DOCUMENT RESUME

ED 378 938 IR 016 933

TITLE	Research Priorities in Networking and Communications.
INSTITUTION	National Science Foundation, Washington, D.C.
REPORT NO	NSF-94-165
PUB DATE	94
NOTE	51p.; Report by members of a workshop sponsored by the Networking and Communications Research Program of the National Science Foundation (Arlington, VA, May 12-14, 1994).
AVAILABLE FROM	National Science Foundation, Washington, DC 22230.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Access to Information; Agenda Setting; *Computer Networks; *Information Networks; Information Theory; Needs Assessment; *Program Development; *Research Needs; Systems Development; *Telecommunications; Workshops
IDENTIFIERS	*National Science Foundation

ABSTRACT

A workshop focused on major research issues in networking and communications. This report defines the context for research priorities and initiatives and deals with issues in networking and communications. Fifteen major research priorities and four research specific initiatives were identified by participants as areas that should be pursued over the next several years. Priorities identified for basic research are: (1) coding and coded modulation; (2) information theory; (3) communications signal processing; (4) protocol theory, design, and engineering; (5) dynamic network control; (6) internetworking; (7) optical networks; (8) security; (9) reliable networks; (10) switching networks; (11) wireless networks and access; (12) data storage systems; (13) video, speech, image, and data compression; (14) mobile and nomadic computing; and (15) satellite systems. (SLD)

RESEARCH PRIORITIES IN NETWORKING AND COMMUNICATIONS

Report to the NSF Division of Networking and Communications Research and Infrastructure by Members of the Workshop Held
May 12-14, 1994
Airlie House, Virginia

NATIONAL SCIENCE FOUNDATION

Arlington, VA 22230

The opinions expressed in this report are those of the workshop panel and do not necessarily represent NSF policy.

In accordance with Federal statues and regulations and National Science Foundation (NSF) policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, denied the benefits of, or be subject to discrimination under any program or activity receiving financial assistance from the NSF.

The Foundation has TDD (Telephonic Device for the Deaf) capability which enables individuals with hearing impairment to communicate with the Division of Personnel and Management for information relating to NSF programs, employment, or general information. This number is (703) 306-0090.

Facilitation Awards for Scientists and Engineers with Disabilities (FAD) provide funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on an NSF project. See the FAD program announcement (NSF 91-54), or contact the FAD Coordinator in the Directorate for Education and Human resources. The telephone number is (703) 305-1636.

The Foundation provides awards for research and education in most fields of science and engineering. The awardee is wholly responsible for the conduct of such research and preparation of the results for publication. The Foundation, therefore, does not assume responsibility for such findings or their interpretation.

The Foundation welcomes proposals on behalf of all qualified scientists and engineers, and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and research-related Programs described in this document.

Ordering by Electronic Mail or Fax

If you are a user of electronic mail and have access to Internet, you may order publications electronically. Internet users should address requests to pubs@nsf.gov. In your request, include the NSF publication number and title, number of copies, your name, and a complete mailing address. Printed publications may be ordered by FAX (703-644-4278). Publications should be received within three weeks after receipt of request.

This program is described in the Catalog of Federal Domestic Assistance, Number 47.070, Computer and Information Science and Engineering.

PREFACE

The Networking and Communications Research Program of the National Science Foundation supported a two-day workshop at Airlie House, Virginia, May 12-14, 1994. This was the third such workshop, following the second by slightly more than two years. The focus of the workshop was to identify major research issues in networking and communications. The goal of the workshop was to produce a vision of telecommunications for the future and a research path to implement that vision.

This report defines the context for research priorities and research initiatives by addressing many important applications of networking and communications in today's telecommunications and information systems. The report deals with issues in networking and communications and identifies and discusses in detail fifteen major research priorities and four research initiatives that the participants concluded should be pursued over the next several years. In addition, a number of research areas were identified that are suitable for collaborative efforts between the various divisions of NSF and between NSF and other appropriate federal agencies.

RESEARCH PRIORITIES IN NETWORKING AND COMMUNICATIONS

Report of a Workshop held May 12-14, 1994, Airlie House, Virginia

Sponsor: NSF Division of Networking and Communications Research and Infrastructure

Participants:

Richard Blahut, University of Illinois, Urbana
Charles A. Brackett, Bellcore
David D. Clark, Massachusetts Institute of Technology
Donald C. Cox, Stanford University
John N. Daigle, University of Mississippi
G. David Forney, Jr., Codex/Motorola Corporation
Robert G. Gallager, Massachusetts Institute of Technology
Zygmunt Haas, AT&T Bell Laboratories
Bruce E. Hajek, University of Illinois, Urbana
Robert Kahn, Corporation for National Research Initiatives
Simon S. Lam, University of Texas
Shu Lin, University of Hawaii at Manoa
Robert W. Lucky, Bellcore
W. D. Sincoskie, Bellcore
* William H. Tranter, University of Missouri at Rolla
Jonathan Turner, Washington University, St. Louis
Jack K. Wolf, University of California at San Diego

* Chair

TABLE OF CONTENTS

1. INTRODUCTION

The NSF Division of Networking and Communications Research and Infrastructure was established in 1987, shortly after the creation of the Computer and Information Science and Engineering Directorate. In 1989, a workshop was convened to assess the program and to suggest directions for communications and networking research. A similar group reconvened in April 1992 to update the suggestions made by the 1989 workshop, to reconsider the assessment of research directions vis-a-vis research opportunities and national needs, and to evaluate the educational and research base in these areas and in the United States.

Since the 1992 workshop the drive toward a National Information Infrastructure (NII) has captured the attention of the general public as well as the research community. In view of rapid changes in the fields of communications and networking, and the pace of policy decisions toward the NII, NSF convened a third workshop in May, 1994. The purpose of this workshop was to identify the major issues affecting networking and communications research for the future, including the NII, and to identify research priorities.

The workshop participants made the following observations regarding the environment of networking and communications research:

- The areas of communications, networking, and information storage constitute major and rapidly growing segments of the U. S. industrial base. A metric of the importance of networking and communications is the 100% per year growth rate of Internet traffic. With the introduction of user-friendly tools, such as MOSAIC (utilization of MOSAIC has initially been increasing at 11% per week) network traffic and requirements for bandwidth and storage are certain to increase dramatically.

- During the last several years the technologies for computing, communications, networking, cable TV, entertainment, and information storage have been converging at an increasing rate. As a result of this convergence, many predict a radical and revolutionary change in the way people live and work.

- Research in communications, networking and information storage has contributed greatly to the development of the Internet. A substantial new industry has emerged in the past four years, in the wake of the Internet, supporting and applying progress in networking technology.

- Critical policy decisions are required to set the stage for efficient, cost effective, and user friendly implementation of the new information and communications based society. Informed policy decisions benefit from the insights generated by networking and communications research.

- The National Information Infrastructure (NII) has been declared a national goal and objective. There is a growing awareness of its influence on education, health care, commerce, and government. Networking and communications are among the core enabling technologies of the NII.

- The increasing level of competition in the telecommunications industry has led to declining basic research funding in the communications and networking areas. In addition, industry has turned to more near-term research and development. These trends have created both an opportunity and a requirement for NSF to take a more proactive role in guiding basic research leading to national strategic goals. Increased funding levels at NSF for research in communications, networking and information storage would contribute significantly to the rapid realization of the NII.

Future communications networks that are suitable for a National Information Infrastructure will be expected to support a wide range of different applications, in a cost-effective manner, with unprecedented reliability and on a very large scale. They will be expected to:

- support applications with a variety of data rates from a few bits per second to gigabits per second, and applications whose data rates may vary widely during a short time period;

- provide a range of quality-of-service options to accommodate different performance and cost requirements;

- support mobile users and users who access the network from a variety of locations;

- provide one-to-many and many-to-many communication channels to support information distribution and collaborative applications.

In addition, the NII will require network architectures that can accommodate evolution across multiple generations of technology and that facilitate heterogeneity in applications, end systems, transmission technologies and switching mechanisms. To allow future networks to be managed effectively they must provide mechanisms for traffic measurement, error monitoring, usage accounting and cost recovery.

Workshop participants identified research priorities and recommended four major research initiatives. In addition, a number of research areas, suitable for collaborative efforts between various NSF divisions and between NSF and other federal agencies, were identified. These priorities and initiatives are detailed in Sections 4, 5, and 6 of this report.

2. THE NATURE OF BASIC RESEARCH IN NETWORKS AND COMMUNICATIONS

The support for basic research in communications and data networks is decreasing both in industry and in the department of defense. Thus it is particularly essential for this support to increase at NSF. Part of the reason for the difficulty in supporting basic research in these areas is a lack of understanding about the nature of basic research. These misconceptions characterize basic research as high risk, long term, and curiosity driven. To the contrary, we maintain that basic research is often low risk, short term, and vital to national priorities.

In understanding the nature of basic research, it is perhaps better to call it discipline - oriented research. This is research in which one builds understanding in a particular discipline by analyzing and studying successively more complex models, using the insights generated at each stage to guide further development. Such disciplinary research develops a body of insights, of generic, well-understood simple problems, of back of the envelope calculations, and of algorithms and methodologies for coming to grips with real problems.

Much of modern communications rests on the well established, but still vital, disciplines of modulation, coding, detection, compression, equalization, signal processing, etc. As application areas and technologies change, new models are suggested for these disciplines, but the new models use and build on the existing discipline, enriching it for the next set of applications. Thus practical needs, combined with discipline-oriented research, ideally work in synergy with each other, the needs enriching the discipline, and the discipline providing the insights to satisfy the current needs. The existence of disciplinary understanding permits rapid short-term solutions of new problems, and the possibility of finding these solutions is low risk because of the existing disciplinary insights.

Along with the above synergism, discipline-driven research in communications has often solved important problems almost by chance. For example, the Lempel-Ziv algorithms (the workhorses of modern data compression) were developed out of a very theoretical study of data sequences without stochastic structure. The Viterbi algorithm (one of the workhorses of modern error correction) was developed out of an effort to understand theoretical error probability bounds. The Reed-Solomon codes (another workhorse of modern error correction, used in compact discs) were developed primarily because of their mathematical elegance. One could view such chance solutions as "high risk," but in fact these were simply added bonuses to work that had to be done to maintain the connection between real problems and discipline-oriented research. Each of these bonuses in fact created entire industries. Research was an investment with extraordinary return in these cases. These bonuses were curiosity driven, but in fact satisfied national needs.

The above picture describes the almost ideal relationship between theory (discipline-oriented research) and practice as it exists in communications today, but this is the result of the rich discipline that has been developed in the past. Because of the past successes of this relationship, communications theorists continue to enrich the discipline while solving the immediate problems of the present.

In the network area, theory and practice have had a less satisfactory relationship than in the communications area. The network area has grown up more recently than communications, and its disciplines, such as congestion control, routing, and distributed algorithms, have had a relatively short gestation period. In addition, the very rapid technological changes in the last 15 years have never allowed a strong connection between theory and practice to develop. It is becoming increasingly apparent, however, that the disciplines of congestion control, routing, and addressing are generic to all networks. The current practice of re-solving these problems in an ad hoc way for every new practical network leads to repeated poor solutions. There is an urgent need for NSF to support the development of sufficiently strong disciplines that can interact harmoniously with new network needs.

Maintaining an appropriate balance between discipline-oriented and application-oriented research is difficult in today's climate. There are strongly perceived pressures to solve practical problems quickly, and there is a temptation for discipline-oriented researchers to give up on the rapidly changing technological scene, and retreat into abstraction and detail. It is important for both application-oriented and discipline-oriented researchers to recognize their mutual dependence upon each other, and to make this dependence clear to administrators and executives. It is particularly important to cultivate that small group of individuals who have a deep understanding of both applications and disciplines. NSF should be proud of its efforts to support both practical needs and disciplines, and to encourage the linkages between them. A major challenge for NSF is to draw on the wisdom of the community to continue to strengthen these linkages.

3. THE NSF PROGRAM

Organizational Base of the Program

Within the Computer and Information Science and Engineering (CISE) Directorate, the Division of Networking and Communications Research and Infrastructure (DNCRI) encompasses two major activities; the Network Infrastructure and the Networking and Communication Research (NCR) activities. The infrastructure activity includes the NSF involvement in the Internet, via the NSFNET and international connections, currently in transition to a commercially oriented focus involving network access points (NAP's), a Routing Arbiter (RA), and a very high speed network (vBNS) serving the research and education community beginning with the supercomputer centers. In addition, the infrastructure activity includes special supporting projects, such as the InterNIC, CNIDR, EDGAR and the Global School House, designed to further the usability and friendliness of the network. NCR activity focuses on basic research projects furthering the knowledge base in information theory, networking, and communications. Where possible and appropriate these two aspects of division activity complement each other.

Goals of the Program

The long term vision of the Networking and Communications program is seen as providing the underlying knowledge leading to a national or international infrastructure with a primarily optical backbone or core, including satellites as appropriate, with cable, copper twisted pair, fiber, or wireless access to provide ubiquitous communications and information services on demand. Transmission links and storage systems require effective and efficient utilization of spectrum, space, and time for the communication of voice, video, text, graphics, or other data from point to point in space and/or time. Distributed or networked communications requires extensive protocols for seamless and interoperable information exchange. Privacy, authentication, and intellectual property rights require expanded attention in a networked environment. Network management of shared resources is a fundamental requirement. Basic research required to enable this vision is the goal of the program.

The focus of the program is to find and support, to the extent of available funding, the very best basic research in the area encompassed by the program. Since computer science, mathematics, electrical and computer engineering, and other disciplines are strong contributors to networking and communications, another goal of the program is to further collaboration among these communities of interest. Catalyzing an atmosphere in which multidisciplinary teams can work together across the spectrum from materials and devices to systems to networking and communications theory is also a goal of the program. Within these goals there is an objective of relating to the infrastructure activities of the Division to enhance the vision of the future and the choice of research projects.

Continuous evolution of educational resources in networking and communications as technology advances is essential. For DNCRI this includes not only involvement of graduate students in research projects, and ongoing curriculum development to include new knowledge developed in the research program, but also the development and implementation of new tools for educational delivery systems, including high bandwidth networks, digital libraries, and networked information resources.

Relationship of the NCR Program to the NII and HPCC Initiatives

The DNCRI activities in creating the NSFNET backbone network, enhancing its capabilities through higher speed access and wider connectivity throughout the research and education communities, and through progress in the development of tools for utilization of network resources focuses attention on the potential for information exchange and resource sharing via networking. The recognition of the value of networking is made clear by the enormous growth rate of network utilization and by the rapid emergence of new national industries based on networking. The rationale for the transition from a government sponsored NSFNET backbone to a commercial focus is the current availability of network services commercially where none existed previously. NSF will continue to push the edge of services at the state of the art and contribute to and accelerate commercial development.

Furtherance of the objective of networking and information exchange has been pressed forward by the Congress through the High Performance Computing and Communications (HPCC) initiative in 1990 and by the National Information Infrastructure initiative of 1993.

Both DNCRI activities in infrastructure development and in research are central to and provide the very core for these important initiatives. The entire division is central to the HPCC initiative. Networking and communications constitute in fact the second "C" and are central to distributed computing applications. The growth of the Internet is dependent upon basic research in networking and communications, and the future NII will also be based on technology developments resulting from basic research.

Operation and Funding History of the NCR Program

The NCR program was created in 1987 as a part of DNCRI and CISE. The initial funding level was very small. Funding for the program has increased from $3.36 M in FY89 to $10.3 M in FY94. The number of active projects has grown from 42 in 1989 to 102 in 1994. A major research initiative in gigabit per second networks, led by the Corporation for National Research Initiatives (CNRI) and jointly funded by DNCRI/NSF and ARPA and involving five testbeds, began in FY89 and ends its initial phase in FY94. In FY93, a special solicitation on All Optical Networking in collaboration with the Engineering Directorate at NSF resulted in jointly funded interdisciplinary projects.

Networking and communications research is supported under one program (NCR). The program is administered by a Program Director and a Program Manager. No a priori division of the program funds is made; the best proposals from all submissions are selected for award.

The program draws insight as much as possible from the research communities it serves, including computer science and electrical engineering and related disciplines, via workshops and symposia, review panel discussions, and personal interactions. A program goal is to enhance interaction between computer scientists, electrical engineers, device physicists, mathematicians, and others to bring about an integrated approach. In particular, the networking and communications aspects of the program are treated in an integrated manner rather than separately. A primary thrust is fundamental theoretical research. Experimental research, drawing upon the network infrastructure as a tool and at the same time contributing to the future development of the infrastructure is a secondary thrust, but is limited by the high cost of such projects compared to available funds.

4. PRIORITIES FOR BASIC RESEARCH

A number of research areas were identified by the workshop participants as being especially important and timely. The priorities identified in the following paragraphs will support the development of future systems, both by providing the underlying theory to guide system development and, more generally, by strengthening the necessary educational base. While research in the areas listed below is encouraged by the workshop participants, the following list of topics is not intended to discourage the submission of proposals in other areas. The order of topics presented in the following paragraphs is entirely arbitrary and no priority assignment is implied.

Coding and Coded Modulation

The future communications network will be digital, and will be based on a fiber-optic backbone with almost unlimited capacity. At the same time, the use of wireless communications will continue to explode, both to provide access to the network for people and computers on the move, and in stand-alone radio networks for specialized applications. Broadcast television will become digital in HDTV. More and more bits will be sent down the existing copper wires that go to the individual home or desk. New satellite systems for high speed digital communications will be developed.

On channels with less than unlimited capacity, it is well understood that coding is needed to achieve the best efficiency at low error rates. Powerful error-correcting codes and error control techniques are now used almost routinely in data communications and storage systems. The recent invention of coded modulation has revolutionized communications over bandlimited channels, and is starting to be used in magnetic storage.

As a research field, coding and coded modulation contains both well explored and newly emerging areas. There are many fundamental problems to be answered and practical problems to be solved. Important current research includes the following areas in code structure, code construction, efficient decoding algorithms, quantization dual of Euclidean-space coding, and combined coding and equalization:

- A unified structure theory embracing block, convolutional, lattice and trellis codes has begun to emerge. There have been exciting advances in developing structural properties of block and lattice codes, such as the trellis and decomposable structures, which allow the use of soft-decision decoding to achieve better performance with reduced decoding complexity. One goal is to have a better understanding of the structural properties of existing codes so that efficient decoding schemes can be devised, and another goal is to construct new codes which have good distance properties and are easy to decode.

- Coding techniques for memoryless channels have been successfully extended to channels with intersymbol interference. New codes and suitable decoding algorithms are needed for other types of channels, such as fading and bursty channels. Application of coding for fading and interference channels to wireless transmission deserves to be investigated.

- There have been exciting developments recently in Euclidean-space group and ring codes, and in Hamming-space algebraic-geometry codes. Our knowledge of these classes of algebraic codes remains far from complete.

- Suboptimal decoding algorithms that can approach optimal or near optimal performance with significantly reduced decoding complexity are likely to be the best choice for high-speed and high-performance applications. New hardware and software architectures are needed for high-speed decoding.

- Work in the past several years on developing quantization duals of Euclidean-space coding techniques (source coding), and vice versa, has been promising. Further development of such dual techniques should enrich both fields.

- Closer ties between synchronization, equalization and coding are needed. As codes improve, synchronization and equalization must be maintained in the presence of more severe errors.

Information Theory

Information theory establishes fundamental limits on the performance of communication systems. In particular, entropy provides an achievable lower bound on noiseless data compression, and channel capacity provides an achievable upper bound on reliable data transmission. All data compression schemes and all coding, modulation, and transmission schemes are bounded by these limits.

The insights gained from the evolution of this theoretical work now thoroughly permeate the design of point-to-point communication systems. In particular, the principles derived from information theory lie at the heart of the practical design and implementation of modern data compression, coding, modulation, and detection. Recently, information theoretic principles have been generating new approaches to magnetic and optical data storage. In addition, information theory has made central contributions to cryptography, public key cryptosystems, computer science, statistics, and pattern recognition.

Multiaccess information theory is by now relatively well established as the appropriate conceptual tool to study wireless systems in which multiple sources share a physical medium to access a single receiver. For any given multiaccess channel, one can compute the capacity region, i.e., the sets of source rates for which all sources can communicate reliably with the receiver. One can also calculate achievable error probabilities, as a function of coding constraint length, for such systems.

For more general networks of sources and receivers, under different types of feedback conditions, power constraints, and source cooperation, there is an impressive collection of theoretical results, but, so far, little cohesive theory and insight about applications exist.

There are a number of areas in information theory that appear to be ripe for increased coupling to applications. Two of these, compression and coded modulation, are treated elsewhere in this report. Others areas include the following:

- Code division multiple access (CDMA or spread spectrum) for cellular radio has evolved rapidly in recent years and is appropriately modeled in terms of multiaccess information theory. Information theory is currently guiding work on multiuser detection, coding, rate allocation, and power control. Much more work is needed on modelling the time varying multipath channels, and on developing multiuser decoding algorithms. It is not yet clear which theoretical results have practical utility, and the theory itself is still in a rudimentary phase.

- Coding and decoding for channels with unknown probabilistic behavior is a problem that is relatively well understood theoretically, but there is a need for applications of this to practice.

There are also many areas where further theoretical development is necessary. A few of these are as follows:

- There is a need for a more cohesive information theory of networks. This involves better understanding of broadcast channels, cooperation, feedback, and interference.

- Formulating information theoretic principles for networks with bursty sources has long been a perplexing problem which needs progress before information theory can truly provide a theoretical basis for data networks.

- Many relations between information theory and estimation, classification, and statistics have been developed over the years, but it appears that these relations only scratch the surface.

The above research areas are merely suggestions. History suggests that the most fundamental and important contributions to information theory come from researchers following their own insights about problems of inherent and general interest.

Communications Signal Processing

Communications signal processing covers the theory and algorithms for the recovery of a stream of data from a waveform that may be deeply buried in noise and interference and may be subject to a variety of degradations and impairments. Although the subject of communications signal processing is closely related to other topics in signal processing, it is a highly specialized topic with many subtopics. This class of signal processing algorithms is in wide use both in communication systems and, in very similar form, in magnetic and optical storage systems. A communications receiver may have limited knowledge of the transmitted waveform, possibly it must learn even the modulation format only from the received signal. Equalization is necessary in almost all receivers because of channel dispersion.

Closely related to equalization is the problem of bit synchronization. Very powerful codes for error control are now well known that allow for reliable communication in the presence of very noisy and degraded channels. It is important that equalization, bit synchronization, and block synchronization techniques be developed to a similar level of noise performance.

Algorithms for channel equalization range from long-standing and straightforward techniques such as the zero-forcing equalizer to more recent techniques such as adaptive equalization, either blind or trained. These are closely related to adaptive filters and may be based on least-mean-squared (LMS) adaptation. However, VLSI technology now provides special purpose digital signal processing chips and custom signal processing chips that can implement sophisticated algorithms. Blind equalization techniques, which deduce an equalizing filter from a received unknown data sequence, are necessary for receivers that may dial into a broadcast waveform in progress.

A variation of the equalization problem is the problem of echo cancellation which arises in problems of heavy multipath such as wideband communication in an urban environment. Equalization may also be partially combined with modulation by means of the modern ideas of prefiltering in the transmitter to anticipate dispersion in the channel. This method protects against the unnecessary waste of signal power that would result from an equalization filter in the receiver. The notions of prefiltering and of equalization may also interact with modulation in another way; this through the ideas of Nyquist pulses, partial response signaling, and intersymbol interference suppression. Maximum-likelihood methods are now in vogue for combining the ideas of data demodulation and equalization, and possibly also decoding. The Viterbi algorithm is the algorithm of choice for implementing the maximum-likelihood demodulator in the presence of all of these effects provided the number of trellis states is not too high. Thus there is an interaction between performance at low signal power on a dispersive channel and the complexity of an affordable Viterbi demodulator.

Signal processing is also important for communication over fading channels, spread spectrum channels, diversity combining channels, and interference channels. Demodulators for these channels may employ nonlinear combining techniques. Waveforms for CDMA applications may require near/far suppression techniques or transmit power control.

Finally, communication systems may employ space diversity. Multiple spatial transducers, such as antennas or acoustic devices, may be trained to create a multitude of individually steered, virtual point-to-point channels.

Protocol Theory, Design, and Engineering

Today's computer network architectures and protocols are inadequate for tomorrow's needs, such as gigabit per second communication paths using high performance fiber-optic links and switching technology. The design and engineering of high performance networks should be well founded upon theory. Towards this goal, we need sound and effective techniques for the design, specification, analysis, implementation, testing, maintenance, and modification of network architectures and protocols. Recent advances in protocol theory are beginning to provide understanding into many facets of protocol behavior and interaction. As the demands for more effective network communication increase, research on protocol theory must continue so that we will have a strong foundation upon which to base future designs. This is especially important when protocols are to be designed and engineered to accommodate new technologies (e.g., wireless access and mobility management) and application requirements (e.g., delivery of variable-bit-rate compressed video for teleconferencing and video servers).

A network, and even more so an internetwork, is necessarily the composition of a large number of protocols. The communication services offered by the network are the result of the interaction of its many protocols which interact with each other through interfaces. To meet the objectives of network architectures, it is important to understand how these interfaces should be defined, specified, and satisfied. The individual protocols should be designed for reuse, portability, efficiency, modification, and maintainability. Furthermore, to build reliable networks, it is important not only to prove the correctness of individual protocols, but also to prove the correctness of the entire collection of interacting protocols. A complete understanding of protocol interaction and interfaces is essential to the development of a sound theoretical basis for composing protocols in the construction of computer networks. We need to understand how to design and manage such complex software structures, in which protocol components are typically designed, implemented, modified, and maintained by different groups of individuals. Most protocols are designed to perform multiple functions and provide multiple services. This is done for efficiency but, in many cases, results in very complex protocols. Usually very little can be proved about their logical behavior. Methods are needed to synthesize multifunction protocols from relatively simple ones that implement individual functions. Specifically, techniques to add new functions to a protocol, without affecting its original functions or correctness, would be extremely useful. Conversely, methods to remove specific functions and unnecessary code from an existing protocol to make it more efficient (lightweight), without affecting the remaining functions, would let us tailor protocols to specific operating environments. In this way protocols could ``adapt'' to changes in underlying switching and transmission technology. The development of such methods would facilitate the structured implementation of protocols, improve understanding of interoperability between different protocol implementations with similar functional components, and address issues of portability of protocols across implementation platforms.

Formal models are therefore needed that provide a clear understanding of the relevant theoretical concepts of protocol refinement, projection and conversion, as well as interface semantics and protocol composition. Research advances are needed in protocol specification notations, and semantic models of protocol behavior. Research is also needed to develop from these notations and models, protocol specification, testing, and verification methods which can be used in protocol design and engineering tools. We envision a future in which network protocols, having been formally verified to work as intended, are cataloged and stored in a library. Their interfaces would be formally specified in such a way that their source codes are portable, reusable and easily modifiable.

New high performance fiber optic channels and communication switches are forcing us to rethink the architecture and structure of computer networks. Perhaps the old layering paradigm is inappropriate and no longer compatible with these new high performance networks. As the price/performance ratio of processors continues to decrease we see more processor capability in front-end controllers and increasing use of coprocessors. It is necessary to rethink the decomposition of functions within computer systems and among their processing elements. Our rethinking should be coupled with foundational developments in protocol theory and formal models. We need to understand the interaction among protocols and the resultant engineering implementation structures. We should also encourage innovative network architecture and protocol designs. Appropriate metrics for evaluating and comparing protocols as well as new models for analyzing the performance characteristics of these architectures and protocols are needed.

Dynamic Network Control

Much research is needed to determine how to configure and control large high-speed networks of the future. In particular, flow control, admission control, and routing algorithms need to be developed. New network control techniques are needed to support new applications with diverse communication requirements, and to exploit new technology for data links and switches. The techniques should be dynamic, adapting the network operating mode in response to changes in network resources or demand.

Future networks approaching gigabit transmission speeds are a driving force for much of the research on dynamic network control. However a largely different set of dynamic network control techniques arise in other contexts, such as in support of ground or satellite-based mobile information networks, and in support of personal communication systems.

New challenges are posed by the increased ratio of propagation delay to bit duration. A coast-to-coast gigabit link contains 15 megabits in transit. Predictive rate control, based on modern automatic control concepts, needs to be developed. At least in the short run, the increase in network speed will also cause a large variation in data-rate requirements. A small number of high-speed sources can generate extremely bursty traffic loads.

High-speed applications of a network are likely to pose stringent requirements, including a need for quality of service (QoS) guarantees on attributes such as end-to-end delay, throughput, support of bursty traffic, and reliability. "Best-effort" delivery by networks is not adequate for some applications, so provisions for negotiating service must be considered. Research directed towards basic control methods should be conducted. Some research may be tied to particular new transmission formats, such as asynchronous transfer mode (ATM). Moreover, some mechanisms for QoS provisioning in the ATM development serve as useful examples. However, a better understanding of the network control problem is also needed for more heterogeneous environments, with a mix of virtual circuit and datagram traffic, and with the concatenation of multiple resource control systems, as envisioned for example in the Open Data Network[1]. Research is needed in the whole spectrum of switching techniques, from datagram packet switching to circuit switching, including a wide variety of virtual-circuit methods.

A closely related issue is the dynamic establishment and management of special purpose virtual networks over an internet. Execution of applications over such virtual internets may consume substantial network resources, especially bandwidth. In addition, transport of the traffic for these applications may be given priority over "best effort" traffic. Specific considerations include protocols for specification of resource requirements of virtual networks, algorithms that access the admissibility of the virtual network under the network's current operating load, algorithms that establish the virtual network connection, and algorithms that manage resources allocated to the virtual network during its lifetime. At issue are the impact of the existence of the virtual network on best-effort network traffic and the performance of the virtual network itself as perceived by the applications.

Cost and complexity considerations may dictate the use of high-speed switches that occasionally drop packets or block circuits in the face of high congestion. Traffic modeling, dynamic traffic control, and network sizing techniques are needed to ensure satisfactory end-to-end performance in the face of possible packet loss.

Research on network control techniques should be driven in large measure by an integrated system viewpoint. First, the control techniques should be designed and ultimately assessed for use in a (possibly large) network, not just for use on a single link or connection. Secondly, the control techniques should be suitable for existing or anticipated technology, to work in conjunction with a complete set of protocols. Implementation requirements, including

---

[1] Realizing the Information Future, National Research Council, 1994.

communication and computational demands, should be assessed, whether the techniques are distributed or centralized. Because of the difficulty in implemerting high speed networks, it is unreasonable to implement every control technique that is evaluated. It is thus necessary to improve the methodologies for applying such techniques. However, analyzing implementation requirements, particularly computational demands, without constructing networks frequently results in inaccurate results that overlook important aspects of the problems. Therefore, to the extent possible, analytic work on high speed networks should be performed in conjunction with physical experiments that verify whether or not the approach is reasonable.

Internetworking

An internetwork consists of a set of host computers whi ι share a lingua franca, or common language, for the exchange of information. This common language, which might be called a bearer service, is mapped onto a set of lower layer transport networks via a set of mappings from the bearer service onto a variety of specific networks. An internetwork is created by concatenation of a set of transport networks, and translation devices, which then enable end to end provision of the bearer service to hosts connected to the internetwork. The bearer service must then in turn support a set of applications which communicate a wide variety of information via the bearer service. A set of mappings from applications to the bearer service are required.

Research in internetworking can then be classified in terms of this model as addressing one or more of the following issues:

- Fundamental internetwork architectural questions deal with the model itself. These relate to issues such as whether there should be one or more than one common bearer services, at what protocol layers should the bearer services exist, how evolvable may a bearer service be, how does one communicate between networks built around fundamentally different bearer services, etc.

- There are research questions dealing with the nature of a specific bearer service. Within the internet community today, the existing bearer service, IP, is being changed in the context of the IP-next generation work. Concepts such as larger address spaces, controllable QoS, and provider selection are being incorporated into the next generation IP. Research is required to understand what set of characteristics a bearer service should possess, and what the limitations of any given set of characteristics might be. Evolvable and non-uniform bearer services are also issues.

- The mapping of bearer services onto specific transport networks is an area of research. This area includes not only mapping onto new and unusual local and wide area networks, but also the mappings necessary within the end systems. Thus, host interfacing issues such as performance and scalability, as well as issues surrounding the translation of the bearer service onto specific transport networks, and the requirements the bearer service may put upon various transport networks are issues. These mappings must be robust enough that all possible combinations of transport networks can still be concatenated into a functional end-to-end bearer service.

- Mapping of applications onto bearer services raises research issues about the applicability of a given bearer service. Today's IP protocol provides an unreliable datagram service which is not able to support real-time services such as voice when the network is congested. This has motivated a body of work which attempts to extend the bearer service to include a controllable quality of service. Once one has this property, the way in which applications map onto the service must be extended to allow the application to specify the service it desires. These type of mappings may be called application to bearer service signaling issues. Signaling research may be concerned with issues of internetwork service management, including naming, addressing, accounting, billing, authentication and access control, resource discovery and management, and performance/QoS management, in addition to the recognized need for common internetwork transport protocols/interfaces.

- Current internetworks often lump together the transport of information (i.e. packet forwarding engines) with the control of this information (i.e. routing protocols which load the address tables of the forwarding engines). It may be useful to consider a more structured division between the low-level information transportation and its control. Signalling then communicates information from the application to the bearer service controllers, while a control interface is used to communicate information between bearer service controllers and the information transport engines. Research into these topics may enable more flexible and extensible implementations of internetworks with complex or evolvable bearer services.

- As internets evolve to support billions of users, scalability of nearly all components of the internetwork become an issue. Research on scaling of routing protocols is of particular near-term interest. Also, procedures once done manually or semi-automatically such as fault isolation, cost accounting, and others require high levels of automation to lower the operational costs. These, and many other issues of scale deserve increasing attention.

Optical Networks

Optical fiber has emerged as the medium of choice for point-to-point transmission systems because the low-loss, low-dispersion properties of single-mode fiber allow transmission of information at much higher rates and over much greater unrepeatered distances than does copper wiring of any form. However, the fundamental architecture of the nationwide telecommunications infrastructure has remained essentially that which evolved during the pre-photonic era; point-to-point transmission systems interconnecting multiplexing / demultiplexing equipment and digital switches, with fiber simply displacing earlier technologies as the physical transmission medium. The advances in transmission speeds not withstanding, the enormous transmission capacity of the fiber remains largely under-utilized.

Research in optical networks is aimed at answering the question of how best to take advantage of the bandwidth and other properties of the optical fiber to enhance the capacity and flexibility of networks. The optical physical layer may, in addition to bandwidth, provide a degree of transparency and flexibility which might result in an optical connectivity layer which is independent of the fiber's physical interconnection patterns, transparent to the data being transported, and rearrangeable to meet changing traffic and service requirements. The essential research issue is to understand the ways in which such an optical connectivity layer can be used to advantage and the limitations to its use. There are two principal modes of accessing this bandwidth; wavelength division multiplexing (WDM) and very high-speed time division multiplexing (TDM).

The cornerstone and key technology that enables optical networks is the wide-band fiber amplifier, which makes possible transmission spans between generation and detection far exceeding systems having repeaters, and has a broadband gain over a spectrum of about 30 nm, or 4 THz, centered around the low-loss region of the fiber transmission. This broad bandwidth, together with the very low-noise amplification of the fiber amplifier, have brought about many proposals for carrying multiwavelength optical signals end-to-end all-optically, with signals being routed by passive components that redirect signals according to their wavelengths. For TDM based networks, optical signals may similarly be directed by processing of routing headers in the time domain. There are many research issues that need exploration before such systems can become practical realities. A number of these research issues are outlined in the following paragraphs.

a. Fundamentals

The are a number of fundamental issues that demand the attention of the research community. Among these issues are the following:

- Research on the fundamental aspects of multiwavelength and time-domain all-optical networks including studies of:

    - limits to numbers of users
    - limits to geographical size
    - limits to network capacity
    - numbers of wavelengths and switching elements required
    - architectures which take maximum advantage of the properties of, while overcoming the limitations of, the physical layer.

- Demonstrations of new network concepts at the local, metropolitan, and wide-area level

- Fundamental limits to physical layer scalability including the following studies:

  - component crosstalk
  - fiber nonlinearities
  - signal to noise degradation in amplified distributed networks
  - multiwavelength amplifier operation
  - optical power regulation throughout a distributed network
  - optical wavelength or frequency registration and regulation

b. Transparency

An important issue is the need for and the role of transparency in optical networks. Questions are: What are the benefits and what are the costs? What are the most severe limitations of transparency and what limits to network application and performance do they impose? Research issues include the following:

- limits to transparency
- applications of transparency
- the role for electronic switching in optical networks

c. Network architectures

The following research agenda appears appropriate:

- Research on network architectures to explore the relationship between connectivity and network performance, and to understand the potential advantages of network rearrangeability at the optical layer

- Studies of the cost effectiveness of optical networks with regard to various applications, and the development of network economic models and tools

- Studies of the potential improvement of network survivability, and methods of implementation

- Research on the distributed network control algorithms to effect wavelength assignment, to perform reconfiguration management, routing, traffic control, signaling and call setup, and to understand the needs and means for scalable network control algorithms in reconfigurable optical networks

- Research on network management and operations in a transparent, reconfigurable network and studies of the interaction of multiple network management schemes on a reconfigurable transparent optical network

- Studies of the architectural role for wavelength translation, and its impact on capacity, connectivity, network scale, and network performance. Is all-optical connectivity required?

- Research to develop an understanding of the role of hierarchy in optical network architectures, and its impact on network scalability and survivability

- Research on architectures and technologies for low-cost optical network access - important questions are: What is the appropriate interface to all-optical networks? How far towards the end user should all-optical networks extend, and what means for accessing the network is most effective? Are the advantages of optical routing maintained if high-speed network access is done electronically?

- Studies of evolutionary strategy: How are the functions of transmission and switching best combined in optical networks, and what are the trade-offs between multiwavelength and multifiber designs? What are the considerations in making an optical network upgradeable to advancing technology such as all-optical switching, for example?

- Studies of optical interconnection networks for high-speed switching fabrics to produce either very high capacity switches or very high switch-interface speeds

In very high-speed time domain networks, research issues include means for doing all-optical regeneration, switching, and routing, and the optimal use of such techniques in networks.

The overall field of optical networking has reached the point where the first enthusiasm of infinite bandwidth available at zero cost on a global basis has begun to run into severe practical difficulties. For example, the notion of end-to-end optical transparency will be difficult to achieve. Overcoming one of these difficulties will require fundamental advances in technologies, while others may be solved by new network designs. All of them will impact that which can be achieved and the extent to which optical networks will become useful as a part of the broadband network of the future. The general research questions to be posed at this time relate to finding the limits of such networks and technologies as imposed by real-world technology and architectural constraints.

Security

A great deal of research has been directed toward techniques for encrypting data to ensure confidentiality; however, networks and communications remain vulnerable. To counter malicious attacks effectively and provide secure services, a coherent security architecture is needed. It is envisioned that such an architecture would include the following components, among others:

- encryption/decryption techniques to ensure data confidentiality, data integrity, and origin authenticity

- techniques and protocols for access control (authentication, authorization, firewalls), ensuring anonymity, and countering traffic analysis and denial of service attacks

- cryptographic protocols to support financial transactions

- electronic money to facilitate the marketing of services

The emphasis here is on open networks with many domains (rather than the security kernel of a host node or military networks). In this environment, different domains typically have their own administrative authorities, different standards of security, and different security mechanisms and protocols (trusted servers, authentication protocols, authorization methods). Thus any overall security architecture should be designed to accommodate heterogeneity.

As more and more hosts are networked together, it is unreasonable to expect that they will all have the same high standards of security. Once an intruder has violated a network at a weak point, it can possibly violate the seemingly more secure parts of the network. To guarantee that the effects of any security violation will not spread to other parts of the network, research is needed into protocols and network architectures for secure information exchange; in particular, protocols for access control functions such as authentication, authorization, and authority delegation are important.

Existing techniques for authorization (e.g., protection bits, access control lists) are highly implementation dependent. Research is needed into new authorization approaches that allow convenient composition of authorization requirements specified by a variety of authorities, or for different authority domains. The formal semantics of such composition must be well understood. The design of authorization servers that off-load the authorization function from network services should be investigated.

Smart cards are now available with which human users can carry out the computations necessary to authenticate hosts. New protocols for authentication (user-host, peer to peer, interdomain), key distribution, and authority delegation are needed. The formal semantics of such protocols must be well understood. In particular, the formal meaning of a network or protocol being "secure" is still an open question. The consequences of security violations are so severe that we should at least verify that protocols for secure information exchange can survive known techniques of attack. In order to deal with the complexity of networks, verification methods founded upon well-defined semantic models are needed.

Reliable Networks

As the role of the communication system in society has grown, user expectations for reliable and continuous operation have out-paced our ability to meet those expectations. Successful realization of a National Information Infrastructure will require a better fundamental understanding of the sources of failure in networks and more systematic approaches to combatting failures.

The sources of network failures can be grouped into three categories: physical device failures, errors in system design and mistakes on the part of users or network operations staff. Significant improvements are needed on all three fronts if the emerging information infrastructure is to meet the demands that will be placed upon it.

The use of redundancy to detect hardware failures and make systems fault-tolerant has been used successfully in communication systems for many years. However, we still lack systematic methods for evaluating alternative redundancy, fault recovery and repair strategies, making it difficult for system designers to make intelligent choices. The trend toward higher speed networks can be expected to lead to lower intrinsic device reliability, making these issues more critical than in the recent past. Intermittent hardware failures are the most problematical in practice and the least understood.

Many of the more spectacular failures of communication systems have been by design errors, usually in system software. Ironically, software intended to improve reliability tends to be the most error-prone, since it must guard against events that are difficult to anticipate. There is a clear need for more systematic approaches to this class of problems.

The increasing sophistication of distributed network control algorithms can lead to greater opportunities for catastrophic failures. We need more effective methods for reasoning about distributed algorithms and simulating them under realistic conditions. Research is also needed on algorithms that are self-correcting, to minimize the potential for errors to propagate.

Even the best trained operations staff are prone to making mistakes that lead to system failures. General techniques are needed that minimize the need for staff interaction with network equipment, and which minimize the opportunities for errors when intervention is required.

Switching Systems

Effective implementation of a National Information Infrastructure extending to individual homes and businesses will require switching systems with tens of thousands of ports operating at gigabit speeds and supporting a wide range of applications, from low speed telemetry and control to high resolution images and multichannel HDTV. These systems will need to support a continuum of channel bandwidths, bursty traffic streams and highly dynamic multipoint channels. It must be possible to cost-effectively engineer them so that users' requests for new channels seldom or never block and so that end-to-end quality of service guarantees can be provided. They should have a flexible control architecture that separates the higher level functionality from particular implementations, so as to facilitate interoperability across multiple switching platforms and evolution over successive generations of technology.

Realization of these goals require significant new advances in switching system theory and design. To date, there are no practical architectures for nonblocking multipoint virtual circuit switches that can meet the theoretical limits on optimal scaling with respect to all the characteristics of practical concern (switching network complexity, routing memory requirements, connection modification cost, delay), and most systems now being used have poor scaling properties.

Our understanding of blocking in switching networks remains incomplete. Effective methods for analysis of blocking probability in multipoint switching systems are only now being formulated. Queueing analyses of large switching networks have focused on steady-state behavior for simple traffic arrival models. Engineering of systems for NII applications will require a better understanding of the transient behavior of large systems in the context of non-stationary and highly bursty traffic. Both simulation and analytical approaches are needed to fully explore these issues.

Recent advances in optical device technology have created new opportunities for novel switching architectures. While the classical tools of switching system theory can be applied to such systems, new issues, such as cross-over minimization and wavelenth re-use require extension of the standard models.

In the last decade, the research community has witnessed a wide variety of proposed switching system architectures, but little definitive evaluation or comparisons. It is time for the community to create evaluation criteria that factor in both performance and cost measures, where technology-dependent and technology-independent factors are clearly separated.

Control architectures for large switching systems have yet to receive adequate attention from the research community. Architectures that can support a variety of switching platforms, can support very high rates of connection configuration, and in which the control processing capacity can be engineered separately from the switching system capacity are clearly needed for NII applications, and equally clearly, are well beyond the current state of the art.

Wireless Networks and Access

Wireless local area data networks are needed to provide high data rates (> 10 MB/s) to low speed and stationary terminals within campus and building environments, and wireless personal communications systems (microcellular systems) are needed to provide voice and moderate data rates (~ 1 MB/s) to moving terminals within all populated areas. These different wireless access applications require different optimizations in the integration into efficient wireless access systems of the large numbers of complex and diverse techniques that make up these systems. These systems need to cope with the contention and mutual cochannel interference among multiple randomly positioned terminals that share frequency spectrum, and access multiple spatially separated access points to fixed networks. The techniques must work together in harmony to maximize spectrum utilization (overall system capacity) and system performance in the time-varying frequency-selective multipath fading environment while minimizing power consumption in battery operated portable terminals.

Fundamental research to optimize these systems is needed on the following topics:

- Combined multi-carrier and coding to mitigate delay spread. Effective implementations of these techniques for time-invariant wireline transmissions need to be extended and advanced for effective use in the multiple-access time-varying wireless environment.

- Interference nulling techniques for increasing the reuse of frequencies and to increase system capacity and improve performance.

- Algorithms and criteria used to determine when handoff from one access point to another is required. Pattern classification and recognition techniques should be applied to identify local signal characteristics to aid in handoff execution.

- Low complexity signal processing algorithms and implementations are needed to reduce power consumption in battery powered terminals.

- Assessments of system capacity and performance of competing access technologies are needed in the time-varying dispersive multipath environment. In order to be useful, these assessments of FDMA, TDMA and CDMA must be done for the same environmental conditions, and must include, for each, the same capacity enhancing techniques, e.g., power control, dynamic channel assignment, optimum handoff algorithms, diversity, and interference nulling.

- Research on new techniques for improving the performance and capacity of wireless access systems should be supported.

- Measurements and modeling of radio propagation between a terminal location and multiple base station locations for different propagation environments.

Large integrated networks that support wireless access must include capabilities not required to support fixed point-to-point or point-to-multipoint communications. Research is needed in large network protocols, signaling, data based management and routing algorithms to provide the following:

- Management of large scale mobility. This includes recognition of the location of mobile terminals away from home, registration of the current location with the network, and efficient routing from data initiation locations to new away from home locations.

- Management of small scale mobility. This requires accommodation of handoff from one network access point to another as terminals move and efficient routing to the new attachment point.

- Accommodation of the wide range of variability of wireless circuits. These circuits are often good (low error rates), but occasionally become bad (bursty errors) with terminal motion or changes in the environment, and may disappear as a terminal moves out of range, or interference becomes excessive. Network intelligence must cope with these realities of wireless access, since mobility is an important feature.

- Security and privacy of the inherently not-secure wireless link. Provisions for encryption of data, security of identification and authentication, and resistance to unauthorized users must be provided in wireless network protocols and signaling.

Networks supporting wireless access must accommodate different wireless technologies with different mobility characteristics including: a) cellular vehicular mobile and high tier PCS (terrestrial and satellite) providing low data rates (<10 kB/s) and high-speed, wide-ranging mobility, b) low-tier personal communication systems (low-power access) providing moderate data rates (~ 1 MB/s) and mobility throughout populated areas and c) wireless local area data networks providing high data rates (>10 MB/s) with limited mobility in buildings and campus environments.

Data Storage Systems

For many years, the growth in linear and areal bit densities in magnetic and optical data recording devices showed a steady but predictable increase. These density increases were largely due to scaling (e.g., lowering the flying height of the head in a magnetic hard disk drive) but keeping the signal processing the same. Within the last year or two, however, this rate of growth in density has shown a dramatic increase. This change in the slope of the growth curves has at least partially been due to the introduction of modern techniques of communications signal processing to these products. An example is the introduction of a so called PRML (partial response, maximum likelihood) system whereby instead of trying to eliminate intersymbol interference by linear equalization (with the concomitant noise enhancement), one accepts but shapes the intersymbol interference into the form of a Class IV partial response channel and then uses Viterbi detection to search for the most likely transmitted data pattern.

The applications of the principles of modern communications to magnetic and optical storage devices is still in its infancy and a multitude of challenging research topics exist in this area. Although research in this area has shown a steady growth, the breadth and depth of research topics to be solved could support a larger number of investigations.

Although magnetic and optical storage systems share many similarities with the more classical communication channels, there are sufficient differences from the classical channel models. For example, in a high density thin film magnetic disk, it is observed in the laboratory that magnetic transitions (which carry the information) tend to partially destroy each other. Furthermore one type of noise (due to the zig-zag nature of these transitions) is only present where transitions exist and thus should not be considered as additive and independent of the signal. Thus, new channel models must be obtained from experimental investigations.

Since classical channel models do not apply, neither do the classical modulation, coding and signal processing solutions. For example, magnetic and optical media usually confine their write waveforms to take on only a small number of values (two, for saturation recording in magnetic media). This precludes the use of the usual bandwidth efficient modulation techniques such as QAM. Yet, we know that these channels should be able to support signals carrying many information bits per unit bandwidth. How to achieve this remains an unsolved problem.

Furthermore storage surfaces are fundamentally multi-dimensional. The idea of a track is an artificial constraint that introduces new problems of tracking and track to track interference.

In addition to the usual implementation constraints, recording channels have some severe additional ones. In order to record and retrieve at rates to support today's high speed computers or networks, one needs to read and write at extremely high data rates (which stresses the capability of the read-write electronics). Furthermore, many storage devices are intended for portable use where power consumption is a severe requirement. Finally, there is the ever present cost pressures which are particularly severe in commodity products.

With the coming capability to transport enormous quantities of information at ever increasing speeds, there will be a corresponding growth in the capacity to store this information. Special purpose storage devices will be needed for particular types of information (e.g., video). Higher input-output speeds will be required to accept data from this information superhighway. Furthermore, it is to be expected that radically different new storage systems will evolve (e.g., holographic memories).

The storage industry is one of the most successful industries in the United States. Although the U.S. is still the dominant force in this very lucrative field, its dominance is under constant threat. It is essential that we aggressively pursue research in this area.

Video, Speech, Image and Data Compression

With the expected onslaught of multi-media traffic on the national data superhighways, the need for the efficient digital transmission of video, speech, images, and data becomes ever more important. Although other NSF programs treat both speech and video compression techniques, the proper choice and marriage of these techniques to the peculiarities of data networks need special attention.

Data compression, for voice, images, video, and data, are old and well established disciplines within information theory, and at the same time are applied to constantly shifting application areas. One might think that the need for compression should disappear with the decreasing cost per bit of communication and storage, but in fact commercial interest is growing. One reason for this is that the cost of computation in compression algorithms is decreasing at about the same rate as communication and storage costs are decreasing. Another reason is that communication capacity is likely to remain limited in wireless communication systems such as cellular radio, personal communication, emergency and military communication systems, and broadcast video. Yet another reason is that one can temporarily avoid adding additional storage to personal computers through data compression. A final reason, for voice, images, and video, is that one can reduce congestion on packet networks by throwing away less important data. The segregation of important from less important data is best done by compression techniques.

Compression can be separated into lossless and lossy compression. In lossless compression, the original data must be exactly retrievable from the compressed form; this is usually required for digital data. In lossy compression, the original data need only be retrieved within some allowable distortion; this is the case with voice, video, and images.

If speech is to be transmitted over a digital communication system, it is necessary to convert the speech into a digital format before transmission. Many methods are available for digitizing speech, and it is desirable to achieve high efficiency while satisfying the user requirements and the communication system constraints. User requirements may be concerned with intelligibility, speaker recognition, and natural voice quality. Some communication systems require a constant data rate into the digital system. Other communication systems benefit from a variable data rate format. For example, CDMA wireless systems take advantage of the voice activity factor (i.e., the fact that user are not speaking 100% of the time.) to decrease interference and increase the maximum number of simultaneous users. ATM networks take similar advantage of variable data rates.

It is to be expected that digitized video will provide a large percentage of the traffic in future high speed integrated networks. The public's present addiction to broadcast video will be enhanced by increasing the number of available channels and the introduction of interactive video. The worldwide adoption of digital standards for higher definition TV is expected and electronic cinema is also on the horizon.

A variety of schemes have been proposed for achieving data compression for video. Both intraframe and interframe coding schemes have been devised and built. Presently, real time compression systems based upon variations of the two dimensional cosine transform exist in hardware that can achieve almost perfect reproduction of video with less than one bit per pixel. Much higher compression ratios have been promised by other techniques.

Image compression and video compression are closely related, of course, and advances in one can usually be applied to the other. It is important in both areas to develop better compression techniques in which important data can be segregated from less important data in such a way that reasonable image quality can be retrieved in the absence of the less important data. This is important in integrated networks where congestion might require eliminating the less important data. It is also important for browsing, where one needs to rapidly acquire a large number of less detailed images, and then focus more closely on very high quality images.

Although much progress has been made in both video and speech digitization, many fundamental questions still remain. New and more efficient means for compressing both speech and video are sought. High priority should be placed on the synthesis and analysis of compression systems that match the requirements of the new emerging communication technologies such as digital cellular, PCN and high speed data networks.

Lossless compression is a mature science, and the sliding window type of Lempel-Ziv algorithm and the adaptive dictionary type of Lempel-Ziv algorithm have become the workhorses of commercial lossless compression. These algorithms will adapt in principle to any source statistics with any kind of memory. In practice, these algorithms are less adaptive because of the fixed window or dictionary size. These limitations, however, allow the commercial algorithms to rapidly adjust to changing statistics. There is a need for a better understanding of adaptive data compression of sources whose statistics occasionally change (or, more generally, sources with multi time scale statistics). There is also need for a better understanding of the relationships between lossless compression and estimation.

Mobile and Nomadic Computing

The technological advances in hardware miniaturization (e.g., portable computers) coupled with the advances in wireless technology have led to the introduction of new computing paradigms - the Nomadic and Mobile Computing. Nomadic Computing refers to the ability of a user to compute, communicate and access information regardless and independent of the user's location; we refer to this feature as "personal mobility." Mobile Computing introduces the element of "terminal mobility" into the working environment; i.e., a user's terminal may be moved while engaged in a communication session or computation processing. Combined, Mobile and Nomadic Computing allow computing and communication from anywhere and at anytime. A mobile and nomadic user may need to communicate with other users, to download files, or to obtain some other network services while accessing the network intermittently, from different locations, and sometimes while on the move. In order to communicate, a mobile terminal requires tetherless access to local-, metropolitan-, and wide-area networks. Thus, Mobile Computing can be implemented through the wireless technology, allowing the users to maintain constant connectivity while in the area covered by the wireless communication. A nomadic user connects (possibly infrequently) to a fixed network and may use different network attachment points and different terminal equipment. The challenge in implementing a Mobile and Nomadic Computing environment is to reduce the effects of personal and terminal mobility on supported services and applications. In other words and for example, the performance of applications running on a mobile platform should only be minimally affected by the terminal mobility. Similarly, nomadic user interfaces should be preserved when subject to personal mobility.

Some of the networking issues falling under the subject of Mobile and Nomadic Computing are:

- effects of the limitations of portable hardware and the effects of mobility on the performance of applications; e.g., how the limited computing power and limited throughput of portable machines, or the frequent changes of associations between mobiles and the fixed network, affect the performance of applications and what needs to be done to reduce (or hide) these effects.

- design of applications for portable and mobile hardware

- novel applications and services for the mobile environment (e.g., location-dependent services)

- effects of mobility on communication protocols

- design of networking protocols for mobile networks (e.g., Mobile IP)

- consistency of data bases and file systems in the disconnected computing environment

- location management and user tracking algorithms for nomadic users

- seamless performance across heterogeneous networks

- preservation and transparency of user interfaces to the connection network and connection hardware (e.g., using different terminal types)

- security in the mobile environment

Some additional non-networking issues, related to the topic of Mobile and Nomadic Computing, are:

- hardware supporting the portable and mobile environment (e.g., portable computers, long-life and large-capacity batteries, low-power subsystems, hardware reliability)

- user interfaces for nomadic computing (e.g., pen-based interfaces, handwriting recognition, voice recognition)

- economical and social implications of Mobile and Nomadic Computing

Satellite Systems

More than a quarter of a century of research has been directed to the development of satellite communication systems. Advances in computer, data, and optical communications have, however, left satellite communications behind in terms of transmission speed, protocol compatibility, and time delay performance. Because of this a number of challenges lie ahead due to changes in satellite applications. A sample of these changes include the following:

- Satellite applications have evolved from point-to-point (trunk traffic) to point-to-multipoint (broadcasting), multipoint-to-point (data collection), and to multipoint-to-multipoint (personal communications).

- Satellite applications have evolved from fixed earth stations to moving earth stations (personal communications).

- In contrast to early satellite applications, a significant change is from a small number of satellite user networks to a very large number of satellite user networks.

From a scientific point of view, communication by satellite involves multi-disciplinary areas with many facets. These include the space segment, the ground segment, orbital dynamics, power sources and the efficient use of power, launch vehicle technology, transponder optimization, and the operational strategy of the system. Within the realm of communications and networking the following research topics are suggested:

- New innovative concepts, ideas, and applications of network theory, access strategies, channel coding techniques, modulation alternatives, techniques for equalization and synchronization are needed for application to satellite systems. Especially important are applications aimed at mobile satellite communication systems, personal satellite communication systems, and very small aperture terminals.

- Asynchronous Transfer Mode (ATM) has been recommended for multimedia, multi-link, and multi-purpose transmission applications. How can satellite networks be compatible with ATM?

- The GPS (Global Position System) is basically for location identification. What are new potential GPS applications?

- Without reference to any particular system, what are the basic underlying principles which impact the design of all low-earth-orbit and medium-earth-orbit satellite communication systems. Of particular importance are low-cost architectures, effective handover procedures, optimal coverage, performance improvement, and reliability enhancement.

- How should new space segments, new earth segments, and the new communication infrastructure be developed in order to meet the rapidly changing user environment?

As can be seen from the preceding list, the optimal design and implementation of a satellite communication system draws from many areas of communications and networking, and therefore profits from research in many of the priority areas previously discussed.

5. RESEARCH INITIATIVES

The following four projects were recommended by the workshop committee as research initiatives for the NSF. Unlike research projects resulting from proposals submitted by individuals or single institutions, the initiatives described in the following paragraphs will require larger funding and will require a significant level of coordination.

Research Initiative on the Open Data Network

There are many visions of the NII, ranging from 500 channels of TV on demand to an open, universal network marketplace of ideas, products and services. This section describes an initiative to explore one particular vision for the NII, the Open Data Network, or ODN, which emphasizes the open nature of the interfaces, and open access by service and network providers. The defining development of this vision can be found in a recent report from the National Research Council, which served as a motivation for this initiative proposal[1].

While there are a number of aspects to the ODN vision, the most central idea is that of a protocol layer called the technology independent bearer service, which is an abstract definition of the basic network service provided by the various technologies out of which the network is built. The bearer service is a low-level protocol layer; current examples include the IP layer of the Internet, or the 3 kHz circuit of the telephone system. The NRC report identifies the key features that such a bearer service should have; it should be an open interface, it should be implementable as a separately priced service, it should be as independent as possible of technology, it must

[1] Realizing the Information Future, Computer Science and Telecommunications Board, National Research Council, National Academy Press, Washington D.C., May, 1994, a study commissioned by the NSF.

support a range of quality of service (QoS), and it should include only those service aspects that cannot be exported to another layer.

The ODN vision of the NII is a broad one; it includes the services and objectives of the Internet community, the cable, the telephone and the entertainment industries. The precise technical details of an integrated bearer service that can serve all these communities is unproven. One effort to produce an advanced bearer service is within the Internet community, which is developing a next generation of the IP protocol, but it remains to be shown if this effort will be broad enough to meet the needs of this diverse set of players. This initiative is intended to provide a testbed for the development and evaluation of any bearer service proposal with this broad set of objectives.

To meet this set of objectives, a successful bearer service must have a range of technical features. For example, it will presumably have to support a range of QoS, it must provide a framework for dealing with congestion, it must provide building blocks for security, it must provide for change and evolution, and so on. This list is not meant to be definitive, but to suggest the range of issues that a suitable bearer service must address.

a. Core research

Because of the range of issues that the bearer service must address, successful definition of such a service must depend on research in a number of areas which were discussed earlier in the report. Most of the Research Priorities discussed in the previous section of this report impact in some way the architecture, implementation, and performance of a data network.

b. Architecture

Integration of all these issues into a unified proposal for a service is an effort in its own right, as can be seen by the level of effort in the development of the next generation of IP. Architectural efforts of this sort are very important, and should be separately recognized from the core research discussed above.

c. Testbed

The key objective of this initiative is to develop a rich enough experimental context, both in the technology options over which the bearer service is implemented and as well in the applications it in turn supports, so that the resulting experiments and demonstrations provide a valid assessment of the breadth of the proposed bearer service. The testbed could be realized as a collection of separate experiments, addressing different technologies and application areas, so long as they are tied together by an unified vision of a common bearer service architecture. The manner in which this unified vision can be articulated and imposed on a set of separate testbed components remains to be defined.

d. Technology in the testbed

For example, the range of technology below the bearer service might include ATM, very high speed networks, wireless, LANs, and emerging subscriber loop technology. These specific technologies are suggested both for their importance and because they have specific technical issues relevant the breadth of the bearer service. Wireless technology raises issues of the management of mobility within the network, intermittent connectivity and variable bandwidth, and the trade-off in function between the wireless technology and the bearer service. Very high speed networks raise issues of achieving real throughput, dealing with high delay, implementation of cost-effective, high performance forwarders, interfacing to end-nodes and so on. Subscriber loops raise special issues, as discussed below.

e. Applications in the testbed

Applications should be selected to illustrate the service requirements of the communities discussed above. Thus, they might include video on demand, interactive multi-media, telephony, and traditional best effort data access. Attachment to the testbed of high-performance video servers, for example, would at the same time demonstrate issues in storage servers, high speed networking, and effective integration of video into a general network, as well as the specifics of the bearer service. Attachment of other application experiments is a validation of the objective of openness, and should be facilitated. One component of the testbed should be the construction of a virtual network over the Internet to support the bearer service, a capability that would permit wide access to the testbed environment.

f. Objectives of the testbed

At a high level, the objective of the testbed must be to demonstrate real interoperation among the technologies and services. The key issues are generality of the bearer service, the cost-effective delivery of key application data (for example voice and video) and scalability of the resulting solution. Related issues such as security should also be demonstrated.

g. Routers for the bearer service

As a part of this testbed, the different technologies must be interconnected as is done with IP routers today. To allow experimentation, it would preferable if the interconnection and forwarding function could be realized in a changeable manner (e.g., software or programmable hardware), so that different proposals for bearer services could be explored on the testbed, and the proposals could evolve as experiments proceed. A key to experimentation at this level is flexible experimental apparatus, which must include the interconnection devices or routers.

h. Technology development

As a part of this testbed, new applications and new network technology could be developed and evaluated. A particular technology example at the present time is new equipment for subscriber loops. Current commercial proposals for a next generation subscriber loop seem to specialize on video delivery and telephony, but do not seem to recognize the potential value of a general service such as is provided by the Internet. Development and deployment of advanced subscriber loop technology would at the same time validate that concept as well as the ability of the proposed bearer service to integrate this technology into the overall network framework. The NRC report identifies a number of specific technical options that might be incorporated into an ODN subscriber loop, including a flexible means to allocate transmission capacity from the subscriber into the network, the integration of transport for specialized and for general sorts of data traffic, and the statistical aggregation of bursty traffic.

Research Initiative on Data Storage Architectures for the NII

Most data storage devices, data storage systems and data storage architectures have been designed for either stand-alone computers (e.g., for PC's or for workstations), for a cluster of computers (e.g., a file server for a cluster of stand-alone computers) or for computer centers (e.g., the disks and tape systems that take up all the space in a super computer center).

The enormous number of users that will be interconnected via the NII, the tremendous data rates at which traffic will flow over the NII, the mismatch of data rates which will occur where the NII interconnects lower bandwidth communication channels and the diversity of services that will be offered over the NII all will impose an entirely new set of requirements for data storage systems. If this important issue is not faced early on, the communications portion of the NII will be designed without properly taking into account the interplay between the storage and the communications network.

In the past, users satisfied their storage requirements either by owning sufficient storage to satisfy their needs or by sharing storage with others within some organization (such as with others at the same workplace). If the NII is to reach a majority of the households in the US, the former approach would be exceedingly wasteful and unduly expensive. The alternative is to create "sharehouses" of storage which would serve a community of users. These users would rent storage space as needed ("storage on demand"). There would probably be a basic monthly charge for this service plus a charge based upon usage. Indeed many users might share the same storage space if they were sharing common information but such a scenario would require the standardization of data formats. Lurking in the background are the issues of privacy, authentication, etc. Models of user storage requirements would have to be developed in order for the storage on demand provider to ensure a specified quality of service.

It is likely that buffering of information will be required to accommodate the high bandwidths of the NII and the lower bandwidths of the "last mile" into the household and/or mobile users connected by a wireless service. Here temporary storage will be required for the information at these interfaces, just as warehouses are used to temporarily store goods delivered by trucks which travel on our superhighways before they are delivered to consumers (or retailers who sell to consumers). The storage devices that interface with the largest bandwidth data channels will have to have the ability to record at much higher data rates than what is available in present storage systems. Indeed a hierarchy of temporary storage devices might be needed with the device closest to the high bandwidth network being a large capacity device with high bandwidth and where devices closer to the consumer would have less capacity and lower bandwidths.

The development of specialized storage devices for specialized applications is an important part of the NII. For example, video on demand will impose an entirely different set of requirements on storage devices than our present requirement for millisecond access for data to a computer. Alternatively, a general purpose, cheap, and easily deployable storage might be developed (say from microdisks) which could accommodate a wide variety of uses.

Although most of the above discussion is concerned with issues of storage architecture, equally important is the need for the continued development of storage systems with increased capacity, lower cost, decreased power, etc.

The following structure is suggested to launch this initiative. First, an advisory panel of experts which span all aspects of this problem should be convened to discuss and refine the scope of the initiative. Then a Call for Proposals should be put forth that encourages response from teams of investigators which are comprised of experts in both storage and data communication networks. Responses from both industry and universities should be sought. Since the problem needs to be attacked both from the standpoint of basic principles and experimental research, proposals for one or more testbeds should be sought.

Wireless Research Initiative

Two objectives of the Wireless Initiative testbed are to: a) configure and demonstrate a high performance wireless data system with technology prototypes that can provide reliable wireless access at rates greater than 10 MB/s to an open data network testbed, and b) exercise mobility management and wireless access capabilities of a prototype open data network testbed that has been built to be compatible with the wireless access system and technology prototype. The accomplishment of these objectives will require the installation in a testbed of several prototype wireless base stations attached to the open data network testbed in at least two geographically separated locations and construction of several portable wireless terminal prototypes that can be moved among base stations within a location and between locations. This is the minimum size experiment needed: a) to exercise the wireless access system and technology and both the small scale and large scale mobility management capabilities of the open data network, and b) to demonstrate their combined performance.

Current research activities (e.g., ARPA sponsored projects at UCLA and UCB) and commercial products aimed at wireless data networking have been hampered by the previously limited choices of radio spectrum available for their use. The focus of these current activities has been on use of Industrial, Scientific and Medical (ISM) frequency bands that restrict use to direct sequence or frequency hopping spread spectrum techniques, and that contain many sources of ISM interference. These techniques are not optimum for wireless transmission of bursty data at rates of 10 MB/s and greater. Also, the available spectrum bandwidths cannot support enough spreading to provide significant spreading gain for such high data rates. Short-burst data-transmissions are not compatible with the tight power control and additional degrees of freedom needing initial synchronizing that are required by high-capacity multiple base-station spread-spectrum. The FCC has recently allocated 20 MHz of frequency spectrum near 2 GHz for unlicensed wireless data systems that must be designed to operate without jamming each other. Also, for time invariant wireline applications, multi-carrier modulation and coding have been combined to provide high capacity and multipath mitigation for data transmission over those bandlimited channels. Research to extend these multipath mitigation techniques for application to time-varying multiple-access wireless multipath channels in the newly allocated frequency bands provides opportunities for advances in the performance and capabilities of high-speed wireless access to open data networks.

The challenge for wireless local area information networks is to integrate into an efficient wireless access system a large number of complex and diverse techniques. Many different techniques are required to cope with the contention and mutual co-channel interference among multiple randomly positioned terminals that share frequency spectrum and access multiple spatially separated access points (base stations) to the fixed network. The techniques must work together in harmony to maximize spectrum utilization efficiency (overall system capacity) and system performance in the time-varying frequency-selective multipath fading environment while minimizing power consumption in battery operated portable terminals. Global optimization of the techniques together is required; optimizing each technique individually does not result in a good overall system solution. Diverse techniques to be harmoniously integrated to accomplish these objectives include interference nulling, multipath mitigation (multi-carrier, diversity or adaptive equalization), radio link level packet data protocols, handoff criteria and algorithms, frequency assignment (dynamic assignment algorithms), error control and modulation, adaptive power control, time and frequency division multiple access, and encryption for privacy, security and authentication.

Challenges for large integrated fixed networks that support wireless access are: 1) to provide large scale mobility management (recognition of terminal location, registration of location, and routing) for terminals used away from home, 2) to provide small scale mobility management for handing off terminals from one fixed wireless access point to another as terminals move, and 3) to accommodate the widely varying error performance of wireless circuits as terminals move.

A workshop should be held to configure some wireless access systems for trial in open network testbeds. Research, planning, and implementation of prototypes for use in the testbed should be supported, along with experiments using the wireless access techniques in an open data network testbed. These experimental prototypes should make use of the new frequency spectrum near 2 GHZ, recently allocated by the FCC for these wireless communications applications. New wireless access information networks will take advantage of the opportunity provided by this new frequency spectrum dedicated for this use, as well as advances in techniques and technologies resulting from research on wireless access technologies.

National Gigabit Network Research Initiative

The NSF/ARPA sponsored project on gigabit testbeds has significantly advanced the field of very high speed networking. The initiative created five relatively small scale, disjoint testbeds to focus attention on distributed high performance scientific computing applications and the use of gigabit networks to support such applications. The Corporation for National Research Initiatives played a very important role in organizing this unique collaboration among academia, industry and government. The visionary, technically ambitious nature of the project appealed to stakeholders in industry and government, as well as academic researchers, and the three communities were able to come together to construct a research project which jointly far exceeded the capabilities of any single community.

The completion of the gigabit testbed projects in 1994/95 leaves open the question of how to maintain the research momentum, and capitalize on the opportunities available for collaborative research in high speed networking in the 1995-2000 timeframe. One of the main contributions of the gigabit testbed project was the development of a unique academia/industry/government collaboration. The landscape has changed significantly, the research agenda is somewhat different, but the opportunity for building upon the unique collaboration already established is, if anything, far more promising than it was in 1988 when the gigabit testbeds project was formed.

NSF has restructured the NSFNET program along the lines of NSF 93-52, to create the very high speed backbone network service (vBNS), the Network Access Points (NAPs), and a Routing Arbiter (RA), and to fund Regional Network Providers (RNPs) for the purchase of inter-regional connectivity on the open market. The NAPs, the RA, the RNPs, and the private-sector inter-regional IP carriers will constitute a coordination mechanism for commodity networking. There will be 4 NAPs deployed, one each in California, Chicago, New York, and Washington, DC. The vBNS will be, in 1995, a high speed backbone service connecting the NSF supercomputer centers operating at 155 Mb/s or greater speeds, with connections to each of the NAPs, initially at 45 Mb/s. The vBNS and the NAPs are likely to employ ATM technology, although some of the NAPs will begin with LAN technology. While the NAPs and vBNS might be used to provide interconnection among experimental subnets, they cannot serve as viable research networks for development of the next generation of network technology, since they must be extremely reliable, and cannot be disrupted for purposes of experimentation.

Members of the telecommunications industry are quite interested in contributing towards a national gigabit research network. Many researchers currently developing applications over limited scale gigabit testbeds could benefit greatly from access to other research sites with end to end gigabit speeds. Achievement of this goal will require a wide range of research activities. Greater gigabit network connectivity through a national gigabit research network which complements the government investment in NSFNET would enable a much richer set of collaborations for network and applications research, and allow a large set of research groups to participate. A national gigabit network complementing and interconnected with the new NSFNET can provide an effective basis for continued advancement of NSFNET capability as well as the technology transfer necessary to accelerate the NII.

A joint ARPA/NSF workshop was convened July 20-21, 1994, to articulate the research agenda for high performance networking and applications and determine the requirements for a gigabit network infrastructure. The workshop was attended by a broad cross-section of researchers in both networking and gigabit applications and the workshop report is expected to present a strong consensus position that carries the endorsement of key research leaders. NSF should work with ARPA and the research community to formulate specific plans for creating a gigabit research network infrastructure that will support the applications and network technology research agenda outlined in the workshop report.

6. COLLABORATIVE PROJECTS

As the power and extent of communications networks have increased, issues in networking have become intertwined with issues having to do with the usage of these networks by human users for the advancement of social and business purposes. The National Information Infrastructure initiatives recognize that issues in networking far transcend the technical and scientific problems usually associated with the communications theory field. If the great breakthroughs in optical transmission, image compression, data network protocols, and in other comparable areas championed by the National Science Foundation in past years are to contribute significantly to the attainment of national goals, then this formative work must be augmented by conceptual advances in the applications, usability, and economics of networks.

Most of the issues in the higher layers of networking, i.e., the "middleware" and applications layers, involve multi-disciplinary considerations cutting across the National Science Foundation organizations and across other government agencies. It is often unclear which organization has the ultimate responsibility to champion these issues, but the work itself is vital. Not only are these issues energizing to the communications field, but it is often true that the applications user interfaces and middleware needs determine the philosophy of network design for the lower layers. The committee strongly recommends that the Division of Networking and Communications Research undertake collaborative efforts with other divisions of the NSF, and with other government agencies, to ensure the development of appropriate research in these areas.

Agents and Filters

Agents are intelligent programs that act on a user's behalf to carry out actions in an autonomous fashion. An example of an agent would be a program that went into the network in search of a particular piece of information, such as the address of another user. In carrying out this task the agent might well clone itself to search out different substructures of the network. In a highly heterogeneous network of networks with ever-changing, locally-maintained directories, this is an important and difficult problem. Since the agent itself has many of the characteristics of a virus, a high priority area of research is the incorporation of protection against harm to the network and user resources.

A useful instance of simple agents is executable email. An electronic mail message could, for example, act as an automated polling mechanism with built-in return forms and statistical analysis of these returns. An executable email message could also serve for catalog ordering of items wherever the recipient clicked his mouse. The uses are obvious, but the protection concerns are paramount.

Filters are programs that select and prioritize information according to the instructions, needs, or customs of a given user. Simple filters today sort and prioritize electronic mail based on subject, author, key words, etc. Other experimental uses include the selection of news items from a passing stream of information. At issue is the effectiveness of user profiles, whether these are pre-selected by the user or learned over time by the filter itself based upon the accumulation of instances. The problems are greatly compounded when the information is in the form of digitized sound or images, as opposed to text. In a world increasingly filled with a flood of information, and with users ever more strained for available time, filters will assume an important role in the acquisition of information.

Navigation Tools

Arguably the most important development the past year in the evolution of the Internet was the introduction of Mosaic, a hypermedia information navigator that has opened up the network both to untrained users and to new and non-traditional information suppliers. Mosaic presents the user with a "point and click" interface that relieves the user from having to know where on the network the information sources are located and how these resources can be obtained. Mosaic also incorporates sound and both still and video images. It was originated at the University of Illinois National Center for Supercomputing Applications under funding in the gigabit testbed program. Through 1993 its use grew at the incredible rate of 11% per week, exceeding one million users by the beginning of 1994.

Mosaic is the latest in a succession of information navigators, or front ends, that includes Gopher, World Wide Web, and the Wide Area Information System. It is now very clear that the expansion of Internet into electronic commerce and into the nation's homes will be critically dependent upon its ease of use as perceived at the customer interface. Although the Mosaic program is currently serving as the model for this interface, and is in fact being rapidly commercialized, there is the likelihood that further exploration of the principles of information navigation might result in new breakthroughs in the application and use of networks. It is recommended that collaborative efforts be undertaken towards establishing models, techniques, and behaviors in the navigation of heterogeneous networks of disparate and ever-changing information bases.

Human Interfaces

Information navigation is one aspect of the user interface to computer networks. In Apple's influential advertisement featuring the "knowledge navigator" the user engages in a spoken dialog with his palmtop computer. "Show me the amount of rainfall in Brazil by year," he speaks to the computer, which responds with charts that exemplify the desired information. Subsequently, the computer negotiates video calls and conferences, relieving the user of the responsibility of placing and administering these multimedia connections.

One of the highly contested and debated areas of network interface today is in the set-top box that connects the user to video-on-demand services. Principles relating to how average users will interact with these highly intelligent appliances are not well understood, yet the utility and economics of broadband services will depend on how well accepted these interfaces become. It took many years, for example, for widespread acceptance of the mouse/windows interface to become nearly universal in personal computing.

Another paradigm for naive user interaction with a network is being pioneered today by General Magic, with an interface that perhaps owes more of its antecedents to the world of video games. In conducting electronic shopping using this interface, the user effectively "walks" down a street filled with shops. Doors can be opened, shops entered, and purchases made. It remains to be seen what level of acceptance this interface can achieve among the untrained populace, as compared with the the point-and-click hyperlinks of the Mosaic interface.

Research problems in the human interface to networks include issues in behavioral science, computer science, speech and handwriting recognition, image understanding, language understanding, and other allied fields. Some of these issues will affect the network design, while others may be regarded as incidental to the philosophy of networking. All of them will determine how and how much networks are used in the future.

Electronic Publishing

Electronic publishing is a field that has its genesis in the emergence of wide scale public networking, but at the moment it is still in an experimental phase. A number of journals and information providers are now offering free or subscription versions of their conventional publications through Internet. Others are inventing new multimedia forms of their journals that take advantage of the unique capabilities of computer networks. All of these efforts seem tentative and experimental at present, establishing a presence for the time when a suitable infrastructure for electronic publishing becomes a reality.

Perhaps the most important issues in electronic publishing lie outside the conventional technological domain. Certainly, copyright and intellectual property protection are at the forefront of discussion and controversy in this field. There are technical considerations that affect these issues however. One important capability in the network is the efficiency of billing. Can the network support the collection of a very small charge per unit of information accessed, or will the cost of billing itself swamp the small charge? Achieving a reliable, secure, and inexpensive billing system will be essential to foster the new electronic publishing industry.

The techniques developed in cyptology for privacy and security have direct applicability to electronic publishing. For example, it may be necessary to ensure privacy or anonymity in the access of information. Copyright protection is another problem that may be amenable to technological solution through the use of cryptographic techniques.

Scientists and engineers themselves have much to gain by building an infrastructure for electronic publishing. Advances in science are dependent upon a fast, efficient, and open publication system. The rate of advancement of technology has now outstripped the conventional publication system, and many scientists and engineers look to Internet to provide the latest information. However, the current culture in network news groups is so inefficient that many knowledgeable users have quit reading these bulletin boards altogether. Collaborative experiments for the publication of technical information might be an especially fruitful area for National Science Foundation participation.

Economics

It is difficult to consider many of the most important issues in networking without confronting issues in economics. In considering the National Information Infrastructure, for example, what is the cost of installing broadband capability? How does this compare with the cost of ISDN? How much more does it cost to make this capability two-way, as opposed to broadcast (and low speed control) only?

Why does Internet appear to be free to industrial users? Is it because Internet has discovered a more cost-efficient architecture for networking, through the philosophy of pushing complexity and cost to the periphery? Is it because Internet has a simple fixed-cost billing algorithm? Or is it because Internet has avoided the regulatory controls that have through many years achieved a socially-accepted distribution of the costs of the telephone infrastructure among the many classes of users?

It is frustrating to technologists that the costs of communications networks are so poorly understood, and seem only weakly dependent upon technology. Where are the economic bottlenecks, and how can technology be used to make network access and services inexpensive?

Collaborative efforts in the economics of networking would only be fruitful for communications technologists if real wisdom bearing on design philosophy was likely to emerge. The track record here does not encourage optimism, but some of the most critical national issues in networking depend on a better understanding of economics. Possibilities for collaboration are worth exploring, but must be carefully assessed.

Workstation Structure and the Communication Environment

Current workstations and personal computers are not designed with the aim of supporting very high speed network interconnections. With the exception of the fastest supercomputers, internal bus structures typically cannot support interfaces that run at speeds much above a few hundred megabits per second or less. Gigabit network speeds will require internal computer bus speeds that are at least several times that rate. Even at speeds of a few hundred megabits per second, the workstation software imposes a set of further bottlenecks for the data to contend with. Operating systems are not designed to handle large data flows in very short time intervals or those with real-time constraints. Applications programs often interface the operating systems in ways that provide barriers to high speed operation to the end application. The applications themselves may need to be rewritten or reconceived to allow such high speed operation, even if all the underlying software can support high data rates. Architectural approaches to the design of workstation hardware and software are needed that will alleviate these limitations and enable the transition from the slower existing network environments to the higher speed network environments of the future.

Middleware

The National Information Infrastructure (NII) will consist of a large number of application specific capabilities that are provided in a nationwide networking environment consisting of many interconnected physical elements. While a user could write software applications for a given piece of hardware without the need for operating system software, there is great advantage in not requiring each such user to separately invent or reinvent file management conventions, memory management techniques, scheduling, and resource allocation procedures. Further, by having common software services that all the applications can share, independent applications have a better chance of being able to interoperate, and multiple processes may even run essentially at the same time.

A similar set of intermediate software services (to be known as middleware) is needed in the NII to assist applications developers in creating applications and systems efficiently. Independent applications development groups may have the ability to interwork their applications over the NII without having to do joint development. Examples of generic middleware capabilities are persistent storage in the network, locator services, and global resource allocation. Examples of application specific middleware capabilities would include codified definitions of common terms, mappings between common terms and discipline specific terms, and widely used application specific functions. Research is needed to define the taxonomy and framework of the middleware and to explore specific instances of middleware in a collaborative network environment.

Optical Systems

Optical networking and communications is very much interdependent on optical devices and systems. This includes the development of integrated optical components, the integration of these components into optical systems, and the integration of optical components and systems with electronic systems and components. Optical interconnects for computer systems and optoelectronic computing systems are closely related topics. In 1993 the Networking and Communications Research (NCR) program, in collaboration with the Electrical and Communications Systems Division in the Engineering Directorate and the Cross Disciplinary Activities Division in the Computer and Information Science and Engineering Directorate, conducted a special solicitation on all-optical networking, with a focus on the interaction of device, systems, and networking researchers. A workshop was held, white papers were solicited and received, followed by a smaller number of full proposals, and five awards to three universities were made. Subsequently a Foundation-wide committee was established to facilitate collaborative efforts in optical systems and NCR is a participant is this effort. Small group research integrating optical networking and communications with optical devices and systems should continue to be supported. Also, it will be especially necessary to encourage partnerships among academic and industrial researchers if progress is to be realized in this field.

Applications of Gigabit Networks

While the technology of gigabit networks enables the interaction of multiple machines at very high speeds, not all network applications require high bandwidth. Applications per se do not normally fall largely within the network research area. Uses which do require very high bandwidth often involve supercomputers and scientific grand challenges; these generally lie at the heart of other scientific disciplines. Design and manufacturing applications are often engineering enterprises and computer intensive. Most tools for collaboration, including those that are specially designed for very high speed networks, are as much a distributed computing activity as they are a networking activity. Thus, we expect that most efforts to apply gigabit networks will require the joint support of multiple divisions of NSF and possibly multiple agencies of the government. These kinds of collaborative activities are encouraged.

Networking of Applications

Techniques are sought that enable scientific research applications that have been developed for operation on a single computer to be made operational in a distributed computing environment on the network. These include techniques for describing how to partition a given computation to run on multiple machines, tools for recrafting software for this environment and for distributing the components to other machines, and mechanisms for linking the components together and for instrumenting the resulting system. Where administrative control of the machines and their software environments cannot be provided easily, techniques need to be developed to locate and negotiate for network resources. Systems that facilitate the dynamic implementation of a shared computational environment are likely to provide implementation and execution efficiencies, but the use of standard execution environments will also be necessary.

7. CONCLUSIONS

This report identifies and describes a number of important research priorities (15 in all) central to the communications and networking area. These areas must be supported if a firm scientific base is to be provided for the design, development and deployment of future communication systems and networks. Many of these priorities represent enabling technologies for the National Information Infrastructure (NII). The four research initiatives, which are broader in scope and will require greater funding than single investigator projects, relate directly to the design and the deployment of future large-scale systems. It is now clear that development of future communication systems and networks is very much an interdisciplinary activity requiring expertise from the fields of electrical engineering, computer science, materials science, mathematics, and a number of other fields. In recognition of this interdisciplinary nature a number of activities are described in this report, which are suitable for collaborative efforts between several NSF divisions and for collaborative efforts between the NSF and other federal agencies.

STIS

The Science & Technology Information System (STIS) at the National Science Foundation

What is STIS?

STIS is an electronic dissemination system that provides fast, easy access to National Science Foundation (NSF) publications. There is no cost to you except for possible long-distance phone charges. The service is available 24 hours a day, except for brief weekly maintenance periods.

What Publications are Available?

Publications currently available include:

- The *NSF Bulletin*
- Program announcements and "Dear Colleague" letters
- General publications and reports
- Press releases, Other NSF news items
- NSF organizational and alphabetical phone directories
- NSF vacancy announcements
- Award abstracts (1989-now)

Our goal is for all printed publications to be available electronically.

Access Methods

There are many ways to access STIS. Choose the method that meets your needs and the communication facilities you have available.

Electronic Documents Via E-Mail. If you have access to Internet e-mail, you can send a specially formatted message, and the document you request will be automatically returned to you via e-mail.

Anonymous FTP. Internet users who are familiar with this file transfer method can quickly and easily transfer STIS documents to their local system for browsing and printing.

On-Line STIS. If you have a VT100 emulator and an Internet connection or a modem, you can log on to the on-line system. The on-line system features full-text search and retrieval software to help you locate the documents and award abstracts that are of interest to you. Once you locate a document, you can browse through it on-line or download it using the Kermit protocol or request that it be mailed to you.

Direct E-Mail. You can request that STIS keep you informed, via e-mail, of all new documents on STIS. You can elect to get either a summary or the full text of new documents.

Internet Gopher and WAIS. If your campus has access to these Internet information resources, you can use your local client software to search and download NSF publications. If you have the capability, it is the easiest way to access STIS.

Getting Started with Documents Via E-Mail

Send a message to the Internet address **stisserv@nsf.gov**. The *text* of the message should be as follows (the Subject line is ignored):

get index

You will receive a list of all the documents on STIS and instructions for retrieving them. Please note that all requests for electronic documents should be sent to **stisserv**, as shown above. Requests for printed publications should be sent to **pubs@nsf.gov.**

Getting Started with Anonymous FTP

FTP to *stis.nsf.gov*. Enter *anonymous* for the username, and your E-mail address for the password. Retrieve the file "index". This contains a list of the files available on STIS and additional instructions.

Getting Started with The On-Line System

If you are on the Internet: *telnet stis.nsf.gov*. At the login prompt, enter *public*.

If you are dialing in with a modem: Choose 1200, 2400, or 9600 baud, 7-E-1. Dial (703) 306-0212 or (703) 306-0213

When connected, press *Enter*. At the login prompt, enter *public*.

Getting Started with Direct E-Mail

Send an E-mail message to the Internet address **stisserv@nsf.gov**. Put the following in the text:

get stisdirm

You will receive instructions for this service.

Getting Started with Gopher and WAIS

The NSF Gopher server is on port 70 of **stis.nsf.gov.** The WAIS server is also on **stis.nsf.gov.** You can get the ".src" file from the "Directory of Servers" at **quake.think.com.** For further information contact your local computer support organization.

For Additional Assistance Contact:

E-mail: *stis@nsf.gov* (Internet)

Phone: (703) 306-0214 (voice mail)

TDD: (703) 306-0090

NSF 94-4
(Replaces NSF 91-10)

NATIONAL SCIENCE FOUNDATION
ARLINGTON, VA 22230

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

RETURN THIS COVER SHEET TO ROOM P35 IF YOU DO NOT WISH TO RECEIVE THIS MATERIAL ☐, OR IF CHANGE OF ADDRESS IS NEEDED ☐, INDICATE CHANGE INCLUDING ZIP CODE ON THE LABEL (DO NOT REMOVE LABEL).

BULK RATE
POSTAGE & FEES PAID
National Science Foundation
Permit No. G-69

00145173 ERIC
ERIC FACILITY
1301 PICCARD DRIVE
SUITE 300
ROCKVILLE MD 20850-4305

NSF 94-165
(Supersedes NSF 92-109)