, and
- Health care providers in remote locations.

If the technology proves to be cost effective, the increase in managed care organizations should expedite the use of the NII for patient care, medical research and education, and health care administration. Though the primary factors relative to ramp-up expectations and determinants are outside the health care environment, government actions regarding Medicare and Medicaid will affect investments in technologies intended for health care markets.

### **PUBLIC AND PRIVATE SECTOR RESPONSES**

Since health care is uniquely the purview of both the public and private sectors, such an application of NII technology is affected by government and industry. Acceptance by health care providers and related organizations in terms of cost-effectiveness and utility is affected by both political and economic considerations.

### **RECOMMENDATIONS**

A comprehensive evaluation based on a realistic demonstration should be conducted to identify the factors related to the utilization of the NII by health care providers and related organizations. The evaluation should be based on stringent criteria that focus primarily on patient care issues such as quality and access, and that measure selected key parameters related to technology, economics, and legal/regulatory and social/cultural factors. The project will require the participation of industry, academia, and government in cooperation with health care providers to develop the evaluation criteria, design the study, and conduct the demonstration of cost-effective systems that will support telemedicine, medical research and education, and health care administration. It is intended that this white paper will provoke serious consideration of health care applications on the NII.

### **REFERENCE**

1. Perednia, D.A., and A. Allen. 1995. "Telemedicine Technology and Clinical Applications," *Journal of the American Medical Association* 273:483–488.

## 30

### The Internet—A Model: Thoughts on the Five-Year Outlook

*Ross Glatzer*

*Prodigy Services (retired)*

We've heard a lot of statistics about the mushrooming growth of the "Net." I believe it's much more than a fad. What I see on my radar is a stunning social change as important as anything in this century. You can see signs of it already.

- One example is in the area of religion. Devout people are increasingly sharing their faith online. They hold religious services, complete with sermons and music. There are online support groups for every need. By the turn of the century, we may have tithing online by way of credit cards.
- In politics, candidates now routinely debate issues online. Citizens have been sending e-mail to their representatives for a long time. Will it be that many years before we have online voting? I think it could be available in a decade. The only barriers are social, not technological.
- Consumer access to online medical information is growing rapidly. How long before you can buy prescriptions online? Make medical appointments? Even get consultations?

Clearly, the online medium is taking to a much higher level what started with radio in the 1920s—that is, the dispersal of knowledge, culture, and democracy more directly into citizens' homes. But, unlike radio—or, later, television—now it's two-way. And therein lies the profound significance of this trend.

I'm going to go out on a limb and say that I think the online medium will become a mass phenomenon in 5 years, or very soon thereafter. No, I don't predict that 100 million Americans will be sitting at PC keyboards every night, squinting at their screens, typing in URLs, and watching the hourglass while waiting for a grainy picture of Jerry Seinfeld to paint—ever so slowly. There's just no way, at present course and speed, that the online experience will become the ubiquitous social glue that television and movies are today. Online resources are too hard to use and too expensive.

So why did I say this will be a mass medium? Because I don't think it's a linear world, and we're not going to continue at present course and speed. There's too much at stake and too many smart people seeking breakthroughs. In other words, I'm factoring in major advances—yes, advances in the underlying technology and infrastructure, but also some major breakthroughs in the human interface:

- Something that will make going online as simple as hitting the power switch on a TV set.
- Something that will replace keyboards.
- Something that will improve the resolution of displays by a couple of orders of magnitude.
- Something that will make navigation as intuitive as dialing grandma's phone number.

It's going to happen. Just look at who's working on it—companies like AT&T, Microsoft, MCI, IBM, the broadcast and cable networks, the Baby Bells, and many others, plus a fair number of smart people in their garages or home offices.

But another type of breakthrough could be the emergence of our industry's Holy Grail: the "killer application." It's happened before without any change in technology. Sony came out with the Betamax in 1975. Everyone thought then that the "killer app" for VCRs would be time shifting. And VCR usage rose steadily—but not spectacularly—toward 10 percent by 1982. That growth curve closely parallels the experience to date of online services. But something changed in 1982. Suddenly, mom-and-pop video stores sprang up in every town. The "killer app" turned out to be movie rentals. Within the next 2 years, home penetration of VCRs approached 50 percent. So we could be nearing a flash point for the online medium. All we need to do is make it easier to use and figure out the "killer app." In my opinion, the "killer app" is communications.

Sure, I know, online services have had communications from Day One. And these services are still a niche product. But consider this: most online services to date have been built around information, not around communications. That's turned out to be backwards. It's backwards because all people communicate; but only a small percentage of the population really cares about any one topic of information—O.J. Simpson excluded!

Consider this. CompuServe has 2,000 categories of information. But how many of those does the average member use? I know from my experience that it's seldom more than four or five applications. At Prodigy, there's a terrific feature called Strategic Investor, yet only a small fraction of members subscribe.

Given these facts, anyone starting an online service today would probably be advised to build outward from a rock-solid communications core, optimizing everything for subscriber exchanges. And you wouldn't just have bulletin boards, chat, and e-mail. You'd have instant messages, 3-D chat, and easily attached files for sound photos, video, and graphics. You'd also let subscribers create their own home pages on the Web, where they could talk about themselves—even show their cars. Prodigy will do this shortly, and the other services will quickly follow.

After you'd established a firm core of communications, you'd want to start adding information and transactions to it. But the information and transactions would be tightly integrated with the communications. That's important, because getting around online services today is like being the guy who explains where he's calling from by saying he's in a phone booth at the corner of WALK and DON'T WALK.

The leaders in online services will be those who can best integrate communications, information, and transactions. They will build on a core of communications to create communities of interest.

Let me give you an example. A company like Ford spends a fair amount of money each year putting information online. So as a user, I can navigate to Ford's advertising section and read about its new cars. Even see photos of them. OK, fine. But how often am I going to come back to see the same information? Now, if I'm interested in cars, maybe I'll log on to a car enthusiasts' bulletin board. But after a while, I'll get bored talking to the same old regulars. On another day, I might order a subscription to *Car and Driver* magazine right from my computer. The problem is, all these are discrete activities that I carry out as an individual. They don't create much excitement. They don't involve me very much.

But what if an online service creates communities of interest built around world-class communication functionality? From a single screen, or with hyperlinks, I can see Ford's cars, chat with other car enthusiasts about them, debate with *Car and Driver's* editors, download model specs, check the archives of the *Detroit Free Press* for an article about Carroll Shelby, ask Shelby a question, place a classified ad to sell my '67 Mustang, look up the price of Ford stocks, buy 100 shares, and send an instant message to a friend's beeper urging him to get online and join a discussion group.

You can do all or most of this today on the online services. But no one has done a very good job of integrating it to create true communities of interest. That's what I think will make online a ubiquitous medium. And if it is, can advertisers be far behind?

Advertising on the Web is a tricky business, however. Very few advertisers really understand it. The environment differs from other media and will become more different over the next few years. It's an environment where the revenue model will become pay-per-view—or pay-per-minute.

Today, for most users, the Internet is essentially untimed. And, as a practical matter, so are the commercial online services, since the majority of their users stay within the flat-rate time limits that go with their

monthly subscriptions. But on the Internet, people will not want to subscribe to multiple features. They'll want to dip in and out. And the information providers will oblige by making everything timed, including ads. Very few companies will offer anything free on the Internet.

So-called "free" applications on the Web today, things like online newspapers and Time Warner's Pathfinder, are really just experiments—beta tests, if you will. All these will be timed when masses of consumers arrive. The implications for advertising and transactions are profound. As commercial applications migrate to the Net, they must change dramatically. Even the best of the online service advertisers—and some are very good—must be better on the Net, because most of their customers will be paying for the time to view their applications. An advertiser or merchant on a commercial online service today pays "rent" to be in the service's "mall." On the Net, advertisers will pay for the number of "footsteps" across their thresholds—and the amount of time those feet stay in their stores. This puts a premium on closing a high percentage of sales. You must be a consummate retailer to prosper.

And it's not easy. You can't just show up and be successful, because your online customers are far more discerning than the typical catalog or retail customer. They simply will not tolerate anything but the highest level of service. Incredible service explains why a few companies have prospered in an online arena where many have not.

Take PC Flowers, for example. In the florist business, a 5 percent error rate is considered normal. At PC Flowers, it is one-tenth of that. And they have a more complex business than any florist. Needless to say, PC Flowers has a fabulous back-end system, one that can easily handle more than \$1 million in business on Mother's Day. And on the rare occasion where they do slip up, they move heaven and earth to make it up to the customer ... and the customer's customer. That's expensive. It's also profitable.

The PC Financial Network is another online success story—\$2.5 billion in stock trades last year with an average trade of \$6,000. But to do that accurately, they've invested in 80 full-time employees on the back end.

So while it's easy to set up a business on the World Wide Web, it's not going to be so easy to make it prosper in the long term. You need to make major investments. Successful advertisers and merchants understand this. They understand that information alone doesn't cut it. It takes true relationship marketing. And the relationship has to be based on communications.

If the commercial online services hope to retain these advertisers—and with them, a key part of their revenue base—they will have to change. One way they can change is to become the preferred access providers to the World Wide Web. In doing so, I think they will eventually have to give up the idea of charging a monthly subscription for a defined collection of content.

Users will pay a flat rate for access, billing, customer services, etc. But they will select content seamlessly from anywhere on the Net without caring or needing to know who is integrating it. The online services will need to establish a revenue model by which they receive a fraction of a penny every time one of their users clicks on an advertiser's Net site. Thus, Prodigy's home page might feature an ad pointing users to the home page of PC Flowers. Netcom's home page might include a paid icon pointing users to America Online. All kinds of synergistic relationships are possible.

To summarize, here's what I see happening by 2000:

- The Internet, especially the World Wide Web, will become the communications/data transfer standard. It will be the backbone for advanced interactive services, including TV shopping.
- There will be breakthroughs in ease of use and input/output technology.
- Commercial activity will be an important part of the Web, but not as important as some people predict.
- Advertising will be substantially more sophisticated and will be communications centered. A click-stream ad revenue model will replace display revenue.
- There will be some consolidation of the online services as we know them, although many companies will offer Net access.
- The online services will be forced to become one with the Internet. Their ability to survive as distinct entities and to prosper will depend in part on their success in creating communities of interest.

- For at least a few years, the online services will retain uniqueness in terms of structure, context, back-end transactions, billing, customer service, marketing segmentation, technology innovation, and distribution. But even these advantages will fade over time.
- Pricing pressures will be brutal. Consider that only a few years ago, CompuServe could charge \$22.50 an hour for access at 9,600 baud. Today, it charges \$4.80 while others charge as little as \$1 for access at 14,400 baud. This trend will continue, which is great for the consumer, but ominous for the online services.
- More than 50 percent of American households will be online. But many of them will use simpler access devices than today's computers.
- Like most successful trends, all these changes will be driven by the competitive market.

And finally, if all this doesn't happen in the next 5 years, just remember what Mahatma Gandhi said: "There is more to life than increasing its speed."

## 31

# The Economics of Layered Networks

*Jiong Gong and Padmanabhan Srinagesh*  
*Bell Communications Research Inc.*

### STATEMENT OF THE PROBLEM

The creation of a national information infrastructure (NII) will require large investments in network facilities and services, computing hardware and software, information appliances, training, and other areas. The investment required to provide all businesses and households with broadband access to the NII is estimated to be in the hundreds of billions of dollars; large investments in the other components of the NII will also be required. In *The National Information Infrastructure: Agenda for Action*, the Clinton Administration states, "The private sector will lead the deployment of the NII."

What architectural and economic framework should guide the private sector's investment and deployment decisions? Is the Open Data Network (ODN) described by the NRENAISSANCE Committee<sup>1</sup> an appropriate economic framework for the communications infrastructure that will support NII applications? A key component of the ODN is the unbundled bearer service, defined to be "an abstract bit-level transport service" available at different qualities of service appropriate for the range of NII applications. The committee states that "bearer services are not part of the ODN unless they can be priced separately from the higher-level services" (p. 52). The rationale for this requirement is that it is "in the interest of a free market for entry at various levels" (p. 52).

What effect will the unbundled bearer service proposed by the NRENAISSANCE Committee have on the investment incentives of network providers in the private sector? Will carriers that invest in long-lived assets be given a fair opportunity to recover their costs? This paper provides a preliminary discussion of the economics of an unbundled bearer service.

### BACKGROUND: CONVERGENCE AND EMERGING COMPETITION

Technological advances are rapidly blurring traditional industry boundaries and enabling competition between firms that did not previously compete with one another. For example, cable TV providers in the United Kingdom (some of which are now partly owned by U.S. telecommunications firms) have been allowed to offer telephone service to their subscribers since 1981 and currently serve more than 500,000 homes.<sup>2</sup> Numerous telephone companies in the United States are currently testing the delivery of video services to households over their networks. In addition, new firms have recently entered markets that were not subject to major inroads in the past. Competitive access providers (CAPs) have begun to serve business customers in the central business districts of many large cities in competition with local exchange carriers (LECs), and direct broadcast satellite services have begun to compete with cable providers. In sum, new entrants are using new technologies to compete with incumbents, and incumbents in previously separate industries are beginning to compete with one another.



## ANALYSIS

### Theoretical Approaches to the Economics of Pricing Under Differing Market Structures

In a pure monopoly, a variety of price structures may be consistent with cost recovery, and the firm (or regulators) may be able to select price structures that promote political or social goals such as universal service or unbundling of raw transport. In a competitive market, this flexibility may not exist. Under perfect competition (which assumes no barriers to entry and many small firms), the price per unit will be equal to the cost per unit (where costs are defined to include the opportunity costs of all resources, including capital, that are used in production). There is no pricing flexibility. When neither of these pure market forms exist, economic theory does not provide any general conclusions regarding equilibrium price structures or industry boundaries. While substantial progress has been made in developing game theory and its application to oligopoly,<sup>3</sup> no completely general results on pricing are available. This is particularly true in the dynamic context where interdependencies between current and future decisions are explicitly considered. Some theoretical work in this area is summarized in Shapiro.<sup>4</sup> An important result in game theory asserts that no general rules can be developed: "The best known result about repeated games is the well-known 'folk theorem.' This theorem asserts that if the game is repeated infinitely often and players are sufficiently patient, then 'virtually anything' is an equilibrium outcome."<sup>5</sup> Modeling based on the specific features of the telecommunications industry may therefore be a more promising research strategy.

The economic analysis of competition among network service providers (NSPs) is further complicated by the presence of externalities and excess capacity. Call externalities arise because every communication involves at least two parties: the originator(s) and the receiver(s). Benefits (possibly negative) are obtained by all participants in a call, but usually only one of the participants is billed for the call. A decision by one person to call another can generate an uncompensated benefit for the called party, creating a call externality. Network externalities arise because the private benefit to any one individual of joining a network, as measured by the value he places on communicating with others, is not equal to the social benefits of his joining the network, which would include the benefits to all other subscribers of communicating with him. Again, the subscription decision creates benefits that are not compensated through the market mechanism. It has been argued that the prices chosen by competitive markets are not economically efficient (in the sense of maximizing aggregate consumer and producer benefits) when externalities are present.<sup>6</sup>

It has also been argued that "[i]ndustries with network externalities exhibit positive critical mass—i.e., networks of small sizes are not observed at any price."<sup>7</sup> The consequent need to build large networks, together with the high cost of network construction (estimated by some to be \$13,000 to \$18,000 per mile for cable systems<sup>8</sup>), implies the need for large investments in long-lived facilities. The major cost of constructing fiber optic links is in the trenching and labor cost of installation. The cost of the fiber is a relatively small proportion of the total cost of construction and installation. It is therefore common practice to install "excess" fiber. According to the Federal Communications Commission, between 40 percent and 50 percent of the fiber installed by the typical interexchange carriers is "dark"; the lasers and electronics required for transmission are not in place. The comparable number for the major local operating companies is between 50 percent and 80 percent. The presence of excess capacity in one important input is a further complicating factor affecting equilibrium prices and industry structure.

To summarize: a full economic model of the networking infrastructure that supports the NII would need to account for at least the following features:

- Oligopolistic competition among a few large companies that invest in the underlying physical communications infrastructure;
- Network and call externalities at the virtual network level; and
- Large sunk costs and excess capacity in underlying transmission links.

An analysis of the optimal industry structure is well beyond the scope of this paper; it is a promising area for future research. This paper focuses on implications of two of these issues (oligopolistic competition and sunk cost with excess capacity) for industry structure and the unbundled bearer service. The Internet is used as a case study to illustrate trends and provides a factual background for future analyses. This paper provides a description of the current Internet industry structure, current Internet approaches to bundling/unbundling and reselling, some recent examples of the difficulties raised by resale in other communications markets, and the increasing use of long-term contracts between customers and their service providers. The implications for the unbundled bearer service are drawn.

### Layered Networks

The Internet is a virtual network that is built on top of facilities and services provided by telecommunications carriers. Until recently, Internet service providers (ISPs) located routers at their network nodes and interconnected these nodes (redundantly) with point-to-point private lines leased from telecommunications companies. More recently, some ISPs have been moving from a private line infrastructure to fast packet services such as frame relay, switched multimegabit data service (SMDS), and asynchronous transfer mode (ATM) service. Specifically, among the providers with national backbones,

- PSI runs its IP services over its frame relay network, which is run over its ATM network, which in turn is run over point-to-point circuits leased from five carriers;
- AlterNet runs part of its IP network over an ATM backbone leased from MFS and Wiltel;
- ANS's backbone consists of DS3 links leased from MCI; and
- SprintLink's backbone consists of its own DS3 facilities.

CERF net, a regional network based in San Diego, uses SMDS service obtained from Pacific Bell to connect its backbone nodes together.<sup>9</sup>

These examples reveal a variety of technical approaches to the provision of IP transport. They also show different degrees of vertical integration, with Sprint the most integrated and AlterNet the least integrated ISP in the group listed above. The variety of organizational forms in use raises the following question: Can ISPs with varying degrees of integration coexist in an industry equilibrium, or are there definite cost advantages that will lead to only one kind of firm surviving in equilibrium? The answer to this question hinges on the relative cost structures of integrated and unintegrated firms. The costs of integrated firms depend on the costs of producing the underlying transport fabric on which IP transport rides. The cost structures of unintegrated firms are determined in large part by the prices they pay for transport services (such as ATM and DS3 services) obtained from telecommunications carriers. These prices, in turn, are determined by market forces. More generally, the layered structure of data communications services leads to a recursive relationship in which the cost structure of services provided in any layer is determined by prices charged by providers one layer below. In this layered structure, a logical starting point for analysis is the lowest layer: the point-to-point links on which a variety of fast packet services ride.

### Competition at the Bottom Layer

For illustrative purposes, consider a common set of services underlying the Internet today. At the very bottom of the hierarchy, physical resources are used to construct the links and switches or multiplexers that create point-to-point channels. In the emerging digital environment, time division multiplexing (TDM) in the digital telephone system creates the channels out of long-lived inputs (including fiber optic cables). The sunk costs are substantial.

There are at least four network service providers with national fiber optic networks that serve all major city-pairs.<sup>10</sup> Each of these providers has invested in the fiber and electronics required to deliver point-to-point

channel services, and, as was stated above, each has substantial excess capacity. The cost structure of providing channels includes high sunk costs of construction, the costs of lasers and electronics needed to light the fiber, costs of switching, the high costs of customer acquisition (marketing and sales), costs associated with turning service on and off (provisioning, credit checks, setting up a billing account, and so on), costs of maintaining and monitoring the network to assure that customer expectations for service are met, costs of terminating customers, and general administrative costs. The incremental cost of carrying traffic is zero, as long as there is excess capacity.

If all owners of national fiber optic facilities produced an undifferentiated product (point-to-point channels) and competed solely on price, economic theory predicts that they would soon go out of business: "With equally efficient firms, constant marginal costs, and homogeneous products, the only Nash equilibrium in prices, i.e., Bertrand equilibrium, is for each firm to price at marginal cost."<sup>11</sup> If this theory is correct,<sup>12</sup> firms could recover the one-time costs of service activation and deactivation through a nonrecurring service charge, and recover ongoing customer support costs by billing for assistance, but they would not be able to recover their sunk cost. Industry leaders seem to be aware of this possibility. As John Malone, CEO of TCI, stated, "We'll end up with a much lower marginal cost structure and that will allow us to underprice our competitors."<sup>13</sup>

The history of leased line prices in recent years does reveal a strong downward trend in prices. According to *Business Week*,<sup>14</sup> private line prices have fallen by 80 percent between 1989 and 1994, and this is consistent with Bertrand competition. During the same period there has been a dramatic increase in the use of term and volume discounts. AT&T offers customers a standard month-to-month tariff for T1 service and charges a nonrecurring fee, a fixed monthly fee, and a monthly rate per mile. Customers who are willing to sign a 5-year contract and commit to spending \$1 million per month are offered a discount of 57 percent off the standard month-to-month rates. Smaller discounts apply to customers who choose shorter terms and lower commitment volumes: a 1-year term commitment to spend \$2,000 per month obtains a discount of 18 percent. The overall trend toward lower prices masks a more complex reality. "There are two types of tariffs: 'front of the book' rates, which are paid by smaller and uninformed large customers, and 'back of the book' rates, which are offered to the customers who are ready to defect to another carrier and to customers who know enough to ask for them. The 'front of the book' rates continue their relentless 5 to 7 percent annual increases."<sup>15</sup> In 1994 AT&T filed over 1,200 special contracts, and MCI filed over 400.<sup>16</sup>

There are some theoretical approaches that address the issues discussed above. Williamson's discussion of nonstandard commercial contracting<sup>17</sup> as a means for sharing risk between the producer and consumers is relevant to the term commitments described above. In addition to risk reduction, long-term contracts reduce customer churn, which often ranges from 20 to 50 percent per year in competitive telecommunications markets.<sup>18</sup> As service activation and termination costs can be high, reduction of churn can be an effective cost-saving measure.

There appears to be an empirical trend toward term/volume commitments that encourage consumers of private line services to establish an exclusive, long-term relationship with a single carrier. There is little published information on long distance fast packet prices. According to one source, none of the long distance carriers or enhanced service providers (e.g., CompuServe) tariff their frame relay offerings. Some intra-LATA tariffs filed by local exchange carriers do offer term and volume (per PVC) discounts, and the economic forces that give rise to term/volume commitments for private lines have probably resulted in term/volume commitments for long-distance, fast-packet services as well.

### Competition among Internet Service Providers

The effect of term/volume commitments in private lines and fast packet services affects the cost structures of ISPs that do not own their own transport infrastructure. It may be expected that large ISPs that lease their transport infrastructures will sign multiyear contracts, possibly on an exclusive basis, with a single carrier. These providers will then have sunk costs, as they will have minimum payments due for a fixed period to their carriers. Competition at this level will then be similar to competition at the lower level, and we may expect to see term/volume contracts emerge in the Internet. A quick survey of Internet sites shows this to be the case. For

example, in January 1995, AlterNet offered customers with a T1 connection a 10 percent discount if they committed to a 2-year term. Global Connect, an ISP in Virginia, offers customers an annual rate that is 10 times the monthly rate, amounting to a 17 percent discount for a 1-year commitment. There are many other examples of this sort.

The Internet is beginning to resemble the private line market in one other aspect: prices are increasingly being viewed as proprietary. ISPs that used to post prices on their ftp or Web servers now ask potential customers to call for quotes. Presumably, prices are determined after negotiation. This development mirrors the practice of long-distance carriers to use special contracts that are not offered on an open basis at the "front of the book" but are hidden at the back.

### Economics of Resale

Kellogg, Thorne, and Huber<sup>19</sup> describe the history of the Federal Communication Commission's decision on resale and shared use. Noam<sup>20</sup> analyzed the impact of competition between common carriers and contract carriers (including systems integrators and resellers) and concluded that common carriage cannot survive the competitive struggle. Recent events lend some credence to this view. According to one recent study, "resold long distance services will constitute an increasing portion of the total switched services revenue in coming years, growing at a compound annual growth rate of 31 percent from 1993 to 1995.... The number is expected to rise to \$11.6 billion, or 19.2 percent of the estimated total switched services market in 1995."<sup>21</sup> The growth of resale in the cellular market suggests that there are equally attractive resale opportunities in this market.<sup>22</sup> In the Internet, some ISPs charge resellers a higher price than they charge their own customers.<sup>23</sup> Other ISPs, such as SprintLink, make no distinction between resellers and end users. Facilities-based carriers have had a rocky relationship with resellers, and courts have often been resorted to by both carriers and resellers.<sup>24</sup>

The pricing model that is emerging appears to resemble rental arrangements in the real estate market. In the New York City area, low-quality hotel rooms are available for about \$20 per hour. Far better hotel rooms are available for \$200 per day (which is a large discount of 24 times \$20). Roomy apartments are available for monthly rentals at much less than 30 days times \$200/day. And \$6,000 per month can be used to buy luxury apartments with a 30-year mortgage. Term commitments are rewarded in the real estate market, where sunk costs and excess capacity are (currently) quite common. The telecommunications industry appears to be moving in the same direction. Contracts are not limited to five-year terms; MFS and SNET recently signed a 20-year contract under which MFS will lease fiber from SNET,<sup>25</sup> and Bell Atlantic has a 25-year contract with the Pentagon.<sup>26</sup> The term structure of contracts is an important area for empirical and theoretical research.

### Implications for Unbundled Bearer Services

Unbundled bearer services have much in common with common carriage: both approaches facilitate greater competition at higher layers (in content, with common carriage, and in enhanced services of all types with the bearer service). The dilemma facing policymakers is that, if Noam is right, competition in an undifferentiated commodity at the lower level may not be feasible. In his words: "The long-term result might be a gradual disinvestment in networks, the reestablishment of monopoly, or price cartels, and oligopolistic pricing."<sup>27</sup> Thus policies promoting competition in the provision of unbundled bearer services among owners of physical networks may ultimately fail. The market may be moving toward contract carriage based on term/volume commitments and increasing efforts at differentiation, and away from the ideal of an unbundled bearer service. Should unbundled bearer services be aligned with this trend by being defined as a spectrum of term/volume contracts? The competitive mode that is emerging is quite complex, and the effects of unbundling in this environment are hard to predict.

## CONCLUSIONS

This paper does not suggest specific architectures or policies for the emerging NII. It identifies some difficult economic problems that may need to be addressed. These are the familiar ones related to resale and interconnection, with the added complication of competition among multiple owners of geographically coextensive physical networks. This paper has provided references to recent developments in telecommunications markets and identified strands in the economics literature that are relevant to the central issues raised by the bearer service.

There is an urgent need for a clearer economic analysis of these issues, and it is critical that the analysis pay close attention to the realities of competition and evolving competitive strategy. Three specific areas appear particularly promising:

- Empirical analysis of evolving price structures that quantifies the movement from pricing by the minute (the original Message Toll Service) to pricing by the decade (contract tariffs);
- Game theoretic models of competition in long-term contracts with sunk costs; and
- Experimental approaches to network economics.<sup>28</sup>

## NOTES

1. Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C.
2. "Cable TV Moves into Telecom Markets," *Business Communications Review*, November, 1994, pp. 43–48.
3. A recent text is Fudenberg, Drew, and Jean Tirole, 1992, *Game Theory*, MIT Press, Cambridge, Mass.
4. Shapiro, Carl. 1988. "Theories of Oligopoly Behavior," Chapter 6 in *Handbook of Industrial Organization*, Richard Schmalensee and Robert Willig (eds.). North Holland, Amsterdam, pp. 400–407.
5. Fudenberg, Drew, and Jean Tirole. 1988. "Noncooperative Game Theory," Chapter 5 in *Handbook of Industrial Organization*, Richard Schmalensee and Robert Willig (eds.). North Holland, Amsterdam, p. 279.
6. These externalities are discussed in some detail in Mitchell, Bridger, and Ingo Vogelsang, 1991, *Telecommunications Pricing: Theory and Practice*, Cambridge University Press, Cambridge, England, pp. 55–61.
7. Economides, Nicholas, and Charles Himmelberg. 1994. "Critical Mass and Network Size," paper presented at the Twenty-second Annual Telecommunications Policy Research Conference, October.
8. Yokell, Larry J. 1994. "Cable TV Moves into Telecom Markets," *Business Communication Review*, November, pp. 43–48.
9. A fuller description of these networks can be obtained from their Web pages. The URLs are <http://www.psi.com>, <http://www.alter.net>, <http://www.sprint.com>, and <http://www.cerf.net>.
10. These networks are shown on a pull-out map in *Forbes ASAP*, February 27, 1995.
11. Shapiro, Carl. 1988. "Theories of Oligopoly Behavior," Chapter 6 in *Handbook of Industrial Organization*, Richard Schmalensee and Robert Willig (eds.). North Holland, Amsterdam.
12. An example of an alternative theoretical approach is found in Sharkey, William, and David Sibley, 1993, "A Bertrand Model of Pricing and Entry," *Economic Letters*, pp. 199–206.
13. *Business Week*. 1995. "Brave Talk from the Foxhole," April 10, p. 60.
14. *Business Week*. 1994. "Dangerous Living in Telecom's Top Tier," September 12, p. 90.
15. Hills, Michael T. 1995. "Carrier Pricing Increases Continue," *Business Communications Review*, February, p. 32.
16. Ibid.
17. Williamson, O.E. 1988. "Transaction Cost Economics," Chapter 3 in *Handbook of Industrial Organization*, Richard Schmalensee and Robert Willig (eds.). North Holland, Amsterdam, pp. 159–161.
18. See FCC Docket 93-197, Report and Order, 1.12.95, page 8 for statistics on AT&T's churn in the business market. See also Ko, David, and Michael Gell. n.d. "Cable Franchise Growth in the U.K.," memo, University of Oxford, for churn in the U.K. cable market.
19. Schmalensee and Willig (eds.), op. cit., pp. 610–614.
20. Noam, Eli. 1994. "Beyond Liberalization II: The Impending Doom of Common Carriage," *Telecommunications Policy*, pp. 435–452.
21. *Telecommunications Alert* 13(2):6.

22. *Forbes*. 1995. "Restless in Seattle," March 27, p. 72.
23. For example, AlterNet's policy in January 1995 was as follows: "Because wholesale customers use more of our backbone facilities and because they also place greater demands on our staff, we charge more for our wholesale services."
24. For example, see *Telecommunications Reports*. 1994. "Oregon Jury Decides Against AT&T in Reseller Case," July 4, p. 34, and "AT&T Sues Reseller for Unauthorized Trademark Use," November 7, p. 26.
25. *Telco Business Report*, February 14, 1994, p. 4.
26. Ray Smith, CEO of Bell Atlantic, in an interview in *Wired*, February 1995, p. 113.
27. Noam, *op. cit.*, p. 447.
28. Plott, Charles, Alexandre Sugiyama, and Gilad Elbaz. 1994. "Economics of Scale, Natural Monopoly, and Imperfect Competition in an Experimental Market," *Southern Economic Journal*, October, pp. 261–287.

32

## The Fiber-Optic Challenge of Information Infrastructures

*P.E. Green, Jr.*

*IBM T.J. Watson Research Center*

This paper discusses the role of optical fiber as the one physical medium upon which it will be possible to base national and global infrastructures that will handle the growing demands of bandwidth to the desktop in a post-2000 developed society.

A number of individual applications today demand large bit rate per user, such as supercomputer interconnection, remote site backup for large computer centers, and digital video production and distribution, but these are isolated niches today. The best gauge of the need for the infrastructure to supply large amounts of bandwidth to individual users is probably to be found in the phenomena associated with the use of the Internet for graphics-based or multimedia applications.

Just as the use of noncommunicating computers was made much easier by the emergence of the icon- and mouse-based graphical user interface of the Macintosh, Windows and OS2, the same thing can be observed for communicating computers with the World Wide Web. The modality in which most users want to interact with distributed processing capability (measured in millions of instructions per second [MIPs]) is the same as it has always been with local MIPs: they want to point, click, and have an instant response. They will in fact want to have a response time from some remote source on any future information infrastructure that is a negligible excess over the basic propagation time between them and the remote resource. They will want the infrastructure to be not only widebanded for quick access to complex objects (which are evolving already from just still graphics to include voice and video) but also to be symmetric, so that any user can become the center of his or her own communicating community. This need for an "any-to-any" infrastructure, as contrasted to the one-way or highly asymmetrical character of most of our wideband infrastructure today (cable and broadcast), is thought by many political leaders to be the key to optimizing the use of communication technology for the public good.

Thus, a dim outline of many of the needs that the information infrastructure of the future must satisfy can be discerned in the emerging set of high-bandwidth usage modes of the Internet today<sup>1</sup>, particularly the Web. The picture that emerges from examining what is happening in the Web is most instructive. Figure 1 shows the recent and projected growth of Web traffic per unit time *per user* assuming the present usage patterns, which include almost no voice, video clips, or high response speed applications such as point-and-shoot games or interactive CAD simulations. As these evolve, they could exacerbate the already considerable bit rate demand per user, which Figure 1 shows as a factor of 8 per year. If the extrapolations in Figure 1 are correct, this means that in the decade to 2005, the portion of the available communication infrastructure devoted to descendants of the Web must undergo a capacity growth of about  $10^9$  in order to keep up with demand.

There is only one physical transmission technology capable of supporting such growth: optical fiber. Fortunately for the prospects of an infrastructure that will provide society what it needs, fiber has been going into the ground, on utility poles, within buildings, and under the oceans at a rapid rate. The installation rate has been over 4,000 miles per day for some years, just in the continental United States alone, so that by now over 10 million miles of installed fiber exist here. Even more fortunately, *each fiber* has a usable bandwidth of some 25,000 GHz, roughly 1,000 times the usable radio spectrum on planet Earth, and quite enough to handle all the phone calls in the U.S. telephone system at its busiest. While this gigantic capacity is underused by at least a

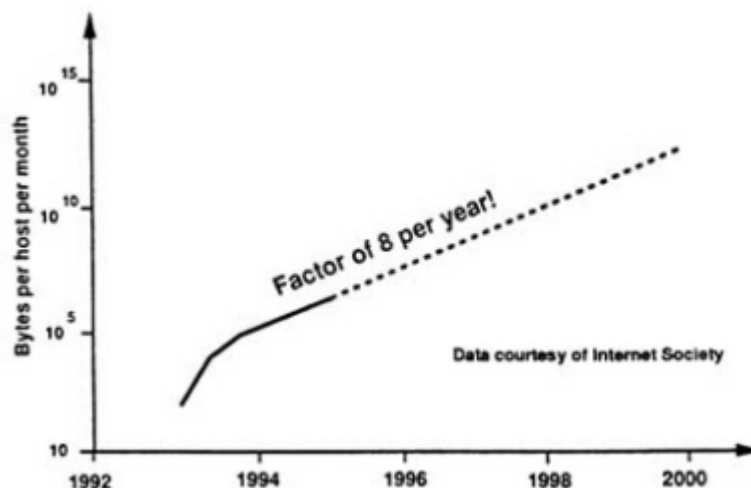


Figure 1 Predicted World Wide Web bandwidth demand.  
SOURCE: Data courtesy of the Internet Society.

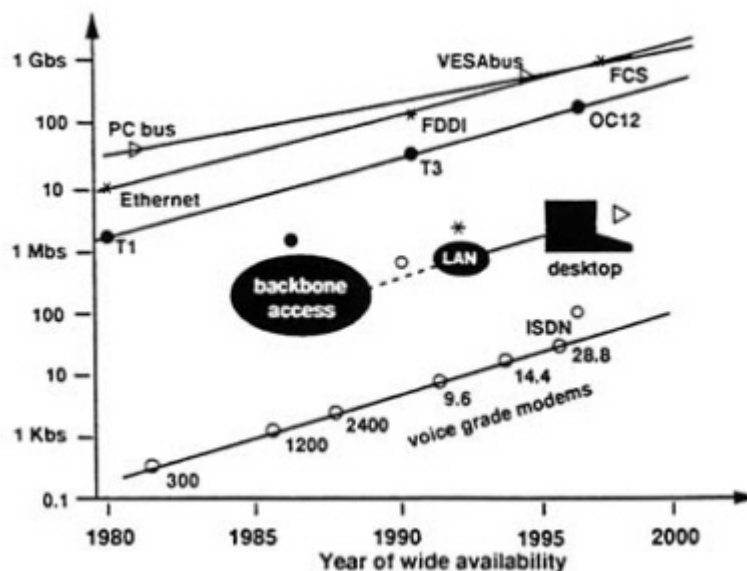


Figure 2 The "last mile" bandwidth bottleneck.

factor of 10,000 in today's practice, which is based on time division multiplexing, the technical means are rapidly evolving to open up the full fiber bandwidth. This is the all-optical networking technology, based on dense wavelength division, in which different channels travel at different "colors of light."

So, why isn't it true that we already have the physical basis in place over which to send the traffic of the future? Most of the answer is summarized in Figure 2<sup>2</sup>. All the communication resources we have been installing seem to be improving in capacity by roughly 1.5 per year, totally out of scale with the 8-times-per-year growth of demand shown in Figure 1. The top curve of Figure 2 shows the capability of desktop computers to absorb and emit data into and out of the buses that connect them to the external world<sup>3</sup>. The next line shows local area network capacity as it has evolved. The third one shows the evolution of high-end access to the telco backbone that allows users at one location connectivity to users elsewhere outside the local LAN environment. The capacity of this third curve has been available only to the most affluent corporations and universities, those that can afford T1, T3, or SONET connections to their premises.



While all three of these capacities are evolving at the rate of only a factor of 1.5 per year, they represent really substantial bit-rate numbers. Current Web users who can afford 10 Mb/s LANs and T-carrier connections into the backbone experience little response time frustration. However, the situation of most of us is represented more accurately by the bottom curve, which shows the data rate available between our desktop and the backbone over the infamous "last mile" using telco copper connections with either modems or ISDN. There is a  $10^4$  performance deficiency between the connectivity available between local and long-haul resources and the internal performance of both these resources.

If one compares the rate of growth of Web traffic in Figure 1 with the data of Figure 2, it is clear that there is an acute need to bridge the  $10^4$  gap of the last mile with fiber and inevitably to increase the bandwidths of the backbone also, probably at a greater rate than the traditional factor of 1.5 per year.

As for bridging the gap between the desktop and the telco backbone, the proposed solution for years now has been "fiber to the home"<sup>4</sup>, expressing the notion that it must pay for itself at the consumer level. The alternative of coaxial cable to the premises, while having up to a gigahertz of capacity, is proving an expensive and nonexpandable way to future-proof the last mile against the kind of bandwidth demands suggested by Figure 1, and the architectures used have assumed either unidirectional service or highly asymmetrical service. What is clearly needed is fiber, probably introduced first in time-division mode, and then, as demand builds up, supporting a migration to wavelength division (all-optical).

Figure 3 shows the rate at which fiber to the premises ("home") has been happening<sup>5</sup> in the United States. The limited but rapidly growing amount of fiber that is reaching all the way to user premises today is mostly to serve businesses. The overall process is seen to be quite slow; essentially nothing very widespread will happen during the next 5 to 7 years to serve the average citizen. However, the bandwidth demand grows daily. Meanwhile, all-optical networks are beginning to migrate off the laboratory bench and into real service in small niches.

What Figure 3 shows is the steady reduction of the number of homes that, on the average, lie within the area surrounding the nearest fiber end. In 1984, when fiber was used only between central offices (COs), this figure was the average number of homes or offices served by such a CO. As the carriers<sup>6</sup>, cable companies<sup>7</sup>, and competitive local access providers<sup>8</sup> found great economies in replacing copper with fiber outward from their COs and head-ends, the number decreased. A linear extrapolation down to one residence per fiber end predicts that 10 percent of U.S. homes will be reached by fiber by about 2005, at best. In Japan, it is quite possible that a strong national effort will be launched that will leapfrog this lengthy process using large government subsidies<sup>9</sup>.

During the coming decade, several things will happen, in addition to ever increasing end-user pressure for more bandwidth to the desktop. Competition between telcos, cable companies, and competitive access providers may or may not accelerate the extrapolated trend shown in Figure 3. Advances in low-cost optoelectronic technology, some of them based on mass production by lithography, could also accelerate the trend, because analyses of costs of fiber to the home consistently show a large fraction of the cost to lie in the set-top box, splitters, powering<sup>10</sup>, and, in the case of wavelength division multiplexing (WDM) approaches, multiwavelength or wavelength-tunable sources and receivers. It is widely felt that the price of the set-top box itself will have to be below \$500 for success in the marketplace. This is probably true, whether the "set-top box" is really a box sitting atop a TV set or a feature card within a desktop computer. By 2005 it should become quite clear whether the TV set will be growing keyboards, hard disks, and CPUs to take over the PC, whether the PC will be growing video windows to take over the TV set, or whether both will coexist indefinitely and separately. In any case, the bottleneck to these evolutions will increasingly be the availability by means of fiber of high bit rates between the premises and the backbone, plus a backbone bandwidth growth rate that is itself probably inadequate today.

Meantime, looking ahead to the increasing availability of fiber paths and the customers who need them to serve their high-bandwidth needs, the all-optical networking community is hard at work trying to open up the 25,000 GHz of fiber bandwidth to convenient and economical access to end users. Already, the telcos are using four-wavelength WDM in field tests of undersea links<sup>11</sup>. IBM has recently made a number of commercial installations<sup>12</sup> of 20-wavelength WDM links for achieving fiber rental cost savings for some of its large customers who have remote computer site backup requirements. The rationale behind both these commercialization efforts involves not only getting more bandwidth out of existing fiber, but also making the

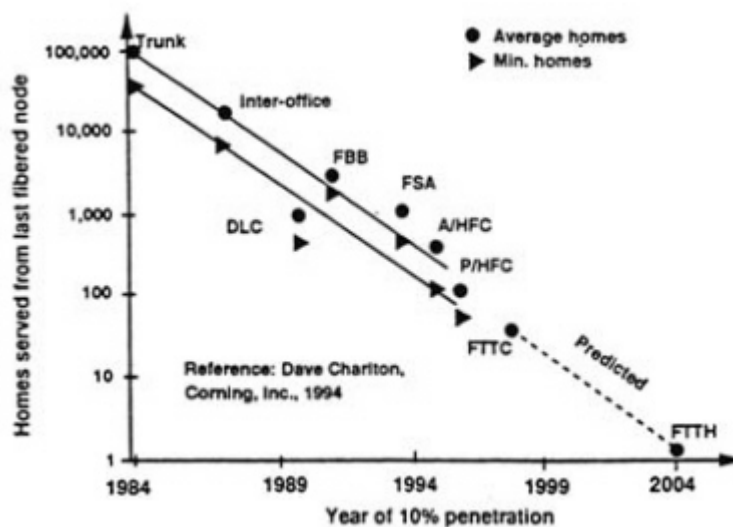


Figure 3 Predicted rate at which fiber reaches user premises.

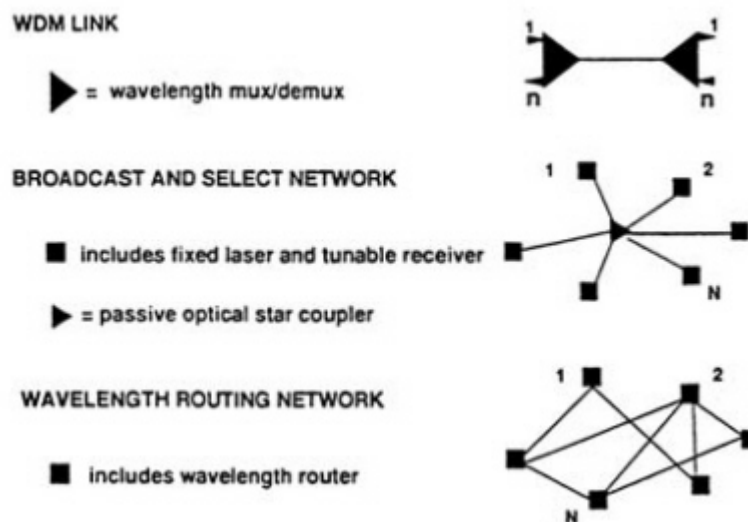


Figure 4 Three wavelength division architecture.

installation "multiprotocol" or "future-proof" by taking advantage of the fact that each wavelength can carry an arbitrary bit rate and framing convention format, or even analog formats, up to some maximum speed set by the losses on the link.

These successful realizations of simple multiwavelength links represent the simplest case of the three different kinds of all-optical systems, shown in Figure 4. In addition to the two-station WDM link (with multiple ports per station), the figure shows the two forms taken by full networks, structures in which there are many stations (nodes), with perhaps only one or a few ports per node.

The second type, the broadcast and select network, usually works by assigning to the transmit side of each node in the network a fixed optical frequency, merging all the transmitted signals at the center of the network in an optical star coupler and then broadcasting the merge to the receive sides of all nodes. The entire inner structure, consisting of fiber strands and the star coupler, is completely passive and unpowered. By means of a suitable protocol, when a node wants to talk to another (either by setting up a fixed light path "circuit" or by exchanging packets), the latter's receiver tunes to the former's transmit wavelength and vice versa. Broadcast and

select networks have been prototyped and, while still considered not quite in the commercialization cost range, have been used in live application situations, for digital video distribution<sup>13</sup> and for gigabit supercomputer interconnection at rates of 1 gigabit per second<sup>14</sup>.

Aside from high cost, which is currently a problem with all WDM systems, there are two other things wrong with broadcast and select networks. The power from each transmitter, being broadcast to all receivers, is mostly wasted on receivers that do not use it. Secondly, the number of nodes the network can have can be no larger than the size of the wavelength pool, the number of resolvable wavelengths. Today, even though there are 25,000 GHz of fiber capacity waiting to be tapped, the wavelength resolving technology is rather crude, allowing systems with only up to about 100 wavelengths to be built, so far<sup>15</sup>. The problems of both cost and number of wavelengths are gradually being solved, often by the imaginative use of the same tool that brought cost reductions to electronics two decades ago: lithography.

Clearly, a network architecture that allows only 100 nodes does not constitute a networking revolution; some means must be provided for achieving scalability by using each wavelength many places in the network at the same time. Wavelength routing accomplishes this, and also avoids wastage of transmitted power, by channeling the energy transmitted by each node at each wavelength along a restricted path to the receiver instead of letting it spread over the entire network, as with the broadcast and select architecture. As the name "wavelength routing" implies, at each intermediate node between the end nodes, light coming in on one port at a given wavelength gets routed out of one and only one port.

The components to build broadcast and select networks have been available on the street for 4 years, but optical wavelength routers are still a reality only in the laboratory. A large step toward practical wavelength routing networks was recently demonstrated by Bellcore<sup>16</sup>.

The ultimate capacity of optical networking is enormous, as shown by Figure 5, and is especially great with wavelength routing (Figure 6). Figure 5 shows how one might divide the 25,000 GHz into many low-bit-rate connections or a smaller number of higher-bit-rate connections. For example, in principle one could carry 10,000 uncompressed 1 Gb/s HDTV channels on each fiber. The figure also shows that erbium amplifiers, needed for long distances, narrow down the 25,000 GHz figure to about 5,000 GHz, and also that the commercially available tunable optical receiver technology is capable of resolving no more than about 80 channels.

With broadcast and select networks the number of supportable connections is equal to the number of available wavelengths in the pool of wavelengths. However, with wavelength routing, the number of supportable connections is the available number of wavelengths multiplied by a wavelength reuse factor<sup>17</sup> that grows with the topological connectedness of the network, as shown in Figure 6. For example, for a 1,000-node network of nodes with a number of ports (the degree) equal to four, the reuse factor is around 50, meaning that with 100 wavelengths, there could, in principle, be five connections supportable for each of the 1,000 nodes.

As far as the end user is concerned, there is sometimes a preference for circuit switching and sometimes for packet switching. The former provides protocol transparency during the data transfer interval, and the latter provides concurrency (many apparently simultaneous data flows over the same physical port, by the use of time-slicing). In both cases, very large bit rates are possible without the electronics needing to handle traffic bits from extraneous nodes other than the communicating partners.

The very real progress that has been made to date in all-optical networking owes a great deal to the foresight of government sponsors of research and development the world over. The three big players have been the Ministry of Posts and Telecommunications (MPT) in Japan, the European Economic Community (EEC), and the U.S. Advanced Research Projects Agency (ARPA). The EEC's programs, under RACE-1 and RACE-2 (Rationalization of Advanced Communication in Europe), have now been superseded by ACTS (Advanced Communication Technology Systems).

In 1992, ARPA initiated three consortia aimed at system-level solutions, and all three have been successful. The Optical Networking Technology Consortium, a group of some 10 organizations led by Bellcore, has demonstrated an operating wavelength routing network using acoustooptic filters as wavelength routers. The All-Optical Networking Consortium, consisting of the Massachusetts Institute of Technology, AT&T Bell Laboratories, and Digital Equipment Corporation, has installed a network that combines wavelength routing, wavelength shifting, broadcast-and-select, and electronic packet switching between Littleton, Lexington, and Cambridge, Massachusetts. With ARPA and DOE support, IBM (working with Los Alamos National

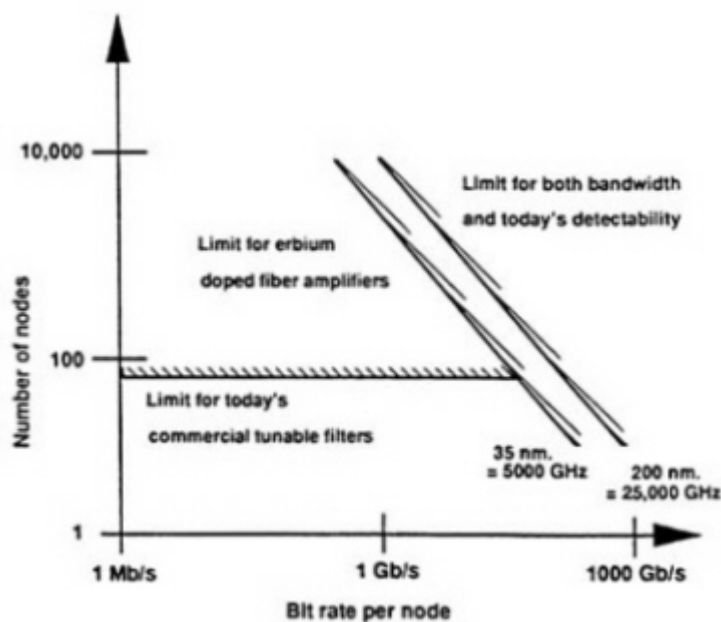


Figure 5 Capacity of broadcast and select networks.

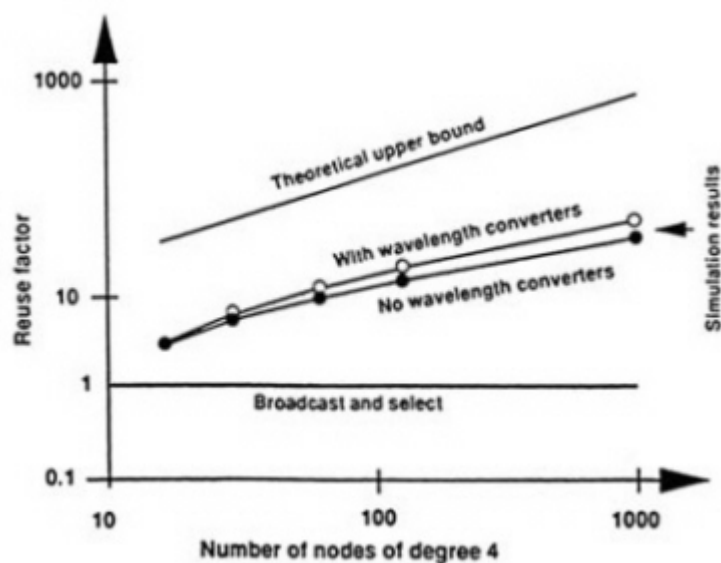


Figure 6 Wavelength reuse.

Laboratory) has developed an extensive set of algorithms for distributed control of very large wavelength-routing networks, and has studied offloading of TCP/IP for supercomputer interconnection in its Rainbow-2 network.

It is fair to say that the United States now holds the lead in making all-optical networking a commercial reality, and that ARPA support was one of the important factors in this progress. At the end of 1995, ARPA kicked off a second round of 3-year consortia in the all-optical networking area, with funding roughly five times that of the earlier programs<sup>18</sup>.

Whether all-optical networking will be a commercially practical part of the NII depends on three factors: (1) whether the investment will be made to continue or accelerate the installation of fiber to the premises and desktop (Figure 3), (2) whether it proves feasible to reduce component costs by two to three orders of magnitude below today's values, and (3) the extent to which providers offer the fiber paths in the form of "dark fiber"—that is, without any electronic conversions between path ends.

This last problem seems to be solving itself in metropolitan and suburban areas of many countries, simply by competition between providers, but the problems of long dark fiber paths that cross jurisdictions and require amplification have yet to be faced. In the United States, the Federal Communications Commission has viewed dark fiber as being equivalent to copper, within the meaning of the Communication Act of 1934<sup>1920</sup>; that is, if the public interest requires making dark fiber ends available, one of the monopoly obligations implied by monopoly privileges is that the public should be offered it at a fair price.

The optoelectronic component cost issue is under active attack. Considering that there are significant efforts under way to use lithography for cost reduction of tunable and multichannel WDM transmitters and receivers, it seems possible to predict a one-order-of-magnitude decrease in price by 2000 and two orders of magnitude by 2005. This implies that the optoelectronics for each end of WDM links of 32 wavelengths should cost \$15K and \$1.5K, respectively, and that the optoelectronics in each node of a broadcast and select network of 32 to 128 nodes should cost \$1,000 and \$100, respectively. If these last numbers are correct, this means that broadcast and select MANs and LANs should be usable by desktop machines some time between 2000 and 2005, since the costs would be competitive with the several hundred dollars people typically spend year after year on each modem or LAN card for PCs.

The sources of investment in the "last fiber mile" are problematical. In the United States the telcos and the cable companies are encountering economic problems in completing the job. In several other countries with strong traditions of centralized telecommunication authority, for example Japan and Germany, a shortcut may be taken using public money in the name of the public interest. So far in the United States it is "pay as you go." This has meant that only businesses can afford to rent dark fiber, and even then this has often been economical only when WDM has been available to reduce the number of strands required<sup>12</sup>.

Whether a completely laissez-faire approach to the last mile is appropriate is one of the problems governments are facing in connection with their information infrastructures. Fiber has ten orders of magnitude more bandwidth (25,000 GHz vs. 3.5 kHz) and can operate with ten orders of magnitude better raw bit error rates that can voice grade copper ( $10^{-15}$  vs.  $10^{-5}$ ), and yet on the modest base of copper we have built the Internet, the World Wide Web, ten-dollar telephones at the supermarket, communicating PCs and laptops, prevalent fax and answering machine resources, and other innovations. It is the vision of those working on all-optical networking that a medium with ten orders of magnitude better bandwidth and error rate than one that gave us today's communication miracles is unlikely to give us a future any less miraculous, once the fiber paths, the network technology, and the user understanding are all in place.

## REFERENCES

- [1] A.R. Rutkowski, "Collected Internet growth history of number of hosts and packets per month," private communication, March 26, 1995.
- [2] A.G. Fraser, banquet speech, "Second IEEE Workshop on High Performance Communication Subsystems," September 2, 1993.
- [3] R. Dodson, "Bus Architectures," IBM PC Company Bulletin Board at 919-517-0001 (download files PS2REFi.EXE, where i = 1, 2, 3 and 4), 1994.
- [4] P.W. Shumate, "Network Alternatives for Competition in the Loop," SUPERCOM Short Course, March 22, 1995.
- [5] D. Charlton, "Through a Glass Dimly," Corning, Inc., 1994.
- [6] J. Kraushaar, "Fiber Deployment Update—End of Year 1993," FCC Common Carrier Bureau, May 13, 1994.
- [7] "Ten-year Cable Television Industry Projections," Paul Kagan Associates, Inc., 1994.
- [8] "Sixth Annual Report on Local Telephone Competition," Connecticut Research, 1994.
- [9] J. West, "Building Japan's Information Superhighway," Center for Research on Information Technology and Organization, University of California at Irvine, February, 1995.
- [10] P.R. Shumate, Bell Communications Research, private communication, March 1995.
- [11] J.J. Antonino (ed.), "Undersea Fiber Optic Special Issue," *AT&T Technical Journal*, vol. 74, no. 1, January–February 1994.
- [12] F.J. Janniello, R.A. Neuner, R. Ramaswami, and P.E. Green, "MuxMaster: A Protocol Transparent Link for Remote Site Backup," submitted to *IBM Systems Journal*, 1995.

- [13] F.J. Janniello, R. Ramaswami, and D.G. Steinberg, "A Prototype Circuit-switched Multi-wavelength Optical Metropolitan-area Network," *IEEE Journal of Lightwave Technology*, vol. 11, May/June 1993, pp. 777–782.
- [14] W.E. Hall, J. Kravitz, and R. Ramaswami, "A High-Performance Optical Network Adaptor with Protocol Offload Features," submitted to *IEEE Journal on Selected Areas in Communication*, vol. 13, 1995.
- [15] H. Toba et al., "100-channel Optical FDM Transmission/Distribution at 622 Mb/s Over 50 km Using a Waveguide Frequency Selection Switch," *Electronics Letters*, vol. 26, no. 6, 1990, pp. 376–377.
- [16] G.K. Chang et al., "Subcarrier Multiplexing and ATM/SONET Clear-Channel Transmission in a Reconfigurable Multiwavelength All-Optical Network Testbed," *IEEE/OSA OFC Conference Record*, February 1995, pp. 269–270.
- [17] R. Ramaswami and K. Sivarajan, "Optimal Routing and Wavelength Assignment in All-optical Networks," *IEEE INFOCOM-94 Conference Record*, 1995.
- [18] R. Leheny, "Advanced Network Initiatives in North America," *Conference Record, OFC-95*, March 2, 1995.
- [19] "Four BOCs Denied Authorization to Cease Providing Dark Fiber Service," Document CC-505, FCC Common Carrier Bureau, March 29, 1993.
- [20] "Dark Fiber Case Remanded to FCC," U.S. Court of Appeals for the District of Columbia, April 5, 1994.

## 33

# Cable Television Technology Deployment

*Richard R. Green*

*Cable Television Laboratories Inc.*

The national information infrastructure (NII) is envisioned as having several components: an interconnected web of networks; information content, consisting of technical and business databases, video and sound recordings, library archives, and images; the software applications necessary for users to access and manipulate these stores of information; and the network standards that promote interoperability between networks and guarantee the security of information transmission. This infrastructure will potentially connect the nation's businesses, schools, health care facilities, residences, and government and social service agencies through a broadband, interactive telecommunications network capable of transmitting vast stores of data at high speed. Because information is a crucial commodity in an increasingly global service economy, the NII is of critical importance to the competitiveness and growth of the United States.

The cable television industry is providing a significant part of the technological infrastructure needed to make this telecommunications network a reality. This white paper discusses trends, predictions, and barriers surrounding the deployment of an advanced cable television network architecture over the next 5 to 7 years. This discussion lays the foundation for the subsequent consideration of the trends, projections, and barriers to the deployment of new services over that advanced cable architecture.

### TRENDS IN CABLE TELEVISION NETWORK DEPLOYMENT

New technological developments within the cable industry are transforming the existing cable architecture into a state-of-the-art, interactive conduit for the NII. These developments are outlined in detail below.

#### The Evolution of Cable

Cable television reached its current form during the mid-1970s when the technology was developed that allowed cable customers to receive satellite transmissions via the cable architecture that had evolved from its beginnings as Community Antenna Television. This new delivery system enabled cable companies to offer customers more channels than standard terrestrial broadcasting companies. Cable then surpassed its original mandate to bring television reception to rural or obstructed areas and became a means for delivering new types of programming through specialty channels for sports, news, movies, home shopping, weather, and so on, and through pay-per-view channels. Cable television is a major video service provider, with 63 percent of all U.S. TV households subscribing to cable. As significant, 97 percent of U.S. households have access to cable facilities, making cable a nearly universally available telecommunications infrastructure.

Cable television historically operated through the technology of coaxial cable, implemented in a "tree and branch" architecture. Video signals, in analog format, from satellites, broadcast transmissions, and local television studios are received or generated at the cable facility's head end, which serves as the point of origination for the signals to be distributed to subscribers via coaxial cable. A trunk cable carries the signal from the headend to the feeder cable that branches from the trunk into local neighborhoods. Signal amplifiers are placed into this

coaxial cable network to prevent the signal from degrading over distance and producing noise or distortion. Finally, a drop cable is run from the feeder cable into a subscriber's home and is attached to the television set.

Channel capacity for cable systems has grown from an average of 12 channels, mostly retransmission of broadcast signals, to an average of over 40 channels today. The number of cable subscribers served by systems with 30 channels or more has doubled from 48.7 percent in 1983 to 95.4 percent in 1993. Channels provided now include satellite delivered cable programming, and a variety of new educational, shopping, and entertainment networks. Using this same architecture as a platform, cable companies are currently exploring their role in the NII by initiating new applications and offering access to other networks and resources, such as the Internet. The expansion of cable's role in the NII requires building on the foundation that was laid over the last 20 years.

### The Role of Fiber Optics

The cable industry has been upgrading its coaxial cable infrastructure into a hybrid fiber optic/coaxial cable (HFC) network. Cable companies have installed fiber-optic trunk lines to replace these major arteries of the cable architecture with wider bandwidth (higher capacity) links. Optical fiber is constructed from thin strands of glass that carry light signals faster than either coaxial cable or twisted pair copper wire used by telephone companies. It allows signals to be carried much greater distances without the use of amplifiers, which decrease a cable system's channel capacity, degrade the signal quality, and are susceptible to high maintenance costs. With further upgrades, hybrid coaxial/fiber technology will also be able to support two-way telecommunications. Therefore, a broadband cable network that is capable of delivering more channels as well as high-quality voice, video, and data can be created without replacing the feeder and drop lines with fiber optic technology. This is the reason that the cable industry is perhaps the best positioned industry to deliver on the promise of the NII with a reasonable and prudent amount of investment.

Cable companies began widespread installation of fiber technology into the trunk of the cable architecture during the late 1980s. This use of fiber improved signal quality and lowered maintenance costs. In effect, fiber upgrades paid for themselves in terms of immediate cost reductions and service quality improvements. At the same time, the installed base of fiber served as a platform for further deployment of fiber to serve new business objectives.

In the early 1990s, cable further pioneered the installation of "fiber trunk and feeder" architecture in some of its markets. This approach runs fiber deeper into the network, segmenting an existing system into individual serving areas comprising roughly 500 customers. Time Warner provided a "proof-of-concept" of this approach in Queens, N.Y., with its 1-gigahertz, 150-channel system completed in 1991.

This evolutionary step offered a number of benefits. Backbone or trunk fibers may carry a multitude of available cable channels out to fiber "nodes," and remaining coaxial cable to the home can carry a particular targeted subset of the available channels. Thus, customers may be presented with more neighborhood-specific programming. Penetration of fiber deeper into the network also reduces the number of amplifiers, or active electronics, remaining between the subscriber and the headend. In some designs, amplifiers may be entirely eliminated, resulting in a so-called "passive" network design. Removal of amplifiers considerably simplifies the use of the coaxial cable for return signals from the home or office back to the headend and beyond. The portion of bandwidth reserved for return signals, usually in the 5- to 40-MHz portion of the spectrum, is often subject to interference and other impairments. Any remaining amplifiers must be properly spaced and balanced, a labor-intensive process that must be performed on an ongoing basis. Other technical impairments are unique to the return path, and technical solutions must be optimized. These obstacles are the focus of current industry research and product development. A Cable Television Laboratories (CableLabs) request for proposals issued in the fall of 1994 has spurred a number of technology companies to accelerate the refinement of technology to address return path issues. It appears that the full two-way capability of the coaxial cable, already installed to 90 percent of homes, will be fully utilized beginning in the next 12 to 18 months.

Full activation of the return path will depend on individual cable company circumstances ranging from market analysis to capital availability. There may be intermediate strategies employed by some of these companies to speed the deployment of two-way or interactive services. Such strategies might include alternative



return paths offering support for asymmetric services, services that may require a narrowband information signal that triggers delivery of broadband downstream information. Narrowband return-path options include wireless technologies such as personal communication service (PCS) or use of the public telephone network for Touch-Tone signaling or use of a narrowband modem.

Hybrid networks, then, are capable of delivering a variety of high-bandwidth, interactive services for a lower cost than fiber to the home, and still provide a graceful evolutionary path to full, two-way, broadband communications. Cable companies estimate that most subscribers will be connected to cable via fiber-optic technology by 1996 to 1998.

### **The Transition to Digital Television**

Digital compression is another technological development that will vastly increase channel capacity in addition to fostering interactivity and advanced services via cable. In contrast to current analog technology, which can collect noise (such as shadows or snow) during transmission over the air and through cable, digital compression technology delivers a signal sharply and clearly while employing a fraction of the bandwidth used by analog technology. Digital technology converts a video signal into a binary form that is stored in a computer, compressing signal information into a fraction of its original size while still permitting its easy transformation into video signals for transmission. The result is that approximately 4 to 10 digital channels can be delivered over the same bandwidth historically required to deliver 1 analog channel.

Thus, compression technology will enable cable customers to have a greater diversity of programming options such as delivering niche programming to narrowly targeted audiences, expanded pay-per-view services that will rival the video rental market, multiplexing channels (carrying a premium movie service on several different channels with varying starting times), and high-definition television. Digital compression upgrades make economic sense for consumers as well, since the converters necessary to decompress digital signals will be provided only to those cable customers subscribing to these new services.

The cable industry has led the development of digital compression technology, and standards for digital compression have been established. The cable industry's innovative work with digital television was spearheaded by CableLabs' 1991 efforts with General Instrument Corporation and Scientific-Atlanta to form a cable digital transmission consortium, which emphasized cable's leadership role in the creation of digital transmission technology. CableLabs has worked with the industry to foster convergence of digital coding and transmission for cable industry application, to provide technical support for the cable industry's work with the consumer, computer, home electronics, and entertainment software industries, and to cultivate awareness of digital compression technology. Cable companies will be ready to commence the delivery of the technology after hardware/software protocols are implemented. The industry is working to encourage its vendors to accelerate development, and CableLabs recently invented a universal analog-digital demodulation technique to enable the use of equipment using different forms of modulation being used by different vendors.

The cable industry also has looked closely at the impact of deploying digital video compression technology in the real world. CableLabs conducted a 2-year digital transmission characterization study, to (1) determine how well cable systems can transmit high-speed digital data, (2) gauge the maximum bit rate that can reliably transmit compressed video and audio, both NTSC and HDTV, and (3) identify the optimum modulation techniques for achieving three goals: maximum data rates, minimum data errors, and minimum costs for terminal equipment. This information has been distributed to vendors so that they understand the world in which their equipment must perform.

### **Advanced Television**

Cable's leadership in digital compression manifested early in the development of high-definition television. This form of advanced display technology allows cable companies to bring subscribers a television picture with greater clarity and definition than current transmission standards permit. In the NII, new forms of

information and video display will be just as important as new and expanded transmission media, and telecommunications companies must design networks that support such display technology.

The Federal Communications Commission convened a committee in 1987 to develop a broadcast standard for high-definition video transmission. Although Japan's NHK-MUSE analog solution was favored owing to its early deployment in Japan, cable industry-led efforts proved that a digital solution was a better choice for reasons of flexibility and efficiency. Since that time, cable industry representatives have worked with the FCC's Advisory Committee on Advanced Television Services to develop an industry digital compression standard. The cable industry presently has the broadband capacity to transmit HDTV; in fact, several cable systems in the United States and Canada already have experimented successfully with the delivery of HDTV in several locations, and CableLabs is working on further testing of HDTV standards of transmission.

### **Regional Interconnection**

An important architectural and economic component of cable's ongoing evolution is the construction of regional fiber optic networks to link headends and "regional hubs," so that cable operators in the same region can share services with one another in order to eliminate headends. Capital intensive network components, such as video storage, signal compression, or advertising insertion, may be shared among operators in a regional hub, thereby permitting operators to offer more services to customers. Beyond the economic benefits, regional hub designs allow cable operators to interconnect with other telecommunications services so that cable can provide video, audio, and textual data to homes and businesses from a variety of sources, and subscribers can request the delivery of specified services (such as electronic newspapers, home shopping, or video teleconferencing). Regional hub systems are being built in San Francisco, Denver, central Florida, Boston, Long Island, and Pennsylvania, among many other markets.

Interconnection of cable headends with each other and with other types of networks raises issues of interoperability. Cable already has the incentive to work toward standards for video transmission and other services in order to link cable systems together. Such standards must be extensible to other types of networks and have global compatibility as well. Both the cable industry and its competitors acknowledge that interoperability is critical to successful deployment of, and access to, the NII; thus, there is a great incentive for industries to cooperate to arrive at standards and otherwise foster open networks.

For example, CableLabs has tested the MPEG-2 (Moving Picture Experts Group) standard for compression and decoding of digital video and audio bit streams, which allows software, hardware, and network components from different manufacturers to communicate with one another. The MPEG-2 standard will likely be implemented in 1995. Cable is also active in the ATM Forum, an international consortium chartered to accelerate the use of ATM products and services. ATM refers to a cell switching technology featuring virtual connections that allow networks to efficiently utilize bandwidth. The cable industry views ATM as a technology with great long-term potential, and the ATM Forum is a useful venue to discuss interoperability specifications and promote industry cooperation. The ATM Forum is not a standards body but works in cooperation with standards bodies such as ANSI and CCITT. And finally, CableLabs has taken a leadership role in the Digital Audio-Video Council (DAVIC) that was recently created to promote interoperability among emerging digital audiovisual applications, such as video on demand, particularly in the international marketplace. This interoperability among technologies will ensure that technological development will be less costly, that the free flow of information is promoted, and that the NII will be brought to consumers more quickly.

### **Information Technology Convergence**

Cable is in the thick of communications and information technology convergence activity. As part of the ongoing development of cable technology applications, the cable industry is launching initiatives with the computer industry to combine high-capacity transmission with the latest developments in software technology. For example, software developments will enable cable carriers to provide the "road maps" customers will need to

navigate the information superhighway (such as interactive program guides). Other developments will allow cable to offer expanded services, such as teleconferencing, transaction processing, and home shopping. Intel and General Instrument, in conjunction with Tele-Communications, Inc. (TCI), and Rogers Cable systems of Canada, are working to create a cable delivery system that is capable of sending data to personal computers at speeds that are up to 1,000 times faster than today's modems. Intel is also developing services tailored to the cable market, including multimedia online services, personal travel planning, software distribution, and Internet access, which will allow the personal computer to become a powerful communications tool in the foreseeable future. Elsewhere, Silicon Graphics is playing an integral role in Time Warner's Full Service Network in Orlando, Fla. And Microsoft recently announced the demonstration of its server software architecture called Tiger for the delivery of continuous media, such as voice and video on demand. Tiger, which is deployed in a cable system's headend and in software for in-home receivers, is being tested by Rogers Cable systems Limited in Canada and by TCI in Seattle.

### **OUTLOOK FOR FUTURE DEPLOYMENT OF CABLE TELEVISION ARCHITECTURE**

Cable companies plan to invest about \$14 billion to institute equipment and plant upgrades through the end of the decade. Some cable companies now predict that the construction of fiber/coaxial hybrid networks will be complete between 1996 and 1998. Upgrade costs for this hybrid network are relatively low since cable's broadband coaxial cable to the home is already in place, making the total investment in broadband/digital technology cheaper for cable than for competitors, such as telephone companies, to develop. The cost of media servers (the digital storage devices that cable systems will use to handle simultaneous requests for data, voice, and video services to the home) and of set-top boxes or home terminals that consumers will use to access multimedia and other digital services are relatively high at present, but will decline as production increases. Set-top boxes with digital decoders, which will bring such services as movies on demand and on-screen program guides to the home, are expected to be widely available in 1996 to 1998. Costs of servers and switching technology that could bring advanced services such as true video on demand or video telephony are expected to continue to fall and will likely be incorporated into cable architectures around the end of the decade.

Today, fiber nodes are being installed, upgrade planning is in progress, regional hubs are being developed, and a number of cable MSOs are conducting near-video-on-demand, video-on-demand, or full service network trials, in places like Orlando, Fla., Omaha, Nebr., and Castro Valley, Calif. Satellite-based video compression, near-video-on-demand services, and some first generation interactivity, such as customer navigation or program selection, will be deployed in 1995 and 1996.

By 1997 and 1998, 750-MHz upgrades should be complete in many areas, with that capacity likely allocated to 78 analog channels and the equivalent of 100+ digital channels. Also in this time frame, near-video on demand will be a mature service, second generation interactivity will begin to appear, and deployment of telecommunications services such as telephony, personal communication service (PCS), high-speed data, and perhaps video telephony will begin. By 2000, broadband, full-service networks will be widespread, featuring ATM/SONET technology, media servers used to provide true video on demand, full motion navigation tools, and advanced television services.

### **BARRIERS TO FUTURE DEPLOYMENT OF CABLE NETWORK ARCHITECTURE**

The ability of cable companies to make necessary investments in network upgrades and new technology may be affected by regulatory issues currently being debated within Congress. The telephone companies' state-regulated monopoly over local telephone service is a primary hindrance in the development of a competitive telecommunications marketplace. Moreover, the rate reregulation mandated by the Cable Act of 1992 has amounted to a loss of \$2 billion in cable industry revenue through 1994. Capital formation critical for technological development requires a stable and competitively neutral regulatory environment. The cable industry's development of the hybrid fiber/coaxial network has nevertheless resulted in cable's increased ability to

act as a competitor with telcos since the hybrid system provides a higher capacity than phone companies' twisted pair wire. Legislation is being debated in Congress that explores the possibility of eliminating the barriers to entry into local cable and telco markets and preempting state regulations that deter phone companies from making their network available for use by the cable industry. Regulatory reform efforts are under way in many states as well. State-imposed restrictions on competition within local telecommunications markets must be lifted so that cable companies can develop competitive services on a national scale; otherwise the information infrastructure may not become national in character. Moreover, competition must be promoted so that consumers can choose among service providers and so that diverse services will be available.

### TRENDS IN TELEPHONY DEPLOYMENT

This discussion has described the historical and prospective evolution of cable television architectures to deliver entertainment and educational video. This deployment has reached a watershed in that evolution, as the existing cable platform becomes capable of delivering a range of new services, including competitive telephony services.

Telephone service has historically been a monopoly service, but this situation began to change with the opening of the old Bell systems as early as the 1950s. By the 1970s, long-distance markets had become competitive, and technological and regulatory forces led to the divestiture by AT&T of its local telephone companies. In the 1980s, further advances in technology led to the birth and rapid growth of competitive telecommunications companies, called Competitive Access Providers (CAPs), that targeted network access services connecting local telephone customers to interexchange carriers. Such competitive options were initially limited to large business customers, but the trend has led inexorably to the need to open up the local loop to competition. This trend has led a number of cable television providers to recast themselves as full-fledged cable telecommunications companies. Several cable companies operate competitive access subsidiaries, and a consortium of cable companies now owns Teleport Communications Group, one of the two largest competitive access companies.

### OUTLOOK FOR DEPLOYMENT OF TELEPHONY SERVICES

Many cable companies are assessing their technological and financial capabilities for competing in the telecommunications business. The advanced capabilities being provided by the evolving cable architecture will provide a platform capable of providing telephony services. A request for proposals (RFP) for telecommunications service was issued by CableLabs in 1994, and the RFP served to announce the intent of six leading cable companies to buy up to \$2 billion worth of hardware and software. This equipment would enable cable operators to provide telephone service to residential and business customers over cable television hybrid fiber/coax networks. The RFP has focused vendors on devising affordable answers to issues of reliability and bandwidth management of multiple services over the same HFC network. Further, cable companies are fully aware of the particular demands of lifeline telephone service, and the RFP stipulates requirements for full network reliability.

Regional hub evolution has already led a number of cable companies to deploy class 5 telecommunications switches. Continental Cablevision in New England, Cablevision Systems on Long Island, and Time Warner in Rochester, N.Y., and Orlando, Fla., are among the companies that have done so, aggregating a large number of customers on a regional basis in order to share costs.

Cable companies Tele-Communications, Inc., Comcast, and Cox Cable formed a consortium with Sprint in 1994 to provide telecommunications services nationwide. The consortium's initial action was to participate in PCS spectrum auctions, but plans also commit the companies to invest in network upgrades and to provide capital to other cable companies that may wish to upgrade plant in order to participate in the telecommunications venture.

Time Warner in Rochester, N.Y., may be the first cable company to provide head-to-head competition with a local telephone company. Regulatory actions have cleared the way for competition, and Time Warner has

devoted resources to upgrading its HFC network in Rochester and installing a complete service management system accommodating back-office functions such as billing.

### **PERSONAL COMMUNICATION SERVICE**

Personal communication service, or PCS, is a form of wireless two-way communication using a lightweight handset similar to cellular telephone technology. This technology is compatible with existing cable network architectures. Using microcells that transfer a user's call from one cell to another as the user travels, PCS uses the cable industry's fiber optic backbone and feeder plant to interconnect the cells, and thus cable has a built-in advantage over other potential providers by virtue of an existing infrastructure. PCS will distinguish itself from cellular telephone service primarily by its lower cost and greater convenience. More than 26 cable companies received FCC approval to test PCS technology. Cox Cable received the FCC's "pioneer's preference" license for spectrum space to explore PCS systems, and Time Warner has successfully completed testing in Orlando, Fla., of PCS technology and its interconnection with the cable infrastructure. The Sprint-TCI-Comcast-Cox consortium successfully bid for licenses covering 29 of the top 50 markets, covering 180 million POPs (points of presence), making it the largest winner in the recent PCS spectrum auction.

### **BARRIERS TO DEPLOYMENT OF TELEPHONY SERVICES**

Current modes of telecommunications regulation hamper cable efforts to enter the telephony business. In particular, a number of states still prohibit or hinder any competition to the entrenched local telephone company monopoly. The cable industry is seeking to safeguard competition by careful removal of regulatory restrictions. The cable industry is supporting telecommunications reform proposals currently before Congress. These proposals will clarify the rules governing the development of the NII. Ultimately, competition will best stimulate development of new technology and services; furthermore, domestic competition will best build U.S. competitive strengths.

Current legislative proposals can provide a rationally managed transition to foster facilities-based competition for telephone service. Permission to enter the business must include provisions for reasonable access to unbundled elements of the public switched telephone network, as it is in the public interest for interconnection to be as fair and as seamless as possible.

The ability to make decisions in a competitive environment in turn will stimulate appropriate investments in R&D and technology deployment. Presumably, appropriate allowances can be made for joint ventures and mergers, especially for small companies serving lower density areas, to permit the capital formation necessary for building a truly national information infrastructure.

Apart from regulatory barriers, cable companies have recognized that providing lifeline and other telecommunications services requires a high standard of customer service. The cable industry in 1994 initiated a comprehensive customer service program including "on-time" guarantees and other elements that it recognizes are crucial elements in playing its part in the NII. Further, the industry is implementing a new generation of network management and business support systems that are an integral part of a multiple service endeavor providing transactional and other two-way services. In fact, cable is deploying state-of-the-art systems for telecommunications that give it an advantage over existing telephone companies, which are hampered by dependence on "legacy" network management and business support systems.

### **TRENDS IN DEPLOYMENT OF PERSONAL COMPUTER-BASED SERVICES**

Cable companies have historically carried a number of data services. These have ranged from news and weather feeds, presented in alphanumeric form on single channels or as scrolling captions, to the X\*Press information service, a one-way transmission of data over classic cable systems. X\*Press has transformed itself in recent years, forming a multimedia educational service called Ingenius in order to respond to the ongoing changes

in the cable infrastructure, which now supports expanded capabilities. X\*Press/Ingenius is now joined by a host of information providers targeting cable as the delivery mechanism of choice for advanced data services.

These changes parallel those taking place in the commercial and residential data transmission markets. The PC explosion of the 1980s was rapidly followed by leaps in computer networking technology. More and more people at work and at home were becoming familiar with computer networking, ranging from commercial services such as CompuServe or Prodigy to the wide-ranging global Internet. Increased awareness has led to increasing demand for service, and for enhanced levels of service. Cable is in a unique position to meet these demands.

The same highly evolved platform that enables cable to provide telephony also supports high-speed data services. Here, cable's high bandwidth pipeline offers a strong advantage in delivering useful data services accessible by residential and commercial customers over their personal computers and other data terminals.

### Information Services

Cable technology can provide customers with access to a variety of information services, including catalog shopping, financial data, and household bill-paying. Because cable systems have a higher capacity for data, audio, and video information, cable subscribers can send and receive information services through their computers at a much faster rate than traditional telephone lines offer. One such service, Ingenius Executive, is offered by cable systems across the country and permits subscribers to connect their personal computers to their cable, using a software interface to receive information services via their computers without tying up their telephone lines. Ingenius Executive subscribers may access news, sports, and weather information as well as stock market and consumer information. Ingenius also operates a service called X\*Change, which provides subscribers with a news feed from journalists in more than 20 locations throughout the world, as well as other educational data. The on-line service Prodigy is offering similar services in several markets across the country via Media General, Cox Cable, Viacom, and Comcast. Some of the informational features available through Prodigy over cable include reference databases that offer *Consumer Reports*, restaurant reviews, and political profiles; children's features such as *Sesame Street*, *National Geographic*, and the *Academic American Encyclopedia*; stock quotes and charts of key economic indicators; travel services; bulletin boards; shopping; and local community information. Several continually updated videotext services are available via cable as well, such as UPI DataCable, which features international and national news, financial reports, and weather, or CNBC, a commercial-free version of the cable channel's programming combined with financial presentations to corporate clients.

Companies may take different or evolving approaches to on-line service access. For some applications, customers may be accessing information stored on CD-ROM databases at or near the cable headend or regional hub. Some forms of information, such as encyclopedias, are particularly suited for such an approach. This may serve as a transitional approach until wide area cable interconnections are in place to allow information access from any remote sites. Some forms of frequently updated material may require such networked access.

### Internet Access

In addition to these information services, upgraded cable networks are now able to provide high capacity access to the Internet to customers with home computers. With a cable connection to the Internet, businesses and consumers can pay a flat monthly fee for access and can receive electronic mail, access to USENET discussion groups, ability to connect to computers around the world via telnet, and access to information archives through file transfer protocol (ftp) and gopher. This option was recently offered to Continental Cablevision customers in Cambridge, Mass., in a joint project with Performance Systems International, Inc. Subscribers receive Internet access for a monthly fee through a cable converter box that does not interfere with cable television service and works at speeds hundreds of times faster than telephone modem calls. The half-megabit per second cable link allows customers to download large data files in a fraction of the time it takes over the telephone, and it even

provides enough bandwidth to watch videos over a personal computer. Cable's Internet gateway not only allows customers faster access to the Internet but also a greater variety of creative applications of Internet multimedia features.

### **LAN Interconnection**

Some larger businesses presently use cable-provided facilities to link their local area networks to transmit high volumes of data. The evolving cable network architecture will soon enable smaller businesses to benefit from advanced telecommunications services. TCI and Digital Equipment Corporation are working jointly to develop new business telecommunications applications using TCI's HFC facilities and DEC's computer networking (Ethernet) technology. Cable's high bandwidth infrastructure combined with DEC's computer networking technology will allow remotely located LANs to share computer resources and data. Cable networking technology also will enhance opportunities for product design and manufacturing as well as science and engineering research. For example, Times Mirror Cable, DEC, and Arizona State University have tested a broadband metropolitan network called Economic Commerce Net (EC Net) that supports manufacturing applications for aerospace businesses in the Phoenix area. EC Net allows these businesses to collectively improve manufacturing by offering desktop videoconferencing on manufacturing processes, a computer aided design (CAD) tool that permits remote businesses to view and manipulate designs simultaneously, and a multimedia storage and retrieval facility for data, video, purchasing specifications, and other information.

### **Cablecommuting**

The cable industry's HFC infrastructure has the potential to increase the already popular notion of working from home via the information superhighway. Approximately 8 million Americans already work through some form of telecommuting, and cable's high-volume, high-speed broadband technology will allow millions more to "commute" to work through cable. This technology will lessen the burden of commuting for both businesses and their employees who may be geographically isolated, physically disabled, or single parents. TCI is currently testing a cablecommuting project that will allow its Denver area customer service representatives to receive customer calls at home. TCI is also testing a service, developed by Hybrid Networks, Inc., in the San Francisco area that transmits high-speed data over a standard cable channel and low-speed data through telephone lines to create an interactive corporate/home network. In addition, technologies such as video teleconferencing and high-speed fax transmission that will be delivered via cable in the future will enhance cable commuters' options.

### **Research Support**

The cable industry's broadband architecture will allow researchers in a variety of fields to share advanced computing applications or scientific instruments from remote locations, resulting in comprehensive research at a fraction of current costs. One such network is being offered by Cablevision Systems in the New York City metropolitan area. FISHNet (Fiber optic, Island-wide, Super High-speed Network) is a high-capacity fiber optic network that links the State University of New York-Stony Brook with the Brookhaven National Laboratory and Grumman Data Systems. The system has been used to revolutionize the efficient use of medical imaging and diagnostic techniques as well as to develop modeling procedures for the transport of contaminants in groundwater.

## **OUTLOOK FOR PC-BASED APPLICATIONS**

Personal computers continue to penetrate the home market. Moreover, most home PCs today include a CD-ROM attachment. This indicates strong demand for content-based services running over PCs. CD-ROMs, however, may be a short-term solution. Network-based content servers interconnected via cable company high-speed

links are a more efficient way for customers to access updated databases, news, and other forms of information.

Further, data networking has been an explosive growth area within information technology in recent years. Business applications require more and more bandwidth, especially in interconnecting disparate LANs, and cable is the provider with the greatest capacity to meet the mushrooming demand for bandwidth.

### **BARRIERS TO DEPLOYMENT OF PC-BASED SERVICES**

There appear to be very few barriers to cable deployment of high-speed data transmission. The cable platform is steadily evolving to a hybrid digital and analog transmission system. Data modems already exist, and improved models are under development by such companies as Intel, LANCity, Zenith, General Instrument, and Hybrid Networks, Inc. Modem prices are dropping precipitously—LANCcity announced in May 1995 that its 10 Mbps cable modem was priced at \$595, compared to several thousand dollars just 8 months earlier. Return path issues that were mentioned previously in connection with cable plant upgrades also will need to receive attention in the context of PC-based services. Customers may not desire totally symmetrical data communication, particularly residential user, but the return path still must be reliable. Bandwidth demand for return signals may be very dynamic, requiring cable systems to be able to allocate frequency in an agile manner. Upstream, broadband transmission over cable is expected to be fully supported within 5 years.

### **TRENDS IN THE DEPLOYMENT OF INTERACTIVE TELEVISION**

The convergence of telecommunications and computer technology is transforming American society. In the cable industry, the convergence of cable's high-speed transmission capability and computer hardware and software intelligence is not only enabling cable to deliver telephony and high-speed data services, but is creating new opportunities in entirely new forms of entertainment, education, health care, and many other areas. Cable companies, such as Time Warner in Orlando, have now deployed operational systems actually delivering such services.

#### **Video on Demand**

Digital compression technology, which enables cable companies to offer a greater number of channels as well as interactive capability, ensures that video on demand will be part of the information superhighway's features. This service allows customers to select from a range of movies, sporting events, or concerts for viewing at their convenience. Many cable operators, such as Paragon Cable in San Antonio, are already offering the less costly option of near-video on demand, which allows customers to view programs with start times every 15 or 30 minutes. True video-on-demand systems are currently being tested in Omaha, Neb., by Cox Cable; in Yonkers, N.Y., by Cablevision Systems; in Orlando, Fla., by Time Warner; and in Littleton, Colo., by TCI. A specific service, Your Choice Television (YCTV), packages and delivers television programs on demand. For about \$1 per program, viewers can order a weekly program up to a week after it airs and a daily program the day after it airs. YCTV has been tested in eight markets, including Continental Cablevision in Dayton, Ohio, and TCI in Mount Prospect, Ill.

#### **Interactive Entertainment**

With the pioneering technology that allows interactive capability, many new information and entertainment options are being created within the cable industry. For example, the ACTV network, which was tested via Continental Cablevision in Springfield, Mass., and is now being offered to subscribers in Montreal, allows viewers to alter the content of the TV screen during the course of a program. Viewers can call up



statistical information on players during sporting events, watch a synopsis of the daily news headlines, and choose the stories on which to receive detailed reportage, or even change the pace of an exercise program. Interactive commercials will soon be available on ACTV so that customers may receive more information or discount coupons for particular products or services.

### **Education and Distance Learning**

Cable technology has enabled the development of interactive educational applications and distance learning that improve not only the way students learn but also the way teachers teach. Jones International's Mind Extension University has been offering college courses via cable television for years, and cable companies are expanding educational opportunities by building networks devoted to education. Students can now learn from national experts or students in different cities, access online libraries from around the world, or take "virtual field trips" to museums while remaining in the classroom. One such interactive learning pilot project, called the Global Laboratory Project and currently run by Continental Cablevision and the National Science Foundation, connects students in Cambridge, Mass., with those from 27 other states and 17 foreign countries to explore environmental problems. The students monitor climate change, pollution, and ultraviolet radiation and share their data among themselves and with scientists to gain a global perspective on the environment.

As part of its effort to expand learning opportunities through the use of cable technology, the cable industry has initiated a nonprofit program to provide schools with basic cable service, commercial-free educational programming, and teacher support materials. The project, Cable in the Classroom, serves nearly 64,000 schools and supplies more than 525 hours of commercial-free programming for educators each month. Highlighting the industry's role in distance learning, Continental Cablevision now operates an interactive learning network called the Cable Classroom in Enfield, Conn. It permits teaching in one school district and simultaneously offering to several others advanced math and language classes that would otherwise not be offered because of low enrollments. Because the system uses two channels, the teacher can also view the students and home cable customers can change channels to watch either the students or the teacher. The Cable Classroom also offers teachers professional development by conducting interactive meetings with teachers in four Connecticut school districts. In another example, the Ohio Cable Foundation for Education and Technology is promoting distance learning through a range of applications throughout the state. For example, TCI in Zanesville provides data retrieval services and other multimedia distance learning applications.

### **Interactive Program Guides and Navigators**

Because of the expected explosion in future cable programming and service choices, customers will seek greater choice, control, and convenience with regard to their viewing environments. These needs will spur demand for a personalized form of channel navigation. Several interactive subscriber guides are being developed and tested. One of these on-screen guides, StarSight, is available in Castro Valley, Calif., through Viacom Cable. Time Warner has developed its own navigation system for the Orlando Full Service Network.

### **Interactive Shopping and Advertising**

Companies like Time Warner are testing interactive shopping services, such as ShopperVision, that enable customers to see and evaluate products over interactive video catalogs before purchasing. Immediate ordering and the requisite billing and payment mechanisms are integrated as well.

Interactivity will permit cable subscribers to request consumer information on businesses, products, and services at the touch of a button. Real estate advertisements have been among the first to make an impact in the burgeoning marketplace. A program called Home Economics, available throughout New England via Continental Cablevision, permits viewers to request specific information about homes. Moreover, an interactive channel devoted to classified real estate advertising is in the works; the Real Estate Network will provide customers with

the ability to view real estate in full-motion video and to access details on contractors, mortgage rates, and lending institutions.

### Games

The Sega Channel, a joint venture between Time Warner and TCI that offers a library of Sega video game titles for download through cable systems, is being tested in several markets. Subscribers use a "tuner/decoder" cartridge that tunes in the channel, provides a menu, and downloads a game choice. Scientific-Atlanta and General Instrument Corporation are presently manufacturing adapters that will permit their set-top boxes to serve as vehicles for the Sega games. Multiplayer games are also a possibility across full service networks such as Time Warner's Orlando system.

### Health Care

The cable industry's bridge to the information superhighway has important consequences for new developments in health care. Through the use of telecommunications technology for medical applications, doctors can consult with researchers or specialists in other locations, conduct video seminars throughout the country, consult visual information from remote libraries or medical centers, and share medical records or images such as X-rays. Paragon Cable in Portland, Ore., uses its institutional network to transport a telemedicine or "cablemedicine" network that connects 12 medical facilities and schools throughout the city to provide information from video conferences held across the United States. Cable also has the potential to deliver a standardized means of electronic health insurance claim filing, and personalized health information systems are being tested that use terminals in the home to connect consumers to a health care databank that advises customers about self care based on their personal medical records.

## OUTLOOK FOR INTERACTIVE TELEVISION

The preceding discussion outlined a variety of potential interactive services that may be provided over cable. The range of experiments, both technical and market trials, appears to bode well for the flexibility of cable to provide the services. There have been no technology hurdles discovered in making the cable systems perform at a level needed to support interactivity. Consumer demand for such services over the next 5 to 7 years is the wild card for all prospective services. But early cable trials point to reasons for optimism. TCI's trial in a Denver suburb compared near video on demand (offering a small selection of hit movies starting every 15 minutes) to video on demand (offering a large library of movies, starting immediately on demand). Indications were that customers are receptive to increases in choice and convenience. The initial reports on the new DBS services also indicate increased buy rates from customers given a larger menu of programming from which to choose. Thus, along the dimension of upgraded entertainment services with elements of interactivity in terms of navigation and control, it appears that consumer demand will support growth of such services over the next 5 to 7 years.

More advanced levels of interactivity also appear to make sense. The same desire for choice and convenience appears to be driving early success for the Sega Channel, which delivers video games to the home. The availability of more video game choices in the home is attractive, and one might assume further that adding elements of interactivity, such as multiplayer games, would drive increased customer acceptance.

In the area of shopping and related services, many providers are optimistic about market prospects. Electronic commerce over the Internet is spawning a flurry of activity by companies trying to perfect a secure means of conducting business over electronic networks. Cable's broadband capability appears to provide significant enhancement to such commerce as a natural showcase for products through video catalogs incorporating customer interaction.

As noted above, the cable platform supporting such interaction is likely to be widely available within the next 5 years.

## BARRIERS TO DEPLOYMENT OF INTERACTIVE TELEVISION

Barriers to interactive services may manifest as financial concerns or as marketing uncertainties. As with data services, interactive services will take advantage of the in-place cable platform capable of flexible, high-capacity support. Deployment of interactive services will require only incremental investments, largely in the customer premise equipment necessary to plug into the cable platform. Such an incremental approach will minimize financial risk given remaining uncertainty about the potential size of the market. As noted above, technology and market trials give reason for optimism. Nevertheless, cable's evolutionary approach offers welcome security in the event that forecasts are too optimistic.

## CONCLUSION

Cable television companies operate a highly evolved platform capable of delivering a variety of telecommunications and information services. Additional technology components that will enable particular services contemplated for the national information infrastructure to run over that platform appear to be coming on stream at reasonable time frames and cost levels. The evolutionary nature of cable allows cable companies to invest incrementally and cost effectively only in those technologies that serve clearly defined needs or market demands. There are no significant technological or financial barriers to continued deployment of cable architecture. However, regulatory uncertainty still remains problematic.

## ADDENDUM

1. For your company or industry, over the next 10 years, please project your best estimate of scheduled construction of new broadband facilities to the total residential and small business customer base, in 2-year increments.

The cable industry already provides broadband facilities to 64 percent of television households, and makes such facilities available to 96 percent of television households. Cable facilities are available to approximately 90 percent of business locations as well, although businesses have not historically subscribed to cable. Beyond this level of service, cable companies are currently upgrading existing systems by migrating fiber optics deeper into networks, creating individual service areas of roughly 500 homes. This expands available bandwidth to around 750 MHz. The deployment of such upgrades may develop as follows:

1995	25 percent
1997	65 percent
1999	80 percent
2005	90 percent

2. Over the same period, please project dates when each family of service will first become available, and its subsequent penetration of the total base, again in 2-year increments.

## Telephony

Capabilities for providing wireline telephony over cable will track the progression of cable system upgrades as described in the response to question 1. Telecommunications switch deployment is under way as systems migrate to regional hub designs. Several cable companies have formed partnerships with Sprint and other companies to provide a full range of telecommunications services, including wireless personal communications services. Actual penetration of such cable-provided services in a competitive market is difficult to project. In the United Kingdom, cable operators offering local telephony service have reported selling telephony services to over 20 percent of homes where the service is available, indicating a significant market for competing providers of telecommunications.

### PC-Based Applications

Digital, PC-based applications may be rolled out almost immediately, based on cable's broadband pipe into the home and the development of affordable cable modems, which are probably less than a year away. Alternate return paths such as telephone lines could be used in a hybrid approach. Availability of a broadband return path for such services will follow the progression of cable system upgrades as described in the response to question 1. Penetration of such services is difficult to project, particularly in the face of phenomenal growth rates in Internet connectivity.

### Interactive Television

- 1996—First generation interactivity (navigation or program selection), satellite-based digital compression for near video on demand.
- 1998—Early deployment of second generation interactivity, such as multiplayer video games.
- 2000—Deployment of full service networks, enabling server-based true video on demand, full motion navigation tools, video telephony types of services.

Again, it is important to note that interactive services may be deployed on an incremental basis matched to customer demand, once an upgraded infrastructure is in place. Thus, availability of these services will closely track the progression of cable system upgrades as described in the response to question 1. Evidence from early interactive field trials indicates a definite customer preference for increased choice, convenience, and control by the customer.

3. Please outline the architecture(s) that will be used to build this broadband plant.

This question is answered in the foregoing white paper.

4. Please outline the peak switched digital bandwidth (in kbps or Mbps) available to an individual residential or small business user when you launch broadband service, and how that bandwidth can evolve to respond to increased peak traffic and to new, high-capacity services (which may not now exist).

Cable's existing bandwidth to the home, combined with upgrades described previously, will reach 750 MHz. Using 64 QAM digital modulation, this implies a total digital bandwidth of close to 3 gigabits per second. Allowing for preservation of existing analog services, available bandwidth would be closer to 1.3 gigabits per second. Current industry plans call for reservation of the 5- to 40-MHz portion of the spectrum for the return path use by customers for upstream transmission. Using QPSK or 64 QAM digital modulation, 35 MHz could handle anywhere from 50 Mbps to 135 Mbps. Thus, the amount of raw bandwidth available over cable is substantial.

Cable modems under development call for effective bandwidths up to 27 megabits per second per 6 MHz channel. The actual amount available depends upon usage within the roughly 500 home nodes described previously. Assuming subscriber penetration of 60 percent gives 300 potential users. For PC-based services, one may further estimate that 40 percent of homes have PCs (numbers are roughly similar for cable or noncable homes). Further, less than 20 percent of PC homes have PCs with modems, so that the range of potential data service users may fall anywhere between 8 percent and 40 percent. Assuming growth in PC penetration, we may assume that 33 percent of subscribers are potential data transmission customers, or 100 homes. And probable peak usage will be less than 33 percent, so that one might expect 33 simultaneous users. Thus, close to 1 mbps would be available given the above assumptions. Of course, data are packetized so that individual customers would use considerably less than a full Megabit per second at any given time.

This is a lower limit. One might initially increase capacity by allocating additional 6 MHz channels on demand. Cable companies may migrate fiber further into distribution networks, creating even smaller nodes and expanding bandwidth from 750 MHz up to the feasible maximum of 1.2 GHz. Cable companies are researching dynamic bandwidth allocation techniques allowing the flexibility to meet demand. Digital modulation techniques will also continue to allow greater amounts of data to be transmitted over existing bandwidth.

5. Please project the capital investment you or your industry plan to make on a per-home-passed basis to install broadband infrastructure, and on a per-subscriber basis to install specific services.

The cable industry plans to spend \$28 billion over the next 10 years on plant and equipment upgrades (according to Paul Kagan Associates). This works out to approximately \$460 per subscriber. Upgrades to evolve cable systems to

multipurpose platforms cost in the range of \$100 to \$150. Costs are rapidly dropping for cable modems necessary for PC-based data services, and should reach the \$200 to \$300 range in the next year or two. Capital investment for such services is incremental, and will be matched to new revenue streams, which may in turn expand capital available for further technology deployment. Costs for telephony service over cable are difficult to quantify because they are highly dependent on penetration assumptions. Interactive television applications may require server and other technology costing as much as \$500 per subscriber when widely deployed. Again, investment will be incremental and matched to new revenue streams.

6. Please respond to the concerns raised in Vice President Gore's letter regarding the ability of users of your network to originate content for delivery to any or all other users, versus the control of all content by the network operator.

The cable platform is steadily evolving to a hybrid digital and analog transmission system. Data modems already exist, and improved models are under development by such companies as Intel, LANcity, Zenith, General Instrument, and Hybrid Networks, Inc. Return path issues that have been raised in connection with cable plant upgrades do not appear to present obstacles to customer generation of content. Bandwidth demand for return signals may be very dynamic, requiring cable systems to be able to allocate frequency in an agile manner. Upstream, broadband transmission over cable is expected to be fully supported within 5 years.

7. Please specifically enumerate the actions that you or your industry believe that the federal government should take to encourage and accelerate the widespread availability of a competitive digital information infrastructure in this country.

Current modes of telecommunications regulation hamper cable efforts to enter the telephony business. In particular, a number of states still prohibit or hinder any competition to the entrenched local telephone company monopoly. The cable industry is seeking to safeguard competition by careful removal of regulatory restrictions. The cable industry is supporting telecommunications reform proposals currently before Congress. These proposals will clarify the rules governing the development of the NII. Regulatory relief from some cable pricing regulation will permit rational capital investment. Ultimately, competition will best stimulate development of new technology and services, and further, domestic competition will best build U.S. competitive strengths.

Current legislative proposals can provide a rationally managed transition to foster facilities-based competition for telephone service. Permission to enter the business must include provisions for reasonable access to unbundled elements of the public switched telephone network, as it is in the public interest for interconnection to be as fair and as seamless as possible.

The ability to make decisions in a competitive environment in turn will stimulate appropriate investments in R&D and technology deployment. Presumably, appropriate allowances can be made for joint ventures and mergers, especially for small companies serving lower density areas, to permit the capital formation necessary for building a truly national information infrastructure.

## 34

# Privacy, Access and Equity, Democracy, and Networked Interactive Media

*Michael D. Greenbaum, Bell Atlantic*

*David Ticoll, Alliance for Converging Technologies*

Networked interactive media have already wrought irreversible changes in the way we live, work, and educate ourselves in America. This transformative set of technologies, like others before it, is the focus of broad hopes and boundless anxiety. For some, it is a singular force for good—empowering the individual, unlocking human potential, and ushering in a new era for wealth creation in a networked economy. For others, it raises dark suspicions—as a tool for control or an enabler of fringe elements in our society.

At the very least, the new interactive media raise serious issues relating to individual privacy, access and equity for the underprivileged, and, ultimately, the impact of these media on the evolution of democracy.

The creation of a new interactive media infrastructure represents yet another "westward migration" for pioneering Americans. Like all frontier environments, the boundaries and structures of the interactive information and communications frontier are still very fluid. As such, efforts to "civilize" it through the force of regulation should be tempered with caution. Past lessons of success and failure argue that we embrace the inevitability of change with our eyes wide open.

The development of too many regulations at this nascent stage of development would dramatically slow innovation and deployment of this new public infrastructure in the United States.

The challenge ahead is threefold: (1) to deliver the promise of these emerging communications and information products and services to the consumer with an element of individual control over the collection, distribution, and use of personal information; (2) to achieve reasonably broad public access to new media and information content, particularly with regard to education and training opportunities; and (3) to develop this infrastructure to its fullest as a force for individual development and expression and the public welfare.

We recommended a coalition of consumer, business, and government interests to engage in a constructive dialogue and work toward the development of guiding principles for privacy, access and equity, and democracy. As a disciplined, self-regulating body we will see an interactive information and communications infrastructure evolve as a major force for economic development and individual opportunity within this generation.

### GENERAL CONTEXT—THE ELECTRONIC FRONTIER

Harold Innis, an early critical voice in the electronic media age and teacher of Marshall McLuhan, pointed out that new media have precipitated political change throughout history, shifting power toward the citizenry. "Monopolies or oligopolies of knowledge have been built up ... [to support] forces chiefly on the defensive but improved technology has strengthened the position of forces on the offensive and compelled realignments favoring the Vernacular."

Libraries based on clay documents enabled priests to monopolize knowledge in ancient Babylon. Papyrus scrolls supported the limited democracy of Greek city-states, and the rule of law in ancient Rome. Paper and the printing press, used to reproduce texts in the language of the common people, precipitated the Reformation, the end of feudalism, and the emergence of parliamentary democracy, capitalism, and the Industrial Revolution.

Technology continues to transform our lives and forward the aspirations of people around the world. Who can forget, for example, the student dissidents in Beijing who used computers and facsimile communications to make their case before the world and launch communist China's first democratic uprising?

Individual technologies like the fax machine can make a difference in our lives and even influence events. The convergence of interactive multimedia and the information highway marks a truly disorienting, potentially liberating step ahead for humankind. This phenomenon, like that of the affordable mass-produced automobile and the interstate highway system, can change the economy, the physical landscape, and the very nature of community in America.

TV, video, computers, networked communications, and a host of new digital, microprocessor-based applications are all, it turns out, pieces of a puzzle that is only now coming into view.

This dynamism of convergence is extraordinary. As people explore its possibilities into the future, the interactive frontier will unfold in ways that its pioneers today can't possibly imagine. In fact, any presumption today about how convergence will ultimately change the lives of our children and grandchildren is not likely to be a sound basis for public policy. Though Vice President Albert Gore (as a U.S. Senator) predicted an explosion of activity and innovation on the "Information Superhighway" 6 years ago, few others gave much credence to the Internet or the then emerging phenomenon of the World Wide Web.

The data are finally catching up with Gore's optimism. A *Newsweek* poll published in February 1995 found that 13 percent of adult Americans said they had gone online, 4 percent had perused the Web, and 2 percent had logged on for an hour or more each day. As interest among consumer and business computer users climbs, online activity accelerates exponentially. Growth of commercial online services (e.g., Prodigy, America Online, and CompuServe) remains in the double digits. Products promising easy World Wide Web navigation, such as Mosaic and Netscape Navigator, are proliferating, and modem ownership doubled in the second half of 1994.

Interactive television, video teleconferencing, work group computing, and other important interactive media are also gaining footholds as the so-called "killer" applications that will catapult them into the mainstream begin to emerge. Together, these applications will justify a national information infrastructure (NII) analogous to basic telephone or cable television service.

### TODAY'S CONTEXT: WHICH WAY TO THE FUTURE?

Today's convergence of interactive electronic media is unique in more ways than one. In addition to the novelty of the new interactive media and their intriguing capabilities, we have witnessed unprecedented public input and participation in their development. In large part these new media are closely linked to the ascent of the personal computer. Unlike other communications infrastructures, most recently cable TV, the new media owe as much to the individual PC developer, engineer, and enthusiast as they do to any corporate, institutional, or government initiative.

This, in part, explains the popularity of new interactive media among educated, middle- and upper-income American families. These are people who use computers and online services at work and at home for both serious tasks and leisure. There is an enthusiasm for interactive media among this group akin to the frontier spirit that propelled individuals and families westward in search of opportunity and adventure.

The new interactive frontier, however, is a "virtual" one, defined by the fact that it is always changing and not limited by traditional constructs of geographic proximity, time, or self-actualization. It is at once infinite in scope and capable of redefining one's sense of community. It provides a forum in which individuals, organizations, and social groups are empowered to create, collect, exchange, and distribute information in previously unimaginable ways.

This virtual frontier is predictably confounding many existing standards and ideas about legal jurisdiction. It is also dredging up new versions of old questions about what constitutes social equality. And

democracy itself is being reexamined. Speaker of the U.S. House of Representatives Newt Gingrich is perhaps the most vocal proponent of an interactive democracy, featuring virtual town halls with direct online, even video teleconferenced connections to local representatives. The key issue is, will such access improve our system, or will it deliver a new tyranny of the majority?

Privacy, access and equity, and the function of democracy are all issues on the table for discussion and debate. As we approach a more pervasive virtual frontier we are well advised to raise the level of public discourse on these issues. Soon, this frontier population, perhaps more diverse and complex than our own, will demand new standards for justice, law, and order that make sense in a world without boundaries.

## PRIVACY

The U.S. Constitution provides no explicit help in determining what privacy means. On the other hand, everyone seems to have a strong notion of what it should mean. For the purpose of this discussion, privacy includes an individual's desire to be left alone, the ability to restrict the disclosure of personal information, and the ability to restrict the way such information is used.

On the virtual frontier, privacy issues abound. Individual privacy claims are running headlong into the administrative requirements of government agencies, the competitive concerns of business, and, increasingly, the best efforts of law enforcement. A carefully considered balance will be required to ensure that reasonable standards of privacy survive the information age.

### The Right to be Left Alone

The right to be left alone would seem straightforward enough in cyberspace, but individuals increasingly find themselves vulnerable to assault. Some of these uninvited assaults are as innocuous as unsolicited advertising messages distributed throughout the net. Others are more serious. In 1992, the FBI arrested a Sherman Oaks, Calif., man for terrorizing women on the Internet. A man in Cupertino and another in Boston were arrested for pursuing young boys over the Net.

Though it may be in decline at the office, sexual harassment is alive and well online. The experience might have an unreal quality for the harasser hidden behind an anonymous online identity, but the threat or uncomfortable approach comes across as anything but virtual for the victim.

Another disturbing development online is the breakdown of the noble "hacker" ethic idealized in Steven Levy's 1984 book *Hackers: Heroes of the Computer Revolution*. A culture coveting access to secure environments for the sake of it is giving way to a more destructive ethos. The new rogue hacker unleashes viruses causing millions of dollars in software and file damage, steals phone and credit card numbers, and uses network access to terrorize individuals.

### The Ability to Restrict Disclosure and Use of Personal Information

Perhaps the most chilling privacy issue is the degree to which we can manage or contain the digital trail we leave behind. Large databases that piece together life histories and personal preferences have been a fact of life for more than 20 years. The sophistication with which that information is applied to marketing programs, human resource department analysis, surveys, and investigations is more recent.

As David Chaum, a leading cryptography expert, writes,

Every time you make a telephone call, purchase goods using a credit card, subscribe to a magazine or pay your taxes, that information goes into a database somewhere. Furthermore, all of these records can be linked so that they constitute, in effect, a single dossier of your life—not only your medical and financial history but also what you buy, where you travel, and whom you communicate with.



The new interactive media will provide many more opportunities for the collection of personal history and preference information, from the TV shows we watch to the groceries we buy. Without knowledge of these records, the ability to verify information contained within them, or control over their disposition and use, consumers are at a distinct disadvantage.

These sophisticated systems, however, also offer many benefits. Information can improve the service that consumers receive in retail environments. It may mean less time is spent filling out forms with redundant information. And it helps marketers trim databases to include only those who are predisposed to their message in their direct mailings.

The danger, of course, is that information shared among companies, nonprofit organizations, banks, credit organizations, and government agencies is inaccurate or used for a purpose that defies a reasonable standard of privacy. In addition, different consumers will have varying standards as this issue becomes increasingly apparent.

The NII's Privacy Working Group has taken on the task of developing guidelines for privacy in the information age. These guidelines, though still in development, are an attempt to update the Code of Fair Information Practices established by the Organization of Economic Cooperation and Development.

The essence of these principles is quite simple:

- Limit collection of information (only collect what you need);
- Where possible, collect information directly from the individual to whom it pertains;
- Inform subjects of the purpose of the collection (tell them why you need it);
- Only use the information for the purpose intended and for which the subject was informed; and
- Give subjects the opportunity to access their personal information and the right to seek its correction.

As a minimum, information about a transaction should be strictly between the end customer and the information user (product/service provider). The network service provider, for example, should not be privy to the content of a transaction.

An important element in the new NII guidelines is the emphasis on information user responsibility. It stresses the need for information users to "educate themselves, their employees, and the public about how personal information is obtained, sent, stored and protected, and how these activities affect others." However, to the extent that the NII places the onus of responsibility on individual consumers, it may be assuming too much. This is an increasingly sensitive area in light of recent highly publicized abuses of personal databased information by federal and state government employees:

- In 1992, Operation Rescue members used connections at the California Department of Motor Vehicles to get the addresses of abortion clinic escorts so they could harass them at home.
- The IRS says the agency has investigated 1,300 of its own employees for browsing through the tax files of family, friends, neighbors, and celebrities since 1989; of these, 400 have been disciplined.

Abuses of personal information have also surfaced in the private sector with rogue employees misusing access to private records including personal and financial information, occasionally leading to further criminal acts. In such instances, responsibility for the security of personal information should be clear. Decentralized, interactive, digital communications vastly complicate the issues of privacy and security.

### **Privacy and the Law**

Privacy issues have always loomed large for local carriers in the telecommunications industry. Regional carriers are held to the highest standards for privacy and security of any industry. As the national network infrastructure shifts from analog to digital with an increasing ratio of data to voice traffic, these standards must be reviewed and amended to consider a host of new players and technologies.

The Internet, for example, presents a whole new ball game when it comes to privacy and security. A critical element of the new network infrastructure is that it be a level playing field where both transport and content providers are equally accountable and operate under a common set of rules.

If a comprehensive, equitable framework for privacy and security is not established, a frontier vigilante ethic will take hold among consumers of the new interactive media. We are already seeing this tawdry trend emerge on the Internet.

A defensive reaction to invasive information collection and the perceived threat to privacy posed by government-specified "Clipper chip" technology is the widespread use of privacy tools in the online community. Sophisticated encryption techniques, accessible through shareware on the Internet, allow users to protect conversations, electronic messages, even databases from unauthorized view. This move toward self-protection should put service providers and information users on notice. The continued use and proliferation of privacy tools will have a destabilizing effect.

What happens, for instance, when law enforcement officials face kidnapers, child molesters, and terrorists who use encryption shareware to protect their communications?

Shortly after the tragic bombing of the Oklahoma federal building, law enforcement officials learned of incriminating bulletin board and e-mail messages that had been posted on an online service by a suspect and others with possible knowledge of his plans before the act occurred. Such evidence is invaluable and should be accessible to law enforcement.

Regional telecom carriers have shown that accommodation of wiretaps and other court-ordered law enforcement efforts targeting telephone customers can be achieved without compromising other constitutionally protected communications. Bell Atlantic, for example, has instituted a privacy policy that calls for full cooperation with any government or consumer group to help resolve privacy issues. Volunteer joint-industry consumer panels should develop model standards for joint collaboration to assist with dispute resolution.

Networked interactive media present broad legal, ethical, and technical challenges. For example, there is no physical wiretap capability online. Service providers are required either to grab information real-time from bulletin board conversations or drill down through massive data repositories, which include primarily protected conversations. Cooperative efforts and careful compromise are needed to safeguard privacy and security in these emerging areas in the face of serious public safety demands.

The right of privacy and free speech will always be weighed against the public demand for law and order and for community standards of decency. Although many of the specific technology and jurisdictional issues are different or somehow transformed in today's virtual environment, the act of balancing competing concerns remains the same.

The management at Prodigy Services, for example, has made a number of controversial decisions regarding extreme behavior online—from a suicide threat to a racist hate group forum in which other subscribers were verbally attacked. The decision to pull the plug on free discussion and provide authorities with personal information that saves lives is not always easy. Many such cases will likely end up in court.

In another case, threats issued by one individual subscriber to another on Prodigy led to an investigation in which three federal agencies, three states, and seven municipalities made similar jurisdictional requests for information. This is another issue demanding guidelines as the networked interactive infrastructure expands. If all legal jurisdictions are to be honored, how are transport and content providers to comply with contradictory requests, or even foreign legal inquiries? What if, for example, a transport provider receives a legal requests for private information from a rogue police state within the virtual boundaries of the network? This raises the potential for causing not only a violation of privacy or human rights but also of sovereignty and national security.

In the absence of exacting legal precedents, service providers have established tough but pragmatic standards for privacy. Such individual voluntary efforts are needed as broader privacy guidelines are tested and cooperation among consumer, business, and government representatives brings a workable consensus.

## ACCESS AND EQUITY

As the networked interactive media revolution ramps up in this country of 250 million people, some of us are inevitably better positioned to leverage it for our own development and prosperity than are others. It is the obligation of business, educational institutions, and government to plan for the broadest possible inclusion of people, bringing diversity of perspective and ideas to the virtual frontier.

The widening fissure between the haves and the have-nots in the United States and around the world is one of the most serious challenges to modern society. Though the issue of poverty defies direct assault, efforts to include, educate, and learn from the economically underprivileged among us will chip away at the core problems and improve the quality of discourse in America.

Interactive product and service providers investing in or supporting the development of a networked interactive media infrastructure must find ways to leverage their limited efforts so that the broadest possible inclusion can be achieved. The two major obstacles are access and the capacity to use.

### Access to Networked Interactive Media

The issue of access is perhaps more varied and complex than privacy within the context of networked interactive media. It is certainly more demanding of the available resources and requires that tough choices be made on an ongoing basis toward some ideal of universality. In general, women, children, old people, poor people, those who are legally blind and those who are illiterate, and nearly the entire continent of Africa are disproportionately absent from the virtual frontier.

In the United States we have to determine who among the underrepresented are least likely to get there without direct action on their behalf.

In computer ownership—a characteristic considered most likely to precede active interest in networked interactive services—income, education, and race are the statistically significant factors.

According to a 1995 U.S. Census Bureau report, households headed by people with a college degree are nearly eleven times more likely to own a computer than households headed by those who did not complete high school. White households with incomes above \$75,000 are three times more likely to own a computer than white households with incomes of \$25,000 to \$30,000. Among blacks, higher earners were four times more likely to own a computer, while the likelihood was five times among Hispanics.

These discrepancies are not likely to improve as the digital revolution progresses. The costs of going online at home are not at all insignificant. At least several thousand dollars must be spent on computer and communications equipment before a consumer can hook up to one of the popular online services. These services charge an average \$10 per month plus up to \$150 or more for 30 hours of connect time and content delivery.

Giving people tax credits to buy PCs, as Newt Gingrich has suggested, may be impractical, but something must be done to provide disadvantaged people with the tools needed to succeed in a knowledge economy.

In the future, direct access to networked interactive services will be an integral feature of commonplace consumer electronic equipment. In the interim, schools, libraries, and other public facilities will likely serve as the point of entry to the virtual frontier for many Americans. Interactive product and service providers are well advised to work with these institutions to determine what facilities and related services will produce the best results for those without access.

### The Capacity to Use Networked Interactive Media

A significant finding in recent years has been the correlation between poor and illiterate segments of our population and the lack of both interest in and access to the new online media. This correlation is dangerous in that this group is one of the most likely to benefit from the eventual mainstreaming of interactive media.

Networked interactive media will serve to fill in gaps left by economic disparity. They could help, for example, address the current lack of affordable training in literacy, work, and communications skills that prevents people from getting and keeping employment. And as networked interactive media become less text intensive and more verbal, visual, and highly intuitive, they will be more easily accessible to those who are less educated or able. The opportunity is upon us to deliver everything from comfort food for couch potatoes to information services for learning and empowerment, particularly with the development of Web-based and interactive television (ITV) services.

Access to these services for people with a limited capacity to pay for or utilize them will hinge on ease of use and basic economics. Unfortunately, delivering such slick applications directly into millions of homes in underprivileged communities during the early stages of commercialization is not realistic. The cost could not reasonably be passed on to the remainder of the customer base or the taxpayer, and the investment community would have no incentive to participate without expectation of returns.

As early test markets prove successful, the cost of rolling out network infrastructure drops, and interest in interactive media increases among low-income people, a solid economic motivation to wire these communities will emerge.

In the short term, cost-sensitive alternative solutions can be considered as a means to boost the online population. Again, the schools and other public facilities will be a critical point of access.

ITV trials and high-bandwidth trials are in the early stages of planning and implementation, so it is premature to draw any conclusions regarding deployment of these technologies.

Concerns have been raised about equitable deployment and, in some cases, charges of "economic red-lining" have been raised. Many content and transport providers appear to be seriously addressing the "red-lining" issue. Bell Atlantic, for example, in its initial deployment plans, specifically addresses this concern. Its currently targeted "hot sites" slated for field trials have a population composed of 36 percent of the cited minority categories in comparison with a 24 percent minority population in the total region. Most LECs and RBOCs also have state educational initiatives in place to address access and equity issues.

In addition, service providers, government agencies, and other organizations might consider limited interactive service into homes or public areas that can be delivered at lower, less costly bandwidths than ITV. Many valuable, albeit streamlined, interactive products could be offered at 14.4 or 28.8 kbps or via ISDN.

New York City's United Community Organization, for example, has installed 200 PCs with ISDN connections to the Internet. The new facilities, paid for with \$1.4 million in federal grants and private donations, are used by UCO staff and neighborhood residents.

### **Realizing Access and Equity**

Information technology is becoming part of the popular vernacular. The networked interactive information media will further this evolution. They will bring good (learning, communication, self-help, and entrepreneurialism) and evil (crime, propaganda, and stultifying mass culture). But ultimately, they must deliver substantial inclusion if they are to transcend the vernacular and serve the larger goals of individual and economic development in a free democratic society.

Larry Irving, assistant secretary of the U.S. Department of Commerce and director of the National Telecommunications Information Administration, says, "It is going to require a concentrated effort not to be left behind.... We have to get the technologies deployed in minority communities, make sure our children are technologically literate, and seize the entrepreneurial opportunities."

Vice President Gore said, "This is not a matter of guaranteeing the right to play video games; this is a matter of guaranteeing access to essential services. We cannot tolerate, nor in the long run can this country afford, a society in which some children become fully educated and others do not."

The networked interactive media infrastructure, described by Vice President Gore as the "Information Superhighway," can evolve into an infrastructure as fundamental as the interstate telephone or electrical power systems, enabling individuals of limited means and capacities to meet their own needs. And like these momentous projects, universality will take time and careful allocation of limited investment resources.

## DEMOCRACY

Networked interactive media have many potential implications for democracy. The extent to which formal and informal aspects of the democratic process in the United States will be transformed is impossible to gauge.

If the Internet's early impact is any indication, it will not be insignificant. The following are a few examples:

- A loose confederation of angry online activists is credited as a key catalyst in the downfall of ex-House Speaker Tom Foley.
- Online constituent service and information are provided by Senator Ted Kennedy and more than a dozen other members of the House and Senate.
- World Wide Web sites are maintained by powerhouse political action groups including the Christian Coalition and the NRA.
- Online participatory government, from town meetings to national referendums, has been proposed by opinion leaders including Newt Gingrich and Ross Perot.

In the Third World, information and communications technologies have already transformed the process of political dissent and even revolution. In Mexico, rebels have waged a public relations war via laptop computer, modem, and fax from the remote state of Chiapas. Their assaults via the news media have captured public sympathy and reversed repeated government offensives. Information is replacing the standard weapons of war.

According to Howard Rheingold, online services are making less dramatic, but no less remarkable, transformations possible in American society. These services, he claims, enable the creation of badly needed "virtual communities" based on shared interests, which hearken back to the world before radio and television, both of which diminished social discourse as a pursuit distinct from work and obligation.

At their best, the networked interactive media will give cohesion and community to underrepresented realms of opinion. At their worst, they will enable terrorists and hate groups to reach wider audiences under cover of anonymity. In all likelihood, they will add yet more dimension and subtlety to the social and political landscapes.

### Information and Community Are Power

Unlike other media that have been co-opted and at times successfully manipulated by political leaders, the Internet is too diffuse and decentralized. It is better suited to affinity-based constituency building than to mass communication.

The Internet and other online services enable the formation of virtual communities built on shared interests. In a fast-paced information age during in it is increasingly difficult for people to assemble and share ideas face to face, the ground for virtual communities is fertile. With thousands of forums and news groups proliferating, cyberspace is host to many of these virtual communities. Some of these have come to function as special interest groups. Whether they facilitate constructive dialogue, involve a greater number of people in the democratic process, or just further derail the deliberative aspect of representative government is a matter of personal perspective.

The intimate "back fence" feel of the virtual community is perhaps its greatest attribute as a forum for political communication. Republican presidential candidate Lamar Alexander, for example, followed his first television appearance as candidate with an extended public question-and-answer session online. He became the first candidate to do so.

Both campaigns and special interests turn to the Internet for information. In addition to news and countless databases, the Internet provides an efficient way to collect competitive intelligence. Speeches, position papers, voting records, and a host of other information can be collected through myriad Web sites. Any forum

open to supporters is also open to the opposition. It is even possible to listen in on discussion groups hosted by opposing candidates or interest groups—a new twist on focus group research.

### **Teledemocracy and Control**

In the 1992 election campaign, Ross Perot proposed the use of technology to support electronic town meetings at the national level, with instant plebiscites on a multitude of issues. This concept of the virtual town hall of "teledemocracy" did not die with Perot's presidential hopes. Similar ideas of varying degree have recently been endorsed by opinion leaders, most notably representative Gingrich. But the impact of a virtual polity may go too far, according to many critics who fear that such immediacy will overwhelm the point and purpose of representative government.

For some, the control of the networked interactive media infrastructure is an important issue for democracy. Where there is control of distribution, there is control of content and therefore opinion that is represented. The openness of the Internet has quashed early fears about freedom of speech and access in the new media. But as major content and transport providers begin to position themselves as major players in networked interactive media, concern swells.

However, the interactivity of the new media, the proliferation of competitive service providers, and consumer demand for diversity of content will limit the influence or manipulative power of any one service provider. Competition for consumer attention, in fact, should intensify efforts to identify and cater to specific audiences. In such an environment, content can more easily reflect the views of participating consumers than in the past.

### **RECOMMENDATIONS**

On regulation and enforcement, we recommend the following: Pure voluntary self-regulation by information collectors and users is a good start, but it can and should be supplemented by government standards to ensure that "bad actors" are dealt with appropriately and that privacy, access, and democratic standards are defined and protected. Consumers face a confusing patchwork of self-administered approaches that vary by state, industry, and company or service provider. Many examples of abuse have already surfaced. Legislation should set minimum standards along the lines suggested here. A joint industry-consumer panel should enforce the standards, educate and consult, and provide dispute resolution services. Disputes that cannot be resolved through this voluntary mechanism should be referred to the courts.

35

**As We May Work: An Approach Toward Collaboration on  
the NII**

*Marjorie Greene*

*First Washington Associates*

**ABSTRACT**

An approach is introduced for information sharing and retrieval on the national information infrastructure (NII) that focuses on the communications paths between collaborative organizations in a telecommunications network rather than upon the content of the information that is stored in distributed databases. Direct connections between domains of information are classified in the form of trails through the network where they can be retrieved and shared in natural language. An application of the approach in shared health care systems is discussed.

The ability to connect to global networks and communicate with millions of people has made every user a publisher ... but just as important, it has made every user an editor, deciding what's important and real. In this medium, you get the filter after the broadcast.

—Paul Saffo, Institute for the Future

Our technologies [will] become more a medium for designing how we should work together, rather than merely designing how we share our information.

—Michael Schrage

**STATEMENT OF PROBLEM**

Contemporary organization theories have suggested that interorganizational networks developed to coordinate activities across organizational boundaries will become the "new institution" of the future. Not only will these networks be important mechanisms for providing superior economic performance and quality but they will also survive, largely because of their "adaptive efficiencies"—that is, because they have the ability to adjust more rapidly to changing technology and market conditions, to produce more creative solutions, and to develop new products and services in a shorter period of time (Alter and Hage, 1993).

Public institutions such as schools, hospitals, libraries, social service organizations and state and local governments are also beginning to work together to provide their own services more efficiently, and they view the NII as an important tool for enhancing their efforts. These institutions will serve as catalysts for further developing the NII and will ultimately create the demands for the private sector's vision of an information superhighway that offers more practical services that address a growing and demographically shifting population.

There are, however, formidable barriers to the deployment of the technologies used in collaborative networks. The recent White House Mini-Conference on Aging highlighted a major problem when it pointed out the need for "a standard protocol linking national, local, 'on-line,' off-line, public, non-profit and private databases" in delivering services to the elderly. "Differing classification schemes, confusing terminology, and lack of 'info-glut' screening mechanisms" are limiting access to information and preventing the effective delivery of integrated care ("Accessing Eldercare via the Information Highway," 1995).

The vision of linkages between users of patient information within communities in which each health care facility and practitioner would connect to a network through an information system is greatly hindered by the inability to create, store, retrieve, transmit, and manipulate patients' health data in ways that best support decision making about their care. This is the problem that is addressed in this white paper. It is hoped that the approach presented here for information classification and retrieval through the NII will lead to further investigation of its potential.

## RELATED INITIATIVES

### Community Networks

Several efforts are already under way to promote the widespread use of advanced telecommunications and information technologies in the public and nonprofit sectors, especially at the community level. (See, for example, the U.S. Department of Commerce National Telecommunications Information Administration/TIAP initiatives.) The private sector is also beginning to explore the use of information technology in community networks, including those designed to support and enhance collaboration among health and human services providers (Greene, 1995). Eventually, a system of "global, shared care" is expected to evolve in which the coordinated activities of different people from different institutions will apply different methods in different time frames, all in a combined effort to aid patients medically, psychologically, and socially in the most beneficial ways. Because the ability to move data is considered fundamental to the process of integrated care, attempts have been made to find cost-effective ways to share data among the participants. However, this approach has been fraught with difficulties that are largely unrelated to the ability of the technology to provide solutions. Questions of ownership, confidentiality, responsibility for health outcomes, and semantics are paramount, and clinicians are themselves calling for new solutions that do not require "knowledge" to be formalized, structured, and put into coding schemes (Malmberg, 1993).

### The European Approach

Many Europeans have also recognized that one of the major problems in designing the shared care system is management of the communications process among the different institutions and health care professionals. They are taking a different approach and conducting field studies to evaluate the feasibility of using patient-owned, complete medical record cards, which patients would carry with them and present to the institution carrying out the treatment. Although they recognize the importance of natural language processing and the potential of optical-storage technology to reduce costs, they conclude that the technology will only be available within the respective information systems that contain medical records and that new solutions such as the chip card of the hybrid card must be found in order to extend communication to all health care providers (Ellsasser et al., 1995).

### The Digital Library

Information sources accessed through the NII also represent components of emerging universally accessible, digital libraries. The National Science Foundation, in a joint initiative with the Advanced Research Projects Agency and the National Aeronautics and Space Administration, is supporting research and development designed to explore the full benefits of these libraries, focusing on achieving an economically feasible capability to both digitize existing and new information from heterogeneous and distributed sources of information and to find ways to store, search, process, and retrieve this information in a user-friendly way (National Science Foundation, 1994). It has been suggested, however, that "for digital libraries to succeed, we must abandon the traditional notion of 'library' altogether.... The digital library will be a collection of information services; producers of material will make it available, and consumers will find and use it" (Wilensky, 1995). New research is needed



that is fundamental to the development of advanced software and algorithms for searching, filtering, and summarizing large volumes of data, imagery, and all kinds of information in an environment in which users will be linked through interconnected communications networks without the benefit of preestablished criteria for arranging content.

### VANNEVAR BUSH AND THE NEW TECHNOLOGIES

The concept of a dynamic, user-oriented information system was introduced as early as 1945, when Vannevar Bush suggested that an individual's personal information storage and selection system could be based on direct connections between documents instead of the usual connections between index terms and documents. These direct connections were to be stored in the form of trails through the literature. Then at any future time the individual or a friend could retrieve this trail from document to document without the necessity of describing each document with a set of descriptors or tracing it down through a classification scheme (Bush, 1945).

In 1956, R.M. Fano suggested that a similar approach might prove useful to a general library and proposed that documents be grouped on the basis of use rather than content (Fano, 1956). This suggestion was followed 10 years later by a pioneering contribution of M.M. Kessler at the MIT Technical Information Project, who developed a criterion for such grouping of technical and scientific papers through "bibliographic coupling," in which two scientific papers cite one or more of the same papers (Kessler, 1965). This concept of bibliographic coupling has been extended to other types of coupling and refined to the present day, largely through computer-based techniques that identify sets of highly interrelated documents through "co-citation clustering" (Garfield, 1983).

Although it was recognized that the model of "trails of documents" as suggested by Dr. Bush 50 years ago had useful features that the subsequent partitioning models did not offer, research has not been conducted on its potential for classification and retrieval in modern communications networks. Perhaps this would be a good time to revisit the concept, especially as traditional computer-based systems are merged with communications systems in a network of networks such as the NII. And because citation characteristics are an indication of how scientific doctrine is "built," we might want to combine the idea of trails of documents (represented as "communications paths") with sets of documents (represented as "domains of information") into a more general model that can be used for both classification and retrieval of information. Such a model has been developed for military "command and control" and is presented here for further consideration by the NII community.

### ANALYSIS

Message traffic among higher-echelon commands during the early part of a crisis situation is extremely difficult to classify. This is because such communications do not generally fall into categories that deal with specific predetermined military tasks, but instead are much less precisely defined, less routine, and consist primarily of the exchanges of information along with recommendations, advice, and other messages that are necessary before any tactical systems can be put into effect. By the same token, these communications are difficult to retrieve in any formatted sense because the unexpected, evolving, and interdependent nature of the information places an even greater emphasis upon natural language communication.

In an attempt to avoid the inadequacies inherent in any classification system while at the same time recognizing that as the amount of available information grew there was a parallel need for a more precise way to retrieve specific data, a technique was developed for associating messages with each other that required no interpretation of the subject content of the messages (Greene, 1967). This technique is based upon the thesis that if a message referenced a previous message, the previous message must have influenced that message in some way. For example, a message might say, "This is in answer to your question in reference A." Often a message referenced a previous message that referenced a yet earlier message. Still other connections of messages through their references are possible.

In **Figure 1**, if each number represents a message and if an arrow from 2 to 1 means 2 referenced 1, then we can interpret **Figure 1** as follows: message 2 references message 1, message 4 references message 2 but also references message 3, and message 5 is another message that references message 3. Thus, we can speak of a "reference-connected" set of messages  $S = (1, 2, 3, 4, 5)$ —that is, a set of messages that are connected in any way through their references. (This concept is analogous to the one of "joining" in directed graph theory.)

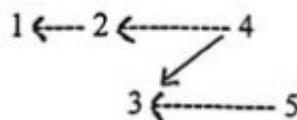


Figure 1 A reference-connected set of messages.

It is noted that in **Figure 1**, messages 4 and 5 are "bibliographically coupled." Another type of coupling occurs if two papers are cited by one or more of the same papers (e.g., 2 and 3). And finally, there is the simple citation relationship between 1 and 2, 2 and 4, 3 and 4, and 3 and 5. These three basic types of reference connectivity have been used as separate partitioning criteria for retrieval systems in the past. However, they have not been combined into a single dynamic system for both classification and retrieval, nor have they been used to link databases for interorganizational collaboration, as this white paper suggests.

It was found that during the early part of a crisis situation when messages throughout the command structure and in different locations were put into reference-connected sets, these sets in most cases uniquely identified particular events during the crisis. For example, one set that was constructed from crisis-related message traffic found in files at three command headquarters contained 105 distinct messages that dealt with the preparations for landing airborne troops. Other sets of messages represented the communications related to other events such as the provision of medical supplies, the preparation of evacuation lists, and sending surgical teams. All of these events were represented by unique message sets in the investigated files of crisis-related traffic.

Reference-connected sets proved to be valuable tools in analyses of command information flow as well as of the operations they describe. Deficiencies in flows and use of information were much more easily identified when focus was placed upon a specific event represented by communications throughout an entire command structure. The natural application of these sets to information retrieval was also noted because it was possible to file messages automatically into appropriate message sets by noting only the references that were given. These sets then represented events during a crisis and were available for answering queries regarding their status. Predetermined subject categories were not required, nor were any restrictions placed upon the format of the messages. The method simply provided a way of quickly locating a message that had the information (as it was expressed in natural language) that was necessary to make a decision.

### Automatic Classification

A simple filing method was used in the analysis for automatically classifying messages into reference-connected sets. If a message referenced a previous message, it was put into the file of the previous message. So, for example, in **Figure 2**, message 2 would be filed with message 1 because it referenced message 1. Message 3 does not reference a previous message and would thus begin a new file 2. However, message 4 referenced messages in both files and therefore connected the two. Two subsets were identified in this way. One subset (assigned the number 1) contained messages 1, 2, and 4. The other subset (assigned the number 2) contained messages 3, 4, and 5. Message 4 is the link between them and, in the language of directed graph theory, may be considered to be a linking point between two maximal paths in the semipath from message 1 to message 5.

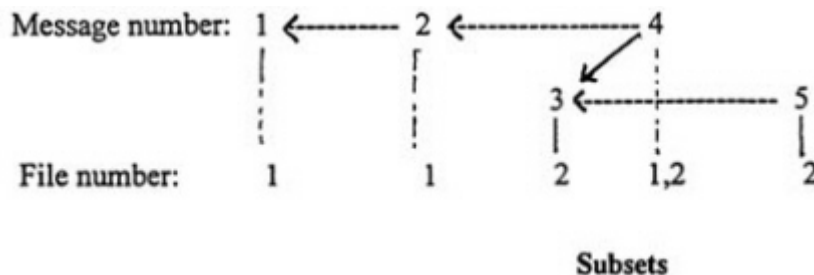


Figure 2 Automatic classification of a reference-connected set.

The structure of a reference-connected set identifies subsets (as in the preceding section) that can be interpreted in a number of ways. For example, it is noted that a subset will occur only if there is a message within the set (such as 3 in Figure 2) that does not reference a previous message but that is eventually linked to the set. Such a message may begin a "new" event that eventually becomes related in some way to the earlier event initiated by message 1. However, the structure of a reference-connected message set is also a function of another important factor—the organizational chain-of-command and the distribution of information throughout this chain. For a message cannot reference a previous message unless its author is cognizant of the previous message. Consequently, the paths in a reference-connected set (and thus the corresponding subsets) will often reflect the information flow between specific commands although the event is essentially the same.

It is easily seen that this model can be extended and adapted to other interorganizational networks in which information is exchanged to meet a common goal, such as provision of health care. The application of the model also becomes more complex as additional nodes are included and multiple addressees are allowed. Nevertheless, two important characteristics should be noted that illustrate this model's potential in supporting the collaborative process:

- The subsets of a reference-connected set often correspond operationally to the sets of messages received at various nodes for the same event and therefore define "domains of information" stored and processed at each node in the network for that event or, in the case of health care, patient episode. (This concept can be pursued further in order to address concerns about excessive centralization in integrated systems, to protect patient privacy, and to measure the benefits of collaboration—i.e., health outcomes—in terms of the costs assumed by individual participants. These are all potential advantages of this approach when applied to health information networks.)
- Rules of referencing could be established that would guarantee that certain sets and subsets would appear. For example, if each node in the network has a record (through the references) of all previous communications dealing with an event, all of the messages would automatically form a reference-connected set at each node even though all of the messages were not processed at every node. (Again, there are advantages here for social service agencies that would like to "track" clients without sacrificing client confidentiality or losing control over their own administrative processes.)

### CONCLUSION AND RECOMMENDATIONS

In a medium in which "the filter comes after the broadcast" and in which users everywhere have direct access to the full contents of all available material, finding information will be a key problem. How can a classification system be developed for a communications-based system in which the unexpected, evolving, and interdependent nature of the information places even greater emphasis on natural language? New approaches will have to be found that avoid both the problem of describing the content of information and the problem of integrating new information into a predetermined classification code. The collaborative networks of the future will focus on information flows. They will lead to dynamic user-oriented information retrieval systems that are

based on communications paths and direct connections between distributed information sources rather than upon technologies that mechanically or electronically select information from a store. New paradigms of interaction appropriate for multimedia distributed systems will be the focus of new technologies, and automated, intelligent search agents will be found that help consumers as well as providers to find and use what is important and real.

New technologies, combined with the concept of reference-connected sets, may offer another potential solution to the management of the communications process among different institutions in collaborative networks. Future research on community networks should be focused on the operational level rather than the administrative level by linking users of information from the "bottom up" and by searching through communications paths rather than through the content of the information that is stored in distributed databases. This would give communities an opportunity to assess the role of the NII without large investments in technology and would allow participating organizations to gain the economic benefits of the network only in so far as there is a need to collaborate.

An approach is presented here that does not attempt to guide users through the vast domains of information that will be available through the NII. Instead, it helps them to find quickly the others user within their community of interest that may have the information they are seeking. This approach could provide the protocol needed to link national, local, "on-line," off-line, public, nonprofit, and private databases for increased access to collaborative networks. It could also enable providers of health and human services to work together to aid patients medically, psychologically, and socially in the most beneficial ways. It is a tempting approach

## REFERENCES

- "Accessing Eldercare via the Information Highway: Possibilities and Pitfalls," a 1995 White House Mini-Conference on Aging, March.
- Alter, C., and J. Hage. 1993. *Organizations Working Together*. Sage
- Bush, Vannevar. 1945. "As We May Think," *Atlantic Monthly* 176(1):101-108.
- Ellsasser, K-H., Nkobi, J., and Kohler, C.O. 1995. "Distributing Databases: A Model for Global, Shared Care," *Healthcare Informatics*, January.
- Fano, R.M. 1956. *Documentation in Action*, Chapter XIV-e, pp. 238-244, Reinhold Publishing Corporation, New York.
- Garfield, E. 1983. *Citation Indexing—Its Theory and Application in Science, Technology, and Humanities*. ISI Press, Philadelphia.
- Greene, M.J. 1967. "A Reference-Connecting Technique for Automatic Information Classification and Retrieval," Research Contribution No. 77, Operations Evaluation Group, Center for Naval Analyses, The Franklin Institute, March.
- Greene, M.J. 1995. "Assessing the Effectiveness of Community Services Networks in the Delivery of Health and Human Services: An Economic Analysis Model," research conducted under HRSA Contract No. 94-544 (P), March.
- Kessler, M.M. 1965. "Bibliographic Coupling Between Scientific Papers," *American Documentation* 14(1):10-25.
- Malmberg, Carl. 1993. "The Role of Telematics in Improving the Links Between Primary Health Care Providers," Annual Symposium on Computer Applications in Medical Care.
- National Science Foundation, Digital Library Initiative, FY 1994.
- U.S. Department of Health and Human Services. 1993. *Toward a National Health Information Infrastructure*, report of the Work Group on Computerization of Patient Records, April.
- Wilensky, R. 1995. "UC Berkeley's Digital Library Project," *Communications of the ACM*.

## 36

# The Use of the Social Security Number as the Basis for a National Citizen Identifier

*W. Ed Hammond*  
*Duke University Medical Center*

### STATEMENT OF THE PROBLEM

Although there is little disagreement that some type of unique universal citizen identifier is necessary for creating a complete, lifetime, patient-centered, computer-based health record, there is considerable disagreement over what that number should be. This paper makes the argument that a number derived from the Social Security number (SSN) and administered by the Social Security Administration (SSA) is the best and most economical solution to this problem. Arguments against the SSN are, for the most part, arguments about any identifier that might be universally used to identify individuals for bringing together all data relating to their health care.

New models for health care delivery, particularly managed care, can be fully supported only through an integrated, electronic information system. The concept of a lifetime, patient-centered health record containing, at least logically, all data from all sources are key to delivering high quality, cost-effective care. Patients receive that care from a variety of providers in a variety of settings. The information system must be able to aggregate data about a person into a single, logical record. To do this integration, the identity of a person must be unequivocally established in the sending and receiving systems.

There are two different problems in establishing patient identity. The first problem is to establish the identity of a person with respect to a presented identification number. This process is called authentication, and several options are available. In the past, authentication has usually been accomplished by a person presenting a card with an identification number. Biological identifiers, such as a thumb print reader, are becoming affordable and can establish a person's identity with a high degree of certainty. The other problem occurs when data are being transferred between two systems, and the patient is not available.

Some people propose the use of demographic data such as a person's name, date of birth, mother's birth name, and/or address. Inconsistency in these data parameters are a source of trouble. In the case of a name, comparison of databases shows inconsistency in the use of name order, full names versus initials, nicknames, and the occasional omission of suffixes, such as Jr. or Sr. Many people have multiple addresses, mailing addresses, home addresses, and incomplete entries. The listed date of birth, particularly in the medical records setting, may be in error. The literature and my own experience suggest that approximately 30 percent of the entries in two databases that are being merged require resolution by a human.

### BACKGROUND

The Social Security Act was signed into law on August 14, 1935, by Franklin D. Roosevelt. The Social Security Board recommended the adoption of a nine-digit numbering system for identification purposes and was granted authority by the Treasury Department on November 1936 for the assignment of numbers to people who were employed. The SSN is a nine-digit number broken into three groups. Its form is 999-99-9999. The first three digits, called the area number, are determined by the address shown on the application for the SSN (now based on Zip code). Initially, the United States was divided into 579 areas numbered from 001 to 579. Each state was assigned certain area numbers based on the number of people in the state expected to be assigned an SSN. At present, area numbers 001 through 647, with the exception of 588, have been assigned. In addition,

area numbers 700 through 728 were assigned to railroad workers until 1963, at which time the practice was discontinued. The area number has little meaning today due to the mobility of people. The next two-digit group, called the group number, has no special significance except to break the numbers into convenient blocks. The last four-digit group, called the serial number, is assigned sequentially within each group. Note that no group contains only zeroes.

In a study done at Duke University, examining the SSNs of approximately 150,000 individuals, the last six digits of the SSNs were uniformly distributed. This uniform distribution is particularly valuable for certain hash-code indexing techniques.

In the 1960s, the use of the SSN spread to the Internal Revenue Service for tax purposes, the Department of Defense for military designation, and the Civil Service Commission for employee identification. In 1976, states were authorized to use the SSN for tax purposes, public assistance, and for driver's license or motor vehicle registration. A number of states use the SSN on the driver's license.

## ANALYSIS AND FORECAST

### Value of a Universal Citizen Identifier

Simply put, the most reliable method of integrating data from multiple sources is to have a unique identification number known to all sources. In the absence of such a number, combining data from multiple sources or even reliably identifying a person within a single source is difficult. If we fail to identify a person in the health care environment, that person's data are split into multiple records and valuable data are misplaced.

Community health care information networks (CHINs) and statewide alliances are becoming popular in which health care information about a person is available, with proper safeguards, to those people responsible for a patient's care. Failure to associate known health care data about a patient can lead to serious consequences. For example, if the patient is allergic to a certain drug and he or she is misidentified and that information is not available, that important point could be missed. If, in fact, we believe that information about the patient's health, medications, allergies, problems, and treatment plans is important, then we must be sure that the information is available to the proper health care providers. The highest probability of making that happen is through the use of a unique universal identifier.

### Requirements for a Universal Citizen Identifier

The universal citizen identifier (UCI) must be unique. Each person must possess one and only one identification number. A UCI number, once assigned, can never be reassigned. A UCI should be assigned at birth or when a person becomes a resident of this country.

The UCI should be context free. The UCI is a pointer to data about a person. It should not attempt to convey any information about gender, age, or geographical area where a patient was born or now lives. Its sole purpose is to link the number to one or more data banks.

A system must be established for creating an identification number for foreign visitors and illegal aliens. Such a number must also possess the characteristic of uniqueness and must never be reassigned. We now have international telephone numbers that use a country code. These numbers are of various lengths and format. We might use a similar scheme for personal identifiers. The popularity of international travel and the availability of the Internet make it particularly feasible to transmit a person's health record to any country. A known identification number would make that process more reliable.

One of the commonest errors that results in the misidentification of a patient, even with the use of a patient identification number, is the transposition of two numbers. The use of a check digit would provide a solution. There are several check digit algorithms. Generally the check digit is generated by multiplying each digit of the identifier, in order, by a weighted multiplier. The resulting product is divided by some number and the remainder is taken as the check digit. This digit becomes part of the identification number and is entered into the

computer. The computer, in turn, calculates the check digit and compares it to the entered number. If they match, the entered number is assumed correct. If it is different, the number is rejected. ASTM recommends the use of a lookup table to determine the check digit.

The UCI should use both letters of the alphabet and numerals to make up the identification number. Certain letters, which might be mistaken for numerals, should be omitted. Examples are the letters "O" and "Q," which might not only be mistaken with each other but also with the numeral "0." If lower case letters are used, the letter "l" might be mistaken for the numeral "1." In any case the number of unique combinations of some 30 elements—and with lower case some 62 elements—would more than handle the population of the world for a long time. For economic reasons, I recommend that numerals be used as long as unique combinations are available, and that letters then be added one position at a time. Most legacy systems could accommodate numerals without a problem, and there would be ample opportunity to plan for the accommodation of letters.

### Validation of the UCI

The biggest problem with any personal identifier system is establishing and maintaining an error-free link between the actual person and the associated number. The Internal Revenue Service recently reported 6.5 million cases of missing, invalid, or duplicate Social Security numbers (Fix, 1995). Most of these errors were the result of recording errors. Other duplications occurred in connection with an attempt to defraud the IRS. In one case, an SSN was used more than 400 times. There is no question that duplicate SSNs exist. One story suggests that when the announcement of the SSN program was published in the newspapers, a sample SSN was included. Many people apparently thought that this number was what they were supposed to use and accepted that published number as their SSN. Another story is that many people, in purchasing a new wallet that included a dummy SSN card, accepted that number as their SSN. In some cases, the SSA apparently reissued SSNs. In other duplications, people have simply made recording errors and have been using incorrect numbers for many years. Increased use of the SSN has resulted in a significant reduction in these duplications for a number of years.

Validation of the UCI will require the creation of a database containing demographic and identifying data about every resident of the United States. Considerable thought is required to define this database, and it will ultimately be a trade between what is required to identify an individual uniquely and what should not be included to protect the rights of the individual. This database could be used for other purposes as well. Certainly, the existence of such a database would reduce the effort of producing a census and of being able to do population-based statistics. Many citizens would not be concerned about the existence of such a database; others would consider any database an invasion of privacy. Nonetheless, everybody is already in many databases and the anonymity of these databases permits easy abuse. Legislation would be required to protect the contents and use of such a database. This topic is explored below.

Keeping a UCI database up to date would be a difficult challenge. Some items should never change, others might change infrequently, and others might change with some frequency. Elements in the database would include a person's name, gender, marital status, race or ethnicity, date of birth, and address. Persons would be responsible for informing the agency of change, perhaps as part of some annual event.

### Arguments for Advantages of Using the SSN Over Other Proposals

Under the assumption that a personal identifier system is selected, that system would have to be administered by some agency. One possibility is that a private, trusted authority could be given the responsibility of assigning the UCI and maintaining the accompanying database. Another possibility is that a new government agency could be created to administer the UCI. Another option is to use the existing SSA to administer the UCI program. Setting up a new agency with the accompanying bureaucracy would take longer and cost more than using an existing agency.

There are over 1300 Social Security offices distributed around the United States where a person can apply for and receive a Social Security number. Evidence of identity, age, and U.S. citizenship or lawful status is required. All applicants over the age of 18 must apply in person. Individuals under age 18 or those seeking replacement cards may apply in person or by mail. Nonwork SSNs may be assigned to illegal aliens if they receive benefits payable in some part from federal funds.

SSNs are assigned at the SSA's central headquarters in Baltimore. Key data elements are a person's full name, date and place of birth, mother's maiden name, and father's name. These elements are used to screen the SSA database to prevent the assignment of more than one number to the same person. If no match occurs, a new SSN is assigned. If a significant match occurs, a replacement card is issued. The current system assigns an SSN within 24 hours of receipt of the application. Cards are sent by mail and usually require 7 to 10 days for delivery.

Beginning in 1989, the SSA began a program in which an SSN can be assigned to a child as part of the birth registration process. This procedure currently requires a parent's approval. The percentage of birth registrations including a request for an SSN is more than 75 percent and is increasing.

As of March 3, 1993, 363,336,983 SSNs had been issued. The number of currently active SSNs (of living people) is estimated to be approximately 250 million. It is estimated that approximately 4 million individuals may have more than one number.

The Privacy Act of 1974 (5 U.S.C. 552a) states that it is unlawful for any federal, state, or local government to deny an individual any legal rights or benefits because the individual refuses to disclose his/her SSN. There is no legislation concerning the use of the SSN by nongovernment entities.

The SSA recently announced that the agency was undergoing a reorganization. My recommendation is to give the SSA the tasking authority and the required funding to administer a UCI program.

### **The Case for a Single Identifier for All Purposes**

The increasing use of the SSN for identification purposes supports the argument that a universal, unique identifier has value. An individual's having only one number that he/she would use for any identification purpose would represent a considerable savings for federal agencies, vendors, health care agencies, and any other organization that creates a database. The suggestion that a single number could be used to access patient data in any of these databases or to join data from any database regardless of purpose or owner is frightening. Yet, in this age of connectivity and computerization, it is a trivial problem to link any number system, particularly if 100 percent accuracy is not sought. Anyone who thinks that confidentiality is preserved by requiring different numbers is misinformed. I would argue the opposite. Given a single number, it would be possible to provide more positive controls in making sure that the number is not misused. I therefore recommend that the UCI be permitted to be used in any legal operation subject to the individual's approval.

### **Confidentiality, Privacy, and Security**

In a recent opinion poll conducted by the Louis Harris organization, 85 percent of those polled agreed that protecting the confidentiality of people's medical records is essential (Louis Harris and Associates, 1993). In that same pool, 67 percent indicated a preference for the SSN as the preferred national health care ID number.

There can never be any security in a publicly known personal identifier. Security and protection of an individual's privacy must be provided through each database and the supporting applications. All individuals have certain rights relating to who sees data about them, how those data are used, and the opportunity to review and correct errors in the database. In the case of health care data, the patient should be able to define, in writing, by whom and under what circumstances those data may be used. On the other hand, a health care provider should be told when data are being withheld and, except in emergency situations, should be able to refuse treatment. In an emergency situation, if the provider makes an incorrect decision due to lack of complete information, that provider should be protected from malpractice lawsuits. Individuals should be able to request a list of all persons who have accessed their data.



The inability to correctly identify a patient's data from some type of patient ID might actually result in less, rather than more, protection of confidentiality. For example, if a patient indicated that "my primary care physician" could see the record and the patient's ID did not match the record, such a discrepancy would permit inappropriate access to data. Overly strict rules and computer-enforced rules are risky where patient care is involved. Blocking health care providers from access could lead to serious consequences in the case of an emergency. Proper education of users of data and emphasis on the need to preserve confidentiality are of particular importance.

Federal legislation must be passed making it illegal to acquire data of any type against a person's written wishes. If legally or illegally acquired data are used for purposes for which they were not intended, the individual acquiring the data should be punished by law. Such action should be considered to be as serious as a bank robbery, and punishment should be similar. Individual confidentiality can only be assured through legal constraints. It cannot be achieved through confusing identifiers that might prevent databanks from being accessed or linked.

Federal legislation should also spell out the security requirements required for each organization that would use the UCI as the pointer to data contained within the databank. Each of those organizations should be required to have an information security officer who would ensure that confidentiality and security requirements were met.

### RECOMMENDATIONS

I recommend that legislation be passed that will task and fund the SSA to be the administrator of a universal citizen identifier, which may be used for a variety of purposes as a patient identifier. Use of this number for a databank must be requested by an organization and approved by the SSA. Access to data must be logged by individual and organization, date and time, and purpose. The UCI would be based on the SSN and would be the currently assigned SSN plus a check digit. The SSA, in establishing the validating databank, would eliminate duplicates. An added advantage of this approach would be eliminating errors in calculating and paying Social Security benefits.

New UCIs would be issued electronically to newborns and to individuals moving to this country, either as citizens or as legal entrants. Illegal aliens would be assigned a number from a selected and identifiable set. Foreign visitors would also be assigned a permanent number. Legislation protecting the use of the UCI and guaranteeing protection of the rights of an individual would be simultaneously introduced.

Electronic access to a regional office would be by Internet, a state information network, or even by modem. Information would be transmitted electronically. That information would be verified before the assignment of the UCI was made permanent. Special efforts would be made to avoid fraud. SSN cards would be coded to make creation of false cards very difficult.

The American College of Medical Informatics, of the American Medical Informatics Association (ACMI, 1993), the Computer-based Patient Record Institute, and the Working Group for Electronic Data Interchange have all recommended the use of the SSN as a UCI. Several states are now using the SSN for identification purposes, including in the management of health care benefits. Many third-party payers use the SSN as the basis for the subscriber identification.

We recognize the emotional issues associated with the use of a UCI (Donaldson and Lohr, 1994; Task Force on Privacy, 1993). Those emotions are correct and understandable. Unfortunately, the suggested solution of not having a universal identifier, or even of restricting such an identifier to use only in the health care setting, will provide little protection. Instead, open use of an identifier with safeguards and audits will provide greater protection. The advantages of being able to integrate personal health care data over a variety of settings and systems far outweigh the risks of such a system. The important thing is to recognize that the use of a universal health care identifier, and specifically the SSN, does not in itself mean a lack of concern for patient confidentiality or an inability to preserve that confidentiality.

Already we are paying a penalty for the lack of such an identifier. Time is important. Now is the time for action.

### ACKNOWLEDGMENT

Much of the information relating to the SSA was taken from an early draft of an ASTM document (ASTM, 1995), "Guide for the Properties of a Universal Healthcare Identifier," written by Dr. Elmer R. Gabrieli and provided by Andrew J. Young, deputy commissioner for programs, Social Security Administration.

### ADDITIONAL RESOURCES

- American College of Medical Informatics (ACMI). 1993. "Resolution for the Use of the SSN as a Universal Patient Identifier," ACMI, Bethesda, Md., February.
- ASTM. 1995. "Guide for the Properties of a Universal Healthcare Identifier," draft proposal developed by ASTM, Philadelphia, Pa., January.
- Donaldson, Molla S., and Kathleen N. Lohr (eds.). 1994. *Health Data in the Information Age: Use, Disclosure and Privacy*. Institute of Medicine, National Academy Press, Washington, D.C.
- Fix, Janet L. 1995. "IRS Counts 6.5 Million Errors So Far," *USA Today*, April 5.
- Louis Harris and Associates (in association with Alan Westin). 1993. *Health Information Privacy Survey 1993*. A survey conducted for EQUIFAX Inc. by Louis Harris and Associates, New York.
- Task Force on Privacy. 1993. *Health Records: Social Needs and Personal Privacy*. Task Force on Privacy, Office of the Assistant Secretary for Planning and Evaluation and the Agency for Health Care Policy and Research, Washington, D.C., February 11–12.
- Work Group on Computerization of Patient Records. 1993. *Toward a National Health Information Infrastructure*. U.S. Department of Health and Human Services, Washington, D.C., April.

## 37

# Estimating the Costs of Telecommunications Regulation

*Peter W. Huber, Manhattan Institute*

*Boban Mathew, Yale University*

*John Thorne, Bell Atlantic*

The licensing schemes embodied in federal telecommunications regulation forbid people to sell goods or services without the say-so of a federal regulator. They frequently reflect the conclusion that regulators can allocate goods and resources and set prices more efficiently than can market forces. In some instances (declining in number as new technology and the globalization of most markets spur competition) this may be true. Unregulated "natural monopolies," for example, may price too high and produce too little.

But the beneficial effects of regulation (such as they are) can only be realized if regulators perform their functions efficiently, on schedule, on the basis of up-to-date information, and to protect the public, not industry incumbents. The record of federal telecommunications regulation is not good. Routine licensing decisions take far longer than they should. When issued, licenses are loaded up with restrictions and conditions that serve mainly to promote bureaucratic control. Regulation often solidifies the economic status quo, protects incumbents against would-be competitors, and deprives the public of new services at lower prices.

This paper examines the costs that restrictions on the use of both the electromagnetic spectrum and the wireline impose. The largest component of costs attributable to such zoning is the lost opportunity cost of preventing higher-value uses of the airwaves, and of restricting socially desirable uses of wireline media. Outright restrictions and delays in approving new uses of the two media also hamper competition in existing markets for telecommunications services, to the detriment of consumers in those markets. Finally, zoning often leads to the existence of large economic rents actively sought by market players through socially wasteful activities. It is, of course, impossible to measure precisely the total social cost of zoning restrictions.<sup>1</sup> But conservative calculations of specific restrictions suggest that the costs are very high.

In this paper, we describe the zoning process and estimate the economic costs that several of the restrictions entail.

### WIRELESS ZONING

The airwaves are "owned" by the federal government. The government licenses private users for fixed periods. With few exceptions, the licenses are given away; they are not sold. For the most part, the licenses strictly prescribe how the spectrum is to be used.

---

NOTE: Peter Huber is a senior fellow, Manhattan Institute for Policy Research; Boban Mathew, M.A./M.Phil Economics 1995, Yale University, is a J.D. candidate 1996, Harvard University; John Thorne is vice president and associate general counsel, Bell Atlantic. The authors wish to thank Evan Leo for significant assistance and research in the preparation of this paper. The views expressed in this paper are those of the individual authors; they should not be attributed to any company or client.  
Copyright © National Academy of Sciences. All rights reserved.

The Modified Final Judgment (MFJ)—the consent decree that broke up the old Bell system—imposes further zoning requirements on the networks of the Bell operating companies.<sup>2</sup>

### **Zoning the Airwaves**

The federal government began to nationalize the airwaves in 1912, when Congress gave the secretary of commerce authority to license broadcasters.<sup>3</sup> But most empty airspace could still be occupied freely. "Homesteaders" simply had to register their claims with the Department of Commerce. No exclusive rights were assigned.<sup>4</sup> By the mid-1920s, courts were beginning to affirm private property rights in spectrum.<sup>5</sup>

The Radio Act of 1927, however, placed almost every aspect of radio broadcasting under the control of the newly created Federal Radio Commission (FRC).<sup>6</sup> Seven years later, the provisions of the 1927 act were rolled, largely intact, into the Communications Act of 1934.<sup>7</sup> The FRC became the Federal Communications Commission.

The licensing of broadcasters is conceptually straightforward. The FCC first zones the real estate, allocating blocks of spectrum for particular uses such as AM radio, FM radio, VHF TV stations, UHF TV stations, and so on. Within each block, it then assigns licenses to particular users. The commission has virtually unbounded discretion in both regards. The law simply requires distribution of broadcast "licenses, frequencies, hours of operation, and power among the several states and communities so as to provide a fair, efficient and equitable distribution of radio service to each of the same."<sup>8</sup>

However chosen, licensees do not get a formal property right in their spectrum. The 1927 Radio Act expressly declared that licensees were entitled to the "use of channels, but not [to] the ownership thereof."<sup>9</sup> Licenses were to run for only "limited periods of time."<sup>10</sup> (Only in 1981 were the original 3-year broadcasting license terms extended to 5 years for television and 7 years for radio.<sup>11</sup>) Licenses may not be transferred without commission approval.<sup>12</sup> The commission may revoke a station license for any reason that would have warranted refusing a license in the first place.<sup>13</sup>

### **Zoning of Cellular**

The provision of cellular service is zoned in several ways. The allocation of spectrum for cellular services was originally split between telcos and other nonwireline carriers.<sup>14</sup> In 1981, the commission decided that two (and only two) cellular carriers would be licensed in every cellular service area.<sup>15</sup>

A quite different and independent set of zoning requirements has come into existence by way of the MFJ. The MFJ's line of business restrictions preclude Bell cellular affiliates from offering "interexchange" services. Bell cellular affiliates thus may not arrange with a particular interexchange carrier to provide discounted service to their customers.

### **WIRELINE ZONING**

In contrast to the airwaves, wireline networks are privately owned. But wireline media are zoned even more strictly than the airwaves. Local telephone facilities are still "zoned" to provide mostly voice services. For years, cable television operators were strictly "zoned" to supply simple carriage of broadcast video signals; to this day they still operate under an array of quasi-common-carrier and other zoning obligations that sharply diminish the value of cable networks and greatly reduce economic welfare.

### Zoning of Telephone

Section 214 of the 1934 Communications Act prevents phone companies from constructing new facilities or discontinuing existing service without advance permission.<sup>16</sup> Section 201 of the act enables the commission to regulate what sort of devices can be connected to the telephone network, and thereby what kind of add-on services can be offered over telephone lines.<sup>17</sup> The FCC has used its authority under both statutory provisions to zone the telephone to provide basic voice services, and not much else.

The franchising of telephones was initially a way of reducing the confusion and hubbub of competition.<sup>18</sup> The useless duplication of facilities, Congress believed at the time, would lead to higher charges to the users of the service.<sup>19</sup> What started as a means of suppressing excessive telephone competition in an industry marked by declining average costs, however, has become a tool for suppressing telco involvement in broadband services and, when such services are permitted, regulating them in minute detail.

Video services are largely zoned out of the telephone franchise, too, though phone companies have recently been successful in challenging some of these regulations on First Amendment grounds. In 1968, the FCC declared that cable television was an "interstate" service and that telephone companies therefore needed FCC permission to build or operate cable networks.<sup>20</sup> In 1970, the commission barred telephone companies from providing cable TV service.<sup>21</sup> In 1984, Congress codified this prohibition in the new Cable Act.<sup>22</sup> The "cable television service" language of the 1970 rules was replaced with "the provision of video programming."<sup>23</sup> The prohibition extends to everything that can be characterized as "providing" video over the telephone company's own network, including selection and packaging for resale of programming created by others.<sup>24</sup>

A second sphere of telephone-wire zoning involves the FCC's two-decade crusade against allowing "enhanced services" into the telephone network. In 1966, the commission undertook to examine the convergence of computers and communications.<sup>25</sup> In an abundance of caution, the commission insisted that "enhanced services" should not under any circumstances be intermingled with basic phone service. For a 20-year period, from the mid-1960s to the mid-1980s, the commission enforced a policy of "maximum separation" between familiar voice services and network-based data processing, electronic services, and computers.<sup>26</sup> It has since attempted to back off from that somewhat, but has been thwarted (so far) by litigation.

### Zoning of Cable

Cable operators may not operate without a franchise.<sup>27</sup> For many years, the FCC sharply curtailed cable's right to bid for programming, accept advertising, and generally compete unhindered against broadcast television. To this day, cable is required to devote one-third of its channels to carry local TV stations<sup>28</sup> and is required to set aside additional channels for lease and "public access."<sup>29</sup> Like the rules that zone telcos out of video, some of the cable-zoning rules are also under First Amendment attack. For now, however, they remain in place.

### COSTS OF FEDERAL ZONING

Calculating the welfare consequences of zoning restrictions is an inherently speculative task. The most potent criticism of such calculations is that they generally entail a partial equilibrium analysis. The introduction of new products and services and the infusion of greater competition in existing markets will necessarily affect other sectors of the economy. To be theoretically sound, one ought to consider such effects. For example, the introduction of electronic voice mail may adversely affect the answering machine market. Competition-induced reduction in the price of long-distance calls may reduce the demand for overland mail services. To remain tractable, empirical welfare calculations abstract from such considerations. If and when such effects are large and observable, they ought to be included. Otherwise, the welfare calculations must be interpreted cautiously.

The social costs of zoning restrictions can be broken down into four categories:

1. Use of spectrum by low-value users as dictated by zoning artificially reduces social welfare by not allowing higher-value users to employ the scarce resource. Analogous costs do not result from restrictions on the use of wireline, which faces no capacity constraint.
2. Zoning restrictions reduce the extent of competition in existing markets and thus reduce consumer welfare. Spectrum zoning, for example, makes it extremely difficult for a third cellular provider to enter most markets. The prohibition of video delivery by Bell operating companies stunts competition in cable television markets. Even if cost penalties due to scale economies are present in these markets, greater competition would likely improve social welfare (the sum of consumer and producer surpluses).
3. Zoning that prohibits or delays the provision of innovative services reduces the welfare of consumers willing to purchase these services.
4. Zoning creates economic rents pursued by the providers of telecommunications services through socially wasteful rent-seeking activities and reduces government revenues that could be generated through auctions by effectively excluding bids by higher-value users of the spectrum.

Although these various costs are analytically separate, they are not mutually exclusive. For example, a third cellular provider that displaces a current low-value-use occupant of the spectrum could spur competition in the cellular market. Similarly, an innovative video delivery system could generate more competitive pricing in the cable market.

#### Lower-Value Use of Spectrum

In 1992, the Office of Plans and Policy of the FCC completed a study analyzing the welfare implications of reassigning spectrum currently earmarked for use by a UHF channel in the Los Angeles area for purposes of providing cellular phone services.<sup>30</sup> The calculations are based on alternative uses of spectrum during the years 1992 to 2000; numbers are expressed in 1991 dollars. Under plausible assumptions, the study estimates the welfare loss attributable to the loss of a single UHF channel in the Los Angeles area to be approximately \$139 million for the years 1992 to 2000. This figure is the sum of welfare foregone by consumers and the producer of UHF television services. The study also estimates that the welfare gains in the absence of any competitive effects (no price or output effects) of two incumbent cellular providers would be approximately \$118 million. Reassignment of the spectrum under such a scenario would, of course, not be welfare-enhancing.

Consider, however, the strategy of Nextel (formerly "Fleet Call"). The company purchased spectrum previously licensed for other purposes—mostly dispatching taxis. In 1991, Nextel persuaded the FCC to permit it to use that spectrum to provide digital wireless telephone service.<sup>31</sup> Nextel has aggressively developed new digital radio services, and is now positioned to be the third major wireless operator in many urban markets. It has already launched its digital mobile network service in many regions of California, including Los Angeles and San Francisco; in 1995 it will roll out service in New York, Chicago, Dallas, and Houston.<sup>32</sup> With its acquisition of several regional SMR companies and Motorola SMR frequencies across the country, Nextel will have the spectrum to serve all 50 of the largest U.S. markets.<sup>33</sup> In the Los Angeles market, it purchased approximately 9 MHz of spectrum for roughly \$45 million.<sup>34</sup>

Assume that the two cellular incumbents in the Los Angeles market rather than Nextel had purchased equal shares of the 9 MHz of spectrum. If this spectrum had been evenly divided between the two existing cellular providers, the increase in social welfare (inclusive of the purchase price) would have been in the range of \$37 million and \$73 million for the years 1992 to 2000. To the extent that radio dispatching services were continued, one need not consider the foregone welfare of those customers. If spectrum zoning had prevented the use of the previously assigned frequencies for cellular service, it would have reduced social welfare by roughly \$50 million for 9 years alone. Zoning of the spectrum may prevent socially efficient transfers of a scarce resource from low-value to high-value users even in the absence of any competitive effects.

### Zoning-induced Reduction in Competition

The case for allowing transfers from low-value users to high-value ones becomes even more compelling if such transfers have the effect of spurring competition in existing markets. Assuming that Nextel faced no competitive disadvantage vis-à-vis incumbent cellular providers, the presence of a third competitor would likely increase competition and drive down prices. Assuming unitary elasticity of demand for cellular services, the authors of the FCC study estimate that a reduction in cellular prices of 25 percent would lead to an increase in total welfare in the Los Angeles cellular market of \$799 million for 6 MHz of additional spectrum and \$893 million for 12 MHz of spectrum. If such a price reduction were indeed to occur because of competition from a third supplier, net social welfare would increase by approximately \$750 million to \$850 million.

The Nextel story would be replayed hundreds of times if spectrum were de-zoned across the board. The FCC study that forms the basis of the Nextel calculations also found that if a UHF television station in Los Angeles were to shut down and transfer its spectrum to a third cellular provider, the overall public gain would be substantial. The FCC study explicitly took into account the possibility of loss in scale economies in their estimates. Even with duplication of vital inputs by the new entrant, the net welfare gain is substantial: \$63 million for a 5 percent reduction in cellular prices and \$783 million for a 25 percent decline. Comparable gains are almost certainly possible by de-zoning spectrum licenses across the board. In other words, the simple deletion of a few lines of legal boilerplate from FCC spectrum licenses could create a substantial increase in social welfare nationwide.<sup>35</sup> The case for transfers of rights in spectrum is extremely compelling when such transfers are likely to boost competition in existing markets.

Zoning of wireline media, including the cable-telephone cross-ownership rules and the inter-LATA restrictions placed on the Bell companies, has also imposed welfare losses. Cable and telephone companies could make deep inroads into each other's markets if freed to use their wires to compete. Cable operators estimate that their telephone service could achieve a 30 percent penetration rate of their basic cable subscribers within 5 to 7 years. Telephone companies estimate they could acquire 45 percent of cable subscribers.<sup>36</sup>

Restrictions that ban Bell companies from providing video program services have enabled cable television companies to maintain a monopoly position in most markets. In the absence of such zoning, the Bell companies could have provided viable competition to the cable television companies to the benefit of customers. In the few markets where duopolistic competition exists in the provision of video programming, cable television rates have been estimated to be 20 to 30 percent lower than in monopolistic markets.<sup>37</sup>

Assuming demand elasticities of -1.0 to -2.5 and price reductions of 15 to 25 percent, consumer welfare in the years 1983 to 1993 would have been higher by anywhere from \$1.4 billion to \$2.9 billion (1993 dollars) annually in the absence of zoning restrictions.<sup>38</sup> Using average revenue per subscriber yields even higher annual welfare loss estimates ranging from \$2.7 billion to \$5.4 billion. Our calculations are summarized in Table 1.

TABLE 1 Consumer Welfare Loss Attributable to Cable Monopolies, 1983 to 1993

	Dp=-15%	Dp=-20%	Dp=-25%
Basic Rate Calculation (1993 \$billions)			
e=-1.0	15.8	21.6	27.6
e=-1.5	16.4	22.6	29.1
e=-2.0	16.9	23.6	30.7
e=-2.5	17.5	24.5	32.2
Average Revenue Calculation (1993 \$billions)			
e=-1.0	29.3	40.0	51.2
e=-1.5	30.4	41.8	54.0
e=-2.0	31.4	43.7	56.8
e=-2.5	32.4	45.5	59.7

Dp, decline in prices; e, elasticity.

Assumptions most consistent with observed facts (20 percent decline in basic rates and demand elasticity of -2.0)<sup>39</sup> yield foregone consumer welfare of approximately \$2.1 billion annually (basic rates) for the years

1983 to 1993. These savings account for approximately one-tenth of the total revenues in the cable television market in 1993. Even if overbuilding were to lead to nonsignificant cost penalties, it is unlikely that total producer surplus would decrease by enough to swamp the much larger savings to consumers. The true measure of foregone consumer welfare should also take into account the provision of more channels (greater variety of programming) in competitive markets relative to monopolistic ones. Consumers lost out during every year that the Bell companies were prohibited from providing video programming services. Had phone companies been authorized to provide such services two decades ago, they would almost certainly be doing so by now absent the no-video zoning of their networks.

The costs imposed by the cable/telco cross-ownership restriction can be independently estimated using an event study. On the day Judge Ellis of the U.S. District Court in Alexandria, Virginia, declared the telco/cable cross-ownership restriction unconstitutional, the stock price of the plaintiff, Bell Atlantic, rose 4.6 percent above the previous day's close.<sup>40</sup> On the day of the court decision, Reuters released a story about the decision in which it stated that "shares of Bell Atlantic and other Baby Bells soared."<sup>41</sup> On the day after the court decision, a Dow Jones News Wire headline announced, "Cable, Telecom Equip[ment] Stocks Soar on Bell Atlantic Ruling."<sup>42</sup> An event study of the effect of the ruling on the value of Bell Atlantic stock suggests that the market believes Bell companies to be viable competitors in the provision of services provided by cable companies.

Assuming that the judge's decision was unanticipated (a reasonable assumption considering that no other phone company had prevailed in any remotely comparable First Amendment suit before), that the market absorbed the implications of the ruling in a single day, and that no other significant value-affecting information was revealed that day, the excess return attributable to the ruling was roughly 4 percent and accounted for roughly \$856 million of the day's increase in Bell Atlantic's value.<sup>43</sup> Despite the imperfection of this estimate, the rise in the company's value suggests two facts. First, Bell companies can be profitable entrants in the cable market. Second, the restrictions on cross-ownership unduly suppressed viable and potentially welfare-enhancing competition.

Cellular zoning also imposes great welfare losses as well. Richard Higgins and James Miller have calculated that if the MFJ restrictions were removed, Bell cellular customers alone would potentially realize annual savings of \$200 million in long-distance charges.<sup>44</sup> They arrive at this estimate by comparing the prices of toll long-distance and cluster services. They demonstrate that retail prices for toll long-distance services exceed bulk wholesale prices by at least 110 percent.<sup>45</sup>

The Wharton Econometric Forecasting Associates (WEFA) Group has independently estimated that if MFJ restrictions were eliminated, over the next 10 years cellular consumers would save \$107 billion.<sup>46</sup> WEFA estimates that monthly wireless bills would fall 30 percent within 10 years.<sup>47</sup> They base the estimate on the following three factors. First, Bell cellular customers currently pay about 40 percent more for long-distance calls than customers of independent cellular companies.<sup>48</sup> Second, in metropolitan service areas (MSAs) where neither cellular provider is a Bell company subject to inter-LATA restrictions, local cellular service is 7 percent cheaper (because of unrestricted competition) than in MSAs where there is a Bell cellular provider.<sup>49</sup> Third, Bell cellular providers are unable to cluster service areas and offer large local calling areas as can independents.<sup>50</sup> For example, an 84-mile, 10-minute cellular call from Indianapolis to Terre Haute is \$2.10 cheaper using GTE than using BellSouth.<sup>51</sup>

One further example of the costs of cellular zoning involves AirTouch. In 1994, Pacific Telesis spun off its wireless affiliate into a wholly separate entity, renamed AirTouch. A primary reason for the spin-off was to free AirTouch from the MFJ's restrictions.

When Pacific Telesis announced it was considering the move, investment analysts predicted that spin would increase shareholder value. Before the spin, one analyst valued PacTel's cellular business at \$140 per point of presence (POP),<sup>52</sup> while comparable independent cellular companies traded at an asset valuation of \$160 per POP.<sup>53</sup> Using the difference (\$20) between Morgan Stanley's \$140 valuation of PacTel's cellular interests and its valuation of independent cellular companies at \$160 per POP, the domestic cellular operation as an independent company is worth \$700 million ( $\$20 \times 34,893,721$  POPs<sup>54</sup>) more than the cellular operation as a part of PacTel. Assuming that half of the valuation increase is the result of AirTouch's superior financial standing relative to the other independent cellular providers, the company is worth \$350 million (6.25 percent) more when freed from the MFJ.



The MFJ's ban on Bell provision of interexchange services has further facilitated tacit collusion among the three major suppliers of such services—AT&T, MCI, and Sprint. Evidence of high price-cost margins in various interexchange services supports the hypothesis of tacit collusion among these firms.<sup>55</sup> The FCC's tariff process has also facilitated such collusion by allowing competitors to monitor and to match each other's pricing strategies, thereby reducing the value of deviating from a collusive outcome.

Removing the restrictions on Bell provision of long-distance services—even just to permit Bell entry out of region—would eliminate several of the conditions facilitating such tacit collusion. First, the number of viable competitors would rise from three to nine. Second, the current stability of market shares among the three dominant suppliers would be destroyed; Bell companies would likely pursue new customers through aggressive pricing strategies. Finally, the regulation-imposed barrier to entry in long-distance services would simply be eliminated.

The welfare costs of the inter-LATA restrictions imposed on Bell companies are substantial. Such entry would stimulate more competitive pricing of long-distance services, driving down the price-cost margins. The FTC estimates that if Bell entry caused prices to fall to marginal cost, it could lead to the elimination of deadweight loss.<sup>56</sup> The welfare gain would then be between \$17 million and \$119 million per year (0.03 percent and 0.36 percent of industry revenue). Consumer welfare gains would be many times more this reduction in deadweight loss. Furthermore, Bell companies might enjoy cost advantages for some routes because of economies of scope with their existing local networks.

### Restrictions and Delays in the Provision of New Services

Potentially the greatest cost imposed by zoning restrictions is the opportunity cost of the inability to introduce unhampered new services for which consumers are willing to pay. That cost is also the most difficult to quantify with precision. Nevertheless, it is useful to highlight those services and provide rough measures of foregone consumer surplus. One has to estimate not only demand for a service that has never existed but also the "virtual" or "reservation" price at which demand is choked before proceeding with consumer welfare calculations.

Relying on survey data on demand for advanced services that could be offered by Bell companies, WEFA estimated the loss in consumer welfare attributable to zoning restrictions for a number of services.<sup>57</sup> We summarize below their estimates of foregone yearly consumer welfare for services that Bell companies can but may not provide under MFJ restrictions:

- Residential Customers
  1. Advanced portable phone—\$1.6 billion
  2. Return voice mail—\$1.7 billion
- Small and Medium Size Businesses
  1. Fax overflow service—\$1.4 billion
  2. Return voice mail service—\$720 million
  3. Call manager service—\$320 million

If one adds producer surplus from the provision of these services, the total welfare gain will be substantially larger. Furthermore, the above list is only illustrative, not exhaustive.

As described above, one large obstacle to the provision of new services has been the FCC's policy against allowing "enhanced services" into the telephone network. Today, some 1.3 million customers buy voice mail service from Bell Atlantic. The service was first offered in 1988, when both Judge Greene and the FCC finally agreed to dezone telephone company wires to permit them to provide such services. At roughly \$5 a month per mailbox, and 10 million mailboxes nationwide, it is reasonable to estimate that in this single market, excluding Bell Atlantic from the market needlessly suppressed over \$460 million dollars of service a year to willing consumers in Bell Atlantic's area alone. Unable to buy the on-line service from U.S. enterprises and U.S. workers, most consumers probably turned to stand-alone answering machines manufactured in Singapore or

Taiwan. Extrapolating this figure nationwide, the no-voice-mail zoning (part of the much broader no-computer-services zoning) of phone companies cost the U.S. economy some \$600 million a year for at least a decade.

Even if the restrictions are eventually lifted, the delay in the provision of socially desirable services imposes enormous costs to the economy.<sup>58</sup> In 1985, an FCC OPP working paper estimated the costs of a comparative hearing for a cellular license application, breaking it down into cost of the delay in awarding the license, cost to the government, and cost to applicants.<sup>59</sup> The paper found that the typical 18-month delay in awarding cellular licenses eroded the value of the license by \$90,000 and cost the government \$20,000 per hearing and each applicant \$130,000 per hearing.<sup>60</sup>

Finally, the possibly anticompetitive effects of lifting such restrictions must be weighed against the foregone welfare of consumers willing to purchase and of producers willing to provide such services. The net welfare effect may indeed be positive in many or most instances where zoning is imposed because of alleged anticompetitive effects. Professor Paul Rubin, for example, has analyzed the costs and benefits of the MFJ waiver process.<sup>61</sup> Rubin estimated the cost of the waiver process—the delay of waivers that would have had procompetitive consequences, administrative burdens and rent-seeking, and the deterrence of procompetitive activities due to the waiver process—at over \$1.5 billion since 1984.<sup>62</sup>

### Rent-Seeking Activities

In addition to the opportunity costs described above, a number of other social costs can be identified. Zoning has led to the existence of substantial economic rent in a number of telecommunications markets. Evidence of such rent includes estimates of Tobin's Q-ratio for a number of markets as well as the observed decline in stock value of incumbent firms in response to announcements of regulatory measures that enhance competition in their markets. To the extent that much of the economic rent results from government protection from competition, one could expect substantial resources devoted to gaining such protection. Although the expenses directed at socially wasteful rent-seeking activities are difficult to quantify precisely, they are certainly in the hundreds of million of dollars.<sup>63</sup>

## MODELS OF DEZONING

The costs of telecommunications regulation are not immutable. Limited dezoning has already begun. Recent direct broadcast satellite (DBS) and broadcast regulation are small examples of what can be applied. Moreover, a property-based system of spectrum allocation could replace the current system of government ownership.

### DBS Dezoned

Regulation of the relatively new DBS services broke new ground—it severely blurred the formerly clean lines between private carriers, common carriers, and broadcasters.<sup>64</sup> The owner of a DBS satellite can lease under contract, or sell outright, transmission space to private users. It may operate others on a common carrier basis for independent programmers. And it may send its own programming over others, either scrambled and paid for by subscribers, or "in the clear" and paid for by advertisers.<sup>65</sup> The owner is thus free to use its satellite, and the spectrum the satellite uses, for any mix of carriage, broadcast, or none-of-the-above activities like subscription services or purely private transport. It may change the mix as it pleases.

This is perfectly sensible; it is also a radical departure from 50 years of FCC regulation under the 1934 act. The satellite broadcaster is the first spectrum licensee that has been told, in effect, to use spectrum for any useful purpose it can find. Spectrum licenses have been issued without cumbersome "zoning" restrictions for the first time. No other spectrum licensee has been this free since the enactment of the Radio Act of 1927.

### Broadcast Dezoned

Conventional broadcasters have crept forward in the deregulatory wake of other wireless operators, one small step at a time. They have begun, for example, overlaying some modest carrier-like services on top of their basic broadcast operations. Radio stations, for example, transmit paging services using the "subcarrier"<sup>66</sup> portions of their assigned frequencies.<sup>67</sup> Television broadcasters broadcast video programming, but also broadcast electronic newspapers, data, computer software, and paging services within the otherwise unused "vertical blanking interval" of their spectrum licenses.<sup>68</sup>

### Market-based Allocation of Spectrum

The most significant alternative to the current zoning approach to spectrum allocation is a market-based one. The idea is not new. As far back as 1959, Nobel economist Ronald Coase proposed that property rights be accorded to radio licenses, that license owners pay for the right to use spectrum, and that bidding for that right was the most efficient way to allocate spectrum.<sup>69</sup> In contrast to spectrum, the lack of any substantial capacity constraints in the wireline medium suggests that the costs of zoning can be alleviated by simply lifting most of the restrictions on the medium's use.

In 1991, the National Telecommunications and Information Administration (NTIA) studied the idea, reviewed the considerable literature and past proposals on the subject, and concluded that a "market-based system for spectrum management would improve considerably the efficiency and fairness of the current U.S. system, and, if properly designed and implemented, can fully address concerns about such issues as competitive equity, spectrum 'warehousing,' and the preservation of socially desirable services."<sup>70</sup> As part of its recommendation, NTIA supported private transfers and subleasing of spectrum rights directly from one user to another.<sup>71</sup>

Australia and New Zealand are ahead of the United States in selling spectrum. The Australian Federal Spectrum Management Agency recently auctioned 196 wireless cable licenses.<sup>72</sup> The government will also receive yearly license fees on the auctioned spectrum.<sup>73</sup> The Australian Broadcasting Authority plans to auction new television and radio licenses.<sup>74</sup> New Zealand has auctioned off 20-year rights to new radio spectrum and receives additional revenues from user fees.<sup>75</sup> It has also auctioned off spectrum for cellular service<sup>76</sup> and broadcast television.<sup>77</sup>

### CONCLUSION

The social cost of zoning the electromagnetic spectrum and wireline media is extremely high. While it is impossible to quantify exactly the total social cost of such zoning, the welfare effects of specific restrictions suggest that zoning imposes large opportunity costs on society. Zoning misallocates resources; it reduces competition; and it delays or prevents the provision of desired services. These costs are rarely considered carefully when restrictions are imposed.

### NOTES

1. A study by the Wharton Econometric Forecasting Associates (WEFA) Group attempts to quantify the total costs of all legal and regulatory barriers in the telecommunications, information services, equipment manufacturing, and video programming markets. The study compares a baseline economic forecast with a forecast assuming that all legal and regulatory barriers to competition are removed and that rate-of-return regulation is replaced by pure price cap regulation in all jurisdictions. The study concludes that "competition and the expected lower prices that competition will bring result in nearly \$550 billion in consumer savings cumulatively over the next ten years." In its analysis, WEFA first develops pricing models for long-distance, local, cellular, and cable service, and then estimates the impact that competitive entry will have on prices for long-distance, local, cellular, and cable television service. The study predicts that prices for these services will fall dramatically over the study period (1995 to 2005), with both sharp one-time price adjustments (to reduce prices to competitive levels) and steadily decreasing prices over time due to technological efficiencies. The main impetus

for these price changes, WEFA asserts, will be entry of the Bell operating companies into the various markets, after the lifting of the video programming ban and the MFJ restrictions on inter-LATA service. The WEFA Group, *Economic Impact of Deregulating U.S. Communications Industries*, February 1995 (hereinafter 1995 WEFA Study).

2. For further discussion on the federal regulation of telecommunications, see generally Kellogg, Michael, John Thorne, and Peter Huber. 1992. *Federal Telecommunications Law*, Little, Brown & Company; Thorne, John, Peter Huber, and Michael Kellogg. 1995. *Federal Broadband Law*, Little, Brown & Company, New York.

3. Stat. 302 (Comp. St. §10100–10109) (1912).

4. See Hazlett, T. 1990. "The Rationality of U.S. Regulation of the Broadcast Spectrum," 33 *J.L. & Econ.* 133.

5. See, for example, *Tribune Co. v. Oak Leaves* (Cir. Ct., Cook County, Ill. 1926), reprinted in 68 *Cong. Rec.* 216 (1926).

6. Radio Act of 1927, 44 Stat. 1162 (1927).

7. See Emord, Jonathan W. 1992. "The First Amendment Invalidity of FCC Content Regulations," *Notre Dame Journal of Law, Ethics, and Public Policy* 93, 185.

8. 47 U.S.C. §307(b).

9. Radio Act of 1927 §1. Cf. 47 U.S.C. §301; 47 U.S.C. §304; 47 U.S.C. §309(h)(1).

10. Radio Act of 1927 §9.

11. See 47 U.S.C. §307(c).

12. 47 U.S.C. §312.

13. 47 U.S.C. §312 (a).

14. An Inquiry Relative to the Future Use of the Frequency Band 806–960 MHz, 19 Rad. Reg. 2d (P & F) 1663, 1676–1677 (1970).

15. An Inquiry into the Use of Bands 825–845 MHz & 870–890 MHz for Cellular Communications Systems, 86 F.C.C.2d 469 (1981).

16. 47 U.S.C. §214(a).

17. 47 U.S.C. §201.

18. See 78 *Cong. Rec.* 10314 (1934).

19. See Robinson, Glen O. 1989. "The Federal Communications Act: An Essay on Origins and Regulatory Purpose," in *A Legislative History of the Communications Act of 1934*, Max Paglin (ed.), p. 40.

20. *General Telephone Co. v. FCC* 13 F.C.C.2d 448 (1968), aff'd sub nom. *General Telephone Co. v. FCC*, 413 F.2d 390 (D.C. Cir.), cert. denied, 396 U.S. 888 (1969).

21. Applications of Telephone Companies for Section 214 Certificates for Channel Facilities Furnished to Affiliated Community Antenna Television Systems, 21 FCC.2d 307, 325 (1970).

22. 47 U.S.C. §533(b).

23. Id.; see also, 47 C.F.R. §63.54(a).

24. See Telephone Company-Cable Television Cross-Ownership Rules, 7 FCC Rcd 5781, 5817–18 (1991); see also H.R. Rep. No. 934, 98th Cong., 2d Sess. 57 (1984).

25. Regulatory Pricing Problems Presented by the Interdependence of Computer and Communication Facilities, Final Decision and Order, 28 F.C.C.2d 267, 269 (1970).

26. Amendment of §64,702 of the Commission's Rules & Regulations, Second Computer Inquiry, 77 F.C.C.2d 512 (1981).

27. 47 U.S.C. §541(b).

28. 47 U.S.C. §534(b)(1)(B).

29. 47 U.S.C. §§531, 532.

30. Kwerel, Evan, and John R. Williams. 1992. "Changing Channels: Voluntary Reallocation of UHF Television Spectrum," Office of Plans and Policy Working Paper No. 27, Federal Communications Commission, Washington, D.C., November.

31. Memorandum Opinion and Order, Request of Fleet Call Inc. for a Waiver and Other Relief to Permit Creation of Enhanced Specialized Mobile Radio System in Six Markets, 6 F.C.C. Rec. 1533 (1991).

32. *Edge*. 1993. "Fleet Call Becomes Nextel; New Company Name Reflects New Business Designed to Serve Broader Wireless Communications Market," March 29.

33. *Business Wire*. 1994. "Nextel Reaches Agreement with U.S. Department of Justice," October 27.

34. *Land Mobile Radio News*. 1994. "Motorola Gains Long-awaited Foothold in Los Angeles SMR Market," November 13, at Section No. 46.

35. As Janice Obuchowski, former Department of Commerce assistant secretary for communications and information and former NTIA administrator, has said, "Efficient use of the spectrum will be maximized only if licensees are given the widest possible latitude in determining which services to offer within their assigned frequencies. In

principle, the flexibility granted to licensees should be limited only to the extent necessary to prevent radio frequency signal interference with other users." Obuchowski, Janice. 1994. "The Unfinished Task of Spectrum Policy Reform," 39 *Fed. Com. L. J.* 325, pp. 328–329.

36. Yokell, Larry. 1994. "Cable TV Moves into Telecom Markets," *Business Communications Review*, November, p. 2.

37. Affidavit of Thomas W. Hazlett, attached to Motion of Bell Atlantic Corporation, BellSouth Corporation, NYNEX Corporation, and Southwestern Bell Corporation to Vacate the Decree, No. 82-0192 (D.D.C. July 6, 1994).

38. Following is a brief summary of the data and methodology we employed to make the consumer welfare calculations. (1) We performed two sets of calculations—one based on subscribers and prices of basic service and another based on subscribers and average revenue per subscriber. (2) Basic rate calculations were based on data found in the appendix to Hazlett (July 3, 1994). Calculations were for the years 1983 (the year after the MFJ) through 1993. We assumed that 98 percent of all basic service subscribers were in monopoly markets. This assumption was maintained for all the years under consideration, while in reality competition did not exist every year in markets that were competitive in 1992. This assumption therefore is likely to provide a downward bias in the estimates. (3) Average-revenue-per-subscriber calculations were also based on data found in Hazlett. We assumed that the total number of subscribers was the same as the number of basic service subscribers (since premium service subscribers are also basic service subscribers). We further assumed that 98 percent of total subscribers were in monopoly markets. Again, downward bias should be expected. (4) We used average basic service prices and average revenue per subscriber as the monopoly price. The correct numbers to be used are averages in monopoly markets only. (5) We assumed that income effects due to decline in cable prices are negligible. The assumption is generally valid if the share of cable television expenses is small relative to total income. (6) We assumed demand was locally linear. (7) We converted all calculations to constant 1993 dollars using the Consumer Price Index for all goods with 1993 equal to one. (8) We performed the welfare calculations assuming various elasticities of demand and expected decline in prices due to duopoly competition. We considered own-price elasticities of -0.5, -1.0, -1.5, -2.0, and -2.5 and price declines of 15, 20, and 25 percent. To the extent that a monopolist always operates in the region where elasticity of demand exceeds one, the first two elasticities are valid only if monopoly pricing behavior is somehow constrained.

39. Crandall, Robert. 1990. "Elasticity of Demand for Cable Service and the Effect of Broadcast Signals on Cable Prices," Attachment TCI Comments to the Federal Communications Commission, MM Docket No. 90-4; Levin, Stanford L., and John B. Meisel. 1990. "Cable Television and Telecommunications: Theory, Evidence and Policy," *Telecommunications Policy*, December; Emmons, Willis, and Robin Prager. 1993. "The Effects of Market Structure and Ownership on Prices and Service Offerings in the U.S. Cable Television Industry," paper presented to the Western Economics Association 68th Annual Conference (22 June); Federal Communications Commission, FCC Cable Rate Survey Database (Feb. 24, 1993).

40. "Tradeline Database," Dow Jones News Retrieval, April 21, 1995.

41. Talking Point/Bell Atlantic Court Decision, Aug. 24, 1993.

42. *Dow Jones News Wire*. 1993. "Cable, Telecom Equip[ment] Stocks Soar on Bell Atlantic Ruling," August 25.

43. This study assumed Bell Atlantic's Beta to be 0.78 and that the relevant market was defined by the S&P 500 basket of stocks. It used outstanding shares and prices at the end of calendar year 1992 to calculate the benchmark valuation. A one-day window using the previous day's valuation as the benchmark yielded an increase in Bell Atlantic value attributable to the ruling of roughly \$950 million. Two-day and three-day windows yielded a cumulative excess return of 7 percent and increased valuation of roughly \$1.5 billion. A longer window would yield an even larger estimate of increased return and valuation.

44. Affidavit of Richard S. Higgins and James C. Miller III, attached to Motion of Bell Atlantic Corporation, BellSouth Corporation, NYNEX Corporation, and Southwestern Bell Corporation to Vacate the Decree, *United States v. Western Elec. Co.*, No. 82-0192 (D.D.C. July 6, 1994).

45. Because of the MFJ's interexchange and equal access restrictions, the BOCs cannot obtain bulk wholesale rates and therefore cannot offer these savings to their customers.

46. 1995 WEFA Study, *supra* n. 1, at p. 4.

47. *Id.*, p. 30.

48. *Id.*, pp. 30–32.

49. *Id.*, p. 32.

50. *Id.*, pp. 32–33.

51. *Id.*, p. 34.

52. A POP is derived by multiplying the total population of a service area by an operator's interest in the license.

53. Morgan Stanley & Co., Pacific Telesis Group—Company Report, Report No. 1219882, Apr. 21, 1992.

54. Donaldson, Lufkin & Jenrette. 1994. "The Wireless Communications Industry," Summer, p. 68.
55. MacAvoy, Paul W. 1994. "Tacit Collusion by Regulation: Pricing of Interstate Long-Distance Telephone Service," Working Paper #37, Yale School of Organization and Management, August, pp. 37–47.
56. Ward, Michael R. 1995. "Measurements of Market Power in Long Distance Telecommunications," Bureau of Economics Staff Report, Federal Trade Commission, Washington, D.C., April.
57. 1995 WEFA Study, supra n. 1.
58. Milton Mueller has explained that "as a barrier to entry, getting an allocation through the administrative process can be far more formidable than paying for access to spectrum." Mueller, Milton. 1988. "Technical Standards: The Market and Radio Frequency Allocation," *Telecommunications Policy*, March, p. 51.
59. Kwerel, E., and A.D. Felker. 1985. "Using Auctions to Select FCC Licensees," FCC OPP Working Paper, May, pp. 12 and 17.
60. Id.
61. Affidavit of Paul H. Rubin (June 14, 1994), attached to Motion of Bell Atlantic Corporation, BellSouth Corporation, NYNEX Corporation, and Southwestern Bell Corporation to Vacate the Decree, *United States v. Western Elec. Co.*, No. 82-0192 (D.D.C. July 6, 1994).
62. Rubin averaged data provided in several waiver requests on the fixed and annual costs of delaying approval of the waiver, and extrapolated these costs over all waivers. Because over 96 percent of waiver requests are approved, and thus presumed to be procompetitive, and because individual anticompetitive requests are less harmful than individual procompetitive waivers are beneficial, the waiver process results in significant wasted resources, time, and money. The costs due to rent-seeking and deterrence of procompetitive activities are specified in less exact terms, but Rubin's reasonable estimate places them in the hundreds of millions of dollars.
63. Id.
64. Report and Order, Inquiry into the Development of Regulatory Policy in Regard to Direct Broadcast Satellites for the Period Following the 1983 Regional Administrative Radio Conference, 90 F.C.C.2d 676 (1982), recon. denied, Memorandum Opinion and Order, Regulatory Policy Regarding the Direct Broadcast Satellite Service, 94 F.C.C.2d 741 (1983), aff'd in part sub nom. *National Association of Broadcasters v. FCC*, 740 F.2d 1190 (D.C. Cir. 1984).
65. See *World Communications v. FCC*, 735 F.2d 1465 (D.C.Cir. 1984).
66. Spectrum can be subdivided into a main channel and a number of "subchannels" or "subcarriers," both of which can be transmitted simultaneously.
67. First Report and Order, Dkt. No. 82-536, 48 *Fed. Reg.* 28445 (1983); Second Report and Order, Dkt. No. 21323, 49 *Fed. Reg.* 18100 (1984); Amendment of Parts 2 and 73 of the Commission's AM Broadcast Rules Concerning the Use of the AM Sub-carrier, 100 F.C.C.2d 5 (1984).
68. Amendment of Parts 2, 73, and 76 of the Commission's Rules to Authorize the Offering of Data Transmission Services on the Vertical Blanking Interval by TV Stations, 101 F.C.C.2d 973 (1985).
69. Coase, R.H. 1959. "The Federal Communications Commission," 2 *J.L. & Econ.* 1, pp. 14 and 25.
70. National Telecommunications and Information Administration. 1991. *U.S. Spectrum Management Policy: Agenda for the Future*, February, p. 7. Henry Geller, former NTIA administrator and former general counsel for the FCC, co-authored a report that explained that "charging for the spectrum is particularly appropriate now in light of the weakness of traditional broadcast public trustee regulation and the growing demands on the spectrum overall." Geller, Henry, and Donna Lampert. 1989. "Charging for Spectrum Use," Benton Foundation Project on Communications and Information Policy Options, p. ii.
71. National Telecommunications and Information Administration. 1992. *U.S. Spectrum Management Policy: Agenda for the Future*, February, p. 8.
72. *Market Reports*. 1994. "Australia—Pay TV Overview," September, p. 19.
73. Ellis, Stephen. 1994. "Australia: MDS Pay-TV Auctions Net \$90.6M," *Australian Financial Review*, August 19.
74. Davies, Anne. 1995. "Australia: Facts Lay Groundwork for ABA Legal Challenge," *Sydney Morning Herald*, January 27.
75. *New Zealand Herald*. 1991. "New Zealand: Seven Radio Frequencies for Auckland," January 12.
76. *New Zealand Herald*. 1990. "New Zealand: Telecom Free to Bid for Mobile Phones," May 17.
77. *New Zealand Herald*. 1990. "New Zealand: Space on Airwaves Costs Sky \$2.2M," March 28.

## 38

# Residential PC Access: Issues with Bandwidth Availability

*Kevin C. Kahn, Intel Corporation*

### ABSTRACT

The preeminent interactive information access device in the business world today is clearly the personal computer. Via the PC, individuals access the exploding array of information sources both within their businesses and throughout the Internet. PCs are rapidly penetrating the consumer environment as well, and we strongly believe that the PC will become the center for national information infrastructure (NII) access for the residential consumer. The intelligence encoded in PC applications and their improving human interfaces will make PCs the tool of choice for ubiquitous consumer information access. However, to achieve their great potential in this role, residential consumer PCs must have access to adequate communication bandwidth. This bandwidth must be priced to be attractive to the residential consumer. The consumer must be free to access all information services using this bandwidth.

High-bandwidth access to information services must get the same level of attention with respect to public policy as providing competitive television and telephone delivery systems. This attention is needed to ensure that the bandwidth, access, and consumer choices are made available in ways that promote the growth of consumer NII use.

This paper develops these themes by examining the forces that are driving corporate PC network access to the NII and the various network topologies being discussed for the residential environment (HFC, FTTC, etc.), as well as possible future service directions. We then develop what we feel are the critical requirements for bandwidth availability for residential PCs.

### STATEMENT OF PROBLEM

The personal computer has become a ubiquitous networked communications platform within the business environment. In addition to traditional communications applications such as electronic mail, it is now becoming the base for all sorts of information access and personal communications tools. These tools enable a new level of business activity that includes everything from World Wide Web access for general information acquisition to various levels of electronic commerce and personal conferencing. Key to this development of widespread, cost-effective information and communications applications have been a number of important technologies. First among these has been the deployment of high-bandwidth network connections in both the local and wide areas, utilizing high-volume components, based upon open standards. Free and open access to this bandwidth has permitted any software or hardware developer or information service provider to easily enter these businesses. For example, there are numerous suppliers of network connection hardware and software, a growing number of competing conferencing products, and the beginning deployment of multiple on-line news services from traditional and nontraditional information publishers utilizing the World Wide Web. The resultant competition and interactions are leading to rapid development of the business use of the developing NII.

The same dynamics must be allowed to operate within the critical residential environment to support the development of individual utilization of the NII. While there has been a lot of focus on the larger infrastructure developing to support the NII, particularly in the context of the Internet, there has been less focus on the data access issues of the "last mile" or access network to the residence. What discussion has occurred seems to revolve largely around video-on-demand entertainment services and telephony. We are concerned that while public policy may operate to guarantee competitive entertainment services and telephony over cable or telephony infrastructures, it might not guarantee the deployment of the reasonable levels of openly available, discretionary

bandwidth required to energize an affordable and comprehensive personal computer-based information infrastructure.

The deployment of bandwidth that can be utilized in an open manner to access NII services from the home needs to be encouraged. For data, this access should be via general packet protocols (based on the Internet standards of TCP/IP) that permit users direct interaction with any service they desire. Specifically, data services from the home should come to be viewed like existing telephony services that allow consumers to access any services anywhere in the nation. It may be that for other multimedia services, such access would be better provided via circuit-oriented services such as switched ATM virtual circuits (although promising work is ongoing on utilizing packet protocols for this as well), but in any case the same principles of openness and full connectivity must apply. We also believe that the development of industry-driven standards or implementation agreements for attaching to and utilizing this open bandwidth will be critical to the creation of a competitive environment for consumer information applications. The government role should be to encourage rapid industry convergence on such de facto standards or implementation agreements that can later be the basis of more formal de jure ones. This paper attempts to lay out some requirements for the development of consumer PC information services broadly within the NII.

## BACKGROUND

### The Business Environment as an Illustrative Example

As a point of reference let us first consider the typical business access being deployed to allow PCs to utilize the developing NII. The typical networked PC in the office is attached to a 10-Mbps Ethernet or similar performing Token Ring network. (International Data Corporation estimates that in 1994, 73 percent of the business PCs in the United States were attached to local area networks.) While this bandwidth is shared with the other users of the network in most cases, switched networks are beginning to be deployed that replace the shared access with dedicated access. In addition, technologies such as 100-Mbps Ethernet are appearing that will also greatly improve the bandwidth available to the individual user. Beyond the local environment, most large corporations are deploying reasonable bandwidth into the Internet at large. This combination makes available relatively high-speed access to information and services whether local to the business desktop or remote.

In addition to the bandwidth that is available, another key aspect of the business desktop is the development of standards. Industry-driven network connection standards have driven down the cost of connecting to the network from the PC. For example, Ethernet controller cards for PCs now typically sell for less than \$100, operating systems increasingly come with protocol software built in, and a growing number of applications comprehend networking capabilities. Platform software standards have made it possible for creative developers to build software independent of the nature of the specific network. For example, the Winsock de facto standard for accessing network services on the Microsoft Windows Operating System is allowing application developers to focus their efforts on enhancing function in their products rather than on adapting those applications to a variety of different, incompatible, network services.

Network protocol standards have allowed end-to-end services to operate over a wide variety of network implementations. In particular, general packet networks have allowed a wide variety of data services as well as new applications such as video conferencing to begin to operate without special provisions from the providers of the networks. An entrepreneurial developer need not make deals with a variety of platform and network providers to begin to deploy an application in this environment. Neither is an interested business consumer restricted from beginning to take advantage of such new services by the choices offered him by various network suppliers.

Key aspects of the business environment are that it has been almost entirely industry driven and motivated by competition. De facto standards or implementation agreements have been rapidly developed and only later evolved into de jure standards. Throughout the evolution of the business and the associated standards, the intellectual property of the participants has been respected by the processes. It is interesting to note that the more ponderous de jure first approach to standards represented by the International Telecommunications Union (the official international body for defining standards within the telecommunications industry) has been considerably



less successful in affecting the U.S. environment than have the more chaotic and rapid moving processes of private company consortia or the Internet Engineering Task Force (the de facto group that defines Internet standards).

A couple of other interesting aspects of packet networks are typical in the business data environment. End systems are always available to be connected to—that is, they are always "online." It is thus reasonable to build applications that initiate connection to a user's system from a service at any time. This is different from an environment where, unless the user has initiated a connection to the network, information cannot find its way to the user's end system. Related to this is the fact that packet networks inherently support multiple simultaneous connections. That is, an end system can have an unlimited number of distinct applications running at the same time, each of which is using some of the physical bandwidth to connect to another distinct end system. There is no notion of the line being busy. Again, this flexibility opens up many possibilities for applications that operate in parallel with each other over the same physical connection. It allows service-initiated contact with an end system even while the user is doing other things over his connection.

### **The Current Residential Environment**

We can compare this business situation to that seen in the residential environment, first by looking narrowly at what is really available today, and then more importantly at what is being tested and deployed for the near future. Today's electronic service to the home consists of a number of options.

First among these is the existing plain old telephone service (POTS). This is universally available and provides a completely open but relatively low bandwidth access to data services for the consumer. Using 14.4-kbps and more recently 28.8-kbps modems, the consumer can connect to any information service or network. Traditional choices have been the private information services such as CompuServe, Prodigy, or America Online. However, there are a growing number of general Internet access providers available via this route as well. These provide general packet access into the Internet and thus to any information service in the developing NII.

An improvement over POTS in terms of available bandwidth is an integrated services digital network (ISDN) line, which potentially provides up to 128-kbps access over a single wire. Fewer information service or Internet access providers have thus far deployed ISDN access, and the availability and practicality of utilizing this level of service varies considerably around the country. Like POTS, this is a circuit-oriented service that must be initiated by the consumer. That is, unlike the typical business connection, an information service usually cannot autonomously contact a consumer's PC to provide information or service. Also, while more flexible in some respects than POTS, ISDN is for the most part not useable by multiple separate applications simultaneously, thus further limiting the range of applications that can utilize it. ISDN can be used effectively to connect an end system to a general purpose router that provides a point of presence for the Internet or other general purpose network. Used in this manner, ISDN may provide our best short-term hope for general purpose residential access. Also, since ISDN can provide fast connections to another end system when compared with POTS, it can approximate the speed of being always connected for certain types of applications. While in principle POTS can also be used in this manner, the combination of increased bandwidth and connection speed makes ISDN much more practical in such a configuration.

In contrast to these point-to-point telephony based services, the cable industry has focused on high-bandwidth broadcast services for video. The cable infrastructure is typically highly asymmetric in its bandwidth with much higher speed available into the home than out. Currently, in fact, most cable systems have no return path available. Also, cable is unlike the telephone where a consumer can call anyone or access any service on any other existing telephone system. The services provided today via the cable infrastructure are generally chosen by the cable system operator or according to the "must carry" rules. Cable systems are generally closed systems that encompass everything from the content source to the set-top box, which provides a conventional television signal to the appliance.

There are beginning to be trials of data services over the cable infrastructure. However, in the spirit of the existing industry, these may tend to be for services driven and chosen by the service provider rather than the consumer. In the case of cable TV programming, the limited number of available channels means that the MSO

must select which channels to broadcast. For data services, however, open packet access means that beyond the provision of the data bandwidth, there is no need to limit access to content in this manner. On the positive side, the cable infrastructure brings the promise of relatively near-term residential broadband service. Cable could certainly offer the advantages of high-bandwidth, multisession, always-on data delivery. In particular, by offering high-quality Internet access, cable could provide much of the flexibility we desire. It is important that PC access to broadband information services follow the telephone model rather than the broadcast TV model of access.

## ANALYSIS, DIRECTIONS IN THE RESIDENTIAL ENVIRONMENT, AND FORECASTS

### Access Network Technologies

Two major groups of companies are vying to provide digital services to the residential environment: the existing local exchange carriers and the cable entertainment companies. The former are building on their expertise in operating a highly reliable telephony network by offering higher-bandwidth services through ISDN and through cable overlay systems that are capable of supplying broadband residential service. The latter are leveraging their existing investment in a physical plant oriented to deliver many analog television channels so that they can begin to offer digital entertainment channels and various levels of broadband data communications to the home. As a part of this enhancement of services they frequently wish to offer telephony services via this network as well. This competition should be a positive force for consumers in the long term, provided that it offers real competitive choices to the consumer as well as to service providers.

A number of physical architectures have emerged for deployment by these companies. The most popular of these include the following:

- *Hybrid fiber coax (HFC)*: In this scheme, optical fiber is used to bring signals from a head end to neighborhood nodes. At these nodes, the signals are placed on a coax system for distribution to the residences. The cable industry target for residences served by a single coax segment is 125, although in early deployments the number is larger. Since the coax is a shared medium, the homes on a single coax segment share the available bandwidth. For most current deployments, the amount of bandwidth available from the home (either on the cable or via a separate POTS path) is much smaller than that available to the home. HFC is particularly attractive as an upgrade path from existing analog cable systems.
- *Fiber to the curb (FTTC)*: In this scheme, optical fiber is used to bring signals from a head end to a pedestal on the street that is within a relatively short distance of a collection of served homes. At the pedestal, signals are placed on coax or twisted pair lines to the home. While the optical fiber bandwidth is shared (as in HFC), the drop to the individual home is not.
- *Fiber to the home (FTTH)*: In this scheme, optical fiber is deployed all the way to the home, where it is converted as appropriate to provide digital services. At this point, this scheme is not being pursued to our knowledge in the United States, although it has been proposed at times in other national markets.
- *Asynchronous digital subscriber line (ADSL)*: For areas where the length of the lines from the last point of electronics to the served homes is bounded, ADSL provides a technique for utilizing existing telephony infrastructure to carry higher data rates. It uses more complex signaling electronics over existing copper to provide data rates capable of supplying digital video.

Beyond these wired topologies, experiments are also beginning that use wireless technologies. For example, direct PC utilizing direct broadcast satellite technology provides an asymmetric bandwidth connection similar to using cable combined with POTS for a return path.

### **Bandwidth and Parallelism**

A common theme among all of these approaches is that the allocation of bandwidth into the home is different from the allocation of bandwidth out of the home. From a purely architectural perspective this is unfortunate, since it is an asymmetry that makes an artificial distinction between providers (who need high outbound bandwidth) and consumers (who can get by without high outbound bandwidth). In the full NII where telecommuting will be common, video telephony will be widespread, and a true cottage information industry will become an important part of the national economy, this asymmetry will become a barrier. However, as a practical matter for the short-to mid-term, the key will be reasonably high bandwidth into the home and moderate bandwidth out. This arrangement will cater to information retrieval applications where the bulk of the data flow is inbound with much smaller flows out to make queries. Provided that the outbound bandwidth is high enough, it can also support limited video telephony and telecommuting applications. We and others have demonstrated the practicality of such application at ISDN speeds (128 kbps) via products like ProShare™ videoconferencing and RemoteExpress™ remote LAN access. Nevertheless, higher rates would greatly improve performance as well as support more parallel activities.

For the long term it is critical that residential services support a full range of simultaneous activities within the home. It should be possible for children to be involved in remote education activities while one parent is actively telecommuting, the other is accessing entertainment or shopping services, and power or security monitoring is operating in the background. Solutions that force artificial limits on conducting such parallel activities will make it impractical to depend upon network access as an integral part of the home. Furthermore, each of these individual activities may require simultaneous sessions. A child may be connected both to school and to a library system; a telecommuter may be accessing corporate databases while also checking travel arrangements with an airline. A shopper may be doing comparisons on products offered by competing providers. From a technical perspective, the solution to these multiple access issues is either the provision of multiple virtual circuits or a general packet network interface (or more likely some combination of these technologies).

### **Openness of Access**

A key to allowing these applications, even in the face of the sorts of bandwidth available in the near-term deployments, is to put the use of the bandwidth completely under the control of the consumer. Consider two approaches to providing access to information services. In the first, a cable system contracts to make available one of the online service providers (say, CompuServe) using some of its digital bandwidth to the home. From the consumer's point of view a bundled package is offered that allows direct connection to the service for some fee. The service being provided by the cable operator is CompuServe and not generic online access. If consumers wish to subscribe to a different service, they cannot use the cable bandwidth to do it. Furthermore, if a new provider wishes to enter the business and have good access to customers, it may need to arrange business deals with many different network service providers. The comparable situation in the telephony industry would be that the range of people consumers could call from home would be affected by their choice of local or long distance carrier.

The alternate approach is typified by the existing low-bandwidth telephony access system and the growing Internet service providers. Here, the customer gets access to the general packet switched network and uses that to provide common carriage of data to various information service providers. ISDN can be used in this configuration to provide moderate rate connection to a router from which general Internet connectivity can be provided. Note that the owner of the access network is not involved in the selection of the available services. That selection is between the consumer and the ultimate provider. Today, this access is the norm for telephony-class access. As we move toward higher-speed access networks, it is important that the model exist for broadband as well. Again the comparable situation in the telephone industry is that anyone can establish a business phone number and then advertise directly to potential consumers for business without further involvement of the network providers. Where cable companies choose to provide Internet access on the cable, this fits this model as well.

The argument here is not that cable operators should be precluded from offering packaged services. They should certainly be allowed to offer them as well as local content, provided that they also provide consumer-controlled access to competitive offerings. Openness should exist in both directions. That is, providers should be able to attach to access networks to offer their services to consumers (similar to requirements for allowing video service providers to access "video dialtones"). At the same time, consumers should be able to attach via their access networks to providers who have not chosen to directly connect to that access network.

### **Openness of Content and the Development of New Services**

We do not need to wait for the implementation of all digital networks and the general NII to see examples of why we are concerned with the issue of open bandwidth access. We can find today in analog distribution systems an interesting example of how fragile the relationship between openness of network capabilities, content provider innovation, and application creativity can be. This example appears in the use and distribution of material in the vertical blanking interval (VBI) of broadcast video signals. Note that the issues in this example are more complex than open access to bandwidth since they deal with bandwidth ownership and contracts between MSOs and content providers. However, the example does serve to illustrate how subtle issues in the nature of the distribution networks can affect the creation and deployment of new applications.

Standard U.S. television signals must provide a period of time in each frame to allow the electron beam that paints the screen to move from the end of the screen back to the top in preparation for painting the next set of lines. This period is the VBI and can be seen as a black bar if the vertical hold of a television is misaligned. Since this part of the signal is not used for carrying the picture, it can be used to carry other sorts of data without disturbing the video broadcast. The most common example of this is closed captioning, which is carried in a part of the VBI.

As broadcasters begin to think creatively about their content and work with application developers, other possible uses of the VBI to enhance the broadcast material can emerge. Examples include digital forms of the stock ticker seen on some news networks, digital references to sources of additional information to augment the content of a show, digital indexing information to assist in selective viewing of the material, and undoubtedly many more. However, any such program-enhancing use of the VBI must reach the end consumer to be useful. While the 1992 Communications Act required that program-related material in the VBI should be carried by local cable operators, exactly what constitutes program related is debatable. As a result, there is no guarantee that a content provider who finds a creative use of the VBI to provide, in conjunction with an application developer, an enhanced service will actually be able to deliver that service to the consumer.

This issue can be viewed as an example of the difficulty of defining what openness in access means, even within the existing rather limited sorts of broadband distribution that exist today. A service may no longer be a two-party transaction between the provider and the end customer. It may now involve at least three parties and become dependent upon the cooperation of the access network provider. This substantially raises the hurdle to entering the business with a new service, particularly one that may be of interest to only a small percentage of consumers. Even though a broadcast channel is already carried by the vast majority of access network providers, it may be necessary to reengage all such network providers to even begin to offer a service. This certainly provides a barrier to innovation that need not be present. It is exactly this sort of barrier that we are concerned not be erected as digital services begin to be deployed. It must be possible for innovation to be between the provider and the consumer over an open digital highway. Conversely, it must also be possible for the network service provider to benefit from the expanded use of the network.

### **Openness of Equipment**

Another key difference between the current directions in the cable access networks and the existing telephony networks is important to the rest of this discussion. This is the issue of open equipment connectivity and ownership of customer premises equipment (CPE). Today, the consumer owns the telephone, the TV, the PC,

and the modem for the PC. However, it is generally the case that the cable access provider owns and provides the set-top box. Furthermore, the lack of standards for the physical network may make it more difficult to find an alternative supplier of CPE. This is quite different from the telephony world, where the access network provider terminates his network at a demarcation point, providing a standard form of connection to which customer-selected and customer-owned CPE can be attached. The combination of the ownership and the standards aspects of this arrangement have together spawned a vigorous industry in CPE suppliers. For telephony, the connectivity standard is the RJ-11 connector and the set of electrical and tone signaling standards supported over it. As a result, PC access to networks via modems can be achieved with the same equipment, anywhere in the country.

The importance of the openness of CPE is beginning to be understood. For example, a press release from Congressmen Bliley and Markey issued on March 16, 1995, states,

Restricting consumers' ability to purchase and own cable "set-top boxes" and other communications interface equipment is like putting a straight-jacket on technological development.... There's no incentive for improvement because there's no competition.... Today's advanced phones only happened because of a healthy, competitive retail market—and so did the revolution in computer modems and fax machines that followed.

However, there is more to energizing the broadband CPE industry than just allowing retail purchase of the set-top box. Consumers must see an advantage in purchasing the equipment, and vendors must see a large market to be served in order to invest in developing products that will attract consumers. Neither of these will happen very quickly without the development of standards for the point of attachment of the equipment.

Consider how likely consumers would be to purchase feature-laden phones if they could not expect those phones to work after they moved homes. Likewise, consider how likely a company would be to develop an innovative new phone product knowing that it could be sold only to that set of consumers who were served by some specific phone companies, and that the company would also have to deal with unhappy consumers who discovered that the product ceased to work after they changed providers. Our concern is that consumers be able to buy PC equipment and applications that will effectively attach anywhere in the United States to the NII, and that they be able to freely move such equipment and applications with them. Industry-driven discussions are under way in various industry groups (e.g., the ATM Forum) toward the establishment of such de facto standards or implementation agreements.

### **Affordability**

A considerable concern exists about the affordability of an open consumer broadband data service. Clearly, the bandwidth must be priced at a level that will allow reasonable access to a broad spectrum of users. True open competition should cause this to occur, as can be seen by looking at how other uses of the bandwidth in the access network might be priced. For example, consider a higher-level service that has been proposed, namely video on demand (VOD). This service promises to deliver to the consumer an online replacement for videotapes (and eventually probably much more interactive sorts of experiences). Assuming that one uses MPEG2 to compress a 2-hour movie, then this service must deliver on the order of 4 to 6 Mbps for 2 hours to the consumer at a price competitive with what she can rent the movie for today (on the order of \$3.00). VOD is an individual service delivered to a single consumer at a time. Thus it is reasonable to expect that competition will drive the cost of similar downstream bandwidth of an unrestricted sort to be similar. (Actually, the resources to deliver the unrestricted bandwidth are less since no server is involved.) Clearly, this accounts only for the access network provider part of the unrestricted service, but it at least provides a starting point. It is harder to estimate what a consumer should be paying for upstream bandwidth, but a similar sort of analysis for a bundled service that makes more upstream demands (perhaps shopping or interactive gaming) should provide an estimate there as well. We are not trying to suggest any particular price structure for consumer-controlled bandwidth, nor are we suggesting price regulation. Rather, we are trying to suggest that competitive forces should cause it to be priced in a somewhat similar manner to similar levels of bandwidth utilized for services bundled by the local access network provider.

### AN ARCHITECTURE FOR A LONG-TERM RESIDENTIAL NII

The actual deployment of broadband services to the residential NII will be evolutionary from today's physical networks. Economic and practical considerations preclude any revolutionary replacement of access networks with some uniform and full-service digital infrastructure. As a result of this evolutionary reality many of the requirements we have for a long-term infrastructure cannot be fully met in the short term. Indeed, some of the standards that we believe will be critical to providing the full potential of the NII are still in embryonic stages. Nevertheless, it is important to develop a vision of where we would ideally like to wind up if we are to generate public policies that encourage development to proceed toward an end-point architecture that can realize all the potentials of the NII.

We believe that the architecture should be broken into three parts: core networks, access networks, and home networks (Figure 1). Core networks are those networks that provide generalized connectivity both locally and across wide geographical areas. They correspond in today's telephony infrastructure to the long-distance carriers plus that part of the local exchange carriers' infrastructure that carries traffic between the central offices that serve customers. Access networks are those networks that connect the core networks to residences. For example, the HFC networks that are being considered for deployment to connect regional hub systems to homes are access networks. Finally, home networks are those that operate within the residence.

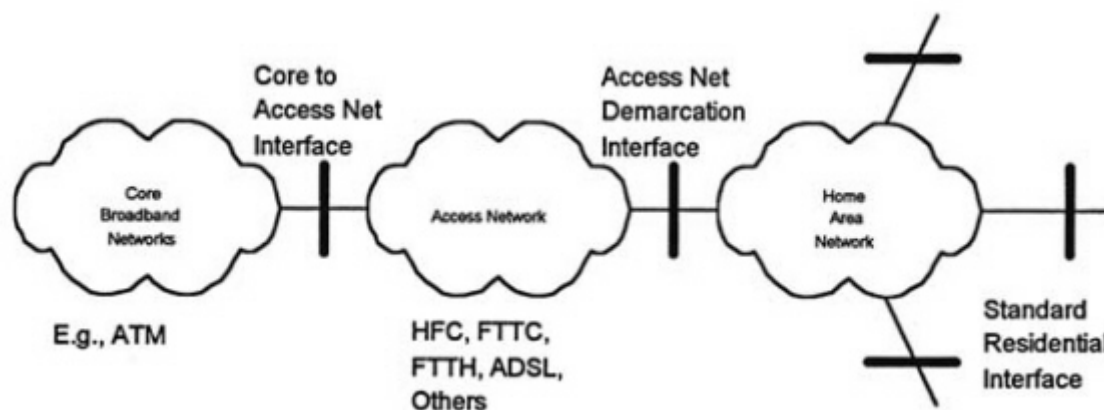


Figure 1 Elements of an architecture for realizing the full potential of a national information infrastructure. ATM, asynchronous transfer mode; HFC, hybrid fiber coaxial; FTTC, fiber to the curb; FTTH, fiber to the home; and ADSL, asynchronous digital subscriber line.

Note that any of these networks can be nonexistent in a specific deployment. In an existing cable system using HFC to deliver only broadcast-type services to dedicated set-top boxes in a home, neither a core network nor a home network may exist. Likewise, one could view the current telephony infrastructure as often not having an access or home network, since for practical purposes it generally appears that the CPE connects directly to the core network.

We expect that core networks will typically be those with symmetric bandwidth. Considering today's trends, these are likely to be ATM-based networks. We also believe that it is in the interest of the carriers that provide core networks to encourage use of bandwidth by consumers. There is today competition among core network providers in the telephony long-distance business, and we would encourage such competition to continue into the broadband world. Given this competition, the easy access by the consumer to multiple core network providers, and the desire of those providers to sell their bandwidth, we believe that the requirement of sufficient, affordable, consumer-directed bandwidth will be met in these networks. Furthermore, there is already much momentum in the standards associated with communications over these networks (ranging from TCP/IP to ATM) so that effective consumer utilization of this bandwidth seems possible.

Home networks are not yet at all well defined. However, as one sees more digital information appliances deployed within the home, the desire to interconnect them will increase, and this will give rise to various sorts of home networks. Displaying images from security cameras on a TV or PC, driving an interactive program on a TV from a PC, or operating a multiplayer game across multiple TVs or PCs in one or multiple homes are all examples of applications that will involve a home network. Home networks are likely to be fairly symmetric, if for no other reason than that there will likely not be a single special "source" to drive any asymmetry.

Access networks may well remain asymmetric regarding bandwidth due to the economics of some of the popular access network topologies. For example, HFC will likely continue to supply much more downstream bandwidth than upstream bandwidth for the foreseeable future. We are most concerned about the evolution of the access networks since they are the critical bridge between the consumer and the core networks. While the consumer can choose what to deploy at home, and there is already a path to active competition in the core networks, the access networks may have much less natural competition.

Once one understands these three related networks, it becomes clear that the critical issue for them is to define stable interfaces that can allow them to evolve independently while still delivering to the consumer a high level of service. The connection standards between the access networks and the core networks are already getting some amount of attention via the discussion of open video service providers. That is, this interface is the one that a service provider will be concerned about, whether the service is a local video store or a general long-distance carrier service. The connection standards between the access network and the home network have received less attention. This is the interface discussed above in the section titled "Openness of Equipment." We believe that it will be critical to elevate the discussion of this interface and build policy that actively separates the home network from the access network via appropriate standards for the reasons discussed above. In addition to the standards, this is also the interface across which we believe we must guarantee reasonable consumer-directed bandwidth.

If we define effective standards at each of these points, then it should be possible for an information industry to develop that supplies consumers with equipment and applications that allow wide exploitation of the NII. In our vision, consumer hardware has a common socket for attaching to the NII. It uses common protocols for interacting with NII services. These protocols operate in the packetized Internet world and in the home to allow easy access to the educational and information resources of the Internet. The typical home has access to sufficient bandwidth to make access to these network resources a pleasant experience. The choices of what network services are available to the residential consumer are essentially unbounded. The local access network provider may choose to package and sell some services, thus making them easier to use, but there are no roadblocks to open consumer access to the NII at large.

The key to all of this is the existence of the interface agreements that allow development on either side of the interface to progress independently and that do not overly constrain the sorts of implementations permissible between the interfaces. Without such industry-driven standards or implementation agreements we will be in danger of one of two extremes. On the one hand, their lack will cause the coupling of what should be distinct parts of the NII, thus slowing development of what should be independent parts. On the other hand, their overspecification will stifle creativity by admitting only a single family of products that can operate within the NII.

With open bandwidth and protocols, intelligence can exist at the periphery of the network as well as inside it. An innovator with a good idea can create an information service as an end point on the network and sell to the consumer an access tool that resides at the consumer's system. In doing so there will not be any impediments due to a need to negotiate business deals with various network providers. This openness will lead to an opportunity for much greater innovation than that possible with an architecture that gives preference to the provision of service intelligence only inside the networks.

#### RECOMMENDATIONS—REQUIREMENTS FOR WIDESPREAD PC ACCESS

1. Provide reasonable levels of bandwidth to and from the home at consumer costs. Data bandwidth into the home should probably at least mirror what is available today for the corporate user accessing the Internet. This argues for a minimum of T1 (1.5 Mbps)-level rates to each home, and more likely, for peak levels of at least Ethernet (10-Mbps) speeds. Obviously more is better, but the key issue is that the consumer must be able to

access rich multimedia content at speeds that make the operation a pleasant one. Initially, it is acceptable that the bandwidth from the home be less than that into it. However, extremely low outbound bandwidth will limit the efficiency of any interaction. Furthermore, there are certainly interesting applications that will be enabled only by relatively high outbound bandwidth as well (for example, work at home or video security). Ultimately, the access network should not overly constrain who can be the providers of services and who can be only consumers. With the right bandwidth levels, the NII can truly energize a grassroots, distributed economy. The first step in this direction can come from the rapid, ubiquitous deployment of ISDN to the residential environment. While it does not meet the full requirements, it can quickly get us moving toward the long term and can allow interesting applications to begin to be deployed as it is embraced by existing service providers.

2. Permit consumer control over the use of this bandwidth to conduct simultaneous digital interactions with multiple services via open packet protocols. It must be possible for intelligence in the user's system to efficiently access multiple services in the NII to generate value at the end system by the combination of services. The openness of the protocols and of the connectivity they provide is absolutely key to enabling the development of new services and to competition among service providers. Consumers should be able to access any service in the NII, just as today they can access any telephone from their residential telephones. The choice of local access provider should not preclude this accessibility.
3. It must be reasonable for multiple PCs and other information appliances within a residence to be simultaneously accessing the Internet or other broadband services. The fact that a child is conducting an educational interaction with some service on a PC should not preclude a parent from conducting a financial transaction at the same time. It is desirable, in fact, that all home information appliances be viewed as equals to that there is at least logical symmetry between originating information inside or outside the home for consumption inside or outside the home.
4. Encourage sufficient standards to facilitate a commodity consumer business on at least the national level for PC connectivity. Competition drives costs down and service levels up. However, meaningful competition cannot take place unless a large marketplace becomes available to the potential suppliers. Consider the differences in corporate networking costs today versus 15 years ago. Standards such as Ethernet have energized an entire industry, with the result that the cost of connectivity for an end system has approached \$100. In the consumer market segment, the standards in place for telephone connectivity have done the same thing for products that attach to an RJ-11 socket. This benefit will not accrue to NII connectivity if every regional or metropolitan market segment requires different consumer equipment for connectivity. For example, the consumer is not served well by the multiplicity of ISDN "standards" across the country. A resident of the United States should be able to move NII equipment anywhere in the country and use it effectively, just as can be done today with other appliances. It is important to note, however, that the most effective standards in the business world are those that have been initially industry-driven, de facto ones that only later were codified into a de jure form. (For example, the nearly ubiquitous Ethernet and TCP/IP came about in this manner, while the OSI protocols represented an attempt to drive standards from the de jure side first.) The role of the government should not be to impose standards but rather to create policy that facilitates the rapid development of appropriate industrial ones.
5. Permit consumer ownership of the periphery of the network in terms of what kind of equipment is connected. (Of course, this should not preclude an access network provider from leasing equipment to the consumer as an alternative.) Similar to the previous point, we can see an entire industry that has innovated based on the ability of companies to develop products to address consumer needs, without needing the permission, or worse, the active participation of intermediaries to become successful. Look at the innovation that has occurred since the *Carterfone* decision opened up access and ownership of end equipment for the telephone network. The evolution of capabilities offered by the personal computer is another example of market-driven innovation. Locking access networks into an environment where only provider-owned or provider-sanctioned interfaces are permitted to be attached to them recreates the old phone network and its constrained equipment competitive environment.



## CONCLUSIONS

It is not clear to us at this point that the natural and policy forces at work in the emerging NII will achieve all the important results outlined above. There is, unfortunately, something of a chicken-and-egg problem concerning open residential use of the NII. The emergence of innovative NII applications utilizing residential broadband capabilities depends on the existence of those capabilities, while justification for the deployment of those capabilities requires hypothesizing the existence of those applications. As a result, we have a situation where to some extent policy choices may need to precede market development. We do not suggest a move toward a highly regulated environment, since ultimately that is completely counterproductive to the development of a new industry. However, we do suggest that policymakers need to find ways to encourage the development of an eventual architecture that supports the full potential of the NII.

To some extent this notion is already present in the types of trade-offs being made to balance the cable industry's and telephony industry's developing competition in each other's businesses. We are simply encouraging policymakers to take a broader view of this set of problems that looks beyond this level of competition to include the rest of the infrastructure that the consumer will need to become a full citizen on the NII. Policymakers should see that more is involved than allowing content providers open access to consumers or considering what the competitive trade-offs should be between allowing cable systems to offer dialtone and telephone companies to offer video. They also need to look at the provision of general data services from the perspective of the consumer. We believe that more attention should be given to the provision of open, standardized, commodity-priced network access to the NII at large. We believe that only with this capability can we tap the PC's full potential to become the consumer's interactive access point to NII services and bring to the residential consumer the dynamics that have so dramatically benefited the corporate PC user.

## 39

# The National Information Infrastructure: A High-Performance Computing and Communications Perspective

*Randy H. Katz, University of California at Berkeley*

*William L. Scherlis, Carnegie Mellon University*

*Stephen L. Squires, Advanced Research Projects Agency*

### ABSTRACT

Information infrastructure involves more than the building of faster and larger computers and the creation of faster and better communications networks. The essence of information infrastructure is a diverse array of high-level information services, provided in an environment of pervasive computing and computer communications. These services enable users to locate, manage, and share information of all kinds, conduct commerce, and automate a wide range of business and governmental processes. Key to this is a broad array of rapidly evolving commonalities, such as protocols, architectural interfaces, and benchmark suites. These commonalities may be codified as standards or, more likely, manifest as generally accepted convention in the marketplace.

Information technology has become essential in sectors such as health care, education, design and manufacturing, financial services, and government service, but there are barriers to further exploitation of information technology. Pervasive adoption of specific service capabilities, which elevates those capabilities from mere value-added services to infrastructural elements, is possible only when value can be delivered with acceptable technological and commercial risk, and with an evolutionary path rapidly responsive to technological innovation and changing needs. Private- and public-sector investment in national information infrastructure (NII) is enabling increased sectorwide exploitation of information technologies in these national applications areas. Although the private sector must lead in the building and operation of the information infrastructure, government must remain a principal catalyst of its creation, adoption, and evolution.

This paper explores the barriers to achieving NII and suggests appropriate roles for government to play in fostering an NII that can be pervasively adopted. The main locus of government activity in research and early-stage technology development is the federal High Performance Computing and Communications (HPCC) program. This program is evolving under the leadership of the National Science and Technology Council's Committee on Information and Communications.

### INTRODUCTION

Information technologies are broadly employed in nearly all sectors of the economy, and with remarkable impact. Nonetheless, there are still enormous unrealized benefits to be obtained from effective application of information technology, particularly the intertwining of multiple distributed computing applications into national-scale infrastructural systems. In many sectors, including health care, education and training, crisis management, environmental monitoring, government information delivery, and design and manufacturing, the benefits would have profound significance to all citizens (as suggested in IIS, 1992, and Kahin, 1993). These sectors of information technology application have been called national challenge (NC) applications.

The pervasiveness and national role of the NC applications prevent them from developing dependency on new technologies, even when those technologies offer important new capabilities. This is so unless the risks and costs are manageable, and there is a clear trajectory for growth in capability and scale,

and that growth is responsive to new technologies and emerging needs. For this reason, while computing and communications technologies have taken hold in numerous specific application areas within these sectors, in most cases the challenge remains for advanced information technology to take on a significant sectorwide infrastructural role.

The NII, in the simplest terms, consists of available, usable, and interoperable computing and communications systems, built on underlying communications channels (the bitways) and providing a broad range of advanced information technology capabilities (the services). These services provide the basis for a wide range of use (the applications), ranging in scale up to national challenge applications. A key point is that NII is far more than communications connectivity; indeed it is generally independent of how communications connectivity is supplied.

Generally speaking, infrastructural systems consist of ubiquitous shared resources that industry, government, and individuals can depend on to enable more productive and efficient activity, with broadly distributed benefit (Box 1). The resources can include physical assets, such as the national air-traffic control system or the nationwide highway system. The resources can also include key national standards, such as the electric power standards, trucking safety standards, railroad track structure, and water purity standards. The resources are ubiquitous and reliable to an extent that all participants can commit to long-term investment dependent on these resources. This also implies a capacity for growth in scale and capability, to enable exploitation of new technologies and to assure continued value and dependability for users. The value can be in the form of increased quality and efficiency, as well as new opportunities for services.

It is therefore clear that a critical element of NII development is the fostering of appropriate commonalities, with the goal of achieving broad adoptability while promoting efficient competition and technological evolution. Commonalities include standard or conventional interfaces, protocols, reference architectures, and common building blocks from which applications can be constructed to deliver information services to end users. A fundamental issue is management and evolution, and in this regard other examples of national infrastructure reveal a wide range of approaches, ranging from full government ownership and control to private-sector management, with government participation limited to assurance of standard setting.

The Clinton Administration, under the leadership of Vice President Gore, has made national information infrastructure a priority (Gore, 1991; Clinton and Gore, 1993), as have other nations (examples: NCBS, 1992, and Motiwalla et al., 1993). The NII vision embraces computing and communications, obviously areas of considerable private investment and rapid technological change. The definition of the government role in this context has been ongoing, but several elements are clear. At the national policy level, the NII agenda embraces information and telecommunications policy, issues of privacy and rights of access, stimulation of new technologies and standards, and early involvement as a user (IITF, 1993). The federal High Performance Computing and Communications (HPCC) program, and the advanced information technologies being developed within it, play a key role in addressing the research and technical challenges of the NII.

In this paper we examine several aspects of the conceptual and technological challenge of creating information infrastructure technologies and bringing them to fruition in the form of an NII built by industry and ubiquitously adopted in the NC applications sectors. Topics related to telecommunications policy, intellectual property policy, and other aspects of information policy are beyond our scope.

In the first section below, we examine the federal government's role in fostering NII technologies and architecture. We then analyze the relationship between high-performance technologies and the NII and describe our three-layer NII vision of applications, services, and bitways. This vision is expanded in the next two sections, in which the NC applications and the technologies and architectural elements of the services layer are discussed. We present the research agenda of the Federal High Performance Computing and Communications Program in the area of information infrastructure technologies and applications, followed by our conclusions.

## ROLE OF THE FEDERAL GOVERNMENT IN INFORMATION INFRASTRUCTURE

New technologies are required to achieve the NII vision of ubiquitous and reliable high-level information services. Many of the envisioned NII services place huge demands on underlying computing and communications capabilities, and considerable energy is being applied in industry, government, and research to creating these new capabilities. But there is more to NII than making computers faster, smarter, and more widely connected together.

Creation of national infrastructure entails delivery of services with sufficient reliability, ubiquity, and freedom from risk that they can be adopted sectorwide in national applications. The challenge to achieve this is considerable in any infrastructural domain and particularly difficult in information infrastructure. These goals usually involve rigorous standards and stability of technology, which appear all but precluded by the extremely rapid evolution in every dimension of information technology.

In the development of other kinds of national infrastructure, government has had a crucial catalytic role in fostering the broad collaboration and consensus-building needed to achieve these goals, even when industry has held the primary investment role in creating the needed technologies and standards.

In the case of national information infrastructure, it is manifestly clear that it should not and indeed cannot be created and owned by the government. But the catalyzing role of government is nonetheless essential to bring the NII to realization. The government has an enormous stake in the NII as a consequence of its stake in the national challenge applications. Information infrastructure technologies play a critical role in the federal government's own plan to reengineer its work processes (Gore, 1993). Vice President Gore draws an analogy between the NII and the first use of telegraphy:

Basically, Morse's telegraph was a federal demonstration project. Congress funded the first telegraph link between Washington and Baltimore. Afterwards, though—after the first amazing transmission—most nations treated the telegraph and eventually telephone service as a government enterprise. That's actually what Morse wanted, too. He suggested that Congress build a national system. Congress said no. They argued that he should find private investors. This Morse and other companies did. And in the view of most historians, that was a source of competitive advantage for the United States.

Government fostered the technology through initial demonstrations and encouragement of private investment. But the U.S. telecommunications infrastructure has been built with private funds. And analogously, the NII implementation must be a cooperative effort among private- and public-sector organizations.

What are the specific roles for government? Addressing this question requires understanding how the NII differs from other major federal research and development efforts. The following characteristics summarize the differences:

- The scale of the NII is so huge that government investment, if it is to have an impact, must be designed to catalyze and stimulate investment from other sources rather than subsidize creation of the NII itself. The NII will emerge as an aggregation of many distinct entities that compete to provide products and services. Of course, rudimentary elements of the NII not only are in place but also constitute a major sector of the economy.
- The NII will provide long-term infrastructural support for applications of national importance, such as health care and education. Decisionmakers in these application sectors cannot put new technologies in a pivotal role, even when they offer important new capabilities, unless the risks and costs of adoption are manageable. Adoption risk issues include, for example, scale-up, ability to evolve gracefully, competition and mobility among suppliers, and commitment to internal systems interfaces.

- The NII will support applications by delivering common services well above the level of simple access to telecommunications. These services can include, for example, mechanisms to protect intellectual property and user privacy, support for information management and search, and support for managing multimedia objects. The high level of services will continue to evolve in capability and power, but there is nonetheless an essential requirement to achieve elements of commonality in the system interfaces through which these services can be delivered by NII providers to application users. (This issue is elaborated on below.)
- The NII depends on advanced computing and computer communications technologies whose evolution is, in large measure, the result of continued government research investment. Government basic research investment continues to be a primary source of the ideas and innovations that stimulate U.S. industry, sustain a high level of competitiveness in the market, and provide a national competitive advantage.

These considerations yield a four-pronged strategy for government investment in research and development related to the NII:

- Research and new technology creation;
- Interoperability, commonality, and architectural design;
- Application demonstration and validation; and
- Aggressive early use.

The first of these elements, research, is clear. Government has a traditional role as farsighted investor in long-term, high-risk research to create new concepts and technologies whose benefits may be broadly distributed. In the case of the NII, the government needs to invest both in problem-solving research, to fulfill the promise of today's vision, and also in exploratory research to create new visions for tomorrow. Government investment in research and development can support the rapid and continual transition of new NII capabilities into commercialization and adoption. Basic research can yield paradigmatic improvements with marketwide benefits. Intensive discussions among leaders from academia, industry, and government have been under way to develop a view of the technical research and development challenges of the NII (Vernon et al., 1994).

The second element involves stimulating commonalities within the NII that can achieve economies of scale while simultaneously creating a foundation for a competitive supply of services. Interface and protocol commonalities foster conditions where the risks of entry for both users and creators of technology are reduced. We use the term commonality because it is more inclusive than the conventional notion of standards. It covers routine development of benchmarks, criteria, and measures to facilitate making choices among competing offerings. It also encompasses the definition of common systems architectures and interfaces to better define areas for diversity and differentiation among competing offerings. Common architectural elements help both developers and users decouple design decisions. Of course, inappropriate standards can inhibit innovation or predispose the market to particular technological approaches. A critical issue for the NII is the speed of convergence to new conventions and standards. In addition, conventions and standards must themselves enable rapid evolution and effective response to new technology opportunities. These are familiar issues in the realm of conventionalization and standards generally; but they are also among the most fundamental considerations in achieving new high-level NII services, and are in need of specific attention.

Demonstration, the third element, involves government sponsorship of testbeds to explore scalability and give early validation to new technology concepts. Testbeds can span the range from basic technologies coupled together using ad hoc mechanisms, to large-scale integration projects that demonstrate utility of services for applications in a pilot mode. These latter integration experiments can bootstrap full-scale deployments in applications areas.

### **BOX 1 INFORMATION INFRASTRUCTURE—SHARED RESOURCES**

The analogy between information infrastructure and the interstate highway system has helped bring the concept of the NII into the popular consciousness. The analogy is apt, and not only because of the role Vice President Gore's father played in drafting the legislation that led to the interstate highway system.

The fundamental commercial utility of the highway system (and the railroads) is the substitution of transportation for local production, enabling new economies of scale and resource sharing in many industries. Economies of scale in manufacturing industries can be such that the cost of large-scale remote manufacture combined with transport can be significantly less than the cost of local production.

The highways are infrastructural in that they make entire industries more efficient, resulting in better value for the customer and expanded markets for producers. It is this value that justifies the public investment in the infrastructure.

More importantly, it explains why public investment and policy support are necessary stimuli for infrastructure development: because shared infrastructure provides advantage to all who use it, there is no particular competitive incentive for specific infrastructure users (producers or consumers) to invest directly in its creation. On the other hand, expanded markets benefit all users, and so user investment in infrastructure is justified if it is distributed equitably. For example, public highways may receive their initial funding from bond issues combined with direct taxpayer support, with operating costs and loan service costs funded by users through tolls and licensing.

The highway system infrastructure is sustained through a continually evolving set of interrelated infrastructure elements, such as roadbed engineering standards, speed limits, roadside amenities, interstate trucking regulations, tolls, and routing architecture. National-scale users are able to rely on this structure and thus commit the success of their enterprises to its continued existence without expanding their risks.

Information infrastructure development involves an analogous set of elements. Network protocols, application interfaces, interoperability standards, and the like define the class of mechanisms through which value is delivered. Value comes from high-level services such as electronic mail, information services, remote access, and electronic commerce all supporting a rich variety of information objects. Reliance and commitment from national-scale users depend on breadth and uniformity of access, common architectural elements, interoperability, and a reasonably predictable and manageable evolution.

The importance of common architectural elements to infrastructural utility must not be understated. Rail-gauge standardization is a canonical example. But commitment to common architectural elements must also include commitment to a process for evolving them. Achieving the right balance is a principal challenge to creating an adoptable national information infrastructure.

Finally, acting in the interest of government applications, the government can take a proactive role as consumer of NII technologies to stimulate its suppliers to respond effectively in delivering information infrastructure that supports government applications. Possible government applications include systems for government information, crisis response, and environmental monitoring.

The gigabit testbeds in the HPCC program offer a model for research partnerships among government, industry, and academe and represent a resource on which to build prototype implementations for national applications. Each testbed is cost-shared between government and the private sector and embraces the computer and telecommunications industries, university research groups, national laboratories, and application developers. The key function of the testbeds is to experiment with new networking technology and address interoperability and commonality concerns as early as possible.

### **RELATIONSHIP BETWEEN HIGH-PERFORMANCE TECHNOLOGIES AND THE NII**

The federal HPCC program supports the research, development, pilot demonstration, and early evaluation of high-performance technologies. HPCC's focus in its initial years was on the grand

challenges of science and engineering, with a strategy of developing a base of hardware and software technologies that can scale up to large-scale processing systems, out to wide-area distributed systems, and down to capable yet portable systems (FCCSET, 1994; CIC, 1994). These scalable technologies will contribute strongly to the NII, as will the legacy of cooperation between government, industry, and academia. These can greatly accelerate the establishment of an evolvable information infrastructure architecture, with testbed development, protocol and architecture design, interoperability experiments, and benchmarking and validation experiments. This legacy has helped facilitate adoption of HPCC-fostered technologies by independent users by significantly reducing their costs and risks of adoption.

This twofold HPCC stimulus, of research and cooperation, combines with a program emphasis on demonstration, validation, and experimental application to create a framework for government technology investment in NII. For this reason, HPCC was expanded in FY1994 to include a new major program component, Information Infrastructure Technology and Applications (IITA), focusing directly on creation of a universally accessible NII, along with its application to prototype NC applications. (These activities are described in more detail in the section below titled "The Federal HPCC Program and the NII.")

Each of the other HPCC program activities contributes to IITA. For example, emerging large-scale information servers designed to provide information infrastructure services are based on HPCC-developed, high-performance systems architectures, including architectures based on use of advanced systems software to link distributed configurations of smaller systems into scalable server configurations. The microprocessors used in these large-scale systems are the same as those found in relatively inexpensive desktop machines. High-performance networking technologies, such as communications network switches, are increasingly influenced by processor interconnection technologies from HPCC. Networking technologies are also being extended to a broad range of wireless and broadcast modalities, enhancing mobility and the extent of personal access. Included in this effort are protocols and conventions for handling multimedia and other kinds of structured information objects.

NII can be viewed as built on a distributed computing system of vast scale and heterogeneity of an unprecedented degree. HPCC software for operating systems and distributed computing is enhancing the interoperability of computers and networks as well as the range of information services. The software effort in the HPCC program is leading to object management systems, methodologies for software development based on assembly of components, techniques for high assurance software, and improvements to programming languages. These efforts will contribute to the development and evolution of applications software built on the substrate of NII services.

### THREE-LAYER NATIONAL INFORMATION INFRASTRUCTURE ARCHITECTURE

Within the HPCC community, a much-discussed conceptual architecture for the National Information Infrastructure has three major interconnected layers: National Challenge Applications, supported by diverse and interdependent NII communication and computation Services, built on heterogeneous and ubiquitous NII bitways (see [Figure 1](#)). Each layer sustains a diverse set of technologies and involves a broad base of researchers and technology suppliers, yielding a continuously improving capability for users over time. By delivering utility to clients in the layers above through common mechanisms or protocols, a rapid rate of evolution of capability can be sustained in a competitive environment involving diverse suppliers. Thus, developments in each of these layers focus both on stimulating the creation of new technologies and on determining common mechanisms or protocols—the commonality—through which that capability can be delivered. For example:

- The keys to scaling up in national challenge applications are often in the choice of common application-specific protocols. For example, manufacturing applications require shared representations for product and process descriptions to support widespread interoperability among design systems and tools.

- Services such as multimedia multicast can be provided to developers of Application capabilities through proper adherence to common protocols. With well-designed protocols and interfaces, rapid growth in multimedia capability and capacity can be delivered to end users and applications developers without requiring major reengineering of the whole applications-level systems. Services are also interdependent and themselves evolve in this manner.
- The diverse bitways technologies deliver communications capability in a uniform manner through use of standard protocols, such as SONET/ATM. This has the effect of insulating developers of NII services from the details of the rapidly evolving communications technologies used to deliver information capabilities to end users and applications.

This architecture addresses directly the challenge of scale-up in capability, size, and complexity within each of the three layers. Ongoing validation of concepts can be achieved, in each layer, through large-scale testbed experimentation and demonstration conducted jointly with industry, users, and suppliers of new technologies and information capabilities. If the evolution of the NII architecture proceeds as envisioned, the result will be the integration of new capabilities and increased affordability in the national challenge applications. Each layer supports a wide range of uses beyond those identified for the specific national challenge applications. For example, generalized NII service and bitway technologies can also support applications on a very small scale, extensions of existing services, ad hoc distributed computing, and so on.

The national challenge applications are described in more detail in the next section, with the issues addressed by the services layer in the succeeding section titled "Services." Bitways technologies are well covered in other sources, such as *Realizing the Information Future* (CSTB, 1994), and are not discussed here.

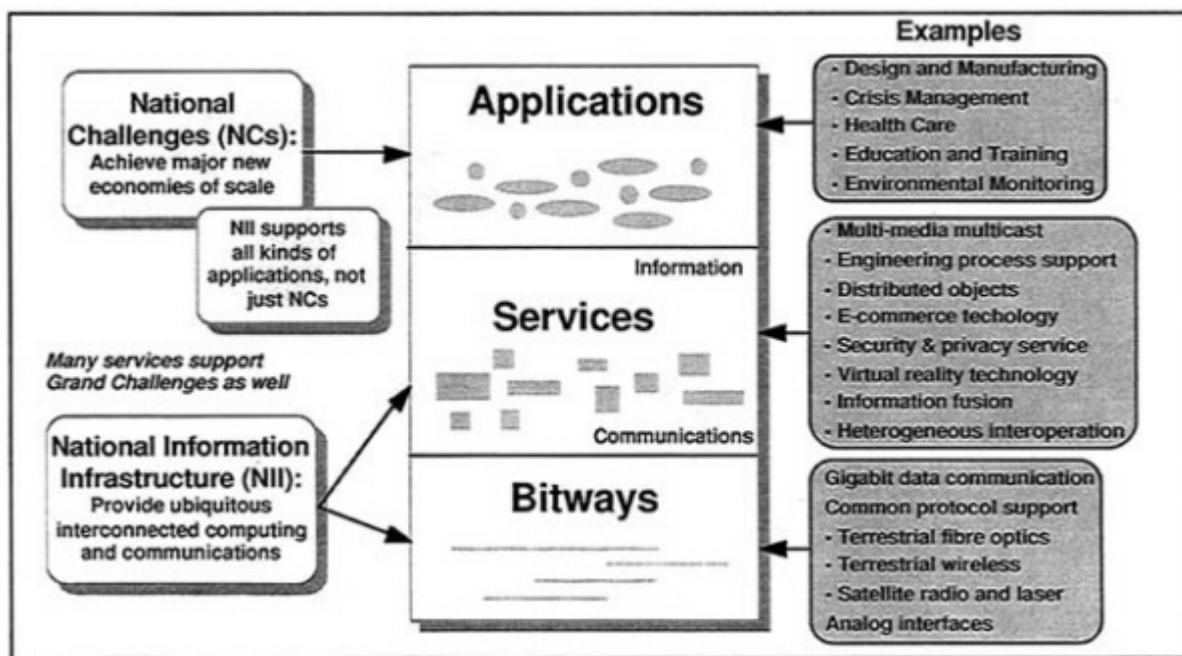


Figure 1 A model three-layer architecture for the NII. Bitways provide the communications substrate, the applications layer supports the implementation of the NCs, and the services layer provides the bridge between communications and information. (a) Despite the need for similar application-enabling services, each NC sector might reimplement these from scratch, yielding overly expensive stovepipe systems if there were no common services layer. (b) A common services layer coupled to toolkits for building applications.



## NATIONAL CHALLENGE APPLICATIONS

Numerous groups have developed lists of critical applications, characterized by the potential for a pervasive impact on American society and exploitation of extensive communications and information processing capabilities. For example, in 1993 the Computer Systems Policy Project identified design and manufacturing, education and training, and health care as the national challenges (CSPP, 1993). A more exhaustive list has been developed by the Information Infrastructure Task Force, representing the union of much of what has been proposed (IITF, 1994). Among these NC applications are the following:

- *Crisis Management:* Crisis management systems exploit information technology to ensure national and economic security through various kinds of crises. This is accomplished by providing timely data collection and intelligence fusion, advanced planning tools, rapid communications with defense forces spread around the globe, and a command and control ability to respond quickly to crises. The same basic capabilities can be deployed on a smaller scale to respond to local emergencies, such as devastating hurricanes, earthquakes, or fires.
- *Design and Manufacture:* These systems integrate engineering design with product manufacturing, to reduce the time to create new products, to lower production costs, and to increase product quality. In a wider sense, a pervasive design and manufacturing system should couple suppliers to their customers throughout the production chain. Goals are more responsive product design, manufacture, and just-in-time warehousing and product delivery.
- *Education and Training:* These systems provide access to online instructional and research materials, anywhere and anytime, as well as more direct communication among students and educators. Once created and made accessible, instructional materials may be reused and evolved by instructors around the country. For example, educational use of the information infrastructure can enable distance learning, where students in remote locations can gain access to specialized instruction. Training could exploit simulation coupled with remote access to actual apparatus.
- *Environmental Monitoring:* These systems integrate data from ground, airborne, and space-based sensors to monitor (and potentially respond to) environmental changes. They may be used to discover a nuclear accident in progress, oncoming climatic effects such as smog conditions, or can be exploited for longer-term studies such as climate change.
- *Government Information Delivery:* Citizens have a right to ready, low-cost access to government information that they have already paid for, including economic statistics, trade information, environmental and land use information, and uniform one-stop shopping for government services such as veterans' and social security benefits.
- *Health Care:* These systems use information technologies to improve the delivery of health care, by providing ready access to patient records, remote access to medical expertise, support for collaborative consultations among health care providers, and rapid, paperless claims adjustment that can help reduce health care costs.

There are two additional applications that sit at the interface of the national challenges and the underlying service layer: digital libraries and electronic commerce. In a sense, these are fundamental enablers for information access and electronic exchange of value and will be extensively used by virtually all of the other NC applications described above.

- *Digital Libraries:* A digital library is a knowledge center without walls, accessible from anywhere through networked communications. These systems are leading to significant advances in the generation, storage, and use of digital information of diverse kinds. Underlying services and technologies range from advanced mass storage, online capture of multimedia data, intelligent information location and

filtering, knowledge navigation, effective human interfaces, system integration, and prototype and technology demonstration.

- *Electronic Commerce:* Electronic commerce integrates communications, data management, and security services to allow business applications within different organizations to automatically interchange information. Communications services transfer the information from the originator to the recipient. Data management services define the interchange format of the information. Security services authenticate the source of information, verify the integrity of the information received by the recipient, prevent disclosure of the information to unauthorized users, and verify that the information was received by the intended recipient. Electronic commerce applies and integrates these infrastructure services to support business and commercial applications, including financial transactions such as electronic bidding, ordering and payments, and exchange of digital product specifications and design data.

In each of these applications there is an unmet challenge of scale: How can the service be made ubiquitously available with steadily increasing levels of capability and performance? The applications communities depend on information technology for solutions but are facing scaling barriers, and hence the NII goal of crossing the threshold of ubiquity. In the absence of common architectural elements, such as interfaces, methods, and modules, it may be possible to demonstrate prototype solutions to specific applications problems through monolithic stovepipes. But these solutions may not give any means to pass this threshold of pervasiveness and dependability.

## SERVICES

### Overview

As we have noted, information infrastructure is more than bandwidth, switching, and ubiquitous communications access. It is (1) the common service environment in which NC applications are built. All applications share generic service needs: human interfaces (e.g., graphical user interaction, speech recognition, data visualization), application building blocks (e.g., planning subsystem, imaging subsystem), data and process management (e.g., search and retrieval, hyperlink management, action sequencing), and communications (e.g., IPC, mobile computation). Also, the engineering of applications requires (2) tools in the form of development environments, toolkits, operational protocols, and data exchange and action invocation standards from which service solutions can be combined, integrated, and reused. Finally, the engineering of applications becomes more efficient (as is already occurring for shrink-wrap software running on personal computers) in the presence of (3) a marketplace of reusable subsystems; in this manner, applications systems can be assembled from competitively acquired subsystems rather than built directly from the raw material of lines of code.

We elaborate slightly some of the elements of the common service environment:

- *Tools, Libraries, and Databases:* There already exist major, complex software systems that provide implementations for portions of the national challenge applications. For example, large collections of computer-aided design (CAD) software are already used extensively in engineering design domains. Similarly, relational and object-oriented database management systems provide extensive capabilities for structured data storage, indexing, and management. Diverse sets of software tools and subsystems can be integrated into coherent applications development environments to form the development base with which to assemble the national challenge applications. Similarly, diverse libraries of program components and databases of data elements can be composed and integrated into the development environment.
- *Composition and Integration Frameworks:* Toolkits already exist in certain specific domains to assist in the composition and integration of tools, libraries, and databases. For example, the CAD

Framework Initiative (CFI) accomplishes this by providing interface specifications for tool-to-tool communications and tool-to-database communications. In addition, the CFI has developed prototype implementations of these capabilities. These can form the basis of value-aided and commercially supported packages and software toolsets. Commercial vendors of applications software for desktop computers are developing a variety of frameworks (such as CORBA, OLE, OpenDoc, and others) for integration of software applications. Users expect that commercial pressures will eventually result in some degree of integration of these various frameworks. This issue of multiple standards is discussed further below.

- *Building Block Object Sets:* The commonality that characterizes many of the service needs of the national challenge applications naturally yields an evolving shared market of software objects (that is, actions, operations, and protocols as well as data structures) to emerge that can be reused across multiple application development efforts. For example, a schedule object, which provides operations for allocating limited resources to critical tasks, could be used as a component of several different applications.
- *Application Customized Objects:* Leveraging the evolving building block object sets, we expect the objects from which the applications are implemented to be customized and extended for the application at hand. For example, though there is much in common in terms of command, control, communications, and intelligence (C<sup>3</sup>I) for an intensive care unit and an environmental spill, we would expect the details of sensor integration, strategies for alerting, and demands for real-time response to be somewhat different. The elements of the underlying object base will need customization for their use in specific national challenge applications. The degree of commonality across applications, which we hope is large, remains to be discovered.

### Considerations in Constructing the National Information Infrastructure

**Common architectural elements.** The national challenge applications obtain service capabilities delivered through common protocols or interfaces (known commercially as APIs, or applications portability interfaces). Though service capabilities may evolve rapidly, to the benefit of users, they are delivered through particular interfaces or protocols that evolve more slowly. This insulates the client architecturally from the rapid pace of change in implementations, on the one hand, but it enables the client to exploit new capabilities as soon as they appear, as long as they are delivered through the accepted interface. A competitive supply of services hastens the processes of convergence to common protocols and evolution therefrom.

**Industry standards, stovepipes, and risk.** We have asserted that commonality among the protocols, interfaces, and data representations used in the services layer of the NII will be critical for its success. To the extent that emerging or evolving industry-standard commonalities are replaced by ad hoc or proprietary stovepipe approaches for the national challenge areas, applications developers place themselves at risk with respect to delivery of capability and future evolution path. In particular, in return for complete ownership or control of a solution, they may give up the opportunity to ride the curves of growth in rapidly growing underlying technologies, such as multimedia, digital libraries, and data communication. The challenge of the national challenge applications is how the applications constituencies can have both control of applications solutions and participation in the rapid evolution of underlying technologies. Government, supported by research, can invest in accelerating the emergence of new common architectural elements, and in creating technologies that reduce the risk and commitment associated with adoption of rapidly evolving standards.

**Evolution of commonalities.** Accepted protocols naturally manifest a certain stickiness independent of their merit, because they become a stable element in determining systems structure and

develop associated transition costs and risks. The history of TCP/IP and OSI is a good example of this well-known phenomenon, as is the recent introduction of de facto standards relating to the World Wide Web (URLs and HTML). In particular, research and government can take a leading role in establishing new commonalities that foreshadow industry standards.

**Rapid evolution and multiple standards.** There are numerous standards presently in use for image representation. Most, but not all, are open standards; several are proprietary or otherwise encumbered. Regardless of the degree of acceptance of any one these standards, the pace of change is such that it would be foolish for a major software application developer to lock itself into accepting or producing images according to just one of these standards. Indeed, most major software applications building-blocks accept multiple such standards, thus increasing the robustness of the client applications with respect to either the technical characteristics or market acceptance of any one of the particular standards for bitmaps. In addition, tools are readily available for converting among the various representations for images. Thus, from the standpoint of applications architecture, a robust design can be created that does not depend on the fate of any one of the many standards, but rather on the evolution of the entire suite. The multiple commonalities emerge as customers and producers seek frameworks for competition in service niches. However, experience suggests that over time multiple related standards may begin to coalesce, as the commercial focus (and margins) move to higher levels of capability and the differential commercial advantage of any specific standard diminishes or even evolves into a liability. Anticipation of the process can yield robust scalable designs for major applications even when there is volatility in the markets for the subsystems they depend on.

**Competition and layering.** With the right approach to standards and infrastructural subsystems, diverse underlying technologies can evolve into common, shareable, and reusable services that can be leveraged across multiple NC applications. Alternative implementations of a frequently used service, such as display window management, eventually will lead to the identification of best practices that can be embodied in a common services layer—for example, for human interfaces. And robust designs of the applications layers above will enable this rapid evolution to be accepted and indeed exploited. (Consider, for example, the rapid rate of release of new versions of World Wide Web browsers, and the huge multiplicity of platforms they run on, and the rapid rate of evolution of the many multimedia (and other) standards they rely on. The Web itself, however, evolves at a slower rate and is not invalidated by these changes in particular niche services. The standards on which the Web is based evolve even more slowly.) The conclusion we draw is that simultaneous evolution at multiple layers is not only possible but also needs to be an explicit architectural goal if ubiquity is to be attained at the applications level.

**Concerning layers.** Services depend on other services for their realization. For example, a protocol for microtransactions will likely rely on other protocols for encryption and authentication. This enables a microtransaction system not only to be designed independently of the particular encryption and authentication services, but also to sustain later upgrade of (or recompetition for) those services in a robust manner. In spite of this dependency, services are not organized rigidly into layers as is, for example, the seven-layer OSI model. The term "layering" is instead meant to suggest that services naturally depend on other services. But the exact interdependency can change and evolve over time. The commonalities through which services are delivered thus form a set of multiple bottlenecks in a complex and undulating hourglass (using the analogy of CSTB, 1994).

**Service classification.** A consequence of the above argumentation is that the success of the overall NII does not depend on achievement of a particular master infrastructural architecture. But it must be emphasized that it does strongly depend on emergence of a broad variety of infrastructural service architectures designed with scale-up, and indeed ubiquity, in mind. Ubiquity (as suggested in the comments

above on multiple standards) is in the appearance of representatives of a set of related commonalities, and not in any particular protocol or component. This also suggests there is no ultimately correct layering lurking in the soup of services, but rather multiple candidates and arrangements. Without commonalities there is no national information infrastructure, but the particular need for specific all-encompassing commonalities is mitigated to the extent that technologies and tools for interoperability are available. That is, suites of related evolving commonalities can be supported to the extent that conversion and interoperability tools are available. The issue devolves into finding the right balance in this equation.

The government thus can employ a mixed strategy in fostering national challenge applications through infrastructural commonalities. It can stimulate development of new services, creation and evolution of new architectural commonalities, and development of readily available technologies of interoperability. Direct research and development is the most effective way to stimulate new service capabilities and associated commonalities. The government can also exploit its own market presence (though the leverage is less), taking an activist role in industry forums for conventionalization (informal emergent commonalities) and standards (formalized commonalities).

**An illustrative layered model.** One possible service taxonomy, elaborated below, classifies generic services into categories: human interfaces, applications building blocks, data and process management, and communications. Human interface services include window managers (e.g., Motif, NextStep), tools for speech handling and integration (generation as well as recognition), handwriting recognition, data visualization packages, toolkits for audio and video integration, and so on. Applications building blocks include planning packages, scheduling packages, data fusion, collaboration support, virtual reality support, and image processing and analysis. Data and process management services consist of capabilities for configuration management, shared data spaces, process flows, data integration, data exchange and translation, and data search and retrieval. Communications services include ubiquitous access through various communications mechanisms (e.g., wireless as well as wired connections into the bitways), mobility services to support users as they move through the points of connection into the network, interprocess communications and remote process call mechanisms to support distributed processing, and trust mechanisms such as authentication, authorization, encryption, password, and usage metering.

The service layers themselves evolve as new underlying technologies appear that provide new functionality or better ways of doing things. A construction kit can support the assembly and evolution of applications based on the service suite. Elements of this kit, also elaborated below, could include software environments for developing applications, evolution of standard operational and data exchange protocols, software toolkits and software generators for building or generating well-defined portions of applications, and frameworks for integrating tools and data into coherent, interoperable ensembles.

The value of a common services layer is conceptually indicated by [Figure 2](#). In [Figure 2\(a\)](#), the lack of a common services infrastructure leads to stovepipe implementations, with little commonality among the service capabilities of the various national challenges. In [Figure 2\(b\)](#), a common set of services is leveraged among the national challenges, aided by a collection of toolkits, integration frameworks, and applications generators.

### Information Enterprise Elements

Commonalities usually (but not always) emerge in the presence of a diversity of evolving implementations. A commonality in the form of a protocol is an abstraction away from the details of implementation that allows utility or value to be delivered in an implementation-independent manner to the service client. This suggests a threefold analysis for service capabilities: utility of some kind, delivered through a particular commonality such as a protocol, abstracting away the details of the diversity of implementations. Of course, the commonalities themselves evolve; they just evolve more slowly.

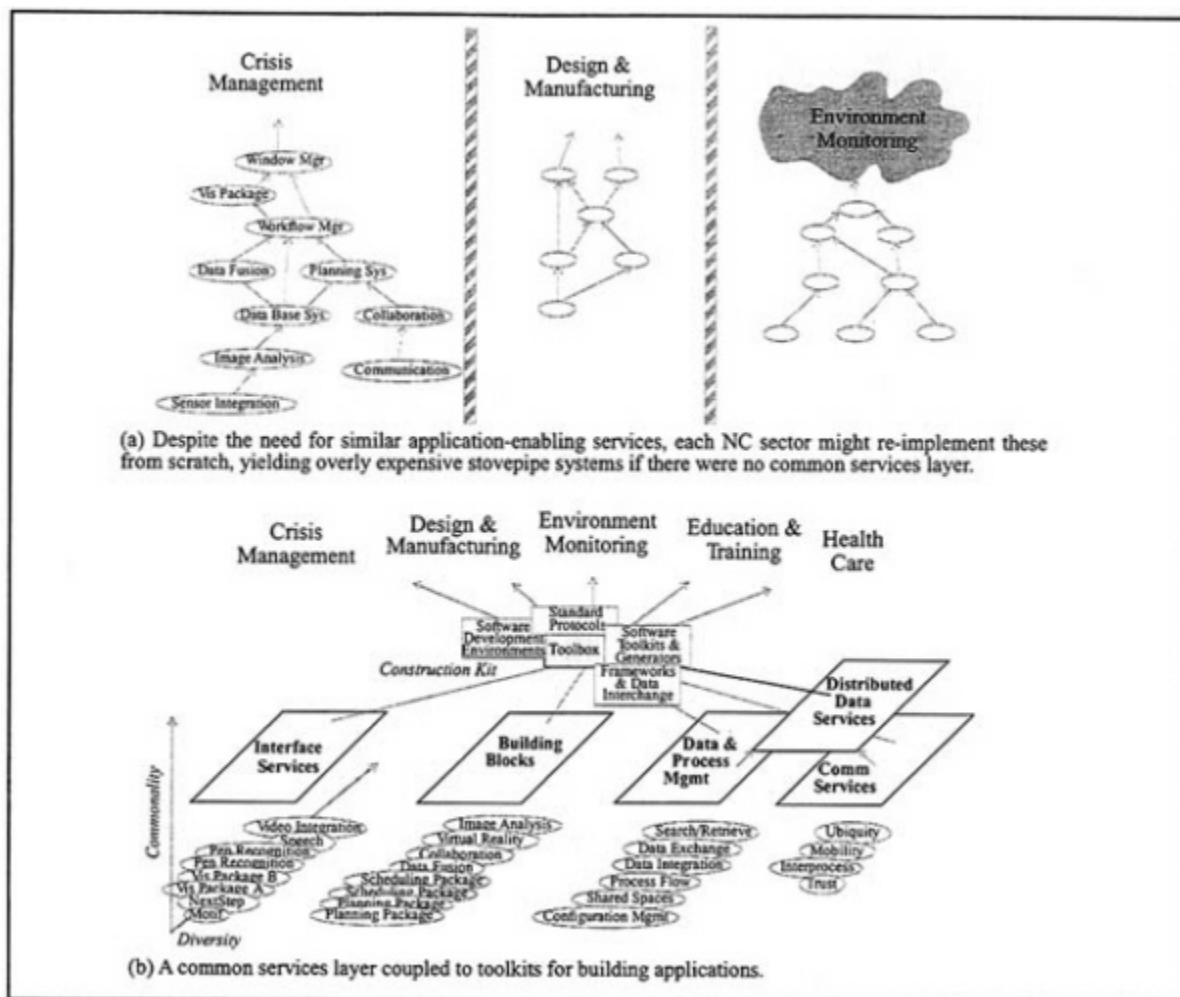


Figure 2 Technical challenges in building a national information infrastructure.

Figure 3 shows examples of elements for each of the three layers of the national information infrastructure architecture. In the figure, the three columns indicate the following:

- *Utility:* Each service provides specific value to users or clients. For example, the bitways are intended to provide ubiquitous data communications, and in a manner such that designers of applications need not know whether the communications links are through fiber, wireless, or some combination of links. The client needs only an abstract rendering of the characteristics of the aggregate link.
- *Commonality:* A common protocol or API creates a framework for delivery of utility. Clients engineer to this framework (and its expected evolution), thereby insulating themselves from the underlying implementation details. Diversification of technology occurs behind the protocol, enabling the technologies to be made accessible to clients with acceptable risk and cost-effectiveness, and also lowering entry barriers both for new users of the technologies and for new sources of capabilities. This is the essence of the principle of open architecture. For example, transport protocols for bitways provide users of communications services with a means to access the service independently of the particular choices of underlying component technologies.
- *Diversity:* These are the areas of implementation technology where innovation, rapid technological growth, and diversity of supply are essential to cost-effective delivery of increasing levels of

capability. For example, a competitive supply of fiber optic connectivity is needed to provide ubiquitous access to high performance bitways. Also, continued improvements in optical and wireless communication improve affordability of high-performance mobile communication.

Figure 3 shows examples of these concepts for each of the layers of the NII conceptual architecture. This organization focuses attention on two critical issues, alluded to in the foregoing, that must be addressed in the design of service commonalities:

- *Scalability*: Testbeds and other mechanisms provide means to assess the degree of scalability of new service concepts and protocols. They also address the extent of dependencies among services. Scalability, for infrastructure, necessarily includes potential for pervasive acceptance. Protocols that are proprietary or encumbered in other ways have a lesser chance of being accepted, because of the degree of technological and programmatic risk associated with them. But, as always, there is commercial advantage to being the first to introduce a successful open protocol, so the incentive persists for commercial introduction of commonalities, even when they are fully open.
- *Legacy*: There are two aspects of the legacy issue, a constraint and a goal. The first is the legacy we inherit, which constrains our architectural design decisions in fundamental ways. The second is the legacy we bequeath in the form of commonalities from which later architectures must evolve.

Opportunities for competition are naturally sought by service clients, and a diversity of implementations indicates success in this regard. At the level of bitways, for example, the space of change is rapid, and there are wide-ranging approaches for achieving a given capability (e.g., physical media may consist of optical fiber, land mobile wireless radios, or laser communications). The challenge for the application developer is how to exploit the continuing innovation while remaining insulated from continuous change; the client wants to ride the curves of growth while avoiding continual reengineering.

One conclusion to draw from this analysis is that research must focus not only on creation and demonstration of new kinds of service capability, but also on the scientific and technological aspects of architectural design: designing and evaluating candidates for protocol and API definitions, looking at both the supplier and client perspectives.

## THE FEDERAL HPCC PROGRAM AND THE NII

### Overview

In FY1994, the federal HPCC program was extended with a new responsibility, to develop Information Infrastructure Technology and Applications (IITA) to demonstrate prototype solutions to selected national challenge applications using the full potential of the rapidly evolving high performance communications and information processing capabilities. The details of the programs evolving goals and research plans are in its annual reports to Congress (FCCSET, 1994; CIC, 1994).

With the incorporation of IITA within its research agenda, the HPCC program is advancing key NII-enabling technologies, such as intelligent system interfaces, real environments augmented with synthetic environments, image understanding, language and speech understanding, intelligent agents aiding humans in the loop, and next-generation data and object bases for electronic libraries and commerce. This is being coupled with a vigorous program of testbed experimentation that will ensure continued U.S. leadership in information processing technologies.

IITA efforts are designed to strengthen the HPCC technology base, broaden the markets for these technologies, and accelerate industry development of the NII. Federal HPCC agencies are working closely with industry and academia in pursuit of these objectives. These objectives are to be accomplished, in part,

	<b>Diversity</b> What technologies require rapid growth, diverse supply	<b>Commonality</b> How capabilities are presented to client systems	<b>Utility</b> What is the value provided to the end user
<b>Applications</b> Capability to users	<ul style="list-style-type: none"> <li>• Universal computing and communication</li> <li>• Decision support and heterogeneous databases</li> <li>• Modeling and simulation</li> <li>• Sensors, teleoperation, advanced displays</li> </ul>	<ul style="list-style-type: none"> <li>• Manufacturing—product and process representations</li> <li>• Education—electronic library protocol</li> <li>• Health care—patient record model</li> </ul>	<ul style="list-style-type: none"> <li>• New ways to do business in critical national applications, leading to fundamental improvements</li> </ul>
<b>Services</b> Information service foundations for applications	<ul style="list-style-type: none"> <li>• Image reprn/compression</li> <li>• Security technologies</li> <li>• Virtual reality, simulation</li> <li>• Systems engineering tools</li> </ul>	<ul style="list-style-type: none"> <li>• Multi-media protocols</li> <li>• Security protocols</li> <li>• Engineering frameworks</li> <li>• Heterogeneous databases</li> </ul>	<ul style="list-style-type: none"> <li>• Information processing and management to enable broad range of applications</li> </ul>
<b>Bitways</b> Communications channels among services	<ul style="list-style-type: none"> <li>• Fiber Optic</li> <li>• Laser Space</li> <li>• Wireless</li> <li>• ISDN</li> </ul>	<ul style="list-style-type: none"> <li>• Transport protocols</li> <li>• Transmission standards</li> <li>• Wireless to wireline interface protocols</li> </ul>	<ul style="list-style-type: none"> <li>• Ubiquitous access to information</li> </ul>

Figure 3 Examples to illustrate the concepts of diversity, commonality, and utility.

by accelerating the development of readily accessible, widely used, large-scale applications with significant economic and social benefit. The HPCC program's original focus of enhancing computing and communications capabilities is thus extended to address a broader set of technologies and applications that have an immediate and direct impact on critical information capabilities affecting every citizen.

As we have described in the previous section, the development of such applications is predicated on (1) creating the underlying scalable computing technologies for advanced communication services over diverse bitways, effective partitioning of applications across elements of the infrastructure, and other applications support services that can adapt to the capabilities of the available infrastructure; and (2) creating and inserting a richly structured and intelligent service layer that will significantly broaden the base of computer information providers, developers, and consumers while reducing the existing barriers to accessing, developing, and using advanced computer services and applications. In parallel with these activities, a more effective software development paradigm and technology base must also be developed, since full-scale implementations in support of the national challenges will be among the largest and most complex applications ever implemented. This will be founded on the principles of composition and assembly rather than construction, solid architectures rather than ad hoc styles, and more direct user involvement in all stages of the software life cycle. The entire technology base developed in this program, including services and software, will be leveraged across the national challenges, leading to significant economies of scale in the development costs.

The intended technical developments of IITA include the following:

- *Information Infrastructure Services:* These are the collection of services provided to applications developers and end users that implement a layered architecture of increasing levels of intelligence and sophistication on top of the communications bitways. Services provide a universally available, network-aware, adaptive interface on which to construct the national challenge applications, spanning communications-based services at the low end, to intelligent information processing services at the high end. These services include network support for ubiquitous access, resource discovery in a



complex distributed network environment, and intelligent support services that can negotiate and adapt to the service quality needs of the application. Information infrastructure services also include system software and services that implement pervasive privacy, security and trust mechanisms for the information infrastructure, persistent object bases with which to build large-scale data repositories, reliable computing technologies to support the mission-critical nature of the infrastructure, and defensive software organized to protect the infrastructure from intrusion and attack.

- *Systems Development and Support Environments:* This area consists of the enabling technologies to develop and support large, complex information systems that exploit a national-scale information infrastructure. Fundamental to this activity is the use of that infrastructure in the software development and support process. Virtual organizations consisting of end users, contractors, and management will synergistically work together to develop software systems that are easy to use, that can be adapted through use to fit human needs and changing requirements, and that enhance end-user productivity, all despite the complexity of the underlying infrastructure. To achieve these goals, the focus is on software architectures, component prototyping, software composition, libraries of reusable and reliable software modules, end-user tailoring, intelligent documentation and online help, machine learning, and scalable compiler and interpreter technology.
- *Intelligent Interfaces:* Many of the national challenge applications require complex interfacing with humans or intelligent control systems and sensors. In addition, these applications must be able to understand their environment and to react to them. Technology in this area consists of high-level, network-capable applications building blocks for real-time planning and control, image processing and understanding, human language technology, extensive use of intelligent computer-based agents, and support technologies for more effective human-computer interaction.
- *National Challenges:* The concept of national challenge applications has already been described above. It is important to distinguish between the implementation of operational systems and the use of challenging applications testbeds to demonstrate the value of high-performance technologies as well as to drive their continued evolution. The government's research and development role is to focus on the latter; the private sector has primary responsibility for the former.

Each of the three technology areas (the first three bullets above) is discussed in additional detail in the following subsections, which include a sampling of technical subtopics. The national challenges have already been summarized in a prior section.

### Information Infrastructure Services

Services provide the underlying building blocks upon which the national challenge applications can be constructed. They are intended to form the basis of a ubiquitous information web usable by all. A rich array of interdependent services bridge the gap between the communications bitways and the application-specific software components that implement the national challenges.

- *Universal Network Services:* These are extensions to the existing Internet technology base to provide more widespread use by a much larger number of users. These include techniques for improved ease of use, plug-and-play network interoperability, remote maintenance, exploitation of new last mile technologies, management of hybrid/asymmetric network bandwidth, guaranteed quality of service for continuous media streams, and scale-up of network capabilities to dramatically larger numbers of users.
- *Integration and Translation Services:* These services support the migration of existing data files, databases, libraries, and programs to new, better integrated models of computing, such as object-oriented systems. They also provide mechanisms to support continued access to older legacy forms of data as the models evolve. Included are services for data format translation and interchange as well as

tools to translate the access portions of existing programs. Techniques include wrappers that surround existing elements with new interfaces; integration frameworks that define application-specific common interfaces and data formats; and mediators that extend generic translation capabilities with domain knowledge-based computations, permitting abstraction and fusion of data.

- *System Software Services:* These include operating system services to support complex, distributed, and time- and bandwidth-sensitive applications. The services support the distribution of processing across processing nodes within the network, the partitioning of the application logic among heterogeneous nodes based on their specialized capabilities or considerations of asymmetric or limited-interconnection bandwidth; guaranteed real-time response to applications for continuous media streams; and storage, retrieval, and I/O capabilities suitable for delivering large volumes of data to great numbers of users. Techniques include persistent storage, programming language support, and file systems.
- *Data and Knowledge Management Services:* These services include extensions to existing database management technology for combining knowledge and expertise with data. These include methods for tracking the ways in which information has been transformed. Techniques include distributed databases; mechanisms for search, discovery, dissemination, and interchange; aggregating base data and programmed methods into objects; and support for persistent object stores incorporating data, rules, multimedia, and computation.
- *Information Security Services:* These services provide support for the protection of the security of information, enhanced privacy and confidentiality for users of the infrastructure, protection of intellectual property rights, and authentication of information sources within the infrastructure. Techniques include privacy-enhanced mail, methods of encryption and key-escrow, and digital signatures. Also included are techniques for protecting the infrastructure (including authorization mechanisms and firewalls) against intrusion attacks, such as worms, viruses, and Trojan horses.
- *Reliable Computing and Communications Services:* These include system software services for nonstop, highly reliable computer and communications systems that can operate without interruption. The techniques include mechanisms for fast system restart such as process shadowing, reliable distributed transaction commit protocols, and event and data redo logging to keep data consistent and up-to-date in the face of system failures.

### System Development and Support Environments

These provide the network-based software development tools and environments needed to build the advanced user interfaces and the information-intensive NC applications.

- *Rapid System Prototyping:* These consist of the tools and methods that enable the incremental integration and cost effective evolution of software systems. Technologies include tools and languages that facilitate end-user specification, architecture design and analysis, component reuse and prototyping; testing and online configuration management tools; and tools to support the integration and interoperation of heterogeneous software systems.
- *Distributed Simulation and Synthetic Environments:* These software development environments provide the specialized underlying support mechanisms for the creation of synthetic worlds, which can integrate real as well as virtual objects, in terms of both their visual as well as computational descriptions. Methods include distributed simulation algorithms; geometric models and data structures; tools for scene description, creation, and animation; and integration of geometric and computational models of behavior into an integrated system description.
- *Problem Solving and System Design Environments:* These environments provide the techniques that support the software and system design process through the use of automated tools, with particular emphasis on maintaining flexibility and tailorability in tool configurations to enable organizations to tailor

their support environments to their needs. Examples include efficient algorithms for searching huge planning spaces, more powerful and expressive representations of plans, operators, goals, and constraints, and the incorporation of efficient methods to facilitate scheduling and resource allocation. The effects of uncertainty must be taken into account as well as the effects of goal interactions.

- *Software Libraries and Composition Support:* These software tools and methods support the development of common architectures and interfaces to increase the potential for reusability across multiple underlying models of computation, the diversity of programming languages in use, and the varying degree of assurance provided by software components. Important elements of this area include the development of the underlying methodology, data structures, data distribution concepts, operating system interfaces, synchronization features, language extensions, and other technology to enable the construction of scalable library frameworks.
- *Collaboration and Group Software:* These tools provide support for group cooperative work environments that span time as well as space. Methods include shared writing surfaces and live boards, version and configuration management, support for process and task management, capture of design history and rationale, electronic multimedia design notebooks, network-based video conferencing support, document exchange, and agents serving as intermediaries to repositories of relevant multimedia information. The technology should be developed to make it possible to join conferences in progress and to be automatically brought up to date by assistants (agents) with memory.

### Intelligent Interfaces

Advanced user interfaces will bridge the gap between human users and the emerging national information infrastructure. A wide range of new technologies that adapt to human senses and abilities must be developed to provide more effective human-machine communications. The IITA program must achieve a high level user interface to satisfy the many different needs and preferences of vast numbers of citizens who interact with the NII.

- *Human-Computer Interface:* This supports research in a broad range of technologies and their integration to allow humans and computers to interact effectively, efficiently, and naturally. Developments in this area include technologies for speech recognition and generation; graphical user interfaces that allow rapid browsing of large quantities of data; user-sensitive interfaces that customize and present information for particular levels of understanding; language corpora for experimental research; and human-machine interaction via touch, facial expression, gesture, and so on. The new IITA emphasis is on integration, real-time performance, and demonstration of these new communication modalities in multimedia, multisensory environments.
- *Heterogeneous Database Interfaces:* This supports development of methodologies to integrate heterogeneously structured databases composed of multiformatted data. To support NII information dissemination, a capability is needed for a user to issue a query which is broadcast to the appropriate databases and a timely response is returned and translated into the context of the users query. Multiformatted data may range from ASCII text to numerical time series, to multidimensional measurements, to time series of digital imagery, etc. Also of critical importance is the integration of metadata with the data and its accessibility across heterogeneous databases.
- *Image Processing and Computer Vision:* This activity supports research in making images, graphics, and other visual information a more useful modality of human-computer communication. Research areas include all aspects of theory, models, algorithms, architectures, and experimental systems from low-level image processing to high-level computer vision. Methodologies of pattern recognition will be further developed to allow automated extraction of information from large databases, in particular,

digital image data. The new IITA emphasis is on integration, scalability, and demonstration of easy access and usability of visual information in real-world problems.

- *User-centered Design Tools/Systems*: This consists of work in models and methodologies leading to interactive tools and software systems for design and other user-centered activities. User-friendly tools that combine data-driven and knowledge-based capabilities is one of the areas for new research. The new IITA emphasis is on supporting the development of ubiquitous, easy-to-use, and highly effective interactive tools.
- *Virtual Reality and Telepresence*: This consists of research that will provide tools and methods for creating synthetic (virtual) environments to allow real-time, interactive human participation in the computing/communication loop. Such interaction may be through sensors, effectors, and other computational resources. The IITA focus is creating shared virtual environments which can be accessed and manipulated by many users at a distance in support of national challenge application areas.

### SUMMARY AND CONCLUSIONS

Much of the discussion of the national information infrastructure has been at the applications level or the level of the bitways. Various groups, including Congress and the Clinton administration, have identified candidate NC applications on the one hand, while others have dealt with the issues of making interoperable the various existing and emerging communications infrastructures. This discussion suggests a shift in focus to the services layer. The right collection of capabilities at this level of the infrastructure will have an extraordinary impact on a wide range of applications.

We have cataloged many of the key technology areas needed for the service layer of the NII: information infrastructure services, systems development and support environments, and intelligent interfaces. The further development of these technologies and their integration into coherent and robust service architectures, incorporating the principles of utility, diversity, and commonality as described here, will be a major challenge for the information technology research community in coming years.

Cost-shared sponsorship of pilot demonstrations and testbeds is a key role for government in accelerating the development of the NII. In each NC application area, opportunities exist to demonstrate early solutions, including the potential for scaling up. We suggest that in the exploration of commonality and conversion issues, testbeds can also help address the fundamental issue of ubiquity. The scale of the enterprise, and the fundamental opportunities being addressed, necessitate cooperation among industry, government, and academia for success. We have suggested appropriate roles and approaches to cooperation, with emphasis on the roles of government and research. This is predicated on the assumption that government, in addition to sponsoring key basic research, has a crucial catalytic role in working with all sectors to address the challenge of the national applications to scaling up to the point of ubiquity and reliance.

### ACKNOWLEDGMENTS

The ideas expressed in this paper have been influenced by discussions with colleagues at DARPA, especially Duane Adams, Steve Cross, Howard Frank, Paul Mockapetris, Michael St. Johns, John Toole, Doyle Weishar, and Gio Wiederhold. Our ideas have also benefited from extensive discussions with participants in the HPCC program from a diverse collection of federal agencies: Howard Bloom (NIST), Roger Callahan (NSA), Y.T. Chien (NSF), Mel Ciment (NSF), Sherri de Coronado (NIH), Ernest Daddio (NOAA), Norm Glick (NSA), Steve Griffin (NSF), Dan Hitchcock (DOE), Paul Hunter (NASA), Jerry Linn (NIST), Dan Masys (NIH), Cherie Nichols (NIH), Walter Shackelford (EPA), and Selden Stewart (NIST).

## REFERENCES

- Clinton, William J., and Albert Gore, Jr. 1993. *Technology for America's Economic Growth: A New Direction to Build Economic Strength*, February 22.
- Committee on Information and Communication (CIC). 1994. *High Performance Computing and Communications: Technology for the National Information Infrastructure, Supplement to the President's Fiscal Year 1995 Budget*. National Science and Technology Council, Washington, D.C.
- Computer Science and Telecommunications Board (CSTB), National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C.
- Computer Systems Policy Project (CSPP). 1993. *Perspectives on the National Information Infrastructure: CSPP's Vision and Recommendations for Action*. Computer Systems Policy Project, Washington, D.C., January 12.
- Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), Office of Science and Technology Policy. 1993. *FCCSET Initiatives in the FY 1994 Budget*. Office of Science and Technology Policy, Washington, D.C., April 8.
- Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), Office of Science and Technology Policy. 1994. *High Performance Computing and Communications: Toward a National Information Infrastructure*. Committee on Physical, Mathematical, and Engineering Sciences, Office of Science and Technology Policy, Washington, D.C.
- Gore, Jr., Al. 1991. "Infrastructure for the Global Village," *Scientific American* 265(3):150-153.
- Gore, Jr., Albert. 1993. *From Red Tape to Results, Creating a Government That Works Better & Costs Less: Reengineering Through Information Technology*, Accompanying Report of the National Performance Review. U.S. Government Printing Office, Washington, D.C., September.
- Information Infrastructure Task Force (IITF). 1993. *The National Information Infrastructure: Agenda for Action*. Information Infrastructure Task Force, U.S. Department of Commerce, Washington, D.C., September 15.
- Information Infrastructure Task Force (IITF), Committee on Applications and Technology. 1994. *Putting the Information Infrastructure to Work*. NIST Special Document No. 857. Information Infrastructure Task Force, U.S. Department of Commerce, May.
- Information Technology Association of America (IITA). 1993. *National Information Infrastructure: Industry and Government Roles*. Information Technology Association of America, Washington, D.C., July.
- Institute for Information Studies (IIS). 1992. *A National Information Network: Changing Our Lives in the 21st Century*. Annual Review of the Institute for Information Studies (Northern Telecom Inc. and the Aspen Institute), Queenstown, Md.
- Kahin, Brian. 1993. "Information Technology and Information Infrastructure," in *Empowering Technology: Implementing a U.S. Strategy*, Lewis M. Branscomb (ed.). MIT Press, Cambridge, Mass.
- Motiwala, J., M. Yap, and L.H. Ngoh. 1993. "Building the Intelligent Island," *IEEE Communications Magazine* 31(10):28-34.
- National Computer Board of Singapore (NCBS). 1992. "A Vision of an Intelligent Island: The IT2000 Report," March.
- Vernon, Mary K., Edward D. Lazowska, and Stewart D. Personick (eds.). 1994. *R&D for the NII: Technical Challenges*. Report of a workshop held February 28 and March 1, 1994, in Gaithersburg, Md. EDUCOM, Washington, D.C.

## 40

# Nomadic Computing and Communications

*Leonard Kleinrock*

*University of California at Los Angeles*

### ABSTRACT

Nomadcity refers to the system support needed to provide a rich set of computing and communications capabilities and services, in a transparent and convenient form, to the nomad moving from place to place. This new paradigm is already manifesting itself as users travel to many different locations with laptops, personal digital assistants, cellular telephones, pagers, and so on. In this paper we discuss some of the open issues that must be addressed in the system support necessary for nomadcity. In addition, we present some additional considerations in the area of wireless communications, which forms one (and only one) component of nomadcity.

### INTRODUCTION

Currently, most users think of computers as associated with their desktop appliances or with a server located in a dungeon in some mysterious basement. However, many of those same users may be considered to be nomads, in that they own computers and communication devices that they carry about with them in their travels as they move between office, home, airplane, hotel, automobile, branch office, and so on. Moreover, even without portable computers or communications, there are many who travel to numerous locations in their business and personal lives and who require access to computers and communications when they arrive at their destinations. Indeed, even a move from a desk to a conference table in the same office constitutes a nomadic move since the computing platforms and communications capability may be considerably different at the two locations. The variety of portable computers is impressive, ranging from laptop computers, notebook computers, and personal digital assistants (or personal information managers) to "smart" credit card devices and wristwatch computers. In addition, the communication capability of these portable computers is advancing at a dramatic pace—from high-speed modems to PCMCIA modems, e-mail receivers on a card, spread-spectrum hand-held radios, CDPD transceivers, portable GPS receivers, and gigabit satellite access, and so on.

The combination of portable computing with portable communications is changing the way we think about information processing (Weiser, 1991). We now recognize that access to computing and communications is necessary not only from "home base" but also while in transit and after reaching a destination.<sup>1</sup>

These ideas form the essence of a major shift to nomadcity (nomadic computing and communications), which we address in this paper. The focus is on the system support needed to provide a rich set of capabilities and services, in a transparent and convenient form, to the nomad moving from place to place.

---

NOTE: This work was supported by the Advanced Research Projects Agency, ARPA/CSTO, under Contract J-FBI-93-112, "Computer Aided Design of High Performance Network Wireless Networked Systems," and by the Advanced Research Projects Agency, ARPA/CSTO, under Contract DABT-63-C-0080, "Transparent Virtual Mobile Environment."

This paper contains material similar to that published by the author in a paper where the emphasis was on the research issues to be addressed in nomadic computing and communications. Copyright © National Academy of Sciences. All rights reserved.

## NOMADIC COMPUTING<sup>2</sup>

We are interested in those capabilities that must be put in place to support nomadicity. The desirable characteristics for nomadicity include independence of location, motion, computing platform, communication device, and communication bandwidth, and widespread presence of access to remote files, systems, and services. The notion of independence does not refer here to the quality of service, but rather to the perception of a computing environment that automatically adjusts to the processing, communications, and access available at the moment. For example, the bandwidth for moving data between a user and a remote server could easily vary from a few bits per second (in a noisy wireless environment) to hundreds of megabits per second (in a hard-wired ATM environment); or the computing platform available to the user could vary from a low-powered personal digital assistant while traveling to a powerful supercomputer in a science laboratory. Indeed, today's systems treat radically changing connectivity or bandwidth/latency values as exceptions or failures; in the nomadic environment, these must be treated as the usual case. Moreover, the ability to accept partial or incomplete results is an option that must be made available because of the uncertainties of the informatics infrastructure.

The ability to automatically adjust all aspects of the user's computing, communication, and storage functionality in a transparent and integrated fashion is the essence of a nomadic environment.

Some of the key system parameters of concern include bandwidth, latency, reliability, error rate, delay, storage, processing power, interference, version control, file synchronization, access to services, interoperability, and user interface. These are the usual concerns for any computer-communication environment, but what makes them of special interest for us is that the values of these parameters change dramatically as the nomad moves from location to location. In addition, some totally new and primary concerns arise for the nomad such as weight, size, and battery life of the portable devices as well as unpredictability and wide variation in the communication devices and channels. The bottom line consideration in many nomadic applications is, of course, cost.

Many of the key parameters above focus on the lower levels of the architecture, and they have received the most attention from industry and product development to date. This is natural since hardware devices must focus on such issues. However, there is an enormous effort that must be focused on the middleware services if nomadicity is to be achieved. We identify a number of such services below, but we must recognize that they are in the early stages of identification and development.

There are a number of reasons why nomadicity is of interest. For example, nomadicity is clearly a newly emerging technology that already surrounds the user. Indeed, this author judges it to be a paradigm shift in the way computing will be done in the future. Information technology trends are moving in this direction. Nomadic computing and communications is a multidisciplinary and multi-institutional effort. It has a huge potential for improved capability and convenience for the user. At the same time, it presents at least as huge a problem in interoperability at many levels. The contributions from any investigation of nomadicity will be mainly at the middleware level. The products that are beginning to roll out have a short-term focus; however, there is an enormous level of interest among vendors (from the computer manufacturers, the networking manufacturers, the carriers, and so on) for long-range development and product planning, much of which is now under way. Whatever work is accomplished now will certainly be of immediate practical use.

There are fundamental new research problems that arise in the development of a nomadic architecture and system. Let us consider a sampling of such problems, which we break out into systems issues and wireless networking issues.

### Systems Issues

One key problem is to develop a full system architecture and set of protocols for nomadicity. These should provide for a transparent view of the user's dynamically changing computing and communications environment. The protocols must satisfy the following kinds of requirements:

- Interoperation among many kinds of infrastructures (e.g., wireline and wireless);
- Ability to deal with unpredictability of user behavior, network capability, and computing platform;

- Ability to provide for graceful degradation;
- Ability to scale with respect to heterogeneity, address space, quality of service (QOS), bandwidth, geographical dimensions, number of users, and so on;
- Integrated access to services;
- Ad hoc access to services;
- Maximum independence between the network and the applications from both the user's viewpoint and the development viewpoint;
- Ability to match the nature of what is transmitted to the network bandwidth availability (i.e., compression, approximation, partial information, and so on);
- Cooperation among system elements such as sensors, actuators, devices, network, operating system, file system, middleware, services, applications, and so on; and
- Ability to locate users, devices, services, and the like.

In addition, the following components can help in meeting these requirements:

- An integrated software framework that presents a common virtual network layer;
- Appropriate replication services at various levels;
- File synchronization;
- Predictive caching;
- Consistency services;
- Adaptive database management;
- Location services (to find people and devices via tracking, forwarding, searching, etc.)—Mobile IP (Perkins, 1995) is an example of an emerging standard here;
- Discovery of resources; and
- Discovery of profile.

A second research problem is to develop a reference model for nomadicity that will allow for a consistent discussion of its attributes, features, and structure. This should be done in a way that characterizes the view of the system as seen by the user, and the view of the user as seen by the system. The dimensions of this reference model might include the following:

- System state consistency (i.e., Is the system consistent at the level of e-mail, files, database, applications, and so on?);
- Functionality (this could include the bandwidth of communications, the nature of the communication infrastructure, and the quality of service provided); and
- Locality, or awareness (i.e., How aware is the user of the local environment and its resources, and how aware is the environment of the users and their profiles?).

A third research problem is to develop mathematical models of the nomadic environment. These models will allow one to study the performance of the system under various workloads and system configurations as well as to develop design procedures.

As mentioned above, the area of nomadic computing and communications is multidisciplinary. Following is a list of the disciplines that contribute to this area (in top-down order):

- Advanced applications, such as multimedia or visualization;
- Database systems;
- File systems;
- Operating systems;
- Network systems;
- Wireless communications;
- Low-power, low-cost radio technology;



- Micro-electro-mechanical systems (MEMS) sensor technology;
- MEMS actuator technology; and
- Nanotechnology.

The reason that the last three items in this list are included is that we intend that the nomadic environment include the concept of an intelligent room. Such a room has embedded in its walls, furniture, floor, and other aspects all manner of sensors (to detect who and what is in the room), actuators, communicators, logic, cameras, etc. Indeed, one would hope to be able to speak to the room and say, for example, "I need some books on the subject of spread spectrum radios," and perhaps three books would reply. The replies would also offer to present the table of contents of each book, as well, perhaps, as the full text and graphics. Moreover, the books would identify where they are in the room, and, if such were the case, might add that one of the books is three doors down the hall in a colleague's office!

There are numerous other systems issues of interest that we have not addressed here. One of the primary issues is that of security, which involves privacy as well as authentication. Such matters are especially difficult in a nomadic environment, because the nomad often finds that the computing and communication devices are outside the careful security walls of his or her home organization. This basic lack of physical security exacerbates the problem of achieving nomadicity.

We have only touched on some of the systems issues relevant to nomadicity. Let us now discuss some of the wireless networking issues of nomadicity.

### Wireless Networking Issues

It is clear that a great many issues regarding nomadicity arise whether or not there is access to wireless communications. However, with such access, a number of interesting considerations arise.

Access to wireless communications provides two capabilities to the nomad: It allows for communication from various (fixed) locations without being connected directly into the wireline network, and it allows the nomad to communicate while traveling. Although the bandwidth offered by wireless communication media varies over as enormous a range as does the wireline network bandwidth, the nature of the error rate, fading behavior, interference level, and mobility issues for wireless are considerably different, so that the algorithms and protocols require some new and different forms from those of wireline networks (Katz, 1994). For example, the network algorithms to support wireless access are far more complex than for the wireline case; some of these are identified below. Whereas the location of a user or a device is a concern for wireline networks as described above, the details of tracking a user moving in a wireless environment add to the complexity and require rules for handover, roaming, and so on.

The cellular radio networks so prevalent today have an architecture that assumes the existence of a cell base station for each cell of the array; the base station controls the activity of its cell. The design considerations of such cellular networks are reasonably well understood and are being addressed by an entire industry (Padgett et al., 1995). We discuss these no further here.<sup>3</sup>

There is, however, another wireless networking architecture of interest that assumes no base stations (Jain et al., 1995; Short et al., 1995). Such wireless networks are useful for applications that require "instant" infrastructure, among others. For example, disaster relief, emergency operations, special military operations, and clandestine operations are all cases where no base station infrastructure can be assumed. In the case of no base stations, maintaining communications is considerably more difficult. For example, it may be that the destination for a given reception is not within range of the transmitter, and some form of relaying is therefore required; this is known as "multihop" communications. Moreover, since there are no fixed-location base stations, then the connectivity of the network is subject to considerable change as devices move around and/or as the medium change its characteristics. A number of new considerations arise in these situations, and new kinds of network algorithms are needed to deal with them.

To elaborate on some of the issues of concern if there are no base stations, we take three possible scenarios:

1. *Static topology with one-hop communications.* In this case, there is no motion among the system elements, and all transmitters can reach their destinations without any relays. The issues of concern, along with the needed network algorithms (shown in bold print), are as follows:

- Can you reach your destination?: **Power Control**
- What access method should you use?: **Network Access Control**
- Which channel (or code) should you use?: **Channel Assignment Control**
- Will you interfere with another transmission?: **Power and Medium Access Control**
- When do you allow a new "call" into the system?: **Admission Control**
- For different multiplexed streams, can you achieve the required QOS (e.g., bandwidth, loss, delay, delay jitter, higher-order statistics, etc.)?: **Multimedia Control**
- What packet size should you use?: **System Design**
- How are errors to be handled?: **Error Control**
- How do you handle congestion?: **Congestion Control**
- How do you adapt to failures?: **Degradation Control**

2. *Static topology with multihop communications.* Here the topology is static again, but transmitters may not be able to reach their destinations in one hop, and so multihop relay communications is necessary in some cases. The issues of concern, along with the needed network algorithms (shown in bold print), include all of the above plus the following:

- Is there a path to your destination?: **Path Control**
- Does giant stepping help (Takagi and Kleinrock, 1984)?: **Power Control**
- What routing procedure should you use?: **Routing Control**
- When should you reroute existing calls?: **Reconfiguration Control**
- How do you assign bandwidth and QOS along the path?: **Admission Control and Channel Assignment**

3. *Dynamic topology with multihop.* In this case, the devices (radios, users, etc.) are allowed to move, which cause the network connectivity to change dynamically. The issues of concern, along with the needed network algorithms (shown in bold print), include all of the above plus the following:

- Do you track, forward, or search for your destination?: **Location Control**
- What network reconfiguration strategy should you use?: **Adaptive Topology Control**
- How should you use reconfigurable and adaptive base stations?: **Adaptive Base Station Control**

These lists of considerations are not complete but are only illustrative of the many interesting research problems that present themselves in this environment. The net result of these considerations is that the typical 7-layer OSI model for networking must be modified to account for these new considerations. For example, we must ask what kind of network operating system (NOS) should be developed, along with other network functions (Short et al., 1995); what mobility modules must be introduced to support these new services; and so on.

This section addresses mainly the network algorithm issues and does not focus on the many other issues involved with radio design, hardware design, tools for CAD, system drivers, and so on. What is important is that the network algorithms must be supported by the underlying radio (e.g., to provide signal-to-interference ratios, ability to do power control, change codes in CDMA environments, and the like). These obviously have an impact on the functionality, structure, and convenience of the appliance that the user must carry around, as well as on its cost.

If we ask what are the great applications of wireless technology that affect the fabric of our society, then education applications stand out among the most significant. In this application, a wireless infrastructure could serve to provide connectivity in a cost-effective fashion to rural areas that are difficult to serve otherwise; it could

serve within a school to provide flexible sharing of devices as they move from location to location. For long-distance wireless access, it seems that direct broadcast satellite (DBS) technology would be great benefit, but it should also provide a decent up-channel as well. For in-building wireless access, the availability of unlicensed spectrum for data—say, the 60-GHz range—would serve a number of education applications nicely.

One might ask what role government could play in helping to bring about some of the advantages just described. The allocation of spectrum is one of the major ways in which government can assist. Currently, most spectrum is assigned for long periods of time to specific types of services; it seems that a more liberal view on the kinds of uses for radio bandwidth would encourage innovative applications, services, and efficient sharing of this bandwidth. Any action (such as spectrum allocation and use) that encourages the introduction of innovative services is to be encouraged by whatever means government has available.

### CONCLUSION

This paper presents nomadicity as a new paradigm in the use of computer and communications technology and outlines a number of challenging problems. It is clear that our existing physical and logical infrastructure must be extended to support nomadicity in the many ways described here. The implication is that we must account for nomadicity at this early stage in the development and deployment of the NII; failure to do so will seriously inhibit the growth of nomadic computing and communications. In addition to those issues we raise here, there are far more we have not yet identified. Those will arise only as we probe the frontiers of nomadic computing and communications.

### REFERENCES

- Jain, R., J. Short, L. Kleinrock, S. Nazareth, and J. Villaseñor. 1995. "PC-notebook Based Mobile Networking: Algorithms, Architectures and Implementations," *ICC '95*, pp. 771–777, June.
- Katz, R.H. 1994. "Adaptation and Mobility in Wireless Information Systems," *IEEE Personal Communications Magazine* 1(1):6–17.
- Kleinrock, L. 1995. "Nomadic Computing—An Opportunity," *Computer Communications Review, ACM SIGCOMM* 25(1):36–40.
- Nomadic Working Team (NWT). 1995. "Nomadicity in the NII," Cross-Industry Working Team, Corporation for National Research Initiatives, Reston, Va.
- Padgett, J.E., C.G. Gunther, and T. Hattori. 1995. "Overview of Wireless Personal Communications," *IEEE Communications Magazine* 33(1):28–41.
- Perkins, Charles. 1995. "IP Mobility Support," an Internet draft produced for the Internet Engineering Task Force; see <http://www.ietf.cnri.reston.va.us/ids.by.wg/mobil>.
- Short, J., R. Bagrodia, and L. Kleinrock. 1995. "Mobile Wireless Network System Simulation," *Proceedings of ACM Mobile Computing & Networking Conference (Mobicom '95)*, pp. 195–205, November.
- Takagi, H., and L. Kleinrock. 1984. "Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals," *IEEE Transactions on Communications*, Vol. COM-32, No. 3, pp. 246–257, March.
- Weiser, M. 1991. "The Computer for the 21st Century," *Scientific American*, September, pp. 94–104.

### NOTES

1. Moreover, one may have more than a single "home base"; in fact, there may be no well-defined "home base" at all.
2. Some of the ideas presented in this section were developed with two groups with which the author has collaborated in work on nomadic computing and communications. One of these is the Nomadic Working Team (NWT) of the Cross Industrial Working Team (XIWT); the author is the chairman of the NWT, and this working team recently published a white paper on nomadic computing (NWT, 1995). The second group is a set of his colleagues at the UCLA Computer Science Department who are working on an ARPA-supported effort known as TRAVLER, of which he is principal investigator.

3. Wireless LANs come in a variety of forms. Some of them are centrally controlled and therefore have some of the same control issues as cellular systems with base stations; others have distributed control, in which case they behave more like the no-base-station systems we discuss in this section.

## 41

# NII 2000: The Wireless Perspective

*Mary Madigan*

*Personal Communications Industry Association*

### ABSTRACT

As a key component of the national information infrastructure (NII), the mobile nature of wireless communications provides consumers with the opportunity to access the NII from any place at any time. Today, as the Federal Communications Commission (FCC) makes available new spectrum for wireless networks that will support a range of new services, both voice and data, wireless communications is poised on the brink of a new era.

However, new spectrum leads to new entrants, and wireless companies of the future will face a much more competitive marketplace. This competition will mean great things to the American consumer, who will benefit from the innovation and lower prices that the increased competitiveness will spark.

With the introduction of more competition into the telecommunications marketplace, public policy decisions need to be crafted to ensure that this vision of a wireless future can be realized.

### INTRODUCTION

For the wireless communications industry, 1994 was a banner year as the FCC launched the first set of spectrum auctions for the narrowband and broadband PCS, giving birth to a whole new era—the era of personal communications.

The vision of PCS is the concept of anytime, anywhere communications—whether that be data communications, voice communications, or both. But what is the real potential for this marketplace? How many individuals are likely to buy into the vision of anytime, anywhere communications?

In early 1995, the Personal Communications Industry Association (PCIA) completed a survey of PCIA members to evaluate the growth, composition, and characteristics of the existing and future personal communications industry and published the results in a PCS Market Demand Forecast. The results indicate that by 2000, combined demand for new PCS, cellular, and paging and narrowband PCS will amount to almost 118 million subscriptions.

To meet this level of demand in the marketplace, the wireless industry must be assured that it will be able to deploy services in a timely fashion. Issues such as site acquisition and interconnection to the local exchange carriers are critical to timely deployment of developing wireless networks and competing effectively. Government must assure that the industry has the opportunity to meet the anticipated demand outlined in the PCS Market Demand Forecast by ensuring a level playing field for all wireless telecommunications service providers and by allowing, where appropriate, competition—not regulation—to govern the marketplace.

### Personal Communications Industry Association

Established in 1949, PCIA has been instrumental in advancing regulatory policies, legislation, and technical standards that have helped launch the age of personal communications services. Through many vehicles—policy boards, market forecasts, publications, spectrum management programs, seminars, technician

certification programs, and its industry trade show, the Personal Communications Showcase—PCIA is committed to maintaining its position as the association for the PCS industry.

PCIA's member companies include PCS licensees and those involved in the cellular, paging, ESMR, SMR, mobile data, cable, computer, manufacturing, and local and interexchange sectors of the industry, as well as private corporate systems users, wireless system integrators, communication site owners, distributors and service professionals, and technicians.

### **Personal Communication Service**

Personal communication service includes a broad range of telecommunications services that enable people and devices to communicate independent of location. PCS networks and devices operate over a wide range of frequencies assigned and authorized by the FCC. There are currently seven different air interface technologies proposed for standardization for the new PCS licensees that will be operating in the 1.8 GHz band. Service providers that will be operating at these frequencies either are new entrants with no established network or are existing telecommunications service providers, such as cable, cellular, local exchange, and long-distance carriers. With the technology choices companies make over the next few months, there will need to be analysis of how and to what extent the various wireless and wireline networks will work together.

### **Interoperability and Interworking**

To facilitate roaming among PCS carriers, some degree of interoperability and interworking needs to be accomplished between the networks. PCIA defines interoperability and interworking as follows:

- *Interoperability.* The ability to logically connect two or more functional network elements for the purposes of supporting shared processes such as call delivery. Service interoperability is defined as the assurance that a service invoked by a subscriber in a network will be performed by the other network in the same way from a user perspective. Network interoperability is defined as the direct one-to-one mapping of services and protocols between interconnected networks. For example, a subscriber may invoke call waiting features exactly the same way in a DCS 1900 (GSM based) network in New York City as in a DCS 1900 (GSM based) network in San Francisco. In this scenario, call waiting network protocol messages map between the two networks on a direct one-to-one basis.
- *Interworking.* The ability to translate between two or more dissimilar networks for the purpose of achieving effective interoperability. Service interworking is defined as the protocol translation that may or may not result in the service being performed in the receiving network in the same way from a user perspective. Network interworking is defined as functional mapping of services and protocols across networks (some services may not be delivered or may be delivered in a different way). For example, a subscriber with a PCS 2000 (Composite CDMA/TDMA) wireless personal terminal may register and authenticate on a San Francisco IS-41 based network, just as he or she could on a home base, DCS 1900 (GSM based) network in New York City. Although the method of registering may not be identical between systems, the end result is effectively the same—the subscriber can be registered and authenticated on both networks, and location services work across both platforms.

Standards should be developed and are currently being worked on in domestic and international standards bodies to facilitate features and services delivered consistently and in similar fashions to an end user, regardless of the air interface and/or network implementation used. All networks do not necessarily need to interoperate or interwork with every other network; those decisions will be made on a company-by-company basis. But the industry is working to make sure that if that choice is made, the technology will be available to support it.

## Market Forecast

Since 1992, PCIA has regularly surveyed wireless communications industry leaders to evaluate the growth, composition, and characteristics of the future of the personal communications industry and has published these results in a PCS Market Demand Forecast. In its yearly surveys, PCIA has asked respondents to provide market size predictions in terms of the number of anticipated subscriptions, not subscribers, anticipating that an individual would probably subscribe to more than one type of wireless service in the coming decade. As in previous years, the 1995 figures show that consumer demand for personal communications services is expected to grow at ever-increasing rates.

Demand growth for new broadband PCS customers is expected to reach 15 million subscriptions by 2000. Total revenues are expected to reach \$8.8 billion by the year 2000, with 7 percent of that revenue coming from data services. Average revenue per subscription is expected to be 20 percent less than that for cellular. Figures for 2005 indicate strong sustained growth to almost 40 million subscriptions and total revenues reaching \$17.5 billion, with 12 percent from data services.

Established voice services such as cellular are expected to grow as well. Respondents expect strong cellular growth during the next 5 years, with the 1994 year-end subscriber count of 23.2 million expected to double to approximately 50 million subscriptions by 2000, with nearly 65 million cellular subscriptions expected by 2005. Thirty percent of the total cellular subscriptions are expected to come from the business segment, representing a presumed growth of the cellular markets into households over the next 10 years. Total cellular revenues are forecast to be approximately \$26 billion by 2000 and \$31 billion by 2005.

In the narrowband PCS arena, market size is expected to reach more than 50 million subscriptions by 2000; by 2005, 71 million one-way and 21 million two-way messaging subscriptions are anticipated. In addition, survey results forecast strong growth from new narrowband PCS and advanced one- and two-way messaging and suggest that these will become established in the wireless world over the next decade. Customer segments will grow due to new narrowband applications and services. Survey results show that by the year 2000, more than 50 percent of one-way and about 65 percent of two-way subscribers are expected to be from business segments. Assuming that businesses will continue to upgrade services, they are expected to remain more than 50 percent of the total subscriber base through the next decade. Total revenues are expected to reach \$4.7 billion for one-way paging and \$1.9 billion for two-way paging by 2000, and \$5.6 billion and \$3 billion, respectively, by 2005.

## DEPLOYMENT OF THE NATIONAL INFORMATION INFRASTRUCTURE

### Site Acquisition Issues

Acquiring PCS antenna and base station sites and gaining the appropriate zoning approvals vary by state and local jurisdictions and are important in wireless network deployment. Furthermore, there are issues regarding site acquisition (such as FAA tower regulations and the lack of a uniform policy regarding sites on federal property) that need to be addressed at the federal level.

### Issues at the Local Level

There are more than 38,000 local jurisdictions throughout the nation, each with the authority to prevent antenna construction; establish standards that can result in site location degrading the quality of service; or prolong site selection, thereby making it unnecessarily expensive. With an estimated 100,000 new wireless antenna sites predicted over the next 10 years, any licensing obstacles present significant problems.

Congress has recognized the need to remove state and local barriers to deploying CMRS facilities by prohibiting state and local government regulation of matters relating to market entry and rates. The current draft of the Senate's Telecommunications Competition and Deregulation Act of 1995 states that no state or local statute may prohibit or have the effect of prohibiting the ability of any entity to provide interstate or intrastate telecommunications services. It further states that if after notice and comment the FCC determines that a state or

local requirement is inconsistent with the legislation, the FCC shall immediately preempt enforcement of the requirement.

CTIA filed a Petition for Rule Making requesting that the FCC initiate a rule-making proceeding to preempt state and local regulation of tower sites for CMRS. The petition states that the state preemption language in Section 332(c) of the Communications Act gives the Commission authority to exercise such preemption, since local zoning could constitute an "indirect" regulation of entry.

Comments on the Petition for Rule Making were due in February 1995. Predictably, service providers filed in support of the petition, while state and local governments and consumer groups filed in opposition. The challenge the wireless industry faces is balancing the recognized needs of the local community to have oversight and fee administration of zoning issues against attempts to meet the ever-increasing demand for new wireless services.

Additionally, the FCC has imposed build-out requirements on the new PCS licensees which mandate that certain percentages of the licensees' markets be covered within set time frames. Potential conflicts between state and federal regulations threaten to delay the entry of wireless services.

### **Site Acquisitions on Federal Property**

Federal property could, in many situations, provide prime locations for PCS base stations. Unfortunately, many agencies of the federal government are not willing or are unable to entertain the prospect of such facilities because of perceived administrative burdens, lack of benefit to local agency staff, or lack of clear policy or regulations for leasing of federal property for such an installation. Additionally, all of the federal agencies that allow private communications facilities on their land have different regulations, lease documents, and processes for doing so. These are often difficult, time consuming, and expensive for both the agency and the communications companies.

Making sure federal land resources continue to be available for efficient delivery of mobile communications services, and ensuring that taxpayers receive a fair price from every communications company with transmitters on public lands, are goals shared by industry, the federal agencies, and the public. However, there needs to be a consistent, government-wide approach for managing the site acquisition process on federal property.

The Executive Branch needs to set a clear directive in order to overcome the obstacles wireless licensees face when trying to acquire sites on federal property. The benefits to the federal government could include increased revenues from the installation of PCS networks above and beyond the auction proceeds, and the potential for improved communications on federal property.

### **FAA Tower Review Process**

PCIA has initiated discussions with the Federal Aviation Administration (FAA) to remove any possible FAA obstacles to efficient deployment of PCS systems. The FCC has established licensing rules that have streamlined the approval necessary to bring systems and PCS cell sites on line. However, due to administrative limitations, the FAA, which must review many requests for towers to ensure air safety, has experienced longer processing times that have delayed carriers' ability to activate certain transmitter sites. With approximately 25 to 30 percent of new wireless sites requiring FAA action and review, PCIA fears that FAA processing delays could significantly burden the industry. Working groups at the national and local levels have been established as a forum to explore methods of educating the industry about FAA procedures and to explore ways to streamline the FAA tower review process.

The FAA, FCC, PCIA, and the Cellular Telecommunications Industry Association (CTIA) have all agreed to participate in this dialogue as part of an antenna work group (AWG) in Washington, D.C. PCIA has also participated in dialogues with the FAA Southern Region and will be working on a local level in other working groups to identify ways to improve the FAA process.



### **Federal Radio Frequency Emissions Standard**

As PCS, cellular, paging, and other wireless carriers build out networks, they are increasingly facing state and local laws and ordinances based on radio frequency (RF) exposure levels, often with conflicting scope and standards, resulting in compliance difficulties. Conflicting standards affect the range of wireless services and can greatly diminish the quality of service consumers receive. This adds greatly to the expense borne by the industry, not only in legal and other business expenses, but also in lost revenue opportunities from long delays in providing services.

The FCC has the authority to preempt local jurisdictions on cell/antenna/tower siting but to date has approached this issue on a case-by-case basis. With as many as 100,000 new wireless sites to be installed, including new PCS sites, and additional sites that will be needed for the expansion and enhancement of service areas for paging, SMR, ESMR, and cellular service, a case-by-case approach to preemption is no longer realistic.

The FCC on April 3, 1993, issued its Notice of Proposed Rule Making, which proposed updating guidelines and methods for evaluating the environmental effects of electromagnetic exposure, and adopting the standard developed by the American National Standards Institute (ANSI) with the Institute of Electrical and Electronic Engineers (IEEE). In December 1994, the Spectrum Engineering Division of the Office of Engineering and Technology of the FCC issued information indicating that levels of exposure to RF at ground level below typical cellular towers are hundreds to thousands of times lower than the proposed standard.

On December 22, 1994, the Electromagnetic Energy Association (EEA) filed a petition with the FCC for a Further Notice of Proposed Rule Making. The petition requested that the FCC preempt state and local regulation of RF exposure levels found to be inconsistent with the FCC proposed ANSI standard.

PCIA favors the establishment of a single, national RF emissions standard that may not be exceeded by local regulations. PCIA encourages the relevant federal agencies to work cooperatively with industry on this issue to develop such a national standard.

### **Interconnection**

#### **Interconnection with Local Exchange Carriers**

Negotiating reasonable rights, rates, and terms under which companies will interconnect with other networks is critical to the success of PCS. With many PCS hopefuls eyeing the local exchange market as a potentially lucrative area in which to compete, the terms of companies' interconnection agreements as a co-carrier will become even more important as they strive to compete with local exchange carriers (LECs), and therefore they will need reasonable interconnection agreements so that they can offer customers low-cost exchange service.

As an example of current interconnection costs, Type 2 interconnection charges for cellular carriers generally are measured on a per-minute basis, with costs ranging from 2 cents per minute to 6 cents per minute, and 3 cents per minute often being considered a "good" interconnection rate.

Interconnection charges have diminished cellular carriers' revenues since the first system came on line, and they remain a high cost to carriers today. Take, for example, a cellular monthly bill of \$59, which includes 86 minutes of air time, at the rate of 3 cents per minute for interconnection. Interconnection charges represent \$2.58 of the bill, or 4.37 percent of revenue.

As air time costs continue to decline in the wireless marketplace, interconnection costs will begin to reduce revenues even more. For example, Cybertel Cellular offers 5 cents per minute of air time in Kauai to compete with the local exchange carrier. At an interconnection rate of 3 cents per minute, interconnection charges could consume 60 percent of the carrier's air time revenue.

Obviously, those wishing to compete at the local loop must achieve lower interconnection costs to compete with established carriers on price. One solution to this problem is mutual compensation, where both telecommunications carriers are compensated for the traffic that terminates on their network.

## Mutual Compensation

Mutual compensation is the concept that a carrier should be compensated for traffic that originates on another network but terminates on that carrier's network, and vice versa. Currently, wireless carriers must compensate wireline carriers for traffic that originates on a wireless network and terminates on a wireline network. Almost without exception, wireline carriers do not compensate wireless carriers for traffic originating on a wireline network and terminating on a wireless network.

The FCC has repeatedly stated that, for interstate traffic, wireline carriers must compensate wireless carriers for traffic originating on a wireline network and terminating on a wireless network. However, states have been reluctant to enforce mutual compensation on an intrastate basis, and therefore wireline carriers have refused to participate in mutual compensation on either an intra- or interstate basis.

Enforcement of mutual compensation rights of wireless carriers is considered to be a key to full competition by wireless carriers in the telecommunications market and will have a significant positive financial impact for the wireless industry.

One potential solution to the high cost of interconnection would be mandating mutual compensation through reciprocal elimination of interconnection charges. One example of this solution is the agreement reached in New York between Time Warner and Rochester Telephone, whereby Rochester Telephone will collect 2.2 cents per minute for traffic terminating on its network and pay at the same rate for its own traffic terminating on other networks. According to the agreement, mutual compensation provisions are eliminated when the traffic flow differentials fall below 10 percent.

## Numbering Issues

The issue of who controls numbers is key to the success of PCS carriers. Traditionally, most national numbering resources have been assigned by the North American Numbering Plan Administration sponsored by Bellcore, which in turn is owned by the Bell operating companies. Generally, the dominant local exchange carrier ends up assigning numbers to wireless carriers in its local telephone market. Wireless carriers usually are charged for activating blocks of numbers in local exchange carrier networks, and the charges vary greatly.

Recently, Bellcore has come under scrutiny for its administration of numbering resources, and actions by wireline carriers within the past few months have brought the issue to the forefront. For instance, in Chicago, Ameritech proposed an "overlay" area code. This would require cellular and paging subscribers to give back their numbers and receive a new area code, thus freeing up numbers in the almost-exhausted code for new wireline subscribers. At a January FCC open meeting, the FCC found this proposal to be "unreasonably discriminatory" against wireless carriers.

The FCC initiated a proceeding more than a year ago to examine whether an independent entity should oversee the assignment of numbers, and it appears as if the Senate telecommunications reform effort might mandate the formation of an independent entity to oversee the numbering assignment process.

## Number Portability

Another key issue for those who want to compete with the local telephone company is number portability, or the ability of an end user, such as an individual or business, to retain its 10-digit geographic North American Numbering Plan (NANP) number—even if the end user changes its service provider, the telecommunications service with which the number is associated, or its permanent geographic location.

With few exceptions, today end users may not retain their 10-digit NANP number if they:

- Switch service providers, referred to as "service provider portability" (e.g., a user switches from an incumbent LEC to a new competitive access provider);

- Change the service to which the number was originally assigned, referred to as "service portability" (e.g., a cellular telephone number becomes the wireline home telephone number);
- Change their permanent location, referred to as "geographic portability" (e.g., an end user moves to a different part of the city or state, and may be assigned either a new 7-digit phone number in the old area code or a new 10-digit number in a new area code).

Service provider portability, that is, moving a number from one service provider to another, is vital for those companies that wish to compete for customers at the local exchange level. It is much easier to gain market share if the customer a company is trying to attract does not have to change his or her phone number when changing service providers.

Currently, 800 numbers are portable between 800 number service providers—an example of service provider portability. This portability allows the 800 service end user to retain his or her individual 800 number, even when switching 800 service providers. Portability of 800 numbers was ordered by the FCC and implemented in 1993.

### **Industry Efforts to Address Number Portability**

The Industry Numbering Committee (INC), a consensus-based industry body sponsored by the Inter-Carriers Compatibility Forum (ICCF), has been actively addressing number portability issues since the fall of 1993. The INC Number Portability Workshop has been addressing a range of issues associated with number portability, including a target portability architecture, network impacts of number portability, and high-level policy issues such as mandated interconnection.

## **PUBLIC SERVICE OBLIGATIONS**

The advent of increased mobility is having an impact on telecommunications public policy. How does wireless technology fit into public policy initiatives such as universal service and access to enhanced 911 emergency calling services? Policies regarding universal service were developed to apply to a strictly wireline environment where competition at the local level was nonexistent. Additionally, wireless technologies present a challenge to the traditional wireline approach to providing enhanced 911 emergency calling. As wireless service providers begin to compete for the local loop, how wireless fits into such public policy provisions will need to be seriously considered.

### **Universal Service**

Universal service, as a public policy concept, is the belief that access to basic telephone services by the widest possible cross section of the American public is in the social and economic interests of the United States. Over a period of many years, Congress has mandated the creation of universal service programs to support universal service public policy goals. The FCC is charged with fulfilling these congressional mandates.

Within the telecommunications industry, universal service refers to a complex system of explicit and implicit charges and cost allocation mechanisms levied on particular carriers and customers in order to provide access to, and subsidize the rates of, basic wireline services for residential customers, high-cost customers and carriers, low-income customers, rural areas, and services for hearing- and speech-impaired consumers.

Estimates of the current total costs of supporting universal service goals and policies range as high as \$20 billion to \$30 billion annually. Congress is intent upon reform of universal service policy and funding mechanisms as part of its effort to reform existing telecommunications law. Any reforms could have a potentially huge economic impact upon the wireless industry.

Universal service reform is a critical part of telecommunications reform and it appears inevitable if Congress passes a telecommunications reform bill. Although it is too early to tell what shape universal service will take, a number of issues need to be considered.

### Wireless Access to Enhanced 911 Emergency Services

The FCC, on October 19, 1994, released a Notice of Proposed Rule Making (NPRM) regarding revision of the Commission's rules to ensure compatibility with enhanced 911 (E-911) emergency services. In many areas of the country, wireline subscribers are provided E-911 service by wireline carriers, which entails transmitting the address and phone number of the caller to the public safety answering point. The NPRM addresses PBX issues and wireless service provider issues. The NPRM outlines proposed requirements on wireless services regarding:

- 911 availability;
- Grade of service;
- Privacy;
- Re-ring/call back;
- Grade of service;
- Liability;
- Cost recovery;
- Access to text telephone devices (TTY);
- Equipment manufacture, importation, and labeling;
- User location information;
- Compatibility with network services;
- Common channel signaling; and
- Federal preemption.

The proposed requirements have considerable technical and economic implications that need to be fully examined. PCIA, in cooperation with representatives of the public safety community, drafted the *Joint PCIA, APCO, NASNA Emergency Access Position Paper*, which was filed with the FCC in July 1994. This joint paper documented the first attempt of the PCS community to comprehensively address the needs of the public safety community. The FCC used the joint paper as a basis for its NPRM addressing enhanced 911 emergency calling systems.

PCIA fully shares the Commission's important objective of maximizing compatibility between wireless services and enhanced 911 emergency calling systems. Specifically, it concurs that subscribers to real-time voice services interconnected with the public switched telephone network ultimately should enjoy the same access to advanced emergency response services as do wireline service subscribers, with due consideration for the unique characteristics of radio-based technology. At the same time, however, PCIA strongly disagrees with the approach toward achievement of the compatibility objective that is set forth in the NPRM.

PCIA believes that full-scale regulatory intervention is not necessary at this time and that the profound technical issues raised by compatibility cannot be resolved through imposition of arbitrary deadlines as proposed in the NPRM. PCIA proposes, as an alternative to arbitrary deadlines, that the industry work to develop technical solutions to the public safety community's requirements and that the FCC require periodic reports from the industry on its progress in meeting the ultimate goals the FCC has set forth.

### CONCLUSIONS

The fact that the NII is a complex web of wireline and wireless service providers providing both voice and data services to the end user at home, in the office, and walking or driving down the street needs to be considered in any telecommunications policy initiative. The new wave of wireless service providers, while providing the

consumer with more choices in services and features than ever before, presents a challenge to the public policymaker who tries to determine how to ensure that telecommunications services are made available to the broadest range of consumers. Competition will take care of that to a certain extent. However, where appropriate, government may need to step in on issues such as interconnection rights, mutual compensation, and numbering to ensure that new entrants are treated as equals by incumbent carriers. Furthermore, revision of universal service and enhanced 911 policies needs to take into consideration both the wireless and the wireline industries.

Additionally, the wireless industry is often faced with federal and state regulatory processes that can slow down the deployment of new networks. Federal guidelines regarding site acquisition and radio frequency emissions are necessary to ensure timely availability of new services. There continues to be a high demand for wireless services, and the industry is poised to meet that demand. However, public policy should be developed such that the promise of wireless services as an integral component of the NII is realized.

### NOTES

1. The Composite CDMA/TDMA system is an air interface technology currently being standardized for PCS in the 1.8-GHz band.
2. Broadband PCS refers to the family of mobile or portable radio services operating in the 1.8-GHz range and providing a wide variety of innovative digital voice and data services.
3. Narrowband PCS services are expected to include advanced voice paging, two-way acknowledgment paging, data messaging, and both one-way and two-way messaging.

## 42

# Small Manufacturing Enterprises and the National Information Infrastructure

*Robert M. Mason, Chester Bowling, and Robert J. Niemi*  
*Case Western Reserve University*

### STATEMENT OF THE PROBLEM

The vision for the future is that an emerging national information infrastructure (NII) and its defense counterpart (DII) will equip U.S. industry to be second to none in the global economy. The NII will enable the U.S. industrial base to become more agile and to operate as a highly competitive, flexible, just-in-time, manufacture-on-demand system that facilitates free competition and specialization among manufacturers and suppliers. All firms, regardless of size, will have ready access to product requirements and specifications and will be able to compete fairly with other firms. Moreover, the NII with the DII will encourage commercial suppliers to respond to defense needs, enabling dual use designs and strengthening the flexibility of the nation's defense infrastructure.

The reality is that many existing small firms are ill equipped to participate in this vision. Moreover, there is concern that the learning cycle for small manufacturing enterprises (referred to here as SMEs) to implement information technology is too long and costly for them to effectively make the transition to the NII environment. The solution to the problem is not simply one of assuring that every SME can purchase and install a new information system. Instead, the solution requires an understanding of how a complex combination of structural, technical, managerial, and economic factors affect the diffusion of information technology in the manufacturing sector, especially among SMEs. From the viewpoint of our national economy, the problem is that this complex set of factors impedes the effective implementation of information technology in SMEs and puts at risk a significant component of the nation's manufacturing base, a component that is responsible for up to nearly 40 percent of the nation's manufacturing employment. Developing nations may "leapfrog" over the United States and other advanced nations if our established enterprises are unable to change quickly enough. The purpose of this paper is to help understand this set of factors and to explore how best to manage the risk associated with a slow rate of diffusion of information technology in SMEs.

### OVERVIEW

The "Background" section provides a synopsis of the different views on the role of SMEs in the nation's economy and manufacturing infrastructure. This section also summarizes the different frameworks within which we can understand the economic, behavioral, structural, and technical issues associated with how SMEs may participate in the benefits of the NII.

The "Analysis" section of the paper provides more detail on the frameworks outlined in the Background section and examines the prospects for SMEs to become full participants in the NII. This section synthesizes adoption and diffusion studies and research on the implementation of new information technologies. The emerging framework is that of organizational learning at the level of the firm and the concept of the supply chain (or value chain). The learning framework enables us to make sense of the range of factors associated with

technology adoption, and the value chain framework illustrates why action by an individual firm is inadequate for that firm to realize the benefits of an NII.

The final section of the paper discusses opportunities for national policy to alleviate the problem of SME's participation in the NII. The paper concludes that a coordinated collaboration among private, university, and government resources offers the best way to assist U.S. SMEs in making the transition to electronic commerce and the benefits of the NII.

## BACKGROUND

### SMEs and the Economy

Small manufacturing enterprises (SMEs) are responsible for an estimated 28 to 40 percent of the employment in the manufacturing sector.<sup>1</sup> Moreover, there is evidence that SMEs are more effective at job creation<sup>2</sup> and job replacement,<sup>3</sup> more innovative in the development of products and process improvements,<sup>4</sup> and more flexible and thus more competitive in terms of the ability to produce small quantities. All these factors may explain the shift to a smaller average plant size.<sup>5</sup> The claims to new job creation are open to question—SMEs also exhibit high failure rates, and thus new jobs may not be long-lived.<sup>6</sup> However, others point out that small firms will continue to add jobs because much growth will take place in industries in which small businesses have relative advantages.<sup>7</sup>

There is no question that SMEs are a crucial element in the nation's manufacturing base. If one believes, as many do, that manufacturing must continue to be a foundation for U.S. economic competitiveness,<sup>8</sup> then SMEs will continue to be a crucial part of this competitiveness.<sup>9</sup> The role for small firms appears to be increasing; there is evidence of a trend toward more of the total production coming from smaller manufacturers.<sup>10</sup> However, the United States is lagging behind Europe and Japan, where small firms account for 45 to 60 percent of manufacturing employment.<sup>11</sup>

### SMEs and the NII

Neither the global competitiveness of U.S. industry nor the future role of SMEs is assured. The NII vision of preserving the tradition of free market competition both among manufacturing suppliers and among international companies is consistent with what Porter<sup>12</sup> suggests are the conditions for global competitiveness: demanding customers who can choose from an intensely competitive local network of suppliers.

The NII and DII are expected to enable this competition and the development of dual use processes and designs. Large manufacturers (including Department of Defense purchasers and contractors) can make their specifications available online, eliminating distribution delays and increasing the scope of distribution. In one vision currently being articulated by the U.S. Department of Defense (DOD), the NII and DII enable the creation of an Integrated Data Environment (IDE) in which information itself (e.g., designs, production methods) becomes a commodity and is traded. With information available on both specifications and designs, firms can work only on those opportunities for which they are most capable, reducing the risk and costs of bidding on marginal opportunities.

For SMEs to participate, they must have access to the NII and they must be able to use computer technology to integrate their business and technical functions. They must understand and use electronic commerce technologies (ECTs). Currently, small businesses are not utilizing computers to the degree necessary to fully participate. A recent survey commissioned by IBM indicated that while 82 percent of small businesses (not just manufacturers) had desktop computers, only 37 percent had local or wide area networks.<sup>13</sup> In a stage model of information technology maturity,<sup>14</sup> almost two-thirds of these respondents would fall into the first, most elementary, stage of maturity.

SMEs are becoming aware of the need to adopt some form of electronic communications. With increasing frequency, prime contractors and large firms have demanded that their suppliers have electronic

capabilities. As one would expect, this has heightened interest in electronic commerce capabilities among small- and medium-sized businesses. The interest is likely to escalate. One software vendor executive, explaining that his company was trying to respond to customers' needs for advice and consultation about electronic commerce, put it this way, "Our executives have been around long enough to tell the difference between a ripple and a wave. This one is a wave."<sup>15</sup>

Engineers from the Cleveland Electronic Commerce Resource Center (Cleveland ECRC) report similar interest but observe that some small firms are satisfied with a "rip and read" solution: they link into a bulletin board or value-added network with a personal computer but use the computer as an expensive fax machine. They "rip off the printed specifications," then read them and insert them into their manual system.<sup>16</sup> This approach works for written specs and to some degree for drawings, but it clearly is limiting. More advanced firms install a computer aided design (CAD) system to enable them to accept design data in digital formats. Often, they too have a manual internal system and do not attempt to use the digitally stored format.

Compounding the technical problem is the lack of a single standard that is widely accepted; Chrysler, Ford, and GM use different, incompatible CAD systems. For most SMEs, the cost of implementing multiple standards is too high, and they either choose a single customer's standard or opt for another market. In either case, the situation does not lead to increased competition and to the increased competitiveness of SMEs. A single standard would help. Standards such as PDES/STEP are being developed, but agreements and adoption take time, and such standards address primarily the technical issues of data sharing.

Organizational (i.e., managerial and cultural) issues are equal to, if not greater than, technical capabilities in importance. In their discussion of agile manufacturing, Goldman and Nagel<sup>17</sup> share the vision of integration of virtual enterprises through the use of information technology, including standards and "broad-band communications channels."<sup>18</sup> They acknowledge the need for flexible production machinery but point out the need for organizational innovations as well. The agile system they envision requires flexible production workers and managers, not just technology. Getting the integration of technology and people into a new, responsive system is a challenge. They conclude, "An understanding of the nature of managerial decision-making is more important than ever before."<sup>19</sup>

Other researchers agree with Goldman and Nagel that the managerial, organizational, and cultural issues are at least equal in importance to the technical challenges of tapping into the benefits of the NII. In a field study of five large firms that were judged to be implementing integrated information systems successfully, a study team found six shared characteristics among the firms, and only one (the goal of capturing all data in electronic form at its origin) was technical.<sup>20</sup> The other five characteristics (vision and clear strategy, vocabulary/language incorporating metrics shared by technical and business staff members, customer focus, and a sense of urgency) were organizational factors.

### Factors Affecting SMEs' Adoption of Technology

One approach to understanding SMEs' use of information technology would be to view ECT as a technology that will be diffused throughout manufacturing. This diffusion approach<sup>21</sup> uses the familiar S-curve to identify the percent of SMEs that have adopted ECTs over time. Factors associated with an individual firm's propensity to adopt technology might suggest strategies for working with innovators, early adopters, and so on.<sup>22</sup> Implications of this type of model for policy are discussed further in the final section.

Another useful approach, the one taken for the remainder of this paper, is to seek an understanding of the decision-making process within the SME. From this viewpoint, we may gain some insight into the economic, technical, structural, and other barriers to adoption as seen by the SME.

The stage model suggested by Venkatraman<sup>23</sup> of firms' use of information technology (Figure 1 shows an adaptation of this model) is used as a basis for identifying the gap between the "as is" state and the "desired" (or "to be") state of SME capabilities.



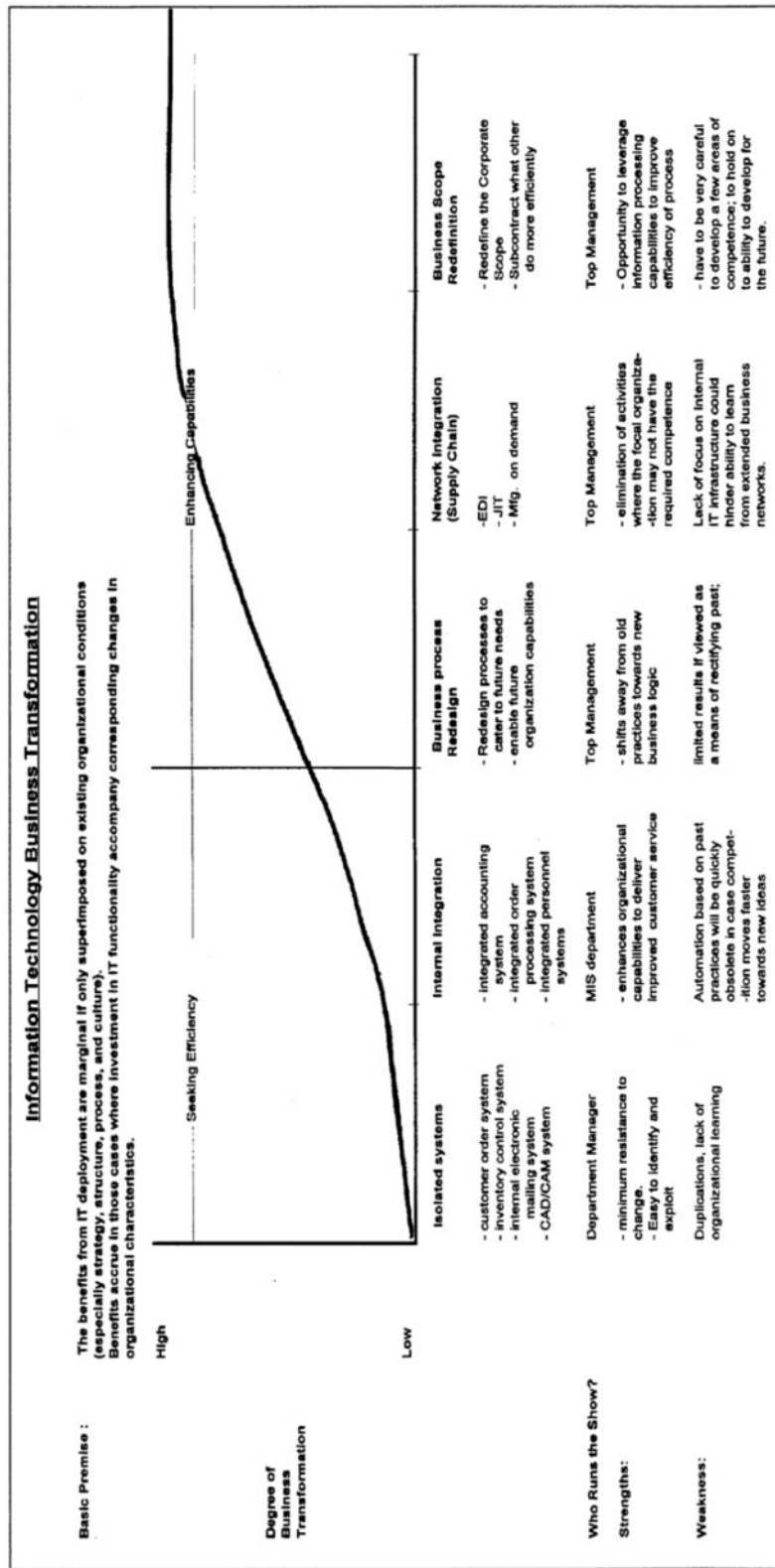


Figure 1 Stage model of information systems use. SOURCE: Adapted from Venkatraman, N., 1994, "IT-enabled Business Transformation: From Automation to Business Scope Redefinition," *Sloan Management Review*, Winter, pp. 73-87.

The data from the IBM survey of small businesses<sup>24</sup> indicate that almost two-thirds of the survey respondents are in the first stage of maturity in applying information technology. Virtually none of them have progressed beyond the second stage, and there is no assurance that they will go beyond this stage. For SMEs to benefit from the NII, they must be at level 3 or above, developing capabilities for network/supply chain integration. Although the IBM survey was not limited to manufacturing firms, our experience with SMEs leads us to speculate that small service firms and those in the retailing and trade sectors may use computers even more than manufacturers, lowering even further the estimate of how many SMEs have moved beyond the first stage of computer use.

This stage model is descriptive, and it only indirectly suggests how an organization moves from one stage to another. Our concern is to understand how the organization, particularly an SME, progresses from the applications of isolated systems to network and supply chain integration and, more importantly, how this process can be accelerated. The relevant fields of research are those of technology policy, innovation adoption, the decision-making process within the firm, and the emerging field of inquiry on organizational learning.

The concept of organizational learning,<sup>25</sup> particularly the use of the human experiential learning model proposed by David Kolb<sup>26</sup> and recently applied to organizations,<sup>27</sup> provides a useful framework to interpret the findings from the other fields. This model, shown in Figure 2, illustrates the different modes by which an individual (organization) learns. Learning takes place in two dimensions: in the concrete-abstract dimension (shown vertically as a continuum) and in the active-reflective dimension (shown horizontally). Individuals (and organizations) have different preferences for learning and processing information in these dimensions.<sup>28</sup> Some prefer more concrete and active learning (e.g., entrepreneurs); others prefer more abstract and reflective learning (e.g., professors).

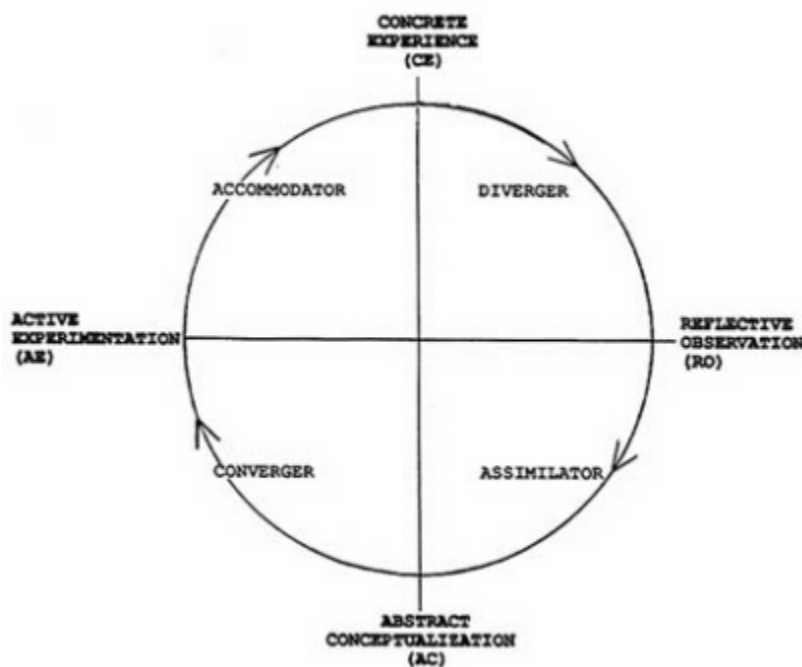


Figure 2 Experiential learning cycle. SOURCE: Reprinted from Kolb, David, 1984, *Experiential Learning: Experience as the Source of Learning and Development*, prentice-Hall, Englewood Cliffs, N.J.

The learning cycle model suggests that only when the organization goes through all four modes is learning complete. For example, a firm may introduce a new process for a customer or product line (*active experimentation*), collect sales and quality data over time (*concrete experience*), interpret these data and compare with prior experience (*reflective observation*), and develop a projection of sales and costs of quality if the new process were applied to all their product lines or to all their customers (*abstract conceptualization*). Based on the model, the firm may choose to switch its other products to the new process, again moving to *active experimentation* and

restarting the cycle. By passing through each of the learning modes, the firm generates new knowledge. The firm *learns*, and the learning is not limited to the simple aggregation of additional data or to thinking about a new idea—the cycle is complete.

Using the concept of the learning cycle, we can frame our concern as that of understanding the predominant learning modes of SMEs and of understanding how SMEs can incorporate all learning modes in their progress toward the higher stages of information technology maturity. For this understanding, we can draw on several areas of research about how organizations adopt technology. In each of the relevant areas, it is evident that one must use caution in applying concepts derived from the large organizational context to the SME.<sup>29</sup> However, some studies have focused specifically on the decisionmaking and policy formulation in the small firm, and these studies are particularly helpful in our efforts to understand how to accelerate learning and ECT adoption in SMEs.

## ANALYSIS

The first three subheads below outline relevant concepts from three distinctive but overlapping areas of inquiry. Each has its own literature base and each offers some insight into how firms, and SMEs in particular, may implement and use information technologies. The fourth subhead outlines the structural issues that may initially inhibit SMEs' effective participation in the NII. The section concludes with a synthesis of ideas about how SMEs may approach the adoption of electronic commerce technologies and realize the benefits from the NII.

### Diffusion of Technology

The diffusion literature<sup>30</sup> characterizes the industry adoption of new products by an S-shaped curve. The curve reflects exponential growth with a rate that depends on the size of the remaining market. The diffusion model has been used with some success in technology forecasting. With good data on when a low level of adoption has been achieved (e.g., 5 percent), the model is effective in identifying the dates by which a specific level of industry penetration (e.g., 50 percent) will occur.

The S-curve model is often used to identify firms according to when (early or late) they make the decision to adopt the technology. The classifications may indicate different organizational characteristics. A modification of this conceptual model<sup>31</sup> classifies the "buyer profiles" as being one of five types: innovators, early adopters, early majority, late majority, and laggards.

Recent research<sup>32</sup> tested the idea that psychological characteristics (e.g., attraction to technology, risk taking) rather than economic variables might be used to discern buyer profiles. The study found that the benefit-cost variables were better predictors. Although one could argue with how the variables were operationalized and with the limits of the study (focus groups on a single product), the researchers' conclusion has face validity: Companies that pioneer new products must focus on the benefits desired by purchasers. Even the early adopters, who are less price sensitive, seek benefits that meet their needs better than current technologies. What is not discussed in the study is the changing nature of the benefits and costs with changes in the organizational characteristics and with changes in risk as the technology matures.

Kelley and Brooks<sup>33</sup> also showed the predictive power of economic incentives in the diffusion of process innovations. Not surprisingly, firms with high wage rates were more likely to adopt labor-saving technologies than were firms with low wage rates. The key is to note that the benefits and costs are established by the firms' perceptions; these perceptions are affected by the organizational values and the firm's particular situation.

As noted by Schroeder,<sup>34</sup> the survival of an SME is linked to the adoption of technology as a regular part of doing business. If it is in the nation's interest for SMEs to thrive, then the diffusion issue is how to accelerate the adoption of information technologies among SMEs. The diffusion model may be a useful metric by which we can track and predict adoption rates as early data become available. However, the diffusion model does not help explain how firm-level decisions are made. Concepts that examine how the individual firm makes a technology adoption decision may be more informative in the early development of the NII.

### SME Decisionmaking

The literatures relevant to an SME's decisions on technology adoption are those on corporate strategy, technology strategy, technology policy, information system implementation and planning, strategic information systems, and investment decisionmaking at the level of the firm. These areas of study are rich in topics that are relevant to technology adoption, but the focus on SMEs and their adoption of technical innovations reduces the scope considerably.

SMEs differ from large companies in how they develop their corporate strategies and their technology policies. Large companies typically have well-defined processes for developing and implementing strategies through a corporate planning process. Small firms often use less structured approaches; strategies and policies may not be formulated but may "emerge" from a set of actions and experiments.<sup>35</sup>

In an SME, the chief executive officer (CEO) often is one—or perhaps the—owner of the firm. In these firms, the CEO's viewpoint is a critical contributor to strategy and policy. A recent study of SMEs<sup>36</sup> showed that implemented technology policies (not just written policies) in SMEs are strongly influenced by how the CEO perceives the world. Even though all the firms in the study were immersed in the same industrial setting in the same Canadian province, the CEOs differed in their view of how hostile and how dynamic their environment was. The firms' propensity to invest in new technology was strongly related to these views. The basis for decisions is not an objective reality but rather a socially constructed reality<sup>37</sup> as reflected in the viewpoint of the CEO.

The social construction of the adoption decision by a firm has other participants as well. For the SME, a strong influence is the supplier, who may be a major source of information.<sup>38</sup>

The innovativeness of an SME is related to the firm's outward orientation (e.g., customer focus) and the participation of the firm's functional groups in the decision.<sup>39</sup> There is evidence<sup>40</sup> that the SME learns with increasing technological capabilities so that, over time, its decisionmaking places more weight on factors that are more closely related to the true potential of the technology.

### SME Learning

Arrow<sup>41</sup> noted that firms learn through experience. This learning normally is considered to be related to process improvements and is the foundation for the concept of reduced costs over time because of "the learning curve." More advanced technologies may have greater productive potential, but the firm has less expertise in implementing such technologies. Knowing it has less expertise, the firm expects greater costs. The firm thus faces a trade-off in its choices of technologies to adopt.<sup>42</sup>

The capacity for learning affects the rate of adoption of new technology. Firms that have existing technological capabilities have higher "absorptive capacity"<sup>43</sup> for new technology; they are able to learn more quickly.<sup>44</sup>

A firm's installed technology also affects the extent and magnitude of benefits the firm experiences from installing new systems. Firms that have more existing technological capabilities—for example, firms that have implemented information technologies in both the administrative and engineering/production operations—enjoy benefits that are greater than the sum of the benefits from individual systems. There is synergy and, because of the added benefits and increased capacity for learning, the "rich get richer" and vice versa. This appears to be the case both for large firms<sup>45</sup> and for SMEs.<sup>46</sup>

When a technology is new to an industry—before its technical and economic superiority has been widely accepted—the learning capacity of a small firm is related to the firm's linkages with other firms and other industrial organizations. These external linkages, many of which provide informal but trusted conduits for sharing of technical know-how, appear to lower the cost of learning for the firm. Kelley and Brooks put it this way: "Small firms' propensity to adopt a process innovation is particularly enhanced by the nature of linkages to external resources for learning about technological development.... Where linkages to such external learning opportunities are particularly well-developed we would expect to find a more rapid rate of diffusion of productivity-enhancing process innovations to small firms."<sup>47</sup>

Organizational learning may be "single-loop" or "double-loop."<sup>48</sup> In single-loop learning, the organization improves its efficiency, becoming ever better at dealing with a prescribed problem or environmental situation. The lowering of costs because of the "learning curve" is an example of single-loop learning. Double-loop learning, by contrast, is characterized by a shift in viewpoint and a modification of basic premises. Double-loop learning requires unlearning prior assumptions and standard operating procedures; it involves developing new paradigms, new frames of reference, and new interpretive schemes. Single-loop learning reduces variability; double-loop learning increases variability in search of more relevant objectives or more effective strategies.

Because prior procedures and paradigms have a history of success, organizations have difficulty engaging in double-loop learning; they actively resist.<sup>49</sup> However, dynamic and turbulent environments demand that firms exhibit more variability in order to meet changing needs. One approach to stimulating variability—and possibly double-loop learning—is organizational restructuring. Restructuring (changing the top management team and/or the CEO) is especially effective when combined with a change in strategy (e.g., new products or markets).<sup>50</sup>

SMEs, especially the smaller ones, are less likely to adopt a restructuring approach. A turbulent environment sometimes stimulates an SME owner to sell or merge with a larger firm. Often, however, the SME that cannot adapt quickly enough to environmental changes simply ceases to exist. The latter outcome contributes to the statistics used by those who argue that SMEs provide unstable employment, even if they do create a significant portion of new jobs.

The learning model in Figure 2 provides a framework that helps synthesize these issues. Since complete learning means that the organization engages in each of the modes, an enterprise may engage in formal or informal collaboration with external organizations to learn. For example, the motivation for close alliances between suppliers and manufacturers<sup>51</sup> is partially explained by the benefits of learning, and the higher rate of innovation adoption because of external contacts<sup>52</sup> may be due to the expanded learning modes made possible by these contacts.

An alternative to restructuring or going out of business is to establish and maintain external relationships that enable learning. Such organizations, which "bridge"<sup>53</sup> sources of knowledge about new technologies (e.g., universities) and the SMEs (as potential users), have been stimulated by federal- and state-level programs that have set up technology transfer centers and assistance networks. Ohio's Thomas Edison Technology Centers, the federally funded Manufacturing Technology Centers (MTCs), and, most appropriately, the federally funded Electronic Commerce Resource Centers (ECRCs) are examples of such bridging organizations.

The value of such organizations was set forth over a decade ago by Trist,<sup>54</sup> who noted that complex societies and rapidly changing environments give rise to "meta problems" that a single organization is unable to solve. The solution is the development of "referent organizations" that mediate the interorganizational collaboration required in the organizational domain of interest.

Although detailed studies of the effectiveness of MTCs and ECRCs are premature given their recent formation, the political judgment seems to be that they are effective.<sup>55</sup> Studies of Ohio's Thomas Edison Technology Centers generally have praised their value and effectiveness.<sup>56</sup> One of the challenges noted is that of "relationship-building."<sup>57</sup> There is the explicit acknowledgment that the relationships and the process of technology solving are equal to, if not greater than, the importance of developing the technology itself. These evaluations appear to support the concept that the bridging, or referent, organizations contribute to learning, and that at least part of the new knowledge created is not migratory knowledge but is embedded in the relationships that are established and maintained.<sup>58</sup> Implicit in the Mt. Auburn report is the notion that the relationship-building role of these organizations is underdeveloped.

### **The Structural Issue: Of What Benefit Are a Few Telephones?**

The current status of electronic commerce technology may be similar to that of the early telephone. Imagine being given the opportunity of purchasing the third or fourth (or even the fiftieth) telephone: unless you are assured that the other people (organizations) with whom you want to talk (trade/communicate) are equipped with compatible technology, the benefits are nil. Unless the advanced technology has its own appeal, a prudent business decision is to "wait and see"—wait until there is a critical mass of manufacturers and suppliers with

whom you can beneficially communicate. Except for the innovators and early adopters, most of the potential SME users of ECTs—if they are aware at all of the NII and its electronic commerce benefits—are likely to think of these as something that may be possible in the future.

One approach to the structural barrier to the diffusion of ECTs is to think of the SMEs in clusters<sup>59</sup> that share a characteristic or interest. Geographic clusters exhibit their own rate of technology diffusion that can be enhanced by bridging and referent organizations in those regions.<sup>60</sup> Other clusters that share other interests (e.g., those firms in a supply chain) may be distributed geographically.

For industries in which the technology is relatively stable (e.g., automobile manufacturing) compared with the dynamism of emerging technologies (e.g., biotechnology), the shared interests of the supply chain may motivate groups of firms to adopt ECT more quickly. Although the relationships of suppliers to the manufacturers has become closer over the past several years, there still are no widely accepted technical standards, nor are there any established social mechanisms for engaging in collaborative efforts.

### Summary: The Key Concepts

The literatures related to SME adoption of information technologies may be summarized in six key points:

1. Business success and implementation of new technology appear to be related.
2. SMEs decide to adopt new technologies based on perceived benefits and costs.
3. SMEs perceive but may not articulate costs of learning to use and integrate new information technologies. This makes evolutionary changes seem less costly and less risky than revolutionary ones.
4. SMEs appear to follow a stage model in implementing information technologies from simple, stand-alone applications to more complex and integrated applications. At the highest level, information technology becomes a component in the strategic definition of the business. There is no evidence (except for greenfield operations) that a firm skips levels, but the movement from level to level may be accelerated.
5. SMEs' benefits from applications of information technologies are cumulative and synergistic, with disproportionately greater benefits as the number of applications (and enterprise integration) increases.
6. SMEs' learning and adoption of new technologies are related to the number and quality of inter-organizational relationships in which they are active.

### Implications for SMEs and the NII

The potential benefits of the NII to SMEs go much beyond simple manufacturing process improvements. If SMEs are to realize the full benefits of the NII, they must advance their level of information technology applications to levels 3 and 4 in the stage model shown in [Figure 1](#).

Once at these levels, manufacturing costs may become lower—for example, firms can more readily specialize and develop core competencies in particular processes. Other benefits, however, contribute to the overall lower costs: shorter administrative lead times, improved risk management through better information about future demands, more flexible (agile) production, and so on. These benefits do not arise because a single firm or even a few firms adopt ECT; they will be realized only if a critical mass of firms in a value chain become interconnected.

The NII is the key element in this interconnection; it is the communications backbone. Even with the backbone, interconnections are not assured. The problem is not merely one of enabling individual firms to adopt ECT; it is one of enabling groups of firms to adopt ECT. This framing of the problem is more than just a change in scale (from one to many); it is a major change in scope and may add significantly to the complexity of the solution. As a minimum, it changes how we approach accelerated learning and adoption of ECT in SMEs.

The SME technology adoption process, as studied by most researchers and as understood today in the United States, presumes independence among the adopters. However, interdependence, not independence, is necessary if the full economic benefits of the NII are to be realized.

This requires cultural changes in SMEs, and the rate at which SMEs change their cultures can be expected to dominate the rate of diffusion of the technology itself (including ECTs) among SMEs. Firms that traditionally have viewed the world through lenses of competition as a zero-sum game now must view competition as a positive sum game: competition as a means of benchmarking and improving one's own performance (e.g., as in organized sports, such as the Olympics). In such a view, technological advances by other firms provide a learning opportunity for their own firm.

## RECOMMENDATIONS

### A Strategy for Setting Priorities for NII Services to SMEs

SMEs, perhaps more than larger firms, have fewer options for second-order learning. For most SMEs, moving to the higher levels of information technology maturity—those levels required for electronic commerce and for realizing the greatest benefits from the NII—will be possible only by evolutionary change. The services available over the NII are expected to be offered by a mix of private, not-for-profit, and government providers. To enable SMEs to benefit from the NII, these providers, to the extent possible, should:

1. Give early priority to encouraging and establishing high-value, low-cost services that SMEs can use as individual firms.

*Rationale:* Most SMEs are at the lowest level of information technology maturity. They will perceive the highest costs (including learning) to be for services that require additional technology and integration. If individual firms can learn to use networks to obtain valued information from read-only services or for simple firm-firm communication (e.g., e-mail), the cultural change is evolutionary and the perceived subsequent costs of moving to more integrated levels will be lower.

2. Match the services offered to the information technology maturity level of the early adopters (e.g., the most advanced 10 percent) of SMEs.

*Rationale:* The perceived costs of moving more than one level make the benefits of adopting a new technology seem "out of reach"; setting the most advanced services just above the capabilities of the early majority balances the need for SMEs to see the possibilities of greater additional benefits with affordable costs of organizational change.

### A Strategy for Public-Private-University Partnerships

Much of the technology will be developed and made available from the private sector. Moreover, the federal government is expected to continue to help establish and encourage the widespread acceptance of international standards for ECT. As established by the summary of research in this paper, the rate at which SMEs adopt ECTs (and benefit from the NII and DII) is dominated by organizational issues rather than purely technical factors. Consequently, the following paragraphs outline high-leverage opportunities for the federal government to improve the capabilities of *existing* public and partnership programs to address these issues.

In particular, DOD and the Department of Commerce programs such as the Manufacturing Technology Centers (MTCs), Electronic Commerce Resource Centers (ECRCs), and Manufacturing Learning Centers (MLCs) provide an appropriate infrastructure for accelerating the changes required in SMEs. These programs comprise geographically distributed networks of centers through which SMEs can receive assistance. From all appearances, these programs are performing their perceived missions successfully and satisfying their constituents. However, there are opportunities to expand these perceived missions and to accelerate the learning

and development of SMEs as participants in the NII. The following is a recommended strategy for expanding these roles:

1. The DOC and DOD should sponsor a series of workshops on SME technology adoption that involves the centers (either across programs or within each program), researchers in SME technology adoption, and SME leaders (and perhaps prime manufacturers) in a particular industry or value chain cluster. The objective of the workshops is to enable the centers to visualize opportunities for additional complementary efforts that would enable the SMEs to develop more quickly.

*Rationale:* These centers should be the leaders in proposing to expand their missions, but they first need to understand the potential roles. Although many of the centers have universities as partners, the partnership may be limited to technological topics. Research results on technology adoption and the management of technology, especially in SMEs, may not be a normal part of the learning environment for the center staffs.

*Possible outcomes:* Expanded roles would emerge, for example, if these programs viewed their centers as referent organizations, operating in a particular interorganizational domain. In addition to the emphasis on technology awareness and technical training that the centers provide to SMEs, the centers could identify opportunities for SME learning that goes beyond the migratory knowledge that an SME can acquire through normal classroom environments or even short-term consulting arrangements. In particular, the center's mission might include convening, facilitating, and maintaining interorganizational consortia that are focused on particular issues. (The ECRCs are doing this to an extent now with regional interest groups.)

2. The DOD and DOC programs, where appropriate, should expand their centers' linkages with educational institutions, including public libraries.

*Rationale:* Research universities (business schools and universities with management of technology programs) can assist the centers in understanding SME adoption of technology and benefit from the centers' experiences. The centers could establish and support ECT facilities in community colleges and public libraries, enabling SMEs to have access to bid information and other emerging NII services until they are prepared to invest in their own facilities.

## REFERENCES AND NOTES

1. Acs, Z. 1992. "Small Business Economics: A Global Perspective," *Challenge* 35(6):38-44; U.S. Small Business Administration. 1993. *The Annual Report on Small Business and Competition*, U.S. Small Business Administration, p. 49.
2. U.S. Small Business Administration, 1993, op. cit.
3. Long, Andrea L. 1984. "Net Job Generation by Size of Firm: A Demographic Analysis of Employer Event-Histories," a report to the U.S. Small Business Administration Office of Advocacy, July 12.
4. Acs, 1992, op. cit.; and Acs, Z. 1994. "Where New Things Come From," *INC* 16(5):29.
5. Carlsson, Bo, and David B. Audretsch. n.d. "Plant Size in U.S. Manufacturing and Metalworking Industries," *International Journal of Industrial Organization* 12(3):359-372.
6. Davis, Steward, John Haltiwanger, and Scott Schuh. 1994. "Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts," *Business Economics*, July 29(3):13-21.
7. Asquith, David, and J. Fred Weston. 1994. "Small Business Growth Patterns and Jobs," *Business Economics* 29(3):31-34.
8. Wheelwright, Steven C., and Robert H. Hayes. 1985. "Competing Through Manufacturing," *Harvard Business Review* 63:99-109; and Wheelwright, Steven C. 1985. "Restoring the Competitive Edge in U.S. Manufacturing," *Calif. Management Review* 27(3):26-42.
9. United States Congress. 1992. *Small Business Manufacturing and Work Force Capability*, hearing before the Subcommittee on Technology and Competitiveness of the Committee on Science, Space, and Technology, U.S. House of Representatives, 102d Congress, second session, Washington, March 9.
10. Carlsson, B. 1994. "Flexible Technology and Plant Size in U.S. Manufacturing and Metalworking Industries," *International Journal of Industrial Organization* 12(3):359-372.
11. Acs, 1992, op. cit.



12. Porter, Michael. 1990. "Competitive Advantage of Nations," *Harvard Business Review*, March–April, pp. 73–79.
13. Mangelsdorf, Martha E. 1994a. "Technology Management in Growth Companies," *INC* 16(14):140.
14. Nolan R. 1979. "Managing the Crises in Data Processing," *Harvard Business Review* 57(2):115–126; Venkatraman, N. 1994. "IT-Enabled Business Transformation: From Automation to Business Scope Redefinition," *Sloan Management Review*, Winter, pp. 73–87.
15. Personal communication with Henry Nelson, Harris Data, April 1995.
16. Personal communication with Mike Kolbe, Cleveland ECRC, March 1995.
17. Goldman, Steven L., and Roger N. Nagel. 1993. "Management, Technology, and Agility: The Emergence of a New Era in Manufacturing," *Interscience Enterprises Ltd.* 8(1/2):18–37.
18. *Ibid.*, p. 29.
19. *Ibid.*, p. 36.
20. Mason, Robert M., A. Thomson, and H. Nelson. 1993. "Implementation of Integrated Information Systems from a Business Perspective," final report on a study performed for the CALS Program Office, Air Force Materiel Command, WPAFB, Ohio 45431. See also Mason, R.M., and H. Nelson. 1993. "Implementation of Integrated Information Systems: Comparisons of Field Studies and the Literature," *Proceedings of the 27th Hawaii International Conference on System Sciences*. Computer Society Press, Los Alamitos, Calif., pp. 987–997.
21. Rogers, E. 1983. *Diffusion of Innovations*. Free Press, New York.
22. *Ibid.*
23. Venkatraman, op. cit.
24. Mangelsdorf, 1994a, and Mangelsdorf, Martha E. 1994b. "Small-Company Technology Use," *INC* 16(12):141.
25. Senge, Peter M. 1990. "The Leader's New Work: Building Learning Organizations," *Sloan Management Review*, Fall, pp. 7–23.
26. Kolb, David. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall, Englewood Cliffs, New Jersey.
27. Dixon, Nancy. 1994. *The Organizational Learning Cycle*. McGraw-Hill, London.
28. Kolb, op. cit.
29. Romano, Claudio A. 1990. "Identifying Factors Which Influence Product Innovation: A Case Study Approach," *Journal of Management Studies* 27(1):76.
30. Rogers, op. cit.
31. Moore, G. 1991. *Crossing the Chasm*. Harper Business, New York.
32. Taylor, James R., Eric G. Moore, and Edwin J. Amonsens. 1994. "Profiling Technology Diffusion Categories: Empirical Tests of Two Models," *Journal of Business Research* 31(1):155–162.
33. Kelley, Maryellen R., and Harvey Brooks. 1991. "External Learning Opportunities and the Diffusion of Process Innovations to Small Firms," *Technological Forecasting and Social Change* 39:103–125.
34. Schroeder, Dean M. 1989. "New Technology and the Small Manufacturer: Panacea or Plague?," *Journal of Small Business Management* 27(3):1–10.
35. Mintzberg, Henry, and James A. Waters. 1985. "Of Strategies, Deliberate and Emergent," *Strategic Management Journal* 6:257–272.
36. Lefebvre, Louis A., R.M. Mason, and E. Lefebvre. n.d. "The Influence Prism in SMEs: The Power of CEOs' Perceptions on Technology Policy and Its Organizational Impacts," submitted and under revision for *Management Science*.
37. Berger, Peter L., and Thomas Luckman. 1966. *The Social Construction of Reality*. Doubleday, Garden City, New York.
38. Preece, David A. 1991. "The Whys and Wherefores of New Technology Adoption," *Management Decision* 29(1):53–58.
39. Lefebvre, L.A., J. Harvey, and E. Lefebvre. 1991. "Technological Experience and the Technology Adoption Decisions of Small Manufacturing Firms," *R&D Management* 21:241–249.
40. *Ibid.*
41. Arrow, K. 1962. "The Economic Implications of Learning by Doing," *Review of Economic Studies* 29:155–173.
42. Parente, Stephen L. 1994. "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory* 63:346–369.
43. Cohen, Wesley M. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly* 35(1):128–152.

44. Kelley and Brooks, op. cit.
45. Virany, Beverly, Michael L. Tushman, and Elaine Romanelli. 1992. "Executive Succession and Organization Outcomes in Turbulent Environments: An Organizational Learning Approach," *Organization Science* 3(1):72–91.
46. Lefebvre, E., Louis A. Lefebvre, and Marie-Josée Roy. 1995. "Technological Penetration and Cumulative Benefits in SMEs," *Proceedings of the 28th Hawaii International Conference on Systems Sciences, Vol. III*. Computer Society Press, Los Alamitos, Calif., pp. 533–541.
47. Kelley and Brooks, op. cit.
48. Argyris, C., and D. Schon. 1978. *Organizational Learning: A Theory of Action Approach*. Addison Wesley, Reading, Mass.
49. Argyris, Chris. 1977. "Double Loop Learning in Organizations," *Harvard Business Review*, September/October, pp. 115–125.
50. Virany et al., op. cit.
51. Helper, Susan. 1991. "Strategy and Irreversibility in Supplier Relations: The Case of the U.S. Automotive Industry," *Business History Review* 65 (Winter):781–824; and Helper, Susan, and David Hochfelder. 1992. "Japanese Style Supplier Relationships in the U.S. Automobile Industry: 1895–1920," chapter in forthcoming book edited by June Akudo for Oxford University Press.
52. Kelley and Brooks, op. cit.
53. Westley, Frances, and Harrie Vredenburg. 1991. "Strategic Bridging: The Collaboration Between Environmentalists and Business in the Marketing of Green Products," *Journal of Applied Behavioral Science* 27(1):65–90.
54. Trist, Eric. 1983. "Referent Organizations and the Development of Inter-organizational Domains," *Human Relations* 36(1):269–284.
55. A "back-of-the-envelope" calculation shows why. If we posit that SMEs employ approximately 6.9 million people and that two-thirds of them are employed at SMEs that are incapable of engaging in electronic commerce today (as noted by the survey cited above), we may assume that these employees are "at risk" of being dislocated or losing their jobs if their firm becomes noncompetitive. If we further posit that the ECRCs and MTCs enable only 10 percent of the employees who otherwise would be dislocated to keep working, and that each one earns at least \$20,000 annually, then the potential savings from unemployment payments alone (one-half of the annual salary) approaches \$4.6 billion—making the DOD and DOC programs seem like a bargain.
56. Mt. Auburn Associates. 1992. "An Evaluation of Ohio's Thomas Edison Technology Centers"; and Commission on Engineering and Technical Systems, National Research Council (NRC). 1990. *Ohio's Thomas Edison Centers: A 1990 Review*. National Academy Press, Washington, D.C.
57. NRC, op. cit.
58. Badaracco, Jr., Joseph L. 1991. *The Knowledge Link*. Harvard Business School Press, Boston, Mass.
59. Porter, op. cit.; and Fogarty, Michael S., and Jar-Chi Lee. 1992. "Manufacturing Industry Clusters as a Logical Way to Structure Technology Deployment," *Center for Regional Economic Issues Report*, Case Western Reserve University, Cleveland, Ohio.
60. Fogarty, op. cit.

43

## Architecture for an Emergency Lane on the NII: Crisis Information Management

*Lois Clark McCoy and Douglas Gillies*

*National Institute for Urban Search and Rescue*

*and*

*John Harrald, NIUSR and George Washington University*

### STATEMENT OF THE PROBLEM

Responses to disasters, both natural and man-caused, are usually multiorganization, multijurisdiction events. This inherent organizational complexity compounds the chances for missed communications, inadequate intelligence regarding the event, and the consequent escalation of both the size and the cost in lives and dollars of that event. For their actions to be consistent, coordinated, and constructive, responders must have a common understanding of the problems they face—a common "mental map." They must also be able to quickly create an organization capable of meeting the disaster-caused needs.

The national information infrastructure (NII) with its new information technologies, properly applied, has the ability to greatly improve disaster preparation and response, thereby reducing the cost of these operations. The technology has the potential of supporting emergency managers in four critical areas: (1) supporting crisis decision making, (2) managing information overload, (3) supporting first responders, and (4) capturing vital information at its source. Within this paper we suggest a strategy to build a command and control, computing, communications, and intelligence (C<sup>4</sup>I) system for civilian crisis management. It will bring with it the new C<sup>4</sup>I doctrine, new emergency organizational networks, and technological standards to ensure seamless interlinking and interoperability. The proposed architecture for the NII that will provide an emergency lane on the information highway is developed in the Recommendations section of this paper.

The implementation of the new technology of information must be combined with a sober realization that immense cultural changes will occur within the emergency management community. If we focus on only the technological aspects of this change and do not consider the users' need to feel in control of the new technology, adequate assimilation of the great benefits to be derived from it could be delayed by years.

### STATE OF PRESENT PLAY

The field of the civil emergency manager is particularly caught up in this technological and cultural change. At the same time that emergencies, disasters, and catastrophes seem to be escalating by the month, the emergency manager is caught between the old and the new. Traditionally, the field of emergency management operated in an area of scarce information about the event, whether hurricane, riot, or chemical spill. Today, emergency managers are overwhelmed with a glut of information. Suddenly, they have neither the training nor the tools to handle this level of data. A well-known paradox of information management is that decisionmakers with too much information act just like decisionmakers with too little information—they make decisions based on their personal experience and expertise, not on an analysis of the current situation. The technology provides no added value if all it does is provide additional information to an already overwhelmed decisionmaker. In the face of this rapid change, emergency response and mitigation have lagged behind the general rate of acceptance of the new information technology. Now, with the support of the Clinton administration and the revitalized Federal Emergency Management Agency, it is full steam ahead. However, the difficulties involved with any dramatic change still remain.

We, living and working at the change of the century, are the "tween" generation. Here is this new technology on the one hand—new, untrusted, and fearsome. On the other hand, there is our intuitive judgment developed through hard-won experience. We are indeed caught on the horns of a dilemma.

History during other times of great change may have a lot to show us about a saner and more harmonious way to survive this frenzied *fin de siècle* we are thrashing through. It is a misconception that the average person is threatened by change. People are threatened only by changes over which they believe they have no control!

Emergency managers are singularly beset by this phenomenon as they are encouraged to adopt new technologies and methods. Most of these managers have had long experience in the field of disaster relief and rely on this knowledge as the basis on which they make their decisions. The more competent the commander and the more complete the knowledge, the better the decision.

But now managers are confronted with a new 21st-century technology—command, control, computing, communications, and intelligence (C<sup>4</sup>I) systems. These systems can provide great support to the emergency manager. They can provide information as concise knowledge that can be quickly understood. They can interpret, filter, and correlate knowledge for making rapid projections and estimates. This is a completely new environment for the emergency manager. Where he or she once operated (and had learned to be comfortable with) uncertainty, now the environment is filled with technologically derived knowledge. Knowledge possessed by experts not at the scene can be made available to the emergency manager either through remote communications or by the use of expert systems. That is a monumental change. It brings with it the often-heard phrase, "That's not how we do it here!" An important part of this rejection of new ways and new tools is a lack of trust. The emergency manager has no sense that he or she owns them and does not feel in control of the technology or of the information it provides.

#### ANALYSIS OF FACTORS INFLUENCING THE REALIZATION OF AN EMERGENCY LANE ON THE NII

After a 3-year program of "town meetings," the National Institute for Urban Search and Rescue (NIUSR) has identified the needs for new-age, interlinking communications in time-sensitive arenas.

A most useful tool in overcoming the lack of trust in new information technology is the use of simulations in exercises and demonstrations. Experience is a great teacher, but where in the real world can one gain that experience in a relatively risk-free environment? In the field of emergency response, too many lives are often in jeopardy to risk trying out new technology in real circumstances. Under stress, emergency managers, even after buying and installing new technology, will tend to revert to their experience and intuitive judgment, which are comfortably familiar and have served them well in the past. One of the statements heard often from the observers in the follow-up days of recovery from the Northridge earthquake was, "We never saw so many computers lining the walls with no one using the information." In times of great stress, we all return to the known ways of management, command, and control.

The new technology continues to become more easily used—and more "fun" to learn. The development of the World Wide Web is only the beginning of an evolution that promises to provide a bridge into the new technology for hard-won experience and intuitive judgment. The seasoned emergency manager now has a tool that lets him or her feel in control. The manager can "see" the links into the information and ask for the piece of information wanted, rather than shifting through reams of data spewing out of a machine. And, most importantly, the manager can see this information displayed in a visual "picture" that allows him or her to quickly assess the situation. The manager has grasped the knowledge within the data.

The first step in achieving an effective conversion of raw data to knowledge is to import that information in a form that is easy to use. "A picture is worth a thousand words" because of the comparable speed with which the brain assimilates one printed word and one picture. With all the picture's background and emphasis, the brain perceives a much greater range of data in microseconds.

The World Wide Web enables the emergency manager to (1) choose the information wanted, and (2) see it displayed on an interactive graphic representation as a picture of the information in the setting of the emergency. To this, add the pre-event simulation training that permits the emergency manager to trust the information being

delivered. At last we begin to convince the experienced, knowledgeable commander that this ability to control and respond to the emergency has been improved to an exponential degree through the sophisticated use of information technology. No longer are we using new technology to merely pave over the cow path (ICMA, 1994). We have moved into another realm of control. We own the future and it is ours.

### THE BARRIER FOR CULTURAL CHANGE

One of the idiosyncrasies of the civilian emergency manager is that both of the words "command" and "control" are forbidden. They send up red flags on the site of any emergency. Who's in charge? has always been a poorly understood question within civil authority. "Battles" can start on Main Street over this question during the disaster itself. So the emergency management community has elected to hide behind the word "coordinate". This euphemism must be discarded. This doublespeak must be reversed. Someone has to bite the bullet and be in charge! Lives are at stake, resources must be expended in increasing amounts, dollars must be spent. Command and control must be present at the disaster scene. Too often, civilian response is a matter of individual heroism—emergency responders, both paid and volunteer, performing over the top in their noble efforts immediately following the event. Uncertainty is the environment of any emergency. The longer this uncertainty remains, the higher the cost in lives lost, property damage, and recovery costs.

Obviously the emergency coordinator must assume command and control. The emergency coordination may be a multilevel, joint command, or a unified command with one spokesman. This type of arrangement must be practiced, simulated, and tested to be quickly effective. One timely adjunct for the civil emergency environment was the adoption, in November 1994, of a national Interagency Command System (ICS) for the multiagency response to large-scale disasters. The National Interagency Command System is built on the original FIRESCOPE and National Interagency Incident Fire Control models of the old Boise Interagency Fire Control Center. It enables the user to limit the uncertainty of the disaster scene quickly. After all, the deciding factor in the successful control of any emergency has always been time. Among other pulses, the new information technology provides the huge advantage of shortened time frames for the development of real information the commander may use in exercising his or her knowledge and judgment. Uncertainty will always be a factor in any response to emergencies. If we have complete control and adequate resources to respond, it is not a crisis! It is merely business as usual for the emergency provider. When there is a lack of trustworthy information and a scarcity of resources, combined with a lack of control, then the true hazards of the emergency environment appear.

One of the continuing problems in the response to any large, multiagency effort has remained the coordination and movement of resources between and among the various responders. The vertical coordination among like agencies (for example, among law enforcement agencies) may proceed with a degree of smoothness. Likewise among the responding fire agencies, coordination tends to remain cooperative and responsive. However, when different agencies such as public works or transportation or social services are brought into the multilevel response, attempts at coordination among dissimilar modes of communication and departmental methods of operation lead to confusion and incompatibility. Further complicating coordination, government and volunteer and nonprofit organizations have different styles of operation. The American Red Cross and the Salvation Army have key roles to play in disaster response. In addition to these formal organizations whose presence is anticipated, it is a well-documented phenomenon that ad hoc groups emerge to deal with unanticipated needs during the disaster response. These organizations do not all have to be tightly linked, but their actions must be consistently coordinated if the needs of the victims are to be met. The inability to coordinate policy, logistics, and operations escalates again when multiple agencies from different jurisdictions and levels of government are involved. The fractionalization of coordinated response is even further adversely affected when military assistance and its chain of command are requested by civil authority, and yet another system is mixed into the command equation.

### FORECAST FOR THE NEXT FIVE TO SEVEN YEARS

A 5- to 7-year forecast is a difficult (and perhaps foolish) undertaking. We do have some current trends, however, that indicate what this future might be. The Internet is growing at 15 percent a month. The World Wide Web is doubling every 53 days. We will take a look at the future, but the traditional rules for predicting future time lines may be another victim of the rate of change in information systems.

One unexpected and highly beneficial spin-off from the cross communication possible on the Internet and the World Wide Web has been the blurring of vertical authority and its consequent concerns for the protection of "turf." Both the Internet and the Web operate across many platforms in a seamless fashion. No one may be sure where or how most of the queries originated. Within the parameters of authorized access, the information is available to all. Interoperability may be the most beneficial of all the various aspects of the new information technology. We have left behind a time when emergencies were synonymous with uncertainty. Now emergencies, and indeed, our daily business lives, are filled with a glut of information. The next big breakthrough in information management will be in software filters such as profilers. Some even predict that there will be an entire new industry of personal media assistants who will filter your information, marking for your attention only that in which you have indicated a major interest. For emergency managers, a system to filter the data glut must have a real time component that may be better served by expert systems, or the reasoning support of knowledge robots ("knowbots"). Such expert systems, simulations, artificial intelligence, intelligent decision support systems, or even some other new tack may be the next exciting advance in the field of information management.

Expert systems are currently pushing the edge of the technology. To date there have been some well-documented failures of attempts at such systems. These failures fall into three classes.

- *Level one.* Errors of commission, in which human operators make a programmatic error of commission—the old garbage in/garbage out problem.
- *Level two.* The programmer forgets to input some data altogether and it goes unnoticed. An example would be the Gemini V space shot that landed 100 miles from where it was supposed to come down. Some programmer forgot to input the motion of the earth around the sun into the re-entry program.
- *Level three.* This is the most difficult, and interesting, of all the sources of error. This is where a group of parameters about the real world in which the program has to operate is input, and where, after these parameters are input, they change. This has been called the "*Sheffield effect*" after the British destroyer sunk in the battle with the Argentines. The ship's program had the Exocet missile (developed by a NATO ally, France) identified as friendly. However, in the arsenal of Argentina, it was decidedly unfriendly, and it sank the *Sheffield*.

Current work on expert systems (not artificial intelligence) is wrestling with two interesting parameters. They are optimal ignorance and appropriate imprecision. Not to delve too deeply into this subject here, let us just say that these are ways to limit the amount of data to something that the machine can handle without thinking itself into a corner and freezing up. The downside of this approach is that the expert system, in many ways, is like an idiot savant. It can do one thing superbly, but only that one thing. The use of parallel processing at a remote site using high-performance computing could provide a solution for the near term. In this way, the smaller storage and performance PCs at the site of a disaster could download the solutions and data summaries provided by high-performance computing on-site processing. Such processing is currently infeasible in real time in the field. Under the current limitations of disk drive storage and the huge amounts of memory needed for image processing and correlation of disparate databases, it is not possible to process such needed information in the field.

And we desperately need another capability for the emergency manager. We need the ability to sort, digest, summarize, and prioritize data into information, and further, into intelligence. We need to be able to convert data into usable decisionmaking tools in nanoseconds. We need the integrated decision tools that will enable an emergency manager to apply this information to the problem at hand.

Information technology has the ability to provide solutions to these old problems, or it will have in the near future. Information technology's true value is in its ability to provide the new while enhancing the old. The new is the rapidity with which real, trusted information can be provided in an easily understood picture for the

commander. The old is the hard-gained experience that provides the knowledge and judgment of the emergency manager.

The ability of the emergency manager to comfortably use the enormous capability of the new technology is not a matter of purchasing the new "goodies" to reside on a shelf as another status symbol. It means learning an entire new set of tools for responding to uncertainty. It means developing a hands-on familiarity to these new tools. Just as the driver of a high-performance race car must learn to skillfully use a stick shift, so must the emergency manager learn to use the tools of the new information technology—not merely learn to use that stick shift, but to be thoroughly comfortable with it and to feel as though it were an extension of his or her own hand.

Some emergency managers will make this effort. Some will not. In all cultural changes, some adapt, and some are overrun by the change. Some continued to lay keels for four-masted schooners after Fulton invented the steam engine. Some continued to make buggy whips after the coming of Henry Ford's Model T. Our emphasis within the NII must be on those millions who are jumping onto the Internet and the Web every day. We must focus on those who take the leap of faith into the new. Every change is cluttered with gatekeepers whose very existence is intended to keep change from happening.

At every crossway on the road to the future, each progressive spirit is opposed by a thousand men appointed to guard the past.

—Maurice Maeterlinck, Nobel Laureate

Rather than worrying about those who will be left, we must focus our efforts on making these new technologies comfortable for those who are joining the new generation. Almost everything negative that one can hear about the computer revolution was said 50 years ago about the radio. All the dire predictions of the cultural changes brought about by the radio—for example, listening to baseball games instead of going down to Wrigley Field—never came true. Baseball is still baseball.

There is now an entire subset of people who are as comfortable with a computer as they are with a telephone. They are lost without it. Millions are carrying smaller and smaller versions of their business and personal lives with them—in the car, on airplanes, and on vacations. Shortly, all those with a computer are going to have to buy another, more powerful one. The World Wide Web, multimedia, graphics, and all those wonderful (and let us not forget, fun) things eat disk space. They need enormous increases in the speeds of transmission. The market is huge.

### **RECOMMENDATIONS: AN ARCHITECTURE FOR THE EMERGENCY LANE ON THE NII**

The National Institute for Urban Search and Rescue has developed an architecture that has great promise. It is simple, involves off-the-shelf components, incorporates tested systems already in place, and uses new technological applications. It is the development side of the equation rather than research that now needs attention and focus.

#### **The Crisis Communications Management Architecture**

Central to the concept of the crisis communications management system is the NII initiative that can provide the opportunity to make dramatic steps forward in crisis information handling. Let us envision this new information and communications support system as succeeding grids laid one on the other. The first layer of the grid is in space. It consists of space imaging and sensor information in all its multiforms. The second layer is a seamless communications grid, transparent to the operator. This communications grid contains virtual networks within the tactical area as well as national networks. The third and final layer is the tactical grid from which we conduct the crisis operations.

How then do we link all this information into one understandable "picture" of the crisis for decisionmakers? To do this our communications architecture must provide both decisionmakers and tactical users with a picture of the crisis that artificially replicates the reality of the emergency. This suggested civilian crisis

communications management system is patterned upon the successful Copernicus architecture of the U.S. Navy (USN, 1993). The brain child of Vice Admiral Jerry O. Tuttle (now retired), Copernicus was developed while Tuttle was director for naval space and electronic warfare. It is considered a communications milestone for the Information Age. Interestingly, the similarities in operational needs and constraints between electronic warfare and crisis management are greater than the differences. The most obvious difference, other than scale, is that the end result of crisis management is safety and recovery, rather than destruction. Although the end result is different, the methods of communication and information exchange are remarkably similar.

The crisis communications management system must supply a doctrinal, organizational, and technological management system applicable across all functions, agencies, and organizations of the crisis management response, regardless of their position in it. The system must be readily adaptable across the levels of participation as the response either escalates or decreases. It must also incorporate the widest space imaging system, including the global positioning system, space imaging and interpretation, weather surveillance, and so on. The interfaces must be synergistic and seamless across the disaster area and operationally transparent to the user regardless of the affiliation, level of response, or component. In addition, the system must integrate command and control, information processing, resources and transportation, levels of responsibility, tactical operations, and on-scene data. Such capabilities must be coordinated across the crisis arena and vertically up and down the levels of stakeholders.

### **The Communications Grid**

Central to the communications grid are the wide-area computer networks that link the commands and activities of the decisionmakers and stakeholders to the response activities at the scene of the disaster. They are configured on a regional or operational area basis and are constructed to transport, standardize, and concentrate sensor, analytic, command, support, administrative, and other data for further passage to incident commanders.

The communications grid will use current and planned common-user communications systems, such as the evolving national communications infrastructure, and the present interlinking media communications networks for multimedia communications. The National Institute for Urban Search and Rescue believes that the emergency environment will become far more data-intensive and require far more technological agility in obtaining, handling, and transmitting data than we have experienced.

A second and equally critical development over the last 10 years has been the growth of small computers, both personal computers and workstations. Although the latest growth in computers has been astronomical, we see an even greater increase in their speed and a great reduction in size. Hand-held computers will soon be in the field with nearly the same processing capability as today's desktop configurations. The incident commander does not necessarily want to sit at a screen and pull up windows. He or she will likely want a "telestrator"-type of system that enables writing on the screen and placing objectives visually to be transmitted. We must be sure that this rapid growth supports industry standards and open systems architecture.

The establishment of "information highways" and the movement toward open systems architecture make possible the aggregation of many disparate agencies and organizations. These entities, potentially involved in catastrophes and disasters, are defined not by physical boundaries but by data addresses and a common software "veneer."

### **The Regional Hubs**

The regional hubs are regionally oriented and contain the major sensor and analytic nodes, both for state and national data. The number and nature of the regional hubs are intended to be dynamic so that the architecture can support the particular desires and needs of the area. For example, to construct a logistics, weather, planning, and/or contingency system simply means developing a software veneer for the common hardware "engines" envisioned as the building blocks of the communications grid. We can also envision contingency hubs as well as the major regional hubs.



## The Command and Control Complexes

A second part of the communications grid includes the command and control complexes. For those agencies with significant areas of responsibility for civil emergency operations, such as the U.S. Air Force as executive agent for inland search and rescue, the U.S. Coast Guard in its responsibility for the maritime area, and the Federal Emergency Management Agency for its role in natural and man-made disaster, command complexes will be dedicated. Significant differences exist between the regional hubs and the command complexes. The regional hubs are an aggregation of "communities of interest"; the command complexes are an aggregation of command structures in the particular area of responsibility.

## The Information Centers

The third part of the communications grid consists of operational information constructs, not communications networks. The regional hubs and the command complexes will share a consistent tactical picture through this series of information constructs. Like the regional hubs and the command complexes, the information centers are not physical but virtual "nets," established at the request and in the mix desired by each incident commander.

The information contained in a single node may be provided via several communications channels or vice versa. These information centers spring from an operational decision about where to send data between the emergency and the regional hub. These nodes will be thought of in three conceptual planes:

- The technological and custom protocols for the exchange of information for technical applications;
- The operation data layering—that is, the doctrinal decision to place the data on a particular distribution network and route it to a particular incident commander's workstation; and
- The transformation of data to information, which is a function of the software interface on the tactical computers.

Communications consultant Charles R. Morris writes in the *Los Angeles Times* that the term "information highway" may not be the most appropriate description for the wireless infosphere of innumerable paths of information to any destination. He believes that "ocean" more realistically describes the process—one where all data perpetually circulate until searched out and plucked down by an intelligent agent embedded in each personal data assistant (Campen, 1994a).

The information centers may support eight formations of communications services and three cases of precedence. Radio frequency (RF) communications will be undetectable, except to the designated and cooperative receiver. A glimpse of the near future includes RF signals that will all operate on top of each other, below the noise and with featureless wave forms. Parasite information will ride on carriers, and antennas will be broadband, high gain, and electronically steerable. They will be used to access multiple satellites simultaneously in various orbital planes along with terrestrial high-capacity data links (Busey, 1994).

The number of information centers will not be fixed. Instead, they will be connected for the length of time necessary to transport the data to the users for the incident and then truncate.

The information centers are classified into three broad categories—a menu is a good analogy—by "communities of interest."

- *Information Center—Command* will service the decisionmakers. These information centers are envisioned as multiformat, including teleconferencing.
- *Information Center—Support* will include such pathways as an environmental, logistics, database-file transfer, imagery, geographic, and narrative message pathway. This is the only information center that is envisioned as carrying narrative.
- *Information Center—Tactical* will be constructed around the tactical needs of the response forces. Since the variety of emergencies can be quite different and are dependent on the mission and the resources at

*hand, these centers are seen to be a mixture of both predetermined information services (perhaps associated with Standing Operations Procedures) and virtual information services created dynamically only for the duration of a particular event. These centers can comprise both preexisting and user-created information groups.*

### The Tactical Grid

The tactical grid is conceived as a wide-area system, a network of small communications links that tie all units operating in one tactical area together regardless of the agency or function. The differences between the communications grid and the tactical grid are functional. The communications grid provides connectivity, which facilitates the movement of information among operators and analysts. The tactical grid, alternatively, connects systems among operational units to provide information across the scene of the crisis. A good analogy for the tactical grid is a power grid. When computers of different makes and operating systems are plugged into electric power outlets, they get a common energy system. By connecting dissimilar systems across the tactical grid, we are connecting platform sensors, fire, medical, and law enforcement units, main computers and electronic subsystems, and so on.

Thinking of the diversity of sensors, communications, and resource systems as grids overlying the tactical crisis arena provides a readily understandable way of viewing the myriad of assets and stakeholders in the crisis communications infrastructure. Operationally, the impact is that dissimilar resources and/or units can be connected in the tactical grid, imposed over the operating area as though they were joining a regional power grid. This link would occur simultaneously, allowing the operators of those resources to plug into the space and communications grids.

The architecture of the tactical command center (TCC) is intended to serve as the "nerve center" for the incident commander and his units in the response arena. This means that the TCC is not only the intelligence center for tactical command, but also the tactical center for individual units and the multiagency incident command areas (MACS). The TCC provides the tactical displays, integrated information management, and accessibility to tactical communications in support of the response missions. It provides the requisite tactical connectivity to units, to other area commanders, and to the command complex of the decisionmakers. Architecturally the TCC is analogous to the command complex. Both will share a consistent tactical picture and connect responsible agencies to the stakeholders, at the tactical level and regional levels.

### CONCLUSIONS

The essence of crisis management is an effective information-handling capability. Command must have it; analysis must have it; tactical operators must have it. Without a true picture of the emergency event, we are playing 1990s scenarios with 1940s technology. Why do we continue to do this? Possibly it is because, for many years, a lack of robust communications has prevented obtaining any understandable "picture" of emergency incidents. Routinely, it was three and sometimes four days before the full scope of a disaster became apparent. Responders did what they could—saved as many lives as possible, shored up as many levees as appeared to be endangered, housed and fed survivors, and hoped that the body count would not be too high when the water receded. Then came the invention of the Minicam, the personal computer, and cellular communications and their proliferation in every aspect of our lives.

Now emergency managers realize that it is possible to obtain a rapid and clearer picture of a disaster. You can watch from space in real time the rise of the Mississippi or a hurricane bearing down on the Florida coast. You can compare snow loads in the Sierra this year with measurable images from last year. You are a visceral participant in each and every disaster through the convenience of CNN! And yet, we still have not applied these tools and capabilities to the actual command and control of emergency response operations.

It is definitely past wake-up time for emergency managers. If they can think of fire trucks, helicopters, and ambulances as "rescue platforms," the electronic equivalent is the communications platform. It is at this level

that systems architecture, programs, and technology converge in a level of detail that does not require specialized expertise.

It is in this tier that critical paths in operations, programming, and technologies converge and become most obvious to managers. This idea that electronic platforms are as essential as other rescue platforms will allow us to approach design and installation of the Crisis Communications Management system. At its simplest, the advent of crisis management is both the recognition of the requirement and the means to operate in the electromagnetic spectra and in space against the increased calamities threatening our traditional economy.

When we think operationally of the Crisis Communications Management architecture as having a space grid of diverse sensors, supported by a dynamic, multinode communications grid, we are conceiving and operating the complex and geographically disparate electronics of today's emergency response as a system to present the crisis scene as it really appears, containing all relevant information in a transparent, easily understood format.

It is the C<sup>4</sup>I system, designed to make communications transparent to the user and all sensors available in common formats, that allows us to conceive of the space and communications grids and of information movement between them. The Crisis Communications Management system has been designed to include both local and wide-area networks that have tied different systems and hardware together, along with higher bandwidth communications capabilities and more efficient software. We have reengineered the work processes for improved information handling.

During the twentieth century, disaster response has moved out of the trenches of land-based operations. With the addition of aircraft to traditional rescue platforms, there was an exponential increase in the space over which we could travel to perform rescue operations. But with this increase in the area of operations, commanders lost the ability to view the catastrophe as a whole. Now, with our latest technology, we can again provide the incident commanders with a visual picture of the scene of operations. To illustrate this, our paper ends with a story quoted from the U.S. Navy's publication *SONATA* (USN, 1993), a presentation of the plan for naval communications through the next century.

Two hundred years ago when Lord Nelson walked out on the deck of HMS Victory, the tactical battle space was obvious. He could see it and share that perception both with his captains and his enemy. The advent of carrier air power in World War II changed that. Because a commander can no longer see the battle space, perhaps hundreds of miles away from where he stands, it must be artificially reconstructed for him.

Today that reconstruct is accomplished by messages arriving over different networks, in diverse formats and with different time delays. Electronic communications, imagery and radar systems, until recently were displayed on separate screens open to mismanagement. Reality was replaced by an artificial view that was too complex, too redundant, and too slow.

Now we think operationally of the battle space as having diverse sensors, supported by a dynamic, multiconstellation communications grid. We are (whether we recognize it or not) conceiving and operating the complex and geographically disparate electronics of modern warfare as a system to present the battle space as it really appears—just as it did to Nelson.

The above paragraphs relate to the use of advanced systems designed for electronic communications support in time of war. These same technologies are available to civilians. Although a few of the concepts are still in development, many are "off the shelf" today. Why aren't we using them to save more of our own people? It seems worth doing to us. To again quote Adm. Jeremiad, "I urge you to leap into another dimension of capability by simultaneously expanding the window of our technological might and our strategic thought." We say, press on!

## REFERENCES

- Busey IV, USN (ret.), Adm. James B. 1994. "Navy's Space, Electronic Warfare Vision Mirrors New Defense Budget Framework," *Signal Magazine*, AFCEA International Press, March.  
*Business Week*. 1995. "Planet Internet," a special report.

- Campen, USAF (ret.), Col. Alan D. 1994a. "Competition, Consumers Are Shaping Information Highway," *Signal Magazine*, AFCEA International Press, March.
- Campen, USAF (ret.), Col. Alan D. 1994b. "Technology Trumps Policy in Information," *Signal Magazine*, AFCEA International Press, February.
- Elliott, Ronald D., and Maj. Scott Bradley, USMCR. 1995. "Effective Command and Control," *JWID '95*.
- Kelly, Maj. Brian J. 1993. "From Stone to Silicon: A Revolution in Information Technology," unpublished white paper available from the Armed Forces Communications and Electronics Association (AFCEA) Educational Foundation, December.
- Harrald, John R. 1993. "Contingency Planning Using Expert Judgment in a Group Decision Support Center Environment," unpublished white paper available from George Washington University, Department of Engineering Management.
- Harrald, John R., and T. Mazzuchi. 1993. "Planning for Success: A Scenario Based Approach to Contingency Planning Using Expert Judgment," *Journal for Contingencies and Crisis Management*.
- International City/County Management Association (ICMA). 1994. "Computer Technology in Local Government, Second Survey," *Government Technology*, March.
- Jeremiah, Adm. David. 1994. Unpublished presentation at "WEST '94", AFCEA and U.S. Naval Institute Conference and Exposition, January.
- Linden, Eugene. 1994. "Burned by Warming," *Time Magazine*, March 14.
- Rockhart, J.F. 1981. "The Changing Role of the Information Systems Executive: A Critical Success Factors Perspective," *Sloan Management Review*, pp. 15-25.
- Signal Magazine*. 1994. "Consortium Sows Seeds for Information Infrastructure," February.
- Starbuck, W.W. 1989. "Clio's Conceit: Looking Back into the Future," unpublished presentation at the Second International Conference on Industrial and Organizational Crisis Management, New York, November.
- U.S. Navy (USN), Office of Space and Electronic Warfare. 1993. *Sonata* .

## 44

### Aspects of Integrity in the NII

*John C. McDonald*

*MBX Inc.*

The national information infrastructure (NII) has a variety of definitions and gives rise to various perceptions. One such perception is that the NII is a rapidly expanding network of networks that, in totality, achieves the national objectives set forth by Congress and the Clinton administration. No matter what the definition, this telecommunications infrastructure is a key element of the expanding information revolution that we are currently experiencing. Plans are being made by many sectors of the economy to further use the resources of the NII to improve productivity, reduce cost, and maintain a competitive edge in the world economy. Included among the sectors with growing reliance on the NII are health care, education, manufacturing, financial services, entertainment, and government.

With society's increasing reliance on the NII, there is a corresponding need to consider its integrity. Integrity is defined as "the ability of a telecommunications infrastructure to deliver high quality, continuous service while gracefully absorbing, with little or no user impact, failures of or intrusions into the hardware or software of infrastructure elements<sup>1</sup>."

Integrity is an umbrella term that includes other important telecommunications infrastructure requirements such as quality, reliability, and survivability.

This paper discusses various dimensions of integrity in the NII. Threats to integrity are presented and lessons learned during the past decade summarized, as are efforts currently under way to improve network robustness. Finally, this paper concludes that architects and designers of the NII must take issues of integrity seriously. Integrity must be considered from the foundation up; it cannot be regarded as a Band-Aid.

#### THREATS TO NII INTEGRITY

Network elements can fail for any number of reasons, including architectural defects, design defects, inadequate maintenance procedures, or procedural error. They can fail due to acts of God (lightning, hurricane, earthquake, flood), accidents (backhoe, auto crashes, railroad derailment, power failure, fire), or sabotage (hackers, disgruntled employees, foreign powers). Architects and designers of the NII should weigh each of these threats and perform cost-benefit studies that include societal costs of failure as well as first-time network costs. Users of the NII should understand that failures will occur and should have contingency plans.

Over the past 10 years, public networks in the United States have experienced failures resulting from most of the threats described above. In May 1988, a fire in the Hinsdale, Illinois, central office disrupted telecommunications services for 35,000 residential telephones, 37,000 trunks, 13,500 special circuits, 118,000 long-distance fiber optic circuits, and 50 percent of the cellular telephones in Chicago<sup>2</sup>. Full service was not restored for 28 days. The failure affected air traffic control, hospitals, businesses, and virtually all economic sectors. Two months later, technicians in Framingham, Massachusetts, accidentally blew two 600A fuses in the Union Street central office. The local switch stopped operation, and calls from 35,000 residential and business customers were denied for most of the day<sup>3</sup>.

In November 1988, much of the long-distance service along the East Coast was disrupted when a construction crew accidentally severed a major fiber optic cable in New Jersey; 3,500,000 call attempts were

blocked<sup>4</sup>. Also in November 1988, a computer virus infiltrated the Internet, shutting down hundreds of workstations<sup>5</sup>.

Several well-publicized SS7 outages occurred in 1990 and 1991 due to software bugs<sup>6,7</sup>. The first had a nationwide impact and involved the loss of 65,000,000 calls. Others involved entire cities and affected 10,000,000 customers.

In response to a massive outage in September 1991, the mayor of New York established a Task Force on Telecommunications Network Reliability. The task force noted that "the potential for telecommunications disasters is real, and losses in service can be devastating to the end user"<sup>8</sup>.

### LESSONS LEARNED THAT ARE APPLICABLE TO THE NII

Network infrastructure architects and designers have used redundancy and extensive testing to build integrity into telecommunications networks. They have recognized the critical role that such infrastructure plays in society and are mindful of the consequences of network failure. Techniques such as extensive software testing, hardware duplication, protection switching, standby power, alternate routing, and dynamic overload control have been used throughout the network to enhance integrity.

A 1989 report published by the National Research Council identified trends in infrastructure design that have made networks more vulnerable to large-scale outage<sup>9</sup>. Over the past 10 years, network evolution has been paced by changes in technology, new government regulations, and increased customer demand for rapid response in provisioning voice and data services. Each of these trends has led to a concentration of network assets. Although additional competitive carriers have been introduced, the capacity of the new networks has not been adequate to absorb the traffic lost due to a failure in the established carrier's network. End-user access to all carriers has been limited by this lack of familiarity with use of access codes.

Economies of scale have caused higher average traffic cross sections for various network elements. Fiber optic cables can carry thousands of circuits, whereas copper cables carried hundreds. Other technologies such as microwave radio and domestic satellites have been retired from service in favor of fiber. When a fiber cable is rendered inoperable for whatever reason, more customers are affected unless adequate alternate routing is provided. The capacity of digital switching systems and the use of remote switching units have reduced the number of switches needed to serve a given area, thus providing higher traffic cross sections. More customers are affected by a single switch failure.

In signaling, the highly distributed multifrequency approach has been replaced by a concentrated common channel signaling system. Also, call processing intelligence that was once distributed in local offices is now migrating into centralized databases.

Stored program control now exists in virtually every network element. Software technology has led to increased network flexibility; however, it has also brought a significant challenge to overall network integrity because of its "crash" potential. Along with accidental network failures, there have been a number of malicious attacks, including the theft of credit cards from network databases and the theft of cellular electronic security numbers.

In regulation, the Federal Communications Commission has mandated schedules for the introduction of network features such as equal access. For carriers to meet the required schedules, they chose to amalgamate traffic at "points of presence" and modify the software at a small but manageable number of sites to meet the imposed schedules. Hinsdale was one such site and, unfortunately, the fire's impact was greater than it would have been without such regulatory intervention because of the resulting traffic concentration.

In my opinion, the most important lesson learned in the recent past regarding telecommunications infrastructure integrity is that we must not be complacent and assume that major failures or network intrusions cannot happen. In addition to past measures, new metrics must be developed to measure the societal impact of network integrity and bring the scientific method of specification and measurement to the problem<sup>10</sup>.

Another lesson learned is that design for "single-point failures" is inadequate. Fires cause multiple failures, as do backhoe dig-ups, viruses, and acts of God. There has been too much focus on individual network elements and not enough on end-to-end service.

Software is another issue. We have learned that testing software to remove all potential bugs is difficult if not impossible. Software does not wear out like hardware, but it is a single point of failure that can take down an entire network. Three faulty lines of code in 2.1 million lines of instructions were enough to cripple phone service in Washington, D.C., Los Angeles, and Pittsburgh in nearly identical failures between June 26, 1991, and July 2, 1991.

### IMPROVING NETWORK ROBUSTNESS

In recent years, efforts to improve network robustness have been redoubled. In addition to the work of individual common carriers, there are many organizations that are addressing these problems, including Bellcore, the National Security Telecommunications Advisory Committee, the FCC, the Institute for Electrical and Electronics Engineers, and American National Standards Institute Committee T1.

Exhaustive testing of new systems and new generic software programs has been instituted by manufacturers and by Bellcore. New technologies have been applied, including "formal methods." New means have been developed and implemented to try and detect "bugs" that previously would have gone undetected.

New network topologies have been implemented using bidirectional SONET rings and digital cross-connect systems. The concept of design for single-point failure has been supplemented to include multiple failures. In cases where economical network design calls for elimination of already sparse network elements, robustness has become a consideration, and the reduction has not occurred.

New metrics have been established to quantify massive failures and reporting means have been implemented by the FCC. Standards have been set to quantify the severity of network outages.

Means have been implemented to detect the theft of cellular electronic security numbers, and new personal identification numbers have been used. There is increased awareness by the employees of common carriers of the need for protection of codes used to access proprietary databases and generic software.

Over the next 2 to 5 years, infrastructure robustness will be enhanced through new procedures and network elements that will soon be in production. Products deploying asynchronous transfer mode (ATM) will give more flexibility in restoring a damaged network. More parallel networks will be deployed which, if interoperable, will add new robustness to the NII.

Current and planned research will enhance NII robustness in the 5- to 10-year window. Some of the research topics were recently summarized in the *IEEE Journal of Selected Areas in Communications*<sup>11</sup>. Open issues addressed in this issue included user survivability perspectives on standards, planning, and deployment; analysis and quantification of network disasters; survivable and fault tolerant network architectures and associated economic analyses; and techniques to handle network restoration as a result of physical damage or failures in software and control systems. These subjects were organized into four categories: user perspectives and planning; software quality and reliability; network survivability characterization and standards; and physical layer network restoration, ATM layer network restoration, network layer restoration, and survivable network design methods.

### CONCLUSIONS

Over the past decade, we have learned many important lessons in the design of telecommunications infrastructure that are applicable to the NII. Although past networks have been designed with high levels of integrity in mind, these efforts have not completely measured up to the expectations of society. Recently, efforts have been redoubled to improve network robustness.

As the NII is defined, it is important that integrity issues be considered from the ground up. Only by these means will an NII be constructed that meets the expectations of society.

## Bibliography

1. Private communication with W. Blalock, Bell Communications Research.
2. National Communications System. 1988. "May 8, 1988, Hinsdale, Illinois Telecommunications Outage," Aug. 2.
3. Brown, B., and B. Wallace. 1988. "CO Outage Refuels Users' Disaster Fears," *Network World*, July 11.
4. Sims, C. 1988. "AT&T Acts to Avert Recurrence of Long-Distance Line Disruption," *New York Times*, November 26.
5. Schlender, B. 1988. "Computer Virus, Infiltrating Network, Shuts Down Computers Around World," *Wall Street Journal*, November 28.
6. Fitzgerald, K. 1990. "Vulnerability Exposed in AT&T's 9-Hour Glitch," *The Institute*, March.
7. Andrews, E. 1991. "String of Phone Failures Reveals Computer Systems' Vulnerability," *New York Times*, July 3.
8. City of New York. 1992. "Mayor's Task Force on Telecommunications Network Reliability," January.
9. National Research Council. 1989. *Growing Vulnerability of the Public Switched Networks: Implications for National Security Emergency Preparedness*. National Academy Press, Washington, D.C.
10. McDonald, J. 1994. "Public Network Integrity—Avoiding a Crisis in Trust," *IEEE Journal on Selected Areas in Communications*, January.
11. *IEEE Journal on Selected Areas in Communications*, January 1994.



## 45

# What the NII Could Be: A User Perspective

*David G. Messerschmitt*

*University of California at Berkeley*

### ABSTRACT

The national information infrastructure (NII) is envisioned as a national public internetwork that encompasses existing networks, such as the Internet, the public telephone network and its extensions, and CATV distribution systems and their extensions, as well as new network technologies yet to be invented. Today, these networks appear to the user to be separate and noninteroperable, in the sense that a user cannot reasonably make a telephone call over the Internet or most CATV systems, cannot reasonably watch video over the Internet or the telephone network (except at unacceptably poor levels of quality by entertainment standards), and cannot send data over the telephone network or most CATV systems (except in the limited sense of using these media for access to data networks or for point-to-point data transmission). It is clear that underlying the NII will be a collection of proprietary networks incorporating a variety of different technologies; indeed, there is general agreement that this is highly desirable. The question addressed in this white paper is what the NII will look like from the user perspective, and how it might differ from today's limited-functionality and noninteroperable networks. We address this question by describing a vision of what the NII could be from a user perspective. In particular, we describe those characteristics of the NII that we believe will be important to users, including connectivity and mobility, quality of service options, security and privacy, openness to new applications across heterogeneous transport and terminal environments, and pricing.

### INTRODUCTION

This white paper is an outgrowth of the planning workshop organized by the NII 2000 Steering Committee. Representatives of a number of industries participating in the NII and its underlying technologies were present. Not surprisingly, given the great variety of industries and their respective largely independent histories and markets, the representatives were often "talking past" one another, not sharing a common vision of what the NII should be, and not sharing the common vocabulary necessary for productive discussion.

In the deployment of a massive infrastructure such as the NII, there is great danger that near-term tactical decisions made by the diverse participants in the absence of a long-term strategic vision will result in an infrastructure that precludes the broad deployment of unanticipated but important applications in the future. Such an infrastructure will not meet the needs of the users and the nation, and will offer its builders a lower return on investment that would otherwise be possible. It might even result in widespread abandonment of existing infrastructure in favor of new technologies, in similar fashion to the recent widespread and costly abandonment of partially depreciated analog communications facilities.

In this white paper, we take the perspective of the users of the future NII and ask fundamental questions about how it should appear to them. It is our belief that, near-term corporate strategies aside, an NII that best meets the future needs of the users will be the most successful, not only in its benefits to society and the nation, but also in terms of its return on investment. Thus, the full spectrum of industrial and government participants should have a shared interest in defining a strategic vision for the long term, and using that vision to influence near-term business decisions.

Looking at the NII from a long-term user perspective, we naturally envision a network that has many capabilities beyond those of any of the current networks or distribution systems. Provisioning such a broad range

of capabilities would have cost implications and is economically feasible only to the extent that it provides value to the user well in excess of the incremental costs. This is problematic if one accepts one of our fundamental hypotheses, namely, that we cannot possibly anticipate all the big-hitting applications of the NII. However, it should be emphasized that it is not necessary that all near-term deployments provide all the capabilities incorporated into a strategic vision. Indeed, one critical aspect of such a vision is that it should be easy and cost effective to add new technologies and capabilities to the NII as unanticipated applications and user needs emerge. If this is achieved, it is only necessary that near-term investments be compatible with a long-term strategic vision, and hence not preclude future possibilities or force later disinvestment and widespread replacement of infrastructure. This is admittedly not straightforward but is nevertheless a worthwhile goal.

One can anticipate the NII falling somewhere on the spectrum from a collection of proprietary and noninteroperable networks (largely the situation today) to a single, universal network that appears to the user to seamlessly and effortlessly meet all user needs. We argue that from the user perspective the NII should, although consisting internally of a diversity of heterogeneous transport and terminal technologies, offer the seamless deployment of a wide range of applications and openness to new applications. Not all participants in the NII may judge this to be in their best interest, and of course they all encounter serious cost and time-to-market constraints. However, if they take into account longer-term opportunities in the course of their near-term business decisions, we believe that both they—the users—and the nation will benefit greatly in the long term. It is our hope that the NII 2000 technology deployment project will move the collective deliberations in this direction.

### TERMINOLOGY

First we define some consistent terminology for the remainder of this white paper.

The users of the NII are people. The NII will consist of a network (or more accurately a collection of networks) to which are attached access nodes at its edge. We distinguish between two types of devices connected to access nodes: information and applications servers, and user terminals (for simplicity, we will abbreviate these to servers and terminals). A networked application is a set of functionality that makes use of the transport services of the network and the processing power in the servers and terminals, and provides value to users. Servers make databases or information sources available to the terminals, or provide processing power required to provision applications. Users interact directly with terminals, which provide the user interface and may also provision processing power or intelligence in support of applications. Examples of terminals are desktop computers, wireless handheld PDAs, and CATV set-top boxes.

There are two generic classes of applications: user-to-user or communications applications, and user-to-server or information access applications. These can be mixed, for example, a collaborative application that combines voice telephony with database access.

The business entities involved in the operation of the NII are network service providers, who provision the transmission and switching equipment in the network, and application service providers, who provision the servers and maintain the databases involved in the applications. These may be one and the same, as is the case for the telephone application in the public telephone network. The users may be the application service provider, as when they load software purchased at a computer store on their terminals. Other entities involved are the equipment vendors, who develop, manufacture, and market the equipment (transmission, switching, terminals, etc.), and the application vendors, who develop and market applications for deployment in the NII.

### CONNECTIVITY ISSUES

#### Logical Connectivity of a Network

The most basic property of a network from a user perspective is the logical connectivity it offers. The network is said to provide logical connectivity between two access nodes if it is feasible to transport data between those nodes through the network. When one access node sends data to another access node, we call the former the

source and the latter the sink. It may be the case that each logically connected access node is simultaneously a source and a sink (a duplex logical connection) or that one may be exclusively a source and the other exclusively a sink (simplex logical connection).

Logical connectivity should be distinguished from network topology. The topology refers to the physical layout of the transmission media used in the network (coax, wire pairs, fiber, radio). Examples are the star topology of the public telephone network and the tree topology of a CATV system. The logical connectivity is determined not only by the topology, but also by the internal switching nodes. Generally, the user is not directly concerned with the topology of the network, although some of the important characteristics of the network (like throughput and quality of service; see below) are affected or constrained by the topology. On the other hand, the network service provider is critically concerned with the topology, as it affects costs.

An important distinction is between the possible logical connections in a network (which may be astronomically large), and the actual provisioned logical connections required by a particular application (typically small in number). A similar distinction must be made between the possible applications (i.e., those that have been developed and made available to users) and those that are actually in use at a particular time. An actual application in use is called an instance of that application, and the actual provisioned logical connections in use by that application are called instances of connections.

### Application Connectivity

There are several important types of connections that arise in the context of specific applications:

- A logical point-to-point connection, in which access nodes are connected in either simplex or duplex fashion. One node in a point-to-point connection may be a source or sink or both, the latter in a duplex connection.
- A logical broadcast connection, in which a single source is connected to two or more sinks. Within the network, this type of connection can be provisioned in different ways. Simulcast implies separate component connections from source to each sink, and multicast refers to a tree structure (where network resources are shared among the component connections). The distinction between these alternatives is generally not of immediate concern to users, who see only indirect effects (cost, quality of service, etc.).
- A logical multisource connection, in which two or more sources are connected to a single sink. A distinction analogous to multicast vs. simulcast does not apply to multisource, since there is generally no advantage to sharing resources among the components of a multisource connection.

Multicast or multisource connections are by their nature simplex. If there are only two access nodes, the connection is necessarily point-to-point. If access nodes are involved, and if for example every access node can send information to and receive information from the remaining nodes, then the connectivity can be thought of as a combination of simplex multisource connections (one to each node) and simplex multicast connections (one from each source). Many other combinations are possible.

From a technology standpoint, multisource connectivity merely requires flexibility in the number of simultaneous point-to-point connections to a given sink, which is a natural capability of packet networks. Similarly, simulcast connectivity requires flexibility in the number of simultaneous point-to-point connections to a source. Multicast connectivity, on the other hand, while beneficial in its sparing use of resources and the only scalable approach to broadcast, requires fundamental capabilities anticipated in the design and provisioning of the network.

## USER PERSPECTIVE

### Connectivity

From the user perspective, it is desirable to have full logical connectivity in a network. Any limitations on connectivity restrict the functionality and availability of both information access and communications applications. For example:

- A user who purchases a telephony application from one application service provider wants the option to call all other telephones, whether they are connected to the telephone network, a CATV network, the Internet, etc. Any application service provider who restricts destinations, let's say to only its own subscribers, will be at a disadvantage.
- The telephone example extends readily to other communications applications. The user will find much less value if the application supplier limits connectivity to a proper subset of those other users who could participate in that application (i.e., who have appropriate terminals, etc.).
- A user with appropriate terminals to access a type of information access application naturally desires connectivity to every available instance of that type of application. For example, a user with the terminal capability to view a video presentation would prefer to maximize the leverage of the investment in terminal equipment by having access to the maximum range of source material.

Similarly, the user would like to see all three types of connections (point-to-point, broadcast, and multisource), since eliminating any one of them will preclude valued applications. For example:

- A "conference telephone call" and "video teleconference" are examples of communications applications that require both multisource and broadcast connections. They are multisource because one participant will want to see and/or hear two or more other participants simultaneously. They are broadcast because any one participant will want to be seen by all the other participants.
- A remote learning class or seminar requires broadcast connectivity because many participants may want to see the presentation, and may also be multisource if the participants have audio or video feedback to the instructor.
- The UNIX X-windows graphical user interface illustrates the value of running applications on two or more servers and displaying the results on a single terminal. This requires multisource connectivity.

Network or applications service providers may view it as in their best interest to restrict the range of applications, information servers, or application service providers that they make available to their subscribers. However, the experience of the computer industry makes it clear that users will choose options with greater flexibility, given the choice and appropriate pricing. For example, restricted-functionality appliances such as the stand-alone word processor quickly lost market share to the personal computer, which offered access to a broad range of applications.

Conversely, in an environment with greater logical connectivity, it becomes more economically viable for new and innovative applications to reach the market. Application service providers with access to a broad range of users (not restricted to the limited market of subscribers to a particular service provider) quickly exploit their economies of scale. Again, the computer industry offers valuable lessons. The personal computer made available an embedded large market for new applications running on widely deployed terminals. Applications vendors targeting the most widely deployed architectures gained the upper hand because of the larger development investments they were able to make.

In conclusion, greater logical connectivity and more connectivity options offer more value to users and hence make the network service provider more economically viable; in addition, there are natural market forces that favor application service providers that target those high-connectivity networks.

### Mobility

The classification of connections is simplest to apply where users are in fixed locations. Users are actually mobile. They may be satisfied with accessing the network from a fixed location, which implies that they can access it only at those times they are physically in that location. Increasingly, however, users expect to be able to access the network more flexibly. There are several cases:

- Fixed location, where an application is provisioned to be accessed from a specific access node. Wired telephony is an example.
- Flexible location but static access node, where the user is allowed to choose the access node from among different geographic locations, but that access node is not allowed to change during a single application instance. An example is wireless access to the network with the assumption that the user remains within range of a single base station.
- Moving location and access node, in which a user with wireless access is allowed to move from the coverage of one base station to another for the duration of an application instance. This allows the user be in motion, such as on foot or in a moving vehicle.

The flexible and moving location options require high logical connectivity in the network. Thus, greater logical connectivity provides great value to users who desire to be mobile. As witnessed by the rapid growth of cellular telephony, this is a large proportion of users, at least for telephone, data, and document applications.

Like multicast forms of broadcast connections, the moving location option requires fundamental capabilities in the network that must be anticipated in its design and provisioning, since connection instances must be dynamically reconfigured. This option makes much more sense for some applications than others. For example, it is reasonable to conduct a phone conversation while in motion, but more difficult and perhaps even dangerous to watch a video presentation or conduct a more interactive application. Even the latter becomes feasible, however, for users in vehicles driven or piloted by others.

### Openness to New Applications

Aside from the logical connectivity of the network, the second most important characteristic to users is the available range of applications. It is a given that the application possibilities cannot be anticipated in advance, and thus the network should be able to accommodate new applications.

Again the evolution of the computer industry offers useful insights. Because the desktop computer was a programmable device, a plethora of new applications was invented long after the architecture was established. Equally important was the availability of the market to many application vendors, which led to rapid advancement. A primary driving force for the desktop computer was that it freed the user from the slow-moving bureaucracy of the computer center and made directly available a wealth of willing application vendors.

The Internet was architected with a similar objective. The network functionality is kept to a minimum, with no capability other than the basic transport of packets from one access node to another embedded within the network. Beyond these minimal capabilities, the intelligence and functionality required to implement particular applications are realized in the servers and terminals. This architecture separates the development and deployment of applications from the design and provisioning of the network itself. New or improved applications can be deployed easily without modifications or added capabilities within the network, as long as they comply with any limitations imposed by the network design (see "Quality of Service," below). This characteristic has been the key to the rapid evolution of Internet applications, and in turn to the success and rapid growth of the Internet itself.

To be of maximum benefit to users, we believe the NII should be designed according to a philosophy similar to that for the Internet (although without some of its limitations). One can summarize these characteristics as follows:

- Provide full logical connectivity among all access nodes, and do not limit the number of logical connections available to any single access node. To do otherwise limits future applications.
- Do not design the NII or portions of the NII around specific applications, thereby limiting its capabilities to support future unanticipated applications. Rather, realize within the network the minimum capabilities required across all present and future applications (to the extent it is possible to anticipate those capabilities).
- Realize the primary application functionality in the terminals or servers, or alternatively at access points to the network (but within the domain of the network service provider), rather than internal to the network itself. This way new applications can be deployed by adding functionality at only those access nodes associated with users willing to pay for those applications, without the obstacle of making uneconomic modifications throughout the network infrastructure.
- Since standardization presents a potential obstacle to the rapid deployment of innovative applications, consciously limit the role of standardization to the basic network infrastructure. Do not attempt to standardize applications, but rather allow them to be provisioned and configured dynamically as needed.

Even when the NII is designed according to this philosophy, there is still a major obstacle to the economic deployment of new communications (as opposed to database) applications: the community of interest problem. Before one user is willing to purchase an application, it is inherent in a network environment that there must be a community of other users able to participate in that application. For example, an isolated user can usefully benefit from a shrink wrapped personal computer application purchased locally, but in a networked environment may depend on other interested users who have purchased the same application. This can place a daunting obstacle in the way of new applications and limit the economic return to application vendors or service providers. Fortunately, there is a solution. If applications are largely defined in software rather than hardware primitives, they can be dynamically deployed as needed to terminals participating in the application. We call this dynamic application deployment.

A crucial element of the NII required to support dynamic application deployment is the ability to transfer software application descriptions in the establishment phase of an application instance. Deployment can also occur during an application instance (if it is desired to change or append the application functionality). This requires a reliable connection to the terminal, even where other aspects of the application (such as audio or video) may not require reliable protocols. Since such application descriptions are likely to be large, the user is also better served if there is a broadband connection for this purpose to limit the time duration of the establishment phase.

Flexibility in deployment of applications also requires a full suite of control primitives as a part of the network control and signaling interface to the user terminal. Anticipating all the capabilities needed here is a key design element of the NII. Such a design also needs to control the complexity inherent in such a heterogeneous environment, for example by defining an independent "universal" signaling layer together with adaptation layers to different network technologies and preexisting signaling systems.

### Quality of Service

Many applications call for control over aspects of the quality of service (QOS) provided by the network. From the user and application perspective, QOS parameters include the following:

- The setup time in establishment, including configuration of the connection instances, transport of the application description to the participating terminals, etc.
- The frequency with which an application is refused by the network (due to failures, traffic overload, etc.).
- The interactive delay through the network (the time from user action to appropriate application reaction).

- The subjective quality of application components like audio and video, which is affected not only by quantization and network loss artifacts, but also the delay introduced in the transport and synchronization of the audio or video. The subjective quality depends not only on network QOS characteristics, but also on the characteristics of the application implementation in the terminals, such as the algorithms used for audio or video compression.

The user is of course also concerned with the pricing of the application, which is likely to be related to the QOS it requires. The QOS parameters of the network itself affect users and applications, and include:

- The throughput of the network, in both directions in the case of a duplex connection;
- The delay, variation in delay, and temporal characteristics of delay variation in transport through the network;
- The frequency with which losses occur, and the temporal characteristics of those losses (such as whether they are bunched together or spread out); and
- The frequency of corruption of data, and the temporal characteristics of that corruption. (For data transport, corrupted data must be discarded, whereas in continuous-media transport such as audio and video, corrupted data are useful but cause subjective impairments.)

There are two distinct philosophies of network design:

- The network provides guarantees on some QOS parameters. The quantitative guarantees are established by negotiation between application and network at establishment, and appropriate resources within the network are reserved for the connection instances to ensure that the guarantees will be satisfied.
- The network provides best-effort transport, in which resources are provided to a connection instance on an as-available basis, without guarantee.

Rarely does a network strictly follow one of these models. For example, the Internet offers as one option guaranteed delivery (zero loss) service, but does not guarantee against delay. Conversely, the public telephone network offers delay guarantees, but does not guarantee against corruption. Even for a single QOS parameter, best-effort and guarantees can be mixed for different connections, by reserving network resources for some connection instances and providing only leftover resources to other connection instances. QOS guarantees have a cost associated with them, principally in reserving resources, making them unavailable to other connection instances even when unused. There is also a substantial increase in the complexity of the network associated with QOS guarantees. The QOS of the network can sometimes be modified more simply in the access nodes, for example by introducing forward error-correction coding to reduce the corruption probability (at the expense of added delay).

There is considerable controversy over the relative merits of best-effort vs. guaranteed QOS transport. It appears that both models have merit and may reasonably coexist. QOS guarantees will be mandatory for some applications: consider the possible consequences of unanticipated interactive delay in a remote telesurgery application! It has not yet been established or demonstrated that best-effort transport can achieve entertainment-quality video. On the other hand, the simplicity and lower cost of best-effort transport seem desirable for other applications, like interactive graphics. The QOS requirements (or lack thereof) vary widely across different applications. Thus, the NII should be capable of provisioning different types of QOS guarantees to different applications on request, and should also offer a lower-cost, best-effort service to other applications.

For both best-effort and guaranteed QOS, an important issue to the users is any inherent limitations on available QOS. There are many network design choices that can (inadvertently or for reasons of cost) limit the best available QOS. Since the NII is expected to support many applications, it is important that fundamental design choices not be made that unduly restrict the best available QOS, although some portions of the NII may deliberately be provisioned in a fashion that temporarily limits QOS for cost reasons. Among the most important of these design issues are the following:

- The network topology can substantially increase the lowest available delay.
- The choice of physical layer technology in conjunction with topology can severely limit QOS. For example, wireless and wire-pair access technologies can limit the highest available rate, as can multiaccess topologies (wireless reverse links or tree-topology CATV distribution system reverse links).
- Achieving high reliability on wireless access links can be expensive (in terms of system capacity), especially in the worst case and especially in the context of moving terminals.

Because of QOS limitations that are either fundamental (like propagation delay) or expensive to circumvent (like wireless corruption), it is important that applications be scalable and configurable to available QOS (see below).

Delay appears to be a particular problem area for the NII. Of all the QOS parameters, delay is the only one that suffers from a fundamental limit, namely, the physical propagation delay. Propagation delay will be on the order of at least 200 to 300 milliseconds round trip for a connection halfway around the world. The desired delay for some applications is actually less than this. For example, desirable round-trip delays for synchronous continuous media applications like voice telephony and video conferencing, as well as interactive keyboard applications, are on the order of 50 to 100 milliseconds, and delays on the order of a few hundred milliseconds are significantly annoying. Thus, there is little margin for introducing delays in excess of the propagation delay without significant impairment at the greater geographic distances. Unfortunately, there are many design choices that can introduce significant delay that are already observed in present networks:

- Packet networks trade network capacity through statistical multiplexing for queuing delay at switching nodes, and this queuing delay increases substantially during periods of congestion. A given connection instance may traverse many such switches in a network with a "sparse" topology, and thus there is an unfortunate tendency for propagation and queuing delay to increase in tandem.
- A high degree of logical connectivity can be achieved in virtually any network topology, including those with sparse physical connectivity, by adding switching. However, as previously noted, this switching can itself introduce queuing delay. Beyond this, the physical path traversed by the data can be considerably lengthened, increasing the propagation delay as well. This is a flaw in any approach involving a collection of "overlay" subnetworks with Internet gateways.
- In packet networks, large packet headers encourage long average packet lengths at high network utilization. For low-throughput applications like voice and audio, the packet assembly time for large packets introduces a large delay (independent of network throughput). An example is the Internet Protocol, which has a large packet header (scheduled to get even larger in the future).
- It is tempting to insert transcoders from one compression standard to another in the network for audio and video applications. These transcoders force delays to add across network links on a worst-case (as opposed to statistical) basis, and also add significant signal-processing delays. For example, digital cellular base station voice transcoders add a one-way signal-processing delay of about 80 milliseconds.

Achieving a feasible delay QOS in the NII (and especially its global extensions) acceptable to the most critical applications will require major attention in the design phase and coordination among the network service providers. Past and present trends are not encouraging in this regard, as many network technologies developed nominally for a limited geographical area have unwittingly introduced substantial delays.

Another troublesome observation is that QOS guarantees will require dynamic coordination among network service providers at connection establishment. A typical connection instance will span at least several network service providers, and possibly many more, for example, local-area network and metropolitan-area network providers at both ends and a long-haul provider. QOS parameters like delay, loss, and corruption will be affected by all the providers' networks; however, the user cares only about end-to-end QOS. Achieving end-to-end QOS will require an allocation of impairments among the providers. Such an allocation should be dynamically determined at establishment, since a static allocation will require that all networks provide a QOS appropriate for the worst-case scenario, an expensive proposition. The only practical approach appears to be dynamic allocation mechanisms that relax QOS objectives for individual links to fit the circumstances, such as



local congestion or wireless access. There are no such mechanisms in place, nor a credible process to establish such mechanisms.

### Security and Privacy

A weakness of some current networks, particularly wireless ones, is lack of security and privacy. It is evident, for example, that insufficient effort has been devoted to this in cellular telephony networks in North America, as evidenced by the ease of eavesdropping and the widespread theft of service. This becomes an issue for both users and network service providers. From a user perspective, the following characteristics of the NII are important:

- Freedom from casual eavesdropping;
- The capability to make eavesdropping infeasible for sensitive applications, acceptably at extra cost;
- Freedom from theft of services (obviously of interest to service providers as well); and
- Inability to surreptitiously track the identity or movements of users.

Achieving all these goals requires careful attention in the design phase of the NII. As an example, transcoders already introduced in cellular telephony preclude privacy by end-to-end encryption.

### Application Scalability and Configurability

As previously mentioned, the maximum benefit will accrue to the user if new applications can be freely deployed and made available to all users, regardless of their terminal capabilities and the transport facilities available. In this model, the application will be dynamically configured to fit the environment (terminal and connection instances), attempting to achieve the best quality consistent with the limitations. Examples include:

- *Scalability to the connection QOS.* For example, a video application may be configured to lower resolution or subjective quality in the case of wireless access, as opposed to a backbone-only connection. It is not desirable for the user that an application is precluded by, for example, a wireless access; rather, the user would prefer that some QOS parameters (and thereby subjective quality) be compromised.
- *Scalability to the terminal capabilities.* For example, a video application will be configured to a compression algorithm requiring less processing (trading that off against lower quality, resolution, or greater transport bandwidth) should the originating or receiving terminal instances have limited processing. It is not desirable for the user that applications be limited to terminals provided by particular manufacturers or with particular capabilities.

Dynamic configuration requires scalability and configurability of all aspects of the application. It also requires a rich signaling and control environment that passes to the application all the information needed to scale to the environment. The mechanisms described above for negotiating and configuring QOS parameters of the transport at establishment do not by themselves provide needed information about terminal capabilities. Thus, there need to be standardized signaling capabilities among the terminal instances at establishment.

### Pricing

The pricing model is a key to the desirability and viability of applications in the NII. It is ultimately in the best interest of the users that both network and application service providers derive revenue related to their costs. This is a difficult issue because of the great heterogeneity of networks and applications.

If the NII provides QOS guarantees as described previously, there must be a coupling of pricing and the cost of resources reserved to provide the QOS, since otherwise applications will always request the highest quality

available. Since the cost of provisioning a given QOS will also depend on current traffic conditions, it is desirable that pricing be traffic dependent. Many connections will involve two or more network service providers, each provisioning identical rate parameters, but possibly contributing quite different impairments such as loss and delay to the end-to-end QOS (based on their technology, local traffic conditions, etc.). Those network service providers should derive revenue that is related to their contribution to end-to-end QOS, since otherwise they will all have an incentive to fully consume the end-to-end impairment objectives.

Thus, we conclude that the pricing to the user and division of revenue should be established based on the rate parameters, the contributions to the impairments of the individual network service providers, and local traffic conditions. This requires a complex negotiation between the application and a set of network service providers to establish an end-to-end QOS that achieves an appropriate trade-off between price and QOS, and a partitioning of that QOS among the network service providers. One approach is a broker that mediates among the application and all potential network service providers. A desirable feature of a brokerage system from the user perspective is that all available network service providers could be considered, choosing the set of providers that is most economic based on their current traffic conditions and pricing strategies.

### CONCLUSIONS

Looking at the NII from a user perspective, we can identify some key challenges for the future:

- To meet a wide range of application needs and provide flexibility for the future, individual network service providers and their equipment vendors need to take a general perspective, as opposed to developing and deploying technologies defined for narrow currently defined applications.
- Major cooperation is needed among network service providers to coordinate their design and deployment strategies in areas like end-to-end transport protocols and signaling capabilities that allow dynamic allocation of end-to-end QOS impairments, support scalability and configurability of applications, and provide desired levels of privacy and security.
- Overall planning is needed, with specific action on the part of individual network service providers, to be sure that near-term decisions do not compromise end-to-end QOS objectives in the NII and especially its global extensions.

The greatest challenge in the NII is to allow for and encourage a variety of technologies, applications, network service providers, and applications service providers to coexist in a dynamic environment, while satisfying the user's desire for interoperability, openness to new applications, and acceptable levels of performance. This will be possible only with initial planning and coordination and ongoing cooperation among all parties involved.

## 46

# Role of the PC in Emerging Information Infrastructures

*Avram Miller and Ogden Perry*

*Intel Corporation*

While most of the talk about the future national information infrastructure (NII) has focused on interactive television, the combination of the personal computer with online services and the Internet is rapidly achieving the promise of ubiquitous information access for education and entertainment, electronic commerce, and work at home. While telephone companies position themselves to enter the television delivery business and cable companies attempt to provide telephone services, the foundation for the information age is actually being built by the computer industry. The thesis of this paper is that if we are to establish an NII of any substance, it will be based on the interconnection of personal computers with online services and the Internet. PCs, not televisions, will be used by people to communicate with each other. Unfortunately, a widely held belief that personal computers serve only a small, technically savvy and wealthy elite—coupled with concern that computers are too difficult for mere mortals to use—has mistakenly put the spotlight on creating interactivity on the television. This, compounded with the fact that TVs are present in virtually every home, has resulted in some erroneous conclusions. In fact, television was designed for one thing only: moving pictures. It is a poor device for handling the high resolution static images required for a high degree of interactivity. With the possible exception of video games (which serve a relatively small segment of the market), it is primarily a passive device. The fact that there are so many TVs is not relevant. If it were, we would have long ago considered "interactive radio." The number of televisions has become a false beacon for media companies, communications companies, equipment suppliers, and government policy makers.

Since the technological and social infrastructure for interactive television does not exist, and its creation would be a truly awesome task, government has stepped into the breach, preoccupying itself with fostering its existence and getting involved in the processes by which it will evolve. Issues of interoperability standards, competitive markets, and universal access have become matters of public policy. This is further complicated by the fact that television and telecommunications services are regulated, implying the need for regulation of interactive television and, by implication, the information superhighway.

It is our contention that this approach of imposing the television as the interactive communication device standard is based on a set of mistaken assumptions. The fact is that the ubiquitous information device exists now, in the form of the personal computer. In fact, while the government and the telecommunications companies argue about how best to deregulate the telecommunications industry, consumers are voting with their pocketbook by purchasing personal computers, the only available interactive information device, at record levels. The role model for the diffusion of personal computers into society may well resemble the adoption of the automobile: it would have been difficult to imagine in the early days of the "Model T" that 88 percent of American households would have one or more automobiles. And in the case of the personal computer, business to a large extent is underwriting both the investment in developing the networks (the role the government had to play in developing the highway system for the car) and, very importantly, the training of a large number of consumers in the use of computers (approximately 30 percent of the U.S. labor force and 47 percent of the white collar labor force use a computer at work, and most children learn to use computers at school).

Telephone and cable companies have been regulated for most of their existence. As a result, they are well positioned to influence government policy with respect to the information superhighway. The computer industry,

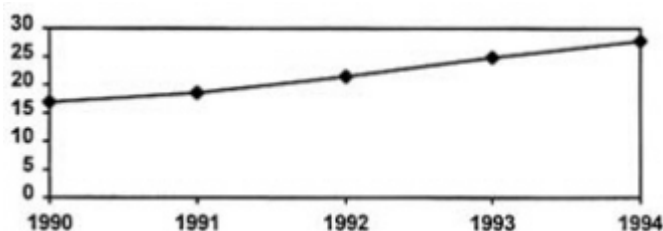


Figure 1 PC installed base in U.S. businesses and government (millions of units). SOURCE: Data from Intel Corporation; Infocorp.

on the other hand, has had almost no regulation and little to no involvement with government. Those of us in the computer industry need to do our part in demonstrating to our policymakers that the information superhighway is already under construction. It is being built primarily on PC technology by private industry, driven by competitive market forces and with little, if any, need for government regulation.

There is a role for government, however; that is to recognize and embrace the personal computer. This can provide America with a foundation for the next century that will improve not only the quality of our lives but also the productivity of our society and the competitiveness of our industries. The government can do many things to move this along. It can facilitate access to computing in schools, libraries, and community centers. It can become one of the largest (maybe *the* largest) content provider of online information. It can encourage business to promote telecommuting. The most important thing the government can do is to recognize what is happening. However, we caution against interference with market forces. The market moved the personal computer industry forward at a phenomenal pace, and we encourage following this model rather than the heavily regulated telecommunications model, which has proven slow to evolve and respond.

## THE PERSONAL COMPUTER INDUSTRY

### Its History

While today the primary use of personal computers is in business, it is interesting to note that they were originally conceived of as a consumer product. The first personal computers were designed either for hobbyists—the Altair—or for consumers—the Apple II. Even IBM broke with its tradition of providing business products (the B in IBM) when it introduced the PC in 1981. This machine, which is the ancestor of over 85 percent of the computers sold today, actually had a game port for joysticks and an audiocassette for storage. Industry leaders at the time, such as Ken Olsen, then CEO of Digital Equipment Corporation, openly referred to personal computers as "toys." Few realized that this toy would completely restructure the entire computer industry within 10 years.

The personal computer overtook the mainframe with its terminals as the information tool of business some time during the 1980s. It is interesting to note that although there are clear reasons for this success, the mainframe was not without its merits. As a centralized facility, a mainframe is easy to control. Each user has access to the same software, and support is easy. Another clear advantage of mainframes is that since they are a shared commodity, it is easy to manage capacity to match the average expected load. On the other hand, under peak usage, all users typically experience slower performance. Probably the main reason for the rapid decline of the mainframe, however, is the slow pace of progress in both hardware and software performance.

While there are many reasons for the success of the PC in business, the most important is its evolutionary nature. The "openness" of the PC allowed for rapid innovation. The open bus inspired hardware companies to add value to the basic PC. They experimented in the marketplace. Successful additions were then integrated into the main computer. It is hard to imagine that the first PC had a game port built in, while the printer port was optional. Application software could be created by small companies. Companies like Lotus and WordPerfect grew from one product, while Microsoft took an early lead in the operating system and Novell provided the next work environment. While this environment was, and still is, chaotic, it provided for rapid evolution. No industry

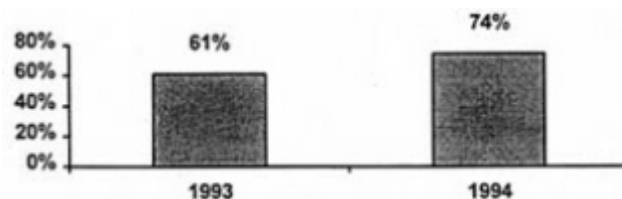


Figure 2 Percent of business PCs with LAN connections.  
SOURCE: International Data Corporation.

or government organization dictated the direction or set the standards, yet standard interfaces evolved out of economic necessity. The customers rewarded companies for innovation. When a company won in the market, others would accept its standard as the de facto standard and would seek to integrate their products into it. In the process, the personal computer became the tool of business. The adoption of personal computing into business in the United States was rapid (Figure 1).

This was followed by the growth in computer networks (Figure 2).

Electronic mail has followed word processing and spreadsheets as a key business application. This has made the PC a fundamental communication device. While E-mail has been historically used within a company, more and more companies are using it as a way of communicating with the outside world (customers, vendors, and partners). Now many individuals are finding that they can also send notes to friends and relatives. E-mail is only the first example of the successful marriage of personal computers and the NII: communication between people through the computer. At Intel, e-mail is a "mission critical" application; the corporation functions through the rapid transaction of issues and ideas using the worldwide electronic mail system.

The term "social computing" is gaining currency—a poor term no doubt, but a powerful concept. And while standard bodies were formed to address the methods by which different electronic mail systems would communicate (x.400 for instance), the Internet became the lingua franca of e-mail. Now, as businesses connect to the Internet to send and receive mail, they are also discovering that they can access a wide variety of information on the Internet.

More and more companies are beginning to use the World Wide Web to communicate with the outside world. The growth of the Internet is staggering. It is estimated that 10,000 companies are adding their presence to the Internet each week.

### The Growth of the Home Market

The growth of computing in the home has been a surprise even to those in the industry. Figure 3 shows historical home penetration in the United States. This year it is expected that 10 million personal computers will be sold to consumers for use at home.

Though these numbers are impressive, they do not do justice to the number of consumers that use computers either at work or at school. While it may be true that the use of computers is not easy to learn (is the use of automobiles?), computers clearly satisfy a compelling need. For the first time last year, consumers spent more money on computers than on TVs and VCRs combined, making the PC the most successful consumer electronic device. Another indication of the success of the personal computer in the home is that of the more than 6 million PCs sold to consumers in 1994, about 25 percent were additional computers to homes already owning one or more PCs.

In the early 1990s a critical mass of home PCs equipped with modems enabled an explosion in a new industry: the online marketplace (Figure 4). Customers are now finding that much of the value of having a personal computer resides in the ability to connect with other users and with a large number of online services.

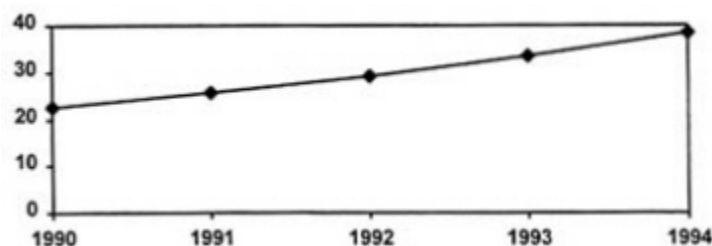


Figure 3 PC installed base in U.S. homes (millions of units).  
SOURCE: *Computer Retail Week*.

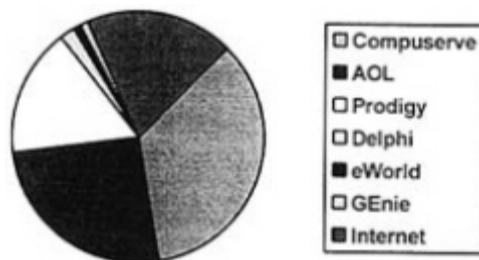


Figure 4 On-line connections in the United States (total = 8 million). SOURCE: *Rocky Mountain News*.

It should be noted that although over 30 percent of U.S. homes are already equipped with PCs (a ramp up faster than many consumer electronic products), the growth shows no sign of slowing down. On the contrary, with other countries joining in, worldwide shipments of PCs are expected to grow from 60 million units this year to over 100 million annually by the year 2000 (Figure 5).

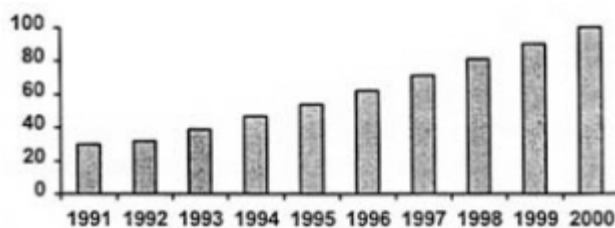


Figure 5 Worldwide PC shipments forecast (millions of units).  
SOURCE: International Data Corporation.

### Today's Capabilities

The power of today's computer is mind-boggling by standards of just a decade ago. The typical computer being purchased has a powerful microprocessor (typically a Pentium) with computing power equal to the mainframe of just 6 years ago. This, combined with a high resolution display, a CD-ROM, and CD quality audio sound provides for a rich multimedia experience. Today's PC uses compression to display video that matches the quality of a typical VCR, with only 0.5 percent of the information required for an analog video stream. The decompression of the data is performed in the microprocessor, without any specialized hardware. In addition, it can act as a fax machine and provide access to online networks and the Internet. It is used for a variety of purposes from entertainment to education, work at home, and home shopping.

### **Tomorrow's Capabilities**

The capabilities of the personal computer are improving at a rate that no other device has ever experienced. Historically, the power of a PC has doubled every 18 months. This rate has recently increased, and the capabilities are now doubling every 12 months. What will we do with such computing power? Here are some of the most obvious areas of use:

- Improvements in ease of use, including natural language input and output,
- Powerful 3D graphics,
- Outstanding audio/video capabilities,
- Video telephony, and
- Multitasking (doing several things at the same time).

### **The Structure of the Industry**

Historically the computer industry was vertically integrated. Computer companies did everything from designing their own power supplies to developing application software. Each "vertical stack" was unique and proprietary. The PC changed all that. The PC industry is basically horizontally integrated, with literally thousands of companies delivering compatible products at every level of integration and in many areas of applications.

The result of the horizontal nature of the computer industry is fierce competition at every level in both quality and price. This competition benefits customers, who have access to the highest performance at the lowest prices. De facto interfaces have emerged at the intersection of the various segments. Periodically, ad hoc industry groups are formed to cooperate in creating new interfaces. An example is the PCI bus, which has become the standard for the high-speed internal bus of personal computers.

The cable television industry still operates in the traditional vertical organization. Cable companies are the only suppliers of both equipment and services, and most consumers don't even have the opportunity to buy their own equipment. As a result, equipment is typically not interoperable between different companies, technical progress is slow, and prices do not decline. This slow progress, in turn, stands in the way of improvements in other parts of the value delivery chain (such as the picture-in-a-picture feature that is disabled by most set-top boxes). As a result, government has recently stepped in to regulate both function and price.

### **KEY DIFFERENCES BETWEEN THE TELEVISION AND PERSONAL COMPUTER INDUSTRIES**

Televisions use CRTs to display moving pictures. Most personal computers also use CRTs. This fact sums up the essential similarity between these two devices. The fact is that in their purpose, their features, their evolution, the way they are sold, and, perhaps most importantly, the way they are used and are connected to the outside world, televisions and personal computers have very little in common.

#### **Purpose**

The television was designed to bring visual entertainment, from the movie theater and the stage, into the home. This emphasis on the visual is seen, for example, in the fact that until recently the quality of television speakers was far below that of other consumer electronics devices and that, even today, few programs are broadcast in stereo. To its credit (depending on your point of view), the television also enabled some new forms of entertainment such as game shows and brought us news footage. A common element to all these experiences is that they are typically enjoyed in the family room, from a comfortable distance, in a reclined position. The PC, on the other hand, was originally designed as a work tool and a platform for gaming. Two of the first applications

for the PC were Lotus's 1-2-3 spreadsheet and Microsoft's Flight Simulator. Although these applications would seem to be very different, they both require the user to sit close to the PC and to interact with it on a continuous basis (whether the device used is a keyboard or a joystick is not important).

### Features

The NTSC standard for television broadcast in the United States specifies 483 lines of video. In practice most broadcasts and most VCRs are capable of only about half that. By contrast, the minimum resolution for any personal computer monitor sold today is  $1024 \times 768$ , with many exceeding even that. There are many reasons that explain the "poor" resolution of televisions—the most significant being that it is adequate for the purpose at hand and the cost of upgrading the system is huge. But resolution is only one of the differences. Probably much more crucial is the fact that the PC is inherently an "intelligent" device capable of processing information locally and of interacting directly with the user. The PC also has local storage that can be used for programs, data, video, or whatever the user thinks is important.

### Evolution

But the single most important difference between the TV and the PC has nothing to do with any one feature but rather with the fact that the PC is constantly changing, in effect adding features on a regular basis. For those who live in the world of telecommunication and television where things move very slowly (witness integrated services digital network and high-definition television [HDTV], the PC hypergrowth may be difficult to comprehend. Now that HDTV is being proposed, the TV is being asked to do something it has not done in 50 years: evolve. With the exception of screen size, televisions have remained fundamentally the same since the introduction of color TV almost 40 years ago (even that change was a painful one that resulted in a compromise that still hurts the picture quality of television). While some features have been added to television, and the VCR gave it a limited linear memory, most new features are seldom used.

### Interactivity

We define interactivity as an activity where individuals are fully engaged in a process. Interaction involves an interplay between people or between people and devices such as a computer. We wish to contrast that with control and navigation. While there may be a brief moment or two where we interact with our remote control and TV in order to select a program of interest, success usually means long stretches of passive involvement with the program of choice. While the telephone can be used for some form of interactive information retrieval or even to perform a transaction by using the keys and following a menu of choices, most would agree that this is not a very enjoyable experience and is unsuitable for sustained periods of time. This should be contrasted with the PC, a device with which individuals routinely interact for hours at a time.

### Connectivity

The TV was built from the start to be a "communications" device—albeit one-way communications. After all, a TV without a connection to the outside world is useless. In fact, it is the television network (especially cable) rather than the TV itself, that has expanded to offer more services. So it is this paradigm that has motivated the communications companies to approach the problem of providing interactive service to American consumers in a similar way. It is therefore very ironic that the PC, which gave users the independence to work on their own and to break away from the mainframe, is, in fact, the superior communications device. This is because on the PC, which is capable of producing, processing, and storing information, communications evolved from the very start as a two-way interaction.



The TV is easy to use only if it is used in its simplest form. To make the TV more intelligent in order to be interactive, something has to be added to it. This is typically called an intelligent set-top box. The idea is to add to the TV many of the functions of the personal computer and the game machine—but at a price that is consistent with television pricing (typically, \$300 is the target). Since \$300 does not buy a lot of computing power, much of the intelligence needs to be somewhere else in the network where it can presumably be shared. If this sounds familiar, it is because it copies the mainframe/dumb terminal concept. The computer industry has the advantage of having seen that movie and of being familiar with its pitfalls. Since the intelligent PC was so successful in replacing the mainframe more than 10 years ago, there is no need, this time, to go through the same steps.

### **A Word About the Telephone Industry**

Much of what was said above about the television is also true of the telephone. This device, too, has been slow to evolve (touch tone being the last—and probably first—breakthrough in the device itself). Here too, it is the network that is responsible for most innovations such as 800 service and call waiting. The awaited breakthrough on the telephone side (equivalent to HDTV) has been some form of video telephony. All attempts have failed because the cost and quality have not been there. On the other hand, personal computers have offered video telephony, incorporating application sharing, on ISDN for some time. It is expected that the personal computer's current trajectory will allow it to offer a marketable quality video telephony product over plain old telephone service (POTS) within the current year with the simple addition of a camera.

As the television and telephone industries eye each other's heartland but consider the massive cost of doing battle with each other, they each hope to create an advantage by offering a new class of service: interactive TV services. Unfortunately, they both have a poor track record in evolving their business. And this time they must build their new infrastructures in a matter of years, not decades.

### **HOW WILL THE PC BECOME UBIQUITOUS**

At the current rate of purchase, between 50 and 60 percent of homes will have computers by the end of the decade. While that is an amazing penetration, the other 40 percent need to be addressed. This group can be broken up into those who would like to have a computer at home and those who have no interest. There is an obvious economic barrier facing a portion of the people who would like to own a computer. Similarly, while almost everyone has access to public transportation, not everyone owns a car. We strongly believe that access to the information highway via PCs will be made available to all people via computers at schools, community centers, libraries, and other government and private sites where the cost is spread over many users.

This "economic barrier" should not be construed as an opportunity for the television to evolve into interactive television. We believe that the cost of interactive TV (both the device itself and the network) would limit its availability to pockets of high-income families. Since the PC is a multifunction device, its cost is, in fact, less of an obstacle—as can be seen by its fast ramp so far.

The following key characteristics of the personal computer industry ensure that the PC will continue to be the primary interactive device of the information superhighway:

- It is a highly capable device designed from the start for interactivity.
- Its adaptability allows for rapid market experimentation with the market setting the standard.
- Businesses have invested trillions of dollars in a computer infrastructure.
- Consumers, in turn, have benefited from this investment in the form of training, cheaper equipment, and access to infrastructure.
- The Internet, which is the only national information infrastructure, was specifically built for and with computers and is a common network for business and consumers.

### Requirements for NII

The first requirement is an access point with an interactive device. By this we mean interactivity as defined by the current and evolving personal computers, as opposed to improvements in navigation and control capabilities of TV. We assume that PCs will not only become easier to use but will be able to handle "rich" media (video, audio, images, in addition to text), with increasingly better quality.

Next, we need a communications infrastructure that connects these access points to services and to each other. This connection must provide adequate bandwidth in order to enable the use of rich, multimedia data types. The current POTs environment, even with the latest 28.8-Kbps data rate, is extremely limiting. ISDN is a short-and intermediate-term opportunity to increase bandwidth to the consumer. Over time, broadband capabilities will be needed. It should be noted that bandwidth symmetry is not required as long as "adequate" bandwidth is available in the return path. For example, the current PC-to-cable connection experiments, while highly asymmetric in bandwidth, would be quite adequate for accessing existing Internet sources. Clearly, we will need higher bandwidth in the return path if we are to expand beyond current modes of use and allow every user to become a contributor of rich data.

In an important departure from the current circuit switched networks that connect most businesses and individuals, the network for the NII needs to be packet switched. This requirement is borne out of the desire for an "always on" mode of operation and for the ability to have multiple sessions active at the same time. The Internet is packet switched, but much of the potential is lost in the circuit switched POTS network that connect homes to the Internet over the "last mile."

If we are to broadly utilize the NII for applications such as electronic commerce, security considerations are going to become important. However, we consider this to be an issue between the two end points of the transactions and not of the network. Current industry efforts, we believe, will solve this problem quite adequately.

### What Can/Should the Government Do and Not Do

The government should do the following:

- Recognize that the PC is the interactive device and that the Internet is the information superhighway. While neither is perfect for the ultimate tasks for NII, they are the closest in satisfying our future needs. Furthermore, they are open standards with industry momentum behind them; they are the winners of this Darwinian process of evolution.
- Encourage employers to make work at home a major thrust. Telecommuting will have economic, social, and environmental benefits. It will also provide computers at home for many who might not be able to afford them otherwise.
- Develop programs making networked computers available to all Americans via schools, libraries, community centers, and other government facilities.
- Encourage a competitive communications environment with specific advantages to those who provide the low-cost, two-way, high bandwidth required by personal computers at home.
- Avoid trying to set standards or dictate economic behavior. The government should also avoid any procedures that slow down the creation of de facto, market-developed standards.
- Encourage states and municipalities to deregulate the local loop so that advanced telecommunication services can be made more readily available to the consumer. ISDN, for instance, should not be looked upon as something for the techno-rich, but as a service available to all Americans.
- Finally, capitalizing on the enormous amount of information it collects but has not been able to redistribute efficiently, the government should become a major provider of content and services on the Internet.

### SUMMARY

The ubiquitous interactive device is here, and it is the PC. While telephone companies and cable companies have focused on television and telephone, the consumer PC market has gained momentum. PCs will be in a majority of homes by the end of the decade. The Internet will provide for an open information environment. Government should understand and facilitate these trends. The computer industry has to play an active role in this process.

## 47

# NII Evolution—Technology Deployment Plans, Challenges, and Opportunities: AT&T Perspective

*Mahal Mohan*  
*AT&T Corporation*

### STATEMENT OF THE PROBLEM

This paper presents an AT&T view of the evolution of the national information infrastructure (NII). As a leading supplier of long-distance communications services and wireless services and a major force in the research and development of communications and information technologies used by customers and service providers, AT&T has a combination of perspectives on this topic that is unique in the industry.

I briefly review the current state of information infrastructure in terms of certain key attributes, outline AT&T's key new technology deployment initiatives, and recommend actions by the public sector to facilitate faster, smoother evolution of the NII.

### BACKGROUND

Several attributes of emerging communications networks offer promise for supporting new modes of collaboration, information access, and exchange:

- Bandwidth and the physical medium and technology used to provide the bandwidth;
- The set of intelligent features that enhance the convenience or usefulness of the networking capability or service;
- Messaging, involving storage, processing, and forwarding of information in different forms;
- Mobility, enabling users to communicate from anywhere using wireless technologies and personal reach numbers; and
- Interoperability and openness, which is an enabler for competition and rapid, ongoing innovation in each of the above areas.

#### **Bandwidth, Physical Media, and Technologies**

Two prominent trends in the current evolution of communications networks are digital technology deployment and optical fiber deployment. Most long-haul backbone communications networks have evolved from analog to digital transmission. Digital deployment in local loops to end users, however, has been proceeding slowly. However, the recent increase in the deployment of integrated services digital network (ISDN) services shows some promise for bringing the benefits of digital transmission all the way to the end user, resulting in the provision of expanded bandwidth with superior performance characteristics.

During the last decade, optical fiber capable of supporting large bandwidths (ranging from tens of megabits per second to several gigabits per second), with clear, nearly error-free transmission over long distances, has been extensively deployed in backbone communications networks that support aggregate traffic from many users. Beginning in the early 1990s, the fiber deployment has extended closer to residential neighborhoods and businesses, mostly still supporting aggregate traffic from many contiguous user locations, and in some cases

supporting aggregate traffic from a single large business location. The "last mile" that delivers information to or from individual locations has some dominant technologies today. These are twisted pair copper typically supporting symmetrical, two-way, narrow bandwidth (less than 56 kilobits per second) analog transport or, less commonly, medium bandwidth (56 kilobits per second to 1.5 megabits per second) digital transport; and coaxial cable, typically supporting broad bandwidth (the equivalent of 1.5 megabits per second or higher) analog transport—mostly one-way, with two-way expansion widely planned.

### **Intelligent Features**

Intelligent features constitute a broad category, and one that tends to get less than the attention it deserves in many discussions of the next generation infrastructure. A classic example of such a feature in voice (and some circuit-switched data) networks is 800 number service (and the myriad related services that have emerged in the past decade that involve database or routing-table lookup and translation in the network). The tremendous utility of such services is demonstrated by the degree of their widespread use today, mostly for voice services. In the case of data and multimedia networks such a concept again applies, albeit with different implementation details. Directory services, database services, and network-based security services are examples of intelligent service capabilities that vastly enhance the value of the underlying connectivity to users. Currently these features are offered in rudimentary form as part of data and multimedia services, often to a limited base of users. As described below, efforts by AT&T and other industry players are slated to substantially increase the deployment and use of these features beginning this year, expanding rapidly over the next several years to offer more robust and useful sets of features supporting a broader user base.

### **Messaging**

Messaging, involving the storage, processing, and forwarding of information, is becoming widespread and accepted as a mode of information exchange. Voice messaging using premises-based equipment such as answering machines or computer-based voice mail systems is common now. Network-based voice messaging services are available in some areas but are less widely used; their features, functionality, and price structure need to evolve further to provide full-fledged competition to premises-based systems. Data messaging, or e-mail, is now widely used in corporations and is used by the more technically oriented consumers. Substantial progress needs to be made to provide simplified user interfaces, build user awareness, and provide user training, before e-mail can become a commonly accepted form of information exchange for a broad cross section of society. Text-to-speech conversion and vice versa are being actively worked on in research laboratories, with early implementations being used in today's commercial applications.

### **Mobility**

One of the major trends in communications during the 1990s is the explosive growth of wireless services. Driven by the needs of a mobile society, greater availability of wireless spectrum, and technologies that allow increasingly more efficient and cost-effective use of the spectrum, wireless services will continue to expand rapidly.

Another trend in serving mobile users is the concept of a personal number that follows users no matter where they are, if they wish to be reached. The first generation of such services has been available for a few years. The next stage in their evolution is likely to link wired and wireless access to a user via a single number.

### **Interoperability and Openness**

This attribute is worth singling out because it leads to competition, increased user choice, and innovation. Standard, open interfaces between local telephone networks and long-distance networks, and between information appliances and telephone company networks, have enabled substantial competition in the long-distance communications and customer premises equipment industries, even while local communications is predominantly provided in a noncompetitive environment today. Open platforms have likewise facilitated vigorous competition and innovation in many facets of the personal computer industry. As we are poised on the threshold of broader bandwidth services and an expanding range of information services, industry consensus on, and implementation of, an expanded suite of open, critical interfaces are of vital importance.

Examples of interfaces that are open today are the customer premises device interface to local telephone networks, and local telephone network interfaces to long distance communications networks. Typically closed interfaces today include the cable network interface to set-top devices at homes, and the cable network interface to electronic content or programming.

### **AT&T'S PLANS, CHALLENGES, AND OPPORTUNITIES**

This section reviews some of AT&T's key new initiatives and plans in the areas of communications services and information services. It concludes with a discussion of issues that industry, users, and governments need to work together on to create a framework for rapid growth of the national information infrastructure (NII).

As its corporate mission, AT&T is dedicated to being the world's best at bringing people together—giving them easy access to each other and to the information and services they want and need—anytime, anywhere.

As indicated above, AT&T has multiple roles in the evolution of the NII—a major long-distance and global service provider; a major wireless service provider; a product vendor for builders of the communications infrastructure; and a provider, to consumers and businesses, of information appliances.

### **Communications Services: Initiatives and Directions**

Communications services lie at the core of AT&T's business. Our worldwide intelligent network provides the bedrock on which we are building a wide variety of communications services, driven by ever more demanding user needs. Over the past decade, our worldwide network has been transformed into one that carries the vast majority of its information in digital form. Interconnecting high-capacity digital switches are high-capacity fiber-optic links that offer a combination of large bandwidth and clear, error-free transmission. To provide the highly reliable services needed by today's users, we have developed and deployed systems such as FASTAR to reroute and restore facilities handling hundreds of thousands of calls automatically, within seconds, in the event of any malfunction or failure in any portion of our network.

Using the worldwide network as a basic platform, we are building families of communications services, aimed at businesses and consumers, that meet the specific needs of different user segments. These services are differentiated, each in terms of features discussed in the previous section, namely bandwidth, intelligent features, messaging capabilities, mobility, and openness and interoperability.

Let us begin with a description of communications services for businesses, because many leading-edge services are first introduced to business users and, whenever appropriate, are adapted to address consumers at homes.

### **Transport Technologies and Services—Alternatives**

Using ISDN digital transport, AT&T has been offering video communications services, ranging from personal conferencing on a desktop to group conferencing with a variety of speeds and picture resolutions. The

most recent developments in this service family, known as AT&T WorldWorx Solutions, are global reach and multipoint capability. It is now possible to conduct video and multimedia conferences between, for example, Washington, Singapore, and Paris simultaneously, using a mix of desktop computers and group video-display systems.

High-speed packet transport technologies, known as frame relay and cell relay technologies, have emerged in recent years and are spreading rapidly in use. For instance, we offer a frame relay transport service (Interspan Frame Relay Service) that connects computers at high speeds using virtual circuits that can be reconfigured as user needs change. To address user needs for simultaneous communication of voice, text, image, and video at high speeds, we have recently begun offering Interspan ATM Service, based on asynchronous transfer mode (ATM) technology, which is emerging as a worldwide industry standard for multimedia networking.

A discussion of ATM from our perspective is not complete without a mention of product development initiatives. At the core of our ATM service is an ATM broadband switching product we have developed at AT&T Bell Laboratories called GlobeView-2000. GlobeView-2000 is one of a family of ATM switches under development that will enable integrated switching of multimedia signals. ATM technology is experiencing rapid growth in the field of local area networks (LANs). We believe that it has the potential to be a new unified infrastructure for communications, coexisting for decades to come with the current large embedded base of circuit-switched technology.

### **Intelligent Features and Collaboration Tools**

ISDN, frame relay, and ATM services, as outlined above, offer transport alternatives that can interconnect and interwork with each other, and support high-bandwidth applications. Communications services are also growing in terms of features such as directory, security, and user interfaces. For example, this year AT&T is introducing a new family of public data services that will build on and expand the capabilities of the advanced transport services mentioned above. These will provide "multimedia dialtone" and offer a flexible applications environment for innovators. The public data services will enable companies to go beyond having sophisticated internal ("private") networks, and to connect to their suppliers and customers with data and multimedia information. AT&T NetWare Connect Services and AT&T Network Notes are part of this new family that we are beginning to beta test with selected customers and that we expect to make widely available starting later this year.

We are developing AT&T NetWare Connect Services in collaboration with Novell, Inc. The service will enable high-speed LAN access and interconnection, both within an enterprise and between enterprises. It will connect to the global Internet but will offer much higher levels of security, ease of use, and directory and database capabilities. By being an open platform with standard interfaces, this service will in turn become the infrastructure for new services, of which AT&T Network Notes is an example. AT&T Network Notes, which we are developing in collaboration with Lotus Development Corporation (now part of IBM), incorporates Lotus Notes, the popular work-group collaboration software, within the AT&T network. As a result, far-flung work groups can work together on shared documents, incorporating multimedia information into them. Users can rely on the AT&T network to update the software and to readily incorporate new features. Directory services, multiple levels of user security, and navigation capabilities will be part of the common services platform and will be offered with new services as we develop them—individually or with partners in the information industry. Taken in combination with a variety of Internet services for businesses and consumers that we will be offering soon, these represent a service creation platform for a new generation of data and multimedia services.

### **Consumer Service and Product Initiatives and Challenges**

The description above outlines some of the emerging communications service options for businesses to enter the multimedia networking era. What about consumers? First, businesses often tend to be early adopters of leading-edge services, and the services and technologies are subsequently adapted for use by consumers. Second,

many of our business communications solutions involve businesses communicating with their customers, who often happen to be consumers. For example, toll-free 800 number services are business solutions, but the people who make the 800 number calls are mostly consumers who have come to accept the services as an integral part of their daily lives.

There are significant challenges in providing multimedia services to consumers on a broad scale. The most notable is the limited availability of two-way broadband access for connecting to our high-bandwidth services. As is well known, cable TV networks offer broadband connections to many homes today. What restricts their utility for most interactive multimedia services is the fact that they are currently offered as part of a closed, horizontally integrated scheme; in other words, they reach many homes, but they are generally not open to connect to everything we would want them to connect to. In addition, the access links are generally designed for one-way delivery of analog information. They also have a history of modest reliability and performance, though significant new investment is going into installing fiber optics closer to clusters of homes to aggregate coaxial cables and to provide improved performance.

In our role as a major provider of telecommunications infrastructure products and capabilities, we are actively working to bridge this gap. We are working with cable companies as well as telecommunications companies to provide new solutions based on a combination of fiber-optic and coaxial cable links from homes to switching or information hubs, combining high reliability with high bandwidth for interactive applications. We are providing technology and integration capabilities that are a major part of projects by telephone companies such as Pacific Bell, Bell Atlantic, and Southern New England Telephone, as well as by cable companies such as Time Warner, to redefine their communications infrastructure and to offer reliable, broadband access to homes in the immediate future.

Two other areas in communications services deployment are worthy of note. The first is the rapid growth and digitalization of wireless communications networks. AT&T Wireless, created by our acquisition of McCaw Cellular, is investing substantially in creating an equal access mechanism for wireless access to all long distance carriers; in expanding the reach of wireless services to serve major markets nationwide; and in expanding the digital capability in our wireless access network and enhancing its ability to serve more users with higher service quality. The second area is globalization. Communications companies such as ours are entering into partnerships with their counterparts in several foreign countries to offer integrated services and one-stop shopping to customers.

### **AT&T Initiatives and Directions—Information Resources**

The two best-known types of electronic information content today are (1) electronic online services, which typically provide digitally encoded information delivered through narrowband access networks and accessed using computers, and (2) television programming, which is typically analog-coded information delivered through one-way broadband networks and accessed using TV sets. The promise of digital convergence reflects the vast potential that exists to create and store digital information and deliver it through two-way broadband access networks.

AT&T is a newly emerging player in the provision of online services. Our new AT&T Interchange Online Network offers its subscribers a rich collection of online information services on a range of special-interest topics from different publishers. In addition to general news and reference services, Interchange is working closely with publishing partners, such as the Washington Post, to offer online news and information for their target customers. Interchange's graphical user environment, hypertext links, and powerful searching capabilities make it a precursor for online services that will utilize the emerging broadband access media.

On another front, AT&T Bell Laboratories has developed interactive TV technologies that enable consumers to interact with the content that is delivered via their TV sets. Video server technology, developed and delivered through IDS, an AT&T partnership with Silicon Graphics, Inc. and Time Warner, Inc., is capable of storing huge quantities of multimedia information in digital form in the network for interactive uses.

An important aspect of making information resources useful for people is offering users the ability to navigate among different applications, to help find and retrieve the kind of information that users need most, when



they need it. To this end, we have recently developed AT&T PersonaLink Service, using intelligent agent technology developed in collaboration with General Magic, Inc. The service enables users with a small handheld device to determine how specific kinds of information should be dealt with—for example, high-priority electronic mail messages from chosen persons or on chosen subjects can be automatically forwarded to a specified computer, and the user's paging device will be notified automatically to alert him or her of the message. On the other hand, lower-priority messages can be sorted and stored for future retrieval, or forwarded as needed. We are at the very beginning of intelligent agent technology, and we expect that capabilities in this area and their use will grow rapidly over time.

At the nexus of information resources and communications networks is an idea that we call hosting, and this involves matching up information from a variety of content providers (information producers) with users (information consumers) no matter where they are. Key to hosting are wide reach, with wireline and wireless technology; open interfaces that interconnect multiple information resources with communications networks; and the navigation technologies referred to above, enabling users to easily sort through and obtain information they need when they need it. We are incorporating these concepts as we develop new products and services, and we intend to continue to support the principles of open, public interfaces so critical to customers and so necessary for competitive markets. We are actively participating in the Information Infrastructure Standards Panel (IISP), which is an industry group sponsored by the American National Standards Institute to identify standards gaps in today's evolving NII.

### Challenges and Uncertainties

#### Access Alternatives

Access technology alternatives for broad deployment deserve special attention, because they are such a fundamental enabler for many new service capabilities. Consumer applications of many of the above services will benefit immensely from the deployment of higher bandwidth access capabilities than those that exist today for supporting interactivity. The current installed base of access in telecommunications networks is predominantly twisted pair copper supporting symmetric two-way, low-speed analog transport. The current installed base for cable TV access is coaxial cable supporting high-bandwidth (hundreds of megahertz) analog transport one way (downstream only).

The most straightforward extensions of telephone network access involve leaving the twisted pair copper plant in place and digitizing the transport over them. Using basic-rate ISDN, up to 144 kilobits of aggregate bandwidth can be brought to homes and businesses. ISDN technology can thus support two multiuse (voice, data, or limited-speed video) channels to the home and one or more packet data channels. These would enable access to information resources with text and graphics. Basic-rate ISDN falls short of being suitable for full-motion, large screen video applications. Local telephone companies are beginning to offer basic-rate ISDN for residential consumers, though price packages and ordering processes are complicated, and user awareness and therefore "take-rates" are limited.

Another existing technology for extending the bandwidth of twisted-pair copper loops is asymmetrical digital subscriber line (ADSL). ADSL involves installing matching equipment at both ends of the customer loop that extend the bandwidth to the user substantially (typically ranging from 1.5 Mbps to 6 Mbps), while less bandwidth (typically one or two 56 kbps channels) will be available for upstream communication. ADSL's key advantage is that it can bring video services with some interactivity to users without replacing the embedded twisted-pair loop; since the installation and changes occur at the ends of the loop, ADSL devices can be disconnected from one user and moved to another user as, for example, when the first ADSL user decides to upgrade to a higher bandwidth medium such as fiber or fiber-coax access.

Hybrid fiber-coaxial cable combinations are gaining in popularity for new deployment. The reasons for their popularity for new deployment are that they offer abundant bandwidth capable of supporting integrated communications and entertainment applications; are comparable in initial deployment cost to new twisted pair copper access deployment when done on a large scale; and, by offering multiple services in an integrated

operation, administration, and maintenance environment, can help minimize ongoing costs relative to potential service revenues. By using the combination of fiber-optic feeders and coaxial cable "drops" to user locations, the strengths of each of these media are maximized. For example, fiber provides clean, high-bandwidth transmission; but coaxial cable (unlike fiber) can be used to supply electrical power to user terminals from the communications network and is also more flexible and robust for ease of deployment.

Hybrid fiber-coaxial cable combinations can be deployed in a variety of configurations supporting one-way, asymmetric two-way, and symmetric two-way services; these involve pure "bus" or "star" configurations, and combinations of the two.

As a developer of a range of access technologies, we believe that broadband access systems (such as hybrid fiber-coax) can be cost-effective alternatives to twisted pair copper access for new installations. They can be deployed for about \$1,000 to \$1,500 per home based on a broad-scale installation. Once deployed, they enjoy the advantages of being able to support traditional telephony, traditional TV, and emerging interactive multimedia applications on a single platform with a unified operations infrastructure. The actual deployment rate will depend on the extent of competition for provision of these services—which in turn will be determined by user needs and the evolution of the regulatory climate.

### Other Challenges

Among the obstacles and challenges, we have so far focused our attention on access capabilities and their deployment. We must also point out that there is significant uncertainty relating to user demand and willingness to pay for new interactive consumer services. Experience with early deployment of services will teach industry a lot about how the emerging services should be priced, packaged, and offered to users. It is hoped that revenue from new access services will be adequate to recover the investment associated with deployment of new access network capabilities. Based on preliminary analysis, this appears to hold true for integrated access capabilities.

Other uncertainties and challenges include laws and practices relating to privacy, security, liability, and intellectual property matters. An informed dialog between the public and private sector is essential to the emergence of appropriate public policy for the information age. This process has already begun, largely within the framework of the NII initiatives in industry and government.

## RECOMMENDATIONS

The challenge of developing the NII industries into a number of viable market areas and addressing user needs with new products and services remains, and should continue to remain, the province of private industry. We in AT&T, along with our counterparts in industry, are actively engaged in developing new capabilities and addressing user needs. However, the public sector (at the federal and state levels) does have a key role to play in allowing these capabilities to develop toward their full potential in a speedy fashion.

A major constraint on our ability to offer advanced service capabilities to consumers and small businesses is the lack of availability of full-fledged access alternatives. The public sector needs to remove today's overt and subtle impediments to the deployment and interconnection of competitive local exchange and access capabilities for consumers and businesses. The transition from the monopoly provision of local telecommunications services to an environment of multiple competing and interconnecting providers needs to be facilitated by legislators requiring the removal of constraints such as franchise restrictions, lack of interconnect points and standards, exchange resale prohibitions, lack of local number portability, and numerous other impediments. The public sector needs to work with the private sector to develop criteria and metrics to determine when a market is competitive. Regulatory efforts should be focused on opening markets to competition and doing so in a manner that inhibits the abuse of monopoly power where it exists. Hand in hand with enabling the emergence of competitive markets, the public sector needs to support industry in the development of open interfaces in critical NII locations where interoperability is necessary, and support industry-led standards for ensuring such interoperability.

The public sector needs to facilitate the ongoing availability of enabling resources such as wireless spectrum, numbers, and rights-of-way as new developments in the market strain limited resources. Flexibility and the ability to support market-driven solutions should in general be the guiding principles in these areas.

As a major user of the NII, the public sector needs to adopt open industry standards and leverage its considerable market power as major commercial users would, to advance innovation. It should avoid creating special networks and requirements without compelling reasons, as such efforts drain resources from the mainstream development of products and services in the commercial marketplace.

The public sector needs to enact laws that recognize the need for individual privacy and security of information in electronic form, and that protect intellectual property rights for information created and disseminated electronically. Although the United States can lead these efforts by example, we must recognize that these efforts are truly global in scope.

Regulatory efforts should be focused on opening markets to competition and doing so in a manner that inhibits the abuse of monopoly power where it exists.

## 48

# Enabling Petabyte Computing

*Reagan W. Moore*

*San Diego Supercomputer Center*

### STATEMENT OF THE PROBLEM

One expectation for the national information infrastructure (NII) is that information be readily available and accessible over the Internet. Large archives will be created that will make various types of data available: remote-sensing data, economic statistics, clinical patient records, digitized images of art, government records, scientific simulation results, and so on. To remain competitive, one must be able to analyze this information. For example, researchers will want to "mine" data to find correlations between desperate data sets such as epidemiology studies of patient records. Others will want to analyze remote-sensing data to develop better predictive models of the impact of government regulations on the environment. Still others will want to incorporate direct field measurements into their weather models to improve the models' predictive power. In each of these cases, the ability to manipulate massive data sets will become as important as computational modeling is to solve problems.

A new technology is needed to enable this vision of "data assimilation." It can be characterized as "petabyte computing," the manipulation of terabyte-size data sets accessed from petabyte-size archives<sup>1</sup> The infrastructure that will sustain this level of data assimilation will constitute a significant technological advantage. Because of recent advances in archival storage, it is now feasible to implement a system that can manipulate data on a scale a thousand times larger than is being attempted today. Unfortunately, the software infrastructure to control such data movement does not yet exist. An initiative to develop the corresponding software technology is needed to enable this vision by 2000.

### BACKGROUND

Advances in archival storage technology have made it possible to consider manipulating terabyte data sets accessed from petabyte storage archives. At the same time, advances in parallel computer technology make it possible to process the retrieved data at comparable rates. The combination of these technologies will enable a new mode of science in which data assimilation becomes as important as computational simulation is to the development of predictive models.

Data assimilation can be viewed as combining data mining (in which correlations are sought between large data sets) and data modeling (in which observational data are combined with a simulation model to provide an improved predictive system). These approaches may require locating a single data set within a data archive ("data picking") or deriving a data subset from data that may be distributed uniformly throughout the archive. The latter requires supporting the streaming of data through data-subsetting platforms to create the desired data set. Therefore, handling large data sets in this manner will become dependent on the ability to manipulate parallel I/O streams.

Current capabilities are exemplified by commercial applications of data mining in which companies maintain "just-in-time" inventory by aggressively analyzing daily or weekly sales. The analysis of sales trends allows a company to tune purchase orders to meet the predicted demand, thus minimizing the cost and overhead of maintaining a large inventory of goods. The amount of information analyzed is limited by current storage and database technology. Data sets up to 5 terabytes in size, consisting primarily of short transaction records, can be analyzed.

Examples of similar transaction-based data sets include bank check logging archives, airline ticket archives, insurance claim archives, and clinical patient hospital records. Each of these archives constitutes a valuable repository of information that could be mined to analyze trends, search for compliance with federal laws, or predict usage patterns.

The size of individual such archives is expected to grow to petabytes by 2000. Part of the growth in size is expected from the aggregation of information over time. But an important component of the size increase is expected to come from incorporating additional ancillary information into the databases. Clinical patient records will be augmented with the digitized data sets produced by modern diagnostic equipment such as magnetic resonance imaging, positron emission tomography, x-rays, and so on. Insurance claim archives will be augmented with videotapes of each accident scene. Check archives will be augmented with digitized images of each check to allow validation of signatures.

In addition, virtually all scientific disciplines are producing data sets of a size comparable to those found in industry. These data sets, though, are distinguished from commercial data sets in that they consist predominantly of binary large objects or "blobs," with small amounts of associated metadata that describe the contents and format of each blob. A premier example of such data sets is the Earth Observing System archive that will contain satellite-based remote-sensing images of the Earth<sup>2</sup>. The archive is expected to grow to 8 petabytes in size by 2006. The individual data sets will consist of multifrequency digitized images of the Earth's surface below the satellite flight paths. The multifrequency images will be able to be analyzed to detect vegetation, heat sources, mineral composition, glaciers, and many other features of the surface.

With such a database, it should be possible to examine the effect of governmental regulations on the environment. For example, it will be possible to measure the size of croplands and compare those measurements to regulations on land use policies or water usage. By incorporating economic and census information, it will be possible to measure the impact of restrictions of water allocations on small versus large farms. Another example will be to correlate crop subsidies with actual crop yield and water availability. Numerous other examples can be given to show the usefulness of remote-sensing data in facilitating the development of government regulations.

Remote-sensing data can also be used to improve our knowledge of the environment. An interesting example is calculating the global water budget by measuring the change in size of the world's glaciers and the heights of the oceans. This information is needed to understand global warming, better predict climate change, and predict water availability for farming. All these examples require the ability to manipulate massive amounts of data, both to pick out individual data sets of interest and to stream large fractions of the archive through data-subsetting platforms to find the appropriate information.

Further examples of large scientific data sets include the following:

- *Global change data sets.* Simulations of the Earth's climate are generated on supercomputers based on physical models for different components of the environment. For instance, 100-year simulations are created based on particular models for cloud formation over the Pacific Ocean. To understand the difference in the predictions of the global climate as the models are changed, time-dependent comparisons need to be made, both between the models and with remote-sensing data. Such data manipulations need support provided by petabyte computing.
- *Environmental data sets.* Environmental modeling of major bays in the United States is being attempted by coupling remote-sensing data with simulations of the tidal flow within the bays. Projects have been started for the Chesapeake Bay, the San Diego Bay, the Monterey Bay, and the San Francisco Bay. In each case, it should be possible to predict the impact of dredging policies on bay ecology. Through fluid dynamics simulations of the tides, it should be possible to correlate contaminant dispersal within the bay and compare the predictions with actual measurements. Each of these projects has the capability of generating terabytes to petabytes of data and will need the petabyte software infrastructure to support data comparisons.
- *Map data sets.* The Alexandria project at the University of California, Santa Barbara, is constructing a digital library of digitized maps. Such a library can contain information on economic infrastructure (roads, pipes, transmission lines), land use (parcels, city boundaries), and governmental policy (agricultural preserve boundaries). Correlating this information will be essential to interpret much of the remote-sensing data correctly.

The capacity of today's hardware and software infrastructure for supporting data assimilation of large data sets needs to be increased by a factor of roughly 1,000, as shown in [Table 1](#).

TABLE 1 Infrastructure Needed to Support Data Assimilation

Technology	Current Capacity	Needed Capacity
Data archive capacity	Terabytes	Petabytes
Communication rates	Megabytes/second	Gigabytes/second
Data manipulation	Gigabytes of data stored on local disk	Terabytes
Execution rate on computer platform	Gigaflops	Teraflops

Furthermore, increased functionality is needed to support manipulating data sets accessed from archives within databases. The development of the higher-capacity, greater-functionality petabyte computing system should be a national goal.

### ANALYSIS AND FORECAST

Technology advancements now make it possible to integrate analysis of massive observational databases with computational modeling. For example, coupling observed data with computer simulations can provide greatly enhanced predictive systems for understanding our interactions with our environment. This represents an advance in scientific methodology that will allow solution of radically new problems with direct application not only to the largest academic research problems, but also to governmental policymaking and commercial competitiveness.

The proposed system represents the advent of "petabyte computing"—the ability to solve important societal problems that require processing petabytes of data per day. Such a system requires the ability to sustain local data rates on the order of 10 gigabytes/second, teraflops compute power, and petabyte-size archives. Current systems are a factor of 1,000 smaller in scale, sustaining I/O at 10 to 20 megabytes/second, gigaflops execution rates, and terabyte-size archives. The advent of petabyte computing will be made possible by advances in archival storage technology (hundreds of terabytes to petabytes of data available in a single tape robot) and the development of scalable systems with linearly expandable I/O, storage, and compute capabilities. It is now feasible to design a system that can support distributed scientific data mining and the associated computation.

A petabyte computing system will be a scalable parallel system. It will provide archival storage space, data access transmission bandwidth, and compute power in proportional amounts. As the size of the archive is increased, the data access bandwidth and the compute power should grow at the same rate. This implies both a technology that can be expanded to handle even larger data archives as well as one that can be reduced in scope for cost-effective handling of smaller data archives. The system can be envisioned as a parallel computer that supports directly attached peripherals that constitute the archival storage system. Each peripheral will need its own I/O channel to access a separate data-subsetting platform, which in turn will be connected through the parallel computer to other compute nodes.

The hardware to support such a system will be available in 1996. Data archives will be able to store a petabyte of data in a single tape robot. They will provide a separate controller for each storage device, allowing transmission of large data sets in parallel. The next-generation parallel computers will be able to sustain I/O rates of 12 gigabytes/second to attached peripheral storage devices, allowing the movement of a petabyte of data per day between the data archive and the compute platform. The parallel computers will also be able to execute at rates exceeding 100 gigaflops, thus providing the associated compute power needed to process the data. Since the systems are scalable, commercial grade systems could be constructed from the same technology by reducing the number of nodes. Systems that support the analysis of terabytes of data per day could be built at greatly reduced cost.

What is missing is the software infrastructure to control data movement. Mechanisms are needed that allow an application to manipulate data stored in the archive through a database interface. The advantages provided by such a system are that the application does not have to coordinate the data movement; data sets can be accessed through relational queries rather than by file name; and data formats can be controlled by the database, eliminating the need to convert between file formats. The software infrastructure needed to do this consists of a library interface between the application and the database to convert application read and write requests into SQL-\* queries to the database, an object-relational database that supports user-extensible functions to allow application-specific manipulations of the data sets, an interface between the database and the archival storage system to support data movement and control data placement, and an archive system that masks hardware dependencies from the database. SQL-\* is a notation that has been developed in a previous project for a future ANSI-standard extended Standard Query Language.

Each of these software systems is being built independently of the others, with a focus on a system that can function in a local area network. The petabyte computing capability requires the integration of these technologies across a wide area network to enable use of data-intensive analyses on the NII. Such a system will be able to support access to multiple databases and archives, allowing the integration of information from multiple sources.

Technology advances are needed in each of the underlying software infrastructure components: database technology, archival storage technology, and data-caching technology. Advances are also needed in the interfaces that allow the integration of these technologies over a wide area network. Each of these components is examined in more detail in the following sections.

### Database Technology

A major impediment to constructing a petabyte computing system has been the UNIX file system. When large amounts of data that may consist of millions of files must be manipulated, the researcher is confronted with the need to design data format interface tools, devise schemes to keep track of the large name space, and develop scripts to cache the data as needed in the local file system. Database technology eliminates many of these impediments.

Scientific databases appropriate for petabyte computing will need to integrate the capabilities of both relational database technology and object-oriented technology. Such systems are called object-relational databases. They support queries based on SQL, augmented by the ability to specify operations that can be applied to the data sets. They incorporate support for user-defined functions that can be used to manipulate the data objects stored in the database. The result is a system that allows a user to locate data by attribute, such as time of day when the data set was created or geographic area that the data set represents, and to perform operations on the data.

To be useful as a component in the scalable petabyte system, the database must run on a parallel platform, and support multiple I/O streams and coarse-grained operations on the data sets. This means the result of a query should be translated into the retrieval of multiple data sets, each of which is independently moved from the archive to a data-subsetting platform where the appropriate functions are applied. By manipulating multiple data sets simultaneously, the system will be able to aggregate the required I/O and compute rates to the level needed to handle petabyte archives. Early versions of object-relational database technology are available from several companies, such as Illustra, IBM, and Oracle, although they are designed to handle gigabyte data sets.

A second requirement for the object-relational database technology is that it be interfaced to archival storage systems that support the scientific data sets. Current database technology relies on the use of direct attached disk to store both data objects and the metadata. This limits the amount of data that can be analyzed. Early prototypes have been built that store large scientific data sets in a separate data archive. The performance of such systems is usually limited by the single communication channel typically used to link the archival storage system to the database.

The critical missing software component for manipulating large data sets is the interface between the database and the archival storage system. Mechanisms are needed that will allow the database to maintain large

data sets on tertiary storage. The envisioned system will support transparent access of tertiary storage by applications. It will consist of a middleware product running on a supercomputer that allows an application to issue read and write calls that are transparently turned into SQL-\* queries. The object-relational database processes the query and responds by applying the requested function to the appropriate data set. Since the database and archival storage system are integrated, the application may access data that are stored on tertiary tape systems. The data sets may in fact be too large to fit on the local disk, and the query function may involve creating the desired subset by using partial data set caching. The result is transparent access of archived data by the application without the user having to handle the data formatting or data movement.

The missing software infrastructure is the interface between object-relational databases and archival storage systems. Two interfaces are needed: a data control interface to allow the database to optimize data placement, grouping, layout, caching, and migration within the archival storage system; and a data movement interface to optimize retrieval of large data sets through use of parallel I/O streams.

Prototype data movement interfaces are being built to support data movement between object-relational databases and archival storage systems that are compliant with the IEEE Mass Storage System Reference Model. These prototypes can be used to analyze I/O access patterns and help determine the requirements for the data control interface.

### **Archival Storage Technology**

Third-generation archival storage technology, such as the High Performance Storage System (HPSS) that is being developed by the DOE laboratories in collaboration with IBM, provides most of the basic archival storage mechanisms needed to support petabyte computing. HPSS will support partial file caching, parallel I/O streams, and service classes. Service classes provide a mechanism to classify the type of data layout and data caching needed to optimize retrieval of a data set. Although HPSS is not expected to be a supported product until 1996 or 1997, this time frame is consistent with that expected for the creation of parallel object-relational database technology.

Extensions to the HPSS architecture are needed to support multiple bit file movers on the parallel data-subsetting platforms. Each of the data streams will need a separate software driver to coordinate data movement with the direct attached storage device. The integration of these movers with the database data access mechanisms will require understanding how to integrate cache management between the two systems. The database will need to be able to specify data control elements, including data set groupings, data set location, and caching and migration policies. A generic interface that allows such control information to be passed between commercially provided databases and archival storage systems needs to become a standard for data-intensive computations to become a reality.

Other national research efforts are investigating some of the component technologies: the Scalable I/O Initiative at Caltech and the National Science Foundation MetaCenter. The intent of the first project is to support data transfer across multiple parallel I/O streams from archival storage to the user application. The second project is developing the requisite common authentication, file, and scheduling systems needed to support distributed data movement.

### **Data-Caching Technology**

Making petabyte computing available as a resource on the NII to access distributed sources of information will require understanding how to integrate cache management across multiple data-delivery mechanisms. In the wide area network that the NII will provide, the petabyte computing system must integrate distributed compute platforms, distributed data sources, and network buffers. Caching systems form a critical component of wide area network technology that can be used to optimize data movement among these systems. In particular, the number of times data are copied as they are moved from a data source to an application must be minimized to relieve bandwidth congestion, improve access time, and improve throughput. This is an important



research area because many different data delivery mechanisms must be integrated to achieve this goal; archival storage, network transmission, databases, file systems, operating systems, and applications.

Data-intensive analysis of information stored in distributed data sources is an aggressive goal that is not yet feasible. Movement of a petabyte of data per day corresponds to an average bandwidth of 12 gigabytes/second. The new national networks transmit data at 15 megabytes/second, with the bandwidth expected to grow by a factor of 4 over the next 2 to 3 years. This implies that additional support mechanisms must be provided if more than a terabyte of data is going to be retrieved for analysis. The implementation of the petabyte computing capability over a wide area network may require the installation of caches within the network to minimize data movement and support faster local access. Management of network caches may become an important component of a distributed petabyte computing system. Until this occurs, data-intensive analyses will have to be done in a tightly coupled system.

### Future Systems

A teraflops computer will incorporate many of the features needed to support data-intensive problems. Such a system will provide the necessary scalable parallel architecture for computation, I/O access, and data storage. Data will flow in parallel from the data storage devices to parallel nodes on the teraflops computer. Most such systems are being designed to support computationally intensive problems.

The traditional perspective is that computationally intensive problems generate information on the teraflops computer, which is then archived. The minimum I/O rate needed to sustain just this data archiving can be estimated from current systems by scaling from data flow analyses of current CRAY supercomputers<sup>3</sup>. For the workload at the San Diego Supercomputer Center, roughly 14 percent of the data written to disk survives to the end of the computation and 2 percent of the generated data is archived. The amount of data that is generated is roughly proportional to the average workload execution rate, with about 1 bit of data transferred for every 6 floating-point operations<sup>3</sup>. Various scaling laws for how the amount of transmitted data varies as a function of the compute power can be derived for specific applications. For three-dimensional computational fluid dynamics, the expected data flow scales as the computational execution rate to the 3/4 power.

Using characterizations such as these, it is possible to project the I/O requirements for a teraflops computer. The amount of data movement from the supercomputer to the archival storage system is estimated to be between 5 and 35 terabytes of data per day. If the flow to the local disk is included, the amount will be up to seven times larger. A teraflops computer will need to be able to support an appreciable fraction of the data movement associated with petabyte computing.

This implies that it will be quite feasible to consider building a computer capable of processing a petabyte of data daily within the next 2 years. Given the rate of technology advancement, the system will be affordable for commercial use within 5 years. This assumes that the software infrastructure described previously has been developed.

The major driving force to develop this capability is the vision that predictive systems based on data-intensive computations are possible. This vision is based on the technological expertise that is emerging from a variety of national research efforts. These include the NSF/ARPA-funded Gigabit Testbeds, the Scalable I/O Initiative, NSL UniTree/HPSS archival storage prototypes, and the development of the MetaCenter concept. These projects can be leveraged to build a significant technological lead for the United States in data assimilation.

### RECOMMENDATIONS

A collaborative effort is needed between the public sector and software vendors to develop the underlying technology to enable analysis of petabyte data archives. This goal is sufficiently aggressive and incorporates a wide enough range of technologies that no single vendor will be able to build a petabyte compute capability. The implementation of such a capability can lead to dramatic new ways for understanding our environment and the impact our technology has upon the environment. Building the software infrastructure through a public sector

initiative will allow future commercial systems to be constructed more rapidly. With the rapid advance of hardware technology, commercial versions of the petabyte compute capability will be feasible within 5 years.

### REFERENCES

- [1] Moore, Reagan W. 1992. "File Servers, Networking, and Supercomputers," *Advanced Information Storage Systems*, Vol. 4, SDSC Report GA-A20574.
- [2] Davis, F., W. Farrell, J. Gray, C.R. Mechoso, R.W. Moore, S. Sides, and M. Stonebraker. 1994. "EOSDIS Alternative Architecture," final report submitted to HAIS, September 6.
- [3] Vildibill, Mike, Reagan W. Moore, and Henry Newman. 1993. "I/O Analysis of the CRAY Y-MP8/864," *Proceedings of the 31st Semi-annual Cray User Group Meeting*, Montreaux, Switzerland, March.

## 49

# Private Investment and Federal National Information Infrastructure Policy

Organization for the Protection and Advancement  
of Small Telephone Companies (OPASTCO)

### INTRODUCTION AND STATEMENT OF THE PROBLEM

The Organization for the Protection and Advancement of Small Telephone Companies (OPASTCO) is a national trade association representing nearly 450 small, independently owned and operated telephone systems serving primarily rural areas of the United States and Canada. OPASTCO membership includes both commercial and cooperative telephone companies, which range in size from fewer than 100 to 50,000 access lines and collectively serve nearly 2 million customers.

This white paper examines network deployment, specifically, the networks used to access information providers, the location of customers trying to access the information, and the problems associated with connecting to the information providers from rural areas. The information providers' databases are unique entities that are separate from the network providers and their networks used to access them and are beyond the scope of this paper.

Further, this paper suggests solutions to networking technology investment problems. The following five such problems, or barriers, typical to small, rural telephone companies are discussed in this paper:

1. Limited local calling areas requiring toll access,
2. Long loops with load coils and station carrier,
3. Interconnection availability and distance,
4. Limited but great needs, and
5. High cost due to poor economies of scale.

### EXISTING NETWORK DEPLOYMENT AND DEFINITIONS

The information market consists of information providers, network providers, and customers. While OPASTCO and its members fall into the category of network providers, they may have some freedom that the regional Bell operating companies (RBOCs) do not because of Modified Final Judgment (MFJ) restrictions that apply only to GTE and the RBOCs. Information providers rarely provide their own network. Instead, they lease access from network providers that add a surcharge based on the number of minutes connected or packets transferred. In metropolitan areas, customers dial a local number or lease an inexpensive high-speed local direct connection to the local node for higher speed to information providers. In most rural areas, the customer is forced to dial a long-distance number and pay per-minute toll charges directly, or dial an 800 number and pay an hourly surcharge to the information provider. Having toll and 800 number access usually doubles or triples that cost of access depending on the minimum hourly or subscription cost that the information provider charges.

Today, data networks fall into three categories: private data, data packet, and high-speed open interface. The first type, private data, are networks that companies and government units build or lease. These networks may be privately owned and built or leased from local exchange carriers (LECs), interexchange carriers (IXCs), and data packet network providers, or any combination of the three. Private data networks can be truly separate

networks or virtual private networks sharing equipment, but they cannot be truly interconnected except through an optional gateway node. These networks are relatively expensive, with high start-up costs. However, such networks exist, and if the company or government unit can justify the need, there are few technical problems or problems obtaining return on investment by small telephone companies in providing this type of network.

The second type of networks are first generation, usually x.25, special data packet networks. Tymnet and Telenet are examples of special data packet network providers. These networks are provided by IXCs, large regional LECs, and special data packet network providers. The high cost of the initial investment, limited speed and capabilities, and limited willingness to interconnect small nodes to the existing networks have limited the existing participation in providing data packet network access by small, rural telephone companies. High-speed open interface is the third type of networks. Router and ATM switched networks, as well as the Internet, are examples of high-speed open interface networks. The Internet is an interconnection of numerous networks supported by regional, nationwide, and worldwide networks that provide access to a host of information sources, as well as electronic mail to anyone connected to the "net." However, the Internet in its existing configuration and capability is not the complete functional network that is needed to provide the full range of services desired by information providers, private industry, universities, and other small business and residential information users.

## ANALYSIS OF PROBLEMS AND ALTERNATIVE SOLUTIONS

### Barrier 1: Limited Local Calling Areas Requiring Toll Access

The key to affordable information access for residential and small business customers is local access to the networks that provide access to the information providers. Unfortunately, few rural communities have network nodes. Only those small exchanges close enough to metropolitan areas to have extended area service can access the network providers with a local or toll-free call. Many special data packet network providers place nodes in calling areas that have a population of at least 100,000, but some require areas to have a population of at least 250,000.

Special data packet providers are telecommunications providers that use local networks to connect customers with various information providers for a profit. Normally, they order standard business lines from the LEC and install banks of modems concentrated with packet assemblers and share the backbone transmission link between metropolitan nodes. IXCs are required to pay the LEC for originating and terminating access on a flat rate per trunk or on a per-minute-of-use basis. The packet data network providers currently are exempted from paying access revenues to the LEC.

One factor to consider when planning to provide local data access from a small telephone company is the lost toll revenue from current subscribers to information service providers. In addition, if local access is provided and the volume of local calling increases significantly because of the highly elastic nature of information services, there will be an increased investment in equipment and a decrease in revenue. For small telephone companies that have cost-based toll settlements through the National Exchange Carrier Association, the impact of the loss in direct toll revenue and the jump in local calling dramatically shifts the revenue requirement from the toll to local, which forces local rates to increase. Average schedule companies experience the loss in toll revenue but are not aware of the negative double shift in revenue impacts of increased local usage. Any national policy to foster investment in infrastructure should assess the impacts to both cost and average schedule companies.

Integrated services digital network (ISDN) has not been deployed to any extent in rural markets because the switching systems commonly used by small telephone companies do not have ISDN available yet or they have just recently made it available, or the cost to add ISDN services is too high for the limited rural market. Even overlay networks that allow a small initial investment to provide small quantities of ISDN service are high in cost when considered on a per-line basis. As the ISDN market penetration increases throughout the country, increased pressure to provide ISDN in the rural markets will occur. Unfortunately, ISDN services are basically local services, and unless the network provider has a local ISDN node, the high cost of ISDN toll access remains a barrier. Fortunately, in the last year, 28,800-baud modems with error correction and compression have improved voice-grade lines' data access speeds considerably.

The only technical solution to the local access barrier that exists today is to provide local access. The following is an examination of the barriers that should be reduced to make local access more feasible.

### **Barrier 2: Long Loops with Load Coils and Station Carrier**

Both load coils and analog carrier limit data capability to voice grade circuits. The maximum bandwidth of all voice grade lines is 4,000 Hz, which may limit data transmission to 7,200 baud or less. Analog carrier in addition to bandwidth limitations also has limited custom local area signaling service (CLASS) capability and for 5 years has been in an industry-wide phaseout. Only within the last 2 years has there been any trend toward eliminating loaded loops.

According to the *USDA-REA 1992 Telephone Loop Study*, the average loop length was 4.11 miles, and 17.84 percent were greater than 6.1 miles. In 1969, the average loop length was 3.3 miles, and about 20 percent of the loops were greater than 6.1 miles; at the same time, less than 2 percent of the Bell companies' loops were greater than 6.1 miles, and they averaged only 2.16 miles. When there was an effort to reduce loaded loops, the increase in average loop length was probably due to a conversion from multiparty lines to all one-party service.

The USDA-REA study also showed that in 1992 38 percent of the REA loops studied used load coils to enhance voice transmission, which was down only 1 percent from 39 percent in 1985. Fortunately, in 1992, only 2.1 percent of the loops studied were served via analog station carrier.

Long rural loops with load coils and analog station carrier limit dial-up access only slightly, but they do not allow for ISDN or higher data speeds. If only one or a few lines of high-speed data are required, then special data circuits can be built, but only at high per-circuit cost. If ubiquitous service is required, an entire system redesign may be required.

For the last 10 years, there has been a trend to minimize loaded loops, but only within the last 2 years have rural, independent telephone companies considered totally eliminating loaded loops with remotes or digital serving areas and loops limited from 12,000 to 18,000 feet. Until now, these designs have been implemented to reinforce for high growth or to transition fiber in the loop to replace deteriorated air core cables.

If ubiquitous high-speed data service is the objective, then considerable investment in digital serving area facilities will be required. An evolving definition of universal service and a continuation of the Universal Service Fund (USF) for high-cost areas will be necessary to foster the added investment in the infrastructure.

### **Barrier 3: Interconnection Availability and Distance**

One factor in availability is the distance from rural exchanges to the nearest network providers. Network providers have established access nodes in most metropolitan areas with a population of 100,000. The distance from these access nodes to rural exchanges can vary from 20 to 500 or more miles. The cost for transport of 56 kbps and T-1 or higher speed data circuits is distance dependent. If multiple incompatible or proprietary network connections are required, then the already high cost for transport multiplies.

Another factor in availability is finding a willing network access provider to connect to an independent telephone company network. To date, only regional Internet providers have been willing to provide direct independent data network connections.

One solution to the lack of available network connections would be a single standardized network enabling a common transport link, as well as lower costs to build in the desired redundancy. Another solution would be to establish regional hub nodes that would decrease the distance of transport, thus allowing a number of companies to share the cost of common transport to the metro nodes. One example of a regional hub network has been implemented by Minnesota Equal Access Network Services (MEANS) and Iowa Network Systems (INS). Together the companies have built statewide networks to share in the cost of nodes and transport.

#### **Barrier 4: Limited But Great Needs**

While there may not be a great demand in rural areas for communications in terms of numbers and volume, there is a great demand for communications, and data communications in particular, due to the travel distances and the need to be competitive with urban areas. Rural areas strive to expand employment and increase economic development, and providing up-to-date communications is imperative for high-technology jobs and is needed to attract business to rural areas.

There is no solution to the problem of travel distances and the need for communications in rural areas beyond providing for the basic communications needs of rural customers.

#### **Barrier 5: High Cost Due to Poor Economies of Scale**

The average OPASTCO member company serves approximately 6,000 customers, and the average exchange or wire center has approximately 1,200 access lines; compare that to a population of 100,000 or even 250,000, which is necessary for some data packet network providers to invest in a node. The high minimum cost of investment coupled with comparatively few customers and a low volume of data traffic make the cost per customer or volume of traffic much higher than in urban areas. In addition to the high cost of investment in equipment and transport, an even higher cost may be required for test equipment, training for maintenance staff, and training and staffing for customer support. In addition to the training of support staff, an additional cost or barrier to marketing data network technology in rural areas is educating customers to the benefits and general use of computers.

One partial solution would be standardized modular node equipment manufactured in volumes large enough to reduce the cost per node. Again, if single network and regional hubs were implemented, the costs per customer could be reduced. Shared networks and even a shared pool of test equipment and support personnel could further reduce costs.

Continued application of dial equipment minutes (DEM) weighting and further application to data networking equipment are existing mechanisms for recovering the cost of investment and keeping the costs to rural customers at an affordable level.

#### **RECOMMENDATIONS AND SUMMARY**

Standards development, standard equipment, regional hubs, and a single common data network are partial solutions to the problems of providing affordable data network access to rural areas. Further, an evolving definition of universal service and the continued or expanded application of the USF coupled with continued or expanded DEM weighting are necessary to provide affordable information services to rural America.

## 50

# Thoughts on Security and the NII

*Tom Perrine*

*San Diego Supercomputer Center*

### STATEMENT OF THE PROBLEM

The rapid introduction of the Internet into U.S. (and global) society has challenged the knowledge, ethics, and resources of the culture. Educational activities, both traditional and in many new forms, are rapidly making information about the Internet and personal computing and communication widely available. The ethical considerations are being addressed by organizations such as the Electronic Frontier Foundation (EFF) and the Computer Professionals for Social Responsibility (CPSR). There is also a renewed emphasis on ethics in the technical communities, as well as a growing understanding of technical issues in legislation and law, as these areas struggle to adapt to and codify new issues raised by emerging technologies.

The Internet has many of the characteristics of a frontier, including a dearth of security and law-enforcement services. This discussion focuses on the security mechanisms that must be developed over the next 5 to 10 years to make the Internet (and its successors) a safe computing and communications environment for individuals and to protect the commercial interests of the businesses beginning to establish themselves on the Internet.

### BACKGROUND

The Internet is becoming an increasingly popular medium for delivering products, services, and personal communications. Unfortunately, none of these commercial or personal activities were anticipated by the original design of the Internet protocols or by the architecture of the new class of common carriers, the Internet service providers (ISPs).

The Internet has become a new frontier for many Americans. Like any frontier, most of the inhabitants are peaceful, interested only in exploration and settlement. But, like any frontier, a minority of inhabitants are more interested in exploiting the more peaceful inhabitants. Another inevitable consequence of a frontier is the (initial) inability of law enforcement to keep pace with the rapid expansion in the number of inhabitants. If all of this sounds like the American Old West, it is not a coincidence.

Networking and computing as communications services have created new problems, and put a new spin on old problems, in the security and law-enforcement resources of the American society. These problems can be addressed on three levels: threat and protection models, deterrents, and law-enforcement resources.

#### Threat and Protection Models

All security practices depend on the development of a "threat model," which details foreseeable risks and threats. Then a "protection model" is developed to address the perceived threats and risks, tempered by additional factors such as law, policy, and costs. Traditional models of both threats and protection have had flaws that have increased the cost of secure computer systems and networks.

Threat models that have been developed for computer and network security in the past have reflected a "laundry list" of potential threats, with no regard for the cost (to the attacker) of any particular attack method. In

other words, all threats have been considered equally likely, even if the cost of producing an attack might be prohibitive. If a threat is considered "possible," it must be addressed by the protection model.

Protection models have not been without their problems, as well. Historically, most attempts at building secure computer systems and networks have followed the "castle" model: build high, thick walls with a few well-understood gates. This paradigm is reflected in the terminology used in information security: firewall, bastion host, realm, password, domain, and Trojan horse.

This mind-set limits the ideas that can be discussed and thus the tools that will be developed. Furthermore, approaches focused on prevention are limited to the scope of the modeled threats and typically are strictly reactive to demonstrated examples of these threats. But, to date, no sufficient threat models have been developed. This approach is the epitome of passive defense, which is not a viable strategy in the long term as advances in offensive technologies will always overwhelm a static defense. To go beyond this focus on prevention to encompass investigation and prosecution, we need to consider alternate modes of thought about information security.

### **Deterrents**

A deterrent is anything that deters a person from performing some undesirable action. It can be as simple and direct as a padlock, or as indirect as strict punishments if a person is caught and convicted.

Traditional, technical, computer and network security has focused on building better "locks," stronger "doors," and so on. Until recently, crimes committed via computer or network were almost impossible to prosecute. The laws were silent on many issues, the courts (including juries) were uneducated concerning computers and networks in general, and law enforcement for such white-collar crimes was seen as less critical than that for violent crime.

With more awareness of the Internet, the spread of home computers, and increasing reliance on computing resources for day-to-day business, there has been a popular push for more legal deterrents (laws) and for better education for judges, attorneys, and law-enforcement personnel. As a result of increased media attention to the Internet and more computers in homes, schools, and business, it is now no longer impossible to get a jury capable of understanding the cases.

### **Law-Enforcement Resources**

Law-enforcement resources will always be at a premium, and crimes against property will always (rightfully) be of less importance than violent crime. As a result, computer and network crimes will always be competing for resources against violent crimes and other, more easily prosecutable ones. In other words, only the largest, most flagrant computer crimes will ever be considered in a courtroom.

### **Analysis and Forecast**

Over the next 5 to 7 years, the Internet will most likely become the de facto national information infrastructure (NII). Talk of hundreds of channels of TV, videophones, and so on will continue; but it is access to people and data on demand that has driven and will continue to drive the growth of the Internet. The Internet is here, and it works. New technologies such as integrated services digital network (ISDN) and asynchronous transfer mode (ATM), higher-speed links, and new protocols such as "IPng" (Internet Protocol—Next Generation) will become part of the Internet infrastructure, but it is unlikely that a separate, parallel network of networks will be constructed.

The problems of making the Internet a safe computing environment will require significant research and development in the areas discussed above: threat and protection models, deterrents, and law-enforcement resources.



### Threat and Protection Models

Tsutomu Shimomura (San Diego Supercomputer Center), Whit Diffie (Sun Microsystems), and Andrew Gross (San Diego Supercomputer Center) have recently proposed a completely new approach to computer and network security. This new model actually combines the threat and protection models into a new model referred to as the confrontation model.

A new research activity at the San Diego Supercomputer Center, undertaken as a cooperative venture between academia, government, and industry, will soon begin exploring an approach to information security based on confrontation in which we engage the intruder by using winning strategies within the scope of policy. Hence, we call our model the confrontation model. As alluded to above, many of our ideas come from conflict-type situations such as might be found in business, intelligence work, law enforcement, and warfare, and so we draw on all these areas for ideas and examples. The research for this new paradigm will require developing both strategies and tactics.

Using the paradigm of an intrusion as a confrontational situation, we can draw from centuries of experience in warfare. The network and other infrastructure are the "terrain" upon which our "battles" are fought. From a tactical viewpoint, certain resources will be more valuable than others (e.g., fast CPUs for analysis, routers to change the topology of the terrain, and critical hosts near the activity for intelligence gathering). We need to know the terrain, make it easy to monitor, and use it to our advantage against intruders. Once we understand the terrain, we can plan infrastructure changes that allow us to control it or position ourselves strategically within the terrain, and thus make it easier to counter intrusions.

Executing strategies within the terrain is complicated by the need to adequately identify an intruder's intent. Confused users may at first appear to be hostile, while real intruders may try to hide within the terrain. To represent this, traditional threat models must be amended to incorporate the extended terrain.

A proactive approach is needed that simultaneously considers the "terrain" in which the engagement is occurring, the disposition of resources to counter intrusions most effectively, and a cost-benefit analysis of countermeasure strategies. Such an approach to information security proved successful in the apprehension of wanted computer criminal Kevin Mitnick. Note that all conflict occurs within the scope of policy. Such policies include criminal law and its rules of evidence. In business, they include contract law, civil procedure, and codes of business ethics.

In addition to understanding the "warfare" context, there is also a need to communicate with and become part of existing law-enforcement structures. Instead of trying to adjust law enforcement to fit the peculiarities of computer crime, we need to adjust the way we think about computer security to more accurately match the law-enforcement model to facilitate prosecution of computer crimes.

### Deterrents

New deterrents will be developed over the next 5 to 7 years. Many of these will be in the form of stronger doors and locks. These technical advances will come from research in many different areas and can be expected to proceed at a rapid pace.

It is expected that such proactive technical measures, leading to identification and prosecution of intruders, will be an effective deterrent. If intruders are aware of the risk they incur when attempting to compromise computer systems and networks, they may modify their behavior.

More important, however, are the societal deterrents: ethics and law. A more vigorous campaign of educating business and the public will need to be undertaken. This education will need to focus on privacy rights, intellectual property rights, and ethics in general. It is not unreasonable that every computer education course of study include an ethics component. This is already starting to happen in many engineering and computer science curricula.

The law of the land will require updating, not wholesale change, to accommodate the digital landscape. However, instead of knee-jerk reactions to highly publicized events (child pornography on computers, etc. that have resulted in laws dealing specifically with the Internet and computers, we need expansion or reinterpretation

of existing laws in the light of computers and networks. If something is already illegal, why should there be a separate law making such an act illegal when a computer or network is involved?

Increasing the ability of existing enforcement structures to initiate and carry through successful prosecution of crimes that happen to involve computers and networks will indirectly increase the deterrence to commit such crimes. This will require educating existing judicial personnel, as well as changes in policies and procedures, and increased resources as well.

### **Law-Enforcement Resources**

As already noted, law-enforcement resources will always be at a premium. There will never be enough law-enforcement resources to fully investigate every crime, and crimes against property, including computer crime, will always (rightfully) be of less importance than violent crime. But this limitation primarily refers to government law-enforcement resources.

As on the American and other frontiers, one solution will almost certainly be resurrected: private security forces. Just as the American frontier had its Pinkerton agents and Wells Fargo security, the Internet will soon have private investigative and security organizations. In fact, the Internet already has the equivalent of private security agents: the consultants and companies that deal with computer and network security. These agents perform such work as establishing "safe" connections to the Internet for companies and providing security software, intrusion-detection, and auditing software and hardware, and so on.

But what about the investigative side?

As part of the research on a confrontation model mentioned above, there is growing commercial interest in private investigative services to perform intrusion analysis and evidence gathering, for use in civil or criminal proceedings. The confrontation model will lead to technical solutions (tools) that will be available to both governmental and private investigative services.

A recent Defense Advanced Research Projects Agency (DARPA) Broad Area Announcement (BAA) stressed the desire to commercialized computer security services, including the detection of intrusions and the tracing of intrusions to their source (perpetrator). At least two existing companies are investigating entering this field.

## **RECOMMENDATIONS**

The government must support open, public security standards and mechanisms. It must remove inappropriate impediments to private-sector development of security technologies, including encryption. This approach will require support of research activities, legislative changes, and increased awareness of how digital communications change the law-enforcement landscape.

### **Research Activities**

The government must foster more research into new protection strategies, and this work must be done in conjunction with the private sector. The computer industry is well aware of the problems and is (finally) being driven by market forces (consumer demand) to increase the security of its products.

However, the computer industry does not always have access to the proper theoretical groundwork, and so academia and government must find ways to cooperatively develop open standards for security software and hardware. This will inevitably lead to more joint research efforts, which may require revisiting the current interpretations of some antitrust laws.

As part of cooperative research and development, testbeds need to be built to provide a better understanding of the battleground. This understanding will enable us to predict the types of intrusion strategies that can be expected and will allow us to develop appropriate counter strategies. A better understanding of intrusions will allow us to better predict the intruder's intent. Given what we believe the intent to be, we then need

mechanisms to identify an appropriate response, appropriate for the chosen policy. For instance, if we identify an intent to access critical resources, then the response may need to support more comprehensive data collection to facilitate prosecution.

To be a viable platform for analyzing the confrontation paradigm, any proposed testbed must be a collection of hardware and software systems that encompass the complexity and extent of today's networking infrastructure. The testbed will be a heterogeneous collection of vendor computer platforms, network routers, switches, firewalls, operating systems, and network applications. These are the terrain in which a confrontation occurs.

Understanding the range of intrusions is required to build credible defenses. Insight must be developed for both the feasible intrusion mechanisms and the types of countermeasures that should be pursued. This insight must quantify the cost of an intrusion, the cost of the countermeasure, and the level of risk that is being reduced. A cost-benefit analysis is needed to understand the best possible response. A testbed serves as a tool to quantify the risk associated with providing desired services and allows the development of mechanisms to reduce that risk. Once the risks are quantified, it should be possible to create systems of graduated resilience as a function of the provided services.

The testbed will be used for "war games," actual intrusion attempts against both current and emerging technology. One person can develop an intrusion mechanism and distribute it widely on the network, resulting in a widespread problem that puts our entire infrastructure at risk. An equally wide distribution of defensive abilities is needed to counter this. Evaluation of successful intrusions from the games will show where effort should be put to best bolster system security. The system bolstering can be in the form of cryptography, better programming standards, and a better understanding of the actual system functionality. Vulnerabilities can be created when a developer's perception of the function of the system differs from its actual function.

As a product of the analyses done in the security testbed, prototype mechanisms will be developed. An application of the confrontation paradigm was used by Shimomura and Gross to analyze the flaws exploited in the intrusion of Shimomura's computers on December 25, 1994. Their analysis resulted in an understanding of the "address-spoofing" technique that was used. The tools, most of which they developed on the fly, focused on two areas: noninvasive examination of the preserved system state of the compromised computers and packet-trace analysis. Understanding the initial intrusion mechanism and the goals of the intruder required analyzing the situation with minimal disruption to the traces left by the intrusion. These tools enabled an appropriate response to this particular intrusion. Other intrusions may require a different tool set.

It is important to note that although tools exist to examine the integrity of a suspected compromised host (for example, TRIPWIRE), they all rely on computing cryptographic check sums. This computation requires reading all the critical files, which destroys all access time stamps in the file system. In some cases, it may be appropriate to have a toolset that examines the system kernel memory and all on-disk structures noninvasively, preserving all available information for further analysis (and as evidence).

The confrontation paradigm provides a framework that can be used to understand intrusions. The actual mechanisms may be built from scratch, such as reconstructing data sets that were "deleted" from a disk. Or they may be built by modifying existing security tools such as logging mechanisms. For example, logs of packets seen on a network were constructed to reproduce all the simultaneous sessions, either keystroke by keystroke or at least packet by packet. (These tools are capable of correlating the simultaneous activities of multiple sessions to trace their interactions on a target computer system or network.) Playback of the sessions in real time was helpful in understanding what the intruder was trying to accomplish and his relative level of sophistication. Analysis of other intrusion mechanisms may require the construction of a different set of tools. In this case, loss of packet logs necessitated a more subtle and thorough analysis.

Analysis tools have been developed that extract relevant log records from centralized log facilities. Sophisticated pattern matching tools were built to monitor for suspicious or unusual events. Such pattern matching tools constitute a software implementation of the knowledge that was acquired. The particular implementation is only valid for a specific set of tactics.

### Legislative Activities

The state legislatures and Congress must become more aware of the impact of digital technologies on the citizens, residents, and businesses of the United States. This will necessarily include education, briefings, and technical information from researchers and users of the Internet.

All computer and network security methods rely on cryptographic technologies in one form or another. Congress must remove impediments—such as the current classification of all cryptographic technologies as munitions—to domestic production of cryptographic methods. If the technologies cannot be exported, then U.S. companies are at a disadvantage in the world market.

Recognition of digital communications as "protected speech" as defined in the Constitution would significantly clear the currently muddled waters and greatly simplify the legislative and law-enforcement burden.

"Jurisdiction" is also a current problem. Consider the case of Kevin Mitnick: He was a fugitive from the Los Angeles area, allegedly intruded into computers in the San Francisco area, but was actually in Seattle and Raleigh.

### Law-Enforcement Landscape

The law-enforcement landscape is going to change. Along with new technologies for fighting computer crime will come an increased burden for investigation. Education of law-enforcement agents to include computer crimes and methods will help, but it seems inevitable that private computer security investigators will play an increasing role in the prevention, detection, and investigation of computer-related crimes.

### ADDITIONAL RESOURCES

Hafner, Katie, and John Markoff. 1991. *Cyberpunk*. Simon and Schuster, New York. Farmer, Daniel, and Eugene H. Spafford, "The COPS Security Checker Systems," *Proceedings of the Summer USENIX Conference*, pp. 165–170, June 1990. Stoll, Clifford. 1989. *The Cuckoo's Egg*. Doubleday, New York. Tzu, Sun. 1963. *Art of War*. Oxford University Press, Cambridge.

## 51

# Trends in Deployments of New Telecommunications Services by Local Exchange Carriers in Support of an Advanced National Information Infrastructure

*Stewart D. Personick*

*Bell Communications Research Inc.*

### STATEMENT OF THE CHALLENGE

The telecommunications industry, the computer industry, and other industries have been performing research and development for more than two decades directed toward the realization of "information age" applications of computer and communications technologies. In the last few years, this vision has been articulated in the contexts of the "information superhighway," the "national information infrastructure" (for the United States), and more broadly as the "global information infrastructure." While the definition of a national information infrastructure (NII) is subject to some differences in viewpoints, several consensus definitions have been published. For example, the definition published by the private sector Council on Competitiveness includes the following paragraph:<sup>1</sup>

The infrastructure of the 21st century will enable all Americans to access information and communicate with each other easily, reliably, securely, and cost effectively in any medium—voice, data, image, or video—anytime, anywhere. This capability will enhance the productivity of work and lead to dramatic improvements in social services, education, and entertainment.

Although definitions of the emerging NII may differ in detail, they appear to be quite similar in spirit and intent in their focus on the following features:

- Applications developed for all sectors of the economy;
- Affordability—enabling *all* members of society to derive an improved quality of life and a higher standard of living;
- Ubiquity (anytime, anywhere);
- Support for easy-to-use multimedia applications (voice, data, image, video), which provide intuitive user-friendly interfaces to people; and
- Dependability (reliable and secure).

A key component of such an NII is the underlying communications fabric, which allows users to connect with other users via their communicating/computing appliances (telephones, computers, personal digital assistants, fax machines, set-top boxes, etc.). In the United States, this underlying communications fabric is composed of the diverse networks of local exchange carriers, cable TV providers, wireless (cellular) providers, alternate access

providers, interexchange (long-distance) carriers, and value-added networks that are built on top of these (e.g., the Internet).

Traditional telecommunications networks have satisfied many of the requirements implied by the vision of the NII, and, indeed, form the communications fabric of today's information infrastructure. They are affordable, ubiquitous, easy to use, and dependable, and they have supported a wide and increasing range of applications including telephony, data communications (using modems), fax, access to the Internet, voice messaging, e-mail messaging, voice-response services, and access to variety of information services. In addition to the applications listed above, which are supported by ubiquitous dial-up telephone services subscribed to by 94 percent of households,<sup>2</sup> there is a variety of higher-speed and/or specialized telecommunications services provided to businesses and institutions for such things as high-speed data transport and video teleconferencing, and for interconnecting Internet routers (packet switching nodes).

The ongoing challenges in telecommunications networking today focus on the following:

1. Realizing affordable, higher-speed communications networking capabilities to support multimedia applications for residences and small businesses, starting with the widespread availability of integrated services digital network (ISDN) access. The challenge is driven by the convergence of the telecommunications, computing, information services, and broadcasting industries.
2. Realizing the ability to offer customized telecommunications services to residences and businesses (e.g., calling name delivery, personal telephone numbers, personalized call screening and routing) by using the emerging advanced intelligent network (AIN) capabilities of public telecommunications networks, and supporting customers' needs for mobility by using combinations of wireless access technologies and AIN functionality in core public network platforms.
3. Meeting the challenges of "information warfare" as U.S. telecommunications networks increasingly become the target of hackers, criminals, and terrorists seeking to exploit the increasing dependency of U.S. citizens and institutions on network-based applications.
4. Making increasingly complex and diverse telecommunications networks appear seamless and easy to use from the perspective of users and their applications.

Meeting these challenges in providing an advanced communications fabric for NII applications requires the investment of billions of dollars of research and development funds, and the investment of hundreds of billions of dollars in new network facilities on a nationwide basis over the next two decades. These investments include the installation of combinations of optical fiber, coaxial cable, wireless technologies, and network software throughout the United States. One cannot overestimate the challenges associated with making networks and network services reliable, secure, and easy to use, and doing so at costs that are compatible with the expectations and ability to pay of residential and small business consumers. The vast majority of these software investments are directed at meeting these challenges. Since the demand of residential and institutional consumers for the newer applications that are envisioned within the framework of the NII is highly uncertain, and by implication the demand and associated revenues for the telecommunications services that the advanced communications platform can support are uncertain, these investments involve high risk, except in situations where a combination of *existing* revenue streams and cost savings can justify the investments independent of the demand for speculative new services. The rapid depreciation of computer and communications technologies, in terms of rapidly improving performance/price ratios, makes these investments even more risky because investments made in advance of market demand may never be recovered in a competitive marketplace.

Further compounding the risk associated with the large investments required to put in place the telecommunications fabric of the NII is the uncertainty associated with the regulatory and legal framework within which network providers must operate. The regulatory and legal framework of the past is ill suited for an environment of large investments targeted toward highly uncertain market needs using rapidly depreciating technologies in a competitive marketplace. For example, the requirement of a network interface device erects an artificial barrier that prevents local exchange companies from providing complete services to their customers. The regulatory and legal framework of the future is still being defined in a slow-moving set of processes. These

processes are, to a large extent, setting the pace for the deployment of the next-generation telecommunications fabric needed to support the full range of envisioned NII applications.

In summary:

- The regional Bell operating companies (RBOCs) are fully committed to supporting the development of the NII and its infrastructure and services.
- The RBOCs are fully committed to providing ubiquitous, reliable, secure, and easy to use state-of-the-art telecommunications products and services, particularly mass market telecommunications, at the lowest cost consistent with providing the required quality of service and operations.
- The RBOCs are making major investments in research and development and in deployments in their networks that are consistent with customer demands for enhanced functionality and customizability of network services, greater bandwidth in communications channels, improved responsiveness to changing customer needs, and improved reliability, usability, and affordability of services.
- A major challenge is the resolution of regulatory and public policy issues consistent with local exchange service competition and the rapid deployment of new technology to meet market demand.

### BACKGROUND: EVOLUTION OF THE NETWORKS

Since the creation of the former Bell system and passage of the Communications Act of 1934, and until the past decade, the business of telecommunications has primarily been focused on cost reduction and improving the quality of telephone calls in terms of the ease and speed of setting up a call and the quality of communication after the call is set up. In the late 1940s and early 1950s, capabilities to allow direct dialing of long-distance calls by calling customers were created and deployed on a nationwide basis. In the 1960s and 1970s the emergence of modern electronic circuitry made possible the introduction of microwave and coaxial long-distance cable systems that provided higher-quality connections and achieved lower transmission costs. In this same period, and continuing throughout the 1980s and 1990s, the introduction of computing technology, both in stored program-controlled switching systems and in the automation of network operations to streamline and facilitate manual tasks, resulted in dramatic increases in efficiency and the ability to serve customer needs quickly. The introduction of fiber-optic systems, starting in 1979, further improved the quality of local and long distance connections and further reduced the costs of transmissions. The introduction of local digital switching systems in the 1980s dramatically reduced maintenance costs.

Figure 1 shows a trend in the total number of local exchange access lines per employee in a typical U.S. telephone company over the last two decades. In the past decade, this number has increased from 167 access lines per employee to 250 access lines per employee. This improvement in efficiency has been enabled by the ongoing investment in new network technologies such as fiber optics and software-controlled remote (unattended) electronic systems, and in software-based systems that are used to facilitate all aspects of the business, including negotiating with customers to take orders for new or changed telephone services, determining the availability of equipment that can be assigned to new customers, assigning installation staff to connect customers to equipment that has been reserved for their use, and determining the causes of service problems and arranging repairs. With state-of-the-art computerized systems, which involve tens of millions of lines of code, many of these functions can be substantially or completely automated.

In addition to the cost reductions that have been achieved by the continuous investment in advanced technologies (hardware- and software-based), there has traditionally been an emphasis on the use of subsidies to make basic residential services universally affordable. Business services and long-distance services have traditionally subsidized residential service. Services in locations that have lower associated costs (e.g., urban areas) subsidize services in locations that have higher associated costs (e.g., rural areas).

These subsidies have resulted in low prices for basic residential services. As mentioned, this has resulted in the ubiquitous availability of affordable telephone services. However, in many places in the United States, basic telephone service is priced substantially below cost. As the nation moves into an environment of

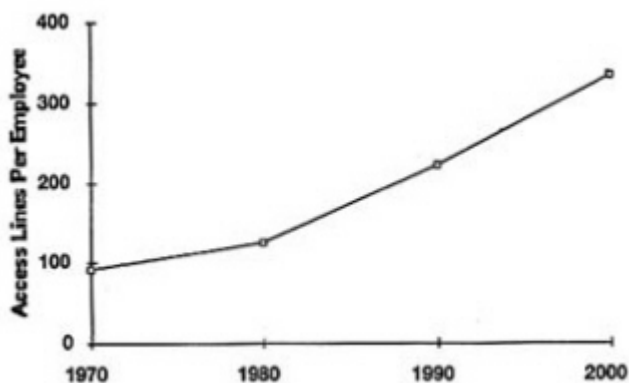


Figure 1 Trend in enterprise efficiency, as indicated by total number of local exchange access lines per employee in a typical U.S. telephone company, 1970 to 2000.

SOURCE: Courtesy of Bellcore.

competitive provision of new telecommunications networking services, the historical subsidies will slow down the widespread deployment of new services for the following reasons:

- Traditional providers will not offer new services below cost in an environment where competition makes it impossible to charge prices substantially above cost for other services in order to subsidize these below-cost services; and
- Consumers, who have become accustomed, through public policy, to below-cost, subsidized services and their associated low prices, will be reluctant to subscribe to advanced services that are substantially more expensive, even if they have the ability to pay for those advanced services.

In the last 10 years, the traditional focus on reducing costs and improving the quality of telephone service has been supplemented with a focus on providing new telecommunications services that are associated with the vision of the emerging information age, as captured recently by the vision of the NII. Three of the main thrusts that have emerged in the context of services directed toward meeting the needs of mass markets (residential and small business customers) are as follows:

- Personalized and customized telecommunications capabilities enabled by the advanced intelligent network (AIN);
- New digital two-way access capabilities, enabled by integrated services digital networks (ISDN) in the near term, and by broadband access in the mid- to longer term; and
- Mobility services based on wireless access and the AIN to support people on the move.

### Conversion of Networks from Analog to Digital Technologies

Over the last several decades, starting in 1962, digital transmission technologies have been introduced in public telecommunications networks. Initially, this took the form of T1 carrier service on relatively short-distance connections between telephone switching machines. Since 1979, it has taken the form of fiber-optic systems that link switching systems and that reach out directly to business customers, and to unattended remote terminals that serve residential customers. In addition, since the second half of the 1970s, analog switching systems have been upgraded to newer systems that employ digital switching technologies. The net result is the ability to provide end-to-end digital services to customers to support emerging multimedia applications.

### The Advanced Intelligent Network

The introduction of stored program-controlled (computerized) switching systems starting in the 1960s made it possible to go beyond direct dialing of calls to offer customized telecommunications services to meet users' needs.



The earliest customized services to be offered included customer-controlled call forwarding, three-way calling, and network-based speed dialing. Recently these have been supplemented by services that depend upon the calling party's number such as caller identification, automatic recalling of the last call received, and call blocking. These services depend upon the combination of low-cost memory (storage of information) and processing power that is enabled by state-of-the-art electronic technologies. However, these types of services have traditionally been implemented by making changes and additions to the large software programs that run the switching systems. Making changes to these large mainframe-like systems is very costly and time consuming. Furthermore, since switches are purchased from a multiplicity of suppliers and come in a multiplicity of types, implementing new services has traditionally required the development and deployment of new generic switching software by multiple suppliers for multiple switch types. This costly and time-consuming process is not consistent with the rapid deployment of a wide range of new telecommunications services that are customized to meet users' needs.

Thus, the local exchange carriers have implemented the AIN as a client-server approach to creating new services. In this approach, the switches act as clients that interface with software-based functionality in server nodes called service control points (SCPs), service nodes, and intelligent peripherals. The switches, service nodes, and intelligent peripherals implement building block capabilities that can be mixed and matched by the SCPs to create new services for network users. Since all switches implement comparable building block capabilities, new services can be created and deployed quickly by implementing new functionality in the server nodes. Following are three examples of new services that can be implemented in this way:

- A large chain of pizza stores requested a new service where its customers could call a single 7-digit number anywhere in a large geographical area and have their call directed to the nearest pizzeria. The 7-digit number is easy to remember and has the "feel" of a local number, which gives customers the confidence that their pizza will be delivered warm and fresh. Using the AIN, this service was implemented as follows. When a pizza customer calls the 7-digit number, the call is delivered to a specific telephone switch that recognizes this number as a special number. The switch sends an inquiry to an SCP, along with the calling number. The SCP accesses a geographical information system that determines the nearest pizzeria based on the 9-digit zip code of the caller and a map associating zip codes to pizzerias. The telephone number of this nearest pizzeria is returned by the SCP to the switch, which then forwards the call to the nearest pizzeria. This example can be generalized within the AIN capability framework to include area-wide abbreviated dialing for businesses with multiple locations and a wide variety of special called-number translation capabilities based on varying numbers of digits.
- Individuals have many telephone numbers associated with them. These include their home (residence) telephone number, their office telephone number, their fax telephone number, their car telephone number, etc. Some individuals would like to have a single telephone number that would make them accessible wherever they are. Using the AIN, one can implement what is sometimes generically referred to as "personal number calling." One way to implement personal number calling is to utilize a special "access code" such as 500 to signify a call that requires special handling. For example, a personal telephone number might be 500-CALL-ANN (500-225-5266). When such a number is called, the switch would recognize it as a personal number, temporarily suspend call processing, and send an inquiry to an SCP to determine the current physical telephone number (wireline or wireless) to which this customer's calls should be directed. Furthermore, depending on the preferences of the called party ("Ann"), the call might be routed to her or to her voice mail or to her secretary, depending on the time of day and the number of the calling party. All of this service logic can be implemented in the SCPs and their associated database and processing systems.
- A service that appears to have very high market demand is voice-activated dialing. This service allows users to record speaker-dependent (and eventually speaker-independent) templates of numbers they wish to dial. Subsequently, those numbers can be dialed by repeating the utterance that was recorded. For example, a family might have each child record the message "Call Mom," and arrange that this utterance be converted into one number or a sequence of numbers to be tried. With AIN, this service can be implemented for end users who pick up any ordinary phone in their home or office. When the child lifts the handset (i.e., goes "off hook"), the switching system automatically connects the line to a centralized AIN intelligent peripheral (IP) server, which stores the recorded speech templates and performs the speech recognition function. This IP returns the telephone number to be called to the switching system and can also work with other AIN functionality to prompt the switching system

through a sequence of numbers that may depend on the time of day, day of week, or other parameters chosen by the subscriber.

The key advantages of the AIN are that it can provide customized telecommunications services to end users in a manner that is easy to use and that works with a wide variety of appliances ranging from simple telephones to advanced personal digital assistants.

### **Advanced Digital Access**

Today's residential and small business telephone customers typically access the network using dial-up telephone services delivered over the copper wires that connect them to the network (except wireless customers as described below). Using modern electronics, it is possible to extend the range of uses of these wire-pair-based connections far beyond what was originally contemplated. For example, the latest modems allow for digital communication at 28.8 kbps on dial-up connections. In order to support multimedia applications such as access to stored images, video clips, compact-disk-quality audio clips, and two-way multimedia teleconferencing/collaborative work, customers need more than a modem and a dial-up line. One of the largest technical, economic, and regulatory challenges facing the telecommunications industry is how to create a path forward toward a ubiquitously available, affordable (by residential users), high-speed digital capability to support multimedia applications.

One of the steps in this direction is the widespread deployment and availability of integrated services digital network (ISDN) capabilities. In most cases, ISDN uses existing telephone wires to provide two-way, simultaneous digital connectivity at up to 128 kbps, plus a 16-kbps digital channel for network control signaling and additional packet data. Users of ISDN have reported greatly facilitated access to the Internet World Wide Web and other sources of multimedia information, and the ability to carry out multimedia teleconferencing/collaborative applications that include face-to-face video that is of reasonably high quality.

Because ISDN uses the existing wire pairs for most users (the distance from the user to the terminating point in the network is the key factor, with roughly 6 km being the limit using existing wire pairs), it can be provided with a relatively moderate (but still large) initial capital investment. The existing switches must be upgraded with hardware and software to support ISDN. Analog switches must be replaced with digital switches, or the ISDN customer must be reterminated on a digital switch. Note that approximately half of the telephone lines today terminate on stored program (computer)-controlled analog switches that still offer high-quality service for wire-pair telephone lines.

The software-based operations support systems that are used to automate operations in the network must be upgraded to accommodate ISDN. The individual customers who order ISDN must employ special terminations that are compatible with the ISDN interface, whether built into an ISDN telephone, a special interface board of a computer, or an ISDN terminal adaptor. The decision of a RBOC or a local exchange carrier (LEC) to make ISDN available throughout its territory (make it available to its roughly 10 million to 15 million subscribers) is a multibillion-dollar investment decision. The capital investment by a customer for an ISDN interface is a commitment of several hundred dollars at this time, but this cost will drop rapidly as ISDN usage rises over the next several years.

The ongoing charges for ISDN access vary throughout the country and are based on a combination of a flat monthly fee and a usage charge that may depend on minutes of use, packets sent, or a combination of these. As mentioned above, the challenge for the telecommunications provider is to recover investment costs in an environment of traditionally subsidized, below-cost pricing of basic residential telephone services.

Beyond ISDN, in the intermediate and longer term, is the challenge of providing residence and small businesses with digital access that is capable of supporting applications such as high-quality, full-motion video. Such applications require more than 1 Mbps and can range up to 20 Mbps or more for full-quality compressed high-definition television signals. To provide such services requires the deployment of a new physical infrastructure consisting of optical fibers, coaxial cable, and broadband wireless technologies. The investment cost is likely to be in the range of \$1,000 to \$2,000 per subscriber location served, amounting to several hundred billion dollars on a

national level. Since the applications and the user demand are speculative at this time, creating an economic, regulatory, and legal framework that will encourage investors to take the enormous risks associated with such a deployment is a national challenge. Network providers are reluctant to deploy higher-cost architectures and technologies 5 or more years in advance of speculative market needs for such things as very high bandwidth (more than a few hundred kilobits per second) upstream capabilities on a per-user basis. Nonetheless, some RBOCs have announced plans to deploy broadband networks throughout their service areas over the next 20 years, and all RBOCs are deploying broadband networks on a limited basis to test market demand and technology capabilities.

In a competitive marketplace, other firms can learn from mistakes made by the first entrant and can then enter the marketplace later with newer technologies and a better understanding of customer needs. The potential advantages gained by waiting may give rise to a "getting started" problem, as all potential investors wait for someone else to make the first move. Long-term, large-scale projects like the NII may not be allocated adequate capital. To offset this risk partially, network providers would like to begin recovering their capital investments as soon as these new networks are deployed. Existing revenue streams for traditional telephony and entertainment video services are less risky than unproven and speculative new services. By offering old and new services on a new shared platform, network providers can reduce their revenue risk and also benefit from economies of scope. The need to reduce risk and share the costs of network deployment across many users and services may be an important factor driving many telecommunications companies' interest in entertainment video and many CATV companies' interest in traditional telephony.

### **Services to Support People on the Move**

Since its introduction in 1984, cellular telephony has grown approximately 40 percent per year to serve nearly 20 million customers today. Paging services and cordless telephones are also highly popular. In the next several years, new kinds of personal communications services based on digital technology and supporting both traditional voice and computer/multimedia applications are expected to be widely available. It has been estimated that by the year 2003, there will be 167 million U.S. subscriptions to personal communications services, with many customers subscribing to multiple services.<sup>3</sup>

While wireless provides the physical access mechanism for an untethered telephone and other appliances, the advanced intelligent network (AIN) provides the software-based functionality to people on the move. Home and visitor location registers (AIN service control points) keep track of where nomadic users are and provide the information required to direct incoming calls to those users. AIN can screen or block incoming calls according to the calling number, time of day, day of week, or other parameters specified by the called party. AIN functionality allows "multitier" telephones to access cordless base stations, high-power vehicular cellular base stations, low-power pedestrian cellular base stations, and, in the next several years, low-Earth-orbiting (LEO) satellite systems, depending on which is most economical and available at any given time. As wireless telephony transitions toward nomadic multimedia computing and communications, the advanced intelligent network will provide access control (security-related) mechanisms, interworking functionality, screening, customized routing, media conversion, and other "middleware" functionality to support people on the move.

## **FORECAST**

### **Advanced Intelligent Network**

Based on demographics, it is probable that all RBOCs and most other local exchange carriers in the United States will deploy the advanced intelligent networking capabilities described above nearly ubiquitously over the next 5 to 7 years. Some carriers have already made these services widely available. These services will represent to telecommunications subscribers what the advent of the personal computer represented to its user community. Users will be able to define and obtain customized call processing capabilities to support both voice and data/multimedia applications such as customized screening and routing of calls, automated media conversion to facilitate the delivery

of messages, personalized telephone/network numbers, and access control (security-related) services. These will be provided in a way that is easy to use, reliable, affordable, and capable of interworking with a wide variety of appliances and terminals, ranging from simple telephones to personal digital assistants. AIN will enhance multimedia communications by enabling users to control multiple channels in a single communications session, and by interfacing with a variety of user terminal devices in a user-friendly way.

### **Integrated Services Digital Network**

ISDN is widely deployed and available today. A detailed deployment schedule for ISDN is shown in [Figure 2](#). ISDN is a major step forward in enabling two-way, interactive access to multimedia information, multimedia messaging, and multimedia teleconferencing and collaborative work. It will be the backbone of the transition of residential access toward broadband over the next 20 years. Along with today's dial-up modem-based access, ISDN will be a principal access technology for residential and small business users accessing the Internet over the next 20 years. ISDN, in both its basic rate (two 64-kbps "B" channels) and primary rate (twenty-three 64-kbps "B" channels) forms, will be used by businesses to meet their traditional telecommunications and Internet access needs, and it will be used by cellular and emerging personal communication service (PCS) providers to connect into the core telecommunications networks. ISDN will be a principal access mechanism for K-12 schools, libraries, and community centers to connect to the national information infrastructure.

### **Higher-Speed Switched and Nonswitched Services**

Until recently, the primary method by which businesses and institutions obtained nonswitched private line connections between their locations was to use dedicated 1.5-Mbps T1 lines, and dedicated 56-kbps digital private lines rented from telecommunications carriers, including the local exchange carriers. Some larger businesses and institutions have used higher-speed 45-Mbps private lines for point-to-point connections. Recently, new types of digital services, including frame relay, switched multimegabit data service (SMDS), and ATM cell-relay, have been introduced by telecommunications carriers, including local exchange carriers. SMDS is a packet-switched, connectionless data service that allows the destination to be specified independently for each packet. Frame relay and ATM are currently nonswitched services that utilize predetermined destinations for traffic; switched versions of these services are under development. All these services offer the advantages of improved sharing of facilities (fibers, terminations on electronic equipment, etc.) through statistical multiplexing. These new services, particularly ATM, can also support advanced multimedia applications that require high data rates and low delay variability between communicating endpoints.

These higher-speed services are being deployed in concert with market demands and are expected to be widely deployed and available over the next 5 to 7 years.

### **Wireless**

Cellular networks are widely deployed in urban and suburban population centers, and coverage and connectivity are steadily improving. These networks are being expanded to meet the growing user base with the deployment of smaller cells and newer technologies. Low-power cellular (personal communications services) to support people on the move is being implemented and will be widely deployed over the next 5 to 7 years. Wireless networks are being upgraded to employ digital technologies that support data and multimedia applications. In addition, these digital technologies enable the incorporation of encryption methods to improve the resistance of wireless services to eavesdropping. The use of advanced intelligent network functionality and services will allow for improved roaming and mobility for wireless users and will enable access to multiple wireless networking

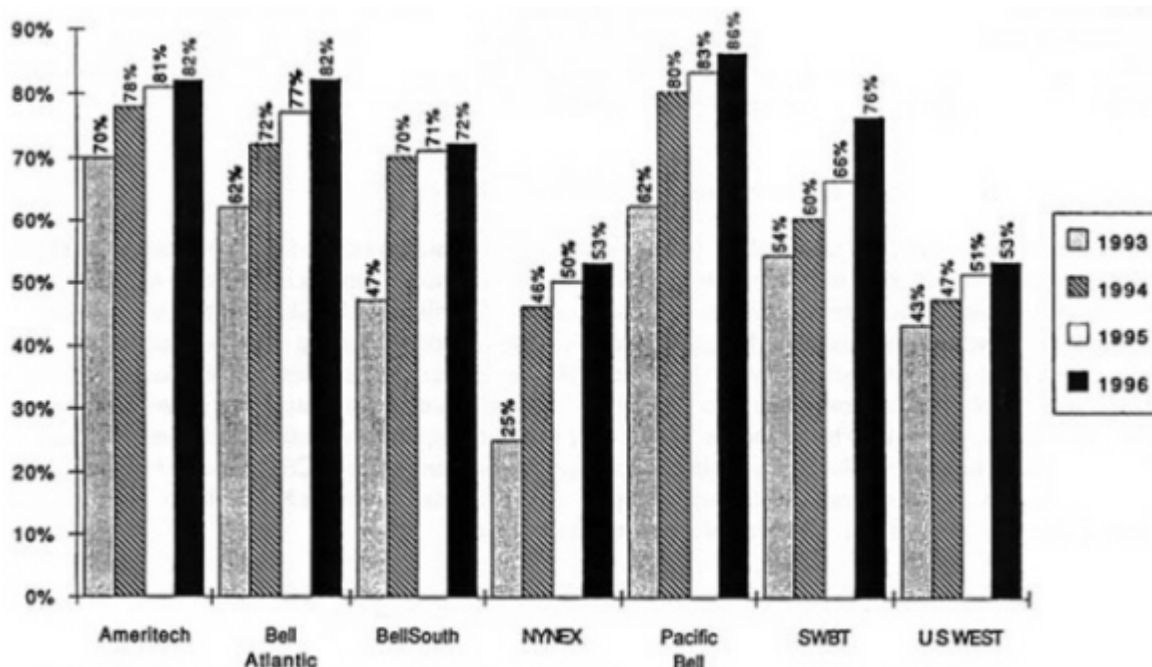


Figure 2 Percentage of lines with actual or planned ISDN availability through 1996. SOURCE: Courtesy of Bellcore, based on SR-2102, Issue 5, November 1994.

services (e.g., cordless telephony, high-power cellular, low-power cellular, and satellite-based services) from a single telephone handset. Such "multitier" applications are being deployed now by some local exchange carriers, and they are expected to be widely available over the next 5 to 7 years.

### Internet

The Internet, as it exists today, is built on services provided by local exchange carriers and interexchange (long-distance) carriers. Users access Internet routers (switches) through dial-up telephone lines and 56-kbps or 1.5-Mbps T1 private lines leased from telecommunications network carriers, primarily local exchange carriers. Routers are interconnected with 56-kbps, T1, and 45-Mbps private lines, typically leased from telecommunications carriers. Increasingly, fast packet services (such as frame relay and SMDS) are being used to replace point-to-point links.

Recently, several local exchange carriers have announced offerings of complete Internet Protocol (TCP/IP) offerings, including routing functionality, mail boxes, and support services. It is likely that most local exchange carriers will offer complete Internet service product lines in the next several years. However, there are regulatory issues that can delay the RBOCs' offerings of Internet services. The Modified Final Judgment (MFJ) prohibits the RBOCs from carrying traffic that crosses local access and transport area (LATA) boundaries; such traffic must be handed off to a long distance carrier selected by the consumer. It is not clear whether, and if so, how, the restriction applies to the provision of Internet service. In testimony before the House Subcommittee on Science, George Clapp, general manager of Ameritech Advanced Data Services, made the following statement:

Offering a ubiquitous Internet access service with the burden of the long distance restriction would increase our capital costs by 75 percent and expenses by 100 percent. The following factors contribute to these additional costs:

- LATAs in which there is low customer demand cannot be served from other sites in other LATAs.
- Customers of our switched data services frequently demand redundancy within our own network to assure service availability. Because of the long-distance restriction, we cannot use sites in other LATAs to provide redundancy.
- Current Internet routing technology requires us to dedicate a router to each long-distance provider in each LATA.<sup>4</sup>

At this point, some RBOCs are interpreting the MFJ restrictions to apply to their Internet service offerings. This is an area where regulations need to be changed to allow the RBOCs to compete on an equal basis with other carriers that are not subject to MFJ restrictions.

### **Broadband Access**

As described above, the provision of ubiquitous, affordable broadband access to residences is one of the most difficult challenges facing telecommunications carriers. All RBOCs have expressed a commitment to deploy broadband access services as quickly as the market demand, technology cost trends, and regulatory and legal environment permit.

The RBOCs have collectively invested approximately \$20 billion per year in upgrades to their networks since divestiture in 1984. They have committed to increase their investments substantially if regulatory reforms are enacted that enable them to be full-service providers of telephony, video, and multimedia interactive services in an environment that is conducive to the high-risk investments required to deploy broadband access networks. Several of the RBOCs and other local exchange carriers have market trials of broadband access under way or planned.

The deployment of symmetrical two-way capabilities, which permit residential users to originate individualized very high speed (greater than several hundred kilobits per second) upstream communications is a major challenge. One must differentiate between the concept of symmetrical two-way access, which has been raised as an issue by the government and other stakeholders, and the concept of two-way capability. The most demanding two-way capability that has been identified in the context of networks that serve residences is two-way multimedia collaborative work, also called multimedia teleconferencing. Research has shown that two-way multimedia collaborative work can be supported, to a large extent, by basic rate ISDN, and that nearly all needs can be met with a two-way capability of 256 to 384 kbps. Most broadband access network architectures being considered for deployment by the RBOCs can support this capability on a switched basis for all subscribers. At issue is whether there is demand for still higher speed two-way capabilities, comparable in speed to the one-way capability needed to deliver entertainment-quality video to residential customers. The data rate associated with entertainment video ranges from 1 Mbps for VHS quality to 20 Mbps or more for HDTV quality video. The ability to deliver entertainment-quality video both downstream to residential users as well as upstream from residential users is what is called symmetrical two-way access.

Although a large number of alternative architectures have been extensively studied from a capability and cost perspective, it appears that in many situations substantial incremental investments are required to provide symmetrical two-way capabilities. It is unlikely that these incremental investment costs will be recovered in a competitive marketplace if they are made many years ahead of the demand for such high-speed upstream services. The details of the trade-offs among alternative broadband architectures vary from RBOC to RBOC depending on such things as the density of housing units.

### **Dependable, Usable Networks**

The tradition of the telecommunications industry has been to provide network services that are highly reliable, secure, and usable by the widest possible range of telecommunications services customers. As new, interactive, multimedia networking services and applications are deployed, using a wide range of new and

heterogeneous technologies, it will be a great challenge for all industry participants to maintain this tradition in the context of the NII. If individuals, corporations, and institutions are to reengineer themselves to become dependent on networked applications, then those individuals, corporations, and institutions must be provided with network-based services and applications that are even more dependable than today's telephony services. They will expect those services and applications to be easy to use, to work all of the time, and to be secure from intrusions and other security threats. The RBOCs are committed to maintaining their tradition of reliable, secure, and easy-to-use services through a combination of technological and operational methods. In particular, the use and sharing (in public forums such as the National Security Telecommunications Advisory Committee) of best practices among network providers are essential to help prevent and minimize such threats. Cooperative testing between networks to detect incompatibilities, particularly of management protocols that protect faults from propagating into large outages, is an essential ingredient of this process.

As we move into the future, the role of telecommunications networks in facilitating interoperability and ease of use will become increasingly important to consumers. While early adopters and those who create new technologies have a relatively high tolerance for complexity and unreliability and are willing and able to invest substantial amounts of time in learning to use applications and in resolving problems, mass market users expect their applications and services to be extremely dependable and intuitive. In theory, software-based functionality can be placed in end users' terminals to enable interoperability, to resolve incompatibilities that would be perceived by customers as application failures, and to make complexity transparent to end users. In reality, this is achieved today by forcing end users to be systems administrators of their complex terminal software, or to engage others to administer their systems for them. Traditionally, the telephone networks have hidden complexity from end users and have resolved incompatibilities among end user terminals by employing "middleware" in the networks. For example, an end user in New York can make a call from an ISDN telephone to an analog cellular phone in London. As applications such as multimedia teleconferencing, multimedia messaging, and remote access to multimedia information become increasingly important in mass market applications, telecommunications networks will play a critical role in resolving incompatibilities between different types of user terminals and between user terminals and servers, in facilitating the location of resources, in helping users manage their communications services, and in providing capabilities such as multimedia bridging.

### BROAD RECOMMENDATIONS

Most of the technology-related challenges in creating the national information infrastructure can be best addressed by the private sector, with the cooperation of the public sector.

- From the point of view of local exchange network providers, the regulatory impediments discussed above must be addressed to enable an investment climate that is appropriate for the high risks associated with the large deployments of network infrastructure needed to provide broadband access. The public sector should work with the private sector to address issues related to universal access and service and to promote open systems and interoperability among networks, systems, and services. The public sector should remove barriers to full and fair competition, such as the MFJ restrictions and the network interface device requirement discussed above, and should avoid creating new barriers in the future.
- The public sector should use networks provided by the private sector rather than building networks in competition with the private sector's commercial service providers.
- The public sector must also address the myriad legal issues related to intellectual property, liability, application-specific law (e.g., practicing medicine across state lines), and other issues that are impediments to the emergence of new applications. Many of these issues have been identified in forums such as the National Information Infrastructure Advisory Council and the private sector Council on Competitiveness.
- The public sector should be a role model user of the emerging NII and should continue its initiatives to create a broad awareness of the potential of the NII to address many of society's challenges in education, health care, and government.

- The public sector should support and fund precompetitive research and development targeted toward enabling NII dependability, interoperability, and ease of use by the broad population. The public sector should also support and fund precompetitive research and development on advanced technology for next-generation networks and advanced applications.
- The public sector should collaborate with the private sector on programs that will lead to the realization of the goal of having K-12 schools and libraries connected to the NII by the year 2000.
- The public sector should work with the private sector to protect U.S. interests in matters related to the global information infrastructure, with particular emphasis on intellectual property protection and trade reciprocity.

#### NOTES

1. Council on Competitiveness. 1993. *Vision for a 21st Century Information Infrastructure*. Council on Competitiveness, Washington, D.C., May.
2. Federal Communications Commission. 1994. *Statistics of Communications Common Carriers, 1993/1994 Edition*. Federal Communications Commission, Washington, D.C., Table 8.1.
3. Personal Communications Industry Association. 1994. *1994 PCS Market Demand Forecast*. Personal Communications Industry Association, Washington, D.C., January.
4. Clapp, George H. 1994. *Internet Access*. Testimony before House Subcommittee on Science, October 4.



## 52

# The Future NII/GII: Views of Interexchange Carriers

*Robert S. Powers, MCI Telecommunications Inc.*

*Tim Clifford, SPRINT, Government Systems Division*

*James M. Smith, Competitive Telecommunications Association*

### SUMMARY

We are not starting a new national information infrastructure/global information infrastructure (NII/GII) from a blank page; we are building on an information infrastructure, corporate structures, regulatory practices, billing practices, services, and public expectations that already exist. So the challenge is not so much "what is the ideal world that we would write down on a blank page?" but rather "how do we get there, starting with today's realities?" This paper presents views of the interexchange carrier community, as to the future NII/GII potentials, the stones we must step on to cross that tricky creek from "now" to "then," and the stones we could trip on and end up all wet. Our principal emphasis is on how to achieve fair and effective competition, utilizing the regulatory process as necessary to help us get from "now" to "then," and assuring fair access to networks and services, for all Americans—individuals, businesses, governments, educational institutions, and other entities.

### INTRODUCTORY ASSUMPTIONS

This paper begins with some postulates upon which we think we can all agree and then discusses where those postulates lead us, and how we can actually accomplish the things that our assumptions imply.

*Postulate 1:* There already exists a "national/global information infrastructure." It has existed, in some degree, since telegraphy became widely implemented. But now we're facing enormous enhancements: The focus in the past has been largely on the "communications" aspects of an NII/GII—getting voice or data (information) from one place to another. The future NII/GII will broaden to include vast improvements in creation, storage, searching, and access to information, independent of its location. To achieve these improvements will require sophisticated customer-premises equipment and skilled users. The government may find it appropriate to assist in developing user skills and providing customers' equipment; but the associated costs must not be imposed on the telecommunications industry.

*Postulate 2:* It is economically feasible, in many instances, for there to be competition in the form of competing local telecommunications facilities, as well as services. But there will surely be some locations in which there is only one local facilities-provider. Furthermore, even if there are multiple local facilities, there may be only one *active* facility connection to a given home or office. In that case, will the multiple competing providers of services of all kinds have open and fairly priced access to that same end-link, so that any service provider can have fair access to customers? Our postulate: Such bidirectional open access can be achieved by facilities providers' unbundling their services and providing them for resale on a fair basis; and effective competition can bring about those results. But it will be a gradual transition process from regulation of

monopolies to effective market competition, varying in speed of accomplishment in different locations. The associated gradual relaxation of regulation must be done carefully, in such a way that monopolies are freed from regulation in proportion to the reality of effective competition.

*Postulate 3:* A fundamental difference between the information superhighway and certain other widely used facilities such as the interstate highway system is in the mechanism for paying for their use. If you pay for something out of the public purse, and give it away for free or for a flat-rate charge, you create the potential for enormous waste of that resource if excessive use of that resource is of some benefit to the users. In the case of asphalt highways, that turns out not to be much of a problem: a person can "waste" the resource only by spending time driving on the roads. But in the case of the NII, one could send megabyte files to thousands of folks who don't want them, just by pressing an <ENTER> key, if the service were free. So Postulate 3 says that although there will be specific instances of flat-rate billing for services, usage-based billing will be widely used, to limit wasteful use of the resources and to correlate costs with benefits.

*Postulate 4:* As often stated by federal representatives, the NII/GII will be built by private-sector investments, not by governments. That clearly does not exclude governments from building facilities for their own use when specific needs cannot be met by private-sector providers, or from exercising their regulatory power to assure fair competition and to achieve other public interest goals, or from having their purchasing power influence the direction and speed of development and implementation. The purchasing power of the top 10 to 15 federal agencies is comparable to that of the Fortune 500 companies, so that power can surely influence the marketplace. But governments won't build the basic infrastructure.

*Postulate 5:* There are, however, public-interest goals that governments must participate in achieving. Aside from relaxing current laws and regulations that prevent competition in networks and services, we postulate that a major governmental role will be in assuring implementation of what we prefer to think of as the "successor to universal service." The successor to universal service could take either of two quite different forms, or some combination of the two. It could mean simply equitable access, meaning that the NII should be sufficiently ubiquitous that anyone can have access to the network at equitable rates and at some accessible location. Or it could be an advanced universal service, which would enable qualified persons or entities to use selected *services* available via the NII. In either case, any subsidy should be targeted to the needs of end users, not to specific providers. All providers must be able to compete for the end user with equal access to the subsidy funds or vouchers.

In whatever combination evolves, the mechanisms for funding and distributing the subsidy pool will be different from today's mechanisms, with multiple network and service providers being involved instead of today's situation with monopoly providers of networks and services. The mechanisms for creating any subsidy pools that may be required must be fair and equitable to all of the contributors to those pools. Contributors could be network and service providers, but could also be governments, using income from some form of taxes. Clearly, some level of regulation will be required during the transition from today's "universal service" to tomorrow's version of that concept.

*Postulate 6:* The anticipated vast increase in use of the GII for personal and business purposes offers potential for enormous compromises of security (of both personal and business confidential information), personal privacy, and protection of intellectual property, unless preventive measures are implemented. And since people don't want their privacy invaded and businesses simply cannot afford to have their proprietary information exposed to others, people and businesses will adopt encryption and/or other mechanisms to prevent such intrusions and exposures. Government and law enforcement agencies must recognize that reality and must not unduly restrict the use and worldwide trade of encryption technologies.

### THE VISION: WHAT IS THE NII/GII?

We begin with a high-level description of what we are expecting will evolve and then discuss what must be done to get there. One of the most concise definitions of an NII/GII is found in the 1994 report *Putting the Information Infrastructure to Work*, from the National Institute of Standards and Technology, "the facilities and services that enable efficient creation and diffusion of useful information."

The concept of NII/GII obviously encompasses a "network of networks," including telephone networks, cable TV nets; the Internet; satellites; and wireless-access networks, domestic and foreign. We also suggest that the broad concept of NII/GII includes not just these basic networks, but also the end equipment, the information resources, and the human beings that interact with the equipment, networks, services, and resources.

The GII should and, we believe, will provide at least the following high-level functions:

- *Person-to-person "virtual proximity"*—that is, the ability of individuals to communicate with each other independent of location.
- *Access to both one-way and two-way communication channels*, narrowband or broadband as required, on demand.
- *Multiple mechanisms for such access, depending on location*. In high-density areas there will be nonradiating channels (glass or metallic) as well as wireless access to allow for personal mobility. Lower-density locations will gravitate more to wireless access, since wireless will be less costly than wires or fibers in those low-density areas. And the areas of very low density will have access to both low- and high-capacity services by means of satellites. (That's not to imply that satellites will be used only in low-density areas.)
- *New mechanisms for protecting the privacy of individuals and the security of financial and other transactions*. These mechanisms will surely include encryption, and also new mechanisms for transmitting or obtaining information anonymously.
- *Highly reliable physical and logical networks*. Network providers are already implementing circular-path networks, rapid restoration capabilities, and even arrangements to use networks of their competitors for enhanced reliability. The marketplace demands such actions, when the entire operation of more and more business depends on the availability of real-time communication among themselves, with suppliers, and with customers.
- *Access to vast amounts of information, with the ability to search those treasure troves for specific information*, enabling new services such as automotive navigation, location of the nearest gas station, and many others. That information access, and associated services, will further support fundamental societal changes such as telecommuting, telemedicine, enhanced education opportunities, and others. Indeed, the success of a GII might well be measured by the proliferation of new services and their effects on society. Today's "carriers" have opportunities to provide easy access to such services, if not the services themselves.
- *Global interoperability*. It is not yet true, and probably never will be, that precisely the same equipment, radio spectrum, and operational protocols are used in every location around the globe. But it is true—and the pressures to make it true are constantly increasing—that gateway mechanisms to achieve transparency from the user's point of view are being and will be implemented, both domestically and globally. We do note, however, that technical feasibility is not the only barrier to be overcome: business arrangements among competitors, to achieve compatibility, are not always easy to achieve. But such arrangements are in the long-term interests of those competitors, and we believe they are achievable.
- *Billing mechanisms that reflect the cost of providing both the network services and the information services to which the users have access*. Especially when the cost of usage-based billing is for some reason significant when compared with the cost of the actual service, flat-rate billing mechanisms or even free services may be provided. Asphalt highways provide an example here: the cost—and the nuisance factor—of billing for every entry and exit and every ton-mile of use of roads is so great that we've found other ways to pay for the facilities, in such a way that does not invite excessive waste of those facilities.
- *New mechanisms for protection of intellectual property rights, and new assumptions as to what those rights actually are*. This is one of the most difficult characteristics to achieve, since once information is transferred electronically, it is virtually impossible to prevent the recipient from passing it along to others. It's cheaper to buy a second copy of a conventional book than to produce a hard copy of it, so the author gets paid twice, by the two purchasers. But if each of us can send the electronic book to dozens of friends just by pushing an <ENTER> key, then how can an author be assured of being compensated in proportion to the distribution of his or her writing? And even if we postulate such a far-out suggestion as to say that the network could or should detect the transmission of copyrighted material, how could the network do that if the transmission were encrypted? We suspect that (1) hard-copy books will continue to be in demand, for the foreseeable future, but that (2) the

basic concept of protection of intellectual property, when that property is available electronically, will evolve in an unpredictable fashion over time, and that (3) protection mechanisms will more and more be confined to commercial use of intellectual property, rather than simply forbidding the copying of such information.

## THE ROLE OF INTEREXCHANGE CARRIERS

### Today's Role

Today, the basic role of interexchange carriers (IXCs) is carriage of narrowband and wideband transmissions over "large" distances, where the difference between "large" and "small" distance is defined by regulators and lawmakers, and the term "large distances" certainly includes global delivery.

In addition to simply providing those services and dramatically lowering their costs to users as a result of competition, IXCs have been instrumental in the development and implementation of new telecommunication technologies and services. These include implementation of digital transmission techniques, single-mode optical fiber, synchronous digital techniques, and "intelligent network" services such as "800" service, which not only provides for the called party to pay, but also allows for such flexible services as time-of-day routing and routing on the basis of the location of the caller. The Internet backbone is, of course, an example of a major IXC role. This is not simply to brag about past achievements, but to point out that IXCs have competed in the past and will continue to compete in the future with each other and with other telecommunications entities. In that process, they will push development and implementation of new technologies and services as hard and as fast as they can.

### Tomorrow's Roles

In tomorrow's GII, the terms "long-distance carrier" (LEC) and "interexchange carrier" will become less used. Yes, there will be companies that concentrate more on building and selling local services, and companies that concentrate more on transcontinental and global networks. And both of those types of companies will provide at least some of their services using network capacities leased from the other types of companies. But the customer/user will hardly care about the details of what network is used; the user will care about the nature, availability, reliability, and price of the services provided.

As the conventional boundaries between local and long distance companies eventually disappear in the eyes of consumers, it will become the strategy of many retail providers to offer end-to-end services that ignore this traditional distinction. The critical element is the introduction of "wholesale" local exchange services that long distance companies can resell as part of their end-to-end packages. Wholesale long distance products are already available for local telephone companies to resell in combination with their local services to provide end-to-end service (to the extent that such resale is not currently restricted by the Modified Final Judgement). With a limited number of local networks expected to develop, it is especially important that a comparable wholesale local service be introduced to foster competitive diversity.

But that's the long view. How do we get there from here? How do today's IXCs play in that game? Here are our basic views:

- There is room for a large number of specialized IXCs in this country obviously—hundreds exist now. Only a limited number do now and will in the future provide nationwide facilities-based networks and services; a vastly larger number can and will provide various services using those facilities.
- Current IXCs will dramatically increase their emphasis on end-to-end services, intelligent-network services, information services, transaction processing services, and global coverage. They will also take on the role as "hosts" to various information services that they do not actually create, but do deliver to end users.
- As IXCs increase their emphasis on global end-to-end services, they will require more and more wideband capabilities in the "last-mile" links, to match the high capacities they already have in their long-distance networks but that are now generally missing in those last-mile links. The demand for data and video services—

often requiring higher bit-rates than voice—is growing rapidly and is expected to continue to do so. It is critical that the most efficient providers of these services be the ones that survive in the marketplace. To achieve this goal, it is important to prevent cross-subsidy of these services by LEC's still-monopoly local voice services.

- As the emphasis changes from simply distance to access to information, today's IXC's will surely move to providing more and more information-related services, rather than just transport. That does not imply that present IXC's will put libraries and catalog marketers out of business. But IXC's are in an ideal position to assist users in gaining access to information services independent of location of those services. Already, we see competing provision of nationwide telephone-number information services, for example. Why would it be a surprise to see today's IXC's beginning to offer information-based services as helping people find information that they want in government databases? Or helping medical practitioners to locate other medical practitioners that may know something they suddenly need to know about an obscure disease? Or assisting schools and universities in finding just the right other school or classroom in which the same problems, issues, or history lessons are being addressed?
- And as more and more business is conducted by means of electronic cash of various forms—credit cards, debit cards, "smart cards" with built-in computing power—who would be surprised to see today's IXC's providing secure, private, reliable, and real-time transaction information? For example, an IXC could provide the validation of the gas station customer's smart card; assure the customer that the gas station's card acceptance device is telling the truth about how much money is getting withdrawn from the customer's bank; and transmit the resulting information from the gas station to the gas station's bank, the customer's bank, and any intermediate parties that have any fiduciary responsibility. The carrier might even offer some functionality in terms of real-time execution of monetary exchange rates, even though the banks have the final fiduciary responsibility.

### ARCHITECTURE

A discussion of architecture is of course closely linked to the above discussion of the future role of today's IXC's. They will clearly have a key role in developing and employing tomorrow's network architecture; however, it may differ from today's.

#### Network Intelligence

The requirements of such services as "find me anywhere, anytime," sophisticated and flexible billing services, call routing dependent on time or day, store and forward, messaging, and others that we have yet to imagine will clearly require more and more sophisticated network intelligence and functionality. IXC's have already made great progress in these areas in the last decade or so; but the demands of video/multimedia services and increasing demands for flexibility and value-added features will require considerable expansion of "intelligent network" functionality. There is no doubt that such functionality will be implemented. And when it is implemented in any bottleneck portion of the network, all providers must have access to that monopoly functionality.

#### Long-Distance Transport

The architecture for tomorrow's long distance transport will not be all that different from today's, from a functional point of view. Clearly it will have even more capacity than today's already high-capacity two-way high-quality digital networks. Network protection and restoration capabilities will be enhanced. Newer multiplexing technologies such as SONET, and packet technologies such as frame relay and ATM, will be employed to provide ever faster and easier ways to insert and pick off individual transmissions from high-bit-rate bulk transmissions and to increase reliability and lower costs for certain services. And mechanisms to protect the privacy and security of users' information will surely be widely incorporated into the networks.

Some have asked whether the interexchange networks will have the transmission capacity that will be needed, if data traffic continues to grow at an exponential rate. Data traffic could, it is said, be equal to the volume of today's voice traffic in a few years, requiring a doubling of total network capacity. It is our view that capacity simply is not a problem in the long-distance networks, although it clearly is a problem in the "last mile." Long-distance networks are already almost entirely optical fiber. The capacity of a given pair of those fibers has been doubling roughly every three years since 1983, based only on changes in electronic and photonic equipment attached to the fiber, not changes in the buried fiber itself. The bit-rates that can be shipped through a given fiber, using a given optical wavelength, have increased as follows, with an even bigger leap expected in 1996:

<i>Year</i>	Rate (megabits per second)
1983	405
1986	810
1988	1,800
1993	2,400
1996	9,600 (expected)
2000	40,000 (expected)

Further, wave division multiplexing (WDM) is becoming practical, so that a given fiber can carry two or eventually even four or more signals, each transporting the 9,600 (or more) megabits per second. So as early as 1996 or 1997, using four-window WDM, it could be feasible for a given fiber pair to carry 38,400 megabits per second, compared to the 405 megabits per second available as recently as 1983. In terms of the conventional 64-kilobit-per-second circuits, the ratios are not quite the same because of overheads and other technical factors, but the result is just as impressive: a fiber pair in 1996 or 1997 could be carrying 616,096 circuits, compared with 6,048 circuits that that same glass could carry in 1983—an increase by a factor of almost 102! Capacity is simply not a problem, in the long-distance network, for the foreseeable future.

#### **Last-mile Connections**

Clearly, today's IXC's are moving toward providing end-to-end services. The reasons for this include the desire of many customers—large and small—to deal with a single service provider, and the need for uniform protocols and services, end to end. The move to end-to-end services is also driven by the interest of both IXC's and their customers in cutting the costs of the last-mile links, which are now such a large portion of the costs of providing long-distance telecommunications and are priced significantly higher than actual costs. IXC's pay access charges to LEC's, on a per-minute-of-use basis, for calls delivered or originated by the LEC's. Those access charges amount to approximately 45 percent of the gross revenues of IXC's. Various estimates made by LEC's indicate that the actual cost of local access is about a penny a minute or less. But the access charge for interexchange carriers is three (or more) cents per minute at each end of the link. Competition will help eliminate these excess non-cost-based access charges now imposed on IXC's and, finally, on end users.

There are two potential mechanisms for cutting those last-mile costs—and they are not mutually exclusive. One is for regulators to require that the access charges be cost based. The second is the introduction of effective competition in the local networks, which cannot develop if the new entrants are forced to pay a non-cost-based access charge to keep the existing monopolist whole. It should also be noted that there is potentially a structural barrier that could dilute the effectiveness of access competition: Obviously, the "access business" of the interexchange carrier must go to whichever network the customer has selected to provide its "last-mile" connection. An IXC's leverage to bargain for lower access prices would be limited by the IXC's ability to influence the subscriber's choice of local network provider. Therefore, until and unless either regulation or effective local competition causes the local network providers to offer open interconnection to all interexchange service providers, on efficient and competitively priced local networks, there could still be limitations to the ability of every service provider to reach every end user, with fair access costs. Because we expect that these local network providers will also be offering their own line of retail local and long-distance services, it is not clear that these network owners will have a great deal of interest in reducing their rivals' costs. Until such time as true

competition removes such barriers, regulation will be needed to assure that access to the end user by unaffiliated retail providers is fully available and priced fairly. One possible regulatory approach is to require that new entrants provide equal access to last-mile links, with rates capped at the rates of the incumbent provider.

We further note that simply the introduction of competition in local networks will not be fully effective in bringing down the last-mile costs if existing LECs maintain a dominant position in service provision and are allowed to continue charging non-cost-based rates for carrying services that originate or terminate on competitors' networks. And even if the new entrants build networks that are significantly more efficient and less costly than existing LECs' networks, those reduced costs cannot be fully passed through to end users if the LEC networks remain inefficient and those inefficiency costs are passed through to those new-entrant networks when the dominant LECs provide the last-mile carriage. Existing LECs will have minimal motivation to become more efficient and lower their actual carriage costs as long as they hold the dominant market share so that they can charge their smaller competitors the LECs' full last-mile costs.

Some argue that the access charge income, over and above actual cost, is used to support "universal service." However, studies indicate that the excess access charges—over and above actual cost—are far greater than what is needed to support universal service. A study by Monson and Rohlf<sup>1</sup> concludes that IXCs and other ratepayers are charged about \$20 billion per year over and above the actual costs of providing the access services needed. And a study by Hatfield Associates<sup>2</sup> estimates that the actual subsidy that LECs need to support universal service is \$3.9 billion. The excess income goes to support LEC network inefficiencies, domestic and foreign investments by LECs, and other LEC activities in which neither IXCs nor the public at large have interests.

Our bottom line suggestion here is that competition in the local communications market—which will include not only today's LECs and IXCs but also today's cable TV companies and perhaps local utility companies that have access to rights-of-way—could go even further toward providing low-cost access than does the current subsidy by means of access fees. But to achieve that goal will require some mechanisms for motivating existing dominant LECs to improve their network efficiencies and lower their costs and therefore their access charges.

### **Numbering Plans**

Numbering plans will clearly change dramatically in the future, as individuals demand multiple numbers for their own individual use and as the population grows. Already, the use of 10-digit numbers is required for more and more local calls, as both overlays and geographic splits are implemented because of exhaustion of numbers in a given area code. The demand for nongeographic "find-me-anywhere" numbers, and for separate numbers for home phones, wireless phones, office phones, fax machines, and other special services, will surely make our grand kids gawk when we tell them we used to be able to make phone calls with only 7 digits! (Some of us remember using only 4 or 5 digits, but never mind....)

## **NEW SERVICES**

### **Enhanced "Phone Calls"**

#### **Personal Numbers**

There's debate about whether everyone will want a "personal number for life," so that the same number will ring her or his phone as long as she or he is around. For example, there's some benefit in a West Coaster knowing that 703 is in Virginia, and it might not be polite to ring that number at 10:00 PM Pacific Time. A geo-independent number would not give the caller such a hint. But clearly there's a market for such geographically independent and permanent numbers. The IXCs, responding to their competitive market environment, have put in place intelligent network capabilities nationwide and therefore are likely to be in the forefront of making such a service practical.

### **"Find-Me-Anywhere" Service**

"Find-me-anywhere" service may or may not be tied to the "number-for-life" idea. The number for life may or may not be set up to follow the called party wherever s/he tells it to, on a real-time basis. But clearly such a tie could be executed and would be popular. For the same reasons as above, today's IXC's are in a fine position to pioneer such service, as they are now doing.

### **Messaging Services**

We already have messaging services, in the form of remotely reachable answering machines as well as services that can be provided by both local and long-distance carriers. As more and more people use the personal number and find-me-anywhere services, there will be more and more need to cut off such access when people go to bed at night, say, six time zones away from where their callers expect them to be. So we expect messaging services to grow rapidly in popularity, for this as well as other obvious reasons of convenience and reachability. There will be services offering messages of limited length (short-message services) as well as the ability to leave longer messages.

### **Multimedia Services**

Aside from whatever multimedia services today's IXC's themselves provide, it is likely that one of their future roles will be in providing transparent interfaces between other service suppliers and end users, when the users and the service suppliers use different equipment or protocols for storage or transport. Also, IXC's will provide billing and security services associated with multimedia services for which they themselves are not the originating providers.

### **Digital Services**

Do users care whether the connection network they are using is analog or digital, voice or "data"? No, they just want to talk, or have their screen filled with pictures or thought-provoking words or good graphic diagrams. It is the network providers and spectrum managers and equipment builders and capacity-theorists like Claude Shannon that notice and care about the differences. But in practice, there are clearly lots of differences, in terms of efficient use of network and spectrum capacities, flexible services, accuracy of data to be passed from one place to another, and ability to protect privacy. And since the entire IXC network is going digital anyway, for technical and cost and quality reasons, whether the transmission is voice or data, the result has huge benefits in terms of cost, speed, and reliability of the data transmissions that will be so important in tomorrow's global economy and lifestyle.

In this case, one could argue that there is no special role for IXC's compared to local providers, since we are all going digital in the end and will therefore eventually be able to provide the needed speed and reliability. But from the practical point of view there is indeed a major role for today's IXC's, and that is back to the point of providing competition and therefore much faster progress in building the digital high-bit-rate, last-mile networks that will be so important to customers who need the features that can be provided with broadband two-way digital networks.

### **Financial and Other Transaction Processing**

The United States led the world in implementation of credit cards and debit cards, to replace paper cash and checks in many financial transactions. But now, Europe is clearly ahead of the United States in its implementation of smart-card technology and services, to take advantage of the enormous capability and



flexibility of cards with computer chips built in. We are not at all out of the running—we do know about computer chips, and there is growing interest here. But at the moment European entities are in the lead in the implementation of smart-card technologies. IXC's will surely not be the only players in this game; but they do have an enormous opportunity to incorporate smart-card technology into many of their service offerings, and thereby bring the United States back into a leadership role in this new technology and service platform.

### **New Billing Mechanisms**

Billing, as has been hinted at above, will be an interesting issue in the future GII. We are rapidly learning—with the help of smart cards and other technologies—how to reliably establish that someone who wants to use a service is indeed the person she or he claims to be and will pay the bill when it comes. But we do have the fascinating challenge of learning how to bill proportionately to either the use or the value of "bursty" digital bit-streams. If a user is hooked up to the network for an hour and hits the <ENTER> key only once, does she or he get billed for 1 hour? Or for sending the 100 or the 100,000 packets that resulted from the touch of the <ENTER> key? Obviously, the provider does not bill separately for each packet, the way it writes a separate line on today's bill for each phone call. Is there a way to make some kind of average or statistical count of the number of packets or bits that flow to or from a given user and bill on that basis? Of course there is. The question is how it will be done to be understandable and fair to both the user and the provider, and how it can be confirmed by either party. Again, this will not be the role of long-distance carriers alone but will evolve on the basis of trials by all the carriers competing in both local and global services.

### **THE SUCCESSOR TO "UNIVERSAL SERVICE": ACCESS TO THE NETWORK**

As suggested in Postulate 5, above, the current concept of "universal service," in the sense of having essentially everyone be reachable by phone, will surely evolve as technologies and new markets and services evolve. There is no public interest in everyone's having access to every single "service" that the NII/GII will provide, any more than there is a public interest in everyone's driving a Rolls Royce. Clearly it is in the public interest for everyone to have at least some access to the GII, just as it is in the public interest for all of us to have access to streets and roads and telephones. But with the dramatic expansion of the types of services that will arise on the GII, we must learn to be more specific about who gets subsidized for use of what services, and how that cross-subsidy can be managed in a competitive structure. This seems clearly to be a case where all the service providers, as well as consumers and the public at large, have a deep interest in designing and executing mechanisms to provide "equitable access" and/or "advanced universal service" in a way that gives all of us the benefits of achieving those societal goals, as they may eventually be defined. Today's IXC's are as anxious as anyone to have the networks and services provided in such a way that students, workers, and all others have access to the GII, although there will clearly be services for which a subsidy to users is totally inappropriate. But we do insist that the contributions to any cross-subsidies that are required must be managed in such a way that the purchasers of any given service or network access are not unfairly required to subsidize some other unrelated service or access. A cross-service subsidy system would not only unfairly tax the users of a particular service; it could even prevent the provision of some desirable service or prevent the survival of a potentially valuable service or network provider.

### **FRAUD PREVENTION**

Fraud was not too great a problem when each subscriber was permanently connected by hard wires to a given port on a given switch, so that it could reliably be established what subscriber made what phone calls. But as the cellular industry has dramatically demonstrated, once the subscriber is no longer so readily identified, fraud sky rockets. Now, however, based principally on smart-card technology but also potentially making use of other technologies such as fingerprints, voiceprints, and other biometrics, we know very reliable ways to uniquely

identify the person or entity utilizing the network or network services. The big question here is not whether fraud can be effectively prevented, but rather what balance we should settle upon, between fraud prevention and the customer's convenience and privacy. We—the IXCs as well as wireless and other service providers—will have to gain some experience in order to reach this balance. But there is no doubt that a balance appropriate to both customers and the business world can be reached.

### REGULATORY ISSUES

In a marketplace situation truly like Adam Smith's, there would be essentially no need for regulation of telecommunications services. But that is not where we now sit. During the transition from monopoly telecommunications markets to effective competition, a new regulatory framework is needed. That framework must simultaneously grant the incumbent LECs the flexibility to respond to competition in an appropriate fashion *and* at the same time protect consumers and potential competitors from anticompetitive abuse of the LECs' substantial remaining monopoly power. We must move incrementally from the monopoly-based assumptions and regulations with which we have lived for many decades to as close to the free-market situation as we can practically get. The goals to be achieved include the following:

- A fair chance for all providers,
- Equitable access,
- Spectrum management,
- Elimination of restrictions on who offers what services, and
- International regulations.

Why can't the marketplace settle these issues? Let us look at them one at a time.

#### A Fair Chance for All Providers

After living in a local monopoly environment for many decades, and building local infrastructure with many billions of dollars, and establishing rights of way that are difficult and expensive to duplicate, there is no way to simply drop all local regulation and declare that everybody has a fair chance at the market. One of Adam Smith's postulates to describe a "market" was that there should be easy (equitable) entry and exit from the market. Over the long term, that could be accomplished in local telecommunications, especially with the opportunities provided by wireless transport. But it cannot be done as quickly and easily as renting a building and starting a new restaurant. So we do have a challenge here, to provide a fair chance for all providers, at least for the foreseeable future, under a regulatory framework that simulates competitive marketplace conditions.

The existing local bottlenecks not only give the LECs potential leverage in providing local services; they also provide significant unwarranted leverage to regional Bell operating companies (RBOCs) in provision of long-distance services, if RBOCs are permitted to provide long-distance services before those bottlenecks are effectively eliminated. At least in the short run, those local bottlenecks will continue to exist. Indeed, there may be some network functions for which competition is infeasible for the indefinite future.

#### Equitable Access

Equitable access—one of the successors to "universal service"—has been discussed above. But clearly it must be in this list also, as it will require some form of industry and public agreement as to what is required and how to achieve it in a manner that is fair and equitable to all players—providers and customers alike. That agreement could either be by "regulation," in the usual sense of having it imposed by government, or it could be by mutual agreement among the parties to the process. It remains to be seen what balance of these two approaches will succeed in this case, which is more complex than was the case with plain black phones. A major

challenge will be to achieve a reasonable level of access for essentially everyone, without having to provide such a high level of subsidy that valuable services are priced out of the market and multiple providers cannot survive in that market.

An equally important aspect of equitable access is, of course, access by competitors to bottleneck functionalities, whether those bottlenecks are owned by other competitors or by entities that are not direct competitors but that have the ability to favor some competitors over others by means of their bottleneck functionalities.

### **Spectrum Management**

The management of radio spectrum is one instance in which we believe Adam Smith would agree that there is no practical marketplace. Mr. Smith wisely postulated that in a proper marketplace both the buyer and the seller must be able to understand what they are buying and selling. The radio spectrum is just too complex to be clearly defined in a market situation. I can buy a shirt from you, and we know who owns what. You can buy an acre of land from me, and we can draw clear lines around it. Those are simple two- or three-dimensional objects, which we can clearly define. But how many dimensions does the radio spectrum have? One dimension of frequency; three dimensions of space; one dimension of time, which could be measured in days, weeks, hours, or nanoseconds; one dimension of power; a few (not clear how many) dimensions of modulation technique. We could argue all day about how to count the number of dimensions, or variables, that it takes to describe the radio spectrum. But it is clear that in general the spectrum is too complex an object to be readily handled in a marketplace, especially since it is not easy to establish who might be illegitimately using some piece of it at a given moment.

Therefore, we must continue to have some governmental management of spectrum use. There are some specific cases, as has been recently demonstrated in the case of personal communication service license auctions, where a marketplace can be created. But even there, it was not "spectrum" that was being bought and sold in the marketplace. The process was first to create radio licenses, using the conventional engineering and policy process; then the spectrum managers determined that there was no significant public interest in who (among the qualified applicants) held the licenses; and only after that determination was made could the spectrum manager in good faith put those licenses (not the "spectrum") up for bids. Surely, future spectrum managers could broaden the specifications of particular licenses, so that the license-holders could have more flexibility in the services provided. But the basic spectrum-management process must be maintained.

### **International Regulations**

As we develop into a truly global economy, with global transactions taking place by the millions every hour, we have more and more need for global interoperability of telecommunications systems—not necessarily globally identical systems, but surely systems that can readily talk to each other. The current problems associated with the new wireless communications systems are a fine example. The global system for mobile communications standards are being implemented in most of the nations of the world. If they were implemented everywhere, a person could carry a smart card—called a subscriber identity module (SIM) in this particular case—wherever she or he went, rent a phone at the airport if the wireless frequencies happened to be different from those at home, insert the SIM into the phone, and have access to the same account and many or all of the services available at home. But right now there is no certainty that such interoperability will be achieved in the United States.

The question is, Should there be international regulations to impose standards that would enforce full interoperability? We believe the answer is no, at least in the case of equipment, although international standards are certainly appropriate in certain instances of spectrum management. We suggest that it is up to the providers to decide whether to adopt voluntary standards and be compatible, or bet on some allegedly better technology, and either win or lose in the marketplace. In any case, it seems highly unlikely that the world is ready right now to have an international standards or regulatory body make mandatory equipment standards and enforce them across

national borders. There is a clear need for some level of international spectrum management, but not such a clear need and practicality for international regulation of the details of how internal telecommunications issues are resolved within nations.

### GOVERNMENT ROLES

The federal government's role as a regulator and spectrum manager has been addressed in the above section. But in addition to the issues mentioned above, we must recognize that the federal government does have a significant potential influence as a major purchaser of telecommunication services. That purchasing power could be used—intentionally or not—to influence the direction and speed of development of the NII. We urge the federal agencies to be conscious of this potential, but we insist that the specific needs of specific agencies must not be distorted in order to fit into some effort to use that purchasing power to influence network development.

Another major step that must be taken by the federal government, to affect not only telecommunications but also many other aspects of U.S. business and its global competitiveness, is to relax or eliminate the current restrictions on export of encryption hardware and software. Encryption technology already exists globally, and so the current restrictions have little or no long-term effect on national security but have a major effect on U.S. manufacturers and NII/GII participants.

We also recognize major concerns with the roles of state and local governments. There are major potential barriers in that domain just because of the structure of local regulation. There are, for example, over 30,000 local franchising agencies in the United States, which exercise some control over cable TV systems. Obviously, such a structure could give rise to major problems of compatibility and implementation, for networks and services that will be far more complex than today's cable television.

### THE NEXT 5 TO 7 YEARS?

It is our position that how far we can advance toward the long-term NII/GII goals in the next 5 to 7 years is primarily dependent on how well we do at the incremental process of implementing competition in the local telecommunications marketplace. If we do start implementing that competition quickly, at the state and local levels as well as at the federal level, then we can expect several fundamental changes within that time period:

- We will have fiber to the neighborhoods, in many areas of many cities, but surely not ubiquitously.
- That fiber will provide high-capacity local access, to match the already high and increasingly higher bit-rates and capacities that exist in the long distance networks.
- When and if local competition becomes real, there will no longer be a requirement for the restrictions of the Modified Final Judgment, in manufacturing, information services, and long-distance services by RBOCs.
- As the services offered on those networks become clearer, and their importance to various segments of the education and health care and other communities as well as the general public becomes clearer, we will be able to better define what we mean by "equitable access" to the NII/GII. And with that better definition, and the cooperation of all providers in creating and funding a pool to provide support for those who need it, we feel confident that we will be able to achieve equitable access without imposing undue burdens on ratepayers and service providers.

The ultimate NII will take longer than 7 years. We will not have a fiber to every single home and office in 7 years. But we are confident that, with proper leadership and cooperation at all levels of government and in the laws and regulation, we can move so far toward those goals that we would want never to return to the situation of 1995.

### NOTES

1. Monson, Calvin S., and Jeffrey H. Rohlf. 1993. "The \$20 Billion Impact of Local Competition in Telecommunications," United States Telephone Association (USTA), Washington, D.C., July.
2. Hatfield Associates Inc. 1994. "The Cost of Basic Universal Service," Hatfield Associates Inc., Boulder, Colo., July.

## 53

# Technology in the Local Network

*J.C. Redmond, C.D. Decker, and W.G. Griffin*

*GTE Laboratories Inc.*

The local telephone network is that part of the telephone network that connects individual subscribers, homes, businesses, and so on, to an end-office switching center and includes the end office itself. In its early embodiment, the local network was simply a telephone (or other instrument) at a subscriber's residence or place of business connected to a pair of wires that led to a switching office (Figure 1). At the switching office, connections were made between local users, or the signal was sent via a tandem or long-distance path for subsequent connection through another part of the telephone network.

The early design goals placed on the local network were relatively simple (at least from our perspective today): to provide reliable transmission and switching of voice signals that could be easily understood at the receiving end. There were other considerations, such as ringing the phone, that were also necessary, but, mainly, the subscribers just wanted to hear intelligible voices.

In concept, the local network is not much different today than it was in the past except that the termination at the subscriber's premises is made at a standard interface that does not include the customer's on-premises wiring or telephone (or other equipment). Of course, things are much more complex now, because the demands for added bandwidth, new services, and overall cost efficiency have greatly changed the design goals used to plan and implement the network.

The local telephone network is evolving rapidly from its historical manifestation as a narrowband connection of physical addresses to a more complex network of networks that includes narrowband, broadband, and variable-band transmissions to physical and logical addresses. The added capabilities and increased efficiency of the telecommunications network have allowed the introduction of new data services such as frame relay and switched multimegabit data service; developed new, intelligent features across a broad spectrum of users; and positioned the network for significant growth in the future.

### HISTORY OF THE LOCAL NETWORK

At the time the telephone was introduced, the telegraph was regarded as far more important to commerce. The product of the telegraph was a written record of the communicated message. This tangible record provided a link to the familiar handwritten discourse of the commerce of the day. Because it did not provide a record of the message, the telephone was initially regarded as a novelty.

However, as the need for communications increased, the telephone soon surpassed the telegraph as the medium of choice. In fact, having a telephone became so popular that the proliferation of the supporting wires became objectionable. Engineers were forced to find a more compact means of running the wires from point to point. The engineers found that wrapping the copper pairs with paper insulation and encasing the resulting bundles of pairs in lead allowed a more compact transmission medium, called a cable.

It was also obvious that it was impractical for users to string a pair of wires from their location to each person to be called. The solution was to run all of the pairs of wires from the customer's premises to a central location. There, "operators" could connect sets of wires together according to instructions from the customers. This created a center where customers were switched at the central point of the wires. This was the origination of the old-timer's call, "Hello, Central!"

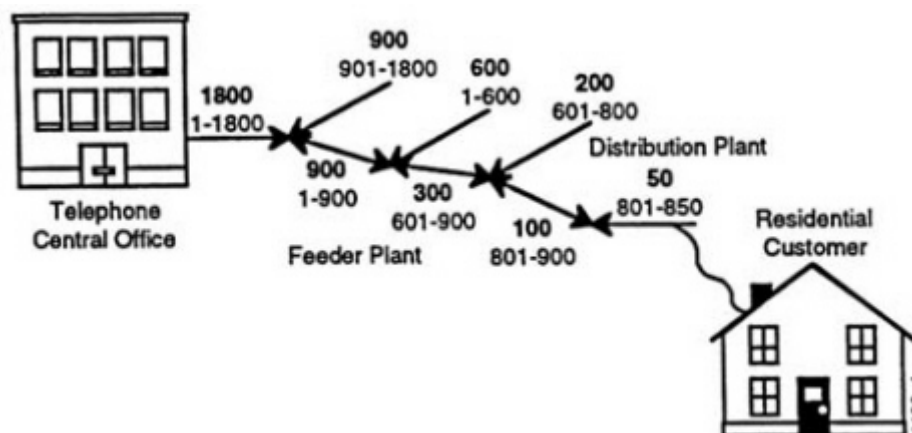


Figure 1 Traditional access to customers.

The switching at Central was originally accomplished by human operators who manually interconnected the customers' calls. The transmission medium was copper wires either cabled or open. Control was provided by the customers' verbal instructions and the operators' manual actions. This division in function (i.e., transmission, switching, control, and terminal equipment) still exists in today's telecommunications networks, albeit in radically different forms.

### Transmission Media and Multiplexing

Transmission equipment for telephony has evolved from simple open wire carrying a single conversation to optical fibers that can carry many thousands of conversations.

Signals from a home or business are carried over a twisted pair of copper wires, called the local loop, to a centrally located local office. Hundreds and even thousands of wire pairs are carried together in a single large cable, either buried underground in a conduit or fastened above ground to telephone poles. At the central office, each pair of wires is connected to the local switching machine. The transmission quality of the early installations was highly variable. Today, however, the plant that is being installed has the capability to transmit at least basic rate ISDN (144 kbps). This is true even for long loops (greater than 12,000 feet) that require loop extension equipment or lower-resistance wire. (Future plans will reduce the number of those long loops.)

In order to reduce costs, methods were developed to combine (multiplex) a number of subscribers on a single transmission medium from the central office, with the individual wire pairs split off nearer to the subscribers (Figure 2). As advances in technology progressed, multiplexing kept pace by increasing the number of conversations carried over a single path. Only a few years ago, multiplexing provided tens of conversation paths over a pair of wires. Initially this was accomplished by shifting each telephone signal to its own unique frequency band.

A major advance in multiplexing was accomplished when normal voice signals were converted into a coded digital form. In this form, the digital signals could be regenerated repeatedly without loss of the voice or other information content.

With today's time division multiplexing, each telephone signal is converted to a digital representation. That representation is inserted into fixed time slots in a stream of bits carrying many digitized telephone signals, with the overall stream operating at a high bit-rate. (An uncompressed voice signal requires 64,000 bits per second [bps] in digital form.)

The multiplexed signals can be transmitted over a variety of transmission media. The most common multiplexing system, called T1, operates over two pairs of copper wires carrying 24 telephone signals, at an overall bit-rate of 1.544 million bits per second (Mbps). First installed in 1962, the system is still widely used today.

With optical fiber, a beam of light is transmitted through a very thin, highly pure glass fiber. The light travels in parallel rays along the axis of the fiber. Many telephone signals are multiplexed together, and the light

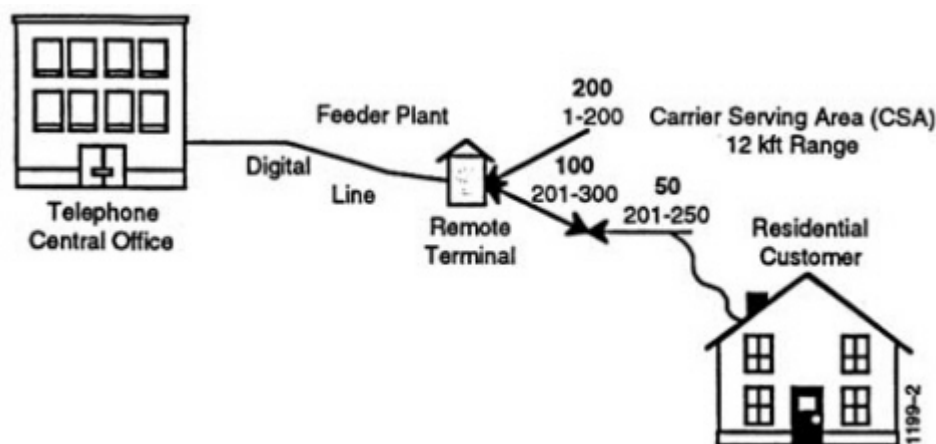


Figure 2 Using pair gain for access to customers.

source is simply tuned on and off to encode the ones and zeroes of the digital signal. A single strand of optical fiber used in today's telecommunication systems can carry 30,000 telephone signals. Also, the repeaters in an optical fiber can be separated much farther (tens of miles as opposed to 1,000 feet) than in an electrical system (see Figure 2).

The greatly increased capacity of fiber links at relatively low cost has led to the practicality of "bypass," in which high usage subscribers bypass the local provider and feed directly into the telecom network (Figure 3).

The history of transmission media and multiplexing shows an ever increasing progression of the total number of conversations that can be carried over a specific generation of the technology. The more conversations carried, the lower the cost per call, or the greater the bandwidth available per subscriber.

### Switching Equipment

The telephone network is a switched network. The connection from one telephone to another is created and maintained only for the duration of each individual telephone call. In the early days, switching was performed manually, by operators who used cords to connect one telephone line to another. The automation of switching was first accomplished by allowing customers to directly control electromechanical relays and switches through a "dial" attached to the telephone instrument. Electromechanical switching equipment reduced the need for human operators. However, the equipment's capacity and capability for supporting new features was limited.

Most of today's switching machines switch signals that are in digital format. Digital switching interfaces well with the time-division-multiplex technology of today's transmission systems.

As technology has advanced, it has blurred some of our old categorizations in the local networks. We now have remote units in the feeder network of various sizes and capabilities that combine the transmission, multiplexing, and switching roles previously accomplished by discrete systems. Initially, these changes were done to reduce costs, but now this added complexity has given the networks much greater flexibility to grow and expand in capability.

### Signaling

The telephone system uses a myriad of control signals, some of which are obvious to the customer and others of which are unknown to him or her. The early signals between the customer and Central were a crank on the magneto to summon the other end. Instructions on whom to call were exchanged by verbal commands.

Today's customers are very familiar with the telephone's ring, dial tone, dialing keypad tones, and busy tone. However, control signals that are sent between switching offices, over circuits separated from the voice



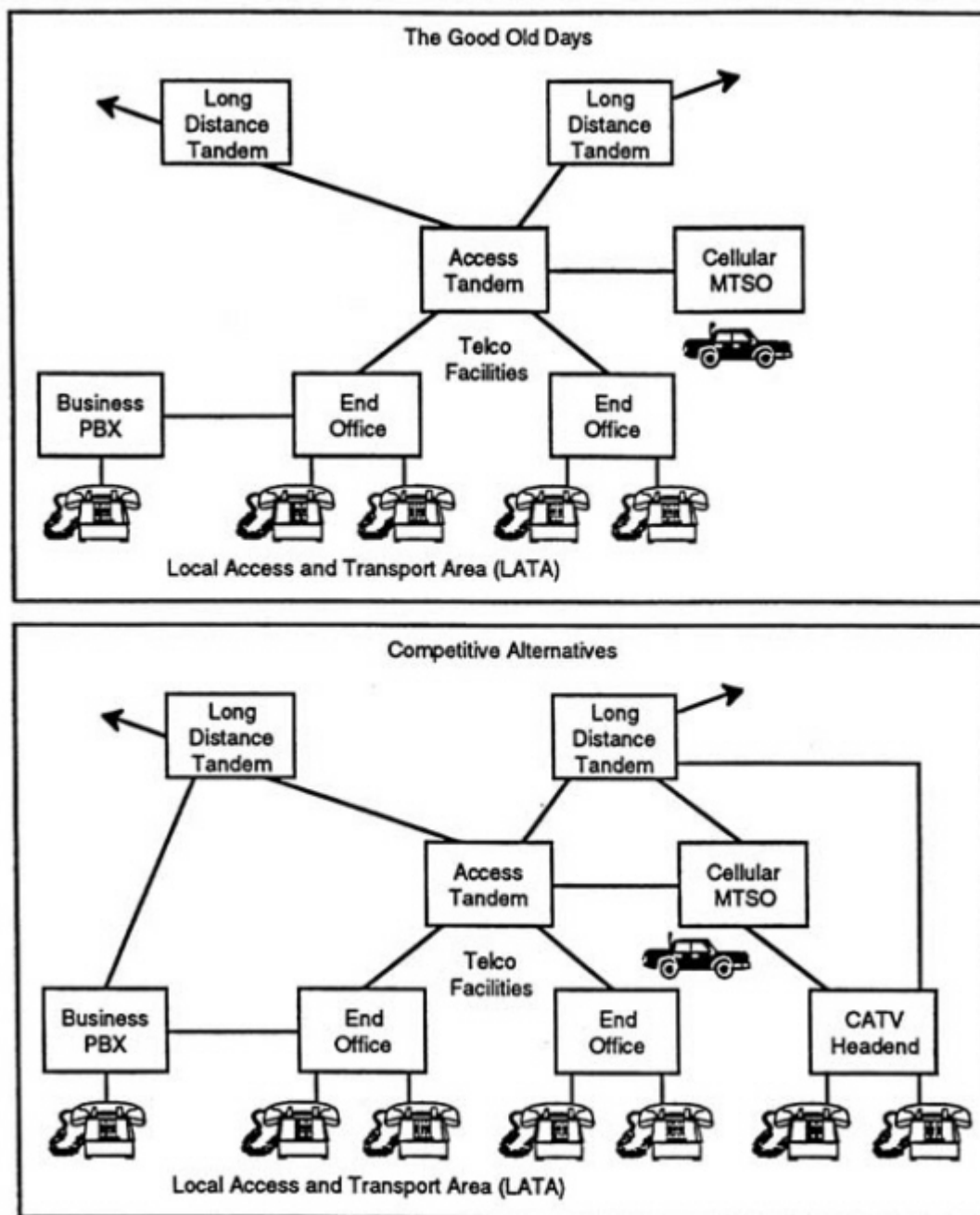


Figure 3 Local access and transport area.

channel, do not directly involve the customer's attention. These separated control signals used within this separate network are called "common channel signaling system number seven," or SS7.

Although the initial motivation in the introduction of common channel signaling was an improvement in call set-up, this change has supported the movement of network intelligence out of proprietary switching systems and into a set of distributed processors, databases, and resource platforms connected through well-defined industry standards.

Thus, we have seen the implementation of the intelligent network in which intelligence is added to the network to implement new features and services such as personal number usage, virtual PBXs, voice response

services, and others (see Figure 4). Further, this intelligence in the network has the potential to evolve to be the service infrastructure for future broadband and multimedia applications such as interactive video games, remote learning, and others.

### Telephones and Other Station Apparatus

The telephone itself is a rather simple appliance. A microphone (the transmitter) and an earphone (the receiver) are contained in the handset. The modern keypad dialer sends unique combinations of two single-frequency tones to the central office to indicate the particular digits dialed.

Newer instruments, especially personal computers, are now common to the network. Their capabilities include simultaneous voice and picture communications and computers with telephone capabilities, and they will include features yet to be invented.

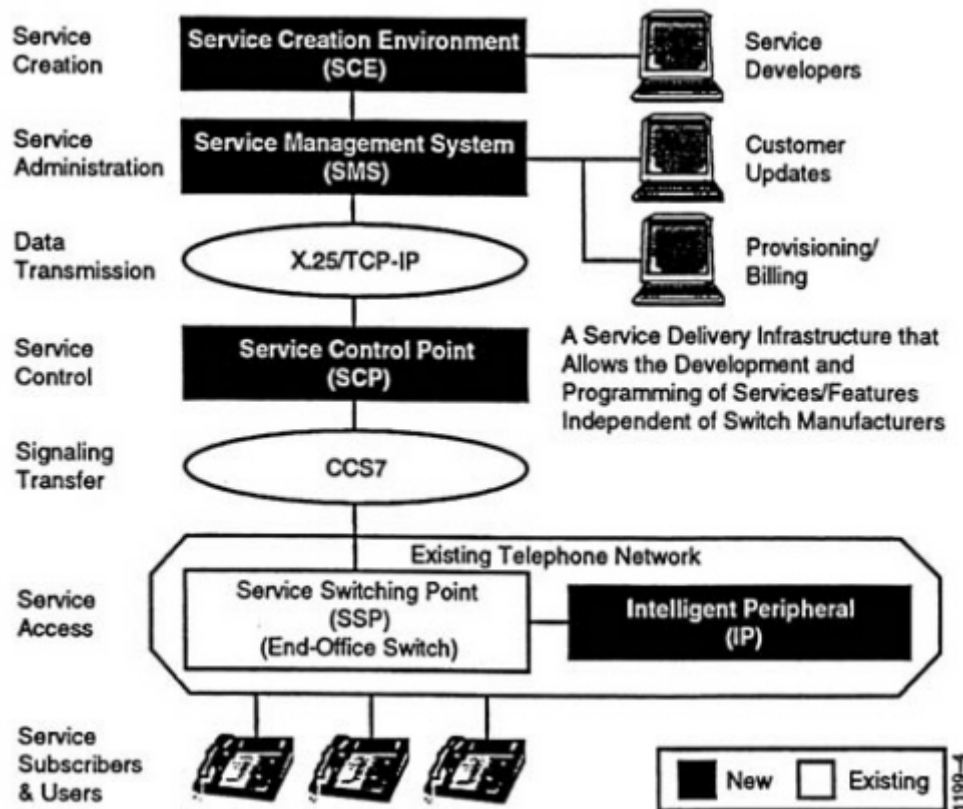


Figure 4 The advanced intelligent network adds significant capability to the network by using intelligence at a service control point or intelligent peripheral to provide new features and services.

### FUTURE EVOLUTION OF THE LOCAL NETWORK

There is currently a revolution under way in the local telephone network that is being brought about by changes in the technical, competitive, and regulatory arenas. From the technical perspective, there has been a great advance in the ability of networks generally to handle and process information. More precisely, there has been a digital revolution brought about by the ever increasing power and ever decreasing cost of microelectronics.

One of the real drivers in this revolution is electronics technology, specifically progress in semiconductor fabrication capabilities. Today's 0.5- $\mu\text{m}$  feature size for complex production integrated circuits is expected to

shrink to 0.25 mm by 1999 and 0.18 mm by 2002. This trend will allow microprocessor speeds to increase from today's approximately 200 MHz clock rates to 500 MHz over the same time frame (see Figure 5). In addition, DRAM (dynamic random access memory) chip capacity will also increase from today's 16 Mb chips to an estimated 1024 Mb capacity in 2002. Putting these advances in electronics to work in development of RISC (reduced instruction set compiler) processors, the heart of desktop workstations or set-top boxes, for example, will mean that these processors can be expected to perform calculations such as those needed, for example, for video compression ten times faster in the years after 2000 than they do today.

While the text files transferred today between users are normally about 100 kb, it is not uncommon for files with graphics information to routinely exceed 1 Mb. It is now becoming common for users to attempt to send 1 Mb files over the existing telephone modem lines, with the result that the commonly used techniques are seen to be quite inadequate. Thus, at least a factor-of-ten improvement in available data rate is required.

The data services that users will soon demand certainly exceed the capability of the existing telecommunications network. As traffic begins to include full-screen, high-resolution color images, files will become of the order of 1 Gb in size, dictating a further increase in capacity of three orders of magnitude. This sort of increase in capability will require some fundamental changes. Growth will be required not only in the pipelines that provide the data but also in file server technology, network management infrastructure, and user software to enable rich new services.

Though we are dealing here explicitly with the local telephone network, the impact of this revolution also affects local networks of all kinds, including telephone networks, CATV networks, private data networks, cellular radio networks, and so on—a profound technological convergence.

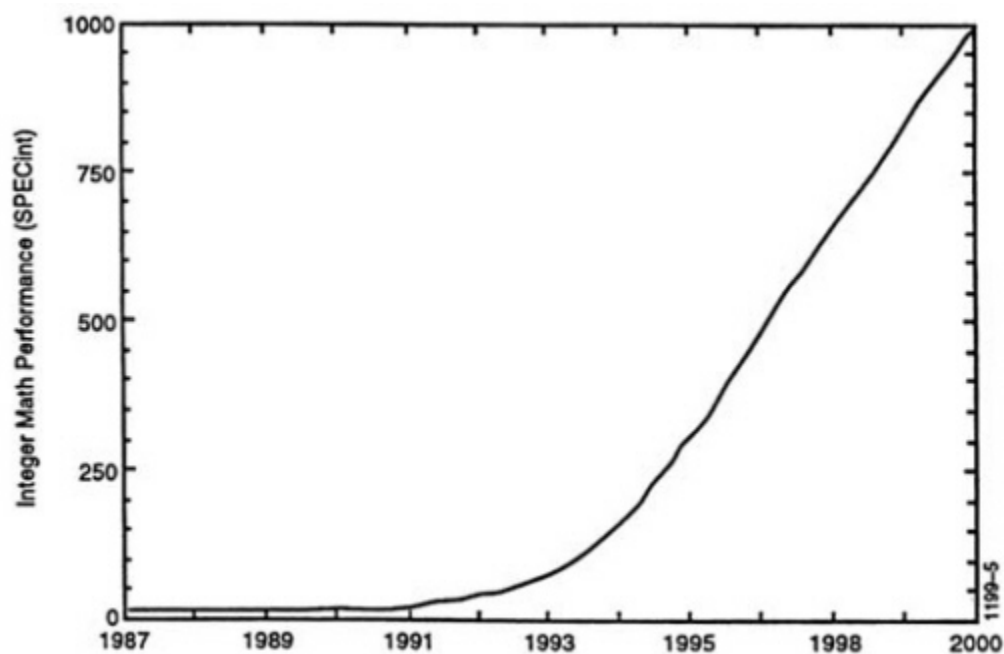


Figure 5 RISC (reduced instruction set compiler) processor performance trends. Data for this graph were taken from manufacturers' specifications as well as from an article by Ohr<sup>1</sup>.

As an example of the data growth envisioned for the network, the increase in DS-1 and DS-3 access lines is enlightening. Figure 6 shows this growth to the year 2003.

Key elements in this robust capability for digital processing are that it is independent of content (such as voice or video) and permits the distribution of the processing to the periphery of the network, thus allowing computing power to migrate to the user. These developments have enabled the conception of a whole range of interactive services that meld and blend the areas of communications, computers, and video. In fact, it is this commonality of technology that has led to other implications in the areas of competition and regulation.

It is the desire to offer broadband video and interactive services that has created the incentive for the local exchange carriers to evolve their plants to provide bidirectional broadband access. Actually, the build-out of the broadband network is a process that has been going on for nearly two decades, beginning with the first introduction of optical glass fiber for carrying trunk traffic in the telephone company (telco) network around 1980. In the intervening years, fiber has replaced virtually all the metallic cable in the interoffice plant and has begun to migrate into the feeder portion of the distribution plant.

The most costly, but at the same time the most restrictive, portion of the access network, the subscriber loop, at this point remains copper. This is key in considering evolution toward a broadband infrastructure. With the digitization of the switching infrastructure, the state of the current network includes a totally fiber interoffice plant, a fully digital narrowband switching infrastructure, and a partially fiber feeder plant.

The approach to a broadband network must be formulated from this vantage point. There are two main technological thrusts that are enabling the digital video revolution. The first is the ability to compress digital video with high quality to the point where one can deliver video streams in a cost-effective way to individual subscribers (Figure 7). Even with the high capacity of optical fiber, uncompressed digital video would have remained a challenge in terms of transport, transmission, and switch capacity.

The other technical event contributing to the availability of digital video has been the development of broadband switching technology in the form of asynchronous transfer mode (ATM). Key features of the development of ATM include the ability to multiplex and to switch together (in a packet or cell format) the content of mixed streams of multimedia traffic, and, what is more, to do this isochronously, so that the time information of each stream retains its integrity. It is anticipated that ATM switches will become dominant after the turn of the century. One plan, shown in Figure 8, predicts 100 percent deployment by 2015.

The first plant upgrade enabled by the digital video revolution has been the migration of the existing CATV network to a fiber-fed technology where fiber is brought to within two or three radio frequency (RF) amplifiers of the subscriber. The fiber-fed bus architecture, or so-called hybrid fiber coaxial (HFC) system, provides capability possible to provide approximately 80 analog and 150 digital channels over a 750-MHz HFC network. If such an

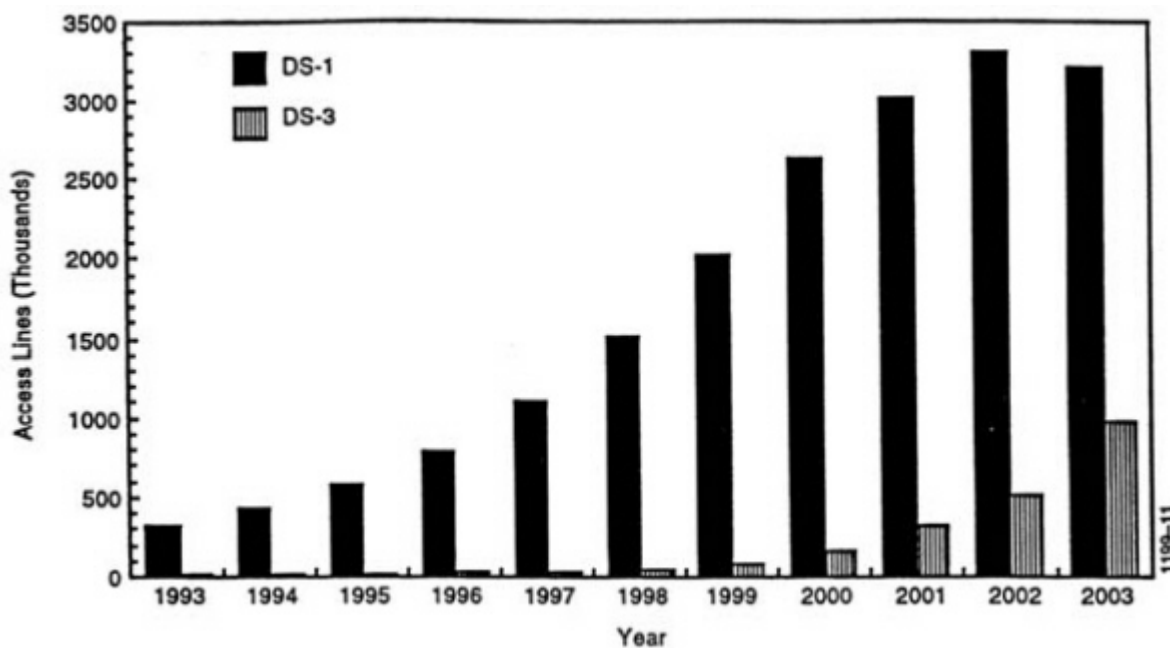


Figure 6 Growth of DS-1 and DS-3 lines. SOURCE: Reprinted from Ryan et al.<sup>2</sup>.

Digital Video Signals	
• Uncompressed broadcast quality	>100 Mb/s
• MPEG 2 (VCR quality)	3–6 Mb/s
• MPEG 1 (VCR quality minus)	1.5 Mb/s
• Video conferencing (movement restricted)	64 kb/s+

Figure 7 Compression of digital video.

HFC network serves areas of 500 homes, there are clearly enough channels available with some statistical concentration to provide selected channels for individual subscribers.

It is GTE's plan to offer 80 analog video channels in the initial rollout of the HFC system to about 400,000 customers in 1995, followed by 500,000 more in 1996. The cost should be in the range of \$700 per customer. Subsequent upgrades will include adding 150 digital channels of broadcast MPEG in early 1996 at an added cost of about \$200 per customer (set-top box). This will allow delivery of near-video-on-demand. In late 1996 or 1997, switched MPEG will be added for video-on-demand and other interactive services for a further incremental cost of \$100 to \$200 per customer.

Further HFC additions beyond 1997 will depend on results obtained with the initial system.

With a large number of channels available, even though the bus architecture is a shared medium, it provides most of the functionality of a switched star-star architecture that is typical of most telephone networks. The enhanced upstream connectivity allows the addition of voice and data as integrated services along with video.

The approach of the local telephone carrier to bringing broadband services to the loop involves the evolution to a broadband distribution network, which includes fiber that will go closer and closer to the subscriber and ultimately to the premises itself. The particular approach to bringing fiber to the loop is a function of cost. At the present time, fiber to the home is too expensive a solution. Bringing fiber to some intermediate point is the preferred option. The hybrid fiber coaxial system, while being implemented initially by CATV operators, is clearly one such approach.

While several local exchange carriers have embarked on network rollout programs with hybrid fiber coaxial technologies for their initial thrust into video distribution, it is clear that there is no straightforward way to integrate HFC with the existing twisted pair copper loop plant (e.g., power, ringing the telephone, etc.). The additional costs of managing and maintaining two networks appear to be a distinct disadvantage for this approach. In some cases, where aging loop plants are in need of full replacement, HFC can be put forward for integrated services and replacement of the existing copper plant.

The challenge is to find a mode of migration that will provide a sufficiently robust set of services to meet customers' needs in the coming broadband environment while maintaining an effective and elegant migration path from the existing copper plant. One such approach is asymmetric digital subscriber line (ADSL) technology (Figure 9). A pair of modem transceivers on each end of the copper loop provides digital line coding for enhancing the bandwidth of the existing twisted pair.

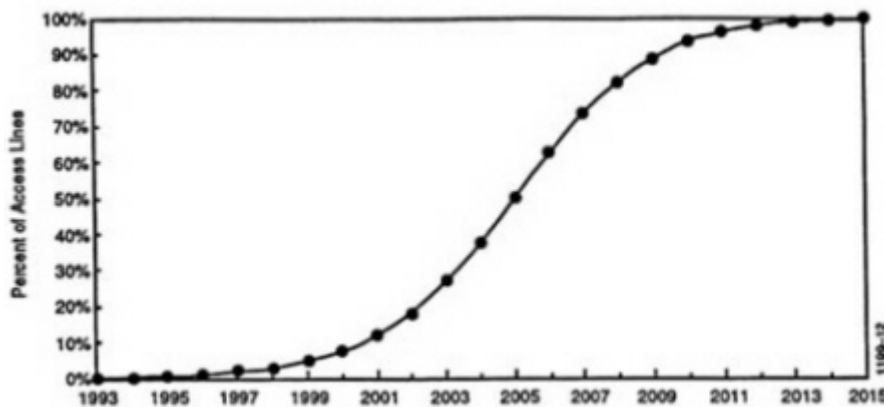


Figure 8 ATM switch deployment.

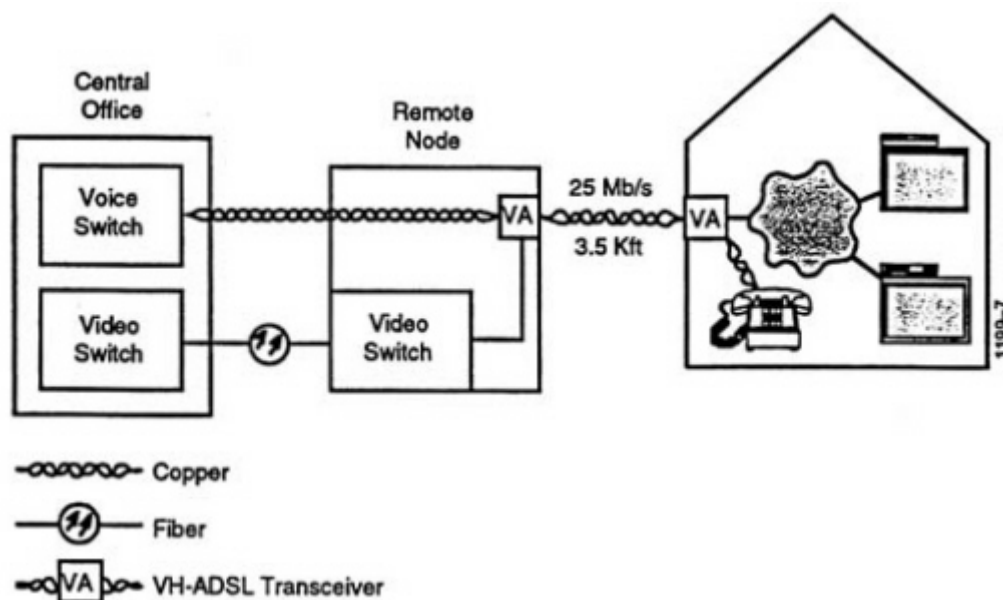


Figure 9 Very high speed ADSL (VH-ADSL) to provide video to the home.

The approach is asymmetric because the capacity or bandwidth in the downstream direction, toward the subscriber, is greater than that in the upstream direction. One ADSL (very high speed ADSL, or VH-ADSL) approach for particularly high bandwidth in the 25 to 50 Mbps range involves serving areas consistent with fiber being brought into the loop within several thousand feet of the subscriber, and is thus consistent with the migration path that brings fiber close to the subscriber premises.

One of the transceivers is therefore installed at a fiber node. Eventually this may go to a fiber-to-the-curb system and ultimately, when economically justified, to fiber to the home. There is a continuum of serving-area sizes. But for the present, this is an integrated network that provides all services over a single plant and is competitive for the range of bandwidths required.

One of the major advantages of this ADSL approach is that it can be applied only to those customers who want it and are willing to pay for the added services. Thus, this method allows for an incremental buildup of a broadband capability depending on market penetration (Figure 10).

This, then, is the infrastructure we will be looking at, but what of the services and programming? It is clear that broadcast services need to be provided in an ongoing way. Additionally, various forms of on-demand or customized video programming formats are anticipated to be important. True video-on-demand commits a port and content to an individual subscriber, and the viewer has VCR-like control of the content. Near-video-on-demand shows a program frequently enough to simulate the convenience of video-on-demand, but without the robustness of true on-demand services. Clearly, other services will be more akin to the interactive features that have grown up on the personal computer (PC) platform. These include various information and transactional services, games, shopping, and, ultimately, video telephony.

The subscriber platform is also worthy of note. There are clearly two converging sources of services here. One is the cable television (CATV) environment, with the set-top box and television set as the platform, and the second is the PC. While the former has been almost exclusively associated with the domain of entertainment services, information and transactional services have clearly been the domain of the PC.

It is clear that the carrier must be prepared to provide services that are consistent with both platforms, because one or the other will likely continue to be favored for specific applications or classes of applications. An example of such a service is embodied in an offering called "Main Street." Main Street uses the TV with a set-top box to provide a visual presentation (currently stills and sound). A telephone line is used to send signals to a

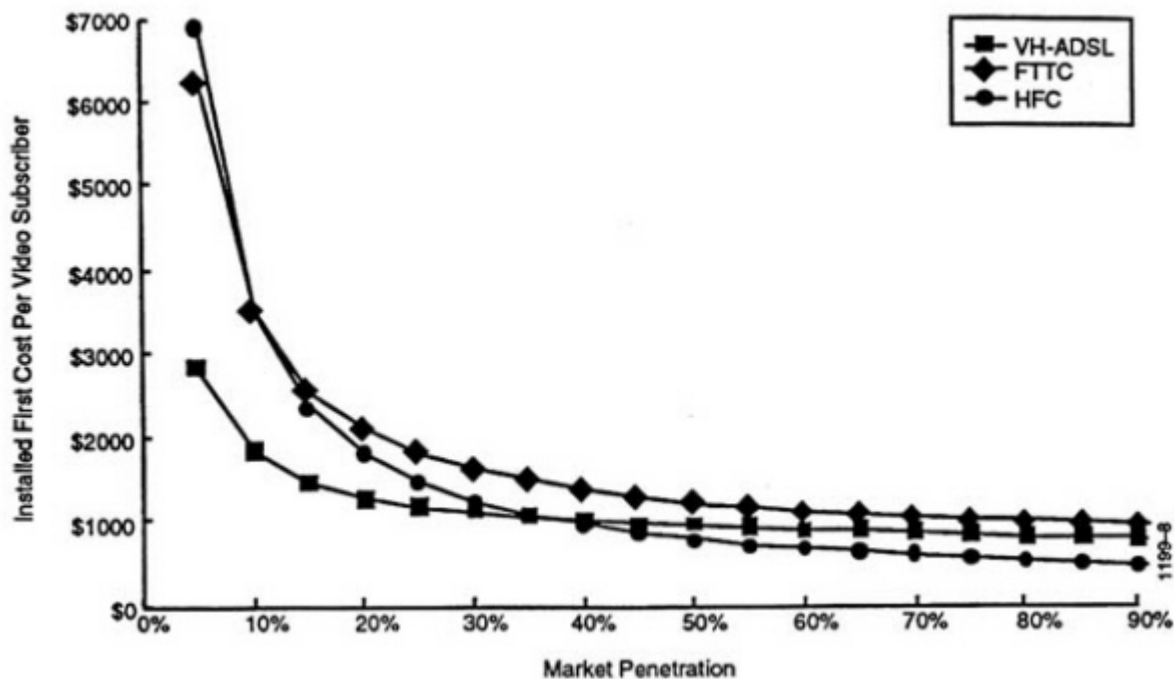


Figure 10 Cost comparisons for very-high-rate ADSL, fiber to the curb, and hybrid fiber coaxial cable versus market penetration.

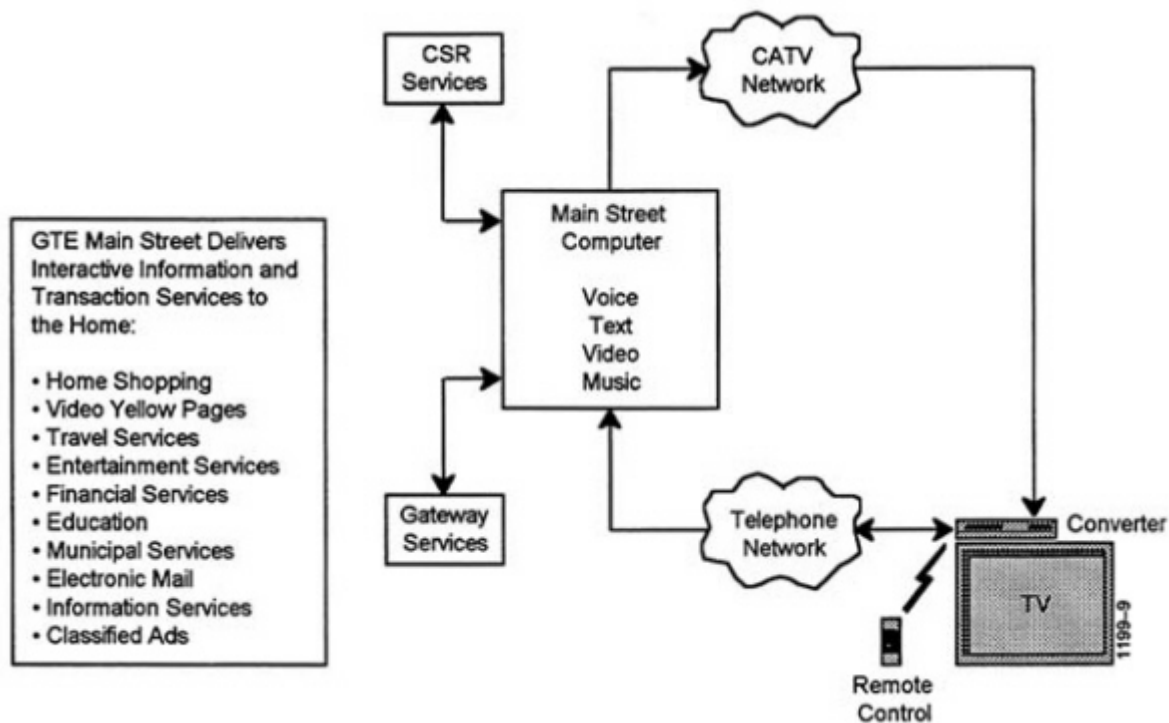


Figure 11 Main Street.

head-end server (Figure 11). Services offered include scholastic aptitude test study, home shopping, access to encyclopedias, games, etc. The service is currently offered at a few locations around the United States.

These new networks of the future are much more complex and require a corresponding increase in intelligence. Intelligence (defined here as the ability to self-inventory, collect performance data, self-diagnose, correct faults, respond to queries, etc.) no longer resides exclusively in the central office but is spreading into the local access network, thanks to the plummeting cost and increasing reliability of processing power (Figure 12). The new distributed intelligent network elements will enable a revolution in the efficiency with which telcos can perform core business functions needed to administer customer services. For example, telcos have historically billed for service based on time of connection, bandwidth, and distance. This approach has little meaning in the case of connectionless data transmission, and new approaches need to be formulated.

Similarly, monitoring and testing network performance will attain new levels of efficiency as digital performance monitoring and automatic alarm collection/correlation move down to the individual line card level. Probably the most exciting aspect of these new intelligent access elements is the new services they will make cost effective. Reducing the cost of digital services such as ISDN and frame relay, and of higher-bandwidth services such as interactive multimedia, will require intelligent elements in the access networks. Dynamic service provisioning, whereby services are delivered in real time on an as-needed basis, will similarly rely on intelligent network elements. As these examples show, the addition of intelligence to the access network represents a new paradigm for the rapid, efficient, and cost-effective addition of new services.

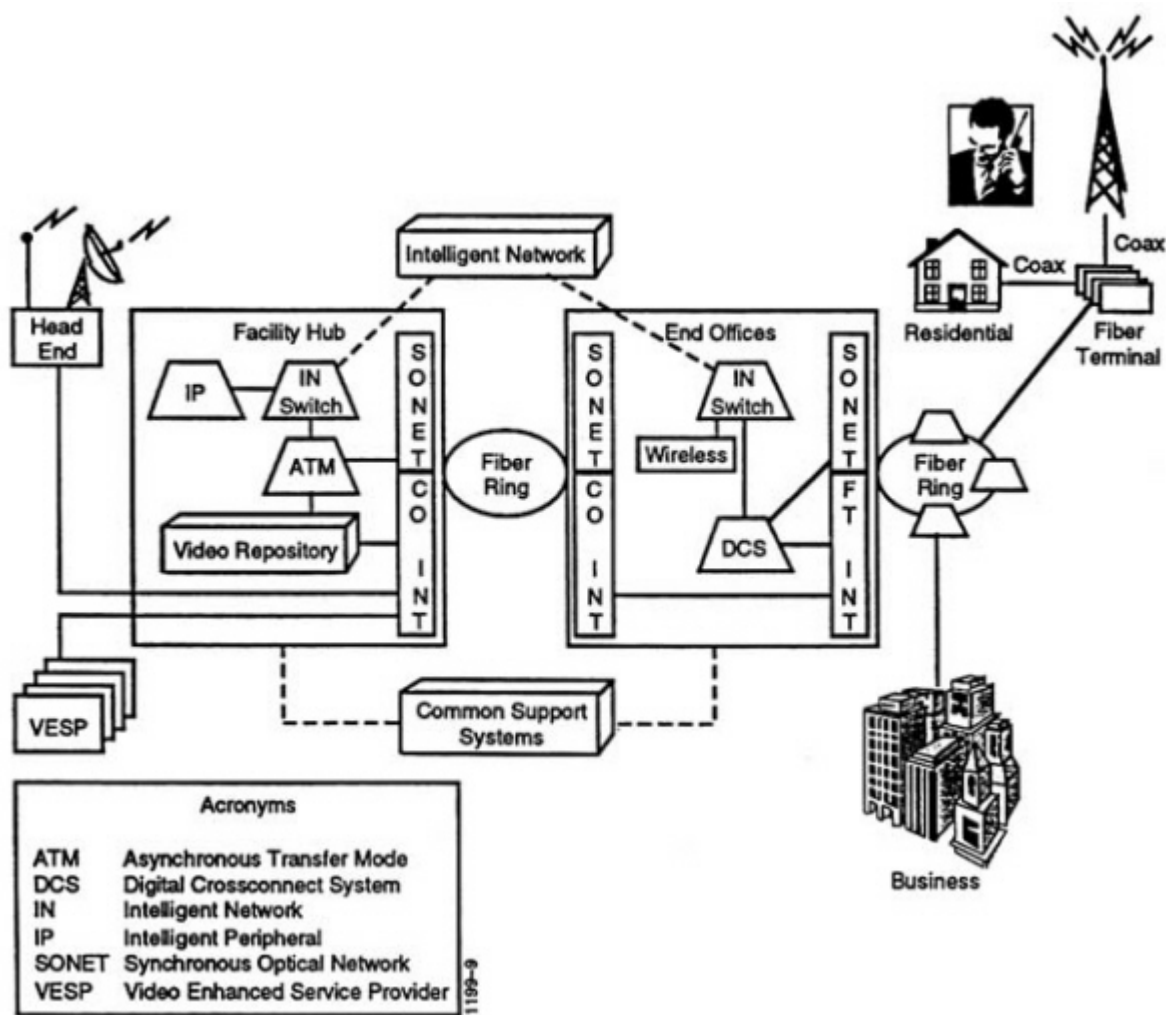


Figure 12 Future network.



Telecommunications management networks (TMN) is another emerging technology that will reduce operational costs by enabling more efficient network management<sup>3</sup>. concept first appeared in the early 1980s, but activity remains primarily in the standards arena, where organizations such as the International Telecommunications Union, the American National Standards Institute's T1, and the European Telecommunications Standards Institute continue to evolve a complex set of TMN standards.

In the existing situation, each network (e.g., those of long distance carriers, local telephone networks, wireless networks, etc.) has its own specialized, proprietary management network, with little or no interoperation. Indeed, at present each network element has a unique interface to the management systems. The TMN approach, which has support across the communications industry, will result within the next decade in an open, multivendor environment with the benefits of reduced costs for operations systems and improved management efficiency.

### THE LOCAL NETWORK AND THE NATIONAL INFORMATION INFRASTRUCTURE

In order for the greatest number of individuals, residential or business, to access the NII, there must be support to allow a variety of types of customer premises equipment (CPE) to gain access to the network as well as to enable interoperability<sup>4</sup> among the pieces of the network. Users are going to want to be attached to the NII from any of various types of CPE (e.g., a plain old telephone system telephone, screen-phone, personal digital assistant, PC, or TV). They will connect to the network either by dial-up (wired or wireless), leased line, telco-provided video network or cable service, or by interfaces provided by the utility companies.

The interoperability of these network access types is going to be dictated by the need for users (and their applications) to get access to other users without needing to have multiple types of CPE and/or set-top boxes and without having to know what's on the "other end of the line"<sup>5</sup>. In addition, the NII will need to support a vastly increased degree of interoperability among both the attached information resources (and their providers) and the active components (e.g., personal "agents") that wish to make use of these resources.

Just as the NII is often discussed in the context of an extrapolation of today's Internet, so also the problems of interoperability in the NII can be thought of as an extrapolation of the simpler problems of interoperability in the context of information processing currently being extended to distributed computing. The current information processing infrastructure involves a vast legacy of networks of heterogeneous, autonomous, and distributed computing resources, including computers, applications, data (files and databases), and the underlying computing and communications technology. We endorse the Computer Systems Policy Project report, *Perspectives on the National Information Infrastructure: Ensuring Interoperability* (CSPP, 1994), which focuses on the importance of interoperability.

We want to add to that report's statements that developers of NII technology cannot assume a clean sheet of paper for new applications. Most current data are, unfortunately, tightly coupled to legacy applications. In order for the current vast repositories of data to be available to future NII applications, interoperability mechanisms must be developed to provide access to these data and, in some cases, to allow current legacy applications to participate in the future. These legacy applications and associated data represent huge investments and are unlikely to be rewritten to upgrade their technology base.

A major consideration for users on the local network is the NII interface. This interface requires an underlying communications model that includes how a user or application can connect to another user, service, or application by using customer equipment in a network-independent manner, and with a set of relatively simple formats for accessing the various kinds of data that will be available in the NII. To accommodate functionally limited CPE, transformation gateways may have to be provided at the central office close to the end user.

### PLANNING FOR THE FUTURE OF THE LOCAL NETWORK

The capabilities of the communications networks have expanded significantly so that, from a technical standpoint, there are major overlaps. Telcos are now attracted toward video delivery, cable companies toward

providing telephony and data services, and broadcasters toward participating in cable networks, with all involved in programming to varying degrees.

This state of affairs begs the obvious question: Why do other businesses seem so attractive? This appears to be based on two assumptions. The first is that, for the current participants, their networks can be modified to handle the total communications needs of customers (voice, data, video) by relatively modest incremental investments, and the resulting, greatly expanded network can be managed with essentially the same management team.

The second assumption is that the communications environment brought about by the digital interactive age promises a growth in and demand for new services that have not been experienced in recent times, if ever before. Thus the stage is set, at least on the telco side, for an entry into traditional entertainment and content-based services as well as an extension from voice and data into multimedia communications.

The evolution of the local network is also being affected by the changes that are taking place in the regulatory arena. There has been a general social trend toward a more competitive market-driven environment as well as a technical basis supporting deregulation, given that technology was establishing the basis for multiple providers for the same service.

This has begun with the Federal Communications Commission's video dialtone rulemaking, which provided the basis for telcos to offer transport services for video within their franchise areas. The most problematic business areas with respect to regulation have been those associated with content and programming. These have been driven by the FCC, based on the traditional concern with control of communications being in the hands of a single entity.

This has been a particular issue for the local telcos for several reasons. First, the video dialtone enabling regulation initially has proscribed the involvement of a telco in providing content over the video dialtone (VDT) network. This has been challenged by several of the local exchange carriers in federal court, with universally successful results to date. While there are still appeals processes to go through, it is clear that the First Amendment right stands significantly behind the carrier's positions.

The concern here has been the availability of extensive competitive programming for initial rollout of the network, in order that there be a basis for competing with the current incumbent. For this to happen, a local carrier must be allowed to prearrange to some degree the initial availability of programming, or the exercise will be one of building the video dialtone infrastructure and hoping that sufficient programmers and/or subscribers will arrive to pay for the cost of the network.

A second incentive for involvement in content and programming is the apparent structure of the current business, whereby significant leverage and profitability remain with the content, and delivery may be a less profitable part of the business. This may or may not be true in the future. It is clear, however, that access to competitive content and programming is essential in an era of video competition.

As the executives responsible for managing the companies that provide local service ponder the actions they should take to remain competitive (and to grow), they are faced with a number of uncertainties that are external to their companies, such as technology advancements, competitive actions, and governmental regulation.

This is a normal state of affairs for many businesses, and any business opportunity is normally undergoing a number of dynamic changes on a more or less continuous basis. In technology, the changes include such things as cost-per-processing capability, data storage capacities, and the like. These elemental changes may allow higher-level changes that dramatically affect the cost and performance of the business opportunity. In regulation, such items as allowing competition into new markets, pricing freedoms, and others likewise have an effect. And, of course, competitive changes occur as well.

With highly dynamic business opportunities, the executives running the competing businesses have to decide what approach to take to address the opportunity and when to make a commitment. If the commitment is made too early and something significant occurs, a competitor may come in slightly later and address the same opportunity with a product or service that costs less and does more. In the event of delay, on the other hand, there is the danger that the market will be taken and will be difficult to penetrate.

There can be little doubt that there will be many approaches to providing communications services in what is now the local loop. Indeed, it is the vitality of this massively parallel approach that is one of the great strengths of our economic system.

If the playing field is truly competitive and without undue restrictive regulations, the best approach, with the most efficient use of technology, should win out. Whether this winning approach will be a natural evolution of the present telephone network, an expansion and upgrade of the cable network, a wireless approach, or some hybrid of all of the above, in a competitive environment, the ultimate winners will be those who provide what the market wants at the most economical price.

Many of the decisions to be made in this arena entail large capital and work force expenditures that, once committed, are expensive to change. Indeed, much of the investment in the local networks has not been recovered, and large write-offs could occur if decisions are not made well.

The executives who are faced with the decisions on how best to address these markets can protect themselves from being blindsided by new technology development by becoming aware of the possibilities raised by new developments. They can likewise measure the competition and feel that they all are at least playing by the same economic rules. It is on the regulatory front that a more aggressive effort is needed to establish a level playing field. The network managers have to be confident that they are in a fair competitive match, and that there are no underlying rules that favor one competitor over another. If this occurs, then the decisions can be made.

The timing of the changes we can foresee, as discussed in this paper, is uncertain. The nature and timing of regulatory reform play a large role in advancing our telecommunications network, and regulatory reform must be completed.

Currently, there is a lot of asymmetry in the application of regulatory rules. The telcos, for example, have a utility company obligation to provide basic telephone service at tariffed rates to anyone who requests it in their serving territory, and to have sufficient reserve capacity to act as a carrier of last resort if some small facilities-based reseller experiences a network failure. For years this mandate has been financed by subsidizing individual local telephone service from business service and long-distance service so that billed local revenues are less than the cost of providing service. This situation is slowly changing as access charges are being lowered and local service rates are being increased in some jurisdictions. However, the situation that persists in many states is that residential local service rates are deliberately set below the actual cost of providing service. Moreover, in several states, the only way the local telco has been able to get any pricing flexibility for competitive services is to agree to price freezes for residential service.

While regulators have decreed that alternative access providers be granted at least virtual collocation in telco central offices to facilitate local telephone competition, they have not yet decided what obligation these new competitors have to share in the local universal service mandate. In that connection, GTE has proposed to the FCC, to the National Telecommunications and Information Administration, and to the committees in the House and Senate that are writing communications reform legislation that the entire array of explicit and implicit universal service funding mechanisms be reviewed as a whole, rather than piecemeal, and that a process be established whereby multiple providers could become carriers of last resort (COLR) and eligible to receive universal service funding. Whether or not this proposal is accepted and implemented, it is clear that until the ground rules for local telephone competition are acted upon and settled, the regulatory environment could continue to discourage some competitors from rapidly building out or evolving to universal broadband networks in the local loop.

The ultimate solution is the establishment of a highly competitive communications environment that would be market-driven to provide the information any customer may want or need, whether it be "just plain old" lifeline telephone service or a broadband, interactive multimedia connection. The challenge for legislators, government regulators, and business leaders is to come up with a process that moves from today's situation to this desired goal. The evolution of the local plant is really the key and, as discussed above, occupies much of the thought and planning of local telephone company operators and others who want to participate in that market.

## NOTES

1. Ohr, S. 1995. "Fast Cache Designs Keep the RISC Monster Fed," *Computer Design*, January, pp. 67-74.
2. Ryan, Hankin, and June Kent. 1994.
3. Glitho, R.H., and S. Hayes. 1995. "Telecommunications Management Network: Vision vs. Reality," *IEEE Communications*, Vol. 33, March, pp. 47-52.

4. Generally, we say that there is interoperability between two components  $X$  and  $Y$  when they can interact based on a (degree of) mutual understanding.
5. For example, you may want to listen to a video on a POTS phone, interact with a database service using your TV, or have a video conference with two of your colleagues at the same time when one of you is using a computer, one is using a screen phone, and the third is using a TV. The situation today is that each user would use a distinct CPE depending on the purpose.

## 54

# Recognizing What the NII is, What it Needs, and How to Get it

*Robert F. Roche*

*Cellular Telecommunications Industry Association*

### STATEMENT OF THE PROBLEM

The national information infrastructure (NII) offers to strengthen the U.S. economy and promote social and educational development, but those contributions depend on its deployment. In turn, the ability of wireless providers to deploy the facilities necessary to provide the wireless component of the NII depends on the government's recognizing—at all levels—that restrictions on deployment also restrict these benefits. The Cellular Telecommunications Industry Association (CTIA) has urged policymakers to adopt policies that will promote deployment of the NII. These policies include (1) an Executive Order from the President directing federal agencies to make available federal lands and sites for telecommunications facilities; (2) an affirmation by the Federal Communications Commission (FCC) of the primacy of the national technical standards applying to radio frequency (RF) emissions over local standards; and (3) an affirmation by the FCC of the primacy of national telecommunications policy over local policies that are hostile to competition.

### SUMMARY

Wireless telecommunications is making great contributions to the deployment of the NII. It has already met the mobile needs of over 25 million consumers in the United States. Wireless services are meeting the need for wireless in-building services as well, and their potential is phenomenal. For example, the ability of our schools to offer students a rich experience and access to a broader information base often runs up against the fact that most schools are not currently wired for telecommunications and computing, and that wiring these schools may pose the risk of exposure to asbestos or the expense of extensive renovation and removal operations. Wireless telecommunications and computing offer, in these cases, more cost-effective and efficient alternatives to wired systems<sup>1</sup>.

Wireless telecommunications is successful because it flourishes in an environment of competition in lieu of government regulation. This wireless paradigm has resulted in more than 200,000 new jobs over the past 10 years, and almost \$19 billion in private-sector investment<sup>2</sup>. In spite of these gains, and the promise of as many as 1 million new jobs and another \$50 billion in investment over the next 10 years, there are impediments to total success<sup>3</sup>. Wireless service is dependent on the deployment of antenna facilities—cell sites—and the ability of wireless companies to deploy the facilities for new systems, greater capacity, and broader coverage is at risk. Some local jurisdictions are preventing the deployment of antennas, either through outright bans, extensive delays, or application of unscientific "local technical standards" to radio frequency emissions. CTIA has called for action to redress these problems and to permit wireless to assume its full effective and efficient role in the NII.

### BACKGROUND

Much of the discussion of the NII has focused on wired technologies—predominantly fiber optics—as the core of the NII. That focus fails to recognize that wireless technologies already make up a significant part of the

national telecommunications network. Wireless already reaches across the United States, with hundreds of wireless companies competing to offer voice and data services, innovative applications, and value to millions of consumers. These hundreds of wireless service providers have been developing, funding, and deploying a *wireless* NII for over 10 years, since the first cellular system began operating in October 1983.

For all of that, wireless has almost been the secret success story—perhaps because it is the success of private enterprise. Over the past 12 months, there have been 19,043 references to the NII or the information superhighway in the media<sup>4</sup>. Of those stories, only 2,139 mentioned wireless or cellular. Of course, the reality is sinking in that the NII—or the information superhighway—is and must be more than a high-fiber diet (of fiber optic cable and other hard-wired systems). The reality is that people are mobile, and mobility implies being wireless. But being fixed does not necessarily mean being wired. Indeed in many environments—urban and rural—fixed services are better delivered by wireless technology than by wired technology.

CTIA, as the industry association for wireless providers—ranging from cellular to enhanced specialized mobile radio (ESMR), satellite, and personal communication services (PCS)—has been relentless in pressing this message. CTIA and its members also have been relentless in making it a reality. Indeed, the CTIA Foundation for Wireless Telecommunications has cosponsored and cofunded wireless education and wireless medical projects across the country (Box 1 gives two examples)<sup>5</sup>.

Increasingly, wireless is being recognized as a vital part of the NII. This forum is one example of that recognition. Last year the National Institute of Standards and Technology's Committee on Applications and Technology requested—and received—comment on the demand for an NII, and on the role of wireless in the NII<sup>6</sup>. The Office of Technology Assessment issued a report on "Wireless Technologies and the NII" in August 1995. This recognition, however, is only the beginning of the battle.

#### BOX 1 EXAMPLES OF CTIA-SPONSORED WIRELESS PROJECTS

##### **Wireless at Work in Education**

On May 2, 1995, the CTIA Foundation, Bell Atlantic Mobile, and Cellular One donated state-of-the-art wireless telecommunications systems to two elementary schools in the District of Columbia. The ClassLink<sup>SM</sup> initiative intends to improve education by bringing wireless telecommunications and information to now-isolated classrooms, allowing schools to link with the Internet via wireless modems.

##### **Wireless at Work in Medicine**

The CTIA Foundation is funding a project at New York's Columbia-Presbyterian Medical Center where wireless is providing a system of coordinated care to tuberculosis patients. The project, done in conjunction with the New York City Department of Health and the Visiting Nurse Services of New York City, enables visiting nurses equipped with laptop computers and wireless modems to treat patients in their homes.

Most of the wireless components of the NII—cellular, ESMR, and PCS—require the deployment of cell sites as their basic building blocks. These sites comprise antennas and towers, as well as base station equipment. The cellular industry alone, composed of two carriers per market, constructed almost 15,000 cell sites between 1983 and 1994. By the end of 1994, almost 18,000 cell sites had been constructed (Figure 1).

As Figure 2 indicates, cell sites have traditionally supported service to between 1,000 and 1,200 users per site. As the number of subscribers increases, the number of cell sites must likewise increase in order to meet demand and preserve service quality.

Another 15,000 cell sites may be required for cellular systems alone in the next 10 years, based on the projections of Barry Goodstadt of EDS Management Consulting that cellular might achieve subscriber levels between 38.2 million and 55.1 million by 2006<sup>7</sup>. (Although the deployment of digital technology might reduce the absolute number of additional cell sites required to meet demand because of capacity restrictions, the number of cell sites required must still increase in order to improve geographic coverage. Thus, the precise number of such cell sites is a matter of speculation and is not definitively predetermined by subscribership.)

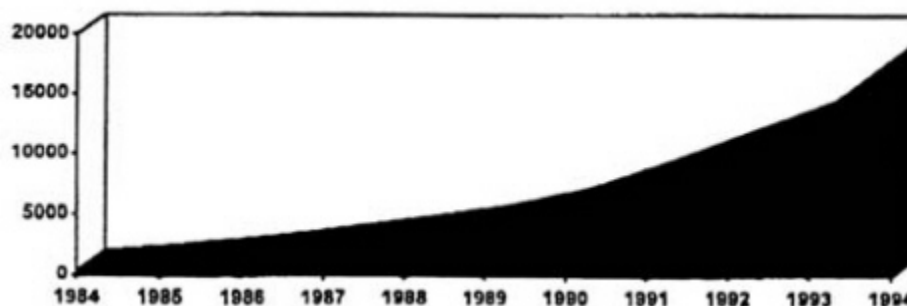


Figure 1 Cell site construction (reported as of December 1994).

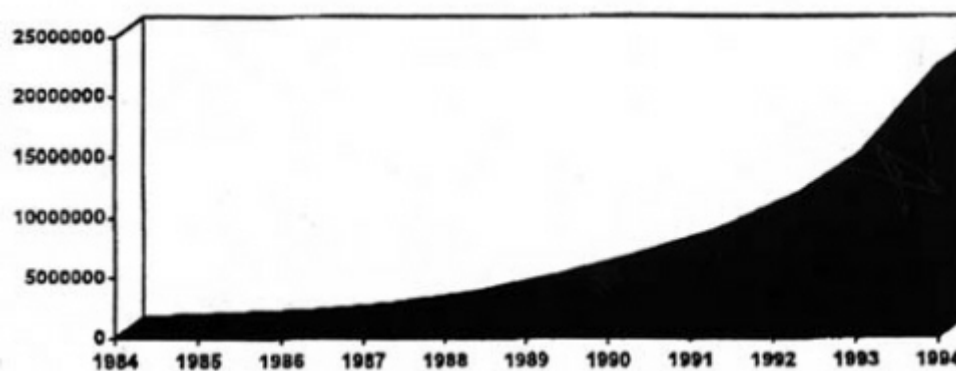


Figure 2 Reported subscribers (as of December 1994).

The construction of Nextel Communications' ESMR network demonstrates that the need for such sites is not limited to cellular systems. By late 1994, Nextel was constructing 2,000 cell sites for its ESMR network, and it had over 130 cell sites in Los Angeles alone. As of year-end 1994, Nextel planned to construct 4,000 cell sites over the next 5 years to complete its network<sup>8</sup>.

With the deployment of PCS, the number of cell sites will increase dramatically. During the FCC's PCS proceeding, would-be licensees estimated that they would have to construct between four and seven cell sites in order to provide coverage identical to that of one cellular cell site<sup>9</sup>. The adoption of higher base station power limits will facilitate the use of wide-area cells in some areas, thereby permitting a one-to-one relationship for PCS and cellular cell sites in those limited areas. However, the need to deploy micro- and picocells in order to provide capacity and coverage in other environments (e.g., urban settings) means that the number of cell sites will still increase many times.

Indeed, the rules contemplate six licensees per service area, with 5-year deadlines obliging three PCS licensees (each holding 30 MHz) to construct systems covering one-third of their service area population, and three PCS licensees (each holding 10 MHz) to construct systems covering one-quarter of their service area population over that period. Thus, the number of cell sites required will necessarily be some multiple of a figure based upon a single cellular licensee's system—as PCS licensees must either build out their systems or forfeit their licenses.

#### ANALYSIS AND FORECAST: WHAT BUILD-OUT WILL PCS REQUIRE?

As previously noted, the FCC has imposed build-out requirements for both major trading area (MTA) and basic trading area (BTA) PCS licenses, such that three licensees will be required to cover one-third of their population in 5 years, and three licensees will be required to cover one-quarter of their population in 5 years. This means that, even starting with a static model, and assuming that the size of the population to be covered is established by 1994 population figures, the equivalent of three nationwide carriers will be obligated to build out

systems so that each covers 86.5 million people, and the equivalent of three more nationwide carriers will be obligated to build out systems so that each covers 65 million people.

The number of cell sites required to provide coverage for current cellular systems may be used to form the basis for extrapolating PCS system build-out. Using the midyear 1994 number of cellular sites—over 15,000—we may calculate that *as many as 22,500 cell sites may be required to provide coverage for the equivalent of one hypothetical nationwide PCS system*<sup>10</sup>. Since there are, in fact, six PCS systems to be licensed, including both the MTA- and BTA-based systems, the numbers of such sites may be as much as six times that base figure. (On the other hand, the number of sites may be lower if operators obtain multiple licenses in a single market, such that an MTA licensee also holds a BTA license in each of its component markets and thereby simply increases its available spectrum resource to 40 MHz across its service area.)

Assuming that the MTA and BTA licensees' coverage requirement can be used to derive a ratio for cell site construction, we can make the following projections:

- The two (theoretically nationwide) 30-MHz MTA licensees and the similar 30-MHz BTA licensees together could be obligated to build a minimum of 22,500 cell sites (towers or antenna sites) within 5 years. (This assumes that each builds 7,500 cell sites or one-third of the theoretical 22,500 maximum required for one nationwide PCS system.)
- The three 10-MHz licensees could be obligated to build a total of 16,875 cell sites over the same period. (This assumes that each builds 5,625 cell sites, or one-quarter of the 22,500 maximum required for one nationwide system. This assumption may not be accurate, depending on the applications that these companies seek to deliver to the marketplace. In fact, the 10-MHz licensees may have to deploy three times as many cells as 30-MHz licensees to achieve equivalent maximum capacity through frequency reuse.)

Thus, the broadband PCS licensees could build about 39,275 cell sites over the next 5 years. Including cellular, ESMR, and PCS, wireless licensees could require as many as 58,275 new cell sites within the next 5 years. (This assumes build-out of multiple PCS systems, build-out of Nextel's system in the projected time frame, and cellular build-out in line with recent growth rates.) In fact, this may be a gross underestimate of the number of antenna sites required, in light of projections by Paul Kagan Associates (Figure 3) and other analysts that over 124 million people—almost 45 percent of the U.S. population—will subscribe to wireless services in the next 10 years<sup>11</sup>.

Indeed, the number of new cell sites may range as high as 100,000 or more for the complete build-out of all these licensees, if the equivalent of four nationwide PCS systems are deployed<sup>12</sup>.

### WHAT'S THE PROBLEM?

The fact of the matter is that in order to create the wireless component of the NII, these towers and antennas must be deployed. Wireless service is dependent on their existence. But the ability of service providers to bring these services to the public is handicapped by (1) *the lack of a process*, in some jurisdictions, for granting the necessary permits to build these facilities; (2) *a process*, in some jurisdictions, that actually hampers deployment by imposing unnecessary delays and transaction costs; and (3) *some shortsighted actors* who, in some jurisdictions, actually seek to prohibit competition or restrict the build-out of wireless services by imposing unscientific "local technical standards" on RF emissions.

It is important to note that these jurisdictions have failed to recognize that national policy puts competition to work for their own citizens' interests<sup>13</sup>. Other jurisdictions have erred by applying local "technical" standards that conflict with the national RF standards adopted by ANSI and accepted by the FCC<sup>14</sup>. Indeed, some of these decisions go beyond what even their own technical experts recognize as valid<sup>15</sup>.



We are not singling out a specific level of government as the culprit. Federal, state, and local governments can help or hamper the creation of the NII. In fact, state governments like those of Maryland and Connecticut have demonstrated a clear understanding of the benefits of promoting deployment—and have moved to help extend the wireless superhighway. Maryland, for example, has solicited Requests for Proposals and has issued a list of state government-owned antenna sites that the state will make available to applicants for suitable fees<sup>16</sup>. The state of Connecticut has essentially established a statewide standard and process applicable to antenna siting, under the auspices of the Connecticut Siting Council.

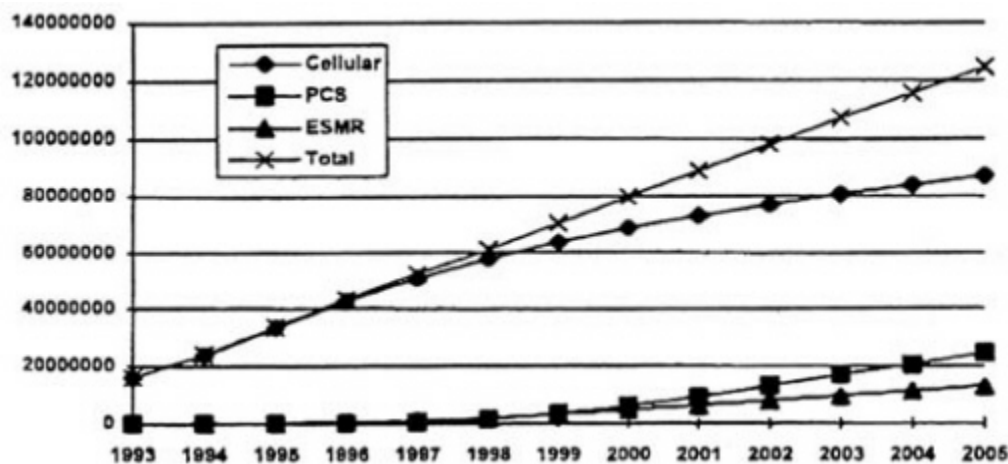


Figure 3 Paul Kagan Associates' wireless subscriber projections (October 1994).

### WHAT SHOULD BE DONE?

CTIA is not advocating a national zoning policy. CTIA is advocating that consumers be helped and not hampered by government at all levels.

Zoning rules that act as a direct bar to entry—such as moratoria on tower or antenna siting—do not help consumers because they obstruct competition and actually reduce the quality of service available to consumers. Fortunately, such rules are properly preempted under Section 332(c) of the Communications Act. Likewise, zoning rules that indirectly bar entry—running counter to the build-out requirements established by the FCC—also are properly preempted under the Communications Act. Since such unreasonable zoning and land use regulations that directly or indirectly bar entry are so clearly inconsistent with the consumer interest, the sole problem is ensuring that they are recognized as inconsistent with national telecommunications policy. Fundamentally, whether those issues are resolved at the federal or state level is immaterial.

Of course, the reality is that the national telecommunications policy—which puts competition to work for the consumer—must be recognized as paramount and must be implemented. Otherwise, the NII will simply be a grand illusion and never a real producer of jobs, education, or security.

CTIA has acted to make the NII a reality. Indeed, cellular companies have been building the wireless lanes and on- and off-ramps of the NII for over a decade. Now, when the means of making the information superhighway are debated, the role of wireless must be recognized—and the needs of consumers given full measure. CTIA has pressed for action to ensure that consumers have access to what they want, when they want it.

On October 27, 1994, Thomas E. Wheeler, president and chief executive officer of CTIA, wrote to FCC Chairman Reed Hundt, warning that

Ubiquitous wireless service—including wireless service to schools, libraries, and public safety agencies—requires ubiquitous cell sites. Too often the process of building the wireless network is

thwarted or delayed by "not in my backyard" activists seeking to block the construction of the national wireless pathway for parochial or ill-informed reasons.<sup>17</sup>

On December 22, 1994, CTIA filed a petition for rulemaking, requesting preemption of state and local regulation of tower siting for commercial mobile service providers<sup>18</sup>. CTIA also supports, through the Electromagnetic Energy Association, rules preempting state and local oversight of exposure to RF emissions from FCC-authorized transmitters<sup>19</sup>.

On March 22, 1995, Thomas E. Wheeler wrote to President Clinton, urging him to issue an Executive Order "directing federal agencies to expeditiously facilitate the location of wireless antennas on property over which they have control"<sup>20</sup>. Such an Executive Order would take a "giant step in reinventing government as a force to encourage innovation in competitive telecommunications services"—fostering hundreds of thousands of new jobs and billions of dollars in private capital investment, facilitating the deployment of the NII, and generating revenues for the Treasury by leasing access to federal properties<sup>21</sup>.

The states have been called the laboratories of democracy—but the interest of consumers, and the interest of the national economy, are not limited by state boundaries. As House Speaker Newt Gingrich has said,

We have to look seriously at those areas where the national economy requires preemption. The reason we went from the Articles of Confederation to the Constitution was to allow preemption where necessary. As a general rule, I want to decentralize decisions as much as I can, but clearly, for example, when you are in a cellular system you ought to be able to be in any cellular system in America and have it work. You cannot suddenly arrive in a dead space that has been created by a local politician for cronies who happen to own an obsolete investment.<sup>22</sup>

## NOTES

1. See "Gingrich Praises Entrepreneurship in Bringing Technology to Schools," *BNA Daily Report for Executives*, May 2, 1995, p. A84. See also "CTIA Foundation, Bell Atlantic Mobile, Cellular One Donate New Wireless Systems to Bring State-of-Art Communications to D.C. Public Schools," CTIA Foundation for Wireless Telecommunications press release, May 1, 1995. See also "Southwestern Bell's ClassLink<sup>SM</sup> Wireless Phone Concept Improves Life at a Texas School," Southwestern Bell Mobile Systems press release, February 1, 1995.
2. See "Reinventing Competition: The Wireless Paradigm and the Information Age," *CTIA Monograph Series*, 1995, p. 2.
3. Ibid.
4. Lexis/Nexis search of the "current news" database for references to the NII or information superhighway, May 9, 1995.
5. See, for example, Note [1] above. See also, "NYNEX Teams Up with Thirteen. WNET to Provide On-Line Anytime, Anywhere Math Education," *Business Wire*, January 10, 1995. Wireless carriers around the country are also contributing resources to support medical applications of wireless telecommunications. For example, CommNet Cellular Inc. has supported rural medical services in the form of airtime contributions to the Sioux Valley Hospital Outreach Program of Sioux Falls, South Dakota.
6. See letter dated December 23, 1994, from Randall S. Coleman, Vice President for Regulatory Policy and Law, CTIA, to Arati Prabhakar, Chair, Committee on Applications and Technology, NIST.
7. See "Evaluating PCS Markets," *PCS News*, January 20, 1994, p. 4. See also Roche, Robert F. 1994. "PCS Predictions and Prescriptions: Highlights from 32 Studies and Reports on the Prospects for PCS," filed in GEN Docket No. 90-314, April 13, p. 12.
8. See "Nextel Installs All-Digital Integrated Wireless Communications Network in Los Angeles," *RBOC Update*, September, 1994. See also Harmon, Amy. 1994. "Nextel Launches New Wireless Service in State," *Los Angeles Times*, September 23, p. D3.
9. See, for example, "US WEST Petition for Expedited Partial Reconsideration and for Clarification," filed December 8, 1993, in GEN Docket No. 90-314, pp. 7–12 (arguing for higher power limits to facilitate competition between cellular and PCS).

10. This assumes that PCS companies must deploy approximately three cell sites for every cellular cell site in order to achieve comparable coverage and that a hypothetical single nationwide cellular company would have deployed 7,500 cell sites to provide coverage.

11. Paul Kagan Associates. 1994. *Wireless Telecom Investor*, October 20, p. 9.

12. This assumes that cellular companies will roughly double the number of cell sites they had deployed as of midyear 1994 in order to improve coverage and capacity, and that Nextel will build out its 4,000 sites. The equivalent of four 30-MHz PCS systems would require a theoretical total of 90,000 cell sites ( $4 \times 22,500 = 90,000$ ). In fact, the number of antennas may be higher regardless of the number of service providers, simply to ensure adequate coverage and service quality.

13. See, for example, "Blairstown Township Zoning Board of Adjustment Resolution Memorializing the Denial of a Certain Use or 'Special Reasons' Variance Sought Pursuant to N.J.S.A. 40:55D-70(d)(1) to the Application of Pennsylvania Cellular Telephone Corporation Seeking Approval for the Erection of a Cellular Telephone Tower on Block 2003, Lot 14.01, on the Blairstown Township Tax Map Application ZB-2-94," pp. 23-24 (dated October 25, 1994; revised November 3, 1994).

14. See, for example, "Village of Wilmette Resolution 93-R-34." See also zoning ordinances of Jefferson County, Colorado, and the City of Stamford, Connecticut, which provide that more stringent state or county standards may supplant the 1992 ANSI standard. See Jefferson County Reg. Section 2, P(1)(a), and City of Stamford Ordinance No. 527 Supplemental.

15. See Ryser, Rob. 1994. "Tarrytown Extends Ban on Installation of New Cellular Antennas," *Gannett Suburban Newspapers*, December 6, p. 3A: "We have been surprised by the board's action from the beginning. The expert that Tarrytown hired to study (antenna transmissions) came back and found our cellular installation safe."

16. See Cody, Michael. 1995. "Bay Bridge Is Potential Antenna Site," *The Capital*, March 30, p. A1.

17. Letter dated October 27, 1994, from Thomas E. Wheeler, President and Chief Executive Officer, CTIA, to Reed Hundt, Chairman, FCC, p. 2.

18. See "CTIA Petition for Rulemaking," RM-8577, filed December 22, 1994.

19. See "Reply Comments of CTIA," RM-8577, filed March 6, 1995, p. 7 (referring to "Petition for Further Notice of Proposed Rulemaking in ET Docket No. 93-62, filed by EEA on December 22, 1994).

20. Letter dated March 22, 1995, from Thomas E. Wheeler, President and Chief Executive Officer, CTIA, to President William J. Clinton, p. 1.

21. Ibid.

22. Speech of House Speaker Newt Gingrich at *Wireless '95*, New Orleans, February 1, 1995.

## 55

# Electronic Integrated Product Development as Enabled by a Global Information Environment: A Requirement for Success in the Twenty-First Century

*Thomas C. Rochow, George E. Scarborough, and Frank David Utterback  
McDonnell Douglas Corporation*

This paper discusses the needs manufacturing organizations will face in the near future in working electronically and collaboratively to develop, market, and support products for the global marketplace. One of the necessary conditions for success is a rich and robust information environment. It is essential that this information environment be implemented to keep pace with the practical needs of the organizations that use it. Users and enablers must mutually support each other if the journey is to be made in time for the travelers to survive and prosper.

### STATEMENT OF THE PROBLEM

Manufacturing organizations today face a confusing and constantly changing global marketplace. Customers have fewer resources to procure products and services. They are less tolerant of performance below expectations, and their expectations are higher. Budgets are tight and getting tighter as economic constraints force industrial companies and customer organizations to become lean. Competition is fierce. Outsourcing requires an integration of services that is far more difficult than when the services resided within the same organization. Product delivery times must be reduced, with no loss in quality. Cost has joined performance as a primary discriminator in competition, and the cost time frame encompasses the entire life cycle of a product.

It is the rare company that can address the competitive needs of the marketplace entirely with its own resources. Many organizations are reverting to a simple and proven approach, with a modern twist that offers great promise but also brings great difficulties. The simple approach is generally referred to as concurrent engineering—getting the right people together, in a timely manner, giving them the right information and tools, and supporting them as they work. The modern twist is that the people, information, tools, and support elements are not usually cohesive or localized, and tend to change in composition over time.

The people are from different organizations, and seldom can more than a few be assembled in the same place. The time frames in which they must work have been greatly collapsed. The information they need to do the job is in various electronic and paper forms scattered in all directions. The tools are often software applications that should, but usually do not, play well together. The automation infrastructure necessary to tie all together is often reminiscent of a railway system spanning countries with different gauges of track.

The need is for integrated development of a product, executed globally and electronically, by what is referred to as a virtual enterprise. Industrial organizations are beginning to work this way, albeit in an embryonic or at best adolescent sense. But enough has been done to prove it is possible. The challenge is to make it practical. The compound question to be answered is this: What is required to emplace an environment that facilitates globally distributed teams—of people and organizations, customers and providers—to collaboratively develop products and support them worldwide throughout their useful life, at costs that are affordable, with performance that meets or exceeds expectations? It is the classic engineering problem of producing optimal results within practical constraints. In this case, these constraints are resources, technology, culture, and time.

This paper is written from the perspective of an aerospace manufacturer. Although some differences with other industries exist, the essence of the needs, problems, and solutions for electronic integrated product development should be similar across industry sectors.

## BACKGROUND

Aerospace is a high-technology industry that necessarily pushes the envelope of the possible. It has been said that aerospace therefore is not a good example of the manufacturing industry as a whole. If that were ever true, it is certainly not so today. Although aerospace manufacturing technologies are often state of the art, they must increasingly be applied within constraints of practicality; there must be a business pull. We must be innovative in how we work, do more with less, and concentrate on making an affordable product, not a masterpiece. These are common needs among manufacturing organizations today. Fortunately, aerospace can return to its roots for salvation, and other industries should be able to do the same.

### The Early Days

In the early decades of this century, airplane manufacturers assembled their resources in a hangar and loft. Design, manufacture, and assembly were carried out by the same group of people, who often had close ties with those who would operate and support the product. Experienced staff and funds were scarce. Innovation and teamwork were necessary conditions for success. The processes by which work was accomplished were identified implicitly by the workers, who were usually considerable stakeholders in the enterprise. Processes were reengineered on the fly, and all concerned knew why and with what results. This scenario can differ little from those in other industries in their early days of growth.

### Growth and Decline

What happened, of course, is that as airplanes grew more complex, and specifications from different customers began to diverge, specialties grew and companies reorganized to gain efficiencies. As markets expanded commercially and governments began to focus on aviation, performance often became more important than product cost; and product cost was considered that of acquisition only rather than including support through a number of years. Upstream innovation to help downstream functions was mostly lost. Information of a "total-product" nature was not available, as functions concentrated on what was needed from their narrow perspectives. A world war followed by the beginning of the Cold War put increasing emphasis on survival, and thus on product performance in increasingly complex missions. Cost retreated further as a major consideration.

The boundary conditions began to change in the last decade or so. As the capability of the threat increased, the likelihood of drawn-out conflict decreased. The ability to keep weapons systems available for action a larger percent of the time became increasingly important. As product complexity grew to meet the growing threat, cost regained its place as a prime factor in product selection, and a product's life-cycle aspect emerged. With the disappearance of the threat to survival in the last few years, coupled with the lethargy of the global economy and its effect on commercial aircraft sales, aerospace was suddenly presented with a completely new marketplace, with new rules and conditions, not yet stabilized. The question of survival began to apply not to the nation but rather to the company.

The resulting gyrations of downsizing and becoming lean to meet market realities are well known. Though perhaps in some cases the scope and degree are less, other industries are experiencing similar trauma. There are fatalities, mergers, and a return to a focus on base expertise. Increasingly organizations are analyzing where they need to go, from what basis; when they have to arrive to survive and prosper; and how they might proceed within practical constraints.

### Return to Basics

Aerospace has begun, and has actually made considerable progress. The experience of McDonnell Douglas Corporation (MDC) over the past several years is probably typical of many high-technology companies. We have recognized that affordability of products from a life-cycle perspective is a paramount consideration. This of course assumes that the products will meet or exceed customer expectations. We have realized that improved business processes are a most important element and have concentrated on the documentation, analysis, and improvement of those processes key to product development. Increasingly, management of our business is based on management of our processes.

We have also realized that we must return to the base values and practices of our industry. We have reorganized along the lines of concurrent engineering, MDC's term for this being integrated product development (IPD). This has been done in our functional support units as well as in our aircraft and missile programs. We have also returned to our base expertise, divesting ourselves of interests not closely aligned with the core elements of our products. We have outsourced much of our automation support and have aggressively struck teaming relationships with companies around the world, some of which continue to be our competitors in other acquisitions. Our IPD teams include customers, teaming partners, subcontractors, and suppliers. As we execute IPD, we have come to understand the prime importance of our information assets, and we are taking steps to analyze and evolve these assets—old and new—to make their provision to IPD teams adequate for the purposes of decisionmaking and task execution in an electronic IPD environment.

### Status and Challenge

MDC is pleased with its progress, but far from satisfied. The F/A-18E/F Hornet approaches first flight in December 1995 with the program on schedule and within budget, and with the aircraft meeting technical performance parameters and being below its weight specification. The program critical design review (CDR) found fewer than ten action items to be addressed—none critical. This compares with several hundred required actions found historically across the industry at CDR. The major credit for this performance is given to the collaborative IPD teams that are developing the aircraft with the highly visible, aggressive, and positive participation of an involved and enthusiastic Navy customer. There are no surprises on the F/A-18E/F.

As part of this initial success story, MDC has applied practical automation enablers in key areas. The aircraft is all digital. We execute an electronic development process that substitutes digital geometry models for the traditional physical mock-up. We define structural and systems interfaces with our major subcontractor, Northrop Grumman, and perform design reviews and resolve fit and function issues in real time, with data models shared electronically between St. Louis and Hawthorne, California. The product is defined in terms of assembly layout, build-to, buy-to, and support-to packages, containing all the information necessary for a qualified source to provide the part, assembly, system, or service. Discipline in the execution of the processes by which these packages are created, reviewed, approved, and released for manufacture or acquisition is applied by a control and release system that also ensures the integrity, completeness, and consistency of the data. Customers, co-contractors, subcontractors, and key suppliers are linked electronically.

Although pleased with this progress, we have also learned how far we must still go to be truly competitive in tomorrow's world. We must be able to execute electronic IPD as effectively and as efficiently as Charles Lindbergh and the folks at Ryan Airlines applied traditional concurrent engineering to the design and construction of *The Spirit of St. Louis*, in a hangar in San Diego during a critical 2 months in the spring of 1927. We must do this with teammates from organizations distributed around the world, including sophisticated giants and simple machine shops that nevertheless provide individual links that will set the strength of the chain. We must be able to collaborate, in real time, on unforeseen problems in short time frames. We must have immediate access to the information critical to the task at hand, be it a management corrective action or making an information package deliverable against a tight schedule. The information access must be controlled, the data exchanges must be secure, and the information provided must be correct, complete, consistent, and pertaining to the right configuration of the part or system.

This future state is necessary for our continued participation as a major player in our marketplace, just as it must be necessary to the success of every modern organization for which collaboration and information are essential to survival. And the time frame in which we must emplace the enabling environment is short. Certainly by the turn of the century, successful organizations will work in this way.

We should be optimistic rather than overwhelmed, since enough capability has been demonstrated to prove the concept. Yet the issues are substantial. Progress must also be substantial, immediate, and continuing if we are to arrive at our destination in time to remain competitive.

Not the least of the elements that enable electronic IPD is the provision of information in a cost-effective manner to diverse teammates around the world—information that resides in a vast heterogeneous "information space" itself distributed globally. However, all the elements are interrelated, and a certain level of understanding of them all is necessary to provide integrated solutions to specific needs.

### FORECAST

Common among successful manufacturing organizations will be their ability to assemble themselves in whatever virtual configurations are necessary to identify, scope, pursue, and capture business, and then perform to specifications within budget and schedule constraints. Business will be won and executed by virtual enterprises. These will not be static over the life of a product but will change with conditions and the ebb and flow of requirements and capabilities. Participants will include customer organizations, high-technology partners and subcontractors, and suppliers with various levels of capability and sophistication.

The mechanisms by which teams interface and operate must in large measure be common, as solutions specific to only one opportunity will be too costly for most players. Work processes will be executed in a collaborative manner, in real time where appropriate, by teammates distributed physically and in time. Collocation will be electronic, and information needed for the task at hand will be immediately accessible and usable with little or no preparation. Information will be provided through a variety of electronic media—text, graphics, voice, and video. Huge volumes of data will be stored, accessed, viewed, and moved. This will have to occur efficiently, accurately, securely, and affordably.

There is little question that this scenario must be realized in the immediate future for successful enterprises. The prosperity of a nation will be proportional to the success of its commercial and government entities as they participate in the interaction of the global scene. The obstacles to the emplacement and effective operation of this environment are significant, and the cost of solutions in time, labor, and money will be as well. It is essential that government and industry, while retaining the essence of fair competition that is a core American value, become partners in providing cost-effective common solutions to shared obstacles that prevent establishing the proverbial level playing field.

### ANALYSIS

Some of the elements necessary for success in the collaborative global marketplace that is upon us are discussed in the following sections. We identify some obstacles to their successful emplacement, suggest an approach for implementation, and discuss options for joint pursuit of common objectives.

#### Elements Necessary for Success

There is a close relationship between the collaborative processes by which we must do business and the policies, procedures, standards, and automation by which the necessary information is provided in a timely way to the team members executing (and evolving) the work processes. Processes and information are both necessary elements for success. Neither processes nor tools are sufficient by themselves. They must not only coexist but also actively support each other. Innovation in identifying better ways to work is often materially dependent on the availability of key information and the timely participation of dispersed teammates. The many opportunities

for innovation in providing essential information quickly, or electronically collocating key participants, are beyond all practical means of simultaneous pursuit. The opportunities need to be prioritized, scoped, and scheduled based on practical requirements for business success.

This suggests that the current emphasis on business process reengineering is appropriate and that it should proceed with a clear appreciation of the importance of providing information to process steps and players. It is hard to identify points of process integration, or key collaborations or decisions, in which information flow and content do not play a deciding role. Processes must change, but in a manner that is consistent with the practical ability of information technology to provide the right data to the right participants at the right time.

Information in electronic form is growing at a tremendous rate. This growth has now reached the point that it can easily drown a team as it attempts to find and use information related to its objectives. Being able to effectively locate and access pertinent information is important, but it is equally important to do so with the user in control, and protected from the onslaught of unwanted data and the uncertainty and inefficiencies of unstructured searches.

### **Automation Enablers**

There are a number of ways to categorize the automation enablers of information provision. We have chosen for our purposes to identify four aspects:

- Software that an individual user or team applies to create, analyze, or change data pertinent to a work step. This is often specific to a discipline, and we refer to such software as end-user applications.
- Software and hardware that provide controlled access with appropriate security to information and related software applications, by individuals according to assigned privileges.
- Software that manages information (i.e., that ensures its integrity, correctness, completeness, currency, and relevance to the specific product configuration or version being addressed by the user).
- The underlying elements, or infrastructure, that enable physical access to, retrieval or view of, or operations on the data permitted to the user. These elements include the hardware and software of networks; various hardware devices for storage and display of, and operation on, data; software such as operating systems, and software that supports desktop capabilities, file transfers, and electronic mail; and the various data and communication standards that facilitate data transfer.

All these enablers must function well and in an integrated sense if the information provision environment is to be effectively available to individuals and teams executing business processes. In the desired near-future state, they must also perform in much the same way for the variety of virtual enterprises in which individual organizations may find themselves.

### **Lessons from Experience**

To help identify specific requirements and obstacles, we now summarize MDC's practical experiences to date in our initial exercise of electronic IPD.

As mentioned previously, MDC has concentrated heavily on the electronic development process for the F/A-18E/F, collaborating with Northrop Grumman. This has given us valuable experience in the sharing and exchange of large amounts of data in various electronic forms. Policies, processes, procedures, automated tools, and infrastructure have evolved. Security and communications requirements have been identified and met. For a close and intensive interaction with a large sophisticated partner, we are comfortable with our ability to perform.

To extend interaction to our customers, and to subcontractors and suppliers with whom we interact more simply or less often, we have established a simple mechanism that we call the CITIS Node. It had its genesis several years ago in our new business area, as a proof of concept of the spirit of controlled access to data and applications suggested by the draft MIL-STD-974 for a contractor integrated technical information service



(CITIS). Based on a database of user profiles, within the constraints of a security firewall and the data access granularity of specific applications, we allowed execution of applications, and through them access to data. In some cases, this was access "in place"—that is, access at the source of the data store. In other cases, data were uploaded to the CITIS Node server and either accessed there or transferred to the requester's system.

No attempt was made initially to include the basic or optional functions identified in MIL-STD-974; this functionality was to be added instead as required for a CITIS under specific contracts. What we found as we demonstrated this proof of concept to various customers and suppliers was that there were immediate opportunities to use the CITIS Node, in all its simplicity, for exchange of production data. Suppliers were especially eager to do so. Over the space of several years, initially with individual suppliers to our new products area but increasingly with customers and suppliers in all our current programs, we have added users to the CITIS Node and have augmented its operation.

Today the CITIS Node supports over 750 users, with growth to 2,000 users expected by the end of 1995. A strategic alliance has been formed with a small disadvantaged business, Aerotech Services Group, for the Node's administration and support. Aerotech also provides hookup and help services to suppliers if they so desire. Over the next few years the CITIS Node will be expanded to accommodate the needs of all our programs and production centers, with additional functionality added to address the kinds of operations specified by both military and commercial perspectives of CITIS.

The significance of this story is not MDC's current capability. Other organizations today can present a similar status. The significance is our growing understanding of the obstacles encountered and the lessons we are learning, through the dual experiences of F/A-18E/F electronic collaboration and the extension of electronic IPD to a broad range of suppliers and customers through the CITIS Node.

Here are some our findings:

- A clear understanding of the work process is essential if it is to be reengineered to benefit from electronic IPD.
- Nothing is harder to facilitate than the change of a work process. This is a cultural change, often a change from what has been a successful process in the past. Human beings, perhaps as a result of a built-in predisposition to protect the species from the unknown, are strongly resistant to change.
- If a change is to be made, it can seldom be dictated. It must occur through producing clear evidence that it is beneficial to objectives in which the organization has a vested interest. This is best done through execution of the new process by real users doing production work, with proof of the benefits captured through accepted metrics.
- New procedures, techniques, and the application of automated tools must be straightforward and simple, and accompanied by sufficient training and on-site support.
- Production implementation to any reasonable extent must clearly pay for itself in the near term, usually within a year.
- Collaboration cannot succeed unless all participants have a compatible automation environment. This is difficult enough to obtain within a single organization. It is much more so when extended to customers and suppliers in various stages of automation implementation. Concepts of operation—who will do what, when, involving what data and with what basic software and hardware—for all participants are essential.
- Organizations can begin simply by exchanging digital data. However, to realize the maximum benefits from electronic IPD, real-time collaboration is necessary. This often entails viewing large amounts of data, for example in an electronic development fixture. Current costs to support this are prohibitive except in specific instances.
- Electronic collaboration requiring large volumes of data, such as electronic development of product geometry, requires broad-bandwidth digital services. Shared screen sessions are usually asymmetric, and a narrow back channel will suffice. When videoconferencing is used, the back channel must be symmetric.
- Current interoperability among networks leaves much to be desired, especially internationally. Connections must be initiated well in advance. Short-fused collaboration is next to impossible.
- The Internet is not viable for conducting the more demanding aspects of product development. It is unreliable, not secure, and cannot support large-volume data transfers with acceptable performance.

- Standards and supporting software are far from what they must become if data sharing and exchange are to be effective as enablers of the virtual enterprise. The processes used to define and validate standards do not primarily serve the needs of the user community. Standards too often do not support real needs and take far too long to mature; vendor-specific implementations are common.
- Suppliers, especially small and medium-size ones, can be overwhelmed by the need to interact with multiple contractors, as most will have to do to survive. They require help to identify the basic environment to put in place, as well as training, advice, and support at affordable cost.
- Few manufacturing organizations can afford to take uncertain risks as they move toward electronic IPD. Progress must be made with proven technology that can be extended as business conditions allow.
- Finally, cost structures of automation elements and services must accommodate the need of organizations to pay as they go. Unless business cases can be developed through production pilots to validate real benefits and the manner in which to capture them, organizations will be inhibited from moving forward at a reasonable pace.

### RECOMMENDATIONS

The need for U.S. organizations of all kinds and sizes to be able to participate in the collaborative global marketplace is obvious. Many of the obstacles are too large for specific solutions worked by individual players, or even by industry sectors or government agencies. For reasons of cost sharing and timely progress, U.S. industry and government must work together and must jointly cooperate with the rest of the world. The problems are global in nature. All concerned will profit if the solutions are also global.

The information environment must be defined and implemented with a clear understanding of its use, in the sense both of providing support for current production needs and providing for future growth by timely identification and validation of extended capabilities.

There has been much discussion of the information superhighway and of the associated visions of the defense information infrastructure, the national information infrastructure (NII), and the global information infrastructure (GII). These are well worth pursuing, but it is also useful to consider the example of the transportation system in defining a practical information environment, setting priorities, and supporting the implementation and integration of the various elements.

Besides superhighways, a national transportation system must have access and exit ramps, outer roads, local roads, and streets leading to neighborhoods of homes, schools, stores, and industrial parks. It must have laws, regulations, and tests to accommodate experienced drivers, beginners, and those with special needs. It must have stop lights, stop signs, and informational signs. It must provide maps of various levels of detail, and triptychs where appropriate. It must provide a measure of free advice and assistance, as well as affordable services for fuel, maintenance, repair, and other needs. And its users must be prepared for travel in similar but different environments in other countries around the world, with common regulations negotiated at the national level, as well as reciprocal acceptance of national standards.

It is beyond the scope of this paper to complete a detailed comparison of the information provision environment with the transportation environment we all know so well. One difference is worth noting: The buildup of the transportation system was mostly serial, expanding from a local base, and it occurred during a more leisurely time. The elements of the information environment must be addressed in parallel because of the pressure of fast-moving events, and no elements can be ignored, no matter how insignificant they may seem. All are necessary for its successful operation.

### A Collaborative Information Environment

Our primary recommendation is that industry and government together must define and describe an information environment to support electronic integrated product development—and most other kinds of business interactions—as carried out in virtual enterprises. They must then jointly implement and enhance this

environment over time, concentrating on addressing key aspects that can best be worked in common for the benefit of all, and taking special care to enable small and medium-size businesses to become active players.

This will require traditional planning and project management efforts. It will also require an innovative approach to gathering requirements, validating elements of the environment, and aggressively addressing common problems and issues in much shorter time frames than those experienced today. A proactive collaborative approach is necessary to quickly identify and emplace a basic information environment that can then be developed as aggressively as available resources, technology, and culture will allow.

There follow recommendations on specific obstacles, based on McDonnell Douglas's experience to date. The concluding section discusses mechanisms for a joint effort to define, emplace, and enhance the overall information environment.

### **Small and Medium-Size Businesses**

We need to offer small and medium-size businesses special help. This should include help with understanding the environment and how to get started and grow in capability at a practical rate. It should include advice and training opportunities. One aspect should be a free initial consultation; a second should be identification of a set of service providers who offer continuing assistance at a reasonable cost.

Several government-funded providers of such services exist. Two of these are the NIST/Manufacturing Technology Centers and the Electronic Commerce Resource Centers, funded by NIST. There are undoubtedly others. One essential task is to identify such existing entities and recommend how they might best cooperate to meet the need and minimize redundancy.

### **Standards**

Standards bodies need to be driven by users and to acquire a sense of the urgency of the real world. There should be effective mechanisms to collect and channel real requirements and to facilitate rapid prototyping and accelerated validation of draft standards. Users need to take back the initiative from the standards bodies and the vendor community. Standards are also information products, and they will benefit from collaborative development.

### **Affordable and Timely Implementation**

The investment in new automation to execute electronic IPD is daunting for large organizations. There is a legacy of software and hardware that cannot easily be replaced without hard evidence of reasonable return on investment in a short time frame. In addition, however inefficiently they may be supported with current processes and tools, production work and new business efforts are ongoing and cannot be put at risk during the transition to a better world. Often there is a catch-22. For example, Department of Defense production contracts include no funds for government upgrades in processes and tools to complement contractor internal upgrades. Business cases to obtain funding usually require verified benefits in a production environment. Automation vendors are often reluctant to support contractors and their customers in such pilots, merely in the hope of eventual broad production use.

This does not mean progress is not made, but it is slow and painful to line up the participating parties and their resources. The completion of desired connections through existing physical paths can also be exceedingly slow, taking a minimum of 60 to 90 days, often much longer. This inhibits aggressive expansion and prevents the flexibility needed to take advantage of short-fused opportunities.

There is a need for an innovative sense of joint partnership to define and validate new processes and tools quickly, with shared risk, among the players. Ways must be found to emplace the overall environment in an affordable manner and to pay for implementation, enhancement, and ongoing support in a shared way from captured business benefits.

### **Access and Security**

Electronic collaboration requires ready access to information. But of equal importance is limiting access to and operations on information in accordance with assigned privileges—not only for reasons of commercial or national security, but also to ensure the integrity of the data against malicious or inadvertent alteration. Access and security make opposing demands on an information environment, and we must find ways to strike a practical balance. Understanding the requirements of electronic collaboration in support of production work, and collaboratively addressing them, is the preferred approach.

The logical course is to move to the Open Software Foundation Distributed Computing Environment, supplemented with security products based on public-key encryption such as RSA (named after its authors, Rivest, Shamir, and Adelman). Robust firewall systems must provide security for controlled entry into an organization's data environment, with systems such as those provided by product data management vendors to further limit access to, and operations on, specific data.

### **Self-Diagnosis and Prescription**

We are entering a complex and fast-changing world. All the involved organizations, perhaps especially the smaller players, should continually evaluate where they are, where they need to be, and how to get there in a reasonable manner. There is a need for a practical way to make this evaluation and to plan corrective action and future growth. A capability model is required that can be self-applied. It should be as simple as possible and still give the needed answers. It should support the growth of capability in a series of finite steps from initial simple participation to attainment of the full sophistication of a virtual enterprise, with easily understood entrance and exit criteria for each step.

Two related and complementary efforts along these lines come to mind. The first is a CALS Enterprise Integration Capability model under development by the EDI World Institute in Montreal, under the leadership of Susan Gillies. The second is a CALS Electronic Commerce Continuum model proposed by Mike McGrath at the Defense Advanced Research Projects Agency. The CALS Industry Steering Group is currently serving as a forum to support and relate both efforts.

### **An Information Environment Atlas**

Industry and government need an atlas of the near-term global information environment that identifies and explains the elements and their relationships, and how they can be used to support electronic integrated product development. This should be a natural by-product of a plan to emplace the information environment. It should identify and describe the elements and their interrelations. It should give guidance for connecting into the environment initially and for growing in ability to use its sophisticated capabilities. It should identify where to go for advice and help, from both government and commercial sources. It should include models and techniques for self-assessment of status and for determining readiness to move to a higher level of capability. It should include references for more detailed information. Finally, it should describe the forums and mechanisms by which the information environment is supported and advanced, and how individuals and organizations can become active in the forums and participate through the mechanisms.

### **Forums and Mechanisms**

We need forums and mechanisms to define and execute the plan to emplace and support the information environment we have been discussing. This includes defining practical scenarios of how we must work and from which we can draw requirements for policies, processes, and automation enablers valid across industry sectors. We then need to validate these elements in support of real production work and capture the metrics and lessons learned with which to justify broad implementation.

It is difficult for individual industries or government agencies to address this need. We should instead have a means to assemble the appropriate representatives and do the job for all. This is not to say that individual elements should not be addressed by the most appropriate organizations. But the description of the environment, the generation of an overall work plan to achieve and sustain it, and the monitoring of its implementation and enhancement should be done by an entity that is representative of all the players.

Two ways come to mind to accomplish this. One is to convene an ad hoc team, and the other is to use an existing organization of a multiplayer nature. In either case, a description of the environment and a work plan should be generated quickly. The work plan should be of sufficient detail to allow selection of the best means to achieve each major deliverable.

An existing organization that would be a viable candidate is the CALS Industry Steering Group (ISG). It is chaired by USAF Lt Gen (retired) Jim Abrahamson and has members from many industries as well as close liaison with government agencies. It has recently restructured to emphasize the requirements of both the commercial sector and the government, and seeks to enable the reengineering of business processes through the effective and efficient sharing and exchange of data electronically. It has begun to negotiate agreements with organizations having similar interests to support a focus on common problems and to avoid redundant and perhaps competing uses of resources.

One advantage an existing organization has is that it would be positioned by its charter and interests to act in a coordinating, monitoring, and steering role, as well as offering a mechanism for actually executing parts of the plan. It could collect common requirements and issues and act as an "honest broker" in facilitating common solutions. It could also coordinate production validation of solutions through pilots in its member organizations, as well as support prototyping of new technologies. It is user driven, which is essential if the new environment so facilitated is to be practical and responsive.

If an organization such as the CALS ISG were chosen as a prime mechanism, it would be important that full-time staff be provided to complement the mostly volunteer resources that currently support its efforts. There is no substitute for the enthusiasm and ability of the dedicated and talented volunteer, but the realities of the business world necessarily limit the time and energy these individuals can provide.

## 56

# Interoperability, Standards, and Security: Will the NII be Based on Market Principles?

*Quincy Rodgers*<sup>1</sup>

*General Instrument Corporation*<sup>2</sup>

### BACKGROUND

The United States stands on the verge of a quantum leap in the capabilities of its national information infrastructure (NII)<sup>3</sup>. Widespread use of fiber optics is one element of that leap, as is the incredible advance in microprocessing, which has supported breakthroughs in digital compression. These, when applied to video, have made it possible to dramatically increase available communications bandwidth.

Just 5 short years ago, all-digital video was seen as impossible before 2005, if then<sup>4</sup>. The amount of data or information required for full-motion video is orders of magnitude greater than that required for voice or nonvideo data transmissions. The data for video demands bandwidth beyond that required for voice and nonvideo data and requires effective compression techniques, to use the bandwidth efficiently or to meet some existing bandwidth limitation. In June 1990, General Instrument Corporation announced that it had solved this problem and would present an all-digital, high-definition television (HDTV) broadcast transmission system for consideration by the Federal Communications Commission and its Advisory Committee for a U.S. advanced television standard. One writer called this the technology equivalent of the "fall of the Berlin Wall"<sup>5</sup>.

The conversion of video into a usable digital data stream is the third leg of the multimedia stool. The other two legs are data and voice telephony, the first being inherently digital, the latter having undergone a transition from analog to digital over many years. Carriage of video, voice, and data constitutes the foundation of the advanced broadband networks now being developed for both telephone and cable television networks.

Behind this simple concept, however, is a roiling pot of players and special interests, each trying to invade the other's turf or at least protect its own. Convergence is real because digitization breaks down the old barriers between service providers<sup>6</sup>. It is increasingly hard to tell computer companies from video-equipment providers. It becomes increasingly more difficult to tell cable television operators from local exchange telephone companies. The growing community of Internet users adds to this mix, some of whom have an almost religious fervor about its uses and potential.

All of these players approach the information superhighway differently. Their perspectives and biases, at least initially, have been parochial and have caused them to overlook matters and issues that are important to other groups. The well-worn parable of the blind men and the elephant is apt here. The NII presents different characteristics and different issues, depending on how it is approached. When these converging players sit down with each other, they often find that they not only think or see things differently, they even tend to use a different vocabulary. What one set of players views as communication is mere babble to another.

Put another way, the convergence of technology is proceeding well in advance of the "cultural" convergence of the various segments of the industries. The resulting different perspectives influence the public policy debates that surround the NII. This paper addresses some of the issues of those debates, based on the following assertions:

- The NII is not synonymous with the Internet;
- Open systems do not have to be, and perhaps should not be, nonproprietary—interoperability is not the only or even the major public good that must be served;

- The private sector, not the government, should set standards; de jure standards, whether set by government or the private sector, are not preferable to de facto standards; and
- Unless security and privacy are protected, the NII will not reach its full potential as a platform for electronic commerce.

These assertions may be controversial; many run counter to conventional wisdom, particularly that of the Washington, D.C., community. However, all are market-oriented. And they are fundamental to U.S. economic goals.

### THE NII: NOT SYNONYMOUS WITH THE INTERNET

Entertainment is the engine that has already pulled broadband networks into over 60 percent of American homes. Entertainment will drive the investment necessary to upgrade those networks as well. This is not a fact easily accepted by some Internet aficionados and some in the computer industry. Many in those groups join millions of other Americans who have a low opinion of much of current television programming. Many view the Internet as a liberator from TV's vast wasteland. But there are also other reasons why the role of entertainment is not more widely acknowledged by the computer world<sup>7</sup>.

Computer businessmen are aware that today's dominant entertainment terminal, the television set, is an extremely cost-sensitive (and low-cost), relatively simple piece of electronics, geared to nonbusiness consumers, with a life of over 10 years. This is a long way from their preferred business model. The consumer electronics industry is dominated by foreign-owned companies; the computer industry is U.S. based. The television industry has relied on interlace scanning and sees it as important to keeping down the cost of its investment<sup>8</sup>; the computer industry wants progressive scanning formats.

Likewise, those in the computer business look at other players and find them to be quite different from themselves. The cable television industry has only recently emerged from its "pioneer" phase and still revels in its "cowboy" image<sup>9</sup>. Cable operators have an obsession with cost control, based on experience with mass consumer marketing. Where the computer industry has traditionally sold into the business community, cable operators have focused on residential customers. Regional Bell operating companies (RBOCs), in contrast to both, have previously been large, sleepy, and cumbersome bureaucracies, typical of utilities, and the antithesis of the computer industry<sup>10</sup>.

The computer and Internet communities are not alone in their suspicion of the alien cultures that have suddenly entered their world. About 15 months ago, a high-level business manager<sup>11</sup> for a leading supplier to the cable television industry told his staff, "Somebody is going to have to explain this Internet to me!" He was reflecting not only his lack of knowledge of the phenomenon but also his exasperation at the whole range of new factors he had to consider as he developed his current core businesses.

So it is not surprising that computer business managers tend to gravitate toward an Internet model of the future NII. They will naturally have less enthusiasm for the role of entertainment than those who have worked in that field<sup>12</sup>. But to recognize entertainment as the primary engine of the deployment of advanced broadband networks does not denigrate the role or the importance of the Internet. The NII is and should be about a lot more than just selling pay-per-view movies or making it possible for people to watch reruns of *Roseanne* or *Baywatch*. The advanced broadband networks that will share the task of serving as the backbone of the NII are about making video an integral part of all communications. The addition of video capability has major, positive implications for education, health, and business efficiency.

What entertainment can do is bring this broadband capability to every home and business. It can and will carry the major load of the investment needed to do that. When cable television operators begin to deploy digital decompression terminals, they will be putting into each user's home a level of computing power that is the equivalent of yesterday's mainframes. Far from detracting from or conflicting with the Internet, the broadband pipes of these networks will make Internet access via high-speed connections available to an increasingly wider range of Americans<sup>13</sup>.

## ACHIEVING OPEN SYSTEMS

Open systems do not have to be, and perhaps should not be, nonproprietary; interoperability is not the only or even the major public good that must be served. For 2 years, the rallying cry has been, "We must take steps to ensure that the new technologies result in 'open and interoperable' networks." The General Accounting Office has proclaimed that "interoperability will be difficult to achieve"<sup>14</sup>. Members of Congress have echoed this belief<sup>15</sup>.

But is the situation really dire? Is there really a problem here? Is government intervention required? Thankfully, the conclusion is "no," because the cure could easily be worse than the disease.

Of course, interoperability is actually quite simple to achieve. All you have to do is standardize all elements of advanced broadband networks and the equipment that will use those networks. This can resolve all issues of compatibility, portability, and interoperability. Requiring compulsory licensing of any intellectual property that underlies those standards then provides unlimited competition in the supply of equipment for the networks. It is that simple, at least in concept. But even laying aside the political and legal barriers to such a program, achieving interoperability through central planning is a daunting, even overwhelming, technical task.

The view that interoperability is a major problem and that it might require such stultifying and intrusive steps rests on an outmoded view of communications capacity. In an environment of limited bandwidth, legitimate questions have been posed about the allocation of that scarce resource. When a cable television system carries 36 channels, issues about who gets "shelf space" in that system can take on major importance. Broadcasters demand "must-carry rules," because they are worried about retaining access to viewers. Programmers express concern that their products will receive less favorable treatment than those developed by the operator on whose system they seek to be carried. New services are concerned that they must go through an established multiple system operator to obtain the critical mass of viewers necessary for viability.

In an environment of bandwidth abundance<sup>16</sup>, the incentives of network providers look quite different than they do in an environment of bandwidth scarcity. This is particularly true where the network provider faces competition. To pay for the network, providers have an incentive to include as many products and services as possible. They also have an incentive to make interconnection as easy as possible, and to develop consumer-friendly approaches to equipment design. For some purposes, network operators will continue to function in a broadcast mode. As channel capacity dramatically increases and fewer homes are served per node, each home or business served has the equivalent of one or more dedicated 6-MHz channels<sup>17</sup>; this service, which resembles carriage more than broadcast, can be expected to develop as a result of consumer demand.

The incentives created by competition and broadband abundance will include incentives to achieve interoperability. Moreover, interfaces on these networks will remain open and do not have to be nonproprietary for this to occur. Network providers will avail themselves of the benefits of proprietary technologies but will, in their own interest, require that they be open to ensure competitive markets for equipment<sup>18</sup>.

Just as the technology, by altering incentives, makes interoperability more likely or more easily achieved, so also is it inherently more conducive to interoperable solutions. Problems that existed in an analog environment are more easily solved in a digital environment.

During the early days of NII debate, Vice President Al Gore used a metaphor that was intended to justify a government role in setting ground rules for the NII. The vice president referred to the need for setting standards for interfaces on the NII and drew an analogy to the standardization of the gauges of connecting railroads<sup>19</sup>. Analysis of this metaphor suggests a major distinction that tells us something important about the digital environment.

Railroad gauges, in fact, provide the perfect illustration of the digital environment's inherent ability to provide new answers to old problems. The chief difficulty created by the use of different gauges by connecting railroads was that, to move from one to the other, freight and passengers had to be transferred from one car to another car suitable for the different gauge. People and goods moving in interstate commerce could not be efficiently transported under such circumstances<sup>20</sup>.

However, because of the essential nature of digital communications, data can be off-loaded, reloaded, and moved from one type of network or environment to another with comparative ease. Indeed, this flexibility is fundamental to the digital revolution, the foundation of convergence, and the basis of multimedia. In this



environment, the question of standards can often be replaced by another set of questions that relate not so much to the difficulty but rather to the cost of conversion. That places the matter in the hands of the market, not of regulators.

Moreover, in the digital world, breakthroughs continue. When addressing the issue of selecting the format for its high-definition television system, the Grand Alliance<sup>21</sup> discussed whether it should employ an interlace scanning format or a progressive scanning format. Upon investigation, it concluded that receivers could easily and economically convert the signals from one to the other. The result was that the system will support six formats and provide both interlace and progressive scanning.

A similar experience is occurring in the case of the modulation schemes used by different networks. Satellite and cable systems use different modulation. The Grand Alliance system currently in testing would use yet a third method for broadcast television. The development of multiple systems is not surprising, given the fact that network operators want to maximize their networks for their specific needs and use. There are actually good and sound reasons to tolerate the lack of interoperability that this could create. But there, it appears, potential issues may evaporate because CableLabs has indicated that it thinks dual or multiple modulation chips are feasible and economical<sup>22</sup>. Thus, because of the inherent characteristics of digital technology, there is reasonable promise of avoiding inconsistency among networks.

Where interoperability is being addressed in the marketplace as a result of competition<sup>23</sup>, incentives, and the technology's ability to support conversion, it avoids negative costs of standardization. That is indeed a good thing because standards are a mixed blessing.

Standards can and do inhibit innovation. The problem was outlined in a recent speech by GI's CEO, Dan Akerson:

The lesson of the personal computer success story is that the government should not prescribe technological standards in dynamic industries. Such standards freeze the current level of technology in place and they stifle the development of new technologies. When the government lets the marketplace operate, innovators innovate, competition flourishes, and consumers' choices increase. And, finally, equipment prices plummet. When the time is right, the technology will mature and the market will set standards and insist on interoperability, the need for competitive pricing, and the availability of compatible equipment.<sup>24</sup>

Where standards cannot be avoided, they should be narrowly focused. For that reason, interface standards are preferable to standardization of functionality, such as compression, transmission, and transport. In effect, interface standards at least restrict the "zone" where innovation can be stifled and provide some opportunity for innovation to develop around them<sup>25</sup>.

There is yet another element to standardization that should give pause to those public officials who would apply it to these emerging industries. Standardization tends to commodize a product or technology. In their book *Computer Wars*, an insightful study of IBM and business strategies in the computer industry, Charles H. Ferguson and Charles R. Morris describe the effect of standardization:

Japanese companies, and Asian companies generally, have succeeded, by and large, by being superb *commodity implementers* within well-defined, stable, open, nonproprietary standards—standards, that is, that are defined by regulatory agencies, other government bodies, industry standards-setting organizations, or very slow-moving incumbents, such as IBM has been in mainframes in recent years. Nonproprietary standard products, such as memory chips, printers, VCRs, CD players, or facsimile machines, are brutally competitive businesses, with high investment requirements and razor-thin margins.

But industries that are fast-moving, where standards are constantly evolving, and where the standards themselves are within the proprietary control of an individual company, are hostile environments for commodity implementers. And the computer industry in the 1990s, under the technological impetus and creative impulse of the American start-ups, has been transmuting into just such an industry, shifting the ground out from under *both* the slow-moving Western giants and the commodity manufacturing-oriented Japanese giants.<sup>26</sup>...

The race over the next decade will be between the Japanese and American styles of industry development. The Japanese will drive to componentize and commoditize every sector of industry so their great monolithic and lean manufacturing skills can define the industry's future. American companies must keep accelerating the pace of technological change to avoid giving the Japanese a stable target to shoot at, and at the same time develop their own manufacturing skills to the point where low-end market share is not conceded too easily.<sup>27</sup>

So the problem is not merely the stifling of innovation. Rather, it is that the very industries of critical importance to the U.S. economy are most susceptible to the negative effects of standardization. The digital revolution is uniquely the product of U.S. research and development. It represents a national asset, no less because it arrived on the heels of a period of "technology pessimism" in the late 1980s. American leadership was widely viewed as losing technology leadership for U.S. industries. Frequently, the problem was seen not as a failure of American invention but as a failure to capitalize on that invention. Many saw government-led industrial policy as the answer to regaining such preeminence.

Commoditizing the new technologies of the digital revolution will demonstrate that we have learned nothing from that earlier debate.

### WHO SHOULD SET THE STANDARDS?

To the extent that standards are used, there is an issue about how they should be established. Here again, among the converging industries, differences in experience and practice create different perspectives. The broadcasting industry, for instance, lives in a world of government standardization, arising out of the practical problems created by potential interference among users of the radio spectrum<sup>28</sup>. In addition, standards and regulations have pushed beyond this limited rationale to govern some features and functions of receivers. To an extent still being debated at the FCC, Congress has applied this form of standardization to the cable television industry as part of the cable television rate regulation legislation of 1992<sup>29</sup>.

Similarly, the telephone industry is no stranger to standardization. In a unified Bell system, standardization was achieved within the company. But it has been the history of the telephone industry not only to live with standards but also to live with and even thrive under government regulation. Tell a telephone company executive that the government is contemplating standardizing some new product and the first question he asks is whether that isn't already regulated in his business; if it is, he dismisses it as a problem and moves on to the next issue.

The computer industry traditionally takes a different view of both standards and regulation than do the broadcast or telephone industries. Appropriately, it devotes significant energy and resources to remaining free of regulatory schemes and government-imposed requirements. There are industry standards, but they rise and fall very quickly and are constantly subject to attack and replacement. In the words of one broadcast official and part-time humorist: "The computer industry loves standards; it has a million of them."

The result of these differing attitudes, springing from different cultures and experiences, is a general policy approach that says the government should stay its hand and take a minimalist attitude toward standards and regulation, intervening only where there are overriding public interest concerns. However, this merely removes the debate to whether a given standard, regulation, or other proposal falls on one or the other side of that line.

The vagueness of this trigger for intervention and the fact that different biases are brought to its application by different groups of government officials invites politicizing of technology decisions. Thus, these issues become part of the political bazaar of Washington, a place ill-suited to making difficult technical choices. The private sector, not the government, should set standards; *de jure* standards, whether set by government or the private sector, are not preferable to *de facto* standards.

The computer industry is wise to oppose government-mandated standards<sup>30</sup>. Even where congressional and regulatory officials can clearly formulate and enunciate their policy goals, their ability to penetrate to the next level of analysis and understand the technical ramifications of a given set of choices is severely limited. Government intervention is rich in the possibility for mistakes and for the operation of the "law of unintended consequences." And government almost always seeks compromise and consensus<sup>31</sup>.

In recognition of these weaknesses in government intervention, many in the private sector, and some in government itself, have been advocating the position that these issues should not be addressed by government but should be left to private, standards-setting organizations. Decisions made in this fashion are preferable to decisions made by government officials. But there are good and valid reasons to ask whether substituting industry bureaucrats for government bureaucrats is, in the long run, a major improvement. There are valid reasons to ask whether this is a procedure that will yield the best results for the U.S. economy.

One problem is in the multiplicity of groups vying for the privilege of writing these standards. In the wake of the hype that accompanied the focus on an information superhighway, there are no fewer than 49 organizations or associations<sup>32</sup> claiming to have a role in the development of standards for digital compression or for the "set-top box"<sup>33</sup>. Some of these groups are established standards organizations. Others are entrepreneurial in the Washington, D.C., sense: find an issue, organize around it, raise some money, and exercise influence. Some groups are national. Others are international. And in some you cannot be sure whether participants represent a U.S. or a foreign interest.

These different groups represent different perspectives, depending upon whether their members come exclusively or in part from the broadcast industry, the cable television industry, the consumer electronics industry, the telephone industry, the programming and production community, or the computer industry. This level of activity is one indicator of the reality of convergence; they are all converging on the "set-top box."

Another problem arises because private standards-setting efforts are as susceptible as is the government to attempts to use the process for some narrow end or to benefit some special interest. After all, losers love standards. The techniques vary but the result can be the same. Bureaucratic standards setting is slow moving, Ferguson and Morris complained<sup>34</sup>. Pending standards can freeze the market and delay investment. Parties that are behind in the development process can use this delay to catch up to the industry leader. Frequently, those behind are numerous; the leader is, by definition, alone.

Like government, private standards-setting organizations generally rely on consensus<sup>35</sup>. This kind of compromise yields a lowest-common-denominator approach<sup>36</sup>. The fact that favoring consensus and not innovation can harm the economy is difficult for such groups and for the political process to comprehend; the effects are felt over time and diffused across the economy in general.

But the major problem arises because standards setting by bureaucratic bodies, whether public or private, is the antithesis of entrepreneurial development<sup>37</sup>. The entrepreneur invents something, innovates. The product may or may not be technical. Then the entrepreneur "undertakes"<sup>38</sup> to bring together all of the things needed to take a product or service to market: financing, distribution, and so on. The entrepreneur's idea is then accepted or rejected in the marketplace.

By contrast, bureaucratic standards setting invents nothing. Its conclusions are never really tested in the marketplace (although, if they are too far afield, they might be rejected outright).

These standards-setting activities are a drain on resources for any company that participates<sup>39</sup>. The smaller the company, the bigger the drain. Such activities require the diversion of valuable and scarce engineering resources. Thus, there is an inverse relationship between the size of the company and its ability to stay involved and protect its interests. Such private standards setting can work to the detriment of the entrepreneur to the extent that it prevents or limits his or her ability to obtain the fruits of the innovation. It can delay or even prevent the deployment of technologies<sup>40</sup>. And, in addition to these factors, the entrepreneur can frequently end up outgunned and outmaneuvered. If he or she is the inventor, the other players seek to benefit from the invention.

In this environment, the entrepreneur gets little or no help from his government, which tends in any dispute to favor the side with more and larger players<sup>41</sup>. What is more, government worries about de facto standards because they are messy: People might complain.

The most common argument against de facto standards is that "... we do not want to have another Beta/VHS situation!" How one feels about the Beta/VHS competition seems to be a matter of perspective, akin to the way that some people look at a glass of water. The "half-empty" crowd looks at Beta/VHS and calls it bad because it stunted investment<sup>42</sup>. But the "half-full" crowd looks at the competition between Beta and VHS and concludes that it was a good thing because it resulted in improved VHS technology and ultimately provided consumers with more choices.

Rather than engaging in government standards-setting activities or in bureaucratic standards setting activities<sup>43</sup>, what is needed is a new willingness (or a renewed willingness) to permit the market to work, to permit people to strike out and establish de facto standards, and to endure the uncertainty and even the temporary incompatibility that this sometimes creates. In the words of Joseph Shumpeter, "Let the gale of creative destruction blow through this industry."

**UNLESS SECURITY AND PRIVACY ARE PROTECTED, THE NII WILL NOT REACH ITS FULL POTENTIAL AS A PLATFORM FOR ELECTRONIC COMMERCE**

The deployment of advanced broadband systems will depend upon private investment.<sup>44</sup> The success of these systems will depend upon finding applications that will support that investment.

Deployment requires two conditions:

- A robustly competitive market for digitally based products and services, with strong economic benefits to those that develop the intellectual property that manifests itself in such products; and
- Constant innovations in technology that facilitate this competition—innovations fueled by similar market forces.

These are the forces that drive a system of dynamic competition in this rapidly evolving digital world. These forces need to be encouraged and protected. This can occur only if there is (1) a recognition that security for protecting intellectual property is a critical element, and (2) a recognition that security systems must be renewed, or evolve, through continued technological innovation if they are to keep pace with those who would seek to violate property rights.

Based on experience over many years, there are some fundamental facts and principles that should drive all policy:

- No matter how good the security system is, it will eventually be penetrated if the value of the material being protected is great enough. For this reason, security must be renewable. The fact that security is renewable is itself a disincentive to attempts at signal theft;
- For security to be renewable, government policy must not hamper innovation in the development of new responses to security breaches and in the development of new forms and methods of security;
- A single, national, uniform security standard, which is frequently advocated under one guise or another, is a dangerous idea. Not only does it provide attackers with a single target with enormous return, but it also would stifle the innovation necessary for security to stay ahead of attackers. A single, national, uniform security standard should not be advocated, advanced, or supported by the government;
- Published ("open") standards for security systems tend to weaken rather than strengthen security. Thus, unbundling and open-interface requirements, where they are employed, should be limited to functions that pose no threat to the intellectual property of programmers;
- Security functions should be placed in the hands of those who have an incentive to protect intellectual property. Proposals either to permit or to mandate their placement elsewhere should be resisted; and
- Although software-based security may be adequate for some applications, hardware-based security may be needed for others.

Security for and protection of intellectual property are fundamental needs for the success of advanced broadband networks, but no less important is the ability of those networks to protect personal privacy—that is, the privacy of personal information. If advanced networks are to become "platforms for electronic commerce," they will need to support a wide variety of applications, including delivery of electronic products, home shopping home banking, and medical assistance. The willingness of citizens and consumers to employ these networks for these purposes will depend upon their sense of confidence that their personal information is secure from prying eyes.

Proposals to require that equipment on these advanced broadband networks (and current cable and telephone networks, as well) be made available at the retail level<sup>45</sup> are an example of the way that regulating one aspect of a network can distort the operation of all others. In this case, the negative implications are for security. Consumer electronics retailers would, not surprisingly, make retail availability of equipment the central tenet of all policy. Naturally, proponents of this point of view ignore other problems that occur when this happens, such as freezing technologies or interfaces or creating problems for maintaining security. They would turn technology decisions into political decisions.

These proposals to mandate retail sale of consumer equipment used with the network are supported by dubious analogies and assertions. One such is the attempt to apply to video distribution and developing broadband networks the model currently used for voice telephone service, whereby an interface specification was developed and used to separate network functions from the functions performed by "consumer premises equipment"<sup>46</sup>. The telephone model is viewed as a way to create competitive markets for equipment used to supply broadband services<sup>47</sup>.

A major problem is that these proposals ignore the fact that voice telephone service requires only access control; video providers have traditionally (and for good reason, given the problems of theft of service) insisted on scrambling as well. Nevertheless, there are indications that the markets will develop such separation, and probably more quickly, without government interference<sup>48</sup>. Another claim made by proponents of this type of government intervention is that new "smart-card" technologies are available. They cite draft (not even finalized) Electronics Industry Association standards for such security systems<sup>49</sup> and seem oblivious to the contentiousness regarding the adequacy of such standards and what features are needed to bring them up to acceptable levels of security<sup>50</sup>.

The principle that underlies these proposals is that the government can and should determine how much security is adequate and that that determination will bind owners of intellectual property and network providers. That principle is unacceptable<sup>51</sup>.

Owners of intellectual property and those acting for them should have the right to determine the level and method of security appropriate for their needs. These decisions should be made on the basis of business needs and realities, not by government.

As network providers experiment with different types of broadband networks<sup>52</sup>, some of which have different implications for security requirements, the government should let these technologies work themselves out in the marketplace, free from the distortions that could be created when the government picks out one factor<sup>53</sup> among the many that must be considered, thereby affecting all other choices<sup>54</sup>.

## CONCLUSION

The assertions set forth here have in common a faith in and a reliance on market forces. The role of entertainment in driving deployment of advanced broadband networks and the need for security for those networks are related. Given adequate government recognition of the rights of property, and given government restraint from imposing other regulations that undermine those rights, these networks can develop. This will not happen instantaneously but will proceed incrementally over the next decade. The end result will be the extension of new broadband capabilities to all parts of the country.

Interoperability issues would also benefit from an understanding of the way markets are working to address problems far more rapidly than government or even private-sector standards organizations can accomplish. Market resolution recognizes the importance of incentives that motivate entrepreneurs. Failure to understand these markets and these incentives can have negative consequences for the economy and for international competitiveness. Conversely, if the government refrains from micromanagement and can successfully provide sound macroeconomic policy, the revolution that is the information age could bring a period of progress on the scale of the industrial revolution.

## NOTES

1. Quincy Rodgers manages government affairs for General Instrument Corporation (GI). A graduate of Yale College and the Yale Law School, he served as Executive Director, Domestic Council Committee on the Right of Privacy during the Ford Administration. He has also practiced law in New York City and Washington, D.C., and has served on the staff of the Senate Judiciary Committee.
2. GI is a world leader in broadband transmission, distribution, and access control technologies for cable, satellite, and terrestrial broadcasting applications. In 1990, GI was the first to propose and, in 1991, the first to deliver an all-digital HDTV system for testing in accordance with FCC procedures. GI has delivered all-digital compression systems for broadcast, satellite, and cable systems.
3. The national information infrastructure has been recognized as a separate subject for policy attention for at least 20 years. See *National Information Policy, Report to the President of the United States*, submitted by the Staff of the Domestic Council Committee on the Right of Privacy, Honorable Nelson A. Rockefeller, Chairman, National Commission on Libraries and Information Science, Washington, D.C., 1976.
4. This was the conclusion of at least one respected and generally knowledgeable business leader in the entertainment/communications industry. Japanese and European HDTV systems were apparently based on this false premise. The Japanese analog MUSE HDTV system was outmoded even as it was launched. The implications of this experience for government-led "industrial policy" are beyond the scope of this paper. See Cynthia A. Beltz, *High Tech Maneuvers, Industrial Policy Lessons of HDTV*, AEI Press, Washington, D.C., 1991.
5. Mike Mills, "A Digital Breakthrough," *Congressional Quarterly*, January 15, 1994, p. 66.
6. Computer Science and Telecommunications Board, National Research Council, *Keeping the U.S. Computer and Communications Industry Competitive: Convergence of Computing, Communications, and Entertainment*, National Academy Press, Washington, D.C., 1995.
7. Of course, entertainment includes games, and the computer community is very much a part of that product line.
8. Some television executives and engineers believe that interlace provides a better picture than does progressive, although admittedly not for text.
9. "Cable-TV Industry Loudly Likens Itself to Wild, Wild West," *Wall Street Journal*, May 10, 1995, p. B2.
10. Former FCC Commissioner Ervin Duggan, now president of PBS, captured the cultural diversity of at least two players in the converging arena when, on a panel at a Cable Television Show, he said that, despite their size and cash, telephone companies posed no danger to the current Hollywood and programming communities because telephone executives "still sleep in their pajamas." Bell executives may yet prove him wrong as they make major investments in creative material. Are the Bells going Hollywood? See "Staid Phone Giants Try Marriage to Hollywood," *Wall Street Journal*, May 24, 1995, p. B1.
11. Who will remain anonymous.
12. Following are some important differences between the computer and the entertainment worlds: computer is point-to-point communications, carriage, password-oriented access control where most messages have little intrinsic value; entertainment is point-to-multipoint, broadcast, and scrambled on broadband, and the material transmitted has great intrinsic value.
13. "Datacom Emerges As Next Wave," *Multichannel News*, May 15, 1995, p. 1.
14. *Information Superhighway, An Overview of Technology Challenges*, Report to the Congress, U.S. General Accounting Office, Washington, D.C., January 1995, p. 31.
15. H.R. 3626 (103rd Congress, 2nd Session), Section 405; S. 710 (104th Congress, 1st Session).
16. George Gilder, "When Bandwidth Is Free," *Wired*, September/October 1993, p. 38; George Gilder, "Telecom: The Bandwidth Tidal Wave," *Forbes ASAP*, Supplement, December 5, 1984, p. 163.
17. Speech by Daniel F. Akerson, Chairman & CEO, General Instrument Corporation, at Lehman Bros. Conference, San Francisco, California, April 28, 1994, Slide 5:

Fiber Optics Plus Compression = Personalized Channels Per Home

Bandwidth (MHz)	Homes/Node	Compression Ratio	Channels	Channels per HP
550	20,000	1:1	80	0.004
750	500	1:1	110	0.02
1,000	200	10:1	1,500	7.50

... By compressing 6 digital channels into the space of 1 analog channel, you can get 660 channels per today's node of 500 homes—or one personalized channel for each home. By further compressing at 10-to-1 ratios, using compression technologies such as GI's digicipher, a broadband system can provide 7 personalized channels per home. At this level of service, the consumer has the ability to be online from multiple TVs and PCs in the home.

18. General Instrument Corporation's digital television technologies are licensed under terms of agreements that it has with cable television operators. GI has licensed three U.S.-owned manufacturers (Scientific Atlanta, Zenith, Hewlett-Packard) to build its complete systems, including access control (security). GI has also indicated that it is prepared to more widely license its digital decompression technology. Parenthetically, the case seems strong for more narrow licensing of security technology than for other technologies. Although these licensing arrangements were established, in the first instance, as a result of vendor agreements, there are also strong economic incentives for vendors to license their technologies in order to drive their acceptance in the market. Charles H. Ferguson and Charles R. Morris, *Computer Wars*, Time Books, Random House, New York, 1993, p. 143 (hereinafter *Computer Wars*). In addition, the complexity of digital communications systems requires that someone perform the functions of disseminating information, answering questions relating to the system, and conducting compliance and interoperability testing. GI's licensing program addresses these needs.

19. Remarks prepared for delivery by Vice President Al Gore, Royce Hall, UCLA, Los Angeles, California, January 11, 1994.

20. Imperial Russia turned this shortcoming into an asset by maintaining gauges different from those of its Western neighbors as an additional deterrent to invasion. Lessons for the NII to be drawn from the experience of the U.S. railroads do exist: "Between 1865 and 1873, 35,000 miles of track were laid, a figure that exceeded the entire rail network of 1865, and inspired a boom in coal and pig iron production ... and the rapid spread of the new Bessemer process for making steel. Railroads opened vast new areas to commercial farming.... Their voracious appetite for funds absorbed much of the nation's investment capital ... contributed to the development of banking ... and facilitated the further concentration of the nation's capital market...." Eric Foner, *Reconstruction: America's Unfinished Revolution, 1863-1877*, Harper & Row, New York, 1988, p. 461.

21. The Grand Alliance is made up of the companies and organizations that were proponents of HDTV systems as part of the FCC's Advisory Committee on Advanced Television Service (ACATS) program for developing an advanced television standard. Participants were AT&T, General Instrument, MIT, Philips, Sarnoff, Thomson, and Zenith. AT&T and Zenith jointly submitted a progressive scan system for testing. Philips, Thomson, and Sarnoff jointly submitted an interlace system. GI and MIT jointly submitted two systems: one progressive and one interlace.

22. "Universal Demodulator Handles QAM, VSB and NTSC," *Communications Technology*, December, 1994, p. 100.

23. An interesting paper prepared for the American Enterprise Institute calls into question much of the conventional wisdom about interoperability. Milton Mueller has taken a historian's look at the development of the telephone companies in the United States and examined the effect that interconnection requirements had on the rollout of telephone service in the late nineteenth and early twentieth centuries. Mueller maintains that the absence of interconnection requirements was an important element of the rapid penetration of telephone service in the United States. The inability of competing telephone systems to interconnect with each other, according to Mueller, forced competition between the Bells and independent telephone companies to focus on providing access. In the absence of interconnection, the desire of potential customers to have access to telephone service and to communicate, not with everyone or anyone, but with specific groups of other users, became the central focus of competition. This resulted in more rapid penetration of service, driven by suppliers of that service. It became a race between the Bells and other providers to see who could sign up the most customers and most rapidly expand the network. This thesis may have implications for public policy and the rollout of new, digital services. Milton Mueller, "Universal Service: Competition, Interconnection and Monopoly in the Making of the American Telephone System," prepared for the American Enterprise Institute for Public Policy Research, working paper presented March 31, 1995.

24. Remarks by Daniel F. Akerson, Chairman and Chief Executive Officer, General Instrument Corporation, at the Washington Metropolitan Cable Club Luncheon, Washington, D.C., April 11, 1995. See also Peter K. Pitsch and David C. Murray, "A New Vision for Digital Telecommunications," A Briefing Paper, No. 171, The Competitiveness Center of the Hudson Institute, Indianapolis, Indiana, December 1994.

In addition, see Stanley M. Besen and Leland L. Johnson, "Compatibility Standards, Competition and Innovation in the Broadcasting Industry," Rand Corporation, November 1986, p. 135. According to Besen and Johnson, "The government should refrain from attempting to mandate or evaluate standards when the technologies themselves are subject to rapid change. A major reason for the Commission's difficulty in establishing the first color television standard was the fact that competing technologies were undergoing rapid change even during the Commission's deliberations. It is

only after the technologies have 'settled down' that government action is most likely to be fruitful, as illustrated in the TV stereo case."

25. Even interface standards have the potential to inhibit development. The Consumer Electronics Group (CEG) of the Electronics Industry Association and the National Cable Television Association (NCTA) have been struggling over the development of a decoder interface specification to be adopted by the FCC. This standard is seen by the Commission as a solution to issues of compatibility between televisions/VCRs and cable television systems. The proceeding is a response to the Cable Act of 1992 (Cable Television Consumer Protection and Competition Act of 1992, P.L. No. 102-385, 16 Stat. 1460 (1992), Section 17). NCTA and some cable industry suppliers are seeking an interface that can utilize a more extensive command set and more connector pins, maintaining that this will be needed to support future services, including QWERTY keyboards, an air mouse, and other yet-to-be-developed but envisioned capabilities. Alternatively, the cable side has argued that the interface should permit the "pass through" of consumer remote control commands so that the consumer can utilize all features of the network. CEG is resisting these requests, arguing, inter alia, potential interference with the features its members provide in consumer electronic equipment. Both sides have made extensive filings with the Commission in an ever growing record (In the Matter of Implementation of Section 17 of the Cable Television Consumer Protection and Competition Act of 1992; Compatibility Between Cable Systems and Consumer Electronics Equipment, FCC ET Docket No. 93-7). Recently, elements of the computer industry have entered the proceeding, voicing concerns over the adequacy of the proposed interface and at least one feature previously agreed to by CEG and NCTA. A Silicon Valley company has mounted a campaign against the proposed solution, charging it can become a bottleneck.

26. *Computer Wars*, pp. 113-14; see also fn. 17. During the early days of the Clinton Administration, this book was reportedly widely read in the White House. Current public policy debates provide a rich set of opportunities for the Administration to heed its lessons.

27. *Computer Wars*, p. 221; see also fn. 17.

28. Even this long-standing and "hallowed" justification for government regulation is being called into question. See Adam D. Thierer, "A Policy Maker's Guide to Deregulating Telecommunications, Part 1: The Open Access Solution," Heritage Talking Points, The Heritage Foundation, December 13, 1994; Paul Baren, "Visions of the 21st Century Communications: Is the Shortage of Radio Spectrum for Broadband Networks of the Future a Self-Made Problem?," keynote talk transcript, 8th Annual Conference on Next Generation Networks, Washington, D.C., November 9, 1994.

29. Cable Television Consumer Protection and Competition Act of 1992, P.L. No. 102-385, 16 Stat. 1460 (1992), Section 17.

30. Computer Systems Policy Project, "Perspectives on the National Information Infrastructure: Ensuring Interoperability," white paper, February 1994; Alliance to Promote Software Innovation and the Business Software Alliance, "The Information Marketplace: The Perspective of the Software and Computer Industry," special focus paper, Spring 1995.

31. Government standards, like all standards, need to take into account the international circumstances in which the standard will operate. The U.S. market remains the world's most attractive, thus giving a U.S. standard additional impetus. In the digital video arena, however, there are signs that Europe or Japan or both may gang up and try to protect their domestic industries by promulgating a standard different from that used in the United States. At a recent meeting in Melbourne of DAVIC, a group working on an international cable standard for digital video, the Japanese and European representatives joined forces in opposing the position of most U.S. representatives. It is too soon to tell whether this kind of activity will ultimately undermine the leadership position that the United States has held as the home of digital development.

32. This number was used by FCC Chairman Reed Hundt in a speech to the National Cable Television Association Annual Cable Show, May 9, 1995.

33. The origin of this term is unclear. Some early commentator made the statement that the square foot on the top of the television set was the most valuable piece of real estate on the information superhighway. This immediately caught the attention of at least one member of Congress who worried that it could become a bottleneck. It also brought the attention of many companies that announced that they planned to make "set-top boxes," long before they had thought through what functions the set-top box would perform or what applications consumers wanted. The term "set-top box" is almost certainly derived from the model of current cable television converters. The convenience that the use of the term provides is more than offset by the way it misdirects focus from the real issue, which is where different functionality is to be performed in a network.

34. *Computer Wars*; see also fn. 17.

35. Sometimes these efforts operate on a principle of unanimity. At other times, minority views are outvoted and rejected. Some groups are incredibly undemocratic; prudence requires not identifying them here, now. The United



States government should acquaint itself with the characteristics of such organizations prior to embracing or affirming their policies.

36. "... The bureaucratic solution is most frequently 'one size fits all.'" George Gilder, speech at the Virtual Democracy Conference, Progress and Freedom Foundation, January 10, 1995. In remarks also applicable to this subject, Gilder said that the answer to NII development lies not so much in competition as in diversity.

37. "... Most entrepreneurs who make breakthroughs are almost by definition psychologically uninterested in working in a dull, slow process in which most of your time is spent listening to people who don't know as much as you do." Speaker of the House of Representatives Newt Gingrich, speech at the Virtual Democracy Conference, Progress and Freedom Foundation, January 10, 1995.

38. Translation provided by Speaker Gingrich, in remarks to the Computer Science & Telecommunications Board, Washington, D.C., May 15, 1995.

39. From a slide offered by CableLabs at a recent Cable/IT Convergence Forum (and, one hopes, intended only to wake up the participants):

Membership Requirements:

- Subordinate your company's interests to those of a large consortium.
- Waste hundreds of man/hours in unproductive meetings in distant locations.
- Lock in your development process with a rigid set of technical standards.
- Spend money on large, unaccountable dues.
- Engage in complex, stressful political maneuvers with people in competing companies.

40. General Instrument Corporation's DigiCable product, which will allow cable systems to offer the equivalent of 350 to 500 channels, is approximately 1 year late to market. Six months of this is attributable to technical problems that the company and its vendors encountered; these digital consumer decoders are the most sophisticated pieces of consumer electronics hardware ever deployed. But 6 months of the delay was caused by a redesign to meet the specifications under development by the Moving Picture Experts Group for MPEG-2, even though MPEG-2 does not constitute a complete transmission system while DigiCable does. "Digital Compression: The Holy Grail," *Cablevision Magazine*, March 6, 1995, p. 22.

41. The tendency of politicians to count noses can work against important national resources. As foreign companies locate plants in the United States, they bring valuable employment to Americans. However, jobs are not the only contribution that technology makes to national wealth. In particular, U.S. ownership of and benefits from research and development activity are an important component of economic investment. Among senior advisers to the Clinton Administration, the relative merits of these factors have been the source of debate. See Robert Reich, "Who Is Us?" *Harvard Business Review*, January–February 1990, p. 53; and Laura D'Andrea Tyson, "Strong U.S. Firms Key to Healthy World Economy," *Los Angeles Times*, December 9, 1990, p. D2. To oversimplify, Reich thinks that foreign firms that provide jobs should be considered American; Tyson thinks that they must do R&D in the United States to qualify. Neither expresses concern about foreign ownership in the context of national and international standards setting; European governments and European multinationals, in particular, have demonstrated an all too frequent tendency to use standards to protect domestic industries.

42. Providing a government entitlement or backward compatibility protection for early adapters who purchased Beta machines for \$1,000 and more would not seem to be a major national priority; presumably they got something for their early purchase, namely to be the "first ones on their block...." In any case, the economy provides winners and losers. This author maintains an extensive collection of Beta tapes and offers to buy used Beta machines in good working condition for \$25 to \$35 and \$10 to \$15 for those in need of reasonable repair. Thus, a secondary market, of sorts, has already developed.

43. Or, worst of all, joint government/private-sector standards-setting activities!

44. Much of the material in this section is drawn from testimony of Richard S. Friedland, president and chief operating officer of General Instrument Corporation, before the NII Security Issues Forum, July 15, 1994.

45. H.R. 1555, Sec. 203 (104th Congress, 1st Session); H.R. 1275 & S. 664 (104th Congress, 1st Session).

46. The telephone specification was well established before regulatory rules and separation were imposed. Moreover, the installed base of telephones was huge, covering over 90 percent of the population. The current installed base of digital video equipment is minuscule by comparison and is exclusively one-way satellite. The technical complexity of voice telephone is orders of magnitude less than that of video networks. There is simply no comparison, a fact that did not inhibit members of Congress who showed up at a press conference announcing this legislation with antique, black, rotary-dial telephones.

47. The market for such equipment is currently highly competitive and has numerous providers, including General Instrument, Scientific Atlanta, Panasonic, Zenith, and Philips, among others. In anticipation of its potential

growth from the introduction of digital transmission and the advent of telephone company video services, others such as Hewlett-Packard, Thomson, Sun, and Sony have said they would enter. Today, telephones, like most consumer electronics, are produced by offshore companies, although many telephones carry the Bell label.

Retail sale proposals are advocated as consumer friendly and, as such, are supported by the organized public interest consumer groups. Retailers do not propose to cut off the right of network providers to also provide equipment, by sale or lease, but would prevent it from being bundled. Techniques such as those used so successfully by cellular telephone companies, whereby deep discounts for equipment allow more people to take advantage of service, would be barred. There are legitimate questions as to how "consumer friendly" it is to deprive consumers of the option of low-cost entry to these new services. Likewise, lease of equipment is likely to be an important component of the introduction of new services. It is one thing to buy a telephone for \$15 to \$100 from Radio Shack, knowing that the chances of its becoming obsolete are minimal. Indeed, it will almost certainly break before then. It is another thing to expect consumers to rush down to Circuit City to buy a \$300 to \$500 digital entertainment terminal and bear the risk of obsolescence and perhaps the risk that it will not work properly with the network.

48. Bell Atlantic, Nynex, and PacTel have issued an RFP for equipment, including separation of a network integration module (NIM) and a digital entertainment terminal (DET). The NIM would include network functions, including access control and decryption. The DET would be available for retail sale. See Chris Nolan, "The Telcos' Set-tops," *Cablevision*, April 3, 1995, p. 56. However, earlier this year, AT&T and VLSI Technologies announced a program that runs counter to the thrust of separation. See "AT&T, VLSI to Embed Security into Set Top Chips," *Broadcasting & Cable Magazine*, February 6, 1995, p. 36. AT&T customers, like GI customers, who are developing products, tend to focus, laser-like, on maximizing security. It is doubtful how many are aware that some in government have different plans for them.

49. *Congressional Record*, 104th Congress, 1st Session, April 4, 1995, p. S5143.

50. Although it is generally acknowledged that smart card technologies will be easier to develop and more secure in a digital environment, the experience in the analog environment is that security has been badly compromised. Satellite systems are badly compromised in Europe. See "Sky War Over 'Smart' Pirates," *Sunday Mail*, Financial Section, United Kingdom, October 9, 1994.

51. Interesting questions would be posed by such a government determination: What happens if nontechnologists in the Congress and/or in the bureaucracies (or even qualified technologists, for that matter) turn out to be wrong and the technology is defeated by theft-of-service pirates? Since the government has chosen this technology, would or should the government be liable for the losses incurred as a result of the defeat of this technology? If the technology were defeated, could retailers be liable for losses incurred thereby? Would that liability require a showing of negligence, such as the failure to maintain secure warehouses for the equipment? Given that maintaining the security of these systems is a matter of concern to the government (the success of our communications networks and the investment in those networks depend upon their ability to maintain security and privacy), would it be appropriate that punitive damages be available in the case of a break against a party, including a retailer, whose negligence caused a break? Should tort reform legislation contain an exception for punitive damages in such cases? Could a manufacturer refuse to make products available to a retailer who failed to sign an agreement in which the retailer promised to maintain adequate security procedures or an agreement in which the retailer promised to indemnify against losses from a breach of security?

52. Hybrid fiber coaxial cable systems (HFC) seem the preferred model for cable television operators and some telcos. Other telcos are indicating a preference for switched digital networks (SDN). Ultimate choices will be worked out in the marketplace over time, as network providers experiment with different delivery models. Cost will be a fundamental issue.

53. For example, retail sale, or even including security considerations advanced without regard to cost or convenience.

54. Retail sale has implications for other values as well, such as innovation. To the extent retail sale is used to justify standardization, it can negatively affect continued development.

57

## Technology and Cost Models for Connecting K-12 Schools to the National Information Infrastructure

*Russell I. Rothstein and Lee McKnight*  
*Massachusetts Institute of Technology*

### ABSTRACT

It is presumed that universal connection to the national information infrastructure (NII) will facilitate educational reform within schools and communities. Several developments suggest that this might be the case:

- The federal government is committed to have every classroom in the United States connected to the NII by the year 2000;
- A number of telephone and cable companies have announced plans to connect schools in their service areas at low or no cost;
- Modern, high-speed networks have been installed at a number of progressive, pioneering K-12 schools; and
- The Internet, a global network of networks, has grown rapidly, providing connections to an abundance of educational resources.

However, to date, there is relatively little known about the costs for connecting schools to the information infrastructure. Even if the exact costs are unknown, they are expected to be significant. To reduce these costs, a variety of approaches are being tried.

Some states have implemented the cost-saving program of purchasing telecommunications equipment and services for all schools in the state. For example, North Carolina has saved its schools 20 to 50 percent of the costs for certain items. Some states have passed legislation permitting the state public utility commission to set preferential or fixed intrastate rates for educational institutions.

Using a baseline of service required for connecting to the NII, there will be \$9.4 billion to \$22.0 billion in one-time costs with annual maintenance costs of \$1.8 billion to \$4.6 billion. At the per-pupil level, this is equivalent to \$212 to \$501 in one-time installation costs and an ongoing annual cost of \$40 to \$105.

Hardware is the most significant cost item for schools. However, most of this cost item is allocated for the purchase of PCs in the schools. The value of the PCs goes well beyond their use as networking devices. Therefore, the real costs for PC purchases should be allocated across other parts of the technology budget, and not only to the networking component. If this is done, then the hardware costs for connecting to the NII drop considerably.

Excluding PC expenditures, costs for support of the network represent about one-third of all networking. Support is a vital part of the successful implementation of a school network and its costs must be factored in to the budget. Support and training together constitute 46 percent of the total costs of networking schools. Costs for telecommunications lines and services represent only 11 percent of the total costs. This amount is lower than the costs assumed by much of the technology community, including the telecommunications service and equipment providers.

---

NOTE: Russell I. Rothstein is a research assistant and Lee McKnight is a principal research associate with the MIT Research Program on Communications Policy. This paper is based on a report prepared when Russell Rothstein was a visiting researcher at the Office of Educational Technology in the U.S. Department of Education in 1994. Further work was supported by ARPA contract N00174-93-C-0036 and NSF grant NCR-9307548 (Networked multimedia Information Services). The invaluable help of Linda Roberts, U.S. Department of Education, and Joseph Bailey, Thomas Lee, and Sharon Gillett, MIT Research Program on Communications Policy, is gratefully acknowledged. All rights reserved.

Our research suggests that a number of programs would have a significant impact on the total costs of connecting to the NII. If all schools coordinate purchasing at the state level, cost savings will exceed \$2 billion. Colleges and universities often have the resources to provide technical support to K-12 schools. If a nationwide program were instituted, potential savings would be \$800 million to \$1.8 billion. If schools were given free Internet connectivity, the reduction in total annual costs for school Internet connections would be between \$150 million and \$630 million.

Finally, as the costs of networking schools are better understood, a new question arises: how will these costs be financed? Many states have programs to fund networking in schools. The federal government has a role, although it must become more flexible and coordinated. However, as Vice President Al Gore has continued to state, the NII will be built by the private sector. A number of states have initiated cooperative ventures between businesses and schools. An expansion of these programs may well be the key for successfully connecting K-12 schools to the NII.

## INTRODUCTION

On January 11, 1994, Vice President Al Gore challenged the nation to "connect every classroom by the year 2000" to the national information infrastructure (NII). In testimony before Congress in May 1994, Secretary of Education Richard Riley said, "We may have to go a step further and provide our schools with free usage of the telecommunications lines that will connect school children and young people" to the NII. In an address at the Harvard Graduate School of Education, FCC Chairman Reed Hundt said that "if the Administration's challenge is met by everyone, education in this country will be reinvented, forever and for better." Universal connection to the NII, it is presumed, will facilitate educational reform within schools. However, to date, relatively little information has been available about the costs for connecting schools to the information infrastructure. This paper presents models for evaluating the total cost of full NII connectivity for schools through an engineering cost study of equipment, services, software, and training needs.

### Cost Models of K-12 Networking

Five models for connecting schools to the NII are presented in the next section in order of increasing cost and power to describe the path that many schools may follow. A school will likely begin its connection through the low-cost dial-up option described in model one. As the school builds expertise and develops a need for greater capability, it will upgrade to a higher level of connectivity. It is not until the school acquires telecommunications infrastructure similar to model four that it is able to take advantage of many of the educational services and applications provided on the emerging NII. Model five presents the costs for putting a PC on the desktop of every student, with a high-speed connection to the Internet. Although this setup is not necessary for access to many of the coming NII services, it presents a model of systemic educational reform with information and networking technology.

These models are representations of the network technology used in schools. A level of complexity and detail is omitted from these models, but the simplicity is helpful because the models encompass broad cross sections of network and school configurations. The models provide a clearer view of the costs and choices for networking K-12 schools.

There are numerous ways to define a school network. The models presented below follow the Internet networking model, in which schools have digital data connections that transmit and receive bits of information. The models exclude both analog video point-to-point networks and voice networks including PBX, centrex, and voice-mail systems. Audio and video functions are possible in digital format over the Internet data network. However, many schools will require video and voice networks in addition to the data networks. The costs of these systems are important to consider but are not modeled in this paper.

It should be noted that although voice and video networks have been separated out from data networks in this paper, schools should not consider these three types of networks to be wholly distinct. Some schools have integrated their voice and video networks with the school data network. The sharing of resources among the multiple networks can be effective in providing significant cost savings. At a basic level, it must be understood

that as a school installs a LAN and puts computer data connections in every classroom, there are minimal added costs to concurrently install other types of connections, including telephone lines.

### Assumptions

The assumptions described below are the basis for determining the costs of networking schools. These assumptions are conservative but realistic, to ensure that the total costs are not underestimated.

### Technology Standard for Connecting to the NII

As described by the Information Infrastructure Task Force (1994), the NII "promises every ... school ... in the nation access anywhere to voice, data, full-motion video, and multimedia applications.... Through the NII, students of all ages will use multimedia electronic libraries and museums containing text, images, video, music, simulations, and instructional software." In models four and five, the school has access to these NII services. The following requirements outline the needs for those models in order to have full connection to the NII:

- *A LAN within each school with connections to multiple machines in every classroom.* The power of the network is greatly enhanced as the number of access points increases throughout the school. A classroom with one connection is not conducive to use of network applications in a class of 20 or 30 students. Telecommunications technology will not be a tool for systemic educational reform until network connections are available throughout the school.
- *A connection from each school to a community hub.* From 2 to 10 schools should connect to a single hub, depending on the size of the schools. In most cases, the hub will reside at the school district office. However, in cases where there are many schools in a single district, the schools should be clustered into sets of four to six. Each of these school clusters will have a group hub, probably at the district office, which will contain the center of the network for those schools. The rationale for the use of this architecture is described below.
- *A connection between the school LAN and the district office hub.* With this configuration, every classroom has a connection not only to every other classroom in the school but also to the central school district office.
- *A connection from the school district office to a community-, state-, or nationwide wide area network (WAN).* This connection will allow all schools to connect to the WAN. The Internet is a good example of a WAN and is used throughout this report as a model and a precursor for the coming NII.
- *Sufficient bandwidth for these connections.* With a high-bandwidth connection, users in schools can make use of graphical applications (e.g., Mosaic) and limited video service (e.g., CU-SeeMe and MBONE). For most school districts, the minimum bandwidth, or bit-speed, that will support these services is 56,000 bits per second (56 kbps). Therefore, the connection between the school and the hub must be at or above this level. For the connection from the district office to the Internet, a higher bandwidth connection is necessary because all of the schools in the group connect to the Internet through this line. The suggested minimum bandwidth for this connection is 1,500,000 bits per second (1.5 Mbps), otherwise known as a T-1 line.
- *Symmetric, bidirectional access to the WAN/Internet.* It is important that the connection to a school allow information to flow both in and out of the school at the same rates. In this way, students can become both consumers and providers of information over the network.
- *Adequate remote dial-up facilities.* With a sufficient number of modems and phone lines, faculty, students, and parents can gain access to the school system remotely on weekends and after school hours.
- *Use of established and tested technologies.* Schools have benefited most from mature technologies that have been well tested in the marketplace. Use of cutting-edge technologies has not been as successful in schools due to the instability of the technologies and the large amount of resources required to support them. The models assume the use of mature technology and transmission media. Therefore, modern technologies such as

wireless and hybrid fiber coaxial systems are not considered in this study. However, given the rapidity of technological change and marketplace evolution for networking products and services, wireless and cable alternatives should be evaluated in future research.

Each of the successive models progresses closer to this configuration. The first three models represent the schools and districts that have not reached this level of networking. It is not until the fourth model that these requirements have been incorporated. The fifth model continues on this path and exceeds the baseline requirements. It is assumed that some schools will traverse the path through each of the models to connect to the national information infrastructure.

### Architecture of the District Network

The basic network architecture for these models follows the "star" network configuration as illustrated in Figure 1. This architecture is also used in the nationwide telephone network. In the telephone network, residential telephone lines in an area are directly connected to a single district office. In the school network, each school building is connected to the school central hub. In most cases, the district office will serve as the central hub. However, in cases where there are either few or large numbers of schools in one district, alternative sites must be chosen.

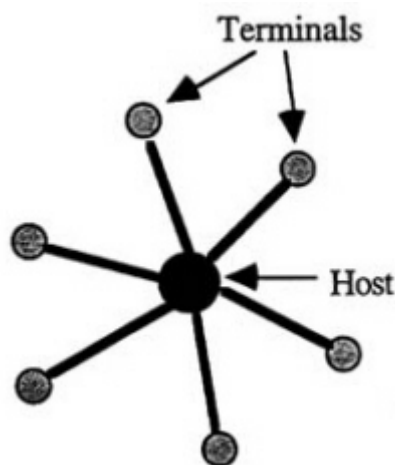


Figure 1 "Star" network. SOURCE: Newman et al. (1992).

The rationale for adopting this architecture is that, when many schools are connected through a single hub, costs can be aggregated among the schools. This gives schools stronger purchasing power as equipment purchases are aggregated by the school district for volume discounts. It also allows schools to share resources—such as the data line to the Internet, training programs, and full-time support staff—that each school might not be able to afford individually. Therefore, there are costs both at the school and at the district level for networking schools across the country. The star network configuration for schools, utilizing a community or district hub, is recommended in Gargano and Wasley (1994) and in California Department of Education (1994).

### COST AREAS

The cost models presented in this paper include four types of costs—hardware, training, support, and retrofitting. The items included in these categories are summarized below.

- *Hardware.* Wiring, router, server, PCs, including installation, maintenance, and servicing of the hardware and telecommunications lines.

- *Training.* Training of teachers and other school staff to use the network.
- *Support.* Technical support of the network.
- *Retrofitting.* Modifications to the school facility to accommodate the telecommunications infrastructure. This may include costs for asbestos removal (where applicable); installation of electrical systems, climate control systems, and added security (locks, alarms, etc.); and renovation of buildings to accommodate network installation and operation.

The following cost area is not included in the models:

- *Educational software.* There are "freeware" versions of many popular Internet applications. However, other educational software may be desired by particular schools. The costs for this software may be high, but they are not included in the models. Further economic analysis of software costs and their evolution in the network scenarios analyzed below is required.

### **Instructional Use of the Network**

It is likely that the type of technology model deployed in the school will greatly affect the use of the network and the educational benefits obtained. A school with multiple networked PCs in every classroom (model four) will reap greater educational benefits from the network than a school with a single PC and modem (model one). Similarly, video and graphical applications (e.g., Mosaic and CU-SeeMe), available in models four and five, add educational value to the text-based applications available in the lower-end models. However, there has not yet been a quantitative analysis of the educational benefits associated with each particular model. This paper is concerned exclusively with the costs for the various network models. Study to determine the educational benefits of various K-12 network models is also needed. When the educational benefits are quantified, they should be compared to the costs outlined in this paper. The synthesis of these studies will generate a cost-benefit curve for connecting schools to the network. That information is vital for determining national policy on connecting schools to the NII.

### **School Characteristics**

The models described are based on a "typical" school and school district, as defined by the U.S. Department of Education (1993), and represent the average costs of all U.S. schools and school districts. Many schools will differ in significant ways from the "typical" school and will therefore face somewhat different costs from those presented in the models.

*Size of School.* The average school has about 518 students and twenty classrooms. It employs 27 teachers and 25 other school staff. The average number of schools in a school district is about six. (These numbers are based on a national enrollment of approximately 44 million students in 85,000 public schools in 15,000 school districts.)

*Existing Equipment.* According to Anderson (1993), as of 1992 there was an average of 23 computers per school at the elementary and middle school levels, and 47 computers per school at the secondary school level. About 15 percent of these machines, or three to seven machines per school, are capable of running the network protocol (TCP/IP) to access the Internet. During the 3 years from 1989 until 1992, the number of computers in schools grew by 50 percent. Given the increasing growth rate of computers in the market, it is safe to assume that the growth rate of computers in these schools has exceeded 50 percent over the past 2 years. Given these assumptions, in an average school there are at least seven PCs capable of running graphical Internet applications (e.g., Mosaic). This number of PCs is sufficient for the first two models, but it is not sufficient for establishing multiple connections in every classroom throughout the school. Therefore, for models three, four, and five, there is a line-item cost for purchasing additional PCs.

## COST MODELS

### Single-PC Dial-up (Model One)

The single-PC dial-up model (Figure 2) represents the most basic connectivity option for a school. The school has no internal LAN. There is a single connection to the district office over a modem and standard phone line. Only one user may use the connection at any time. Since the system is limited, only a few teachers in the school require training. Users of the system will be able to use text-based applications over the Internet (e.g., e-mail, Telnet, Gopher), but will have no real-time access to video or graphics.

Model one is the lowest-cost option for schools. Many of the services and benefits envisioned for the NII will not be widely accessible in schools using this model of connectivity. Table 1 lists the cost items associated with the single-PC dial-up model.

### LAN with Shared Modem (Model Two)

The primary difference between model two (Figure 3) and model one is the existence of a LAN within the school. By connecting the modem to the LAN, every computer on the network has access to the Internet. However, this model supports only a few users at a time, since it is limited by the number of phone lines going out of the school. As in model one, users of the system can use text-based applications over the Internet (e.g., e-mail, Telnet, Gopher) but have no real-time access to video or graphics.

In model two, there is the added cost for a LAN. This model assumes the use of copper wire (category 5) as the medium for the network since it is the most affordable and scaleable option for schools in 1994. The costs for the wiring and network cards run \$100 to \$150 per PC connected. Including the costs for the accompanying hardware and labor, the costs per PC are \$400 to \$500. Therefore, for a school with 60 to 100 connected PCs (3 to 5 PCs per classroom @ 20 classrooms), the total LAN costs are \$20,000 to \$55,000.

Model two is another low-cost option for schools. However, many of the services and benefits envisioned for the NII are still not widely accessible in this model. Table 2 lists the cost items associated with this model.

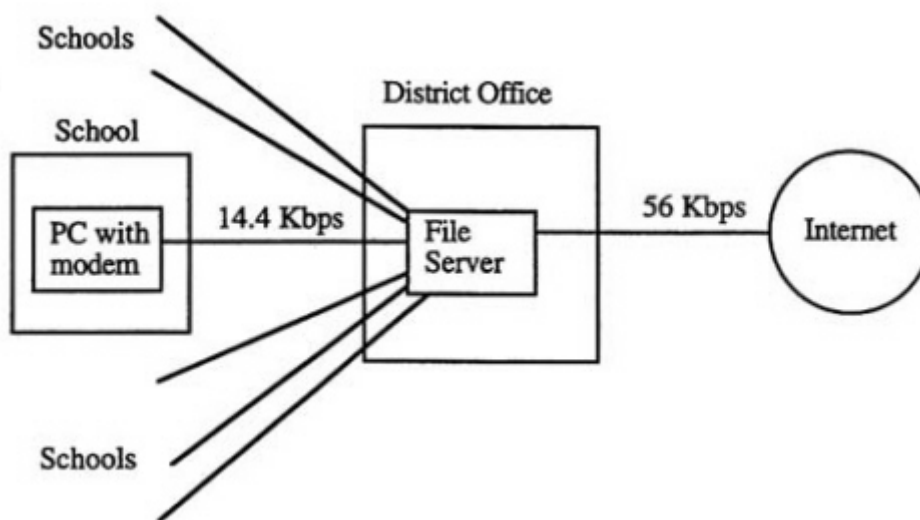


Figure 2 Single-PC dial-up model. SOURCE: Rothstein (1994).



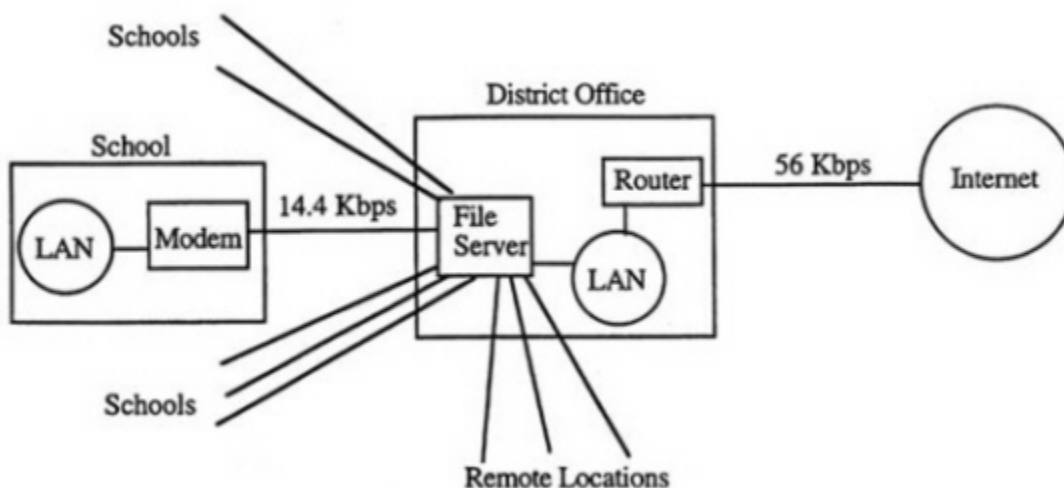


Figure 3 LAN with shared modem model. SOURCE: Rothstein (1994).

### Lan with Router (Model Three)

The primary difference between model three (Figure 4) and model two is a router in place of the modem. With the router, multiple users of the LAN may access the Internet concurrently. Just as in the first two models, users of the system are able to use text-based applications over the Internet (e.g., e-mail, Telnet, Gopher) but have no real-time access to video or graphics.

Since the router allows multiple users of the system, there is an opportunity to expand the entire network infrastructure. With this infrastructure, it is reasonable to support one PC in every classroom. Therefore, there is a requirement to purchase 15 additional PCs for the average school to use in addition to its small initial stock of TCP/IP-compatible machines. It is assumed that the purchasing of these PCs is done at the district level to negotiate better rates (\$1,000 to \$2,000 per PC). Support and training costs are higher since there are additional users of the system. There are additional dial-up lines required to accommodate remote access, as well as significant retrofitting costs for the electrical system, climate control system, and enhanced security. Table 3 gives a line item summary of the costs associated with model three.

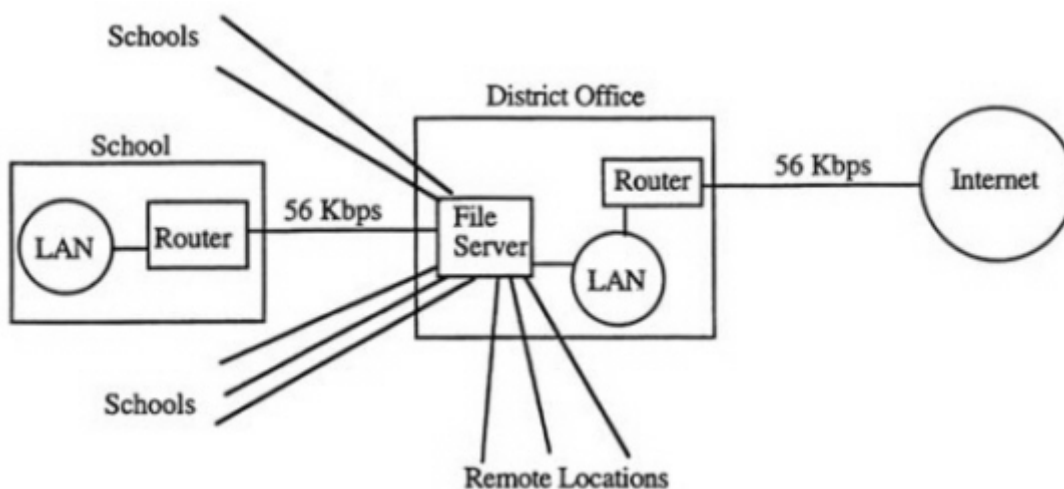


Figure 4 LAN with router model. SOURCE: Rothstein (1994).

### LAN with Local Server and Dedicated Line (Model Four)

The primary difference between model four (Figure 5) and model three is the existence of a file server at the school. The on-site server allows much of the information to reside locally at the school instead of at the district office. This feature provides better performance since much of the information does not need to be fetched over the network. Additionally, the local server allows school administrators to exercise greater control over the information that flows in and out of the school. Higher-speed links from the school enable the use of limited video, graphical, and text-based network applications.

In model four, virtually the entire school is supported on the network. As a result, the training program is extensive and the support team is well staffed. The costs of the connection to the Internet are also higher due to the larger bandwidth connection. There are significant retrofitting costs for the electrical system, climate control system, and better security. Table 4 lists the cost items associated with this model.

The costs associated with using model four are indicative of the costs of connecting K-12 schools across the country to the NII. These numbers indicate that there will be \$9.4 billion to \$22.0 billion in one-time costs, with annual maintenance costs of \$1.8 billion to \$4.6 billion. At the per-pupil level, this is equivalent to \$212 to \$501 in onetime installation costs and an ongoing annual cost of \$40 to \$105.

In this model, hardware is the most significant cost item for schools. However, most of this cost item is allocated for the purchase of PCs in the schools. The value of the PCs goes well beyond their use as networking devices. Therefore, the real costs for PC purchases should be allocated across other parts of the technology budget, and not only to the networking component. If this is done, then the hardware costs for connecting to the NII drop considerably.

If the high start-up costs are amortized equally over a 5-year period, then the breakdown of costs during the first 5 years, excluding PC purchases, is as shown in Figure 6.

Costs for support of the network represent about one-third of all networking costs in model four. Support is a vital part of the successful implementation of a school network and its costs must be factored into the budget. Support and training together account for 46 percent of the total costs of networking schools.

Finally, it is important to note that the costs for telecommunications lines and services represent only 11 percent of the total costs. This amount is lower than the costs assumed by much of the technology community, including the telecommunications service and equipment providers.

### Ubiquitous LAN with Local Server and High-Speed Line

Model five (Figure 7) represents a full, ubiquitous connection to the NII. In this model, there is a PC on the desktop of every student and teacher. There is a high-bandwidth connection to the school to support large numbers of concurrent users of the system. This model supports the full suite of text, audio, graphical, and video applications available over the Internet.

In model five, the entire school is supported on the network. A large portion of the costs for this model is the expenditure for PCs on every desktop. Assuming 500 students, there is a need to purchase 450 new PCs. Since the network is ubiquitous, the training program is extensive and the support team is well staffed. The costs of the connection to the Internet are also higher because of the high-speed line going into the school. The file server is larger than model four's server to accommodate the large number of networked PCs. The dial-up system is larger in order to allow many students, teachers, and parents to access the system remotely. The retrofitting costs are substantial because extensive electrical work must be performed in the average school to accommodate the hundreds of new PCs. In addition, the school in model five must make expenditures on air conditioners and security locks to protect the new equipment. Table 5 lists the cost items associated with this model.

### COST COMPARISON OF MODELS

U.S. expenditures on K-12 education in 1992–93 totaled \$280 billion. Total onetime costs for model four represent 3 to 7 percent of total national educational expenditures. The ongoing annual costs represent between

0.6 and 1.6 percent of total national educational expenditures. For model five, the costs are more significant, with one-time costs representing 18 to 41 percent of total national educational expenditures.

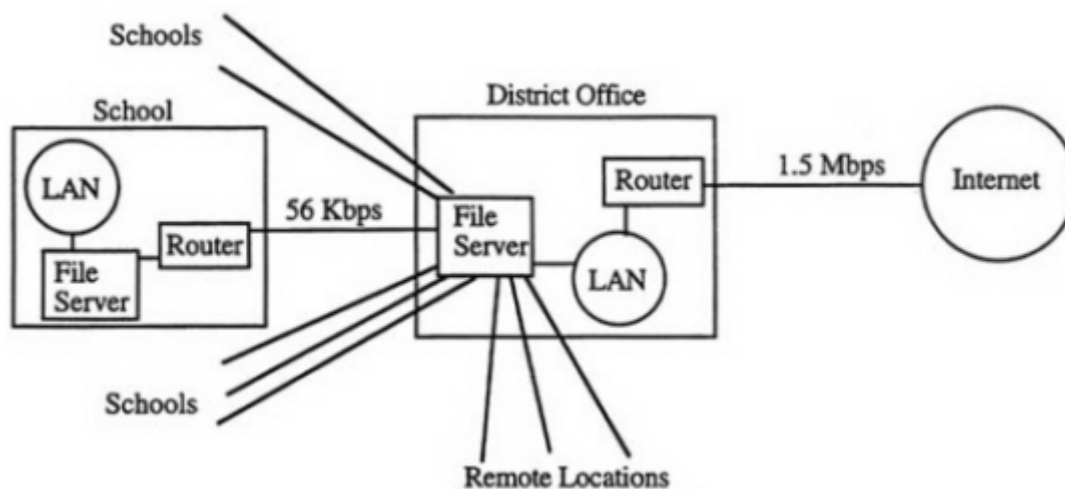


Figure 5 LAN with local server and dedicated line model. SOURCE: Rothstein (1994).

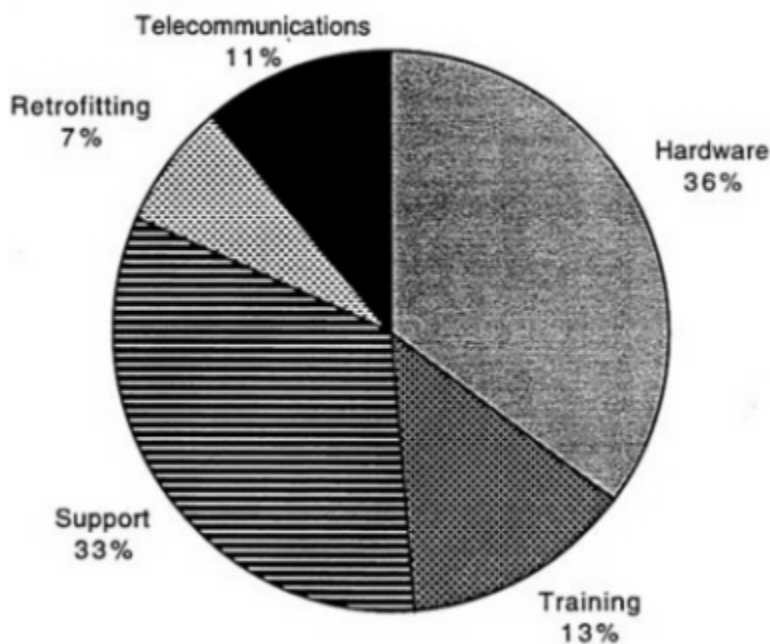


Figure 6 Breakdown of costs for model four. SOURCE: Massachusetts Institute of Technology, Research Program on Communications Policy (1994).

The models with advanced connectivity include significant equipment and training costs, which may be beneficial for educational purposes other than network uses. If these costs are accounted for separately, the difference in costs between models four and five will not be as significant as those presented here.

Table 6 summarizes the associated range of costs for the various technology models.

#### POTENTIAL IMPACT OF COST-REDUCTION INITIATIVES

Much more can be done by the government and the private sector to significantly mitigate costs schools face to connect to the NII. This section examines some possible programs and their impact on costs to schools.

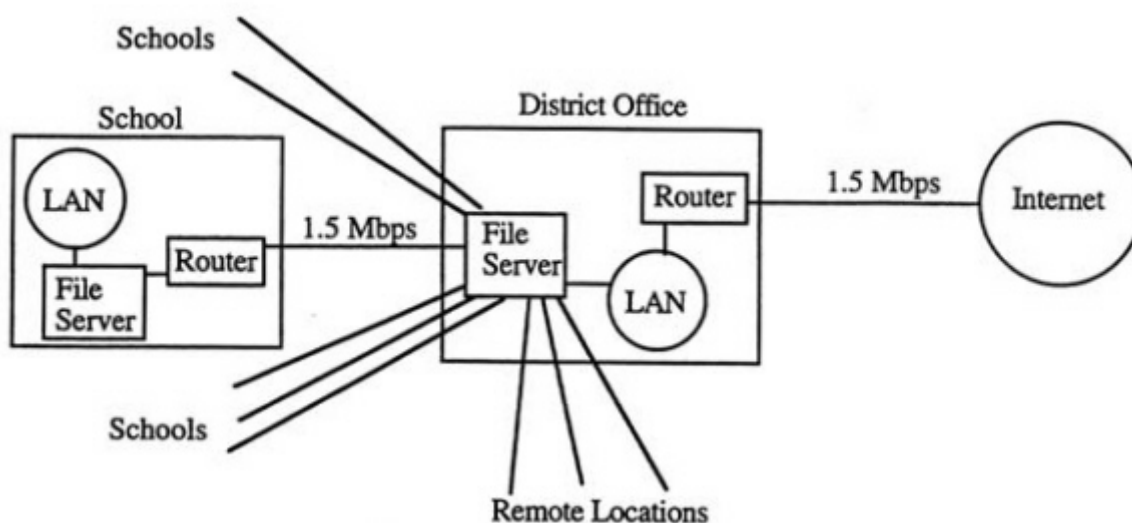


Figure 7 Ubiquitous LAN with local server and high-speed line model. SOURCE: Rothstein (1994).

TABLE 1 Single-PC Dial-up Model Costs (dollars)

	Low	High
<b>SCHOOL COSTS</b>		
	Onetime Installation Costs	
Telephone line	100	250
Modem	150	250
Total	250	500
	Annual Operating Costs	
Replacement of equipment	50	150
Telephone line (10 hr/month)	300	1,500
Total	350	1,650
<b>DISTRICT OFFICE COSTS</b>		
	Onetime Installation Costs	
File server	2,000	10,000
Data line to WAN/Internet (56 kbps)	500	2,000
Training (2 to 4 teachers per school)	1,000	10,000
Total	3,500	22,000
	Annual Operating Costs	
Internet service (56 kbps)	5,000	10,000
Support	2,000	10,000
Training	1,000	5,000
Total	8,000	25,000
<b>U.S. ONETIME COST</b>		
Onetime cost per student	1.68	\$8.47
Total	70,000,000	370,000,000
<b>U.S. ANNUAL COSTS</b>		
Annual cost per student	3.40	11.71
Total	150,000,000	520,000,000

SOURCE: Rothstein (1994).

TABLE 2 LAN with Shared Modem Model Costs (dollars)

	Low	High
<b>SCHOOL COSTS</b>		
	Onetime Installation Costs	
Local area network	20,000	55,000
LAN modem	500	1,200
Retrofitting (minor)	2,000	10,000
Total	22,500	66,200
	Annual Operating Costs	
Replacement of equipment	3,000	8,250
Shared telephone line (40 hr/month)	1,200	6,000
Total	4,200	14,250
<b>DISTRICT OFFICE COSTS</b>		
	Onetime Installation Costs	
File server	2,000	10,000
District local area network	2,000	5,000
Data line to WAN/Internet (56 kbps)	500	2,000
Dial-up capabilities (2 lines)	2,000	4,000
Training (5 to 20 staff per school)	1,000	10,000
Total	7,500	31,000
	Annual Operating Costs	
Internet service (56 kbps)	5,000	10,000
Dial-up lines	300	500
Support (1 to 2 staff per district)	45,000	90,000
Training	10,000	20,000
Total	60,300	120,500
<b>U.S. ONETIME COST</b>		
Onetime cost per student	46.02	138.45
Total	2,030,000,000	6,090,000,000
<b>U.S. ANNUAL COST</b>		
Annual cost per student	28.67	68.61
Total	1,260,000,000	3,020,000,000

SOURCE: Rothstein (1994).

TABLE 3 LAN with Router Model Costs (dollars)

	Low	High
<b>SCHOOL COSTS</b>		
	Onetime Installation Costs	
Local area network	20,000	55,000
Personal computers (15 machines)	15,000	30,000
Router	2,000	3,000
Connection to hub (14.4 kbps or 56 kbps)	50	1,000
Retrofitting (major)	10,000	25,000
Total	47,050	114,000
	Annual Operating Costs	
Replacement of equipment	3,000	8,250
Connection to hub (14.4 kbps or 56 kbps)	500	10,000
Total	3,500	18,250
<b>DISTRICT OFFICE COSTS</b>		
	Onetime Installation Costs	
File server	2,000	15,000
Router	2,000	5,000
District local area network	2,000	5,000
Data line to WAN/Internet (56 kbps)	500	2,000
Dial-up capabilities (8 lines)	8,000	16,000
Training (10 to 20 staff per school)	1,000	10,000
Total	15,500	53,000
	Annual Operating Costs	
Internet service (56 kbps)	5,000	10,000
Dial-up lines	1,200	2,000
Support (1 to 2 staff per district)	45,000	90,000
Training	10,000	20,000
Total	61,200	122,000
<b>U.S. ONETIME COST</b>		
Onetime cost per student	96.18	238.30
Total	4,230,000,000	10,490,000,000
<b>U.S. ANNUAL COST</b>		
Annual cost per student	27.63	76.85
Total	1,220,000,000	3,380,000,000

SOURCE: Rothstein (1994).

TABLE 4 LAN with Local Server and Dedicated Line Model Costs (dollars)

	Low	High
<b>SCHOOL COSTS</b>		
	Onetime Installation Costs	
Local area network	20,000	55,000
Personal computers (60 machines)	60,000	120,000
File server	4,000	15,000
Connection to hub/district office (56 kbps)	500	2,000
Router and CSU/DSU	2,600	5,000
Retrofitting (major)	10,000	25,000
Total	97,100	222,000
	Annual Operating Costs	
Replacement of equipment	3,000	8,250
Connection to hub/district office (56 kbps)	1,000	5,000
Total	4,000	13,250
<b>DISTRICT OFFICE COSTS</b>		
	Onetime Installation Costs	
File server	2,000	15,000
Router	2,000	5,000
District local area network	2,000	5,000
Data line to WAN/Internet (1.5 Mbps)	1,000	5,000
Dial-up capabilities (20 lines)	16,000	32,000
Training (40 to 50 staff per school)	50,000	150,000
Total	73,000	212,000
	Annual Operating Costs	
Internet service (1.5 Mbps)	10,000	42,000
Dial-up lines	3,000	5,000
Support (2 to 3 staff per district)	66,000	150,000
Training	15,000	35,000
Total	94,000	232,000
<b>U.S. ONETIME COST</b>		
Onetime cost per student	212.47	501.14
Total	9,350,000,000	22,050,000,000
<b>U.S. ANNUAL COST</b>		
Annual cost per student	39.77	104.69
Total	1,750,000,000	4,610,000,000

SOURCE: Rothstein (1994).

TABLE 5 Ubiquitous LAN with Local Server and High-Speed Line Model Costs (dollars)

	Low	High
<b>SCHOOL COSTS</b>		
	Onetime Installation Costs	
Local area network	50,000	100,000
File server	10,000	25,000
Connection to hub/district office (1.5 Mbps)	1,200	5,000
Router and CSU/DSU	4,000	7,000
PC on every desk (450 new machines)	450,000	900,000
Retrofitting (major, including electrical)	70,000	250,000
Total	585,200	1,287,000
	Annual Operating Costs	
Replacement of equipment	7,500	15,000
Connection to hub/district office (1.5 Mbps)	10,000	35,000
Total	17,500	50,000
<b>DISTRICT OFFICE COSTS</b>		
	Onetime Installation Costs	
File server	5,000	15,000
Router	2,000	5,000
District local area network	2,000	5,000
Data line to WAN/Internet (1.5 Mbps)	1,000	5,000
Dial-up capabilities (50 lines)	32,000	80,000
Training (all teachers in school)	55,000	165,000
Total	97,000	275,000
	Annual Operating Costs	
Internet service (1.5 Mbps)	10,000	42,000
Dial-up lines	30,000	50,000
Support (4 to 5 staff per district)	112,200	255,000
Training	16,500	38,500
Total	168,700	385,500
<b>U.S. ONETIME COST</b>		
Onetime cost per student	1,163.57	2,580.00
Total	51,200,000,000	113,520,000,000
<b>TOTAL U.S. ANNUAL COST</b>		
Annual cost per student	91.32	228.01
Total	4,020,000,000	10,030,000,000

SOURCE: Rothstein (1994).



TABLE 6 Total Onetime and Ongoing Costs for Associated Models (billions of dollars)

Model	Onetime Costs		Ongoing Costs	
	Low	High	Low	High
Single-PC dial-up	0.07	0.37	0.15	0.52
LAN w/shared modem	2.03	6.09	1.26	3.02
LAN w/router	4.23	10.49	1.22	3.38
LAN w/local server and dedicated line	9.35	22.05	1.75	4.61
Ubiquitous LAN w/high-speed connection	51.20	113.52	4.02	10.03

SOURCE: Rothstein (1994).

In determining the costs savings from programs supporting connection to the NII, it is imperative to define the model of the NII. In this paper, the baseline model for NII access is a school with a LAN, a local server, and a dedicated line to the district hub. This is the fourth model as described above. The costs for this model are summarized in Table 7.

Based on the costs listed in Table 7, we have estimated the potential total cost savings for U.S. schools from various programs.

1. *Preferential telecommunications tariff rates are instituted for schools.* Some state utility commissions have instituted preferential telecommunications rates for educational institutions. These rates are applicable only for intrastate traffic. For interstate traffic, the tariffs set by the Federal Communications Commission are in effect. Currently, these tariffs have no preferential rates for educational institutions. The amount of money that schools will save will depend on the amount of discount if preferential rates are adopted. The following numbers represent the estimated savings with educational discounts of 30 percent and 60 percent.
  - 30 percent reduction—\$39 million to \$150 million (annual)  
 Estimated savings: \$89 million to \$218 million (onetime)
  - 60 percent reduction—\$78 million to \$300 million (annual)  
 Estimated savings: \$179 million to \$435 million (onetime)

TABLE 7 Total U.S. Costs for Baseline Connection to the NII (as in Model Four) (millions of dollars)

Component	Onetime Costs		Ongoing Costs	
	Low	High	Low	High
Local area network	1,730	4,750	0	0
Personal computers	5,100	10,200	0	0
File server	370	1,500	0	0
Telecommunications lines	298	725	130	500
Router and CSU/DSU	221	425	0	0
Retrofitting	850	2,125	0	0
Training	750	2,250	225	525
Internet service	0	0	150	630
Support	0	0	990	2,250
Replacement of equipment	0	0	255	701
Total	9,319	21,975	1,750	4,606

SOURCE: Rothstein (1994).

2. *All information technology purchasing is done at the state level.* When states are involved in purchasing information technology, schools may secure better prices due to volume discounts. Schools in North

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.

Carolina, for example, have enjoyed discounts of 20 to 50 percent for hardware and labor costs. The following figures indicate the possible cost savings across all 50 states, based on an average 30 percent discount.

- Estimated savings: \$1.9 billion to \$4.1 billion (onetime); \$45 million to \$189 million (annual)
- 3. *Universities or other institutions provide technical support to schools.* Universities can also play a role in providing technical support to K-12 schools. Many universities have already undertaken such a project and have provided network support to a number of K-12 schools in their area. With such a program, schools will still require some dedicated support staff. However, it is assumed that schools will be able to function with 80 percent less technical support staff than would be required without university support.
- Estimated savings: \$790 million to \$1.8 billion (onetime)
- 4. *Teachers are trained on their own time.* In the models, a large portion of the training costs are dedicated either to paying substitute teachers to cover for teachers in training, or to paying teachers to be trained after school hours. If teachers agree to attend classes on their own time, there will be costs only for the trainer.
- Estimated savings: \$0 to \$1.5 billion (onetime); \$0 to \$300 million (annual)
- 5. *The LAN is installed by volunteers.* In the models, 65 percent of the costs for installing the LAN are dedicated to labor. If schools can do this work with volunteers, then the cost savings are significant. As an example, Val Verde Unified School District in California laid its wires with volunteers from parents and community members. If such groups provide labor at no cost to schools, schools will reap significant savings.
- Estimated savings: \$1.1 billion to \$3.1 billion (onetime)
- 6. *Personal computers are donated to schools.* In the models, there is a need to purchase a significant number of PCs to provide four to five connections to the network in every classroom. The costs for these PCs can be offset by donations of new machines from PC manufacturers. It is also possible for large corporations to donate these computers to schools. However, the schools will need fairly modern machines to run networking software. The success of a donation program is dependent on the quality of the equipment donated. Donations of obsolete or incompatible equipment may be costly to schools.
- Estimated savings: \$5.1 billion to \$10.2 billion (onetime)
- 7. *Network routing equipment is donated to schools.* This program is similar to the PC donation program. The savings are lower since the routing equipment is less expensive.
- Estimated savings: \$221 million to \$425 million (onetime)
- 8. *Network servers are donated to schools.* This program is similar to the PC donation and router donation programs.
- Estimated savings: \$370 million to \$1.5 billion (onetime)
- 9. *Internet connectivity is made free to schools.* There are great potential cost savings if schools are given Internet access at no cost. This plan could be arranged either by provision from an Internet service provider or from a local university or community college that has its own Internet connection.
- Estimated savings: \$150 million to \$630 million (annual)

Table 8 summarizes the potential savings for U.S. schools nationwide from each of the possible programs.

TABLE 8 Total Estimated Savings for U.S. Schools Benefiting from Various Possible Cost-savings Programs for Network Connectivity (millions of dollars)

Type of Program	Onetime Savings		Ongoing Savings	
	Low	High	Low	High
Reduced telecom rates (30 percent reduction)	89	218	39	150
Reduced telecom rates (60 percent reduction)	179	435	78	300
Purchasing by states (30 percent reduction)	1,889	4,136	45	189
Support from universities	0	0	790	1,800
Teachers trained on own time	0	1,500	0	300
Free labor for installing network	1,125	3,088	0	0
Donation of PCs	5,100	10,200	0	0
Donation of routers and CSU/DSUs	221	425	0	0
Donation of servers	370	1,500	0	0
Free Internet connectivity	0	0	150	630

SOURCE: Rothstein (1994).

### CONCLUSIONS

With a clearer picture of the costs for connecting schools to the NII, a number of conclusions may be drawn:

- *The costs to network a school are complex.* It is not simple to estimate the costs for connecting a particular school to the network. The costs for most schools will fall into a bounded range, but each particular school will vary greatly, depending on its individual needs and characteristics. Although this analysis puts bounds on the cost figures, the numbers are rough estimates at best.
- *The cost of the network hardware is only a small fraction of the overall costs for connecting to the NII.* Initial training and retrofitting are the largest onetime costs for starting connectivity to the network. The costs for the wiring and equipment are typically not as high. Support of the network is the largest ongoing annual cost that schools must face.
- *There are two major jumps in the costs to network a school.* The jumps occur in the transitions from model 1 to model 2 and from model 4 to model 5, as illustrated in [Figure 8](#). The first jump in cost occurs when the school installs a LAN. At that point the school and district must pay to have the network installed (\$20,000 to \$55,000 per school) and employ full-time network support staff (\$60,000 to \$150,000 per school district). The second jump occurs if and when the school decides to purchase computers for all students to use. The number of networkable PCs in 1994 is inadequate for most schools; hundreds of thousands of dollars would be needed to provide multiple PCs in every classroom. Also, many schools will need major electrical work (possibly exceeding \$100,000 each) to support the increased number of PCs in the school. In the intermediate stages between these jumps, the costs are incremental and relatively small.
- *The start-up costs for connection to the network increase at a faster rate than the annual ongoing costs as the complexity of network connection increases.* In the less complex models, the onetime start-up costs are 2 to 3 times the annual ongoing costs of the network. However, for the more complex models (models four and five,) the onetime costs are 5 to 15 times the costs to start connecting to the network. The differences are illustrated in [Figure 9](#). The divergence indicates that the most significant cost hurdle that a school will face is the initial investment in the network and computers. Dispensers of educational funding should be aware of this circumstance, so that they can help schools overcome the initial barrier. Schools should be given flexibility to amortize initial costs, to spread out the burden over a number of years.

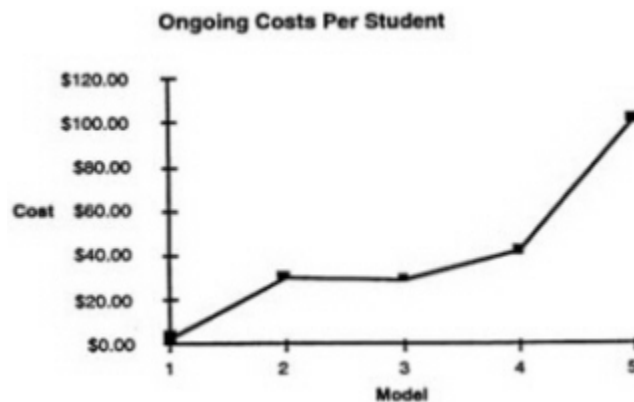


Figure 8 Ongoing costs per student with increasing complexity of network connectivity. SOURCE: Massachusetts Institute of Technology, Research Program on Communications Policy (1994).

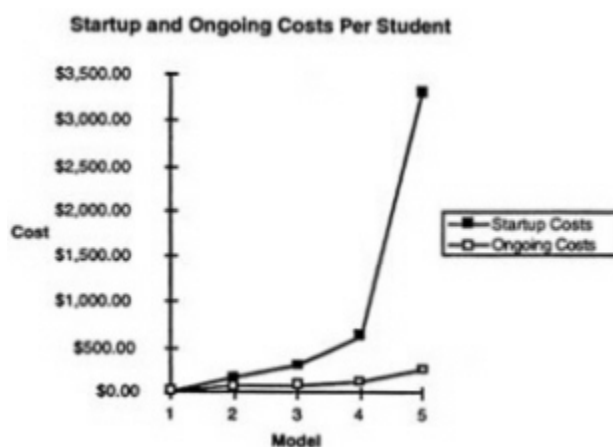


Figure 9 Start-up and ongoing costs per student with increasing complexity of network connectivity. SOURCE: Massachusetts Institute of Technology, Research Program on Communications Policy (1994).

- *Costs are significantly reduced when aggregated at the district and state levels.* Schools stand to save a lot of money by pooling resources and purchasing power with other schools at the district and state levels. When schools share a high-speed data link, or support staff, the per-school costs drop considerably. Schools in North Carolina and Kentucky have saved 20 to 50 percent by purchasing services and equipment at the state level.

Further research on the costs of wireless and cable Internet access methods for schools is recommended to elucidate the costs and benefits of these approaches. In addition, the issue of software and equipment cost accounting requires further analysis. This preliminary assessment of the costs of connecting schools to the NII is intended as a point of departure for analysis of these and other more detailed models of NII connectivity.

## Bibliography

Anderson, Ronald E. (ed.) 1993. *Computers in American Schools, 1992: An overview*. IEA Computers in Education Study, University of Minnesota, Minneapolis.

Burns, P. 1994. *Models and Associated Costs for Connecting K-12 to the Internet*. Draft white paper available from the University of Colorado.

California Department of Education. 1994. *Building the Future: K-12 Network Technology Planning Guide*.

Carlitz, R, and E. Hastings. 1994. *Stages of Internet Connectivity for School Networking*. Common Knowledge: Pittsburgh White Paper.

Computer Science and Telecommunications Board, National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington, D.C.

Gargano, J., and D. Wasley. 1994. *K-12 Internetworking Guidelines*. Internet Engineering Task Force (IETF) draft paper.

Information Infrastructure Task Force. 1994. *Putting the Information Infrastructure to Work*. U.S. Government Printing Office, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

- Massachusetts Telecomputing Coalition. 1994. *Models for Connecting K-12 Schools to the Internet*. Draft white paper.
- Newman, D., S. Bernstein, and P. Reese. 1992. *Local Infrastructures for School Networking: Current Models and Prospects*. Bolt, Beranek and Newman Inc. Report No. 7726.
- Public School Forum of North Carolina. 1994. *Building the Foundation: Harnessing Technology for North Carolina Schools and Communities*.
- Rothstein, R. 1994. *Connecting K-12 Schools to the NII: Technology Models and Their Associated Costs*. U.S. Department of Education working paper.
- Rothstein, R., and L. McKnight. 1995. *Architecture and Costs of Connecting Schools to the NII*. Submission to *T.H.E. Journal*.
- U.S. Department of Education. 1993. *Digest of Education Statistics*. U.S. Government Printing Office, Washington, D.C.

## 58

# Geodata Interoperability: A Key NII Requirement

*David Schell, Lance McKee, and Kurt Buehler*

*Open GIS Consortium*

### STATEMENT OF THE PROBLEM

A wide range of activities depend on digital geographic data and geoprocessing, and these information resources are rapidly becoming more important as commerce expands, information technology advances, and environmental problems demand resolution. Unfortunately, noninteroperability severely limits the use of digital geographic information. Spatial data exist in a wide range of incompatible and often vendor-proprietary forms, and geographic information systems (GISs) usually exist in organizations as isolated collections of data, software, and user expertise.

The potential uses for geodata in the context of the national information infrastructure (NII) reach far beyond current uses, making current limitations even more disheartening. If legacy geodata, data generated by new technologies such as geopositioning systems, and high-resolution, satellite-borne sensors were easily accessible via networks, and if spatial data of various kinds were compatible with a wide range of desktop and embedded applications, the effects would be revolutionary in NII application areas such as emergency response, health and public safety, military command and control, fleet management, traffic management, precision farming, business geographics, and environmental management.

The problem of geodata noninteroperability will be solved in part by institutional cooperation: organizations need to create data locally in standard formats that can be used globally. But we need to plan a technological solution if we want systems in which diverse applications can transparently exchange diverse geodata types and can access remote spatial databases and spatial processing resources in real time. The technological solution is to develop a common approach to using spatial data with distributed processing technologies such as remote procedure calls and distributed objects.

In this paper we look at the components of the geodata noninteroperability problem and describe the technical solution and the unique consortium that is providing that solution. This consortium approach, which involves users in the specification process and which organizationally isolates technical and political agendas, holds promise as a model means of simultaneously developing important technologies and developing and implementing technology policy.

### BACKGROUND

#### **The Importance of Geospatial Information and Geoprocessing Services**

The NII broadly defines an information technology environment for the development and use of many information-based products of vital significance to the nation. Fundamental to the design and function of many of these products is the use of geospatial data, more commonly known as geographical information. So widespread and critical is the use of geographical information and the geographical information systems (GISs) that have been marketed to work with this information in its many forms that President Clinton, in Executive Order 12906 (April

11, 1994), established the National Spatial Data Infrastructure (NSDI). He commissioned the Federal Geographic Data Committee (FGDC; chaired by Secretary of the Interior Bruce Babbitt) to document the NSDI, to give it operational form, and to provide a management structure designed to help it grow and become a focus for the combined efforts of both public- and private-sector organizations concerned in any way with the development and use of geospatial information.

In terms of the clear objectives that give it form and sustain it, as well as its highly focused data development and standardization activities, the NSDI is positioned to be a well-defined and vital component of the NII and a valuable resource for those working in a variety of the application areas targeted for intensive development by NII policy planners. However, although the NSDI was conceived as a distributed data resource, it lacks the operational framework to support real-time access to distributed geoprocessing resources and their associated database archives. Resources such as the Wide Area Information Server (WAIS) and Government Information Locator Service (GILS) are available for catalog access and superficial browsing of spatial data sets. But the problem of remote query against the wildly heterogeneous assortment of spatial data sets that constitute the NSDI's "clearinghouse" environment will not be solved until an operational model for the interoperability of distributed geoprocessing environments is developed and accepted by the geoprocessing community.

The nation's wealth of geospatial data is vast and is used more and more by developers and planners at all levels of operation in both the public and private sectors. Much of the information technology world is rapidly transforming its basis of computation from the tabular domain (the world of spreadsheets and accounting ledgers) to the spatial domain (the world of maps, satellite imagery, and demographic distributions). Applications that merge digital mapping, position determination, and object icons are already being introduced in activities such as overnight delivery service and urban transit. These three technologies and new database management systems capable of handling multidimensional data will play an important role in operations decision support systems, maintenance management, and asset management wherever assets and processes are geographically dispersed.

As the NII concept challenges organizations to adopt more comprehensive, enterprise processing models based on wide-area, multimedia communications technologies, there is a growing need to invest the NSDI with distributed processing capabilities that can ensure the full integration of geoprocessing resources into these models. Commercial need for better spatial data integration is already clear in areas such as electric and gas utilities, rail transport, retailing, property insurance, real estate, precision farming, and airlines. Given the critical nature of applications positioned to combine "real-time" and geospatial attributes—emergency response, health and public safety, military command and control, fleet management, environmental monitoring—the need to accomplish the full integration of NSDI resources into the NII context has become increasingly urgent.

### **The Importance of Interoperability Standards to NII-based Applications**

Fundamental to the development of enterprise information systems is the concept of interoperability, which is used at all levels of information technology development to define a user's or a device's ability to access a variety of heterogeneous resources by means of a single, unchanging operational interface. At the level of chip technology, the interface is a backplane or a bus specification; at the network level the interface may be a hardware-based transmission protocol or a packet specification; at the operating system level, the interface is a set of system calls or subroutines; and at the object programming level, the interface is a specification for the behavior of object classes. Interoperability thus denotes the user's ability to function uniformly and without product modification in complex environments by applying a standard interface to heterogeneous resources.

In the geodata domain, interoperability is defined as the ability to access multiple, heterogeneous geoprocessing environments (either local or remote) by means of a single, unchanging software interface. Interoperability in this context also refers to accessing both multiple heterogeneous data sets and multiple heterogeneous GIS programs. By definition, this view of interoperability assumes the sort of interoperable network environment envisioned by NII policy.

The interoperability profile of the NII is characterized by multiple hardware and software standards, some complete and some under development. Such standards are the product of years of concentrated effort on the part of both public- and private-sector organizations, as well as such recognized standards bodies as the American

National Standards Institute (ANSI) and the International Organization for Standardization (ISO). Within the operational framework established by these standards, it should be possible to exchange diverse information encoded in a wide variety of data formats among all known makes and types of computer, communications, and video systems. Already significant strides have been taken toward integrating such forms of information as database tables, with the result that both database applications and nondatabase applications such as spreadsheet and word processing documents can use database information without switching applications.

To integrate geospatial data into the application environment of the NII and to use it in a coherent way, the FGDC has taken valuable first steps by organizing an approach to standardizing on an exchange format that includes entity attribute schemes for each of the significant geoprocessing application areas in the federal government and a standardized description of the content of geospatial data collections (i.e., their metadata). Associated with this effort is the establishment of the National Spatial Data Clearinghouse, a directory and metadata catalog accessible by the Internet describing a rich assortment of registered databases meant to define the resources of the NSDI. Taken together, the geodata standardization and Clearinghouse initiatives make up the data resources needed to introduce geoprocessing into the mainstream of NII-based distributed application systems. However, they do not address the interoperability issue. They remain available as a "static resource," requiring either that data be accessed by software that is familiar with its native format or that entire data sets be converted to standardized transfer formats and reconverted to the user's processing format before any useful work can be done with them. Many databases are frequently updated, which adds to the cost of acquisition and conversion.

To make the NSDI data both freely available and useful in the NII environment, it will be necessary to introduce a standardized interoperability mechanism specifically designed for remote access of geographical databases. Such a mechanism must embody the interoperability principles on which the NII is based. That is, it must enable transparent, barrier-free access to heterogeneous geospatial data sets in a distributed environment, relying on standardized interfaces for the generation of queries in real time and the consequent retrieval of query-derived data sets to the user's native environment.

A significant issue faced in recognizing the necessity of such an interoperability mechanism for geoprocessing is that the geoprocessing community has been slow to adapt many recent advances in mainstream information technology. Traditionally, GIS packages (whether commercial or public sector) have created highly structured and architecturally closed operational environments, tightly coupling display graphics with spatial analysis mechanisms, and relying on a tight coupling of spatial analysis with proprietary spatial database designs. Such packages tend to be operationally monolithic, lacking the modularity required for an efficient use of the distributed architectures that are more and more coming to characterize enterprise computing. In the past, the GIS user worked primarily in a closed-shop environment on limited data sets that could be archived and maintained in the local environment and only intermittently updated by means of tape import or batch transfer over the network. With the advent of the Internet and the associated surging demand for "processed data" in real time, such closed environments fail to provide geoprocessing professionals with the same ready access to the nation's data resources as people in general are coming to expect of such institutions as libraries, legal archives, news and entertainment programming, and general information and interaction services on the Internet.

## ANALYSIS AND FORECAST

### Contingencies and Uncertainties

The growth of the NSDI subset of the NII, like the growth of the NII itself, is a complex, many-faceted phenomenon. As in a developing economy or ecosystem, there are many complex dependencies. Organized cooperative effort will characterize the growth in some domains; chance, raw market initiatives, and surprise breakthroughs will dominate in others. The rise of geodata interoperability depends on the expertise, commitment, and user focus of those who are directly addressing the problem, but it also depends on the infrastructures that support their efforts and on the vagaries of the market.



In this section we describe a geodata interoperability standards effort in progress and note the other standards and new technologies on which it depends. We also describe the process by which the standard is being developed, and the participation essential to that process.

### **The OGIS**

The Open Geodata Interoperability Specification (OGIS) is a specification for object-oriented definitions of geodata that will enable development of true distributed geoprocessing across large networks as well as development of geodata interoperability solutions. OGIS is like other emerging distributed object-oriented software systems in its basic structure and benefits, but it is the first large-scale application of object technology to GIS and to the management of spatial data in the global information infrastructure (GII) and NII contexts. It is being made sufficiently general so that it can be implemented using software methodologies other than object technology, including remote procedure call (RPC) architectures such as the Open Software Foundation's Distributed Computing Environment (DCE) or application linking schemes such as Microsoft's Object Linking and Embedding (OLE).

The OGIS will specify a well-ordered environment designed to simplify the work of both developers and users. Developers adhering to OGIS-defined interface standards will easily create applications able to handle a full range of geodata types and geoprocessing functions. Users of geodata will be able to share a huge networked data space in which all spatial data conforms to a generic model, even though the data may have been produced at different times by unrelated groups using different production systems for various purposes. In many cases automated methods will bring older data into conformance. The OGIS, and the object technology that underlies it, will also provide the means to create an extremely capable resource browser that users will employ across networks to find and acquire both data and processing resources. Searches will be executed using powerful, object-oriented distributed database technology developed to handle large, complex, abstract entities that cannot be managed in conventional relational database management systems. Queries will potentially return only the data requested, not whole data sets that will require further reduction and preparation before the user can begin.

With the OGIS in place, the GIS industry will have the tools it needs to address the expensive geodata incompatibility problems faced by federal agencies, large private enterprises, and state and local governments. The OGIS will provide a technological boost to organizations committed to spatial data transfer standards. By providing a "smart" generic wrapper around data, OGIS will achieve the objectives of a universal data format while respecting the widely varying missions of data producers. Making good use of distributed processing on high-bandwidth networks, OGIS will multiply the utility of geospatial databases and reduce the size and duration of data transfers.

### **Object Technology**

To understand how the OGIS will work, one must understand the basics of object technology. Object technology will, over the next 5 years, change the way we conceptualize, build, use, and evolve computer systems. It provides a way to integrate incompatible computer resources and a way to build software systems that are much easier to maintain, change, and expand. Its robust components can be quickly assembled to create new or modified applications. It is synergistic with the information superhighway, offering a model in which adaptable agents spawned by a user's local computer can act across the network. In harmony with today's emphasis on sharing of data and resources within enterprises, it breaks out of the earlier model that sequesters proprietary software and data within isolated systems. Object technology will make it much easier for nontechnical people to access and customize information and information systems, because many instances of data conversion and manipulation will become transparent.

The old way of programming, procedural programming, segregates data from the programs that operate on the data. Procedural programming makes sense when processing resources are in short supply and programmers are available in abundance to maintain and "debug" aging software systems that have grown into

"spaghetti code." The world is moving away from this model. Processing power is abundant and cheap, and large enterprises now look to object technology to reduce escalating software maintenance costs (and data conversion costs) that are straining their budgets.

Moving beyond stand-alone programs that operate on specific kinds of data to perform specific tasks, developers of large enterprise systems are now beginning to write software in special modules called objects. Objects typically include both data and behaviors. Data are spoken of as being "encapsulated" in an object. The capsule that contains the data is a set of behaviors (or procedures, or methods) that can act on the data and return a result when requested to do so by another object. If an object is requested to do something it cannot do, it returns an error message. Older, preexisting data and applications can be encapsulated so that they will work in the object environment.

Client objects in distributed object systems can learn other objects' contents and capabilities and invoke operations associated with those capabilities. In other words, objects interact as clients and servers. The object-to-object messaging that is central to object technology gains efficiency through a system by which objects belong to classes with common properties. Classes can be nested in hierarchies. In these hierarchies, classes inherit attributes (data and procedures) from classes that are higher in the hierarchy. Inheritance allows objects to efficiently represent the large amount of redundant information that is common to all the objects of a class.

Large object systems, especially distributed systems, typically have an interface layer called an object request broker (ORB) that keeps track of the objects and activities in the system. The ORB screens for improper requests, mediates between competing requests, makes request handling more efficient, and provides an interface through which dissimilar applications can communicate and interoperate.

Just as a Microsoft Word user can now run analyses on a Microsoft Excel spreadsheet embedded in a Word document without changing applications, a GIS user will access (through the development of the OGIS and the object technology environment) a full set of GIS capabilities while working in an application that may not be a GIS application. Just as the Word user no longer needs to convert the spreadsheet to tab-delimited text before importing the static data into the Word document, the GIS user will not need to convert data formats and projections.

### **The OGIS Architecture**

The OGIS will provide the following:

- A single, "universal" spatiotemporal data and process model that will cover all existing and potential spatiotemporal applications;
- A specification for each of the major database languages to implement the OGIS data model; and
- A specification for each of the major distributed computing environments to implement the OGIS process model.

By providing these interoperability standards, the OGIS will also provide the means for creating the following:

- An interoperable application environment consisting of a configurable user workbench supplying the specific tools and data necessary to solve a problem;
- A shared data space and a generic data model supporting a variety of analytical and cartographic applications; and
- An interoperable resource browser to explore and access information and analytical resources available on a network.

The two main components of the architectural framework of the OGIS are the OGIS geodata model (OGM) and the OGIS reference model (ORM).

The OGM is the core component of the framework. It consists of a hierarchical class library of geographic information data types that make up the shared data environment and unified data programming interface for applications. Cohesive, comprehensive, and extensible, encompassing all current geospatial and temporal descriptions of data, presentation issues, analysis modes, and storage and communication issues, the OGM will be the language of OGIS. The OGM will be the interface to geodata transfer standards such as the Spatial Data Transfer Standard (SDTS), Spatial Archive and Interchange Format (SAIF), and Digital Geographic Standard (DIGEST). It will be interoperable with SQL3-MM, the next generation of SQL, which will be object oriented and will support multimedia entities, including geospatial data.

The OGM will support object queries, read/write of multiple formats, temporal modeling, and long-duration transactions. To accomplish these objectives, it will include sophisticated GIS definitions of spatial objects, fields, and functions.

The ORM describes a consistent open development environment characterized by a reusable object code base and a set of services. The design approach for the OGM determines the set of services that must be supported in the ORM. As a result, the ORM requires directories of services and databases, which will support complex query processing. It also specifies standard methods for requesting and delivering geospatial transformations and processing tasks. The ORM will also facilitate transformations between "private" data and OGM constructs, as well as coordinate conversion and raster/vector conversion. It also manages visualization and display and supports data acquisition.

### **Other Emergent IT Standards and their Relationship to OGIS**

As mentioned previously, implementations of the OGIS will be layered on interfaces such as CORBA, DCE, and OLE. These are, in fact, the three major interfaces that are likely to be used. The success of the OGIS will therefore ultimately depend on the robustness, completeness, stability, schedule, and market acceptance of these standards. CORBA and DCE are the products of consortia; OLE is the product of Microsoft. The success of a consortium-produced standard depends greatly on whether the members put in enough time and money to finish the standard in a timely and complete fashion and whether the members understand the needs of the community that might use the standard. The success of a vendor-produced de facto standard like DCE depends greatly on the vendor's market penetration. The success of any standard depends on whether the standard solves a problem for a sufficient number of users, with sufficient cost-benefit advantage to make the standards-based solution more attractive than alternative nonstandards-based solutions.

The robustness, completeness, stability, schedule, and market acceptance of other aspects of the NII also bear on the success of the OGIS. Everything from the installation schedule of broadband cable to the security of network-based payment schemes will affect the demand for geodata and remote geoprocessing. Reciprocally, working geodata and geoprocessing solutions will help drive the demand for bandwidth, payment schemes, and network-ready devices, as well as the demand for more data and applications.

In the domain of GIS, the standards work of the FGDC and other groups focused on institutional solutions to interoperability will contribute to acceptance of OGIS-based solutions. These groups are focusing the attention of tens of thousands of traditional GIS users. These users will be motivated to publish their data electronically, and their need for data will simultaneously increase the demand for remote access to the data of others. The expense of manually bringing data into conformance will, we believe, often be greater than the expense of using OGIS-based automated methods, thereby increasing the demand for OGIS-based solutions.

### **Factors in Making the Business Case: Who will Build the OGIS and Why?**

The Open GIS Consortium, Inc. (OGC) is a unique membership organization dedicated to open system approaches to geoprocessing. By means of its consensus building and technology development activities, OGC has had a significant impact on the geodata standards community and has successfully promoted the vision of

"open GIS" as the vehicle for integration of geoprocessing with the distributed architectures positioned to define the emerging worldwide infrastructure for information management.

OGC's direction is set by a board of directors selected to represent key constituencies in the geoprocessing community. The OGC board speaks on behalf of both public- and private-sector users interested in finding more integrated and effective ways to use the world's increasing wealth of geographical information to support problem solving in such areas as environmental monitoring and sustainment, transportation, resource management, global mapping, agricultural productivity, crisis management, and national defense.

The OGIS Project Technical Committee of the OGC operates according to a formal consensus process structured to be fair and equitable and to ensure the technical completeness of the specification. To ensure the eventual adoption of OGIS as an official standard, the OGIS Project Technical Committee is represented on key geodata, GIS, and geomatics standards committees, including the International Organization for Standardization (ISO) TC211 GIS/Geomatics Committee and the American National Standards Institute (ANSI) X3L1 Committee. In addition, the OGIS Technical Committee maintains close ties to the Federal Geographic Data Committee (FGDC).

The OGIS Project Management Committee is composed of representatives from the OGIS Project Technical Committee, representatives of the principal member organizations, and others who represent particular constituencies. The OGIS Project Management Committee maintains a business plan for the project and sets overall policy for the project. The dual committee structure serves to separate technical and political issues.

OGC was founded to create interoperability specifications in response to widespread recognition of the following problematical conditions in the geoprocessing and geographic information community:

- The multiplicity of geodata formats and data structures, often proprietary, that prevent interoperability and thereby limit commercial opportunity and government effectiveness;
- The need to coordinate activities of the public and private sectors in producing standardized approaches to specifying geoprocessing requirements for public-sector procurements;
- The need to create greater public access to public geospatial data sources;
- The need to preserve the value of legacy GIS systems and legacy geodata;
- The need to incorporate geoprocessing and geographic information resources in the framework of NII initiatives;
- The need to synchronize geoprocessing technology with emerging information technology (IT) standards based on open system and distributed processing concepts; and
- The need to involve international corporations in the development and communication of geoprocessing standards activity (particularly in the areas of infrastructure architecture and interoperability) to promote the integration of resources in the context of global information infrastructure initiatives.

OGC is not-for-profit corporation supported by consortium membership fees, development partnerships, and cooperative agreements with federal agencies. As of April 26, 1995, the consortium included 40 members. Though organized to manage multiple project tracks, OGC's initial focus is on the development of the OGIS. OGC plans to establish other project tracks in areas related to implementations of the OGIS architecture.

### **Who will Use the OGIS and Why?**

The OGIS will be used by at least the following groups:

- Software vendors have already begun writing software that conforms to the object-based paradigm exemplified by the OGIS. GIS database, visualization, desktop mapping, and application tool vendors will all be able to provide capabilities that will seamlessly integrate into applications targeted for specific end-user communities. Just as the new paradigm will transform monolithic desktop applications into environments of compatible components (such as printing modules that work with a spreadsheet or word processor, making it unnecessary for the spreadsheet or word processor to contain its own printing software), OGIS-based components

will perform specific analysis and display functions, and they will work seamlessly with other, non-GIS applications.

- Integrators working on systems that require the ability to access geodata and geoprocessing resources are already building OGIS compliance into their frameworks.
- Data providers, public and private, who are preparing to serve data across networks, are planning and prototyping encapsulation schemes and database approaches that will depend on the OGIS. This group will ultimately include, for example, public digital libraries; remote sensing data vendors like SPOT and EOSAT; state and federal agencies that need to distribute data to agency offices and/or that are obliged to make their data publicly available; major corporations with spatial data in their corporate databases; and military intelligence, planning, and training groups.

Ultimately, the OGIS will be an invisible but essential enabler of ready access to remote geodata in all of the application areas mentioned here, and probably in others that we have not imagined. Like other standards, it will not be noticed or understood by most people who use it. It will make an extraordinary complex activity simple.

### **What Can the Federal Government Do to Facilitate the Process?**

Many agencies of the federal government have a critical interest in geodata interoperability, and some are already providing support to OGC. Though most of OGC's funding and technical participation comes from the private sector, OGC was started with funding from the U.S. Army Corps of Engineers Construction Research Laboratories (USACERL) and the DOA Natural Resources Conservation Service (NRCS). Both of these agencies are still involved with OGC. Also, the Universal Spatial Data Access Consortium (USDAC), a NASA-funded digital library technology project, has joined the OGIS Testbed Program of OGC. Funding for USDAC comes from a Cooperative Agreement Notice, "Public Use of Earth and Space Science Data Over the Internet," provided by NASA's High Performance Computing and Communications Program. Other federal agencies that are members include U.S. DOC NOAA—Nautical Charting Division, U.S. DOD Defense Mapping Agency (DMA), and U.S. DOI Geological Survey—National Mapping Division.

Other federal agencies with a critical interest in geodata interoperability would benefit from participating in the development of the OGIS, because their specific needs would be represented and they would be in close touch with the organizations best able to help them. Also, their support would help ensure timely completion of the OGIS, and as a result they would begin to reap the benefits of OGIS-based solutions as early as possible.

The federal government can also encourage other distributed computing standards efforts on which the OGIS will depend, such as the Object Management Group's Common Object Request Broker Architecture (CORBA). The OGIS will have value in nondistributed environments, but its greatest potential lies in the potential for remote geodata access.

Perhaps the most important way the federal government can help the cause of geodata interoperability—which is to say, to bring the NSDI into the NII—is to make the OGIS project a special focus for the national Information Infrastructure Task Force (IITF). Technology policy at the IITF level needs to stress private sector support for the standards activities of OGC. The IITF is positioned to strongly encourage major corporations (particularly telecommunications companies) to put significant resources into supporting OGIS development work, testbed activities, and, finally, OGIS-based products and services. The IITF can provide solid leadership in showing major corporations why this kind of interoperability ought to be considered a high-priority development issue.

### **Projections, Including Barriers**

OGIS-based implementations will become available in the same time frame in which underlying and auxiliary NII capabilities will become available. Early prototype work has begun at testbed sites, even though the

specification is not complete. The specification will be complete in about a year, and several testbed projects and independent efforts by vendors are expected to begin yielding useful implementations at about that time. The OGIS merely prepares the world of geodata and geoprocessing to participate in the general progress of information technology (IT) that industry experts forecast. The OGIS is being developed concurrently with other essential software standards and components of the NII that will be developed, released, and tried in the real world of the marketplace in the next 5 years.

Dissension in the software community is the principal barrier to this general progress and to the progress of geodata interoperability. The long history of UNIX shows how industry giants seeking dominance can thwart standards efforts and stunt the growth and acceptance of a useful technology. Every standards consortium confronts the tendency for industry leaders to fly apart because of competitive issues. From the standpoint of the OGIS, it is important that underlying standards like CORBA and DCE succeed and that the members of OGC stay true to the vision of geodata interoperability across the industry.

## RECOMMENDATIONS

### **How Can the Problem of Geodata Noninteroperability Be Addressed by a Public and Private Partnership?**

As described above, the problem of geodata noninteroperability is being addressed by a standards consortium, the Open GIS Consortium. OGC is developing not another geodata standard but rather a standard approach to using distributed computing methodologies, primarily object technology and RPCs, in geodata and geoprocessing applications. Participation by public and private organizations is particularly important in this case because so many public agencies are critically dependent on geodata and geoprocessing, and because most geodata is a product of public agencies. The public agencies have everything to gain and nothing to lose from their support of OGC, because (1) the OGIS Project gives them an opportunity to increase the certainty that the specification, and solutions based on it, will meet their needs, (2) agency representatives have the opportunity to learn about and learn from the most knowledgeable technology providers, and (3) the capabilities unleashed by the OGIS will enable them to vastly improve their ability to fulfill their missions.

OGC members, public and private, recognize the wonderful opportunity afforded by the timing of this effort: No vendor has yet dominated the market with object-based GIS solutions that might hinder the effort to create a rational, comprehensive, standardized approach that can potentially benefit everyone. Through the consortium, vendors and integrators have an opportunity both to involve sophisticated users in their common technology specification effort and to share the costs of this development. All of the members believe that their collective efforts will lead to greatly expanded markets, though they also recognize that their existing products and services will need to change to accommodate the new paradigm. Most understand that change is inevitable in the IT industry whether they embrace it or not, and OGC offers a way for them to exert a degree of control, stay at the leading edge, and make relatively sound business projections.

OGC's goal is to create successful standards for geodata interoperability, and many of the discussions surrounding its founding focused on the failings of other consortia that tried and failed in their attempts to create standards. Leading consortium lawyers and experts in technology standardization contributed to the formation of OGC and its bylaws, and their guidance continues. The challenge is to carefully craft an organization and a set of goals that make all participants winners, and to be sure that all the interests of all affected constituencies are represented.

OGC has a unique organizational structure that resulted from the process described above. It has a board of directors, a full-time professional staff, and separate project tracks. The board of directors has responsibility for approving business plans from each track, for organizing new tracks, and for setting overall consortium direction and strategy. Each project track has its own membership and is managed by a management committee that is responsible for strategic and business planning, technical oversight, and product approval. (OGC products are IT standards related to geodata interoperability.) Each track also has a technical committee responsible for technology development and mediation of the different technical views of the membership. A testbed program in

each track coordinates early implementation efforts and provides feedback to the technical committee. The technical committee is responsible for product development, which it then delivers to the management committee for final approval. This organizational structure isolates technical issues from management issues and allows a better overall progression of product development. It also isolates the board of directors from each project track.

The IITF is positioned to help OGC succeed by encouraging key NII-building companies to participate in OGC's efforts. Clearly, the technology policy mission of the IITF is perfectly aligned with the mission of OGC. But OGC also offers the IITF a higher-level benefit: the opportunity to observe and evaluate a state-of-the-art consortium that may be the forerunner of tomorrow's technology policy planning bodies. The continuing acceleration of technological change forces a new, more anticipatory and proactive approach to technology policy planning, an approach based on community-wide discussion of objectives followed by development of standards that channel vendors' efforts in agreed-upon directions.

### REFERENCES

- Committee on Applications and Technology, Information Infrastructure Task Force. 1994. *The Information infrastructure: Reaching Society's Goals*. U.S. Government Printing Office, Washington, D.C., September.
- National Institute of Standards and Technology. 1994. *Putting the Information Infrastructure to Work: A Report of the Information Infrastructure Task Force Committee on Applications and Technology*. SP857. National Institute of Standards and Technology, Gaithersburg, Md., May.

## 59

# Electronic Commerce

*Dan Schutzer*  
*Citibank Corporation*

Electronic commerce is the ability to perform transactions involving the exchange of goods or services between two or more parties using electronic tools and techniques. It offers many advantages over traditional paper-based commerce:

- It provides the customer with more choices and customization options by better integrating the design and production processes with the delivery of products and services;
- It decreases the time and cost of search and discovery, both in terms of customers finding products and services (e.g., shopping, navigating) and companies finding customers (e.g., advertising, target marketing);
- It expands the marketplace from local and regional markets to national and international markets with minimal capital outlay, equipment, space, or staff;
- It reduces the time between the outlay of capital and the receipt of products and services (or vice versa);
- It permits just-in-time production and payments;
- It allows businesses to reduce overhead and inventory through increased automation and reduced processing times;
- It decreases the high transportation and labor costs of creating, processing, distributing, storing, and retrieving paper-based information and of identifying and negotiating with potential customers and suppliers;
- It enables (through automated information) production of a reliable, shareable historical database of design, marketing sales, and payment information; and
- It facilitates increased customer responsiveness, including on-demand delivery.

A convergence of several factors has recently lifted electronic commerce to a new level of utility and viability. These factors include the increased availability of communications and communications bandwidth, the reduced cost and increased user-friendliness of computers and communications, the growth of the Internet and online services, and the drive toward global competitiveness.

Currently, online purchases account for only 4 percent of total global purchases; online purchases via such precursors as the Internet are practically nonexistent. But electronic commerce is likely to grow dramatically over this decade. For example, it is predicted that within 6 years, global shoppers will use the national information infrastructure (NII) to purchase \$500 billion of goods and services; this represents almost 8 percent of current purchases worldwide. And by 2005, the number of NII-based transactions is expected to rise to 17 billion, which is almost half the number of transactions made in today's credit card market.

### ELECTRONIC COMMERCE TODAY

In recent years, great strides have been made to automate many of the labor-intensive paper-based aspects of commerce. Examples abound of corporations that use electronic data exchange (EDI), electronic mail (e-mail), electronic forms (e.g., for ordering or for contracting), and electronic catalogs, and of electronic financial networks that speed the transfer, settlement, and clearing of funds and other financial instruments. These



electronic tools and techniques provide many benefits to both customers and merchants. EDI standards, for example, enable fast, accurate information exchange between different automated systems in routine, relatively simple business transactions.

Unfortunately, despite the widespread use of numerous methods of electronic support, electronic commerce is still not the most common method of carrying out business transactions. This is so for several reasons. For one thing, most business transactions still require the physical exchange of paper documents and instruments, with the inherent costs and delays this represents. For another, current electronic commerce approaches are not sufficiently well integrated, secure, open, or easy to use:

- *Partial solutions.* Current electronic commerce implementations automate only a portion of the entire transaction process. For example, although ordering and distribution of an information-based product (such as an electronic magazine or a software program) can be nearly simultaneous, the supporting accounting and inventory information, payment, and actual funds transfer tend to lag, often by days. This time lag, and the resulting decoupling of the accounting and payment information from the ordering and delivery of goods and services, increases the transaction's credit risks. It also increases the likelihood of discrepancies between the various information sources, requiring expensive and time-consuming reconciliation. Finally, today electronic commerce implementations are costly to develop and operate. Their high cost of entry does not make them feasible for many of the more spontaneous, high-volume, low-value electronic transactions (e.g., the sale and distribution of electronic information-based products, such as magazine articles and photographs) envisioned for the future. A fully integrated electronic commerce solution would let users maximize their control over their cash flows; for instance, it would allow the majority of their funds to work for them in bank savings accounts and/or investments, and it would minimize cash shortfalls. It would also eliminate the gaps between ordering, distribution, and payment, enabling development of real-time links to recordkeeping and accounting systems with minimal transaction costs.
- *Rigid requirements.* Electronic commerce applications usually require highly structured protocols, previously established arrangements, and unique proprietary bilateral information exchanges. These protocols, arrangements, and exchanges for the most part involve dedicated lines and/or value-added networks (VANs) and batch processing. For example, EDI requires rigid agreements between the two or more transacting parties about the structure and meaning of data. These agreements are often time-consuming to negotiate, inflexible, and difficult to maintain, especially in a rapidly changing business environment. The resulting costs and necessary lead times frequently create barriers to investment in and widespread use of electronic commerce applications by small and medium-size companies and inhibit the expansion of electronic commerce beyond large companies and their major trading partners.
- *Limited accessibility.* The consumer cannot usually communicate or transact with vendors in a simple, direct, free-form environment in today's electronic commerce applications. For example, to access most electronic shopping services, a consumer must subscribe to an online service (e.g., Prodigy or cable TV shopping channels) that then provides proprietary hardware and/or software with which to communicate with the vendors that have also registered with that service.
- *Limited interoperability.* Most current implementations depend on proprietary solutions, which do not easily interoperate, if at all. Internet e-mail and the World Wide Web are notable exceptions. A truly interoperable electronic commerce infrastructure would allow parties to conduct their transactions in private, without paying any fees to intermediaries unless they provide some real added value, such as credit services. This infrastructure would make it easier for any and all interested persons to become service providers as well as consumers.
- *Insufficient security.* The lack of personal contact and the anonymity associated with doing commerce over a telecommunications network make it difficult to authenticate parties and detect intruders; this in turn makes the system vulnerable to fraud and increases the need for security services. Additionally, the speed with which electronic commerce can be conducted leaves parties with less to react, check, and respond appropriately, again creating the potential for system fraud and abuse. Lack of sufficient security inhibits the introduction of direct, secure, real-time electronic payment and settlement systems that can support secure exchanges without prearrangements or third parties.

- *Inadequate search capabilities.* Participants in today's electronic commerce applications must find methods and means of navigating effectively through the sea of rapidly increasing online electronic information and services to find trading partners and items of interest. This problem will only increase as more information and businesses go online.

### **ELECTRONIC COMMERCE TOMORROW: BUILDING THE INFRASTRUCTURE**

Many new systems and service initiatives have been announced for the Internet and the evolving NII that address one or more of these current deficiencies to varying degrees. These include initiatives such as (1) CommerceNet and Secure Hypertext Transport Protocol (SHTTP), (2) electronic catalogs, (3) advanced search engines, (4) e-mail-enabled EDI, and (5) digital cash. Additionally, several alliances and partnerships have been announced that address the need for secure, affordable payment, linked in real time to ordering and billing systems. These initiatives include ventures by Open Market, Microsoft/Visa, Netscape/First Data/MasterCard/Bank of America, Cybercash/Wells Fargo, American Express/America Online, First Virtual/EDS, NetBill, NetCash, Cafe-Digicash, Mondex, NetCheque, NetAccount, Netchex, AT&T, and Internet MCI. More such alliances and start-ups are announced each day. These initiatives promise to accelerate the future growth of electronic commerce and rapidly decrease the overhead and time associated with today's paper- and people-intensive activities.

So, in the near future, we should see a broad range of new value-added electronic commerce services, including the following built around trust and security:

- Authentication over public networks;
- Certification of information, parties, and transactions;
- Performance bonding;
- Electronic escrow;
- Automated dispute resolution;
- Transaction insurance;
- Appraisal services;
- Various electronic broker services; and
- Trusted agents to resolve disputes and claims.

All of these initiatives must be developed within a common framework, or we run the risk of creating isolated, noninteroperable implementations that will inhibit progress toward truly free, open, and spontaneous electronic commerce. The joint ventures listed here, for instance, all vary in their approach to security and privacy, their ability to handle micropayments, and their applicability to various types of transactions. They also differ in their business models—for example, in their pricing strategy and in their assumptions as to who bears the risk in case of insufficient funds or disputes.

The electronic commerce infrastructure must therefore do the following:

- *Allow for interoperability.* The infrastructure must be based on a common set of services and standards that ensure interoperability. Preferably, these services and standards can be used as standard building blocks that service providers and application designers can combine, enhance, and customize.
- *Allow for maximum flexibility to permit innovation.* As the NII evolves, it will grow and mature significantly, possibly in ways not even imaginable today. As it grows, new services and businesses will emerge. For example, NII's electronic marketplace will provide new opportunities for narrow-case marketing to short-lived niche markets. Also, existing services and products will be redefined and modified. The electronic commerce infrastructure will have to be sufficiently flexible to accommodate all of these changes and be able to address new applications and new requirements as they arise.

## ISSUES RELATED TO ELECTRONIC COMMERCE ACTIVITIES

Some related issues important to the design of an electronic commerce framework are discussed below:

- *Nature of information products.* A particularly important class of products on the NII are products that are pure information. In the information age, a large percent of commerce will never be embodied physically. Information products are enabled by information technology and not just distributed more efficiently by it. These products can include not just electronic publications, catalogs, videos, and the like, but also interactive video games, software programs, electronic keys and tokens, customized design specifications, and even electronic keys that can open hotel rooms, cars, storage compartments, and airport boarding gates. Furthermore, these information products are not created entirely by the service provider but can be designed or customized by the customer (e.g., customers can create their own selection of articles to be bound in an electronic book, or their own custom clothing design), adding a customer-driven activity call designed to fit with the purchase cycle. It is also likely that, for these products, ordering, billing, payment, and distribution would likely all happen simultaneously.
- *Advanced revenue collection methods.* The electronic commerce infrastructure will need to support advanced types of revenue collection, in addition to traditional methods (e.g., payment upon receipt, payment in advance). For example, an information product service provider could distribute its product widely and charge on a usage basis—that is, charge the user only when the information (e.g., a software program, a digital document, an electronic key that can open and start a rental car) is used. One innovative approach that permits usage accounting and payment is called "meterware." It provides local hardware and/or software to continuously record and bill customers based on their usage. Meterware and other advanced revenue collection ideas (e.g., payment schemes, such as electronic cash and electronic checks, which do not require the presence of an online payment processor) create opportunities for reaching new customers and for distributing products and services; these make great sense in a low- or zero-distribution cost environment and should be supported by the electronic commerce infrastructure.
- *Transaction devices.* Electronic commerce transactions currently involve all manner of devices (e.g., telephone, fax, point-of-sale device), media (e.g., electronic data, image, voice, paper), and networks (cable, wire, satellite, cellular) over which they will be delivered. As a result, the system infrastructure must accommodate all of these devices, media, and communications networks, without degrading them to the lowest common denominator.
- *Legacy systems.* The electronic commerce domain has a large quantity of legacy systems that it needs to interface to and ultimately phase out of as it evolves to more modern systems, applications, and processes. These legacy systems and processes (e.g., paper checks, mainframe-based settlement and payment systems, and EDI VANs) will not be replaced overnight. A successful electronic commerce infrastructure must allow the user to easily and transparently transfer between and switch back and forth between the new, all-electronic and the older, hybrid legacy systems and processes.
- *Public data.* Finally, libraries of publicly available and accessible files of registered contracts, financial reports, and holdings, as well as catalogs and lists of products, data, and services, if made part of the NII domain-specific infrastructure being offered by various electronic commerce vendors and information providers, could be interfaced with this electronic commerce scenario to further enrich it.

As new initiatives start up under a common framework and set of standards, performed over low-cost commodity computers linked by open public networks, competitive pressures and technology advances should drive down associated costs and time and increase interoperability and competitiveness. New forms of commerce will arise that were impractical under the old cost structure, and the virtual enterprise will be fully realized. Consider the following examples of what life might be like in the future NII if this electronic commerce infrastructure is realized.

## Sample Commercial Scenario

### Commercial Example

Mary wants to lease some warehouse space in a distant city that she will use as a distribution hub. Her main concerns are location, price, and early occupancy. She clicks an icon displayed on her PC that establishes an online connection to several broker icons advertised on the World Wide Web (WWW). She clicks on each of these broker icons in turn to scan their home pages. Soon, she finds a property for lease that seems to match what she is looking for. The property is at the intersection of two major highways, has sufficient square footage, is being offered for a reasonably monthly price, and is available for lease at the beginning of the next month.

Mary checks the broker's credentials by clicking on a certification icon. The broker's credentials are immediately transmitted to a reference service for authentication. Within seconds, the reference service sends Mary an e-mail message with the certification attached; Mary is alerted to the arrival of this message by a special interrupt tone on her workstation.

Mary then clicks on the appropriate button, and an electronic application form to lease the property appears. She begins to fill out this form and inserts her electronic bank card into the PCMCIA slot in her PC, which digitally signs the application with her financial credentials and binds the application data to her signature so that it cannot be altered. Mary then clicks the send button to transmit the application to the broker. The application contains her identification, billing address, and permission for the broker to obtain a credit reference from her bank. Mary then closes the broker page and turns to another task.

Several minutes later, the broker has completed a review of Mary's application, including obtaining and analyzing her bank's credit reference. Most of this work was performed by intelligent software agents that were automatically activated and tasked upon receipt of Mary's electronic application form. Since Mary's background check was routine and acceptable, the application's approval was automatic. Accordingly, the broker can send Mary a standard rental contract. She receives this as an e-mail attachment concurrent with notification of acceptance of her application.

Mary inspects the contract and finds its terms generally acceptable, except for two items:

- Mary wants a 15-day grace period rather than the 5-day period specified in the contract; and
- Although she has no problem with her bank transferring payment electronically to the broker's bank account on the 15th of each month, she does not want this to occur without her first being notified, in case for some reason she needs to override and stop the automatic electronic funds transfer.

Mary e-mails these exception items to the broker, who reads and approves the changes, and makes one additional change: He specifies the conditions under which Mary is justified in stopping payment. This information is communicated via e-mail to Mary, who finds the conditions acceptable and date/time-stamps them, and again digitally signs the digital contractual agreement. She sends it back to the broker, who in turn digitally co-signs it. The process is witnessed by a network notary, who digitally signs the contract as witness; holds the original digital copies; and sends authenticated copies to Mary, the broker, their respective banks, and the appropriate government agency for filing and recording.

Now, each month Mary's bank notifies her by e-mail that it will electronically transfer funds from her account to the broker's account unless it receives override instructions from her within 24 hours. Mary completes the required acknowledgment of these electronic notices and date/time-stamps them. When the bank electronically withdraws funds from her account to meet the lease agreement, this information is automatically transmitted to the broker via e-mail. The broker's financial software agent automatically makes the electronic deposit of the funds. Intelligent agents for the bank and broker then automatically update all appropriate financial records.

### Sample Retail Consumer Scenario

Dave gets an e-mail inviting him to a formal black-tie dinner tomorrow when he returns from his business trip. Unfortunately, he does not own a tuxedo and is 3,000 miles from his home. In the past, he has ordered a rental tuxedo from a neighborhood store via computer. Since his last promotion, however, he has been getting invited regularly to formal affairs and can easily justify buying rather than renting a tux. He dials up his electronic yellow pages, searches for tuxedo manufacturers, explores their interactive multimedia advertisements, and finally finds a supplier to his liking. It allows him to customize his own tuxedo design online and promises 24-hour delivery to any location in the United States.

Dave requests linkage to the tuxedo manufacturer's interactive design facility. A range of charges that vary according to the quality of transmission (image resolution and response time) is displayed, and Dave is prompted to make a selection and to insert his American Advantage Citibank Visa smart card into his PC reader. Dave makes a selection and inserts the card, which issues a payment transfer from his credit card account to the electronic yellow pages, subject to Dave's approval and authorization. Dave approves the charge, digitally signs the authorization sealed with his unique personally encrypted retinal scan recorded by a camera mounted in his portable PC, and inserts his card, thereby transferring the payment. Dave is connected to the tuxedo manufacturer design service.

The tuxedo manufacturer requests that Dave transfer his dimensions. Dave does so easily, his dimensions being stored on a file in his PC. The manufacturer then transmits a number of images of typical tuxedo designs for Dave's selection. Dave enlarges a few of these images to full size, adjusted to his form and dimensions. He rotates them and views them in three-dimensional projection from many viewpoints, and selects one. He then proceeds to manipulate this design using his electronic pen; he makes minor adjustments and changes to the tuxedo, raising the waist and tapering the pants legs to better match his personal taste. He next specifies the material: 100 percent wool, treated for stain-proofing and wrinkle-proofing. When it is finished, Dave transmits the revised design to the manufacturer, along with the desired delivery date and location. The manufacturer replies with a charge appended to the design; this specifies the delivery date and location, and the terms and conditions of payment (either a credit card charge now, or a 20-percent-down cash transfer now with the remainder due in cash on delivery). Dave digitally signs the electronic agreement with his American Advantage Citibank Visa smart card, which he uses as payment, and sends it back to the manufacturer.

The manufacturer receives the order and payment, and issues the design specifications along with a unique order number to its factory for manufacture. At the same time, the manufacturer sends an order to Federal Express for pickup and delivery of the completed suit that afternoon; this order includes the suit's unique order number, delivery address, customer name, and identifying encrypted retinal scan. By 3:00 p.m., the completed tuxedo has been picked up by the messenger.

At 8:00 the next morning, Dave arrives at his office. Within 2 hours, a Federal Express messenger arrives at his office with the suit. Dave authenticates himself, matches his retinal scan, and receives the tuxedo in plenty of time for the evening's formal dinner.

### SCOPE OF PROPOSED ELECTRONIC COMMERCE INFRASTRUCTURE FRAMEWORK

To achieve the vision outlined above for electronic commerce, we need a comprehensive architectural framework and a set of base infrastructure services and standards around which these and future initiatives can be designed. The architecture must permit the flexibility, interoperability, and openness needed for the successful evolution of electronic commerce over the NII. This framework, and its service and products, will offer the consumer a diverse set of interoperable choices, rather than a collection of independent "stovepipe" solutions.

This framework should set the groundwork for developing the electronic commerce infrastructure. To this end, we begin by discussing the basic activities and functions associated with electronic commerce, identifying the key building blocks required to support those functions and activities, and describing key infrastructure services and application programming interfaces (APIs). These services and APIs are placed within an electronic commerce application architectural framework consisting of a services infrastructure layer and an

applications layer that interface with each other and with the physical layer. These layers are defined in terms of APIs and objects from which the various electronic commerce applications can be constructed.

### **Electronic Commerce Activities and Functions**

There are nine key activities in commerce:

- Advertising and shopping,
- Negotiating,
- Ordering,
- Billing,
- Payment and settlement,
- Distribution and receipt,
- Accounting,
- Customer service, and
- Information and knowledge processing.

The specific functions associated with these activities in an electronic commerce setting are discussed below. Note that not all of these activities are performed in every transaction, nor are they necessarily performed in this order; indeed, they may be performed in parallel. Also, the activities are not necessarily all conducted electronically. Finally, these activities can vary in complexity and importance depending on the size and scope of the transaction.

#### **Advertising and Shopping**

Advertising and shopping can include the following:

- A potential buyer browsing electronic yellow pages and catalogs on a network;
- An agent shopping on behalf of one or many buyers and/or sellers;
- A buyer sending an electronic request for proposal (RFP), and sellers responding with various offers;
- Sellers advertising their products and services; and
- Buyers electronically navigating and/or browsing through the World Wide Web's online services.

A major problem associated with the advertising and shopping activity is the cost and time expended in developing, maintaining, and finding relevant information, products, and services, given the plenitude of available information. Obviously, this problem will become increasingly complex as more data and services become available online and the choices and possibilities multiply exponentially. We need new and better ways to find services and information and to publish and update this information.

#### **Negotiating**

Buyers and sellers may elect to negotiate the terms of a transaction (i.e., the terms of exchange and payment). These terms may cover delivery, refund policies, arranging for credit, installment payments, copyright or license agreements, usage rights, distribution rights, and so on. These terms can be standardized for routine commodity use, or customized to suit unique individual situations. Often, in the case of two parties with a well-established business relationship, the terms of exchange are prenegotiated as standing contractual terms for all their future exchanges. Often, this process will also include authentication of the two parties.

### **Ordering**

The buyer eventually issues a contractual agreement of the terms of exchange and payment. This contractual agreement is generally issued as an order, which sets forth the quantity, price, and other terms of the transaction. The order may be verbal, in writing, or electronic. It usually includes an acknowledgment of agreement by the various parties in order to help prevent any future repudiation. This agreement can be confirmed electronically through cryptographic techniques such as digital signatures.

In the case of some commodity purchases, the entire transaction may begin at this ordering stage, bypassing the advertising/shopping and negotiating activities. The ordering activity applies to all transactions, regardless of whether billing will be involved. For example, even requests for free public information should be issued as formal orders so that the service provider can record and account for information requests.

### **Billing**

Once a seller has delivered goods or services, a bill is sent to the buyer. This bill generally includes remittance information that should accompany the payment. Sometimes, a seller may require payment in advance. Sometimes, a supplier sends advance shipping notification, and the customer agrees to authorize payment upon confirmation of the arrival of the products. And in some cases, as with the free information example cited above, this activity is eliminated entirely.

### **Payment and Settlement**

The buyer, or some financial intermediary, eventually sends some form of electronic payment (this could be some form of contract or obligation, such as authenticated payment instructions or digital cash), usually along with some remittance information to the seller. This payment may occur for a single item, on a usage basis, or with a single payment for multiple items or usage. Settlement occurs when the payment and remittance information are analyzed by the seller or the seller's agent and accepted as valid.

### **Distribution and Receipt**

Either before, after, or concurrent with payment, the seller arranges for delivery of the purchased goods or services to the buyer, and the buyer provides the seller with proof of receipt. Policies regarding customer satisfaction and return should be negotiated prior to this activity and made part of the contract between buyer and seller. For larger, more complex orders, distribution may involve more than two parties and entail complicated distribution coordination strategies. An ancillary distribution service involves acting as a fiduciary, and holding goods, certificates, bonds, stocks, and the like in trust.

### **Accounting**

This activity is particularly important to corporate customers and suppliers. Both buyer and seller must reconcile all electronic transactions in the accounts receivable and accounts payable, inventory information, and accounting systems. Account and management information system records must also be updated. This activity can involve third parties, if the transacting businesses outsource their accounting services.

### **Customer Service**

Customer service entails the following:

- Providing the buyer with timely information as to the progress of a transaction;
- Handling customer requirements when transactions go awry—that is, resolving any mistakes, disputes, or complaints concerning product quality, delivery, or payment (this includes managing returns and refunds, further exchanges, and/or repairs); and
- Providing expert advice and assistance in the use of the products and services.

Customer service also involves providing general cash management advice, including addressing foreign exchange imbalances and risk exposures, collection of delinquent payments and late fees, and repossessing products for which payment is long overdue.

### **Information and Knowledge Processing**

A final key activity in electronic commerce is the collection, management, analysis, and interpretation of the various data to make more intelligent and effective transaction-related decisions. Examples include collecting business references, coordinating and managing marketing strategies, determining new product offerings, granting and extending credit, and managing market risk. Performance of these tasks often involves the use of advanced information management techniques, including models and simulations and collaborative computing technologies to support conferencing and distributed workflow processes.

These tasks will become more difficult as the sources of information grow in number and are of increasingly diverse and uncertain quality. Additionally, procurement of this information may raise significant privacy concerns and issues.

### **A MODEL FOR ELECTRONIC COMMERCE**

To support the electronic commerce activities and functions discussed here, and to accelerate the electronic commerce vision described earlier, we need an open electronic commerce architecture and a set of agreed-upon electronic commerce infrastructure standards and services. These elements are detailed in the following sections; here we describe an overall model of electronic commerce.

Major electronic commerce applications can be built by interfacing and integrating, through APIs, elemental electronic commerce building blocks and services; these latter are provided by a variety of service providers and application designers. The enabling infrastructure services include those needed to provide requisite transaction integrity, authentication, and privacy. The enabling services and APIs must be at a low enough level of detail (granularity) to provide open, seamless links to the key electronic commerce building blocks and services; they also must be simple and flexible enough to permit user customization and continuous improvement and evolution over time. To maximize flexibility and modularity while admitting alternative competing implementations, an object model is preferred for describing and invoking these building blocks.

The object model has close parallels with the digital object model associated with the Defense Advanced Research Projects Agency (DARPA) Digital Library Program. Both object models share many of the same needs and attributes (i.e., the need for naming, updating, retrieving, routing, pricing, and maintaining a repository of digital objects, and for addressing copyright and usage concerns). Three key questions remain:

1. Should the application objects and services be self-describing?
2. How should they be standardized, named, accessed, and updated (i.e., should we adopt object request broker standards)?
3. How can they best be interfaced and integrated into existing processes, procedures, and legacy systems?



### Electronic Commerce Architecture

We envision a three-layered electronic commerce architecture, in keeping with that described in a previous Cross-Industry Working Team (XIWT) paper, with an architecture consisting of the following:

- The physical communications and computing infrastructure;
- An enabling infrastructure services layer; and
- An applications layer composed of an API layer and an application object layer.

This architecture includes the infrastructure services needed to support the major electronic commerce activities discussed here and the key electronic commerce services, objects, and APIs discussed below.

#### Infrastructure Services

Several generic infrastructure services are critical to a successful electronic commerce framework:

- Reliable communications services;
- Common security services;
- Access control services;
- Translator services;
- A software agent management and communications infrastructure; and
- Distributed information resource discovery, retrieval, and synchronization and replication services (e.g., search engines, browsing, and publishing tools).

All of these services will probably be needed for most other applications domains as well and have been or will be discussed in other XIWT papers in this regard. In this section, we discuss these elements with particular reference to their application in an electronic commerce setting. We also discuss specific services unique to electronic commerce (e.g., paying and accounting).

#### Communications Services

For electronic commerce, existing communications mechanisms (e.g., virtual circuits, routing and addressing, datagrams, e-mail, file transfer protocol [FTP], HTTP, with image and other multimedia extensions) must be extended to incorporate the following features:

- Reliable, unalterable message delivery not subject to repudiation;
- Acknowledgment and proof of delivery when required;
- Negotiated pricing by usage and/or quality of service; and
- Directory services that can be rapidly updated and that support quick retrieval.

These extensions are either generally available or under development. However, to support electronic commerce, they must work across a variety of information and communications devices (including telephones, personal computers and workstations, set-top boxes, and personal information managers and communicators); human-machine interfaces (ranging from character text to virtual reality, and from keyboard and electronic pen to speech recognition and gestures); communications media (including satellites, cable, twisted wire pair, fiber optics, and wireless, which includes constraints on available communications bandwidth and reliability); and nomadicity (which includes supporting location independence, and remote personal file storage with privacy encryption).

## Common Security Mechanisms

Security is a critical component of any electronic commerce application and must be addressed in designing any electronic commerce service infrastructure. Electronic commerce system security should provide the following types of guarantees to the user:

- *Availability.* The system should prevent denial of service to authorized users—for example, if a third party ties up the network either inadvertently or intentionally.
- *Utility.* The system should ensure that the user obtains and retains value of information. Information can lose its value for the user if it is revealed to unintended third parties.
- *Integrity.* The system should ensure that information is delivered whole, complete, and in good order, and that, where applicable, the information is the same as agreed upon by all parties. Date- and time-stamping along with digital signatures is one mechanism for ensuring the latter.
- *Authenticity.* The system should ensure that the parties, objects, and information are real and not fraudulent or forged. To be sure that users are negotiating and exchanging proper objects with proper parties, the transacting parties, devices, and controlling and exchanged objects all need to be authenticated (i.e., verified that they are who or what they claim to be and that none of the information or objects have been illegally tampered with or modified). This requires mechanisms such as digital signatures, passwords and biometrics, and certification hierarchies.
- *Confidentiality.* The system should ensure that information communicated and stored is kept private and can be revealed only to persons on an approved access list.
- *Usability.* The system should ensure that only the rightful owner or user maintains possession of his or her information. Even if others cannot decode and read the stolen information, if they can take possession of the information, they can deny the rightful owner the use to and access to it.

## Access Control Services

Once authenticated, users need to be authorized for requested services and information. User authorizations can be provided as a blanket binary approval or granted only under or for specified conditions, time intervals, and/or prices. Authorizations can be provided to designated individuals or to designated organizational representatives. It is therefore often desirable to authorize a user in terms of his or her location and organizational function/role, as well as on the basis of individual identity.

## Translator Services

Translators can transform and interpret information from one system into formats more suitable to other interacting objects and systems. Translator services should be able to adapt and evolve automatically. For example, a translator that can interpret a small subset of electronic forms that have been linked to SQL relations and data dictionaries should be able to prompt the user for any needed additional information and update itself accordingly. The translator could then be incrementally expanded by further manual linking of data relations to electronic forms, by direct user query, and by learning from example. This capability for selective incremental expansion would enable a user to customize translators to meet unique needs and to expand the translator easily so as to handle larger vocabularies and collections of electronic forms/documents as needed, as well as incorporate new EDI standards as they evolve and become defined. Finally, such a capability would help simplify and speed up the EDI standards process.

### **Software Agent Management and Communications Infrastructure**

Software agents are intelligent programs that simplify the processing, monitoring, and control of electronic transactions by automating many of the more routine user activities. Software agents may be local programs running on the user's machine, or they may actually transport themselves over the network infrastructure and execute on the service provider's machine.

Agents are a relatively new development. Currently, they can do such things as filter incoming mail, coordinate calendars, and find desired information for presentation to the user. Over the longer term, agents are likely to take over more complex tasks such as negotiating, translating, and overseeing and auditing electronic transactions. Some additional future uses for software agents include personalization and customization of applications, and personalized searching, filtering, and indexing.

Eventually, we may have many different agents working for us, coordinating and communicating among themselves. When this comes to pass, we will need standards and infrastructure to support the necessary management, negotiation, and coordination not only between users and agents but also among agents, and to maintain agent repositories where agents can be stored, retrieved, purchased, and leased.

In an electronic commerce setting, software agents should be able to do the following:

- Control the workflow governing a set of electronic commerce transactions;
- Operate a set of conditional rules specified and agreed to by the involved parties;
- Monitor and enforce the terms and conditions of electronic contracts;
- Provide an intelligent interface or a facilitator to existing proprietary VANs and legacy systems, performing the necessary translations and protocols;
- Help the user find desired products and services, and navigate the NII on the user's behalf;
- Purchase and negotiate on behalf of the user; and
- Operate across a wide diversity of vendor hardware and software.

### **Distributed Information Resource Discovery and Retrieval Services**

**Generic Services.** Distributed information resource discovery and retrieval services help service providers list and publish their services, and help users find services and information of interest. These information services cover the ability to maintain, update, and access distributed directory services. They also cover more advanced navigation services such as maintaining hyperlinks, advanced keyword and context search engines, and software agents, such as Web crawlers, that can explore and index information sources and services of interest. These services should be easy and efficient to update as well as to access and use. They should also be capable of being implemented, maintained, and accessed over a number of locations distributed across the NII.

**Unique Services.** In addition to the generic services just discussed, there are a number of desirable infrastructural services that are unique to electronic commerce:

- Accessing currency exchange services;
- Accessing cash management services;
- Accessing bonding services;
- Accessing escrow services;
- Accessing credit services;
- Accessing various investment services;
- Accessing various insurance services;
- Accessing costing services;
- Accessing financial information and reporting services;
- Accessing notarization services;
- Posting and accessing regulatory notices;

- Performing unique EDI and electronic commerce-specific document translations;
- Linking to existing billing and payment systems; and
- Linking to banks' accounts receivable/accounts payable services.

### **Electronic Commerce Building Blocks: Objects and Object Classes**

In addition to the services cited above, the activities and functions of electronic commerce require certain basic building blocks:

- Unstructured information (reports and freeform text, voice, and video);
- Structured information (EDI messages, electronic forms, contracts and contract rules, design specifications);
- Accounts, account databases, and accounting rules;
- Transactions;
- Records;
- Agents and brokers (information filters, translators, trusted third parties);
- Objects for sale (movies/videos, software objects, contracts, information, documents); and
- Decision support models and simulations.

Over time, these items will probably become increasingly comprehensive and refined.

These building blocks can be best described as classes of digital objects. A digital object is an ordered sequence of bits associated with a handle, or unique identification, that can represent a collection of operations (behaviors) and information structures (attributes); and where an object class represents a collection of objects that share a common set of attributes and behaviors. A digital object can be composed of one or more of these classes; for example, an e-mail object has both structured and unstructured information. Digital objects are particularly useful because they are associated with real objects (e.g., a contract making them easy to understand) and because they can be specified and accessed in an application-independent manner, making them easy to create, reuse, enhance, modify, and replace, and to interface with existing objects, with minimal side effects.

Electronic commerce activities can be specified in terms of the interactions between real objects (e.g., transacting parties) and digital objects (e.g., electronic documents, software agents). An electronic commerce architecture can be defined in terms of how these object classes are defined (e.g., their attributes and behaviors) and how the objects interact with one another. An electronic commerce transaction can also be implemented as an interacting network of these objects, where each object can be dynamically selected as a function of the specific situation.

Electronic commerce digital objects have several important properties that are discussed below.

### **General Properties of Digital Objects**

Several operations and controls can be associated with any electronic commerce digital object.

- *Exchange operations.* Examples of permissible exchange operations include being bought, sold, and transferred (in whole or part). Exchange operations encompass a variety of transport mechanisms, including both existing mechanisms such as e-mail attachments and FTPs, and new and evolving mechanisms such as encasing the object inside a digital envelope (thus making the object opaque to everyone except the intended recipient when he or she opens the envelope).
- *Security operations.* Examples of security operations include making the digital object secure and confidential (e.g., encrypted), annotating and signing the object (e.g., with digital signatures), and making it tamper-proof and/or tamper-evident.

- *Ownership and authentication controls.* Examples of ownership and authentication controls include ensuring the digital object's integrity (that is, showing it to be whole, complete, and in good order); date/time-stamping it; ascertaining copyright status; and linking to owners and collaborators, including evidence of the object's source or proof of origin.
- *Usage tools and controls.* Examples of usage control include allowing the digital object to be created; published; displayed or read; written to and updated; and reproduced and copied (note, however, that for some object classes, it may be desirable to inhibit copying), subject to various restrictions and charges. These controls can restrict use to particular authorized users and with selective access criteria specifying type of use (e.g., read only). Usage controls also include such operations as enforcing intellectual property usage rights and charges, version-control and backup, change control, and group sharing (e.g., collaborative authoring). Objects should be able to be compressed, decompressed, and manipulated in ways appropriate to their format (e.g., images can be rotated, enhanced, have features extracted and/or matched, enlarged and reduced in size; or video and sound can be skimmed and/or played in fast time or slow motion).

### Some Important Electronic Commerce Digital Objects

Several key digital objects for electronic commerce are listed below.

#### Contracts

Examples of contracts include the following:

- Orders,
- Checks,
- Bills,
- Loan agreements,
- Treasury bills and bonds,
- Letters of credit,
- Account agreements,
- Receipts, and
- Electronic money/electronic checks/electronic tokens.

Contracts can include instructions regarding the handling, routing, storing, scheduling, and workflow of the contract itself and of other objects contained in or referenced by the contract. These instructions can address liabilities; acceptable forms of payment (cash, credit card, debit, or check); terms of payment (usage charges, periodic and one-time charges); billing and payment instructions (credit to merchant, automatic debit card deductions, billing and payment addresses, due date); delivery instructions (where and how to deliver); return policies; methods of error and dispute resolution; and conditions of good delivery. Contracts can be negotiated, including prices, terms of payment, penalties, necessary documentation, credit checks or required insurance, and collateral or margin. They can be written, signed, read, and amended. In many instances, contracts can also be bought, sold, and exchanged. In many cases (for example, in the case of electronic cash), contracts should not be able to be altered, reproduced, or copied.

#### Information Documents

Examples of information documents include the following:

- Balance sheets;
- Income statements;

- Official statements;
- Monthly billing statements;
- Stock offerings;
- Credit history;
- Reports;
- Discharge summary;
- Physician orders;
- Process notes;
- Electronic books;
- Movies/videos; and
- Video games.

Information documents can be unstructured, partially structured, or completely structured. They can be browsed or searched, and bought, sold, exchanged, and copied, under contractual constraints. They also can be created, updated, signed, copyrighted, and read, then synchronized, morphed, compressed, and decompressed.

### Accounts

Accounts include the following information:

- User (name, identification, authorizations, preferences, and other profile information);
- Address;
- User profile (e.g., likes, dislikes, personal secrets/information);
- Outstandings;
- Credits and debits;
- Balances;
- Tax computations;
- Receivables and credits;
- Payables and debits; and
- Limits (e.g., credit limits).

Accounts can be opened, closed, linked, updated, blocked, stopped, or attached. They can receive deposits, debit withdrawals, and accept transfers. Also, accounts and account information such as account balances can be verified. Since linked transactions (for example, billing, paying, receipt, and delivery transactions) are not generally simultaneous or one to one, it is often necessary to reconcile account information. The ability to link and associate account and remittance objects to payment transactions helps simplify account reconciliation. Account operations are accomplished through transactions, which are discussed later in this section. It is generally necessary to establish audit trails, so that the consequences of multiple transactions on an account can be tracked.

### Software Agents

Software agents include the following:

- Facilitators, who provide key infrastructure services such as translation and communication/network security management and control;
- Information filters and interpreters;
- Translators;
- Trusted third parties and witnesses;

- Escrow agents;
- Brokers;
- Workflow managers;
- Fiduciaries;
- Expert advisors; and
- Negotiators.

An agent should be able to serve in more than one of these roles.

### **Objects for Sale or Exchange**

Both physical objects (e.g., cars and clothing) and digital objects (e.g., program logic, digital movies, data, electronic design specifications, electronic contracts) can be sold and exchanged.

### **Transactions**

Transactions are generally governed by contracts and update accounts. They can operate on all the other digital objects and generally involve the transmission and exchange of two or more digital objects (e.g., a movie for money, medical services for money, exchange of two currencies, etc.). They can also include the exchange of bills and invoices and of information and services.

Transactions can be designed to be anonymous and untraceable or traceable and auditable. If they are designed to be untraceable, they lose many of their information features. A more satisfying compromise is to execute a transaction that can only be traced with the consent, approval, and active cooperation of the user.

Transactions can, but do not necessarily always, include the following information:

- Who is involved in the transaction;
- What is being transacted;
- The purpose of the transaction;
- The destination for payment and delivery;
- The transaction time frame;
- Permissible operations;
- Date/time stamps; and
- Special information, such as unique identification, transaction path, sequence information, receipts and acknowledgments, links to other transactions, identification of money transferred outside national boundaries, certificates of authenticity, and the like.

Transactions can be reversed, repaired, disputed, monitored, logged/recorded, audited, and/or reconciled and linked (e.g., matched and associated) with other transactions. If the transacting parties want to make the transaction anonymous or untraceable, then the users will forego many of the above features. A compromise is a transaction that can only be traced with the consent, approval, and active cooperation of the user or designated escrow agents.

### **APIs: Infrastructure Standards**

APIs and information exchange protocols are needed for digital object operations and infrastructure services. Many APIs—for example, FTP, HTTP, simple mail transport protocol (SMTP), and multimedia information exchange (MIME)—already exist and could be considered as a starting point for the NII. Additional APIs specifically needed as interfaces between electronic commerce objects and infrastructure services include the following:

- APIs that enable two-way exchange of data between electronic forms and database records into databases, including calling and using translator programs (these APIs would allow the automatic fill-in of electronic forms from databases, and the update of database records from electronic forms);
- APIs, where possible, complying to a plug-in/plug-out model that enables embedding and transmission of electronic forms and documents into e-mail messages, automatic conversion of these e-mail messages into their original format at the receiver site, and subsequent processing of the forms/documents;
- APIs that allow data from an electronic form or document to be transmitted from sender to receiver as a database record update or file transfer via FTP;
- APIs between translators that can translate electronic forms into database commands and/or electronic commerce remote procedure calls, and vice versa;
- APIs that define operations such as writing, reading, certifying, authenticating, and transporting of bill and payment objects; and
- APIs that link electronic orders and electronic form messages with electronic payment messages and exchanges.



## 60

# Prospects and Prerequisites for Local Telecommunications Competition: Public Policy Issues for the NII

*Gail Garfield Schwartz and Paul E. Cain*  
*Teleport Communications Group*

The national information infrastructure (NII) vision embraces three communications components (exclusive of information libraries and other content repositories): internodal networks, access nodes, and end-user access to the nodes. Internodal networks and access nodes are currently subject to competitive supply. End-user access is the only NII component that remains monopolized virtually everywhere in the United States. Such access is costly to provide. Costs of new or improved end-user access facilities must be justified by usage in a two-way mode, in an environment that will probably evidence substantial price elasticity. Nevertheless, many risk-taking access providers are preparing the technologies with which to offer economic, reliable end-user access, confident that a strong market will develop. Still, the in-place, monopoly-provided, end-user access seems to be useful for a great many consumers. Increased consumer demand for the greater capacity and speed of digital technologies within the NII will be conditioned largely by the behavior of public policy makers vis-à-vis the entrenched monopolies. In this paper we discuss the opportunities for and obstacles to replacement of the monopoly for end-user access.

### COMPETITION IN THE LAST SEGMENT OF THE TELECOMMUNICATIONS INDUSTRY

It has long been recognized by economists that a monopoly has less incentive to innovate than does a firm in a competitive market. Firms in a competitive market will seek to gain a competitive advantage over their rivals through innovation and differentiation, thus enhancing consumer surplus as well as producer surplus. Events in the long-distance and equipment segments of the telecommunications industry seem to have confirmed the result predicted by economic theory. In the inter-LATA long distance market, AT&T's market share has dropped from over 80 percent to less than 60 percent since 1984, and the number of carriers has risen to over 400<sup>1</sup>. The facilities-based long distance carriers increased their deployment of fiber-optic transmission facilities fivefold between 1985 and 1993<sup>2</sup>. Long-distance rates have fallen more than 60 percent<sup>3</sup>. Similar gains have been realized in the demonopolized market for telecommunications equipment. Now handsets can be purchased for less than the price of a pair of movie tickets, and businesses and consumers are linking ever more sophisticated equipment (computer modems, PBXs, cellular phones, and handsets that have computer memory) to the public switched-telephone network.

Theoretically, the same benefits would accrue to consumers from competition in the remaining monopolized local exchange market as have been realized in the equipment and long-distance market segments (efficiency, diversity, innovation, and price reductions). But the grip of government-sanctioned monopoly remains powerful. Only slowly has it been understood that local exchange monopolies may temporarily drive down costs but never as far as would competition: They may innovate, but not as swiftly as competitors would; they may improve service quality, but not as readily or effectively as they would under competition. The full benefits of the NII require efficient exploitation of all telecommunications technologies, including photonics, coaxial cable, fiber optics, wireless, and even the existing twisted pair technologies, linking subscribers to one another via a "network

of networks." A competitive market for local exchange services is essential to such efficient exploitation of all technologies.

Moving from theoretical concepts to the practical reality of creating local exchange competition requires certainty and flexibility: certainty to inspire investor confidence, and flexibility to respond to constantly changing—and largely unknown—market realities. Achieving balance between certainty and flexibility means opening up existing monopolized market segments as much as possible without threatening the long-term stability of competition itself. Both industry and public policymakers must take this challenge seriously if the NII promise is to be realized by 2000—or ever.

### **The Current Status of Competition in Local Telecommunications Services**

Revenue for the local telecommunications market in 1993 exceeded \$88 billion<sup>4</sup>. Of that total, 99 percent was captured by mature, traditional local telecommunications carriers. The 1 percent of the market liberated by new entrants was composed almost entirely of private line and special access services; they have not yet begun to reach a mass market<sup>5</sup>.

Still, competitive access providers are optimistic about the future of competition, building or operating networks in some 50 or 60 metropolitan statistical areas. A recent report by the Federal Communications Commission indicates that over 10,070,308 miles of fiber has been deployed in the United States by local, long-distance, and urban telecommunications providers, and shows that the rate of deployment by competitive access providers far exceeds that of the incumbent local exchange carriers<sup>6</sup>. The largest competitive access provider, Teleport Communications Group (TCG), has installed SONET-based, self-healing, two-way ring networks capable of transmitting information at a rate of 2.4 gigabits per second. ISDN is provided over these networks, and TCG offers frame relay at up to 1.5 megabits per second and ATM (switched data) service at up to 155 megabits per second. With more than 167,314 miles of fiber throughout 22 networks and a strong switch deployment program, TCG is technically positioned to serve larger markets<sup>7</sup>.

Cable television operators, whose networks now pass 97 percent of the nation's households and provide television service to more than 65 percent, are the obvious "other" end-user access provider<sup>8</sup>. Cable companies are upgrading their distribution plant and must provide for switching to offer local exchange services. Experiments, such as the Motorola-TCI-TCG trial of residential service using radio frequencies over fiber-optic-coaxial cable, will identify technological requirements. The promise of wireless subscriber loops has also drawn considerable interest from a wide range of industry participants. Bidders in the FCC's recently concluded Personal Communications Service (PCS) spectrum auctions committed more than \$7 billion for licenses to build and operate local telecommunications networks<sup>9</sup>. A recent Merrill Lynch report predicts that one of the successful PCS bidders would have a 5 to 8 percent penetration of the local exchange market by 2004<sup>10</sup>. There is no shortage of potential entrants seeking to provide local exchange services. Meeting consumer demand will mean the employment of a variety of distribution systems, both broadband and narrowband, wireline and wireless.

However, it is not yet certain what customers want and when they want it. Though studies such as a recent Delphi study of industry executives and academics have projected 2010 as the outside year in which a mass consumer market will exist for network-based interactive multimedia products and services provided over switched broadband, market research is notably thin on the subject of what people will pay for "infotainment" or household management services<sup>11</sup>. As with the case of previous telephone, television, computer, and audio products, much investment rides on the premise that supply will create demand.

Despite the limited deployment of broadband networks, consumers have not been especially hindered in their attempts to establish themselves as providers, as well as users, of information. Internet platforms and a plethora of electronic bulletin boards allow consumers to "publish" information available on demand by other consumers. Aside from providing the conduit over which the information travels, the network operator has no role in the content of the traffic speeding over its lines. The development of local network competition will only hasten the development of more information services and more gateways through which consumers can share information.

The prospects for local competition, therefore, depend in part on the perfection of new technologies, but more importantly on reduction of their cost. However, factors other than technology itself pose greater obstacles to deployment of competitive choices and must be dealt with expeditiously if investment is to accelerate. Before the various broadband and other end-user access technologies can be widely deployed, legal, technical, and economic hurdles must be overcome in every state. Furthermore, even though technical interconnection and interoperability of networks have been shown to be feasible without serious difficulties, many aspects of the seamless interoperability of competitive networks remain to be resolved. Among them are a number of portability and central office interconnection arrangements. Finally, economic interconnection of competing local networks has to be achieved so that new entrants can develop their own services and prices that maximize the revenues from their own networks.

### **Legality of Competition**

Regulatory barriers remain the threshold barrier to entry into local telecommunications services. Only seven states have authorized or permit competitive local exchange carriers to provide service: New York, Illinois, Maryland, Massachusetts, Washington, Michigan, and Connecticut. Making competition legal means removing a series of entry barriers imposed by different government agencies whose interests are not necessarily congruent. The primary issues that must be resolved include exclusive franchises, qualifications for entry, franchise fees, and access to rights-of-way and to buildings. Market entry that is conditioned upon demonstrating that the incumbent carrier is providing inadequate service is an unreasonable burden, as are franchise fees or other payments to local government as a condition of operating, if the incumbent does not pay equivalent fees. But the most difficult ongoing issues will be the access issues, because that is the area where the incumbent monopoly has the ability and the incentive to encourage unequal access. Incumbent LECs enjoy generally unlimited access to public right-of-way and often control the rights-of-way needed by entrants. They also have established access to building entrance facilities, at no cost. New entrants must have access to those rights-of-way at the same rates and on the same terms and conditions as the incumbent.

In 1995, 13 states enacted legislation removing barriers to entry and endorsing local exchange competition. Federal legislation preempting state entry barriers and setting guidelines for local competition is again under consideration. The technical and economic aspects of network interoperability are rising to the forefront of public policy issues that will condition NII development.

### **Technical Feasibility**

At the end of the nineteenth century and continuing into the early years of this century, local exchange "competition" did thrive. But rather than a "network of networks," consumers faced a tangle of discrete networks, all of them talking at once but none of them talking to each other. To be certain that they could reach everyone with a phone, customers had to subscribe to the local exchange service of every provider.

As that experience—which gave way to government-sanctioned monopoly—demonstrated, the key to an efficient and effective "network of networks" is interconnection. Adjacent (i.e., "noncompeting") carriers interconnect seamlessly with each other today and have done so for more than 80 years. Now competing local exchange networks must be able to interconnect with the incumbent local exchange network on the same terms and conditions as the incumbent interconnects with adjacent carriers and with itself. At a minimum, the technical interconnection arrangements should include the following.

### **Central Office Interconnection Arrangements**

Central office (CO) interconnection arrangements are the physical facilities that connect a competitor's network to the local exchange carriers' (LEC) network. An efficient cost-based CO interconnection arrangement consists of three elements:

- *Interconnection electronics*. Optical and electronic equipment physically located in the LEC CO, but remotely operated and controlled by the local competitor, used to multiplex, demultiplex, monitor, test, and control transmission facilities;
- *Interconnection cable*. Fiber optic cables, dedicated to the local competitor's use, interconnecting the local competitor's location and the interconnection electronics located in the LEC CO(s); and
- *Interconnected services*. LEC-provided services and facilities interconnected to the interconnection cable and electronics at the CO at cost-based rates.

### **Connections to Unbundled Network Elements**

Although the long-term NII vision presumes that existing copper pair will be replaced (or upgraded) by broadband or at least enhanced narrowband access, a transition period during which local exchange competitors are maximizing their return on other network investment (in internodal transport and access nodes) will occur. These competitors will be able to offer service to a mass market during the interim period only by reselling existing subscriber loops under monopoly control. Currently, whether providing single-line basic local exchange service or multiline local services such as Centrex, LECs usually bundle together the "links" and the "ports" components of the loop. LECs must unbundle the local loop elements into links and ports, individually pricing each element at cost-based rates.

### **Seamless Integration into LEC Interoffice Networks and Seamless Integration into LEC Signaling Networks**

A LEC's tandem switching offices and end offices are interconnected via the LEC's interoffice transmission network. Adjacent LEC switching offices are also interconnected via the same network for the distribution of intra-LATA traffic. Routing through the interoffice network is governed by a local exchange routing guide (LERG) unique to each LATA. The LERG prescribes routing of all traffic between and among all LEC and interexchange carrier (IXC) switching offices. Cost-based interconnection arrangements at tandem offices as well as end offices are necessary, and the Class 4 and 5 switches of local competitors must be incorporated into the LERG on the same basis as (and with the same status as) the LEC's end and tandem offices. This will enable a competing local exchange carrier to efficiently and effectively route calls originated by customers on its network to the LEC for final termination at the premises of a LEC customer, and vice versa. This includes access to Signaling System 7 (SS7) with a line information database (LIDB). In an SS7/LIDB environment, routing, translation, service, and account information is stored in centralized databases that LEC and IXC switches can access through service control points (SCPs). Local competitors must be able to interconnect and query the LEC SS7 databases at the SCPs in a nondiscriminatory manner and under equitable terms and conditions.

### **Equal Status in and Control over Network Databases**

Competitive local service providers must be allowed to have their customers' telephone numbers included in telephone directories, directory assistance, LIDB, advanced intelligent network (AIN), 800-number, and other databases. Their access to such resources must be equal in price, functionality, and quality to those of incumbent local telephone providers.

### **Equal Rights to and Control over Number Resources**

Numbering policy must be broadly developed and administered in a competitively neutral manner. The administration of area codes provides an excellent example of the problems that can arise under LEC-administered code assignments. In recent years, as competition has developed, the incumbent LECs have proposed a deviation from their historic practice of assigning customers to a new area code according to geography when the existing area code numbers were approaching depletion. That is, customers in one part of the existing area code territory would retain their area code, while customers in an adjacent area would be assigned a new area code. The new approach favored by some local exchange carriers is to assign a different area code to customers of the new providers (wireless and competitive access providers), allowing only the LEC's customers to remain undisturbed by the new area code assignments. The anticompetitive effects of such a plan are not hard to imagine: customers of the new carriers are difficult to locate, and only customers of the new entrants must incur the expense of changing their numbers (e.g., letterhead, business cards, advertising). Such expenses could be a significant deterrent to a customer who might have found it otherwise economical to switch to a new entrant provider.

### **Local Telephone Number Portability**

The ability of customers to change service providers and to retain the same local telephone number at the same location (service provider number portability), without having to dial extra digits and without being burdened by "special" actions, is a critical component of the economic viability of local exchange competition. Interim number portability mechanisms (such as remote call forwarding) are an inferior form of number portability that impairs a new market entrant's service, and such impairment should be reflected in interconnection charges.

### **Cooperative Practices and Procedures**

Local exchange telephone companies maintain a detailed set of administrative, engineering, and operational practices and procedures. These coordination practices must extend to new local competitors if competition is to emerge. In addition, the traditional local telephone companies maintain a detailed set of practices governing revenue pooling, intercompany settlements, and other financial interactions between and among established local carriers. These also must be extended to local competitors. Finally, any other intercompany notification procedures, standards-setting procedures, and the like, through which LECs may now or in the future interact with one another, must be extended to include local competitors. These requirements may appear reasonable, but the task of integrating such behaviors into huge corporations with many employees and a monopoly mentality will be daunting.

### **Economic Viability**

The ultimate determinants of the amount and extent of local exchange competition will be neither legal nor technical, but economic. The ability of competitors to meet consumer demand depends ultimately on economic interconnection arrangements for competing local networks. The new entrant must install capacity to accommodate peak-hour traffic, even though initially it has little inbound traffic. Arrangements for reciprocal compensation for the mutual exchange of local traffic should allow both carriers to recover the incremental cost of capacity. This cost is low: about 0.2 cents per minute, on an average basis<sup>12</sup>. Given the trivial cost of supplying incremental capacity, it makes practical sense to implement a "sender-keep-all" arrangement, like that now used by Internet providers, rather than to impose explicit charges on terminating carriers. Sender-keep-all is also administratively simple.

Such arrangements should minimally be in place until data-based number portability is established. The lack of number portability and the substitution for it of measures such as remote call forwarding impair the new entrant's ability to use its network to its full capabilities and to sign up customers. Remote call forwarding and other so-called interim number portability measures entrench the incumbent's position and give it an economic advantage because all calls are routed through the existing LEC's network. Under current regulation, this means that the LEC as intermediary could collect terminating access charges from interexchange carriers for calls destined to a customer on a competing local exchange carrier's network. Although arrangements to overcome this economic inequity might be made, they would be complex and costly. It is far more efficient to plan for and expeditiously implement a long-term solution to the number portability problem. This must be a data-based solution that does not put any telecommunications service provider at a disadvantage and allows all providers to maximize the utility of their networks.

### Universal Service

The long-held policy objective of universal service must be addressed before competitors can compete effectively throughout mass markets for local telecommunications services. This will make it possible for efficient competition to eliminate uneconomic or noncompetitive subsidies embedded in telecommunications pricing structures over a reasonable transition period. This is an expected result of competition. Incumbent LECs argue that retail rates (especially for residential consumers) may be priced substantially below costs and will be subject to upward price pressure. But again, as distribution networks achieve economies of scale, incremental costs are expected to fall, mitigating such price pressures.

Nevertheless, social policy pressures to maintain average prices for similar offerings across the country and between urban and rural subscribers will constrain providers' pricing practices for a transition period of at least 10 years—until competitive facilities-based distribution networks reach a substantial portion (say 30 to 40 percent) of rural subscribers. During this period (and for certain communities, perhaps permanently) universal service must be assured. The ability to obtain access to the public switched network voice-grade service with tone dialing on a single-party line, at affordable rates, together with operator assistance, directory listings, and emergency service, is the minimal level of residential service and may require subsidies for some customers. If public policymakers seek to mandate a higher level of service, such as broadband access, the costs will soar; currently there seems to be a political consensus that market forces should define what "basic" service requires subsidy.

Subsidies to preserve universal service must be explicitly identified and administered through a provider-neutral fund so as to minimize their cost and maximize their efficiency. All telecommunications service providers should contribute to the subsidy fund based on their share of the market. Subsidies should be portable among carriers; that is, all local exchange carriers must have the opportunity to receive a subsidy when selected as a service provider by a subsidized customer.

### PROSPECTS FOR LOCAL TELECOMMUNICATIONS COMPETITION

The prospects for local competition are improving. In 1995, 13 states enacted legislation making local exchange competition possible, and most of these laws directed state regulators to provide for the requisite technical and financial interoperability and equal access. In nine other states, the critical issue of reciprocal compensation for the mutual exchange of local traffic is being addressed and interim agreements have been concluded between new entrants and LECs. Up to 15 additional states will soon develop rules for local exchange competition. Federal legislation setting the standard for local exchange competition and giving all major industry groups something of what they want in a competitive environment has passed the Senate; many observers give it at least a 50 percent chance of becoming law in 1995. Three trials designed to assess the technical requirements for telecommunications number portability are under way. And activity in the courts has sharply accelerated as parties seek to end existing legislative or judicial curbs on their ability to address each other's markets.

Among the restraints under attack are the line-of-business restrictions on the Bell operating companies (BOCs) entering the interexchange market or manufacturing, included in the consent decree that divested the BOCs from AT&T. These restrictions were intended to endure so long as the divested regional BOCs (RBOCs) retain control of the local exchange bottleneck, which gives them the ability and incentive to impede competition in adjacent markets. The "VIIIc" test applied by the court overseeing the decree requires the U.S. Department of Justice to report that there is "no substantial possibility" that a BOC seeking relief from the decree would be able to impede competition in the adjacent, restricted market. Some members of Congress seek through legislation to abrogate the decree entirely, eliminating the role of the DOJ, without evidence that the local exchange bottleneck has actually been broken. Others would substitute a more liberal test—the "public interest" test—for VIIIc and make the FCC rather than the DOJ and the court responsible for its application. Still others would have the DOJ involved in an advisory capacity. And some, like the Clinton Administration, prefer to retain the Justice Department's role <sup>13</sup>.

Several BOCs are seeking relief from the decree directly from the court. On April 3, 1995, the DOJ filed a motion to permit a trial supervised by the DOJ and the court in which Ameritech could provide interexchange service after the development of actual competition, including facilities-based competition, and substantial opportunities for additional competition in local exchange service <sup>14</sup>.

The questions of how much and what type of competition warrants relief will be the key to whether or not durable and robust local exchange competition is possible. Even if all the "paper" requirements for local exchange competition are met, the incumbents can and do manipulate each and every aspect of access to their essential facilities. As BOCs have demonstrated frequently, they can frustrate and thwart competitors in numerous ways: by imposing different and unnecessary physical interconnection requirements for competitors; by engaging in pricing discrimination, tying arrangements, and selective cost de-averaging; by imposing excessive liabilities on customers who wish to change to a competitive carrier; by engaging in sales agency agreements that effectively foreclose markets to new entrants; and more. As a result, a public policy that seeks the full benefits of competition for the greatest number of telecommunications customers must include safeguards that constrain anticompetitive behavior and sanction the parties having market power that engage in it. A rigorous effective competition test prior to BOC relief is such a safeguard.

The other required safeguard is continued regulatory oversight of the firms having market power, with lessening intervention as that market power diminishes. During the development of interexchange market competition, AT&T, the dominant interexchange carrier, endured economic regulation while its competitors were free to price their services to the market. A similar approach is necessary in the local exchange market. Incumbents may be granted pricing flexibility, so long as they do not reduce prices of competitive services below their incremental cost; but cross subsidies from captive customers must be prevented. New entrants must face minimal regulatory burdens because, lacking market power, they cannot harm the public interest. However, new entrants must contribute to the maintenance of universal service.

## CONCLUSION

By any measure, local competition does not exist today, but a more competitive environment is being created. Full-scale competition will require huge investments, which in turn calls for relative certainty as to the pace and outcomes of regulatory and legislative actions. If new entrants have technological innovations to exploit, their ability to complete cost-effective deployment depends on their relative financial strength vis-à-vis the incumbent monopolies. The latter often can manipulate financial faucets by means of their negotiated agreements with state regulators, whereas competitors depend entirely on risk markets for their capital.

Competitors are engaged in a historic process on the legislative, regulatory, and judicial fronts. There is no doubt that the barriers to entry will fall. The safeguards to control residual market power as local telecommunications competition emerges will be lodged in different government agencies, as they currently are, but the guidelines will become clearer as the benefits of competition become more obvious and more widespread.

The final determinant of the pace at which these policy adjustments occur is consumer demand. Demand will surge when "infotainment" is varied and priced to mass market pocketbooks.

## NOTES

1. Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission. 1995. "Long Distance Market Shares: Fourth Quarter 1994," Federal Communications Commission, April 7.
2. Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission. 1994. "Fiber Deployment Update: End of Year 1993," Federal Communications Commission, May.
3. "FCC Monitoring Report, Table 5.5." CC Docket No. 87-339, June 1995.
4. Federal Communications Commission. 1994. *Statistics of Communications Common Carriers*.
5. Connecticut Research. 1993. "1993 Local Telecommunications Competition ... the 'ALT Report.'" Connecticut Research, Glastonbury, Conn.
6. Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission. 1994. "Fiber Deployment Update: End of Year 1993," Federal Communications Commission, May.
7. It has been announced that TCG will be contributing to a new venture of Sprint, Tele-Communications, Inc., Comcast Corporation, and Cox Cable that will package local telephone, long-distance, and personal communications with cable services into a single offering for residential and business consumers.
8. National Cable Television Association. 1993. "Building a Competitive Local Telecommunications Marketplace," National Cable Television Association Position Paper, October.
9. Federal Communications Commission Bulletin Board, March 13, 1995. Wireless Co., a consortium formed by cable television operators TCI, Cox, and Comcast, and long distance carrier Sprint, successfully bid over \$2 billion for licenses in 29 markets.
10. Merrill Lynch & Co. 1995. "The Economics of the Sprint/Cable Alliance," Merrill Lynch & Co., February 10.
11. Kraemer, Joseph S., and Dwight Allen. n.d. "Perspectives on the Convergence of Communications, Information, Retailing, and Entertainment: Speeding Toward the Interactive Multimedia Age," Deloitte Touche Tohmatsu International, p. 13.
12. Brock, Gerald. 1995. "Incremental Cost of Local Usage," prepared for Cox Enterprises, March, p. 2.
13. See S. 652, "Telecommunications Competition Act of 1995"; H.R. 1555 Discussion Draft, May 2, 1995, "Communications Act of 1995"; and H.R. 1528, "Antitrust Consent Decree Reform Act of 1995."
14. *United States of America v. Western Electric Company, Inc. et al.*, and American Telephone & Telegraph Company, Civil Action No. 82-0192.



## 61

# The Awakening 3.0: PCs, TSBS, or DTMF-TV—Which Telecomputer Architecture is Right for the Next Generation's Public Network?

*John W. Thompson, Jr.*  
*GNOSTECH Incorporated*

### ABSTRACT

Universal, equitable, and affordable access to the national information infrastructure (NII) is achievable over the next 5 to 7 years. However, our switched and broadcast telecommunications industries have not been given the national goal and task of pursuing the network architecture and technologies that can provide such access to interactive media public utility services.

At present, these communications industries are pursuing variations of the personal computer and TV set-top box network architecture models. These public network architectures are inappropriate and have economic handicaps that will cause them to fail in the provision of universal access and service for the American public. However, there is at least one network architecture model—DTMF-TV—that does appear capable of delivering on the NII promise. This "architecture," based on the existing and universally available telephone and NTSC video networks, can provide universal and affordable access to the Internet, the Library of Congress, and any other source of entertainment and knowledge<sup>1</sup>. However, the NII needs an encouraging, supporting, and regulating public policy. This public policy should provide for the nationwide common carriage of real-time addressable video to the home, equal access to cable TV head ends for interactive multimedia providers, and the appropriate video dialtone rules.

### STATEMENT OF THE PROBLEM

The national information infrastructure (NII) is totally dependent on the proper convergence of our public switched telecommunications network (PSTN) and mass media public broadcast network (PBN) into a nationwide telecomputer system. Two predominant network architecture models and one emerging model are competing to be "the" information superhighway. The interactive media network architecture deployed by our nation's cable and telephone companies will have a significant impact on whether the general public will begin to have equal, affordable, and universal access to the NII by 2000. Over the next 5 to 7 years we can expect the interactive multimedia industry to continue to improve PC technology and program content. However, the model that posits a PC in every home, connected to an online network of "PC" telecomputer systems, will fail to provide universal access to the NII because of the economic burden of PC hardware and software purchase and maintenance, and their constant obsolescence. Similarly, today's cable and telco interactive TV trials, using the TV set-top box telecomputer system model, will also fail to provide universal access because of the same economic burdens of their underlying telecomputer network architecture. Unless the PSTN and PBN industries change their current architectural focus, we will continue down paths that will lead to greater division between our nation's information haves and have-nots.

To best serve the public with interactive media services will require a fully integrated system of switched and broadcasted telecommunications common carriers. Considering the asymmetrical nature of interactive multimedia networks, the switching technology and broadband distribution media are already in place to integrate these information superhighways in a way that is economical for universal access<sup>2</sup>. Unfortunately, these key NII elements are owned primarily by two competing corporate entities within any given geographic locality. Even

the aborted Bell Atlantic-TIC megamerger reveals the economic shortcomings of integrating the broadband loops of cable TV with the switches and back-office operations of telcos when they are not located within the same geographic locality. As a result, these telco and cable companies are struggling to add to their networks what the other already has in operation. For two capital-intensive industries, this economically taxing competition begs the public policy question: If building the information superhighway is so expensive, why, as a nation, are we trying to build two of them? The problem is that public policy has not charged our PSTN and PBN industries with the goal and task of creating a new and shared "public utility" system nationwide, a *public telecomputer network* (PTN) providing universal and equal access to interactive media on a common carrier basis. This is the strange attractor toward which our communications and computing industries have been converging since the breakup of the Bell system. However, the current chaos will continue and public network engineers will pursue inappropriate network architectures until our country proclaims such a goal with the proper laws to encourage, support, and regulate the enterprise.

### BACKGROUND

Before 1993, the public policy issues surrounding the evolving NII were mostly limited to industrial infighting. It was the telcos versus newspapers over interpretation of the MFJ, broadcasters versus studios over Fin/Syn rules, and everyone taking opposing positions on the FCC's proposed video dialtone rules. These public policy debates were, for the most part, kept within the communications industry establishment and the hallways and offices of Congress and the FCC. Then someone said "500 TV channels," and visions of the entertainment possibilities and public service opportunities moved from the industry trade rags to the front pages and covers of our national consumer publications. This media attention captured the public's imagination as the now infamous information superhighway.

Since then, most of our leading telecommunications and media executives have declared themselves "infobahn" road warriors. They rushed off to announce their respective multimedia trials and megamergers to deliver tele-this, cyber-that, and your own personal virtual-reality Xanadu. Within the next 5 to 7 years, the public will expect the NII, as the next generation's "public network" of switched and broadcasted communications, to deliver on these entertainment, education, and public utility promises. This is a major undertaking for private industry and public policymakers. To understand the network and regulatory engineering paradigm shift that must take place, one needs to comprehend the existing and evolving public network infrastructure within a common context. That context is "video dialtone."

The current state of play within the industry involves two predominant, and one emerging, telecomputer system models. A telecomputer system, in the fuzzy macro context of the NII, is the mass media system that our nation's cable and telephone companies are striving to create to deliver interactive digital everything to the consuming public. It is the underlying hardware infrastructure that will integrate telecommunications, television, computing, and publishing into a seamless national multimedia network.

The older and more familiar of the predominant telecomputer models is that of the PC in every home connected to the Internet and other packet-switched networks of computers. This is the "PC" model. Although this model was a dismal failure as a mass medium during the brief videotext era, it has had a recent resurgence encouraged by flashy multimedia PCs, GUIs, CD-ROMs, and the explosive worldwide growth of the Internet. The champions of this model tend to be the manufacturers and sophisticated users of advanced PCs, workstations, and high-speed data networking gear. The essential NII elements that this model brings to a telecomputer architecture are those that offer the most artistic, creative, and communication freedoms to users, programmers, and publishers.

The other predominant model, getting off to a dubious start, is that of the "smart" TV set-top box (TSB) in every home interfacing with a video server. This is the "TSB" telecomputing model. This model is the result of recent advances in microprocessor, video compression, and network transmission technologies. The champions of this model tend to be the manufacturers of cable converters, microprocessors, and midrange computers in partnership with cable and television companies. In apparent conflict with the PC model, the essential NII elements of the TSB network architecture are those necessary for responsible mass media broadcasting and

network operator control. More important for the NII, the TSB telecomputer model recognizes the public utility potential of an unlimited number of addressable video channels to the consumer.

These two models expose most of the essential NII elements perceived by the converging industries as necessary for the successful introduction of a new mass medium. Between the two, it is possible to extract the inherent objective of a telecomputer system. The objective is to offer all consumers the potential and opportunity for interactive access to all multimedia publications over a public network. The objective is also to provide this access in a responsible and socially acceptable manner. Although each of the incumbent models has technical and philosophical advantages over the other, neither will pass the test of being economically feasible as a mass medium.

Taking lessons from early industry trial experiences and failures, just now emerging is a third telecomputer model. This telecomputer model envisions the public using ordinary telephones and Touch-Tone (DTMF) signaling over PSTN networks, and using only the buttons of an ordinary telephone as an interactive TV (ITV) "remote unit" to access and interact with centralized computers broadcasting user-personalized output over ordinary NTSC video channels. This telecomputer network architecture combines the best of the other two models in a way that can offer universal access to interactive multimedia services. The DTMF-TV model can deliver on the promise of common carrier interactive TV and programmer competition. Whether or not this public utility service will be made available to the public over the next 5 to 7 years will depend on the creation of a new video dialtone policy, a policy that will lead to fully integrated switched and broadcasted services on an equal access and common carrier basis. Such an NII policy should influence the choice of an appropriate telecomputer network architecture by our nation's cable and telephone engineers.

## ANALYSIS AND FORECAST

### Video Dialtone(s)

To use video dialtone as a common context for NII issues, we need to define it. There are three distinct types of video dialtone networks and regulatory models. The first compares to traditional PSTN services because that is all it is. This video dialtone is now finding its way into the marketplace as compressed video teleconferencing and transfers of digital data to office desktops via multimedia workstations and data networking services. As the economics for these switched services improve, this form of video dialtone will likely find its way into the homes of telecommuters. Over the next 5 to 7 years, videophones and PC videoconferencing windows will penetrate the home in the same way that facsimile machines and other home office productivity tools do. This form of video dialtone, VD-1, is only a common carrier's analog or digital switched service offered on demand. Switched point-to-point and bridged two-way point-to-multipoint communications, video or not, are covered by generally accepted PSTN regulations and tariffs.

The second form of video dialtone originates from satellite and local TV transmitters over federally regulated and licensed public spectrum. It also comes from cable TV head ends transmitting over locally franchised rights-of-way. This form of "passive" video dialtone is one means of access to the consuming audiences of the PBN. The public's TVs are now limited to receiving their principal choices for necessary, convenient, and entertaining or interesting passive video program transmissions in this form. This nonswitched one-way, point-to-multipoint video delivery is the most efficient and economical method to distribute high-quality video programming to the viewing public on a "scheduled" basis. Advances in digital video compression and fiber optic transmission technologies have led to the potential for a quantum leap in the number of broadcast video channels that can be delivered to the public. These developments led to the so-called 500 channels and video-on-demand bandwagons. However, this form of video dialtone, VD-2, does not yet have a recognized and accepted common carrier regulatory model. When approached from the common carrier perspective there are some natural, yet severe, technical and network architecture limitations. These limitations relate to channel and time slot availability and social responsibility concerns. If the NII is to include a common carrier infrastructure that would permit any programmer or content creator equal access to America's TV audiences, the evolving public network will require a virtually unlimited number of addressed video channels.

The third form of video dialtone is an integrated media mix of the PSTN and PBN, a "mix" permitting services with favorable economics for a national interactive multimedia medium. It is this evolving new form of public network dialtone (VD-1 + VD-2) that is of particular interest to the NII. This form of video dialtone is VD-3. It will be those "interactive and on-demand multimedia communications" services available from nationwide PTN networks on a common carrier basis. This form of video dialtone will satisfy the promises hyped to the public, thus far, as the information superhighway<sup>3</sup>.

### **Telco (VD-1) and Cable (VD-2) Subscribers: One and the Same (VD-3)**

A major factor affecting the evolution of the NII and the technology to be deployed for it is the need for a clearer understanding of just who are the "public" in the evolving public network. As the now separated wireline network industries converge on the fully integrated network services "attractor," the characteristics that once distinguished video broadcast (VD-2) from switched telephony (VD-1) subscribers are rapidly blurring. This phenomenon will dramatically influence the VD-3 or "interactive video dialtone" regulatory model. Do cable subscribers want to subsidize cable's entry into telephony any more than telco subscribers want to subsidize telco's entry into cable? From an NII public policy standpoint, these subscribers are one and the same. They should not be burdened with paying for a redundant infrastructure in the name of full-service networks' competition in a "natural monopoly" environment. To further compound this issue, as the wireless industries of over-the-air, satellite, and microwave broadcasters also converge on the same "fully integrated network services" attractor over the next 5 to 7 years, the issue of "who are the public" as public utility subscribers accessing this evolving public network will become even more blurred.

Over the next 3 to 5 years we can expect that the quest for spectrum efficiency on the cabled wireline side, through digital compression and multiplexing, will apply equally to the wireless over-the-air side of VD-2 video broadcasting. As the broadcasters of subscription channels (e.g., HBO, Showtime, Cinemax) continue to explore multicasting opportunities (e.g., HBO-1/2/3, ESPN and ESPN2), one can expect the more traditional VHF and UHF networks of the PBN (i.e., CBS, ABC, NBC, Fox) to want access to the same commercial opportunities. This trend, however, will require the NII to set VD-2 standards for compressed audio and video broadcasting in order to encourage a market for the associated consumer electronics (i.e., wireline and wireless broadcast receivers) and other addressable customer premises equipment (CPE). These standards may be based on a version of the Motion Pictures Encoding Group (MPEG) standard, the Asymmetrical Digital Subscriber Line (ADSL) standard, or some future NTSC/FCC or industry digital standard. This trend of subscriber evolution is even more apparent when one considers the eventual societal ramifications of video on demand (VOD), near-VOD (NVOD), and other on-demand "audience targeting" (e.g., billing, screening, selectivity, transactional) functionality. The fundamental NII element at issue here is broadband household "addressability" in the mass media environment of broadcast video networks (VD-2). Beginning with the addition of the ever increasing numbers of pay-per-view (PPV) channels that will eventually constitute NVOD, over the next 3 to 5 years cabled systems will continue to expand the commercial opportunities associated with VD-2 addressability. It is this addressability element and the efficiency of electronic distribution that will eventually attract the direct mail and advertising industry into the interactive multimedia (analog and digital) convergence. Also attracted will be the catalog shopping industry as a natural evolution of the switched telephone (i.e., 800-number service, VD-1) and broadcasted NTSC video (i.e., VHF, UHF, CATV, VD-2) TV home shopping industry. As the electronic publishing industry (i.e., audio, video, and multimedia programmers) converges on a single and fully integrated VD-3 or (VD-1 + VD-2) communications network, the concept of "the subscriber" will evolve from one of being either a VD-1 (i.e., telco) or VD-2 (i.e., CATV) subscriber to that of being a VD-3 subscriber more closely resembling an a la carte magazine subscriber in a "common carrier" postal distribution system.

### **I Have an Address ... Therefore I Am**

The bottom line is that the regulatory model for the NII must acknowledge the emergence of a new public utility network system that will eventually, among other things, enhance if not replace our national postal system and local systems of public and private libraries, bookstores, movie theaters, and video rental outlets. The NII is a public network that carries broadcasted (VD-2) and switched (VD-1) information and entertainment (i.e., electronic communications and publishing) to the "addressable households" of VD-3 (i.e., NII) subscribers.

### **Utility Potential Comes from Controlled Flow**

Assuming a proper interactive video dialtone (VD-3) regulatory model for the encouragement of an NII with public utility potential, the services for these subscribers will have to come from someone. That someone will be those corporate and commercial entities investing in the necessary hardware, software, and "wetware" (i.e., R&D and other creative talents) for some given and recognized, regulated and unregulated, return on investment(s). This brings us back to PTN architecture and the technology deployment issues. The PTN hardware/software, regardless of architecture model (i.e., PC, TSB, DTMF-TV), for delivering and controlling the public utility potential of an information "flow" consists of the following elements:

- "Pipes" (e.g., DS-0, ISDN, SONET, NTSC, and other transmission standards);
- "Valves" (e.g., DXC, ATM, USPS—zip code, Bellcore administered—NPA—addressing plans; and other household routing and switching standards); and
- "Reservoirs" of information (e.g., databases, film libraries, the Internet, the Library of Congress).

One must understand that the metaphor of an information superhighway (VD-1) is only half of the bigger public works picture. The NII (i.e., VD-3) is also an information superpower utility project (i.e., water, gas, electric light, sewage, VD-2).

Although their roles are still to be determined, these corporate entities (i.e., private sector investors) for the next decade or two will nevertheless consist of variations and combinations of LECs, IXCs, MSOs, broadcasters, publishers, and other agents. Their high-level PTN architecture will consist of the following elements:

- "Dams" (i.e., technical, economic, regulatory, legislative, and licensing parameters) to channel the flow;
- "Floodgates" (i.e., LEC COs, MSO head ends, and IXC POPs) to regulate, meter, and bill for the flow; and
- "Generators" (i.e., networked computers) to serve the public with information and entertainment services.

The bottom line is that the dams and floodgates are PTN architectural "barriers and bottlenecks" that will control the flow of information and entertainment to the public. Consequently, the NII will need to provide for dams with equal-access floodgates and pipes.

### **LATAS, MSAS, RSAS, MTAS, AND BTAS: THE ISSUE OF TERRITORY**

Attracting the investment necessary to build this new public network, as in any public utility infrastructure project, will likely require grants of clearly defined market territories to licensed operators. The more exclusive the legal right to exclude nonlicensed competition in a given VD-3 subscriber serving area (SSA), the more attractive these pioneering PTN operators will become to long-term private sector investors. Hence, exclusivity will drive the private sector's funding of the NII.

## RECOMMENDATIONS

### Set a National Goal

Our former Bell system had a simple but broad goal: a phone in every home that wanted one. Since the Communications Act of 1934, this natural monopoly public network was tasked with providing the public utility infrastructure necessary to permit any person in the United States to communicate with any other person in an affordable way. That goal, for its time, was enormously ambitious; but everyone, from planner to engineer, lineman to operator, clerk to executive, public consumer to regulator, intuitively understood it. It took half a century of dedicated work by telephone pioneers to achieve that goal of universal service.

The public sector should set a similar goal that will challenge the private sector to create a ubiquitous interactive multimedia common carrier system to serve the next two to four generations of Americans. That goal should include a public telecomputer network system harnessing the diversified and collective intelligence of our entire country for the public's convenience, interest, and necessity.

### Develop a National Channel Numbering Plan

In the same way that the public and private sector cooperated to develop a channel numbering plan for VHF (i.e., channels 2 through 13) and UHF VD-2 channels, nationwide VD-3 network operators will require a consistent and national CATV (i.e., community antenna TV) channel numbering plan. A video program on a broadband TV channel in one state should be on the same NII channel in any other state. This NII numbering plan should accommodate both the existing wireless and wireline broadcasters in a common carrier environment, provide for equal access by all information and entertainment generators, and reserve channel space for education and government. It should also make allowances for significant channel numbers (e.g., 411, 555, 800, 900, 911) that will assist the public in navigating a VD-3 world. Such a plan is needed to offer the public network channel surfers a "logical interface" to the 500+ channel systems of the relatively near future (see Reference 3 for plan proposal).

### Define the Classes of Service

NII policy should also define the classes of (nondiscriminatory) VD-3 service. These classes in a common carrier environment of addressable VD-3 or (VD-1 + VD-2) households will consist of the classes permitted by combining the PSTN's nationwide automatic number identification (ANI) capability with a numbering plan for the growing number of addressable broadband receivers in the PBN. With the coming evolution of video "mail," the classes of service can, if not should, be modeled after those used in our postal system (see Reference 3 for plan proposal).

### Develop a PTN Test and Demonstration

The PTN will consist of VD-3 network operators providing a public network of common carrier services to the next generation. The public and private sectors should jointly develop the criteria that will define a PTN public utility system. The criteria should set parameters that include demonstrating the inherent capability of the PTN architecture (i.e., PC, TSB, DTMF-TV) both to provide universal and affordable interactive multimedia access and to serve all of America's communities in a nondiscriminatory fashion.

### **Offer the Prize**

There is at least one enabling public utility technology (see Reference 4 for an example) that can deliver on the VD-3 universal access promise. The public sector may need to acknowledge that this form of public utility service, like all others before it, requires a "federal franchise" to operate as a regulated monopoly. Such a franchise will be required in order to attract the necessary capital to build an information superpower system. This approach worked before as the Bell system. With a pioneering spirit, it can work again for a new generation of Americans.

### **REFERENCES**

- [1] Thompson, Jack. 1995. "The DTMF-TV Telecomputer Model Offers the Most Economical Approach to Interactive TV," GNOSTECH Inc., Annandale, Va.
- [2] Thompson, "The DTMF-TV Telecomputer Model," 1995.
- [3] "The Awakening 2.0," the comments of GNOSTECH Incorporated to the FCC's Proposed Rulemaking on Video Dialtone (Common Carrier Docket No. 87-266), 1991.
- [4] United States Patent No. 5,236,199, "Interactive Media System and Telecomputing Method Using Telephone Keypad Signaling."

## 62

# Effective Information Transfer for Health Care: Quality versus Quantity

*Gio Wiederhold*  
*Stanford University*

### STATEMENT OF THE PROBLEM

This paper addresses two related problems created by the rapid growth of information technology: the loss of productivity due to overload on health care providers, and the loss of privacy. Both occur when excessive amounts of data are transmitted and made available. These issues are jointly related to a trade-off in quality versus quantity of medical information. If quality is lacking, then the introduction of modern communication technology will increase health care costs rather than constrain them.

### BACKGROUND AND PROBLEM DESCRIPTION

The online medical record is rapidly becoming a reality. The technology is available and social barriers to acceptance are disappearing. Access to online patient data during a treatment episode will become routinely accepted and expected by the patient as well as by the provider. The image of an expert, in the popular view, is now associated with a computer screen in the foreground, and medical experts are increasingly being included in that view. Eventually, online validation of health care information may become mandatory. A recent court case involving a physician who had failed to use available information technology to gather candidate diagnoses was decided in favor of the plaintiff<sup>1</sup>, presaging new criteria for commonly accepted standards of care.

The rapid growth of the Internet, the improved accessibility of online libraries, and online medical records all provide huge increases in potential information for health care providers and medical researchers. However, most beneficial information is hidden in a huge volume of data, from which it is not easily extracted. Although the medical literature, largely through the efforts of the National Library of Medicine, is better indexed than literature in any other scientific field<sup>2</sup>, the volume of publications, the difficulty of assessing the significance of reports, inconsistencies in terminology, and measures to protect the privacy of patients all place new barriers on effective use and result in what is sometimes called information overload. This overload means that diligent research for any case can require an open-ended effort, likely consuming many hours. We consider that the current and imminent presentation of information is of inadequate quality to serve the practice of health care.

Unfortunately, the pace of development of software to provide services that deal effectively with excessive, convoluted, heterogeneous, and complex data is slow. Since the problem in the preceding years has always been access, there is a lack of paradigms to deal with the issues that arise now. Before, when voluminous medical data had to be processed, intermediate staff was employed, so that the health care provider was protected both in terms of load and responsibility. Staff at research sites filtered and digested experimental results. Staff at pharmaceutical companies filtered for benefits and effectiveness. Government agencies monitored lengthy trials. Publishers required refereeing and editing. Students and interns discussed new reports in journal clubs. Collegial interactions provided hints and validations. But our networks encourage disengagement of intermediaries, and while most of the intermediate filtering tasks are aided by computer-based tools, there is no common paradigm that ensures the quality of information products.



## ANALYSIS

Assessing the demands placed on the national information infrastructure by health care services requires considering the needs of the health care providers and their intermediaries. This analysis is thus based on customer pull rather than on technology push. This approach is likely to lead to lower estimates than would a model focusing on technological capabilities. However, we will assume a progressive environment, where much paper has been displaced by the technologies that are on our horizon.

In our model, information requests are initially generated for the delivery of health care by the providers and their intermediaries. Pharmacies and laboratories are important nodes in the health care delivery system. Education for providers and patients is crucial as well and will be affected by the new technologies. Managers of health care facilities have their needs as well, paralleled at a broader level by the needs of public health agencies. Functions such as the publication of medical literature and the production of therapeutics are not covered here, since we expect that topics such as digital libraries and manufacturing in this report will do justice to those areas.

### Services for the Health Care Provider

The initial point in our model is the interaction of the provider with the patient. Such an interaction may be the initial encounter, where tradition demands a thorough workup and recording of physical findings; it may be a visit motivated by a problem, where diagnostic expertise is at a premium; it may be an emergency, perhaps due to trauma, where the problem may be obvious but the treatment less so; or it may be a more routine follow-up visit. In practice, the majority of visits fall into this routine category.

Adequate follow-up is crucial to health care effectiveness and is an area where information technology has much to offer. Having the right data at hand permits the charting of progress, as well as the therapeutic adjustments needed to improve or maintain the patient's health care status. Follow-up care is mainly provided locally. The majority of the consumers of such care are the older, less mobile population. It is this population that has the more complex, longer-term illnesses that require more information.

The needs for information differ for each of the interactions described above. Initial workups mainly produce data. The diagnostic encounter has the greatest access demands. Emergency trauma care may require some crucial information, but it is rarely available, so that reliance is placed on tests and asking the patient or relatives for information. Note that many visits to emergency facilities, especially in urban settings, are made to obtain routine care, because of the absence of accessible clinical services. For our analysis these are recategorized. A goal for health care modernization should be better allocation of resources to points of need, but here we discuss only the information needs. Information for follow-up visits should summarize the patient's history; unexpected findings will trigger a diagnostic routine.

To assess the need for data transmission we need to look at both the distance and the media likely to carry the needed information. Media differ greatly, and all must be supported. Many physical findings can be described compactly with text. Laboratory findings are compactly represented in numeric form. Sensor-based tests, such as EKGs and EEGs, are time series, requiring some, but still modest, data volumes. Sonograms can be voluminous. The results of ultrasound scans are often presented as images. Other diagnostic procedures often produce images directly, such as x-ray or CT and similar scans that are digitally represented. High-quality x-rays require much storage and transmission capacity, whereas most digital images have larger pixels or voxels and require more modest storage volumes. The practitioner typically relies on intermediate specialists to interpret the data obtained from sensors and images, although for validation access to the source material is also wanted.

The distance that this information has to travel depends both on setting and data source. [Table 1](#) indicates estimated sources of patient care information for the types of clinical encounters listed.

TABLE 1 Sources of Information for Patient Care and Estimated Frequency of Use

Encounter Type	Data Type				
	Text	Numbers	Sensor-based	Images	Estimated Local/ remote Ratio; Encounter Frequency
Workup	Local collection aided by staff	Area clinical laboratories	Area diagnostic services	Area, hospital- based services	Very high; low
Diagnostic	Local and remote reports, experts	Area clinical laboratories	Area diagnostic services	Area, hospital- based services	High; modest
Emergency	Local and remote histories	On-site laboratories	On-site devices	On-site services	Modest; low
Follow-up	Extended local histories	Area clinical laboratories	Area diagnostic services	Area, hospital- based services	Very high; high

We conclude that use of local patient care information dominates. The requirement for remote transmission of data for individual patient care is modest. Instances will be important, as when a traumatic accident requires emergency care and consultation with an expert specialist in a remote locale. Here again, quality considerations will be crucial. It will be important to obtain the right data rapidly, rather than getting and searching through tomes of information. At times images may be required as well. Most x-rays will be done locally, although one can construct a scenario in which an archived image is of value. Any actual medical intervention will be controlled by local insight and information.

In addition to requirements for access to and display of individual data, as shown in Table 1, there is a need for online access to the literature and reference material. Here issues of locality are best driven by economic considerations. If the volume and frequency of use are high, then the best site for access will be relatively local; if they are low, the site can be remote as long as access is easy and latency is small. These parameters are under technological control, and no prior assumptions need be made except that reasonable alternatives will survive and unreasonable ones will not.

### Management and Public Health Needs

The requirements of broad health care information for planning and research are significant. Local institutions must improve the use of the data they have in-house already for better planning. New treatments must be monitored to enable rapid detection of unexpected side effects. Public health officials must understand where problems exist, what problems can be addressed within their means, and what recommendations for public investment are sound.

Today effective use of available health care data is difficult. Standards are few and superficial. For instance, the HL-7 standard does not mandate any consistency of content among institutions; only the format of the access is specified. The data collections themselves also are suspect. Private physicians have few reporting requirements, except for some listed infectious diseases. If the disease is embarrassing, then their concern for the patient's privacy is likely to cause underreporting, because they have little reason to trust that privacy will be maintained in the data systems. In a group practice, the medical record will be shared and potentially accessible, but the group's motivations differ little from those of an individual physician. Physicians working in a larger enterprise, such as a health maintenance organization (HMO), will have more requirements placed on them by administrators who are anxious to have adequate records. Still, there is little guarantee today that data are complete and unbiased. The local users are able to deal with the uncertainty of mixed-quality data, since they understand the environment. Remote and integrated analysis is less likely to be able to use local data resources, even when access is granted.

However, clinical data collection and use is an area where change is occurring. The increasing penetration of HMOs, the acceptance of online access, and the entry of local data provide the foundation. When the local information feedback loops are closed and providers see at the next encounter what information they collected, then quality can improve. Sharing one's record with colleagues also provides an inducement to record the patient's state completely and accurately. As the record becomes more complete, questions of rights to access will gain in importance.

Clinical data collection is broad, but rarely sufficiently deep to answer research questions. Where clinicians collect data for their own research the quality of the variables they consider crucial will be high, but the scope of most studies is narrow and not comparable among studies and institutions. Funded, multi-institutional research studies make valiant efforts to maintain consistency but rarely succeed on a broad scale. Although such data will be adequate to answer focused research questions, little management or public health information can be reliably extracted.

Many funded health care and service programs mandate reporting and data collection. But, again, there is likely to be a narrow bias in collection, recording, and quality control, and, except for administrative purposes, the value is minimal. Biases accrue because of the desire to justify the operation of the clinics and services, and, if the data lead to funding, such biases are strengthened. Public health agencies are well aware of these problems and therefore tend to fund new research studies or surveys rather than rely on existing data collections.

Quality again seems to be the main constraining factor. How can quality be improved? Quality will not be improved by mandating increased transfer of data to remote sites. The only option seems to be to share data that are used locally, and to abstract management and public health information from such local data. Feedback at all levels is crucial, not only from encounter to encounter, but also in the comparison of intervals between treatments in a patient's history (especially for aged individuals), among similar patients, and among physicians using different approaches to practice. Again, it is the actual consumers of the information that need to be empowered first.

The desire to have all possible information will be moderated by the effort and time that health care providers must spend to obtain and record it. Eventually, intelligent software will emerge that can help select, extract, summarize, and abstract the relevant and properly authorized information from voluminous medical records and bibliographic resources. Such software will be accepted by the providers to the extent that its results accord with those experienced in their human interactions and aid their productivity.

We see that the demands for national information infrastructure (NII) services are more in making software available and providing interoperation standards than in providing high-performance and remote communication. Today, access is constrained by problems of interoperation, concern for privacy, and the poor quality of many collections. The effort needed to overcome these barriers is major and will take time to resolve.

### Education

The need for continuing education has been more formally recognized in the health care field than in most other areas. Although unmotivated engineers can spend many years doing routine corporate work until they find themselves without marketable skills, the health care professional is faced with medical recertification, hospital admit privileges, and the need to maintain credibility. In urban areas the patient's choices are many and are often based on contacts leading to referrals. All these factors motivate continuing education. The quality of such education is decidedly mixed. Boondoggles are common, and testing for proficiency of what has been learned is minimal or absent. Few standards exist.

In this instance, access to remote instructors and experts can be a boon. For the rural practitioner, who finds it difficult to leave the practice area, such services are especially beneficial. In urban areas, most educational services will be local. The demand on the NII is again difficult to gauge but may again be modest in the aggregate: Fewer than 10 percent of our health care providers practice in remote areas. Spending a few hours a week on remotely accessed educational services seems to be an outer limit. The services should be fast and of high quality. Since time is not of the essence, the service has to compete with printed material, where it will be

difficult, if not impossible, to match image quality. On the other hand, dynamic interaction has a role and can provide excitement.

There are natural limits to the capabilities of humans to take in information. The rate provided on a television screen is one indication of these limits; few people can watch multiple screens beneficially, although the actual information content in a video sequence is not high. The script for an hour's TV episode is perhaps 200 sparsely typewritten pages, but the story is exciting to watch, and that has added value. Only a few paths are reasonable at any point in time, and the choice of paths represents the essential information. Excessive randomness of events, MTV style, is unlikely to be informative. Eventually, the information retained after an hour of educational video is likely to be even less than that of the script provided.

Technology will allow interaction in the educational process, just as now some choices can be made when reading a book or browsing in a library. Again, the number of choices at each point is limited, probably to the magical number  $7 \pm 2$ , the capacity of the interactor's short-term memory<sup>3</sup> Effective educational systems must keep the learner's capabilities in mind. To what extent intermediate representations expand the material and place higher demands on network bandwidth is unclear.

### Decision Support

The essence of providing information is decision support. All tasks, whether for the physician treating a patient, the manager making investment decisions, the public health official recommending strategies, and even the billing clerk collecting an overdue payment, can be carried out effectively only if the choices are clear. The choices will differ depending on the setting. The manager must give more weight to the financial health of the enterprise than does the physician recommending a treatment.

Customer-based capacity limits can be imposed on all service types provided by the information enterprise, just as we sketched in the section on education. Making choices is best supported by systems that provide a limited number of relevant choices. The same magical number ( $7 \pm 2$ ) raises its head again. To reduce the volume of data to such simple presentations means that processing modules, which fulfill the roles of intermediaries in the health care enterprise, must be able to locate likely sources and select the relevant data. Even the relevant data will be excessive. Long patient histories must be summarized<sup>4</sup>. Similar patient courses can be compared, after matching of the courses based on events in patient records, such as critical symptoms, treatments applied, and outcomes obtained. It is rare that local patient populations are sufficient, and so matching has to be performed with information integrated from multiple sites, taking environment into account.

The presentation to the customer must be clear and must allow for explanations and clarifications. Sources of data have to be identifiable, so that their suitability and reliability can be assessed. Once such services are provided, it will be easier to close the feedback loops that in turn encourage quality data. Having quality data enables sharing and effective use of the technological infrastructure being assembled.

Education and entertainment benefit from a mass market, and so the expansion of information into exciting sequences has a payoff in acceptance and markets. That payoff is much smaller for the review of medical records and the analysis of disease and treatment patterns. The volume of health care information transmitted for decision support will be constrained by the filtering imposed by quality control.

### Protection of Privacy

The public is distrustful of the protection of privacy provided for health care data, and rightly so. In many cases, to receive health care services and insurance reimbursement, patients have to sign broad releases. Once they do, their medical information flows, with nary a filter, to the organization's billing clerks, to the insurance companies, and (in case of a conflict) to legal professionals. Although all these people have ethical constraints on the release of information, little formal guidance and even fewer formal restrictions are in place. In the paper world, loss of privacy was mainly an individual concern, as in the case of the potential embarrassment

to a politician when the existence of a psychiatric record was revealed. In the electronic world, the potential for mischief is multiplied, since broad-based searches become feasible.

The insurance companies share medical information through the Medical Information Bureau. This port is assumed to be a major leak of private information. Unless it can be convincingly plugged, it is likely that health care enterprises will have to limit access to their data if they want to (or are forced to) protect the patient's rights to privacy. Many health care institutions, after having appointed a chief information officer (CIO) in the past decade, are now also appointing a security officer. Without guidelines and tools, such an officer will probably further restrict access, perhaps interfering with the legitimate requests of public health officials. It is unclear how such officials will deal with leaks to insurance companies and their own billing staff.

Legitimate concern for the protection of privacy is likely to hinder use of the information infrastructure. We do believe that there are technological tools that can be provided to security officers and CIOs to make their task feasible<sup>5</sup>. To enable the use of information management tools, the information flow within major sectors of the health care enterprise has to be understood. The value and cost of information to the institution, its major components, and its correspondents has to be assessed. Without control of quality the benefits are hard to determine, and it will be difficult to make the proper investments.

Quality and privacy concerns are likely to differ among areas. That means that those areas must be properly defined. Once an area is defined, access rules can be provided to the security officer. A barrier must be placed in the information flow if access is to be restricted. Such a barrier is best implemented as a system module or workstation owned by the security officer. That node (a security mediator consisting of software and its owner) is then the focus of access requests, their legitimacy, and their correct response. The security mediator must be trusted by the health care staff not to release private information and must also be trusted by the customers (be they public health officials, insurance providers, or billing staff) to provide complete information within the bounds of the rules provided.

The volume of data being transmitted out of the health care institution may be less, but the resulting information should be more valuable and trustworthy.

## RECOMMENDATIONS

The sources and uses of health care information are varied. Technological capacities and capabilities are rapidly increasing. The informational needs of the health care enterprise can be defined and categorized. If quality information can be provided, where quality encompasses relevance, completeness, and legitimacy, then the demands in the NII can be estimated, and it appears that the overall capabilities are likely to be adequate. Distribution, such as access in rural areas, is still an open question. I have recommended elsewhere that the Rural Electrification Services Authority (REA) repeat its success of the 1930s by focusing on the provision of information access to the same customers.

The major point to be made is that, in order to provide health care professionals with the best means for decision making, a reasoned balance of software and hardware investments is appropriate. Software provides the means to abstract voluminous information into decision sequences where, at every instant, the customer is not overloaded. There is an optimal trajectory in balancing investments in the systems infrastructure versus software application support, but we have not spent much effort in understanding it. For health care providers the benefits are to be found in the quality of information—it must be good enough, sufficiently complete, and relevant enough to aid decisionmaking. If the quality is absent, then the effort will be poorly rewarded and the risks of failure will be high. The recommendation from this point of view is therefore to move support to the information processing infrastructure, so that relevant applications can be built easily and the customers satisfied. A happy customer will in turn support the goals of the NII.

An area where government support can be crucial is in helping to define and validate standards. Standards setting is best performed by customers and providers, but the validation and dissemination of standards are precompetitive efforts that take much time and have few academic rewards. Academic insights can help ensure coherence and scalability. Tests performed outside vendor locations are more likely to be trusted and are easier to demonstrate. Infrastructure software, once validated, is easy to disseminate but is hard to market until

the application suites that build on the infrastructure are available. The need to support the communications hardware infrastructure has been recognized. The support of software in that role may well be simpler, since its replication is nearly free.

The desired balance for health information infrastructure support can be replicated in all fields of information technology. We expect the parameters to differ for commerce, defense, education, entertainment, and manufacturing. The common principle we advocate is that, as we move from a supply-limited to a demand-constrained information world, our analysis and actual service methods must change.

### REFERENCES

- [1] *Harbeson v. Parke Davis*, 746 F.2d 517 (9th Cir. 1984).
- [2] Wiederhold, Gio. 1995. "Digital Libraries," *Communications of the ACM*, April.
- [3] George Miller.
- [4] Isabelle de Zegher-Geets et al. 1988. "Summarization and Display of On-Line Medical Records," *M.D./Computing* 5(3):38–46.
- [5] Willis Ware, "The New Faces of Privacy," P-7831, RAND Corporation, 1994.

## 63

# Integrating Technology with Practice: A Technology-enhanced, Field-based Teacher Preparation Program

*Ronald D. Zellner, Jon Denton, and Luana Zellner*  
*Texas A&M University*

### STATE INITIATIVE CENTERS FOR PROFESSIONAL DEVELOPMENT AND TECHNOLOGY

This report describes the development and structure of the Texas Education Collaborative (TEC), which was established at Texas A&M University as a means of restructuring the teacher preparation program and improving public education. The TEC was funded as part of a statewide initiative calling for a series of centers to foster collaboration between K-12 and higher education institutions and the incorporation of technology into the educational process.

Texas has recognized a need, as have most states, for the systematic restructuring of its educational system to meet the needs of a changing society and world community. As part of its response to this need, the Texas Education Agency (TEA) funded a program to develop such centers throughout the state. The program for these centers was established by the Texas State Legislature in 1991, enabling the State Board of Education (SBOE) and the Texas Higher Education Coordinating Board to establish competitive procedures for one or more institutions of higher education to establish a center. The initial allocation for the centers was \$10.2 million. To qualify, centers needed to include a university with an approved teacher education program, members from public schools, regional Education Service centers, and other entities or businesses. Eight centers were approved by the SBOE in 1992, six in 1993, and three in 1994. Currently, this collaborative effort includes 31 institutions of higher education, 7 junior colleges, 15 regional service centers, and 193 campus sites.

#### Focus of the State Initiative

The state of Texas, like most states, is currently undertaking a general review of teacher preparation and certification processes. Texas is a culturally diverse state, with the school population spread out over a wide and diverse geographical area. For example, El Paso is closer to Los Angeles, California, than it is to Beaumont, Texas, and closer to two other state capitals (Phoenix and Sante Fe) than it is to the Texas capital, Austin; and Amarillo is closer to four other state capitals (Sante Fe, Oklahoma City, Topeka, and Denver) than it is to Austin. Texas is divided into 254 counties with 1,063 school districts and 6,322 public schools (972 high schools, 955 junior high, intermediate, or middle schools, 3,618 elementary schools, 743 elementary and secondary schools, and 34 schools in correctional institutions). As a means of aiding these districts, the state is divided into 20 geographical regions, each with an Education Service Center (ESC) that has support staff, services, and training for all of the school districts located within its boundaries.

Consideration, partly related to this diversity, is being given to processes and certification requirements, which may vary from one region to another according to the needs of the particular geographical areas, a

consideration that has potential for a variety of outcomes. Other factors influencing the acceptance of multiple certification requirements are related to philosophical, political, and financial concerns.

The centers were established to take a leading role in the revision of the certification process. They are intended to improve teacher preparation through the integration of technology and the development of innovative teaching practices and staff development programs throughout the state's colleges of education. This initiative was based on the premise that the success of schools and students is directly linked to the success of the state's teacher preparation programs. All centers are to focus on five components: collaboration, restructuring educator preparation, staff development, technology, and addressing the needs of a multicultural student population.

Collaboration between the institutions of higher education and K-12 institutions is to be a major component of the centers' activities. Classroom teachers are to serve as mentors and role models for their interns; teachers, principals, and ESC personnel may help develop courses, teach, or co-teach university courses at their home base or at the university. University professors are to be in the schools as resources and consultants, and to learn from the mentor teachers. The governance structure must reflect the cultural diversity of the state, and no category of representation may be larger than the K-12 teacher representation.

The intended restructuring of educator preparation programs, in conjunction with the collaboration emphasis, is focused primarily on establishing programs and instruction that are field based. New programs are also to include regular collaborative decision making by all partners; an emphasis on teachers as lifelong learners; routine use of technologies (multimedia, computer based, long-distance telecommunications technologies); research and development of new technology-based instructional techniques; and innovative teaching practices in multicultural classrooms. These developments are intended to influence several major components in the structure of teacher preparation programs, including the composition, location, and structure of methods courses, the constitution of the teacher preparation faculty, the role of technology, and the cultural diversity represented.

The following presents a comparison of traditional preparation approaches to program components found in the various centers.

To help ensure complete success in restructuring education, these centers are also to include major staff development components that will provide in-service training to practicing teachers. A major emphasis of this staff development effort is on helping teachers to become lifelong learners. Thus, by keeping current throughout their teaching careers, these teachers will be able to teach their students the skills that they will need to be successful in the twenty-first century. In this way, student achievement will be linked to teacher performance, and both will be enhanced through campus improvement plans.

Another major initiative is the incorporation of technology into the centers to expand the delivery of instruction in the K-12 classroom and in teacher preparation classes. Coupled with field-based instruction, this endeavor will help to prepare students majoring in education to teach in the classrooms of tomorrow. The centers are also to provide for the development of new technology-based instructional techniques and innovative teaching practices. Through direct contact with these activities, in-service and pre-service teachers will learn the appropriate skills and will be encouraged to integrate technology into their teaching practices.

Multicultural education provides information about various groups as well as the skills needed to work with them. To provide cultural diversity and adequately prepare in-service and pre-service teachers, the centers are required to implement public school programs and services that reflect diverse cultural, socioeconomic, and grade-level environments. The inclusion of minority teacher candidates in the centers' recruiting and training efforts will also be a high priority, particularly in subject areas where teacher demand exceeds supply.

### **Systemic Development and Evaluation**

Each center is implemented and evaluated in an ongoing developmental program involving K-12, higher education, and state constituents. This collaboration reflects the program's commitment to systemic change and recognizes that all shareholders must be involved if adequate systemic change is to occur. To provide programs and services throughout the state and prepare teachers to meet the full range of learner diversity and needs, the State Board of Education funding requirements for the centers indicate that funding may be made to centers by geographical areas. This is particularly important in a state as culturally and geographically diverse as Texas.



These centers have also been identified as one of the critical communities to be involved in a program of professional development grants under the Texas Statewide Systemic Initiative. The centers will thus provide a structure for subsequent, related initiatives and help ensure that the needs of education will be met throughout the state.

Each center is required to develop an extensive ongoing evaluation design composed of both internal and external evaluations. Progress reports to the Texas Education Agency are required from each center quarterly as well as annually. Visits by statewide evaluators are conducted to review the progress of the centers regarding the five main components—collaboration, restructuring educator preparation, staff development, technology, and multicultural education. Both qualitative and quantitative data are to be considered in the statewide evaluation of the centers' activities and progress.

### **THE TEXAS EDUCATION COLLABORATIVE**

The Texas Education Collaborative (TEC) was established in 1992 as one of the eight original centers funded by TEA. There are currently 17 Centers for Professional Development and Technology throughout the state. In the first year, the TEC partners included two universities with approved teacher education programs, Texas A&M University and Prairie View A&M University, in collaboration with eight schools representing five independent school districts; two state Regional Education Service centers; a community college; and parent and business representatives. The five independent school districts were Bryan I.S.D., College Station I.S.D., Conroe I.S.D., Somerville I.S.D., and Waller I.S.D. Original business partners included Apple Computer, GTE, IBM, 3M, and Digital Equipment Corporation. The goals and activities of the collaborative were established in conjunction with the TEA guidelines and included the following components.

#### **Teacher Education Programs and Environments for Teaching and Learning**

- The general goals in this area were directed at the establishment of a framework for curriculum development and included the following:
  - Coordinated efforts of faculty across all College of Education (CoE) departments for collaboration on research, proposal writing, course development and delivery, and coordination of course contents related to cultural issues;
  - Integration of technology into CoE course instruction, content, and management;
  - More university faculty delivering courses in schools via partnerships with school faculty;
  - Change from course evaluation to content evaluation by peer groups; and
  - Shift from Carnegie course units to block units over 5 years.
- School faculty, administrators, and counselors will become determine instructional methods and materials, identify and develop interdisciplinary curricular themes, schedule classes, and evaluate student performance.
- Schools will be designated as field-based TEC sites for pre-service and in-service professional development and practicum activities for teachers, teaching candidates and administrative interns (thus providing "real-world" experience and the involvement of expert teachers).
- Each TEC site will become a resource institution for restructuring K-12 education and teacher education at all levels (focusing on issues such as in-service training of new techniques, involvement of parents and community, collaboration, strategies for students from minority cultural groups, mentoring, etc.).

#### **State-of-the-art Teaching Practices, Curriculum, and Instructional Knowledge**

- TEC sites will become centers for intellectual inquiry where everyone is both a teacher and a learner. Academic and clinical experiences will include the following:
  - Technology applications and technology-assisted instruction;

- Integration of interdisciplinary content and pedagogical methods;
  - Collaboration and teamwork among school and university faculty and pre-service teachers;
  - Reflective decisionmaking; and
  - Self-evaluation.
- Pre-service and in-service professional development will emphasize cultural diversity, incorporate strategies for dropout prevention, and encourage students from underrepresented groups in all subject areas.
  - Teaching candidates will demonstrate conceptual understanding by applying information to new situations. Assessment will be done by school and university faculty and will include observation of teaching behavior and candidate-produced portfolios of multimedia projects.
  - New technologies will be used to help students solve problems, synthesize ideas, apply information to new situations, and develop oral and written communication skills.

### **GOVERNANCE STRUCTURE OF THE TEC**

The governance structure will include teachers, administrators, education service center staff, and university faculty and administrators. This structure is to reflect the cultural diversity of the state. The coordinating council is composed according to TEA requirements, including the provision that no one group be larger than the K-12 teacher group.

### **BUDGET**

The initial funding allocation was for 1 year, with subsequent funding potential for 4 additional years. The first 2 years of the TEC were funded at \$2 million; \$1.5 million of this was slated for equipment and the remainder for personnel and operating expenses.

The equipment expenditures represent the need to equip faculty and classrooms with computer and related technologies to foster important, needed changes in teaching and operating activities. A number of compressed video sites for two-way video conferencing and instruction were established and utilized for instruction.

### **SUMMARY OF TEC ACTIVITIES**

During the 1992–93 year, the TEC activities focused primarily on the acquisition and setup of technology hardware and software at university and school sites, training in the use of the technology, and development of collaborative structures among school and university faculty. Activities included technology staff development at school and university sites, technology conferences for school and university faculty, frequent sharing meetings, collaboration within and among schools, team-building activities, a retreat for school and university faculty to develop a shared vision for restructuring teacher preparation, and development of proposals for five programs related to various aspects of restructuring.

### **Technology Hardware and Software Acquisitions**

#### **Year One**

The bulk of the first year's acquisitions were computers, productivity software, network hardware, and support multimedia presentation resources. These resources were considered necessary to prepare teachers and university faculty in the use of technologies in order to begin the process of changing teaching activities in classrooms at both levels. If student teachers are eventually to use such technologies in their classroom activities as teachers, they will need to experience such benefits as learners. All teacher preparation university faculty received computers for their offices to encourage the transformation of their daily activities. In addition, fully

equipped laboratories, networks, and multimedia workstations were established for faculty and student use. Classroom presentation stations were also provided to ensure infusion of these technologies into the actual teaching process. Computer teaching laboratories were updated for undergraduate and graduate educational technology classes. Consequently, student teachers would ultimately learn about the use of technology, learn from technology based materials, and learn with technology through their own developmental projects. A similar model was maintained for teachers and students in the schools.

Table 1 gives a breakdown of the number of computers acquired during this period.

TABLE 1 Computers Acquired During Year One

	Universities	Schools	Total
Macintosh	75	68	143
PowerBook	29	14	43
IBM	37		37
IBM Notebook	2		2
Total	143	82	225

### Year Two

The second year's equipment purchases, using the remainder of the initial \$1.5 million equipment budget, included more computers for faculty, teachers, and classrooms. Additional faculty became involved in the TEC activities and were equipped with the necessary hardware and software. The multimedia workstation labs were enhanced with more and more powerful computers and support hardware. Additional software was provided for the development of presentation and instructional materials. Multimedia presentation stations were acquired for classrooms in the various curriculum units to support the use of technologies in methods classes and additional portable presentation units were provided for instructional use throughout the college. E-mail systems were established and incorporated into a major part of the communication process. Again, parallel resources were acquired for the classrooms in the participating schools and were used in a wide variety of activities by teachers and students. Several local computer networks were established in the schools to support instructional and management functions. Software consisted primarily of productivity packages (word processing, spreadsheet, database), presentation packages, hypertext development programs, and graphics production packages.

A major component for the delivery of instruction and faculty training was the acquisition and establishment of a compressed video teleconference network linking schools in the participating districts with the university and one another. At least one school in each of the participating districts was supplied with a complete teleconferencing station and the necessary support components. These sites were established in association with the Trans-Texas Teleconferencing Network (TTVN), which is centered at Texas A&M and has sites located throughout the state. This association greatly enhanced the ability to quickly create a functional network and provided an extensive network for instruction and conferencing for meetings and supervision of students in the field. This provided additional opportunities for collaboration between university faculty and teachers and students in the schools. Each teleconference site provides two-way video and audio communication with the addition of interactive video and computer-based materials that can be viewed and accessed at any of the participating sites in a session. Table 2 presents the timetable for the establishment and initial use of these teleconferencing sites.

### Year Three (First Quarter)

As the TEC began its third year, the focus was shifting from technology and skill acquisition to the development of applications and programs. The majority of the equipment was purchased in the first 2 years of

TABLE 2 Timetable for Establishment and Initial Use of Teleconferencing Sites

Site	Installation date	Number of Site Sessions		Total
		Dec.—Feb.	March—May	
Anson Jones Academy	Dec. 93	12	7	19
Jones Intermediate	Feb. 94	3	15	18
Somerville Jr. High	Dec. 93	22	20	42
Southwood Valley Elem.	Dec. 93	12	7	19
Washington Jr. High	March 94	0	9	9
COE-PVAMU	April 94	1	0	1
COE-TAMU	Feb. 94	14	27	41

the program. Recent acquisitions were directed at improvement of existing facilities through software purchases and upgrades, memory upgrades, video boards, and so on. Two extensive network facilities and file servers were acquired and are currently being established as a resource for communications and resource delivery. The focus in technology resources is now on keeping up with new advances in hardware and software and on meeting maintenance concerns such as monthly charges for telecommunications and video teleconference activities. Five schools were added during the past year as evolving TEC sites and will continue their involvement, and six more schools were added as evolving TEC sites this quarter. These eleven schools received \$13,000 each in start-up funds for equipment and staff development. Table 3 gives a breakdown of these sources in the current TEC inventory.

TABLE 3 Breakdown of Resources in TEC Inventory

Item	Schools	University	Total
Desktop computers	96	109	205
Portable computers	17	29	46
Network resources	1	1	2
Video teleconferencing sites	5	2	7
Projection systems	13	5	18
VCRs	0	5	5
Televisions	3	4	7
Laser disk players	15	15	30
Compact disk players	7	0	7
Printers	34	15	49
Camcorders	0	8	8

### Staff Development: Teachers, Administration, and University Faculty

#### Year One

Each TEC school site council held at least four meetings during 1992–93. Each council included at least site coordinators, two teachers, a principal, and business, parent, and ESC representatives as voting members. Ex officio members are university liaisons and other university and school district representatives. Main topics of discussion during these meetings were the definition of roles, technology (acquisition of hardware and software, networking, access, demonstration), staff development, budget expenditures, and field experiences for teaching candidates. Discussions at TEC staff meetings focused on issues of alternative, authentic, and integrated assessment as well as implications of such assessment techniques on classroom context and curricular reform.

Several individual faculty and administrators from TEC schools and universities attended seminars and courses on professional development schools to increase their understanding of school-university collaboration. Furthermore, the TEC sponsored several events to foster school-university collaboration and to empower teachers as full participants in restructuring teacher preparation programs.

The TEC facilitated collaboration among school and university faculty and administrators through TEC staff meetings, challenge events, a planning retreat, school site council meetings, and curriculum planning groups at schools and universities. Collaborative meetings and events included the following:

- A series of challenge events were held to bring together school and university representatives to work on team building, trust building, and group problem solving, and to begin conversations about restructured teacher preparation. The challenge events helped break down communication barriers among constituency groups, but little was accomplished in actual planning of restructured teacher preparation programs. Teachers reported using similar team building strategies with their students and colleagues. Approximately 75 people participated in at least one challenge event, 35 people in two events, and 18 people in all three events. Nearly half of the participants were classroom teachers.
- Seven TEC staff meetings were held in 1993, providing opportunities for collaborative sharing and planning among site coordinators across TEC school and university sites. Staff included TEC site coordinators, university liaisons, a director, evaluation personnel, and technology personnel. The TAMU College of Education issued requests for proposals from faculty of various departments who work with schools on restructuring teacher education.

Discussion included arrangements for compressed video setups at school sites, a technology sharing time for site coordinators, the need for site coordinators to visit other school sites, expectations for pre-service teachers in the TEC, development of a TEC newsletter, and the development of multimedia interdisciplinary units.

- Five TEC school partners each had at least five site council meetings in 1992–93. Groups also discussed time involved in planning field experiences, how to implement them, what early field experiences might look like, and what the final product might look like. Teachers received training on restructuring and peer coaching.
- The TAMU College of Education held all-day workshops for faculty across departments to explore the organizational structure of teacher education programs. A technology integration task force held several meetings to redesign upper-division teacher education courses for the integration of evolving technologies. A secondary teacher preparation committee composed of university and school faculty met several times in spring 1993 and developed a training proposal.
- All TEC school site coordinators are kept informed regarding exemplary software through staff meetings, conferences, catalogs, and other listings; in addition, a special technology workshop, "TEC Talk," was held on April 16 for TEC site coordinators. Teachers at TEC school sites received technology staff development training approximately once a week, beginning in October 1992. These sessions were conducted by TAMU technology facilitators. Staff development at school sites included word processing, graphics, spreadsheet and database components of ClarisWorks, Aldus Persuasion, and HyperCard, and multimedia applications, as well as individualized training and instructional applications.

Several projects were funded with TEC funds to address the various goals in the state initiative. One project, the Alternative School Leadership Program, was designed to prepare leaders for the different leadership roles required in an emerging full-service professional development school. Project TIME was designed to prepare teachers who would serve as instructional leaders within their schools. It was also designed to assist colleagues in general education, and teachers who can collaborate with community leaders and health and human service agency personnel. A critical element of this program was early identification of at-risk and special needs students, which would result in referral to appropriate support services in the community.

Both the university faculty and the classroom teachers were provided a series of workshops on the operation of the new equipment, the use of the various software packages, and the application of the software to daily maintenance and instructional materials development. These workshops were conducted in both the university and school labs. Educational technology faculty and a technical staff of graduate students were made

available as needed to assist faculty with questions and problems as they developed to ensure smooth integration of these resources into their teaching.

### **Year Two**

During the second year, the TEC was expanded to include faculty from 12 schools, and plans were made to expand to 18 schools. Training workshops continued to be offered in the schools by both university and school personnel. Workshops and help sessions for university faculty were continued. Many training sessions were offered over the videoconferencing networks to help provide training without interfering with the teachers' daily routines. A number of formal workshops were offered at the university for teachers from participating schools and focused on applying software to their areas of teaching. Some teachers took existing university technology classes that were appropriate to their needs. The TEC sponsored attendance at a number of national technology and educational research conferences for teachers and university faculty.

### **Year Three (First Quarter)**

There are two general changes regarding staff development activities. First, expertise and support have evolved at some of the sites to the extent that recently trained site specialists can now deliver staff development support. Second, there has been a move from general orientation and basic skills training to more advanced training and individualized instruction in specific topics and locations. However, there are several new evolving sites with faculty in need of general orientation. A staff of technologists has been available for trouble-shooting problems and training individuals who have questions or problems. During this quarter, there were 201 participants in group training sessions and 46 people scheduled in individual sessions. More individuals were introduced to video teleconferencing operation and capabilities by participating in sessions and through special training materials that were developed. These activities involved both university faculty and teachers. Also, topics were no longer focused only on the hardware and software needs of the participants. For example, a series of workshops initiated in the fall of 1994 covered the following topics:

- The Role of School-University Partnerships in Creating the Professional Development School of the Future;
- Multimedia Use in Collaboratives;
- National Standards;
- Teacher Tactics, Portfolio Preparation, Classroom Management; and
- Future Roles of Field-based Education.

There were several ancillary activities this quarter that highlighted the collaborative component of the TEC through functions extending its resources and influence to other groups and that brought benefits from these groups for TEC participants. A conference on education, inclusion, and diversity was conducted by the College of Education at Prairie View A&M with partial support from the TEC. Between 300 and 350 educators participated in activities related to cultural diversity in education. Another program was cosponsored by the TEC site school system and a city government to highlight the accomplishments of the schools with support from GTE and the TEC. Approximately 400 citizens and educators attended this activity.

A 2-day educational technology conference, cosponsored by the TAMU Educational Technology program, the TEC, and the Region VI Educational Service Center, was attended by more than 400 teachers, administrators, university faculty, and students. Some participating sites used TEC funds to send teachers as part of their staff development efforts. This conference provided an arena for presentation of projects and activities and an opportunity for teachers and university faculty from all curriculum areas to interact on topics relating to instructional technology.

## Development of Teacher Preparation/Pre-service Courses

### Year One

It is intended that all students involved in the TEC will be "prepared" for using technology in their student teaching experience. Since teachers tend to teach as they were taught, one of the first initiatives here was, as described above, aimed at changing the instructional processes and techniques used in methods classes to model the use of technologies. Faculty training and resource acquisition was a necessary prerequisite for this to happen. In addition, courses in educational technology for pre-service teachers were maintained and closely allied to the equipment and applications needs of the field-based activities. Thus, all of the various activities of the TEC are interlinked and dependent on one another. Regarding the actual restructuring of teacher education, program efforts during the first year were focused on establishing technological and social systems infrastructures for the TEC partner schools to facilitate restructuring of teacher education programs in a collaborative manner. Examples of these activities include the following:

- More than 40 site council meetings, involving over 80 teachers, university faculty, and parents, were held to formulate the instructional programs of 169 teachers and some 3,900 students during the past year.
- Work began on a program for secondary school teacher preparation (grades 7 through 12) that includes a holistic approach to educational issues, sensitivity to the needs of teachers, and particular attention to the needs of students with diverse backgrounds and learning styles.
- Community members, Texas A&M faculty, and pre-service teachers spent quality time each week with at-risk learners in order to provide each one with a support network.
- Pre-service teachers worked one-on-one with learners who have emotional and behavioral disorders.
- Teachers and 80 students from four grade levels worked together in a multilevel learning community and conducted research on an interdisciplinary experience-based curriculum.
- Efforts were made to recruit and support cohorts of pre-service teachers from socially, culturally, and linguistically diverse backgrounds, and to involve them in the governance of teacher preparation programs. Direct field experience in diverse community settings and classroom/teaching experience in schools representing culturally diverse communities are intended.

### Year Two

A college-wide review team identified three program orientations that span the range of professional training for pre-service teachers: Young Learners, Middle School, and Secondary School. Because of the number of pre-service teachers (approximately 1,400) being served by current programs, a gradual transition was planned for the implementation of these programs.

These programs will provide distinct phases for the pre-service preparation of students. The first phase is entitled *Children, Families, and Communities* for the Young Learners and Middle School programs and *Self-Directed Experiences* for the Secondary School program. This phase occurs before actual admission to teacher education and includes experiences where potential candidates learn about children, communities, and families and explore education as a profession. This phase involves field work in the community. Future teachers must demonstrate that they can communicate with children and families who speak a language other than English and they must develop a portfolio that includes meaningful interactions and involvement with a child or children from backgrounds different from their own.

The second phase also occurs before admission to teacher education and is entitled *Children, Schools, and Society* for the Young Learners and Middle School programs and *Teaching and Schools in Modern Society* for the Secondary program. Prospective students study the impact of schooling on society and consider what it means to educate children for their roles in democracy. Both phases require potential teacher candidates to collect evidence of experiences that illustrate their awareness of linguistic, socioeconomic, and cultural diversity.

Students in the Secondary program also complete a third phase, *Developing Skills to Meet the Needs of Society and Students*, before they apply for admission.

Candidates in the Secondary program proceed to two semesters of practicums after being admitted to teacher education, while candidates in the Young Learners and Middle School programs are assigned to cohorts that are guided by teams consisting of a university faculty member, a school-based clinical faculty member, and a doctoral student specializing in teacher education. Candidates in the Young Learners and Middle School programs then proceed through the professional development phase, entitled *Children, Teachers, and Classrooms*.

A number of trial and experimental programs have been conducted at various TEC sites. Twelve university faculty and 24 school based-faculty have designed, implemented, and taught university courses to 65 future middle school teachers on one campus. A pilot program, called *Writing Buddies*, was implemented at a middle school site. Pre-service teachers who enrolled in an undergraduate reading methods course participated in this field experience, which provided a tutorial program for students. The program was supervised by practicing professional teachers. A university professor conducted the lecture on campus while selected teachers conducted seminars "on site" with students enrolled in the course. A similar program was conducted in an elementary school in which undergraduates are paired with elementary students to work on writing skills. The future teachers meet with their "writing buddies" at least once a week. Another elementary school served as a site for bilingual programs in the district. Additional goals include the establishment of a block of field-based methods courses taught there. The first block of courses was offered during the spring semester, 1994, and taught on site by a team of university professors. A high school program involved approximately 80 students in the pilot of an interdisciplinary program where the traditional freshman English and U.S. history courses were integrated. Teachers and student teachers participating in 90-minute instructional blocks used instructional strategies on central themes that integrated topics across the traditional subject areas. Through these activities, pre-service students not only gain classroom experience but are also involved with the conceptualization and planning of the activities.

### **Year Three (First Quarter)**

During the fall of 1994, 100 pre-service teachers registered for an undergraduate course that required them to spend 2 hours a week in lecture at the university for instruction in language arts methods. The students were assigned to 1-hour labs on an elementary school campus. Nine teachers served as lab instructors supporting this methods course. Twenty students were selected from the 100 participants in the reading methods course and *Writing Buddies* program to continue for two more semesters. This is one example of putting the goals of collaboration and field-based instruction into action and will serve as a model for other courses.

## **A COLLABORATIVE MODEL FOR TEACHER PREPARATION**

### **Collaboration**

The "need" for establishing centers for professional development and technology was determined by the Texas Education Agency (TEA). Once established, members of the collaborative focused on the educational goals of the state by specifically addressing educational concerns of the individual site members. It was agreed that they would address these concerns together, working on solutions that would benefit site needs and teacher preparation programs. They did this through a series of meetings that included the following:

1. Site council meetings at school site campuses;
2. Specific program focus group meetings at the university level concerned with teacher preparation; and
3. Development council meetings, which included site council representatives as well as university faculty interested in working with collaborative members on specific programmatic change.



Public school problems addressed by the collaborative included the following:

1. Raising TAAS (Test for Academic Assessment of Skills) scores of children at school sites identified as being at risk for school failure;
2. Lowering the dropout rate;
3. Improving attendance and interest in school;
4. Providing staff development opportunities to improve teacher instructional skills in the use of technology in classroom instruction;
5. Developing inclusion models for teaching strategies that help all learners succeed;
6. Providing time for teachers to collectively plan integrated learning experiences;
7. Providing time for teachers to create computer programs that enhance learning; and
8. Improving teacher retention.

Teacher training program problems addressed by the collaborative included the following:

1. Tying real-world experience with methods courses to accommodate 900 pre-service teachers in a training program per year, 18 TEC sites, and 13 actively involved tenure track faculty plus 2 adjunct faculty;
2. Providing more "practice" in application of pedagogy through field experiences coupled with methods courses;
3. Providing instruction for teacher-training faculty in the integration of technology into course content;
4. Ensuring that pre-service teachers see "best practice" in their field experiences. Examples: integrated instruction by teachers, application of technology into classroom instruction by teachers, and collaboration between teachers in planning instruction;
5. Ensuring that technology is used by teacher-training faculty in instruction;
6. Redesigning methods courses so they can be taught in a multitude of settings. Examples: university class setting (traditional approach), school site setting, through telecommunication, broadcasting from site to site or university broadcasting to site(s); and
7. Developing a system of communication between TEC sites, site coordinators, university liaisons, and teacher trainers.

The following are examples of collaborative efforts:

1. Development of a totally integrated, technology-enhanced curriculum for a fourth grade focusing on the general theme of Texas for interdisciplinary investigations in history, cultures, ecology, commerce, and the arts. This collaborative project was designed by teachers and university faculty.
2. Blocked teacher preparation methods courses taught on site by either tenure track faculty or by trained clinical faculty (teachers) in pilot programs at six TEC sites in the collaborative. The courses include field experiences designed by the classroom teachers in collaboration with university professors. Classroom teachers serve as mentors to pre-service teachers with some projects offering seminars conducted by participating classroom teachers.
3. Several conferences on subjects such as distance learning; a language institute for minorities; linking of families, school, and culture; best practices for inclusion of children with disabilities; and aspects of technology.
4. A monthly newsletter documenting activities taking place among the sites, advertising for pen pals between sites, describing university courses taught through the use of telecommunications, and publicizing staff development opportunities at sites.

### Restructuring Teacher Preparation

The mission of the TEC is to develop and implement restructured teacher preparation programs that are technology enhanced and field based. The TEC is actively engaging faculty from schools, universities, and ESCs to work together as agents of change in creating new approaches for preparing teachers and administrators for all levels of public education. Restructuring was defined by Goodlad in 1990 as reexamining the purposes of education, and reworking the structures and practices of schools and university preparation programs in order to improve teaching and learning. As defined by TEA,

Teacher preparation programs must be developed which integrate field experiences in school and community human service. These programs should include advice and counsel from parents, care providers and governmental agencies. Teachers completing these programs shall be lifelong learners who continue to seek new levels of understanding content, pedagogy and culture.

### Three New Teacher Preparation Programs

In addition to elementary and secondary teacher preparation programs, a third program has been added for specific training in middle school methodology. This was a result of efforts to restructure teacher training programs that followed the 1987 Texas standards for teacher training programs. These programs are currently considered pilots for those to be adopted in the fall of 1995. Final approval of the restructured programs will come about in April 1995 when the TEC conducts its program presentation to the Texas Commission on Standards for the Teaching Profession. The newly restructured programs are described as follows.

**Young Learners Program.** This program is designed for candidates in Texas A&M's teacher training program who seek certification in grades K through 4. The Young Learners Program has methods and field experiences designed to address the needs of educating the young child. There is a preprofessional sequence of course work that teacher training candidates must complete before being admitted to the professional sequence, which is designed to cover three semesters of professional development course work, including field experiences that give progressive teaching responsibility to the pre-service teacher.

*Preprofessional Sequence of Courses I—Children, Families, and Communities*. Methods courses focus on language and literacy, educational psychology, and human development. One course serves as a place holder with a field experience requirement. Students are required to work with families, agencies, and services in their field experience.

*Preprofessional Sequence of Courses II—Children, Schools, and Society*. The second suggested sequence of courses before admission to teacher education focuses on the foundations of education in a multicultural society and understanding special populations.

*Professional Semester I—Admission to Teacher Education: Focus on Young Learners, Teachers, and Schools.* Upon admission to teacher education, pre-service teachers sign up for a 12-hour block of courses with a field experience requirement. Different TEC sites offer different beginning field experiences, but all require students to interact with children in a tutorial arrangement. These tutorial mentoring experiences have a variety of titles and themes. All reflect the needs of the individual campuses. Writing Buddies, Reading Buddies, Aggie Friends, Math Magicians, and HOST Volunteers are some of the programs that serve as vehicles for more than 125 pre-service teachers at a time on each campus. It is a win-win situation for the schools and for the teacher training program. Pre-service teachers are strongly encouraged to take a university course in microcomputer awareness during this semester.

*Professional Semesters II and III—Focus on Young Learners, Teachers, and Schools.* Pre-service teachers enroll in a 12-hour block that includes a practicum experience focusing on working with and teaching small groups of children. Students are given increased experience in instruction that prepares them for the student teaching experience in the third semester.

**Middle School Program.** This program is designed for candidates in Texas A&M's teacher training program who seek certification in grades 4 through 8. The Middle School Program has methods and field experiences designed to address the needs of educating the young adolescent. There is a preprofessional sequence of course work that teacher training candidates must complete before being admitted to the professional sequence. The professional sequence is designed to cover three semesters of professional development course work that includes field experiences at the middle school, junior high, and high school levels. Teaching responsibilities are progressive and offer a variety of settings.

*Preprofessional Sequence of Courses I—Children, Families, and Communities.* The 6 hours of suggested courses focus on language and literacy, and adolescent development. One of the two courses serves as a place holder with a field experience requirement. Students are required to work with families, agencies, and services in a middle school setting. They become part of a *School Families* program, shadowing various school personnel during the semester. This could be working with the school nurse, secretary, counselor, truancy officer, principal, or special education teacher.

*Preprofessional Sequence of Courses II—Children, Schools, and Society.* The second suggested sequence of courses before admission to teacher education focuses on the foundations of education in a multicultural society and understanding special populations. A course in microcomputer awareness is a requirement.

*Professional Semester I—Admission to Teacher Education: Focus on Middle School Students, Teachers, and Schools.* Upon admission to teacher education, pre-service teachers enroll in a block of course work that includes a field experience requiring assignments that integrate the curriculum. These courses are taught at a school site.

*Professional Semesters II and III—Focus on Middle School Students, Teachers, and Schools.* Blocked courses are similar to the *Young Learners* program except that specific methods courses are designed to address the needs of the middle school student.

**Secondary Program.** This program is designed for candidates in Texas A&M's teacher training program who seek certification in grades 9 through 12. It consists of four phases and two practicums (the practicums occur in Phase IV of the program). The use of technology is stressed in Phases II through IV. In Phase II the pre-service teacher learns *about* technology and how to use it. In Phase III, pre-service teachers study how students can learn *from* technology, and in Phase IV of the training program, pre-service teachers study how their students can learn *with* technology as an integral part of their instruction.

*Phase I—Self-Directed Experiences with Children.* Students document their experiences in a variety of settings before admission to the teacher education program in Phase III.

*Phase II—Understanding Teaching and Schools in Modern Society.* The goal is for pre-service teachers to develop an understanding of responsibilities of the teaching profession. Pre-service teachers shadow professional teachers working with students in the middle school, junior high, and high school settings.

*Phase III—Developing Skills to Meet the Needs of Society and Students, Admission into Teacher Education.* The goal is to develop teaching skills responsive to the cultural styles, values, and identity of children and adults from all racial, ethnic, linguistic, class, and religious backgrounds regardless of learning and/or behavioral characteristics.

*Phase IV—Developing and Demonstrating Skills to Organize Content and Technology for Use in Classrooms.* The goal in the first practicum experience of Phase IV is for pre-service teachers to develop skills in organizing content and technology for use in classrooms. The goal of the second practicum experience is to demonstrate skills of organizing content and technology. Pre-service teachers demonstrate proficiency in instructional skill during this practicum (similar to the traditional student teaching experience).

### **Additional Staff and Role Changes**

As stated above in the description of collaborative activities, current pilot teacher training programs are addressing Texas Education Agency's expectations in restructuring teacher education. The need for coordination

of collaborative efforts between TEC sites and university teacher training programs has required the addition of personnel to ensure progress toward meeting the goals of the TEC as well as the different expectations of cooperating teachers, university supervisors, and student teachers:

- *Role and Requirements of the University Liaison.* The university liaison serves as a facilitator of information between the collaborative, the assigned site, and the universities involved. The site assesses its needs; the liaison and site coordinators work collaboratively in finding solutions to meet the needs identified.
- *Requirements of Cooperating Teachers Who Have a TEC Student Teacher.* Cooperating teachers agree to promote and learn along with their student teachers the uses of technology and innovative teaching practices in their classrooms. They agree to assist in guiding the student teacher to plan and implement technology and innovative practices in the classroom.
- *Requirements of Student Teachers Assigned to a TEC Site.* Student teacher candidates agree to the following:
  - Demonstrate a strong interest in learning and using technology and innovative teaching practices in the classroom;
  - Design a unit that has a multimedia component and/or interdisciplinary focus; and
  - Incorporate technology in classroom activities such as record keeping, lesson plans, instructional activities, and student projects.

Following are examples of the number of pre-service teachers affected by new field-based programs at TEC sites:

- *South Knoll Elementary School (125 pre-service teachers):* *Writing Buddies* program, a tutorial program.
- *Southwood Valley Elementary School, College Station (20 pre-service teachers):* Pre-service teachers interested in the *Young Learners* program enroll in two blocked courses taught by a team of two faculty members and a clinical faculty member from the elementary campus.
- *Jane Long Middle School, Bryan (40 pre-service teachers):* As part of a methods course requirement, pre-service teachers observe and assist in all aspects of a school. They shadow different personnel, observing a variety of roles. Some examples include the school office staff, the counselor, a physical education and wellness class, in-school suspension, and/or a classroom one day per week. This fits with the theme of Phase I of the Middle School Program, which focuses on children, families, and communities. These same students observe life at an elementary school on a different day each week.
- *A&M Consolidated High & College Station Junior High Schools (40 pre-service teachers):* Students in this methods course, which is designed for pre-service teachers interested in secondary education, work in two school environments. Twenty students are located at the high school and participate in observation and assigned activities that involve all functions of the school. Another 20 students participate in similar activities at the junior high campus. At midterm, the groups switch. The field experience is designed to give students the opportunity to study adolescent behavior, develop an appreciation for human diversity, and model professional behavior.
- *Crockett Elementary, Bryan (27 pre-service teachers):* Students enroll in 15 hours of methods courses. They are assigned to specific classes with instruction conducted by Texas A&M faculty and clinical faculty. They participate in this experience prior to their semester of student teaching. This is just a short overview of some of our site activities.
- *Somerville Middle School Block, Somerville (27 pre-service teachers):* Students are enrolled in five courses taught by clinical faculty who are mentored by a professor from Texas A&M. The pre-service teachers are in the field every afternoon 4 days a week. They participate in classes and follow-up seminars after school with clinical faculty.
- *Somerville High School:* Integrated units, integrated class curriculum, team planning every day, block scheduling.

Other TEC sites are participating in the collaborative in different ways. They either have student teachers and are actively involved in redesigning student teaching experiences, or they participate in compressed video lectures by doing projects with classes at different sites in the collaborative. The last 3 years have brought about many significant changes in the way we teach at the university and in the way students are taught in the schools.

### **Staff Development**

Activities have included in-service training for university faculty, teachers, student teachers, and undergraduates in the use of Macintosh computers, and computer software such as HyperStudio, ClarisWorks, Inspiration, and FileMaker Pro. Staff development is also offered in e-mail, Internet, Telnet, compressed video, development of interdisciplinary instruction, understanding diverse cultures, and mentoring pre-service teachers. At one site, 5th-, 6th-, and 7th-grade students are trained to be a part of the school's "Tech Team." They help teachers and students with technology in their classrooms. They also learn how to use Telnet and how to access information via the Internet.

### **Technology**

#### **Distance Learning**

Compressed video units located at five TEC sites are currently used for instruction and conferences between Texas A&M University and the sites. They are also used for student projects collaboratively planned between classrooms at different sites.

Some examples of ongoing and future projects include the following:

- The "Global Village" project is currently conducted through the Center for International Business Studies, Texas A&M College of Business Administration. Some of the topics covered in past compressed video sessions have included (1) comparison of communism, socialism, and capitalism, (2) NAFTA, (3) the European economy, and (4) Japan.
- The "Totally Texas" project is a collaborative effort of people from several TEC sites who are creating an integrated multimedia curriculum for 4th and 7th grades. The material will be stored on CD-ROM. Sites connect via the distance learning system either weekly or biweekly to discuss the various units each site will contribute to the curriculum. Teachers share ideas and resources with each other without having to travel great distances to do so.
- Distance learning was used in conducting a presentation on technology usage to legislators in Austin. Legislators were able to talk directly to students and teachers at TEC sites regarding their use of distance learning and computer programs in instruction. Students and teachers were able to demonstrate various student projects and computer programs to TEC sites and legislators at the same time. Eight sites throughout Texas—Lubbock, Laredo, the state capital in Austin, and the five TEC sites—were online at the same time.
- A teleconference via compressed video is planned between students in Sweden and students at a 4th-grade class in Bryan, Texas. The students plan to exchange information about a typical day in their schools and towns. This project is being funded by the Swedish government. Another distance learning project includes students from Somerville, Texas, sharing their multimedia projects with students in Laramie, Wyoming.

#### **Internet**

Several Turkish schools will have connections to the Internet in early March with students in Texas. They will become "keypals."

### **Multimedia Projects for the Classroom**

Undergraduates enrolled in a technology course on the Texas A&M campus meet via compressed video with practicing teachers at a TEC site to develop instructional materials that support interdisciplinary themes. Materials have been created to support such themes as "Africa" and "Westward Expansion."

64

## RegNet: An NPR Regulatory Reform Initiative Toward NII/ GII Collaboratories

*John P. Ziebarth, National Center for Supercomputing Applications*

*W. Neil Thompson, U.S. Nuclear Regulatory Commission*

*J.D. Nyhart, Massachusetts Institute of Technology*

*Kenneth Kaplan, Massachusetts Institute of Technology*

*Bill Ribarsky, Georgia Institute of Technology*

*Gio Wiederhold, Stanford University*

*Michael R. Genesereth, Stanford University*

*Kenneth Gilpatric, National Performance Review NetResults.RegNet and  
Administrative Conference of the United States (formerly)*

*Tim E. Roxey, National Performance Review RegNet. Industry, Baltimore Gas and Electric, and  
Council for Excellence in Government*

*William J. Olmstead, U.S. Nuclear Regulatory Commission*

*Ben Slone, Finite Matters Ltd.*

*Jim Acklin, Regulatory Information Alliance*

### SUMMARY

The premise of this paper is that the pressures of U.S. growth into a national economy, associated societal complexities, and advances in present information technology are rapidly leading the United States out of the post-Industrial Age and into the Information Age. The relationship between the various components of our government and its citizens in an Information Age are substantially different from those found in the post-Industrial Age. The actual structure of our government has been tuned for an Industrial Age level of societal complexity and technology sophistication, and it is being challenged by current conditions.

Given that the preamble to the Constitution of the United States was probably the first written occurrence of a national goal and objective, the original structure of the government was crafted so that the implementation of government, or governance, allowed citizens to achieve the preamble's intent. For individuals for whom this was not the case, governmental process allowed changes to be made.

Regulatory reform initiatives today, especially those involving the implementation of technology, have a significantly greater degree of difficulty in being able to allow citizens to enjoy the intent specified in the national goals and objectives. This is predominantly due to the lack of a national information plan that deals specifically with mapping citizens' needs and rights to regulatory information to national goals and objectives. When an entity consistently fails to deliver to its customers the products that allow the customers to enjoy the achievement of the entity's goals and objectives, then that entity is doomed to fail. If the entity is a federal, state, or local government, then that entity is not permitted to fail and action is required to be taken.

In the following sections, there is a development of several key attributes. These attributes are complexities in regulatory interaction, complexities in implemented levels of technology, and unit of governance involved in the interaction. The discussion shows that when societal complexity and/or technology advances cause government to be engrossed in too fine a level of detail that is too fine, then the structure of government changes to accommodate the complexities. The most dramatic revision so far has been the development of the fourth branch (referring to the creation of the regulatory agencies and programs) of our federal government.

It is time now for the various levels of governance to redefine themselves and their customers and return to a citizen-centered philosophy.

## STATEMENT OF THE PROBLEM

Perhaps to fully understand RegNet, it is best to understand RegNet's customers and the problems that they experience. First and foremost among the customers is the private citizen in the United States who is seeking regulatory information or to have an impact on changing the system. Next there are citizens who form themselves into public interest groups, typically centered on a particular topic or concern. Then there are corporations—both for-profit and not-for-profits—that represent various commercial or private interests. Finally, there are various agencies of government at the local, state, and federal levels.

Each of these customer groups will pursue regulatory information for different reasons. The individual citizen may seek regulatory information to achieve some level of assurance that the regulations are safe and sane and adequately deal with the citizen's concerns. Public interest groups will typically look through regulations in order to find out how the public might better comply with them. In some public interest groups, the intent is to become better aware of the regulation in order to lobby industry or government to more effectively meet the requirements. Corporations will use the body of regulatory information in their day-to-day operations to remain in compliance with the rules as they pursue their corporate goals and objectives. Finally, the various regulators will look at their regulations in one of two different ways. The first is in their role of oversight, or watchdog, and in this case they will review regulations to inspect or enforce them on the industry sector they oversee. The second way the government uses the regulatory information is with the intention of fine-tuning or improving it.

Unfortunately, because of the vast complexity of government today, these RegNet customers typically become frustrated in their pursuit of information. This complexity stems from the many agencies and organizations (international, federal, state, local, and tribal) that issue regulations independently and without consistency. The breadth of this problem encompasses all regulators and regulated entities (industry, government, and the private sector).

This frustration of dealing with government is complicated by the inability to deal with the immense collection of regulatory information using present paper-based methods of access. In addition, the communications between people and their government is limited by technologies (telephones, fax, personal meetings) that have not been able to deal with the enormity of these complex information structures.

The primary cause of this frustration is that the national goals and objectives are not being met by the current implementation of technology.

This discussion, though complicated and fundamental, is centered on the least complex form of governmental interaction—the one-to-one model. This simplest interaction is used to enforce rules and to gather or publish information. The other three forms of interaction, discussed in the background section below, are all more complex than the previous.

All four of these levels of regulatory interaction have serious problems associated with them. How do we know that this is the case? Customer satisfaction is low, costs to maintain the bureaucracy are high, and the effectiveness of many regulations is poor to nonexistent. Even the anecdotal stories of regulatory abuses and excess are making for best-selling novels these days. The solution is to revise the regulatory structure—the current effort is called either reinvention or reengineering. At best, reinvention will blend intelligent technological implementations with thoughtful structural modifications in such a way that quality regulatory services are delivered to as wide a customer base as possible. This is the so-called citizen-centered form of governance.

## BACKGROUND

The six planning and demonstration projects contemplated in this paper are best understood as having three conceptual dimensions, as shown in [Figure 1](#).

One dimension is technological. Each project is defined as having a locus on a spectrum of increasing sophistication in communication and information technology. By accurately situating a given project on the IT-sophistication dimension, we can begin to see the interrelatedness of all six projects. We are also able to resolve questions of how, when, and where—with respect both to conducting the project itself and to operation of the project's end product by its eventual users.



A second dimension relates to regulatory functions. Each project has its space on a spectrum of increasing forum complexity or multiplicity of interested entities. Again, by carefully defining the position of a particular project along a dimension of regulatory-function complexity, we can appreciate project interrelatedness. We can then address the issues of who will or should be involved in project operation and eventual product use, and what exactly they will or should do.

A third conceptual dimension embraces the entire list of governmental units and agencies, graded by increasing scope—local, county and municipal, tribal and state, federal and international—each governmental unit with a discrete set of "customers" and entities subject to its authority. By relating our six projects to such a spectrum of agencies and associated client sets, bearing in mind national goals and objectives, we can approach an understanding of why these proposed projects are crucial and should be funded.

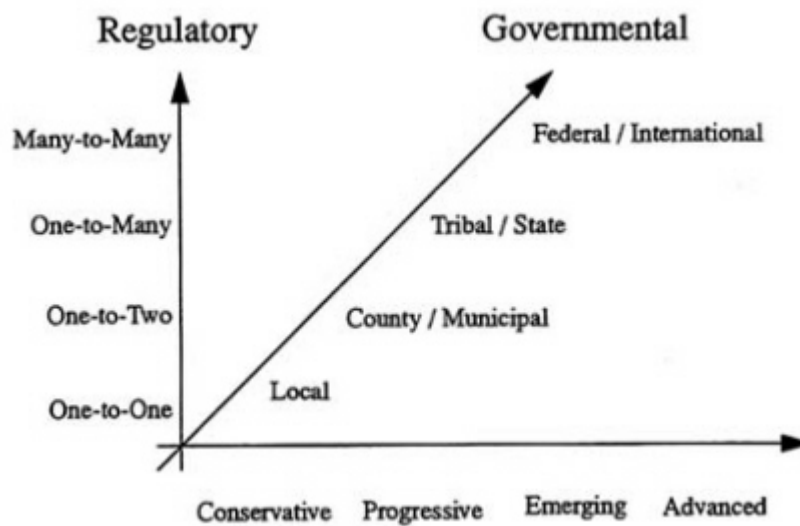


Figure 1 Complexities owing to three conceptual dimensions of demonstration projects.

### X-Axis: Information Technology

Information (the content) and technology (the conduit) have been with us for a long time. No doubt, properties of the conduit have always had a heavy effect on the content, just as qualities and purposes of the content affect conduit developments. Acknowledging the reciprocal relationship and importance of the two, it is helpful to keep them distinct in analysis. To the degree that conduit and content can be considered separately, the former is discussed in this section. The following section, on regulatory functions, treats content. Following that, in a section on governmental units and their customers, we treat context—the considerations that give relevance to conduit and content.

Students of the human history of information often parse the subject into four general periods, reflecting available and utilized technologies. Earliest was an oral or bardic culture, when human memory was the chief tool for information exchange (perhaps supplemented by musical instruments, cave paintings, and carved objects). Later, scribal culture arose, adding to the serviceable toolkit such devices as pen, brush, ink, paper, and expert familiarity with their combined use. In the fifteenth century, the use of movable metal type in printing presses changed human society fundamentally. More recently, the theory goes, a similarly profound shift has started—digitization of information. It will affect human thinking, working, playing, and relationships in unknowable ways—regarding entity boundaries and properties, groups, ownable or meaningful objects, transactions, and disputes about all of the above.

The foregoing historical framework is a useful aid for understanding issues we face today. Our problems stem in significant part from cultural conflict. The entities, habits, tools, methods, and competencies of print

culture are in the process of incomplete and uneasy evolution into those of digitized-information culture. Letter press was the first technological invention enabling the print or book period. Photography, telegraph, typewriters, photocopiers, and faxes expanded the power of printed word. Those methods form the core of a conservative model of IT sophistication.

Technological advances in the recent two decades and growing user familiarity with those methods suggest that there may be a very detailed spectrum of sophistication in information technology, but we offer below a simplified scale for sorting the proposed six projects.

We describe four marks on the spectrum of information technology sophistication:

1. *Conservative model of IT sophistication*—paper documents, photocopying, telefax and hard-copy communications, physical libraries, face-to-face meetings and forums.
2. *Progressive standard model of IT sophistication*—electronic documents and copying, e-mail and file-transfer communications, non-physical libraries and transaction forums (e.g., BBS, commercial on-line services).
3. *Emerging model of IT sophistication*—electronic information objects and copying, graphical data representation, nonimmersive virtual reality techniques, virtual transaction forums (e.g., WWW, Mosaic).
4. *Advanced model of IT sophistication*—virtual 3-D interfaces and data representation, competent AI agents, immersive VR.

### Y-Axis: Regulatory Functions

Before discussing regulatory functions, we should step back and look at functions of government more broadly. Regardless of conduits available, one function of rulership is the gathering and publishing of information. Ancient polities in Egypt, Mesopotamia, and Rome, whatever their varied structures and operational differences, all gathered information from their subjects and the world at large, processed it in some fashion, and disseminated it. Information typically collected would include, for instance, head counts, land boundaries, tallies of wealth and capability, and assessments of intention. Disseminations commonly took the form of rules, orders, decisions, and descriptions.

More modern governments—for example, those of Florence in the 1500s, France in the 1600s, and Britain in the 1700s—were information entities just as the ancient polities had been, but the moderns had a quantitatively different problem. Often they had larger populations and accumulations of wealth, more sophisticated communications technology (conduits), and certainly more complex data, both to gather and to publish. Those modern polities handled the increased volume and complexity, generally, by inventing tiered and nested organizational hierarchies with specified subject matter authority—offices, bureaus, ministries, and committees.

In the 1780s, the founders of the United States chose a three-part form of government that we are all familiar with: the legislative, executive, and judicial branches. One was to write laws, another to administer them, and the third to resolve cases and disputes. Each functional power was separated from the other two, counterweighting and limiting them. Our country's founders were not openly concerned with informational volume and complexity. They aimed instead to secure entities and their properties by averting in the long term both the tyranny and the license, respectively, of too little and too much liberty.

That constitutional division of government functions has served us remarkably well. However, though it may well have sufficed for the degree of informational complexity foreseen by the founders, it did not anticipate the industrial revolution and associated advances in information technology that would unintentionally engender an extensive increase in complexity. When in due course informational volume and complexity became too great for government to handle, threatening to overwhelm resources of the federal executive, judiciary, and legislature with details, a fourth branch of national government was invented: the administrative and regulatory agencies (Figure 2).

Regulation as we understand it today was largely unknown during the first century of this republic. Individual states had engaged in economic regulation for a long time (of turnpikes, bridges, mills, foodstuffs, civic hygiene, and licensing of physicians, for example). But federal involvement with private productive enterprise

was limited to canals, the national highway system, and banks. A strong revenue base, a trained civil service, and systematic economic information—necessary conditions of a modern regulatory state—were lacking until after the Civil War.

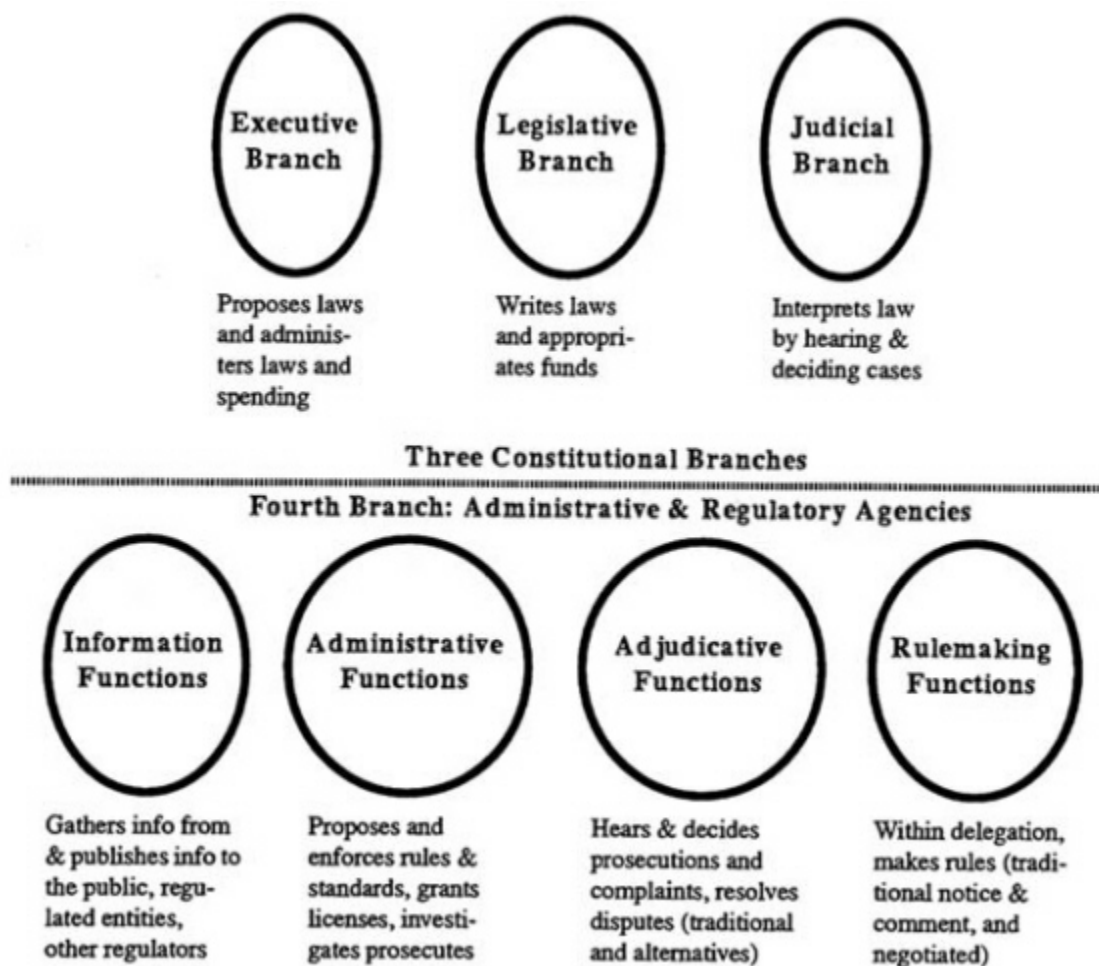


Figure 2 Three branches of government plus administrative and regulatory agencies as a fourth, to handle increasing volume and complexity of information.

Requirements of organized ingenuity, production, and distribution for nationwide conflict combined after the Civil War with private economies of scale and a competitive, almost predatory work ethic (expressed in the doctrines of Social Darwinism and *laissez-faire*). Results of the combination were dense concentrations of wealth and productivity in transportation (rails, shipping), commodities (oils, steel, sugar), and finance. Also, significant abuses of market freedom occurred, some groups were treated with shocking unfairness, and gilded figures flaunted their success and power. The corrective was a wave of regulation. It started tentatively with the Interstate Commerce Commission (1887) and the Sherman Act (1890) and continued, massively, through the Progressive Era (1902–14).

Two other great regulatory expansions happened later. One we call the New Deal (1933–38). Chiefly, it reallocated power among economic interest groups. The other began in the late 1960s and lasted about a decade. In it, political issue groups sought substantial reform of civil, social, and consumer rights, environmental protection, and renewed political responsiveness after Vietnam and Watergate. Between each of the major expansionary waves, there were interglacials (so to speak) when implementation, adjustment, consolidation, and reassessment took place with respect to the preceding regulatory impetus. The oscillatory cycle evidently runs like this: Perceived abuses or inequity prompt disillusionment with a free market; regulatory discipline is applied

to remedy the market failure; inefficiency or ineffectiveness in due course spoils the regulators; and market efficiencies and discipline are pressed forward to correct the regulatory failure.

As we have suggested, administrative and regulatory agencies arose in part as a consequence of this mounting complexity of information. In addition to the function of information gathering and publishing, agencies typically included the three constitutional functions. The traditional and general branches—executive, judicial, and legislative—in effect allowed a limited portion of their respective functions to be delegated to expert and focused subordinate agencies. In concept, these agencies would spare the executive branch, the Court, and the Congress from detailed and repetitive chores of administration, dispute resolution, and rule making. In the ordinary scheme, agencies would report to both the executive branch and Congress, and agency decisions would be subject to review by the Court. An agency's four functions would each connect and operate in different ways with its customer set, from an information base held partly in common.

These four functions are described below.

1. *A one-to-one exchange between entities.* The most primitive level of forum complexity is a one-to-one exchange (Figure 3). Two entities, face to face, engage in a transaction. Entity and face may be virtual or electronic; matter transacted may be hard copy or digitized; an entity may constitute a group of subentities or itself be a subentity of a larger organization; the transaction may be asynchronous or real-time, enduring a millisecond or a decade—none of these local circumstances alters the essential topology of the one-to-one exchange. An agency's information gathering and disseminating function fits cleanly at this primitive level. When a private entity seeks information from an agency (a FOIA request, say, or a general inquiry about policy and procedure), or when an agency asks the private entity to forward information (a tax return or application form or compliance data, for instance), the relationship is one-to-one. Much of an agency's administrative function also fits in this category. Investigation, inspection, granting of licenses, and specific enforcement of rules all can be visualized as an administrative officer on one side of the table and a regulatory subject entity on the other.
2. *A one-to-two relationship.* Slightly more complicated is the one-to-two entity relationship (Figure 4). It is the model for most adjudicatory or dispute-resolution proceedings. A judge or hearing officer or neutral observer sits on one side of the table; on the other side sit two adversarial parties. There may be exchanges between the two adversaries (e.g., discovery or settlement discussions), but transactions with the forum officer tend to be bound by formal requirements. Traditional adjudication and alternative dispute resolution (ADR) are both accommodated in this category.
3. *A one-to-many relationship.* This category comprises traditional rule-making scenarios or any proceedings in which a single officer or official body sits on one side of the table and multiple interested parties sit on the other (Figure 5). Again, it does not matter that the commissioners transact decisional information among themselves or that groups of interested parties form coalitions and exchange information. The topological complexity is formally undisturbed: one to many.
4. *A many-to-many configuration (Figure 6).* The many-to-many relationship can be conceptualized in two ways. It can be seen as an extreme development of the one-to-many configuration, when the single officer on the one side of the table is reduced in authority from central decisionmaker to forum moderator, meeting president, or coequal participant. A legislative assembly or negotiated rule making fits this pattern. Alternatively, this category can be visualized as a roundtable forum without chairing officer or moderator and with all entities having equal power to affect decisions of the forum.

Our account of regulatory development is intended as a summary and overview, and may be too simplified. There have been many efforts over the years to reform regulation—both its substance and its procedures—in narrowly confined areas within a given agency and across broad sweeps of government. Certainly, our short history gives no idea of the extensive variety of regulatory function practiced today and the detailed articulation of those functions with private entities. In important ways, the detailed variety and extensive articulation are aspects of the problem that our six proposed projects are designed to address.

For the purposes of this paper, we have contrived a measuring system for the regulatory function dimension of analysis, a sequence of check marks. It accommodates all of the variety of agency functions as well as the myriad contemporary articulations of regulator with regulate or other external entity. It does so in a

Figure 3 A one-to-one-exchange.

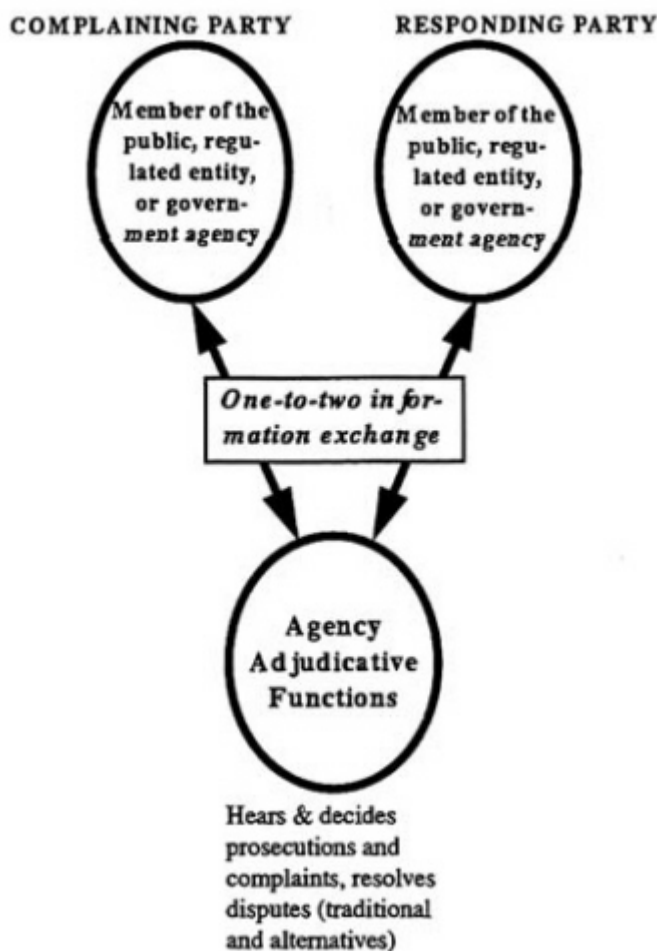
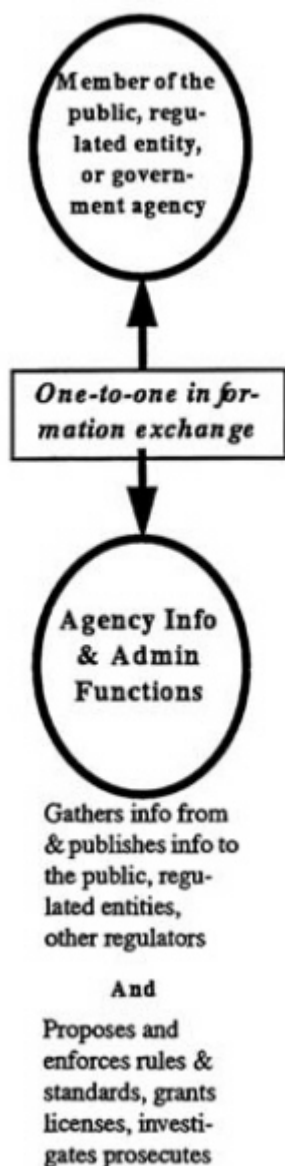


Figure 4 A one-to-two relationship.

fashion that seems sensible and accessible both to students of governance and administrative law and to experts in communications and information technology and software engineering. It may be that there are many workable measuring schemes. We develop one below, borrowing substantially from formulations in microeconomics, information physics, and psychology.

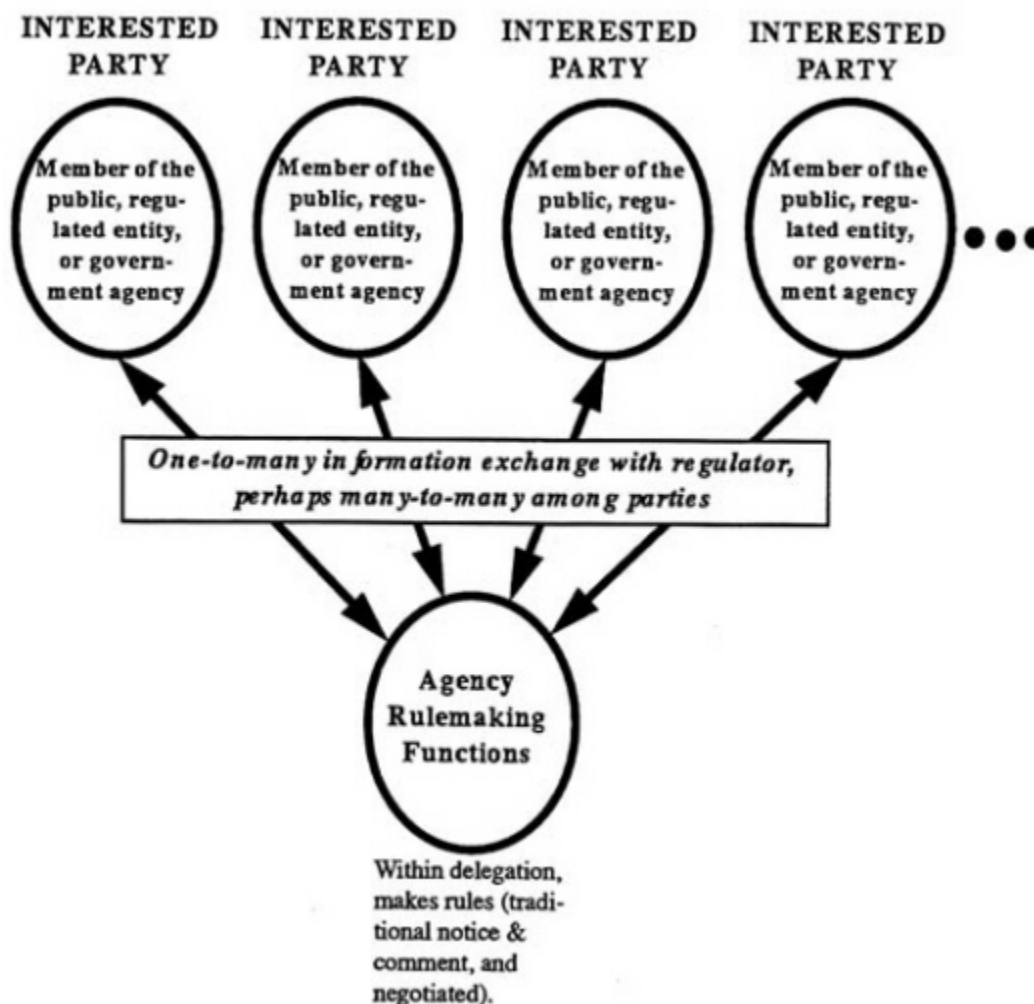


Figure 5 A one-to-many relationship.

### Z-Axis: Government Agencies and Customers

A detailed exploration of business process reengineering or government reinvention is beyond the scope of this paper. But two central principles from those undertakings inform our discussions here: (1) productive entities or units exist, and their current boundaries and methods of operation often reflect historical accident, ingrained habit, or risk aversion; and (2) both entity boundaries and productive processes frequently need to be reengineered so as to be customer driven, market and competition oriented, enterprising, decentralized, and community responsive.

Understandings of public entity "effectiveness" and "efficiency" differ from those applicable to private, profit-making entities. And the concept of a public entity's customers or clients is not the same as that applying to a private entity. Still, there are strong analogies that assist analysis.

We hear anecdotally that some 82,000 governmental units have been identified within the United States (not including metanational entities, such as treaty organization or international dispute-resolution bodies). Of course, each governmental unit has various sets of clients or customers or respondent entities grouped according to subject matter issues for formal proceedings. A comprehensive map of these groups would be dizzyingly complex, but a simplified scale is developed below that allows us to tear apart the complexity, to locate governmental units and their customers along a spectrum of entity scope, and to describe regions where each of the six proposed projects are relevant both as to project operations and as to end-product users.

There are four check marks on the government entity axis:

1. Local governance and associated customers (e.g., PTAs, neighborhood councils, tenant associations, town meetings, and subcounty authorities),
2. Counties, cities, or municipalities (e.g., government offices, school boards, or zoning and natural resources authorities),
3. Tribal or state governance units (e.g., courts, agencies, legislatures, and executives), and
4. National or international governance units (e.g., the federal sovereign and its branches, and international entities).

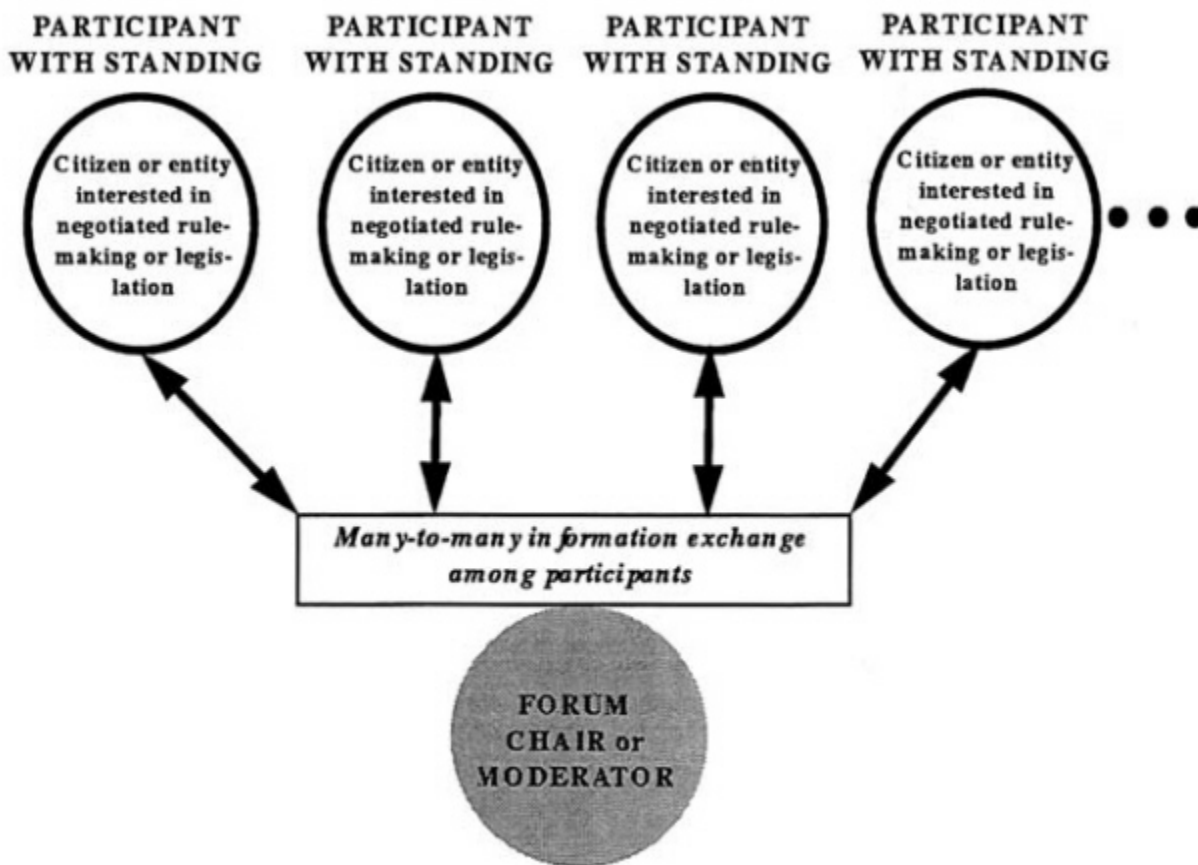


Figure 6 A many-to-many configuration

### THE PLAN

This discussion includes the development of a comprehensive national plan called RegNet that supports national goals and objectives vital to economic prosperity and security. The RegNet national plan integrates (1) ongoing reinventing government and regulatory reform initiatives, and (2) advanced high-performance computing, communications, and information technology with (3) intergovernmental, interacademic, interindustry, and public interests. This plan is an integral part of the proposed technology development and demonstration projects initiated in this paper.

The following is a summary of the RegNet national plan attributes, which reflect human and technological needs, two separate but interdependent requirements to meet in furthering national goals and objectives.

### **RegNet—A Human and Technology-based Definition**

- The RegNet national plan is a system that links human communications and information (HCI) needs to regulatory information and regulatory processes (RIRP) through high-performance computing, communications, and information technology (HPCIT).
- It is multilevel based (intergovernmental, interacademic, interindustry, and intercitizen).
- It supports all users who want or need to access regulatory information and/or to participate in regulatory processes (international; national, including federal, state, local, and tribal governments; industry; and the public)—for example, all regulators and all regulated entities (industry, public, private, and nonprofit sectors; and the public).
- It is a national, integrated, intelligent, distributed, multimedia, World Wide Web (WWW)-based regulatory communications and information system and people network.

### **RegNet—A Citizen-based Definition**

- Under this definition, RegNet is an advanced regulatory information system that provides the public and industry with easy access to one-stop shopping for the regulatory information and participation needed to (1) reduce regulatory burden, (2) increase efficiency, effectiveness, and profitability, and (3) become more economically competitive both nationally and globally for economic growth.
- It provides full access to (1) all intergovernmental regulatory information (existing laws, rules, and regulations), and (2) communication links for participating in regulatory processes (regulatory compliance reviews; making, modifying, or eliminating laws, rules, regulations, policies, standards, guidelines, and positions; and alternative dispute resolution and negotiations).

### **RegNet—A Concept and Initiative for Regulatory Reform and Economic Development**

- RegNet is an interdisciplinary consortium of intergovernmental, interacademic, interindustry, and citizen entities.
- It is a grand challenge (as defined in the High Performance Computing Act of 1991).
- It supports national challenge applications (as defined by OSTP, National Coordination Office for HPCC Information Infrastructure Technology and Applications).
- It is intelligent (as defined by ARPA and others).
- It is distributed (globally).
- It is intergovernmental (international, federal, state, local, and tribal governments), interacademic, interindustry, and intercitizen based.
- It provides citizen and industry access to regulatory information.
- It provides opportunity for intercitizen, interindustry, and intergovernment participation in regulatory processes and in regulatory reform.
- It is of the people, by the people and, for the people.
- It requires cooperation, collaboration through coordination, and integration at all levels (see citizen-centric model).
- It is a comprehensive regulatory reform initiative that plans for the use of advanced computing, communications, and information technology.
- It supports all national and international user levels through the national information infrastructure (NII) and global information infrastructure (GII).

Such a plan implies the full integration of appropriate and applicable data, information, and material in specific domains of interests. This is also the objective of the Vice President's National Performance Review



(NPR) RegNet project—to coordinate the planning, development, and implementation of a distributed, intelligent regulatory information system. This plan will include consideration of both national and global information infrastructure environments.

This plan recognizes that the need for information and material in the private sector is different from that of the public and nonprofit sectors. For example, governments need information to develop laws, rules, regulations, policies, procedures, and practices, whereas the public and industry need access to government processes to become better informed and more able to participate in the decisionmaking processes that affect their future. In addition, the private sector's needs derive from the desire to be in compliance with regulations and associated processes. With information fully distributed to all interested parties, individuals will be adequately prepared to support national goals and objectives.

### Deployment Issues

Only nine participating HPCCIT federal agencies are coordinated by the National Coordination Office for High Performance Computing and Communications (HPCC). This coordination is tied directly to the nation's economic policy in support of national high performance computing (see the High Performance Computing Act of 1991). There are no other federal policy level offices responsible for coordinating HPCCIT national goals and objectives, leaving more than 130 remaining federal agencies, departments, and commissions without HPCCIT policy coordination and integration. This lack of coordination and integration results in inconsistencies, inefficiencies, wasted time, and misdirection of precious human resources and funding. The eventuality and schedule for RegNet deployment rest, in part, on enforcement of HPCCIT policy throughout the federal government, but more importantly on intergovernmental cooperation, collaboration, coordination, and integration, perhaps starting with the National Governors' Association.

As it approaches the twenty-first century, the United States faces challenges that transcend state lines and national borders. This country must shape economic policies for a competitive marketplace. RegNet will contribute to HPCCIT, NII, and GII by playing a vital role in meeting national economic and security goals and objectives.

The NCSA RegNet project addresses and responds to these challenges by underpinning a plan and demonstration project that is based on national goals and objectives (reinventing government, regulatory reform, grand challenge applications, and national challenge applications vital to national prosperity, including NII and GII initiatives).

The issues are important ones for the country's federal, state, and local governments and, more significantly, for its people and businesses. The proposed projects directly support national goals and objectives of NII, GII, and intergovernmental regulatory communication and information needs. Also, the following national challenge applications have been identified by the HPCCIT subcommittee:

- Public access to government information;
- Health care education and training;
- Digital libraries;
- Electronic commerce;
- Advanced manufacturing;
- Environmental monitoring; and
- Crisis management.

This discussion supports the development of a national plan that would be able to encompass and embrace disparate demonstration projects affecting international, federal, state, local, and tribal governments as they affect American citizens. We describe five demonstration projects in federal, state, city, tribal, and rural environments. In essence, a needs assessment of stakeholders has been performed by Congress and the executive branch (e.g., bills and hearings in support of the High Performance Computing Act of December 1991, and subsequent legislation, and executive orders). Regulators, the regulated, and the public need easy access to

regulatory information. Intergovernmental regulatory collaboration that depends on this easy access to information requires the planned use of advanced communication and information technology initiatives. These initiatives have been supported in the Vice President's NPR, the Office of Science Technology Policy (OSTP), industry groups, various federal regulatory agencies, and academic and research institutions. Regulatory reform is under way and needs the vision, coordination, and integration planning recommended here.

An intelligent, three-dimensional visual model will be developed to help communicate this comprehensive RegNet national plan and to illustrate the communications and information interrelationships, including those among the various government, academic, and industrial organizations and the citizen. The collaborative partnership refers to this model as the citizen-centric model since it focuses on the people of the United States. Preliminary visual models have already been developed for previous NPR meeting and proposals. The first is a two-dimensional concentric circle diagram representing the RegNet national plan GII. At the center of this diagram is the frustrated citizen desiring access to regulatory information. Each succeeding concentric circle represents an organizational entity (level) that has access to the same regulatory information, and regulatory process, as the citizen. Combined, these circles (levels of users) represent a national regulatory information infrastructure. In turn, each country of the world (more than 180 countries) will eventually have an infrastructure in support of similar levels of regulatory information users and will become a part of the GII. A pie-shaped sector of these concentric circles represents all regulatory information users in each state of the United States. These levels (international, federal, state, local, and tribal governments in addition to national laboratories, industry, research, and academic institutions) form the model for intergovernmental, interacademic, interindustry, and public coordination and integration of the RegNet national plan. Different organizational entities within each state can have interrelationships both internal and external to that state. For other nations, the model can be replicated with their appropriate organizational entities represented.

A second three-dimensional, intelligent, interface model will be developed for RegNet. This visualization model represents the entire intergovernmental intelligent distributed regulatory information system model. The GII represented here includes (1) an intergovernmental regulatory definition model, (2) an interindustry facility/product definition model, (3) a cost-benefit, risk assessment, economic assessment definition model, and (4) an analytical definition model. All regulations for all governmental bodies (international, federal, state, local, and tribal), all industries (all sea, land, air, and space industries), all hard analytical tools (structural, mechanical, electrical, architectural, illumination, acoustical, etc.), and all soft analytical tools (cost-benefit, economic, risk, and environmental assessments) are incorporated into this intelligent interactive visualization model. The RegNet national plan will incorporate regulations and regulatory processes using intelligent, distributed computer systems, will access the information from anywhere, and will make the information comparable and integrable for the customers. These features include the integration of the enterprise-level information and sophisticated human interface characteristics.

The concept of an intergovernmental regulatory collaborative, called RegNet by the NPR, came about after many fragmented and independent attempts made by government, academic and research institutions, and industry groups to integrate and coordinate regulatory communications and information needs of the regulator, the regulated, and public entities. RegNet, as an intergovernmental regulatory collaborative partnership with the people and their institutions, will supply communications and information technology needs to a broad clientele (including research, development, and application teams). The initial concept of a "collaboratory" was proposed as a "center without walls, in which the nation's researchers could perform their research without regard to geographical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries." This discussion incorporates the "collaboratory" concept for use by all public, private, nonprofit sectors and the public.

The RegNet national plan encompasses all regulators and regulated stakeholders (industries and the public) under the reinventing government and regulatory reform initiatives of NPR. It also will support researchers' needs for such collaboration. Its grassroots base can provide generic distributed benefits across the nation. The plan will depend on a global network of computers to facilitate the performance of its functions and is envisioned to offer a complete infrastructure of software, hardware, and networked resources to enable a full range of collaborative work environments among intergovernmental, interacademic, interindustry, and public sectors.

An NII initiative is being implemented in the United States, of which current, large-scale regulatory movements should become a part. The deployment and use of HPCCIT and the NII are both essential to meeting goals and objectives in the global marketplace. Regulatory reform is thus part of a movement with a great positive inertia behind it.

### THE DEMONSTRATION PROJECTS

For the past two to three decades computing has grown from the dominance of large mainframe computers running in information system department glass palaces to desktop personal computers with slow processors to very high speed performance through special hand-built processors on small compact supercomputers to the era of the microprocessor having extremely high power in a desktop machine. Applications on these new desktop machines scale easily to high-performance computers with many of these same microprocessors tied together. In the past, people for the most part have been able to ignore the use of these desktop machines if they chose to, or if they worked in an environment where computers were absent. Many professionals relied on their clerical staff to use computers for word processing and spreadsheets, but their own involvement beyond this level was usually not required and is still often avoided. Traditionally, only those involved in science and engineering pursued faster and more widespread computing and networking technologies.

Beginning in 1993, when the National Center for Supercomputing Applications released Mosaic, a WWW browser, a fundamental and permanent change began to evolve in all sectors of society, which are now pursuing connections to the Internet and are installing current-generation computers at home and at work. This is not taking place because of the computers themselves, not even because of the Internet, since that has been around for over two decades. What is different now is the human interface. Through the use of WWW technologies and easy-to-use graphical browsers, information and the ways we learn, discuss, and collaborate have become primary commodities on the Internet. Nothing will remain the same in any segment of society from this day forward because our society at large is on the brink of the Information Age. The "customers" within our society are beginning to demand technology and the information it brings: students and parents are driving teachers to be able to access new information sources, teachers are driving administrators to connect schools, purchasers of goods and services are going to the WWW for products and information about products, and citizens are beginning to view and communicate with the government electronically and are expecting services to be available via the Internet.

Regulatory reform has been discussed for many years. Regulations at the local, state, and national level consume citizens' time and money at an alarming rate and create for the nation a complex, unmanageable system that yields no obvious advantage. For example:

- Edward Crane, president, and David Boaz, executive vice president, regulations, Cato Institute, describes regulation as "a costly, wasteful, and job-killing burden on the economy" (*Washington Post*, February 6, 1995).
- Stephen Moore, director of fiscal policy at the Cato Institute, calculates that regulations add as much as 33 percent to the cost of building an airplane engine and as much as 95 percent to the price of a new vaccine.
- Cost to the economy of regulation—\$880 billion to \$1.65 trillion per year, according to William G. Laffer of the Heritage Foundation.
- Cost to the economy of regulation—\$400 billion, or roughly \$4000 per household, according to Thomas Hopkins of the Rochester Institute of Technology.
- A Tax Foundation report found that the tax compliance burden is 10 times heavier on a corporation with less than \$1 million in sales than on one with \$10 million in sales.

Information technologies and the tools of communication across the WWW, which are evolving at an incredible rate, have finally given the nation hope that regulatory reform is possible. It would be total chaos to try to address every possible demonstration project of how technology can be used to solve or overcome obstacles to regulatory reform. Our decision in this paper and in the related documents has been to choose five projects that

will bring together the primary tools as they are now viewed and to apply them to significant but reachable goals and demonstrations. The partners in this set of demonstrations represent industrial, academic, and government organizations that will form the basis for extended collaboration across the nation by all segments of society on regulatory problems in every area of society.

Collaboration software is one cyberspace tool that will become as common as electronic mail for users in just a few years, and rule making is one important regulatory activity that will benefit from collaboration across the Internet. With current and evolving tools, the WWW will allow various segments of society to work together, electronically, to decide on a regulation and to automatically manage the process of consensus and measurement related to a proposed rule. Virtual negotiations and adjudication of regulations, all of which will be electronically available, will allow all segments of society to play a part in the regulatory process. With regulations becoming a part of an openly available electronic database for the nation, it will become essential that access to them begin to be automated and intelligent and that analysis tools—from 3-D visualization tools that are accessible across the WWW to virtual reality (VR) analysis tools, which require immersion and high-performance computing—become automated and effective. The demonstration of these tools in a rural government test bed is essential and puts measurable and meaningful results out for open debate and evaluation. It must be investigated how these tools can be used effectively by the general citizenry of the nation in real everyday situations requiring interaction with government at all levels.

These projects will push the limits of available technology in a regulatory environment and will drive future developments tying technology to information and regulations. The collaborating institutions represent a broad, experienced basis of institutions for other academic, industrial, and government organizations to become involved with in research and demonstration collaborations.

The five technology demonstration projects represent the kinds of technologies that are able to reshape the way people interact with their government (human interface and visualization technologies). These projects, along with the RegNet national plan, described above, provide a success path that governments at all levels can follow to move their regulatory functions effectively into the Information Age. Other methods that do not involve a linking plan and demonstration projects are prone to be less successful.

These projects were selected because there is some existing development under way and because there is a need to integrate their technologies into the overall NII plan. In addition, there is a requirement to formally understand the links between these projects and the national goals and objectives. These projects will leverage the existing NII activities (HPCC, ATP, TRP, etc.) It should be noted that these projects leverage significant investments from industry, not-for-profits, and other publicly interested activities.

### **Demonstration Project One: Collaborative Participatory Rule-Making Environment— RuleNet**

- Regulatory complexity: Significant
- Technological complexity: Emerging
- Government unit: Federal
- Principal investigator: Bill Olmstead, U.S. Nuclear Regulatory Commission (U.S. NRC)

This project will develop and demonstrate the concepts associated with a collaborative rule-making workspace utilizing the Internet and WWW technologies. This type of workspace is being developed by an intergovernmental partnership along with some interindustry support. One of the first uses for this type of workspace (that is, the content for the rule making workspace) is in the area of collaborative participatory rule making in the nuclear energy domain. This places this project at the significant level point on the regulatory complexity axis and at the emerging point on the technological complexity axis. Additionally, since the topic under discussion will be some proposed rule within the nuclear industry domain, regulated by a federal agency (the U.S. NRC), this places the project at the federal level on the governmental unit axis. Being at the federal level of primary governmental involvement means that there can potentially be many thousands of participants on

one side of the discussion and several (regulators) on the other side. Additionally, there can also be participation both by industry consortia groups and by public interest groups.

This type of workspace environment will be central to many of the national challenge applications where public interactive access is a requirement. This project is designed to test technology as it applies to this mode of regulatory interaction.

This first project actually is composed of two primary work efforts, a part that has already been planned and is being funded now (caught on the fly as it were) and extensions to the original planned work scope that will allow this project to more readily fit within the overarching RegNet scheme. By capturing the first part of this project in this way, it is possible to build quickly upon the results of already planned activities with minimal interruptions to work scopes and schedules and to gain maximum benefits for the overall efforts.

The intent as represented here is to more fully support the initial, already planned phase of this project and then add to it the pieces that allow the links to the national goals and objectives, the national regulatory information plan, and the other projects within this grouping.

### **Phase One: Activities Captured on the Fly**

This first phase of the project represents a new initiative with U.S. NRC. But Lawrence Livermore National Laboratory (LLNL) has conducted two previous electronic forums or dialogues in partnership with the NPR: the Vice President's Electronic Open Meeting in December 1994, and the HealingNet initiated after the April 19, 1995, bombing of the federal building in Oklahoma City. A third electronic forum is being built for the discussion of federal acquisition reform (FARNet, one of the several ongoing NPR NetResults activities).

LLNL will provide all of the technical and administrative support for the RuleNet project including, but not limited to, document preparation and conversion to HTML and other formats as appropriate, technical expertise in the form of an expert/facilitator, technical analysis and categorization of comments, "listproc" distributions, WWW home page development, Mosaic-type server with file transfer protocol and Telnet capabilities, telecommunication links, document analysis tools, creation of separate electronic meeting areas where registered members may caucus, and other tasks as necessary. It is necessary for the contractor to assign technical staff to perform the tasks as outlined. Each staff member will have the appropriate expertise to accomplish their function within the assigned task. The key position of expert/facilitator is considered essential. Upon the acceptance of the project by the U.S. NRC and DOE, and acceptance of U.S. NRC funding by DOE, LLNL will mobilize a team consisting of engineers, computer scientists, graphic artists, technical editors, and a facilitator/moderator/coordinator to design and lay out the RuleNet WWW site. This task will consist of six subtasks:

1. Develop and implement a process to identify and register RuleNet participants.
2. Develop and implement a reference library with word and subject search capability—searches will be integrated over
  - WAIS indexed information;
  - Gopher documents; and
  - Inquiry indexed information.
3. Develop and implement RuleNet WWW site, anonymous file transfer protocol, gopher, and listproc capabilities.
4. Receive, categorize, and clarify RuleNet comments.
5. Provide comment analysis and resolution of conflict guidelines.
6. Conduct electronic forum.

### **Phase Two: Extensions for RegNet**

The second phase of the work effort is tailored to three areas:

- Automated management and consensus development;
- Metrics; and
- Implementation and postimplementation debriefs.

In the area of automated management and consensus development, this project is designed to allow flexibility in the way that the single entity (the U.S. NRC) can interact both with the several entities (the regulated industry and their industry consortia representatives) and with the many (the general public). Each of these three groups have differing roles and responsibilities in this type of collaborative rule-making process and each needs to have its interest represented. The automation aspect of this project is geared toward developing mechanisms whereby some degree of moderator interaction can assist in managing the potentially thousands of interactions from the general public while still allowing the primary discussions between the regulated and regulator to take place. In addition, the ability to achieve consensus from the general public by intelligently parsing their comments will be of tremendous use for the regulator and for industry when it comes to finalizing proposed rules.

In the arena of metrics, this project will be of benefit by establishing some demographics on a federal level, rule-making activity. The preliminary numbers of interest would be items such as the number of regulators, industry representatives, and general public participants involved. The impact of rules on each group as perceived by the group correlated to the impact on each group as perceived by the other groups would also be of interest.

Finally, this project would have a post-implementation debriefing phase where an evaluation of all pertinent metrics would be made and folded into the RegNet national plan as well as into the other four projects of this overall effort.

To implement this collaborative workspace, information infrastructure tools are necessary. Since a collaborative workspace allows users to visualize information in new and dynamic ways, users may want to make notes about what they find, or to query an object to obtain specific information about the object in the space. An annotation system would allow users to integrate their notes directly into the collaborative workspace and to obtain detailed textual, graphic, or other information while still within the collaborative workspace. Additionally, there will be times when the threads of negotiation become extremely complex and users may need to have topic-generating "Bayesian systems" to help cross postthreads for evaluation. This capability will be especially important in a regulatory environment where those representing several (sometimes competing) interests must work together, negotiate, or resolve differences about matters of complex and specific detail. With the annotation system, they will be able to compile or exchange notes connected with specific features and actions. The annotations could be passed among remote users or even used in an asynchronous mode, where notes left at one time are collected, organized, and reviewed later.

#### **Demonstration Project Two: Virtual Regulatory Negotiations: Pilot Testing 3-D Technologies**

- Regulatory complexity: Low
- Technological complexity: Emerging to Advanced
- Government unit: Local, State, Federal
- Principal investigators: J.D. Nyhart and Kenneth Kaplan, Massachusetts Institute of Technology

This project seeks to develop, and try out in simulation and then in real life, an electronic, physical, and process environment in which negotiations common in regulatory processes can be carried out on a virtual or dispersed basis. There will be an evaluation of the activities and products associated with each goal identified below. In the first year, existing 3-D technologies will be used in the simulations and compared. In the second, new advanced technology will be developed and deployed based on the lessons and evaluations in the first year.

It is now possible to put into practice combined audio/video/computer-aided communication to enable multiple parties located apart from each other to negotiate agreements in real time, anywhere in the world, anytime they choose, and to have computer-based tools to help in the process. As the Clinton-Gore technology

policy statement of February 22, 1993, stated, "Fast communication makes it possible for teams to work closely on a project even if the team members are physically distant from each other." The observation applies particularly well to negotiation.

Virtual environments are emerging as a revolutionary means of communication. For example, one powerful application currently being explored is the dissemination and display of U.S. government documents, including those concerning federal regulatory processes. If, for instance, the current vast regulatory material could be transformed from text format into a visually dynamic environment, then movement through the material would be translated into an activity similar to walking through a building.

Virtual documentary environments (VDEs), when combined with distance audio, video, and computer capabilities, become cogent negotiating tools or negotiating environments, where representative parties from government, industry, and the public can discuss decisions while simultaneously referencing the easily manageable virtual documentary environment. Through new applications of information technology, RegNet can help push back the frontiers of dispersed (or distance) communication, specifically, the communications of reaching agreement, or negotiation. This project would build multimedia technology—including modeling and other software—as integrative tools to be jointly used by the negotiating parties who are located in different geographic places. The results would be extremely helpful not only to industry, government, and academia but also to the independent citizen trying to understand and negotiate the intricate maze of regulatory legislation.

This project sits at the intersection of the technological advances of the NII and the need to improve, visibly and effectively, the processes of regulation. Underlying issues include the following:

- Enhancing citizens' ability to participate effectively in regulatory decisions affecting them;
- Ameliorating the bilateral mode of most regulatory hearing processes;
- Pending devolution of regulatory power from the federal government to the states, accentuating the desire for effective, long-distance communications and negotiations; and
- Capitalizing on opportunities provided by technology to resolve the above issues.

#### **Project Goals—First Year**

- *Goal 1.* Construct a visualization, demonstrating the virtual or dispersed uses by multiple parties of existing information technology and space for reaching agreements. Two selected scenarios will be developed and evaluated that are typical of those found in the regulatory arena: program initiation and rule making.
- *Goal 2.* Test and evaluate the technology and space in the pilot stage by a simulation or mock run-through of the two scenarios.
- *Goal 3.* Compare and evaluate the information technologies used in each of the two simulated negotiations and the experience, through protocols or interviews of the participants and analysis of any embedded monitoring software. The nature of the data collected will be informed by the evaluative work conducted by NPR and OMB to date; we will seek to be consistent with them. To the appropriate extent, we hope to include existing metrics now in use by NPR and OMB and, based on the experience of the project, to add to them.

#### **Project Goals—Second Year:**

- *Goal 4.* Apply lessons from the simulations and evaluation of existing technology in two parallel directions. The first is the development of new 3-D technology to create an advanced environment for regulatory negotiations of the type simulated in the first year. The second is the testing of that technology, again first in simulation and then in real life.

Governmental and private partners participating during the second year would be secured at the outset, or as close as possible to it, contingent on interim periodic reviews. There would therefore be buy-in at the beginning, reaffirmed throughout the project. We will evaluate the experience, technology, and space in each real-life virtual negotiation.

- *Goal 5.* Analyze the evaluation metrics and protocols used throughout the project for their applicability in NPR's or OMB's continuing monitoring of the NII. The second year will include efforts to provide measures of the effects of advances in the new information infrastructure on different processes of reaching agreement at various points in the regulatory structure. The question is whether new information technology is effective in advancing the processes of reaching agreement. Does the technology work? Is it sought out? Does it seem to make a difference in the eyes of the people who are brought under its mantle?

### **Demonstration Project Three: Virtual Regulatory Environment**

- Regulatory complexity: Low to Significant
- Technological complexity: Emerging to Advanced
- Government units: Local, State, Federal
- Principal investigators: James Foley, William Ribarsky, Larry Hodges, and Nicholas Faust, Graphics, Visualization, and Usability Center, GIS Center, Georgia Institute of Technology

### **Statement of Problem**

To get involved in the regulatory environment, people typically turn to experts for assistance. Sometimes people even use multiple sets of experts because of difficulties in navigating through the vast arrays of printed (or electronic) regulatory information required for understanding even simple issues. When challenged with more complex forms of regulatory interactions (e.g., adjudication or collaborative rule making), current methods all but mandate the use of intermediate "experts" between the "customer" of regulation and the "supplier" of the regulatory services.

The complexities of dealing with paper-based regulation, various experts, four different types of regulatory interactions, and the needs of everyday life typically cause people to just give up in their quest either to access regulatory information or to participate in regulatory processes. These "quests" are fundamental rights of Americans. The intensely complex regulatory arena is responsible for forcing many people out of the regulatory processes that are their right. This project is designed to bring these people back into the process and to act as their own "experts" in accessing information that shapes their lives.

As stated earlier, it is the underlying premise of the RegNet project, and of the NPR, that some problems that stem from complexities, societal or technological, can be dealt with through thoughtful implementation of technologies that are able to render complex systems into simpler constructs.

This project, involving collaborations between several state institutions, federal government agencies, and issues in both rural and urban areas, by its very nature meets the criterion to address nationwide infrastructures meeting the needs of diverse groups of citizens. This project will use significant components of the NII, add to it, and make it more useful. It will be developed by organizations that have had a major hand in using and sustaining that infrastructure.

### **Support for the End User**

Although it will use advanced technology, the proposed system will have an intuitive interface that will make it accessible to a broad range of users at any level of technical expertise. The user will not need specialized knowledge of plans, blueprints, or layouts for sites or structures because these will be rendered as realistic three-dimensional objects among which the user can navigate naturally. Selection will be by directly grabbing (or shooting a ray from a 3-D mouse through) an object. These acts can initiate a query to a geographic information system (GIS) database, so the amount of effort and skill necessary to set up these queries will be reduced. Initial queries to the GIS and cross-referenced regulatory databases will also be automatically structured, based on the



properties of the site and the work to be performed there. As a result, it will be easier for even inexperienced users to get information quickly and, based on what they have found, carry out further queries.

### **Proposed Work**

This project will develop the VReg, a real-time 3-D navigation system for exploring cross-referenced databases. The cross-referenced databases will be formulated and implemented as part of this project and will consist, on the one hand, of regulatory information (local, state, national, international) and, on the other hand, of geographical information (with accompanying layers of geological, economic, local climate, and other data). The databases will be organized as servers with their own sets of calls for placing or retrieving information. Since queries could be made over networks, the database servers could be distributed. Thus communities could maintain their own servers of local regulations or geographical and economic data—they would just use the appropriate calls for organization and data placement. The geographical servers will be extensions of those that already exist and will rely on the massive capability in government (e.g., Army databases) and business (e.g., geographic information companies such as ERDAS). The regulatory servers will be developed and linked with the geographical servers as part of this project. A central index, using the RegNet WWW Server, will be set up with addresses and information about the satellite sites.

### **Serving as a Model**

The VGIS system upon which VReg is based is highly flexible and portable. All it takes is different terrain data, 3-D building plans, and appropriate sets of GIS layers to explore a multitude of distributed regulation scenarios. As part of the project, we will provide documentation on how the terrain data server and GIS layers are set up. We will also provide documentation for the cross-referenced regulation server. In addition, some of the regulatory information (e.g., federal regulations) will be reusable in many situations. Since all 3-D graphics in VReg ultimately use Open GL, which is the de facto standard in 3-D graphics, our software can run on a variety of different platforms. Communications using standard TCP/IP protocols will ensure that our system can be used in a distributed fashion by anyone who has access to the Internet. In addition, standard UNIX socket mechanisms will be used for distributed communication on local area networks.

The VGIS system we have developed has truly immersive capability for navigating and understanding complex and dynamic terrain-based databases. The system provides the means for visualizing terrain models consisting of elevation and imagery data, along with GIS raster layers, protruding features, buildings, vehicles, and other objects. We have implemented window-based and virtual reality versions and in both cases provide a direct manipulation, visual interface for accessing the GIS data. Unique terrain data structures and algorithms allow rendering of large, high-resolution data sets at interactive rates.

VGIS represents a significant advance over previous 2-D, map-based systems or 3-D GISs with decidedly noninteractive rates (having typical delays to produce images of several seconds or more). It can be used anywhere a traditional GIS can be used. Thus it can be used for urban planning, evaluation of geology, vegetation, soil, waterway, or road patterns, flood planning, and many other tasks. In addition, the ability to have detailed 3-D views and to jump to a different location to see the view from there opens new possibilities. Planners for new buildings or other facilities can see full 3-D views from their prospective sites or can see the view from nearby existing buildings with their planned facility in place.

VGIS is implemented using the simple virtual environment (SVE) tool kit, a device-independent library that provides mechanisms and software tools for developing virtual environment (VE) applications. SVE is based on the graphics library (GL) of Silicon Graphics, and VGIS has been run on a number of different hardware systems with GL support.

VGIS can be used either with a workstation window-based interface, or with an immersive virtual reality interface. The conventional window-based interface in VGIS parallels the functionality of the immersive interface. The 2-D workstation mouse can be used to rotate the user's viewpoint, move through the environment,

select objects, and pull down menus. The user's view is displayed in a workstation window. In the immersive environment, users wear a head-mounted display (HMD) and hold a three-dimensional mouse controller, with six degree-of-freedom trackers attached to both the HMD and the 3-D mouse. The head tracking allows the HMD graphics to be generated in accordance with the user's viewpoint, so that the user feels a sense of presence within the virtual environment.

### **System Features**

Users of VGIS have complete freedom in navigating the virtual world described by the databases. Both the window-based and immersive interfaces allow flight through six degrees of freedom (pitch, yaw, roll, and three-dimensional translation). To avoid getting lost or to find specific locations, the user can toggle a labeled grid coordinate overlay or bring up an overview 2-D finder map. The user can then select and jump to locations on either the finder map or in the 3-D world. This provides fast capability to view the 3-D world from any location within it. In query mode, selection of any feature, terrain location, building, or other object in the 3-D world brings up GIS information about the location or object. Since it would be straightforward, within the GIS, to provide links to other data associated with an object, this selection could, for example, bring up a full, 3-D model of a building interior. As an alternative, the user could merely fly through the building wall to enter the interior model. Individual units, such as vehicles, or groups of units can be placed on or allowed to move over the terrain, and symbolic information, such as markers or lines of demarcation, can be laid down or moved around. These are just some of the features available in the VGIS system.

### **Use in this Project**

Our VReg implementation will provide a visual interface for fast and easy retrieval of geography- or location-specific information. This capability itself can be quite useful, especially when one has simultaneous access to a database of regulatory information as we propose here. In addition, we propose to build for various types of facilities and situations a knowledge base that can be used by the system to automatically retrieve appropriate information from the GIS and cross-reference it with relevant regulations. Thus, for example, if one were siting a fuel storage facility, the system would know to look for information on substrate geology, water table and water runoff, whether the site is earthquake prone, whether other sites with flammable or toxic materials are nearby, and so on. The system could then also pull up relevant keyworded local, state, and federal regulations.

This approach offers the possibility of a dynamic regulation model that streamlines the process and reduces the number of regulations. In the above example, it would be fairly straightforward to pinpoint redundant local, state, or federal regulations if they were accessible via the same server structure. One could then have a process of removing or ignoring redundant regulations. Also, the GIS offers the capability to store and retrieve (in an organized fashion) highly detailed inputs for engineering, economic, environmental, or other models; we will extend the knowledge base to enable automatic searches for this input information. Once in place the models could thus be run easily for user-selected scenarios. If validated by government agencies, model outputs could be used to satisfy sets of simple goals, and one could do away with large numbers of regulations. Our approach would also allow the user to play "what if," exploring different siting options, plant output structures, and the like to meet goals, maximize economic gains, or optimize other effects. Since the data are available and organized, it would also be possible for the user to obtain "real-time" environmental and economic impact studies.

### **A Testbed Project**

We will develop distributed GIS and regulatory servers for Champaign County, Ill., (NCSA location) and for the Atlanta, Ga., area (Georgia Tech location). The Champaign County database will build on the regulatory and other information already available via CCNet (the Champaign County network), whereas the Atlanta

database will organize around extensive GIS information available for the area. Thus we will use significantly different locales—one rural and one urban—in our testbed project. We will apply our tools to a complex scenario, such as the decommissioning of a nuclear power plant and the shipping of waste materials to a storage facility. For the purpose of this exercise only, we could place the decommissioned power plant in the urban area and the storage facility in the rural area. We would then have to deal with a welter of different local regulations overlaid with a set of federal regulations for the two sites. We would use the distributed servers (allowing common pictures of the sites) and various collaborative tools (such as the annotation system described below) to do this.

### **A 3-D Annotation System for Note-taking and Note-sharing**

A 3-D environment allows the user to visualize information in new and effective ways. Although visualizing and exploring a complex model or an extended site, the user may want to make notes about what he or she finds, or to query an object to obtain specific information about the object in the space. An annotation system should allow the user to integrate his notes directly into the environment and to obtain detailed textual, verbal, or other information while still within the 3-D space. This capability will be especially important in a regulatory environment where those representing several (sometimes competing) interests must work together, negotiate, or resolve differences about matters of complex and specific detail. With the annotation system they will be able to compile or exchange notes connected with specific features and actions (or time frames for dynamic simulations) in 3-D models and terrains. The annotations could be passed among remote users or even used in an asynchronous mode where notes left at one time are collected, organized, and reviewed at a later time.

#### **Demonstration Project Four: Intelligent Distributed Access to Regulatory Information**

- Regulatory complexity: Low to Significant
- Technological complexity: Emerging to Advanced
- Government units: Local, State, Federal

The recent increase in complexity of our regulatory environment has made it more difficult for

- Affected citizens and businesses to find and understand applicable regulations;
- Monitors to assess compliance with regulations; and
- Regulatory bodies and involved participants to evaluate proposed changes.

A network space for regulatory information and operation, RegNet, could be created as a means to decrease the regulatory complexity. The creation of RegNet would require the cooperation of authorities, businesses, consumer groups, and experts in the domains being addressed and would require academic involvement to study the technological and organizational problems and opportunities. RegNet development would require industrial participation to provide the operational services using appropriate technology and would require the cooperation of local, state, and federal regulatory agencies to put the technology to use.

An application to provide online access to the text of regulations via the WWW could be developed. At this point, citizens, monitors, and regulators would be able to browse and access this text via computer programs like Mosaic and WAIS, moving smoothly from the regulations of one jurisdiction to those of another.

The application would provide linkages among these documents, based on the terms used within the regulations. A modern standard for storage of regulations—perhaps a publicly accessible base such as the Defense Advanced Research Projects Agency (DARPA)-funded Persistent Object Base—would be selected for use. Available search systems would be enhanced with thesauruses to span the application among terms and to

aid nonexperts in completing their searches. Important input to such thesauruses would be obtained by analyzing search strategies and failures from the initial phases.

The ordinary text of regulations with formal logic representations would be augmented. Such an information structure, building on top of the linkages and thesauruses, would enable new, crucial services, such as topic-based retrieval of relevant regulations, automatic conformance checking, and automated determination of conflicts and overlaps among regulations. Such analysis tools would empower workers in their specific domains and allow cooperation without centralization.

#### **Demonstration Project Five: Rural Government Testbed Project**

- Regulatory complexity: Low to Moderate
- Technological complexity: Emerging to Advanced
- Government units: Local, State, Federal

#### **Problem Definition**

Large metropolitan areas often have branches of federal and state government agencies, in addition to county and local government offices. Moreover, these offices are typically located in the same section of the city. Many of the offices that a citizen in a large metropolitan area needs to communicate with for life and business may at least be located near each other. However, in rural areas, there are rarely local branch offices of all the federal and state agencies with which rural citizens need to interact. The rural testbed demonstration project will develop mechanisms for citizens to interact with local, county, state, and federal government agencies seamlessly, utilizing emerging technologies. The regulatory complexity of this project will typically be either one to one, or one to two.

#### **Location for the Project—Champaign County, Illinois**

The rural testbed project will be conducted in Champaign County, Ill. The closest large cities to Champaign County are Chicago, St. Louis, and Indianapolis, each approximately 2-1/2 hours' driving distance, and too far to conduct daily business, even if a citizen were to brave visiting all agencies in person. More than 90 percent of Champaign County is farmland, with a university community at its center. Illinois has a large number of local governments (including city, county, township, and state), so the rural testbed is set in a particularly challenging region.

A virtual regulatory environment will be useful only if the citizenry is knowledgeable enough about the technology to use it. Champaign County was selected for this project because it has been nationally recognized for its innovativeness and advances in the NII. Champaign County Network (CCNet) is a collaborative project between the National Center for Supercomputing Applications, the University of Illinois, and the Champaign County Chamber of Commerce. The mission of CCNet is to create a supportive environment for Champaign County to cooperatively develop applications that use advanced communications technologies. This has been accomplished through the work of a technical committee and six applications task forces exploring applications of the NII in education, health care, agribusiness, government, and libraries.

Since early 1993, over 300 community leaders have been working together to build both the human infrastructure and the technical infrastructure to achieve the goals of CCNet. Today, there are more than 15 sites in the community connected to the Internet at high speeds. Nine sites in the community—including schools, businesses, and a public library—receive data over the cable-TV at 2.5 Mbps and send data over the telephone line at 14.4 kbps. Eight schools are connected to the Internet via ISDN connections sending and receiving data at 56 kbps or 128 kbps. In addition, Ameritech and CCNet recently announced a partnership through which Ameritech is offering an ISDN package connecting individual users to the Internet at 128 kbps. Wireless

connections are being tested. Champaign County also has a free net called PrairieNet, which has more than 8,000 registered users and is rapidly growing. The county recently announced that, thanks to information available online, courts outside the county seat will be able to serve local residents.

The remarkable thing about all the CCNet connections is that they have been community initiated. CCNet serves as a catalyst, bringing sophisticated technologies and Internet connections to Champaign County much sooner than they will be available in comparable communities.

Although the community leadership has been building an electronically connected community in Champaign County, the entire citizenry has become educated about the Internet and tools electronic mail and data exchange. Numerous newspaper and television news stories have covered the local efforts and successes, and there have been countless public presentations and demonstrations to special interest groups.

Participants in the CCNet project have been using electronic mail for nearly 2 years and have been exploring the WWW using NCSA Mosaic. Through the rural government testbed project, the citizens of Champaign County will be able to use the technology for specific purposes and realize the benefits of technology for interacting with multiple government agencies. The level of awareness and of current electronic connectivity in Champaign County optimizes our chances of success in the rural government testbed.

### **Project Concept and Dimensions**

The rural government testbed project will develop a system whereby citizens can use one technology to interact with a variety of governmental agencies at various levels. The goal is to eliminate frustration with government bureaucracy and to return power and independence to individuals. Though the tools developed may be applied to a nearly endless number of situations, we will focus on two primary applications. Both applications can be one to one or one to two, involve a range of governmental agencies, and are a cross between the emerging and complex categories of technology complexity. The second application may lend itself to a one-to-many regulatory complexity.

The basic concept is to allow a citizen to access relevant regulations via an online system (the Internet) and to electronically file forms required for specific real-life tasks. This can be accomplished by providing government information on the WWW, with citizens using a WWW browser or navigator such as NCSA Mosaic. Theoretically, all forms submitted by citizens would be processed automatically and government databases would be maintained automatically. But, for the testbed, individual persons at the government agency will intervene by updating their database with forms and information received from the citizens. Collaborative tools will be built into the browser so that citizens can have advisors helping them to make decisions in regulatory environments, if they wish. In addition, text searching capabilities developed as part of the Digital Library Initiative at the University of Illinois will be incorporated into the tools for ease of use.

### **Rural Government Testbed: Application 1**

Changing residence is one of the most stressful life situations. The stress of moving is exacerbated by the multitudes of contacts that the individual has to maintain separately with different government agencies. For the most part, the individual is responsible for updating the various agencies about his or her whereabouts.

With the rural government testbed we will build an environment where interacting with levels of government before and after a change of residence is drastically simplified. Currently, when a citizen moves, he or she has to communicate with the following agencies at different levels:

- Local/city: register children in school, notify water and power services;
- County: register to vote with county clerk, pay real estate tax with county assessor;
- State: renew driver's license at the Secretary of State, replace license plates, update vehicle registration; and
- Federal: leave forwarding address with Postmaster, send address change to Internal Revenue Service.

In addition to these government agencies, the citizen must also update his or her address with various private organizations such as magazine publications, newspapers, insurance carriers, telephone services, and even online services. If the rural government testbed works to smooth residential transitions, these private organizations could connect to the same system in the future.

### **Rural Government Testbed: Application 2**

A person who wants to establish a small business today has to learn about and comply with an almost unbearable number of zoning, environmental, employment, and other regulations. The time it takes to learn all the regulations to ensure compliance and to fill out all accompanying forms can be daunting, if not completely stifling. Large corporations have in-house attorneys and other experts to handle interactions with government agencies and to keep educated about the most recent relevant regulations. Some potential small business owners have almost no choice but to franchise with one of the large corporations. This difficulty dramatically reduces opportunities for innovation in American business. For example, a person who wants to open a particular type of business—for instance, a dry cleaner—has to interact with a large set of government agencies in order to set up the business:

- Local/city: zoning and sign permit, sales tax regulations;
- County: "Doing Business As" (DBA) name registration;
- State: State Department of Revenue for a retail license; and
- Federal: OSHA for compliance with regulations, the Department of Immigration and Naturalization for each employee, and EPA for an environmental impact statement.

It turns out that many of the state and federal regulations, at least, already exist in digital format. They will need to be reformatted to fit into the testbed project. Even if all the regulations and forms necessary to establish a new small business were available online through the rural government testbed, there would still be two problems: finding relevant sections of regulations, and dealing with the vast amount of information and legal ramifications.

To solve these problems, we will include two additional features in the rural government testbed project:

1. A text-search mechanism developed as part of the Digital Library Initiative (DLI) at the University of Illinois so that users and potential small business owners could search lengthy regulations for the relevant portions, based on proximity and keywords. Since many of the regulations are already in digital format, the search indexes would be generated and available to expedite the search process.
2. A potential small business owner may want to fill out forms and read through regulations with the assistance of professional counsel such as an attorney or an accountant. We will therefore experiment with collaborative features in WWW browsers, including collaborative visualization as well as video conferencing. HyperNews features can be built into the testbed to encourage small business owners to learn from each other.

### **Rural Government Testbed Technology**

Each of the government agencies identified in the applications above will be connected to the Internet at high enough speeds to serve information on the WWW. Some of them are already connected (the local schools). In addition, the public library in Champaign County is connected to the Internet, providing public access, and two other libraries will be connected to provide further public access. One hundred community members will be connected via the Ameritech ISDN solution. These 100 will be selected based on factors that indicate their potential for using the testbed, such as their skill with Internet tools, whether they own a small business, and whether they are located in an area where the service is available.

The first 6 months of the project will be devoted to recruiting individual and agency participants, identifying connection solutions, adapting the browser software to include searching and collaborative tools, and putting relevant information from government agencies online. Extensive training will be conducted for the participants and for the general public during months 5 through 7. The connections will be in place for 1 year for the testbed, and usage will be monitored. The efficiency will be evaluated by examining usage statistics and eliciting continual feedback from participants online, via public forums, and through individual communications. The software will be revised as further needs are identified.

Note that, although the rural government testbed will be developed for two applications, the system could be migrated to another application if the proper agency were involved. For instance, an agribusiness application might be developed by working with county farm bureaus, and with the USDA.

### CONCLUSIONS

The proposed projects, the planning, and the demonstrations will work together to create a synergistic plan with proven capability for the nation. Although the projects' focus will be to empower the rural and tribal communities to function within the intergovernmental system that surrounds them, the results will automatically extend to the other city, state, and federal governments that also must function in the new environment. Information technology is one of the major ways that the current structure of regulatory control can be simplified and made useful to citizens trying to function in the NII. This project will also work with, and solicit resources from, international organizations to include the results of a reinvented regulatory government in a growing GII.

The deployment and use of high-performance computing, communications, and information technology and the NII are essential to meeting national goals and objectives in the global marketplace.

### STATUTES AND EXECUTIVE ORDERS

With the plan proposed here, modified by the lessons learned from the implementation of the demonstration projects also proposed here, future intergovernmental regulatory projects can be developed in such a way that they more effectively support the national goals and objectives and national challenge applications. These goals and applications are found in a series of statutes and executive orders:

- The High Performance Computing Act of 1991, Public Law 102-194, December 9, 1991, and subsequent legislation in support of advances in computer science and technology that are vital to the nation's prosperity, national and economic security, industrial production, engineering, and scientific advancement;
- High Performance Computing and Communications: Technology for the National Information Infrastructure, a supplement to the President's Fiscal Year 1995 Budget; Science in the National Interest, President William J. Clinton, Vice President Albert Gore, Jr., August 1994, Executive Office of the President Office of Science and Technology Policy;
- Future Scenarios, National Challenge Applications, including: Public Access to Government Information, Health Care, Education and Training, Digital Libraries, Electronic Commerce, Advanced Manufacturing, Environmental Monitoring and Crisis Management, High Performance Computing and Communications (HPCC), Information Infrastructure Technology and Applications, National Coordination Office for HPCC, Executive Office of the President, Office of Science and Technology Policy, February 1994;
- The Clinton Administration's development of an advanced national information infrastructure (NII) and the global information infrastructure (GII) as top U.S. priorities; see Global Information Infrastructure: Agenda for Cooperation, Al Gore, Vice President of the United States, Ronald H. Brown, Secretary of Commerce and Chairman, Information Infrastructure Task Force, February 1995;
- Executive Order 12866, Regulatory Planning and Review, September 30, 1993;
- Executive Order 12839, Reduction of 100,000 Federal Positions, February 10, 1993;

- Executive Order 12861, Elimination of One-half of Executive Branch Internal Regulations, September 11, 1993;
- Executive Order 12864, United States Advisory Council on the National Information Infrastructure, September 15, 1993; and
- Executive Order 12875, Enhancing the Intergovernmental Partnership, October 26, 1993.

#### NOTE

1. "Technology for America's Growth: A New Direction to Build Economic Strength," p. 20.

#### REFERENCES

1. The High Performance Computing Act of 1991, Public Law 102-194, December 9, 1991.
2. High Performance Computing and Communications: Technology for the National Information Infrastructure, supplement to the President's Fiscal Year 1995 Budget; Science in the National Interest, President William J. Clinton, Vice President Albert Gore, Jr., August 1994, Executive Office of the President Office of Science and Technology Policy.
3. Information Infrastructure Technology and Applications, National Coordination Office for HPCC, Executive Office of the President, Office of Science and Technology Policy, February 1994.
4. Global Information Infrastructure: Agenda for Cooperation, Al Gore, Vice President of the United States, Ronald H. Brown, Secretary of Commerce and Chairman, Information Infrastructure Task Force, February 1995.
5. Virtual GIS: A Real-Time 3-D Geographic Information System, David Koller, Peter Lindstrom, William Ribarsky, Larry Hodges, Nick Faust, and Gregory Turner, Graphics, Visualization & Usability Center, Georgia Institute of Technology, Tech Report GIT-GVU-95-14, 1995, submitted to IEEE Visualization 1995.
6. The Simple Virtual Environment Library, Version 1.4, User's Guide, G.D. Kessler, R. Kooper, J.C. Verlinden, and L. Hodges, Graphics, Visualization & Usability Center, Georgia Institute of Technology, Tech Report GIT-GVU-94-34, 1994.
7. Graphics Library Programming Guide, Silicon Graphics Computer Systems, Mountain View, California, 1991.
8. Level-of-detail Management for Real-time Rendering of Phototextured Terrain, P. Lindstrom, D. Koller, L.F. Hodges, W. Ribarsky, N. Faust, and G. Turner, Graphics, Visualization & Usability Center, Georgia Institute of Technology, Tech Report GIT-GVU-95-06 (1995), submitted to Presence.
9. The Virtual Annotation System, Reid Harmon, Walter Patterson, William Ribarsky, and Jay Bolter, Georgia Institute of Technology, Tech Report GIT-GVU-95-20.
10. Creating a Government That Works Better and Costs Less: Status Report September 1994, Report of the National Performance Review (NPR) CD-ROM, Vice President Al Gore, 1994.
11. Uniform Resource Locator <http://www.npr.gov/>, NPR home page.
12. Uniform Resource Locator <http://nuke.WestLab.com/RegNet.Industry/>, RegNet Industry home page.
13. Creating a New Civilization: The Politics of the Third Wave, Alvin and Heidi Toffler, Turner Press, 1994-1995.
14. The Death of Common Sense: How Law is Suffocating America, Philip K. Howard, Random House, 1994.
15. Executive Order 12291, Federal Regulation, February 17, 1981.
16. Executive Order 12498, Regulatory Planning Process, January 4, 1985.
17. Executive Order 12612, Federalism, October 26, 1987.
18. Executive Order 12637, Productivity Improvement Program for the Federal Government, April 27, 1988.
19. Executive Order 12839, Reduction of 100,000 Federal Positions, February 10, 1993.
20. Executive Order 12861, Elimination of One-half of Executive Branch Internal Regulations, September 11, 1993.
21. Executive Order 12864, United States Advisory Council on the National Information Infrastructure, September 15, 1993.
22. Executive Order 12866, Regulatory Planning and Review, September 30, 1993.
23. Executive Order 12875, Enhancing the Intergovernmental Partnership, October 26, 1993.



## APPENDIX

John P. Ziebarth, Ph.D., Associate Director  
National Center for Supercomputing Applications  
CAB261

605 E. Springfield  
Champaign, IL 61820  
217/244-1961 voice  
217/244-1987 fax

[ziebarth@ncsa.uiuc.edu](mailto:ziebarth@ncsa.uiuc.edu)

W. Neil Thompson, NPR RegNet. Gov, Coordinator  
Advisory Committee on Reactor Safeguards Rotational Assignment  
U.S. Nuclear Regulatory Commission

Washington, DC 20555  
301/415-5858 voice

[wnt@nrc.gov](mailto:wnt@nrc.gov)

[nthomps@tmn.com](mailto:nthomps@tmn.com)

J.D. Nyhart, Professor of Management and Ocean Engineering  
Massachusetts Institute of Technology

Sloan School of Management  
50 Memorial Drive, E52542  
Cambridge, MA 02142-1437

617/253-1582 voice

617/253-2660 fax

[jdnyhart@mit.edu](mailto:jdnyhart@mit.edu)

Kenneth Kaplan, Principal Research Scientist  
Department of Architecture and Research Laboratory for Electronics  
Massachusetts Institute of Technology

Cambridge, MA 02139

617/258-9122 voice

617/258-7231 fax

[kkap@mit.edu](mailto:kkap@mit.edu)

Bill Ribarsky, Associate Director for Service  
Georgia Institute of Technology  
OIT/Educational Technologies  
Graphic Visualization and Usability Center

Room 229 Himan

Atlanta, GA 30332-0710

404/894-6148 voice

404/894-9548 fax

[bill.ribarsky@oit.gatech.edu](mailto:bill.ribarsky@oit.gatech.edu)

Gio Wiederhold, Ph.D., Professor  
Computer Science Department (also, by courtesy, Electrical Engineering and Medicine)  
Gates Building 4A, Room 433  
Stanford University  
Stanford, CA 94305-9040  
415/725-8363 voice  
415/725-2588 fax  
[siroker@cs.stanford.edu](mailto:siroker@cs.stanford.edu)

Michael R. Genesereth, Professor  
Computer Science Department  
Gates Building  
Stanford University  
Stanford, CA 94305-9040  
415/725-8363 voice  
415/725-2588/7411 fax  
[genesereth@cs.stanford.edu](mailto:genesereth@cs.stanford.edu)

Kenneth Gilpatric, Esq.  
National Performance Review NetResults.RegNet, Consultant  
Administrative Conference of the United States (formerly)  
1615 Manchester Lane, NW  
Washington, DC 20011  
202/882-7204 voice  
202/882-9487 fax  
[ken.gilpatric@npr.gsa.gov](mailto:ken.gilpatric@npr.gsa.gov)

Tim E. Roxey  
National Performance Review RegNet. Industry, Lead  
Baltimore Gas and Electric, Project Manager  
Regulatory Projects  
1650 Calvert Cliffs Parkway  
Lusby, MD 20656  
Council for Excellence in Government (CEG), Principal Investigator  
Suite 859  
1620 L Street, NW  
Washington, DC 20036  
410/495-2065 voice  
410/586-4928 pager  
[timr@access.digex.net](mailto:timr@access.digex.net)

William J. Olmstead, Associate General Counsel for Licensing and Regulation  
Office of General Counsel  
Commission Staff Office  
U.S. Nuclear Regulatory Commission  
Washington, DC 20555  
301/415-1740 voice  
[wjo@nrc.gov](mailto:wjo@nrc.gov)  
[olmstead@tmn.com](mailto:olmstead@tmn.com)

Ben Slone, President  
Finite Matters Ltd.  
2694 Fairground Road  
Goochland, VA 23063  
804/556-6631 voice  
[sloneb@nuke.westlab.com](mailto:sloneb@nuke.westlab.com)  
Jim Acklin, CEO  
Regulatory Information Alliance  
8806 Sleepy Hollow Lane  
Potomac, MD 20854  
301/983-2029 voice  
[acklinj@aol.com](mailto:acklinj@aol.com)

## 65

# Electronic Document Interchange and Distribution Based on the Portable Document Format, an Open Interchange Format

*Stephen N. Zilles and Richard Cohn*  
*Adobe Systems Incorporated*

### ABSTRACT

The ability to interchange information in electronic form is key to the effectiveness of the national information infrastructure initiative (NII). Our long history with documents, their use, and their management makes the interchange of electronic documents a core component of any information infrastructure. From the very beginning of networking, both local and global, documents have been primary components of the information flow across the networks. However, the interchange of documents has been limited by the lack of common formats that can be created and received anywhere within the global network. Electronic document interchange had been stuck in a typewriter world while printed documents increased in visual sophistication. In addition, much of the work on electronic documents has been on their production and not on their consumption. Yet there are many readers/consumers for every author/producer. Only recently have we seen the emergence of formats that are portable, independent of computer platform, and capable of representing essentially all visually rich documents.

Electronic document production can be viewed as having two steps: the creation of the content of the document and the composition and formatting of the content into a final form presentation. Electronic interchange may take place after either step. In this paper we present the requirements for interchange of final form documents and describe an open document format, the portable document format (PDF), that meets those requirements. There are a number of reasons why it is important to be able to interchange formatted documents. There are legal requirements to be able to reference particular lines of a document. There is a legacy of printed documents that can be converted to electronic form. There are important design decisions that go into the presentation of the content that can be captured only in the final form. The portable document format is designed to faithfully represent any document, including documents with typographic text, tabular data, pictorial images, artwork, and figures. In addition, it extends the visual presentation with electronic aids such as annotation capabilities, hypertext links, electronic tables of contents, and full word search indexes. Finally, PDF is extensible and will interwork with formats for electronic interchange of the document content, such as the HyperText Markup Language (HTML) used in the World Wide Web.

### BACKGROUND

A solution to the problem of electronic document interchange must serve all the steps of document usage. The solution must facilitate the production of electronic documents, and, what is more important, it must facilitate the consumption of these documents. Here, consumption includes viewing, reading, printing, reusing, and annotating the documents. There are far more readers than authors for most documents. Serving consumers has a much bigger economic impact than does serving authors. Replacing paper distribution with electronic distribution increases timeliness, reduces use of natural resources, and produces greater efficiency and productivity. It also allows the power of the computer to be applied to aiding the consumption of the document; for example, hyperlinks to other documents and searches for words and phrases can greatly facilitate finding the documents and portions thereof that interest the consumer.

From a consumption point of view, the critical need in electronic document interchange is the ability to view, print, or read the document everywhere that someone has access to it. If the document can also be revised or edited then that is so much the better, but it is not required for use of the document.

### **Final Form and Revisable Form Interchange**

There are two different ways to produce interchangeable electronic documents. Although there are many steps in the production of a visually rich document, the process of composition and layout partitions production into two parts. Composition and layout is the process which takes a representation of the content of a document and places that content onto a two-dimensional space (or sequence of two-dimensional spaces), usually called pages.

In the process of composition and layout, a number of decisions, called formatting decisions, are made: which fonts in which sizes and weights are used for which pieces of content, where on the page the content is placed, whether there are added content fragments such as headers and footers, and so on. These formatting decisions may be made automatically based on rules provided to the composition and layout process; they may be made by a human designer interacting with the composition and layout process; or they may be made using a combination of these two approaches.

The representation of the content before composition and layout is called revisable form. Revision is relatively easy because the formatting decisions have not been made or have only been tentatively made and can be revised when changes occur. The representation of the content after composition and layout is called final form.

The interchange of electronic documents can be done either in revisable form or in final form. If the revisable form is interchanged, then either the formatting decisions must be made entirely by the consumer of the electronic document or the rules for making the formatting decisions must be interchanged with the revisable form electronic document. If the first approach is chosen, then there is no way to guarantee how the document will appear to the consumer. Even if the second approach is chosen, existing formatting languages do not guarantee identical final form output when given the same revisable form input. Some formatting decisions are always left to the consumer's composition and layout software. Therefore, different composition and layout processes may produce different final form output.

The interchange of revisable form electronic documents can meet many authors' needs. Both the Standard Generalized Markup Language (SGML) and the HyperText Markup Language (HTML) are successfully used to interchange significant and interesting documents. But there are cases where these formats are not sufficient to meet the needs of the author. For these cases, interchange of final form electronic documents is necessary.

### **Requirement for Page Fidelity**

The key problem with interchanging only revisable form documents is the inability to guarantee page fidelity. Page fidelity means that a given revisable form document will always produce the same final form output no matter where it is processed. There are a number of reasons why page fidelity is required.

The most obvious reason is that the composition and layout process involve a human designer's decisions. Only in the final form is it possible to capture these decisions. These formatting decisions are important in the presentation of the represented information. This is quite obvious in advertisements, where design plays an important role in the effectiveness of communicating the intended message. It is perhaps less obvious but equally important in the design of other information presentations. For example, the placement of text in association with graphical structures, such as in a map of the Washington, D.C., subway system (Figure 1), will greatly affect whether the presentation can be understood. In addition, formatting rules may not adequately describe complex information presentations, such as mathematics or complex tables, which may need to be hand designed for



Figure 1 Metro system map. SOURCE: Courtesy of Washington Metropolitan Area Transit Authority.

effective communication. (Figure 2 has simple examples of mathematics and tables.) Finally, the composition and layout design may reflect the artistic expression of the designer, making the document more pleasing to read (Figure 3).

The rich tradition of printed documents has established the practice of using page and line numbers to reference portions of documents, for legal and other reasons. These references will work for electronic documents only if page fidelity can be guaranteed. Many governmental bodies have built these references into their procedures and require that they be preserved, in electronic as well as in paper documents.

The final set of cases does not require page fidelity; they are just more simply handled with a final form representation than a revisable one. Documents that exist only in paper form, legacy documents, can be scanned and their content recognized to produce an electronic document. Although this recognized content could be represented in many forms, it is simplest when the content is represented in final form. Then it is not necessary to decide, for each piece of recognized content, what part of the original document content, such as body text, header, or footer, it belonged to. Since the final form is preserved, the reader of the document can correctly perform the recognition function. (See Figure 2 for an example of a legacy page that would be difficult to categorize correctly.)

Finally, preparing a document as an SGML or HTML document typically involves a substantial amount of markup of sections of content to ensure that a rule-driven composition and layout process will produce the intended effect. For documents that are of transient existence, it may be far simpler to produce the composition and layout by hand and to interchange the final form representation than to spend time tuning the document for a rule-driven process.

**B. Conditions and Events**

The input to the machines is the set of environmental conditions (e.g., whether a button is being pressed). All conditions are Booleans, although first-order predicate conditions that can be represented by a finite number of boolean conditions (such as integer ranges and timing constraints) are also expressible.

A system's behavior is defined and controlled by changes to the conditions' values. A change to a condition's value is an event, and events are only detectable at the point in time at which they occur. For example, event

$@T(Cond1)$

specifies the point in time when the value of condition *Cond1* changes from false to true. Similarly, event  $@F(Cond1)$  specifies the time when condition *Cond1* becomes false. In these events, we call condition *Cond1* the event's triggered condition. The occurrence of an event might also depend on the values of other conditions. For example

$@T(Cond1) \text{ WHEN } [Cond2]$

describes the event of condition *Cond1* becoming true while condition *Cond2* is also true. In the above event, we call condition *Cond2* the event's WHEN condition. More complex events can be created from simpler events and conditions using boolean operators.

SCR semantics propose three definitions for an event occurrence and allow the requirements designer to decide which definition pertains to each event [1]. For a given event  $@T(A)$  WHEN [B] there is a defined time *t*, such that if  $t_{BT(A)}$  is the time that triggered condition A becomes true, then the event occurs if

- 1) B is true throughout closed interval  $[t_{BT(A)} - \epsilon, t_{BT(A)}]$ , or
- 2) B is true throughout interval  $[t_{BT(A)}, t_{BT(A)} + \epsilon]$ , or
- 3) B is true throughout interval  $[t_{BT(A)} - \epsilon, t_{BT(A)} + \epsilon]$

We use the first definition for all event occurrences. This definition allows us to distinguish between WHEN conditions (which must be satisfied both before and at the time of the event) and triggered conditions (which must be unsatisfied immediately before the event and satisfied at the time of the event).

**C. Modes and Transitions**

The system state is defined by the current values of the system conditions, and the system state space is the set of all combinations of the conditions' values. To reduce the size of the state space, sets of system states are collected into modes. The use of modes abstracts away details that do not contribute to the system's behavior; the values of all system conditions are not important at all times. In fact, the conditions' values are only important when they can affect mode transitions.

A mode transition occurs between modes in the same modeclass as a result of the occurrence of an event. A transition's transition event specifies the event that triggers the transition. Mode transitions occur at the same time as their transition event and, like the event, take zero time

**TABLE I**  
 MODE TRANSITIONS FOR TEMPERATURE CONTROL SYSTEM

Current Mode	Running	Below-Desired Temp	OK	Above-Desired Temp	New Mode
OFF	@T —	—	1	—	INACTIVE HEAT AC
INACTIVE	@F	@T	—	—	OFF HEAT AC
HEAT	@F	—	@T	—	OFF INACTIVE
AC	@F	—	—	@T	OFF INACTIVE

units to complete. Two transitions from the same mode are simultaneously enabled if their transition events occur at the same time. In such a case, the modeclass is nondeterministic, and the activation of either transition (but not both) satisfies the requirements.

Table I lists requirements specification for a temperature control system. The specification consists of a single modeclass comprised of four modes: OFF, meaning that the system is turned off; INACTIVE, meaning that the system is on, but neither the heater nor the air conditioner is on; and HEAT and AC, meaning that either the heater or the air conditioner, respectively, is on and controlling the temperature. Condition Running indicates whether or not the system has been turned on, and conditions Below-DesiredTemp, TempOK, and Above-DesiredTemp represent the current temperature.

The initial state of the system is mode OFF, in which the system is not Running. The system is undefined if Running is initially true. Each row in the table specifies the event causing the transition from the mode on the left to the mode on the right. Each column in the center of the table represents a system condition. A table entry containing an upper-case letter ("@T" or "@F") signifies that the condition is a triggered condition of the transition event, and must change value (to true or false, respectively) to activate the mode transition. A table entry containing a lower-case letter ("1" or "—") signifies that the condition is a WHEN condition of the transition event, and must have a particular value (true or false, respectively) both immediately before and at the time of the event occurrence. If a condition is neither a triggered condition nor a WHEN condition of a transition event, then the corresponding table entry is marked with a dash ("—").

If the temperature control system is in mode OFF and starts Running when the temperature is within desired limits, then it will enter mode INACTIVE. The AC comes on when the system is INACTIVE and the temperature rises above the desired temperature, or when the system starts Running while the temperature is above the desired temperature. The AC cycles off when the temperature falls so within 3°F of the desired temperature. Transitions into and out of mode HEAT resemble those of mode AC.

Figure 2

**Requirements for a Final Form Document Interchange Format**

For a final form representation to be a suitable electronic document interchange format for the NII, it should meet a number of requirements:

- It should have an open, published specification with sufficient detail to allow multiple implementations. There are many definitions of openness, but the key component of them all is that independent implementations of the specification are possible. This gives the users of the format some guarantee of there being reasonable cost products that support the format. With multiple implementations, the question of interoperability is important. This can be facilitated, although not guaranteed, by establishing conformance test suites for the open specification.
- It should provide page fidelity. This is a complex requirement. For text, this means representing the fonts, sizes, weights, spacing, alignment, leading, and so on that are used in the composition and layout of the original document. For graphics, this means representing the shapes, the lines and curves, whether they are filled or stroked, any shading or gradients, and all the scaling, rotations, and positioning of the graphic elements. For

About this PDF file: This new digital representation of the original work has been recomposed from XML files created from the original paper book, not from the original typesetting files. Page breaks are true to the original; line lengths, word breaks, heading styles, and other typesetting-specific formatting, however, cannot be retained, and some typographic errors may have been accidentally inserted. Please use the print version of this publication as the authoritative version for attribution.



Figure 3

images, this means doing automatic conversion from the resolution of the sample space of the image to the resolution of the device, representing both binary and continuous tone images, with or without compression of the image data. For all of the above, the shapes and colors must be preserved where defined in device- and resolution-independent terms.

- It should provide a representation for electronic navigation. The format should be able to express hyperlinks within and among documents. These links should be able to refer to documents in formats other than this final form format, such as HTML documents, videos, or animations. The format should be able to represent a table of contents or index using links into the appropriate places in the document. The format should also allow searches for words or phrases within the document and positioning at successive hits.
- It should be resource independent. The ability to view an electronic document should not depend on the resources available where it is viewed. There are two parts to this requirement. A standard set of resources, such as a set of standard fonts, can be defined and required at every site of use. These resources need not be transmitted with the document. For resources, such as fonts, that are not in the standard set, there must be provision for inclusion of the resource in the document.
- It should provide access to the content of the document for the visually disabled. This means, at a minimum, being able to provide ASCII text that is marked up with tags that are compliant with the International Standard ISO-12083 (ICADD DTD). The ICADD (International Committee for Accessible Document Design) markup must include the structural and navigational information that is present in the document.
- It should be possible to create electronic documents in this format using a wide range of document generation processes. Ideally, any application that can generate final form output should be usable to generate an electronic document in this format. There should also be a means for paper documents to be converted into this format.



- It should be platform independent; that is, the format should be independent of the choice of hardware and operating systems, and transportable from one platform to any other platform. It should also be independent of the authoring application; the authoring application should not be required to view the electronic document.
- It should effectively use storage. In particular, it should use relevant, data type-dependent compression techniques.
- It should scale well. Applications using the format should perform nearly as well on huge (20,000-page) documents as they do on small ones; the performance on a large document or a (large) collection of smaller documents should be similar. This requirement implies that any component of the electronic document be randomly accessible.
- It should integrate with other NII technologies. It should be able to represent hyperlinks to other documents using the Universal Resource Identifiers (URIs) defined by the World Wide Web (WWW) and it should be identifiable by a URI. It should be possible to encrypt an electronic document both for privacy and for authentication.
- It should be possible to revise and reuse the documents represented in the format. Editing is a broad notion that begins with the ability to add annotations, electronic "sticky notes," to documents. At the next level, preplanned replacement of objects, such as editing fields on a form, might be allowed. Above that, one might allow replacement, extraction and/or deletion of whole pages, and, finally, arbitrary revision of the content of pages.
- It should be extensible, meaning that new data types and information can be added to the representation without breaking previous consumer applications that accept the format. Examples of extensions would be adding a new data object for sound annotations or adding information that would allow editing of some object.

These requirements might be satisfied in a number of different ways. We describe below a particular format, the portable document format (PDF) which has been developed by Adobe Systems Inc., and the architectures that make interchange of electronic documents practical.

### AN ARCHITECTURAL BASIS FOR INTERCHANGE

There are three basic architectures that facilitate interchange of electronic documents: (1) the architecture for document preparation, (2) the architecture of the document representation, and (3) the architecture for extension. These three architectures are illustrated in [Figure 4](#). The architecture for document preparation is shown on the left-hand side of the figure and encompasses both electronic and paper document preparation processes. The right-hand side of the figure shows consumption of prepared documents. The portable document format (PDF) is the architecture for document representation and is the link between these components. The right-hand side of the figure shows consumption at two levels. There is an optional cataloguing step that builds full text indexes for one or a collection of documents. Above this, viewing and printing PDF documents are shown. The architecture for extension is indicated by the "search" plug-in, which allows access to indexes built by the cataloguing process.

#### The Architecture for Document Preparation

To be effective, any system for electronic document interchange must be able to capture documents in all the forms in which they are generated. This is facilitated by the recent shift to electronic preparation of documents, but it must include a pathway for paper documents as well.

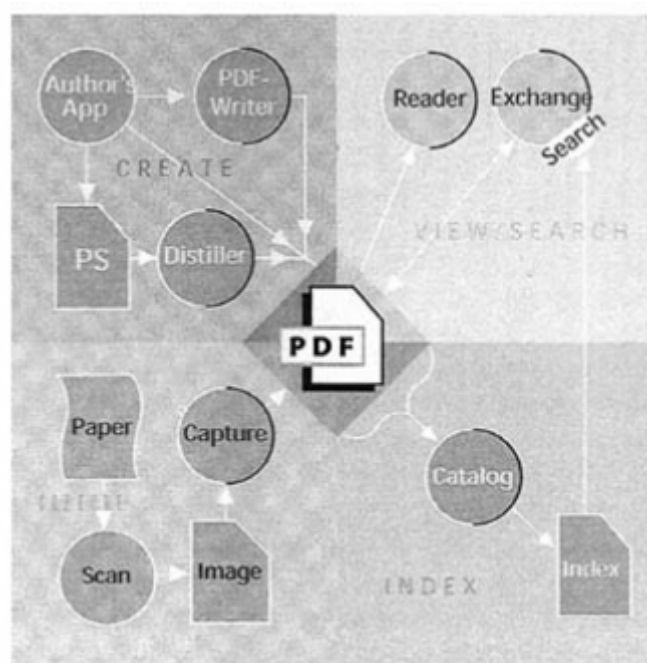


Figure 4

Unlike the wide range of forms that revisable documents can take, there are relatively few final form formats in use today. (This is another reason that it makes sense to have a final form interchange format.) Since, historically, final forms were prepared for printing, one can capture documents in final form by replacing or using components of the printing architectures of the various operating systems. Two such approaches have been used with PDF: (1) in operating systems with selectable print drivers, adding a print driver that generates PDF and (2) translating existing visually rich print formats to PDF. Both these pathways are shown in the upper left corner of Figure 4. PDFWriter is the replacement print driver and the Acrobat Distiller translates PostScript (PS) language files into PDF.

Some operating systems, such as the Mac OS and Windows, have a standard interface, the GUI (graphical user interface), which can be used by any application both to display information on the screen and to print what is displayed. By replacing the print driver that lies beneath the GUI it is possible to capture the document that would have been printed and to convert it into the electronic document interchange format.

For operating systems without a GUI interface to printing and for applications that choose to generate their print format directly, the PostScript language is the industry standard for describing visually rich final form documents. Therefore, the electronic document interchange format can be created by translating, or "distilling," PostScript language files. This distillation process converts the print description into a form more suitable for viewing and electronic navigation. The PostScript language has been extended, for distillation, to allow information on navigation to be included with the print description, allowing the distillation process to automatically generate navigational aids.

The above two approaches to creation of PDF documents work with electronic preparation processes. But there is also an archive of legacy documents that were never prepared electronically or are not now available in electronic form. For these documents there is a third pathway to PDF, shown in the lower left corner of Figure 4. Paper documents can be scanned and converted to raster images. These images are then fed to a recognition program, Acrobat Capture, that identifies the textual and nontextual parts of the document. The textual parts are converted to coded text with appropriate font information including font name, weight, size, posture. The nontextual parts remain as images. This process produces an electronic representation of the paper document that has the same final form as the original and is much smaller than the scanned version. Because the paper document is a final form, the same final form format can be used, without loss of fidelity, for paper documents and for the electronically generated documents.

Page fidelity is important. The current state of recognition technology, though very good, is not infallible; there are always some characters that cannot be identified with a high level of confidence. Because the PDF format allows text and image data to be freely intermixed, characters or text fragments whose recognition confidence falls below an adjustable level can be placed in the PDF document as images. These images can then be read by a human reader even if a mechanical reader could not interpret them. (Figure 2 shows a document captured by this process.)

### The Architecture of the Document Representation

There is more to the architecture of the document representation than meeting the above requirements for a final form representation. Architectures need to be robust and flexible if they are to be useful over a continuing span of years. PDF has such an architecture.

The PDF architecture certainly meets these requirements, as will be clear below. Most importantly, PDF has an open specification: the *Portable Document Format Reference Manual* (ISBN 0-201-62628-4) has been published for several years, and implementations have been produced by several vendors.

PDF also goes beyond the final form requirements. For example, the content of PDF files can be randomly accessed and the files themselves can be generated in a single pass through the document being converted to PDF form. In addition, incremental changes to a PDF file require only incremental additions to the file rather than a complete rewrite of the file. These are aspects of PDF that are important with respect to the efficiency of the generation and viewing processes.

The PDF file format is based on long experience both with a practical document interchange format and with applications that were constructed on top of that format. Adobe Illustrator is a graphics design program whose intermediate file format is based on the PostScript language. By making the intermediate format also be a print format, the output of Adobe Illustrator could easily be imported into other applications because they could print the objects without having to interpret the semantics. In addition, because the underlying semantics were published, these objects could be read by other applications when required. The lessons learned in the development of Adobe Illustrator went into the design of PDF.

PDF, like the Adobe Illustrator file format, is based on the PostScript language. PDF uses the PostScript language imaging model, which has proven itself over 12 years of experience as being capable of faithfully representing visually rich documents. Yet the PostScript file format was designed for printing, not for interactive access. To improve system performance for interactive access, PDF has a restructured and simplified description language.

Experience with the PostScript language has shown that, although having a full programming language capability is useful, a properly chosen set of high-level combinations of the PostScript language primitives can be used to describe most, if not all, final form pages. Therefore, PDF has a fixed vocabulary of high-level operators that can be more efficiently implemented than arbitrary combinations of the lower-level primitives.

### The User Model for PDF Documents

The user sees a PDF document as a collection of pages. Each page has a content portion that represents the final form of that page and a number of virtual overlays that augment the page in various ways. For example, there are overlay layers for annotations, such as electronic sticky notes, voice annotations, and the like. There are overlay layers for hyperlinks to other parts of the same document or hyperlinks to other documents and other kinds of objects, such as video segments or animations. There is an overlay layer that identifies the threading of the content of articles from page to page and from column to column. Each of the overlay layers is associated with the content portion geometrically. Each overlay object has an associated rectangle that encompasses the portion of content associated with the object.

Each of the layers is independent of the others. This allows information in one layer to be extracted, replaced, or imported without affecting the other layers. This facilitates exporting annotations made on multiple

copies of a document sent out for review and then reimporting all the review annotations into a single document for responding to the reviewers' comments. This also makes it possible to define hyperlinks and threads on the layout of a document that only has the text portion present and then to replace the text-only pages with pages that include the figures and images to create the finished document.

In addition to the page-oriented navigational layers, there are two document-level navigation aids. There is a set of bookmark or outline objects that allow a table of contents or index to be defined into the set of pages. Each bookmark is a link to a particular destination in the document. A destination specifies the target page and the area on that page that is the target for display. Destinations can be specified directly or named and referred to by name. Using named destinations, especially for links to other documents, allows the other documents to be revised without invalidating the destination reference.

Finally, associated with each page is an optional thumbnail image of the page content. These thumbnails can be arrayed in sequence in a slide sorter array and can be used both to navigate among pages and to reorder, move, delete, and insert pages within and among documents.

### **The Abstract Model of a PDF Document: A Tree**

Abstractly, the PDF document is represented as a series of trees. A primary tree represents the set of pages and secondary trees represent the document-level objects described in the user model. Each page is itself a small tree with a branch for the representation of the page content; a branch for the resources, such as fonts and images used on the page; a branch for the annotations and links defined on the page; and a branch for the optional thumbnail image. The page content is represented as a sequence of high-level PostScript language imaging model operators. The resources used are represented as references to resource objects that can be shared among pages. There is an array of annotation and link objects.

### **The Representation of the Abstract Tree**

The abstract document tree is represented in terms of the primitive building blocks of the PostScript language. There are five simple objects and three complex objects. The simple objects are the null object (which is a placeholder), the Boolean object (which is either true or false), the number object (which is an integer or fixed point), the string object (which has between 0 and 65535 octets), and the name object (which is a read-only string).

The three complex objects are arrays, dictionaries, and streams. Arrays are sequences of 0 to 65535 objects that may be mixed type and may include other arrays. Dictionaries are sets of up to 65535 key-value pairs where the key is a name and the value is any object. Streams are composed of a dictionary and an arbitrary sequence of octets. The dictionary allows the content of the streams to be encoded and/or compressed to improve space efficiency. Encoding algorithms are used to limit the range of octets that appear in the representation of the stream. Those defined in PDF are ASCIIHex (each hex digit is represented as an octet) and ASCII85 (each four octets of the stream are represented as five octets). These both produce octet strings restricted to the 7-bit ASCII graphic character space. Compression algorithms are used to reduce storage requirements. Those defined in PDF are LZW (licensed from Univac), Run length, CCITT Group 3 and Group 4 FAX and DCT (JPEG).

The terminal nodes of the abstract tree are represented by simple objects and streams. The nonterminal nodes are represented by arrays and dictionaries. The branches (arcs) of the tree are represented in one of two ways. The simplest way is that the referenced object is directly present in the nonterminal node object. This is called a direct object. The second form of branch is an indirect object reference. Objects can be made into indirect objects by giving the (direct) object an object number and a generation number. These indirect objects can then be referenced by using the object number and generation number in place of the occurrence of the direct object.

Indirect objects and indirect object references allow objects to be shared. For example, a font used on several pages need only be stored once in the document. They also allow the values of certain keys, such as the

length of a stream, to be deferred until the value is known. This property is needed to allow PDF to be produced in one pass through the input to the PDF generation process.

### The PDF File Structure

Indirect objects and indirect object references do not allow direct access to the objects. This problem is solved by the PDF file structure. There are four parts to the file structure. The first part is a header, which identifies the file as being a PDF file and indicates the version of PDF being used in the file. The second part is the body, which is a sequence of indirect objects. The third part is the cross-reference table. This table is a directory that maps object numbers to offsets in the (body of the) file structure. This allows direct access to the indirectly referenced objects.

The final part is the Trailer, which serves several purposes. It is the last thing in the file and it has the offset of the corresponding cross-reference table. It also has a dictionary object. This dictionary is the size of the cross-reference table. It indicates which indirect object is the root of the document tree. It indicates which object is the "info dictionary," a set of keys that allow attributes to be associated with the document. These keys include such information as author, creation date, etc.

Finally, the trailer dictionary can have an ID key whose value has two parts. Both parts are typically hash functions applied to parts of the document and key information about the document. The first hash is created when the document is first stored; it is never modified after that. The second is changed whenever the document is stored. By storing these IDs with file specifications referencing the document, one can more accurately determine that the document retrieved via a given file specification is the document that is desired.

The trailer is structured to allow PDF files to be incrementally updated. This allows PDF files to be edited, say deleting some pages or adding links or annotations, without having to rewrite the entire file. For large files, this can be a significant savings. This is accomplished by adding any new indirect objects after the existing final trailer and appending a new cross-reference table and trailer to the end of the file. The new cross-reference table provides access to the new objects and hides any deleted objects. This mechanism also provides a form of "undo" capability. Move the end of the file back to the last byte of the previous trailer and all changes made since that trailer was written will be removed.

The purpose of the generation numbers in the indirect object definition and reference is to allow reuse of table entries in the cross-reference table when objects are deleted. This keeps the cross-reference table from growing arbitrarily large. Any indirect object reference is looked up in the endmost cross-reference table in the document. If the generation number in that cross-reference table does not match the generation number in the indirect reference, then the reference object no longer exists, the reference is bad, and an error is reported. Deleted or unused entries in the cross-reference table are threaded on a list of free entries.

### Resources

The general form and representation of a PDF file have been outlined. There are, however, several areas that need further detail. The page content representation is designed to refer to a collection of resources external to the pages. These include the representations of color spaces, fonts, images, and shareable content fragments. For device-independent color spaces, the color space resource contains the information needed to map colors in that color space to the standard CIE 1931 XYZ color space and thereby ensure accurate reproduction across a range of devices. Images are represented as arrays of sample values that come from a specified color space and may be compressed. Page content fragments are represented as content language subroutines that can be referred to from content. For example, a corporate logo might be used on many pages, but the content operators that draw the logo need be stored only once as a resource.

Typically, however, the resources that are most critical for ensured reproduction are the font resources. The correct fonts are needed to be able to faithfully reproduce the text as it was published. PDF has a three-level approach to font resources. First, there is a set of 13 fonts (12 textual fonts and 1 symbol font) that must be

available to the viewer. These fonts can be assumed to exist at any consumer of a PDF document. For other fonts, there are two solutions. The fonts may be embedded within the document or substitutions may be made for the fonts if they are not available on the consumer's platform. Fonts that are embedded may be either Adobe Type 1 fonts or TrueType fonts and may be the full font or a subset of the font sufficient to display the document in which they are embedded.

The font architecture divides the font representation into three separate parts: the font dictionary, the font encoding, and the font descriptor. The font dictionary represents the font and may refer to a font encoding and/or a font descriptor. The font encoding maps octet values that occur in a string into the names of the glyphs in a font. The font descriptor has the metrics of the font, including the width and height of glyphs and attributes such as the weight of stems, whether it is italic, the height of lower-case letters, and so on. The font shape data, if included, are part of the font descriptor. If the font shape data are not included, then the other information in the font descriptor can be used to provide substitute fonts. Substitute fonts work for textual fonts and replace the expected glyphs with glyphs that have the same width, height, and weight as the original glyphs. If page fidelity is required, then the font shape data should be embedded; but font substitution can be used to reduce document size where the omitted fonts are either expected at the consumer's location or font substitution is adequate for reading the document.

### Hyperlinks

The hyperlink mechanism has two parts: the specification of where the link is and the specification of where the link goes. The first specification is given as a rectangular area defined on the page content. The second specification is called an action. There are a number of different action types. The simplest is moving to a different destination within the same document. A more complex action is moving to a destination in another PDF document. The destination may be a position in the document, a named destination, or the beginning of an article thread. Instead of making another PDF document available for viewing, the external reference may launch an application on a particular file, such as a fragment of sound or a video. All these external references use platform independent file names, which may be relative to the file containing the reference document, to refer to external entities.

The URL (Uniform Resource Locator), as defined for the World Wide Web, is another form of allowed reference to an external document. A URL identifies a file (or part thereof) that may be anywhere in the electronically reachable world. When the URL is followed, the object retrieved is typed and then a program that can process that type is invoked to display the object. Any type for which there is a viewing program, including PDF, can thereby be displayed.

### Extensibility

PDF is particularly extensible. It is constructed from simple building blocks; the PDF file is a tree constructed from leaves that are simple data types or streams and with arrays and dictionaries as the nonterminal nodes. In general, additional keys can be added to dictionaries without affecting viewers who do not understand the new keys. These additional keys may be used to add information needed to control and/or represent new content object types and to define editing on existing objects. Because of the flexibility of the extension mechanism, a registry has been defined to help avoid key name conflicts that might arise when several extensions are simultaneously present.

### The Architecture for Extensions

The third component architecture for final form electronic document interchange is the extension architecture for consumers of the electronic documents. Viewing a PDF document is always possible and the PDF specification defines what viewing means. But, if there are extensions within the PDF file, there must be a

way to give semantic interpretation to the extension data. In addition, vendors may want to integrate a PDF consumer application, such as Acrobat Exchange, with their applications. For example, a service that provides information on stocks and bonds may want to seamlessly display well-formatted reports on particular stocks. This service would like to include the display of PDF documents with their non-PDF information. Providing semantics both for extension data and for application integration can be accomplished using the extension architecture for PDF viewers.

It is reasonable to look at PDF viewers as operating system extensions. These viewers provide a basic capability to view and print any PDF document. By extending the view and print application programming interfaces (APIs), more powerful applications can be constructed on top of the basic view and print capabilities of a PDF viewer. These extended applications, called plug-ins, can access extended data stored in the PDF file, change the viewer's user interface, extend the functionality of PDF, create new link and action types, and define limited editing of PDF files.

The client search mechanism shown in [Figure 4](#) was done as a plug-in to Acrobat Exchange. The search plug-in presumes that collections of PDF documents have been indexed using Acrobat Catalog. The plug-in is capable of accessing the resulting indexes, retrieving selected documents, and highlighting occurrences that match the search criteria. This plug-in is shipped with Acrobat Exchange but could be replaced by other vendors with another mechanism for building indexes and retrieving documents. Hence, PDF files can be incorporated into many document management systems.

### DEPLOYMENT AND THE FUTURE

The process of PDF deployment has already begun. One can find a variety of documents in PDF form on the World Wide Web, on CD-ROMs, and from other electronic sources. These documents range from tax forms from the Internal Revenue Service, to color newspapers, commercial advertisements and catalogs, product drawings and specifications, and standard business and legal documents.

Use of PDF is likely to increase as more document producers understand the technology and learn that it is well adapted to current document production processes. The greatest barrier to expansion of consumption is awareness on the part of the consumers. There are free viewers for PDF files, the Acrobat Readers, available for most major user platforms (DOS, Sun UNIX, Macintosh OS, Windows) and more support is coming. These viewers are available on-line through a variety of services, are embedded in CD-ROMs, and are distributed on diskette.

At this level, the barrier to deployment is primarily education. But there are also opportunities to improve the quality of electronic document interchange. Some examples of these improvements are better support for navigational aids; support for other content types, such as audio and video; support for a structural decomposition of the document, as is done in SGML; and support for a higher level of document editing.

Current document production processes naturally produce the final form of the document, but they do not necessarily enable navigation aids such as hyperlinks and bookmarks/tables of contents. The document production architecture does provide a pathway for this information to be passed to the distillation process and through the print drivers. As producers enable this pathway in their document production products, it will become standard to automatically translate the representation of navigational information in a document production product into the corresponding PDF representation of navigational aids.

Another direction for future development is the inclusion of additional content types within a PDF file. (There is already support for referencing foreign content types stored in separate files via the hyperlink mechanism.) Some of the obvious content types that should be included are audio, video, and animation. There is also a need for orchestrating multiple actions/events when content is expanded beyond typical pages. Much of the barrier to inclusion of these other content types is in the lack of standard formats for these content types. Because PDF is designed to run across all platforms, there is a particular need for standards that are capable of being implemented in all environments. For example, standards that require hardware assists are not as useful as standards that can be helped by hardware assists but do not require them.

A final form interchange format guarantees viewability of the information in a document, but it does not necessarily provide for reuse or revision of the information. Structural information representations, such as SGML and HTML, can simplify reuse, but they do not capture the decisions of human layout designers. Best would be a format that allowed both views: the final form view for browsing and reading, and the ability to recover the structural form for re-purposing, editing, or structure-based content retrieval. PDF will be extended to allow the formatted content to be related to the structural information from which it was produced and to allow that structured information to be retrieved for further use.

Clearly the final form document contains some of the information that is needed to edit the document, but it is equally clear that without extensions to represent structure as well as form the document may not contain information about how the components of the final form were created and how they might be changed. Such simple things as what text was automatically generated, what elements were grouped together to be treated as a whole, and into what containers text was flowed need not be represented in the final form.

The PDF representation was constructed from a set of primitive building blocks that are also suitable for representing structural and other information needed for editing. Augmenting the final form with this kind of information, using these powerful and flexible building blocks, would allow the final form document format to offer revisability. As a simple example, one might use PDF to represent a photo album as a collection of scanned images placed on pages. A simple editor might be defined that allows these photos to be reused—say, to make a greeting card by combining text with images selected from the photo album. The greeting cards thus constructed could be represented in PDF using extensions that allow editing of the added text. More complex editing tasks can be accommodated by capturing more information about the editing context within the PDF file generated by the editing application. For some applications, the PDF file might be the only storage format needed; it would be both revisable and final form.

## CONCLUSION

The business case for final form electronic document interchange is relatively straightforward. There are significant savings to be achieved simply by replacing paper distribution with electronic distribution, whether or not the document is printed at the receiving site. The key success factor is whether the document can be consumed once received. Consumption most often means access to the document's contents in the form in which they were published. This can be achieved by having a small number of final form interchange formats (preferably one) and universal distribution of viewers for these formats. The portable document format (PDF) is a more than suitable final form interchange format with freely distributable viewers.

For practical interchange, there must be tools to conveniently produce the interchange format from existing (and future) document production processes. The interchange format must be able to be transmitted through the electronic networks and included on disks, diskettes, CD-ROMs, and other physical distribution media. It must be open to allow multiple implementations and to ensure against the demise of any particular implementation. Finally, it must be extensible to allow growth with the changing requirements of information distribution. These features all are met by PDF.

PDF provides a universal format for distributing information as an electronic document. The information can always be viewed and printed. And, with extensions, it may be edited and integrated with other information system components.