

# Strategic Directions in Networks and Telecommunications

DAVID CLARK

*MIT Laboratory for Computer Science, Cambridge, MA (ddc@lcs.mit.edu)*

JOSEPH PASQUALE ET AL.<sup>1</sup>

*University of California, San Diego, CA*

## 1. INTRODUCTION

The computer/communication revolution has occurred. The Internet is rapidly becoming ubiquitous. Wireless technology is widely used and is supporting mobile computing. Network bandwidths have increased to the point where effective networked multimedia is now feasible. Yet we have only begun to see the implications of these changes and the changes to come. We can look forward to a future where everything communicates and communication is everywhere. Every person can be connected to the global telecommunications infrastructure wherever he or she is located. Even very small and inexpensive devices will be part of the computer/communications infrastructure. Mankind's information will be digital and accessible via the network.

Existing communications systems are converging rapidly into a ubiquitous information infrastructure. Today we see telephony appearing on the Internet, the Internet running on a mesh of telephone circuits and fast packet switch-

ing, video delivery migrating to a range of new delivery technologies, and shared communications resources supporting a variety of application environments. George Heilmeier, CEO of Bellcore, has been quoted as saying, "The Internet will be larger than the phone system within five years." We are rapidly moving towards an information infrastructure that does not distinguish between computing and communications, but rather provides an integrated set of distributed services to the user. In some cases, the network is a technology to hook computers together; in other cases the computer is just a better way to build a network. The research community must be prepared to deal with these changes, integrating the disciplines of communications, networking, and traditional computer science.

Society needs at least a basic familiarity with this new technology in order to use it effectively. The research community must work with society to assure that networking supports society and society can use the network. Enhanced features are required, such as a greater robustness of the network, better support for an economy of resources and information, and the protection of privacy of personal information. Methods to locate information, people, and resources in the ever-growing network are needed. And network growth itself, both in scale and capability, implies the

---

<sup>1</sup> Members of the committee that produced this report are David Clark (MIT), Deborah Estrin (University of Southern California), Paul Green (IBM), Jim Kurose (University of Massachusetts, Amherst), Barry Leiner (MCC), Larry Masinter (Xerox PARC), Joseph Pasquale (University of California, San Diego), David Sincoskie (Bellcore), Karen Sollins (MIT).

---

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

© 1996 ACM 0360-0300/96/1200-0679 \$03.50

need for new architectural and technical approaches.

The future is driven by societal choices from available technologies. Our job, as the research community, is to enlarge the space of technical possibilities in a way that creates new choices that are of value to society. The choices we make in the research we do will drive the future through creation of these possibilities.

The directions in which networking and telecommunications will evolve are very much shaped by the key lessons learned so far. In this document, we discuss lessons we have learned and the problems and challenges that lie ahead.

## 2. LESSONS LEARNED

As networks and telecommunications have developed, a number of fundamental principles and broad themes have emerged about networks and their applications, lessons that have a great bearing on the directions for the future. In this section we briefly outline the history of networks and some of the principles that have been learned.

### Computers in the Network

The origins of networking as embodied in the telephone system are far older than computing. However, computers have had a profound impact on the shape of networks, both in the character of the telephone infrastructure itself and the advent of totally new sorts of networks to connect computers together. Perhaps the most basic change in the design of networks is the evolution from circuit to packet networks.

In a *circuit network*, at the beginning of a call the switching elements inside the network are configured to provide a path from the source to the destination of the communication, and then that path—the circuit—is maintained for the duration of the call. This may make somewhat inefficient use of the capacity of the circuit, but it is easy to engineer. With the advent of computers, it became

possible to allocate the bandwidth of a physical transmission link to particular communications circuits for shorter periods of time, so as to be more efficient in the use of link capacity, especially for intermittent or bursty traffic sources. The concept of a *packet network* is that the information to be sent is broken into small units—typically a few tens to a few thousand bytes, and each such unit is provided with a header that describes where it is going and how to recombine it into a larger data element at the destination. These small units, called packets, are then forwarded through the network, using a sequence of dedicated computers called packet switches that are positioned at the connection points between the physical transmission links in the network. A packet switch receives packets arriving from a link and examines the packet headers to determine where to forward the packet next. Only with the advent of inexpensive processors was it feasible to dedicate a computer to performing this computation for each packet.

There are two major classes of packet networks. In one class, called a *connection-oriented* or *virtual circuit network*, a preliminary step is performed before starting to send packets, in which each switch along the path is told how the packet should be handled. Each switch remembers this information, and when the packet arrives, the only action required is to locate the relevant information that has been remembered. In the other class, called a *datagram network*, there is no setup of handling information before packets are sent. Instead, the packet header itself contains enough information to direct the handling decision. The advantage of the connection-oriented approach is that the packet header can be small (it need only provide a pointer to the stored information) and the per-packet processing is presumably precomputed. The advantage of the datagram network is that if a packet switch crashes or the route taken by the packets through the network changes, the packets can still

be forwarded, because the header contains all the information necessary to re-establish the forwarding information. Both sorts of packet networks, as well as circuit networks, are in use today.

### Protocols and Network Architecture

Protocols—those conventions that describe how the various components of the network should cooperate to achieve some function—are perhaps the most important contribution of networking research. Protocols represent the codification of required behavior; a well specified protocol will allow two independently conceived implementations to *interoperate* successfully.

Network architecture defines the framework in which such protocols exist. Perhaps the most important architectural principle that has been established is the importance of *layering*, the definition of modular components of a network architecture that successively build upon each other to provide, in the end, services upon which network applications can be built.

Layering provides insulation so that changes in different parts of the system can occur independently. Application-specific requirements are not built into the lower-layer protocols; instead, protocols remain unaware of the application whose data they are carrying. With layering, it has been possible to successfully accommodate vast changes in technology and scale (e.g., the evolution of physical links that underlie the Internet from 56 Kbps to 655 Mbps lines, the increase in the number of Internet hosts from four in 1969 to tens of millions today) with relatively little disruption.

### Partitioning Functionality

A second important architectural principle that has emerged is the importance of properly *partitioning the functions* of the network between the packet switches within the network and the end-systems or computers attached to the edge of the network. In the Internet,

a simple datagram packet service, together with additional functionality provided by software in the attached end-systems, has proven sufficient to build all the Internet applications we see today. This approach contributes to the flexibility and evolvability of the network, since adding a new service often involves only modification of the relevant end-system software, and not the internals of the network. The *end-to-end* design principle [Saltzer et al. 1984] argues that only functions implemented at the endpoints, taking into account the requirements of an application, can fully meet the needs of that application.

### Reliability

The telephone network has taught us that *reliability* is an important concern when billions of people rely upon a large-scale communication system for so many aspects of their daily lives. Redundancy, network monitoring, and network management play key roles here. There is an hypothesis that loosely coupled interoperable systems will function more reliably than tightly integrated, continually consistent ones.

### Scalability

Scalability describes those characteristics of system structure that allow it to grow gracefully. Networks must be designed with scalability in mind from the outset, because successful networks will experience continuous growth in a number of dimensions. A truly scalable network architecture will accommodate growth in bandwidth, number of users, and number of connections, and from lower-level services such as routing to higher-level services such as naming. For example, experience has shown that there never seems to be enough bandwidth. As more bandwidth is provided, users conceive new applications that consume it. At the same time, the number of users continues to grow. As with bandwidth, there never seems to be

enough address space. The next-generation Internet protocol (IP) is going from a 32-bit to a 128-bit address space.

An important aspect of scale is *heterogeneity*. For example, links in parts of the Internet differ in capacity by orders of magnitude, delays across different parts of the network differ by orders of magnitude, and attached devices differ in processing power from pagers to supercomputers. Any successful protocol design, at either the network or application level, must take heterogeneity into account.

### Fundamental Limits

Beyond the current engineering issues lurk some very fundamental limits on performance. A critical factor is the speed of light, which is a very real limit on network performance. The minimum round-trip time across the country is about 30 milliseconds. As computers get faster, an action that takes 30 milliseconds will become more and more a dominating source of overall delay. It is thus critical to develop techniques to mitigate this basic limit, ranging from redesign of applications to modification of lower-level protocols to reduce the number of end-to-end interactions.

### The Centrality of the Application

The innovations in networking that have captured the most visibility are those at the upper layers—new applications and the services out of which they are built. The World-Wide Web is the most visible application framework today; it is being put to countless uses in commerce and human communication. The success of the Web should not eclipse the importance of electronic mail, which at this point has become a basic part of life for many people. Although network designers stress that the network is more than its applications, it is the applications that drive the evolution and deployment of the network.

### The Network Research Process

In addition to the technical lessons discussed, some important observations can be made about the process of performing research to advance the field of networking. Much of the progress in networking research has been a group process, perhaps as a result of the overarching need for interoperability among communicating systems and applications. A lesson that has been reinforced multiple times is the importance of synergy between developing applications and developing network technologies and services. New networking technologies must be developed concurrently with next-generation applications that can drive and exploit those networking technologies.

Another observation is that it is often impossible to develop new network enhancements in the context of an operational network, that is, one that is providing promised levels of service to large numbers of users. An experimental infrastructure is required that will permit these next-generation technologies to be tested. This requirement motivates the continuing need for network testbeds, which in turn is another driver for group research.

### Research in New Applications

The role of classical academic research in the development of new applications is sometimes problematic. A new application, once conceived, can often be put into practice very quickly. This means that a good idea can show up as a product before it emerges from the research lab, since product developers can devote more resources to development and the basic novelty is the core concept itself.

This suggests that academic research in applications should focus on those aspects that individual commercial developers cannot immediately develop. Sometimes an overall architecture or framework is needed in order for an application area to emerge, and individ-

ual implementors cannot develop this framework in isolation. Examples include directory services, common standards for data types, security architectures, namespaces for objects, and so on. These sorts of frameworks, however, cannot be as effectively explored by one worker in isolation as in a collaboration among a range of players in the field. This factor provides yet another impetus for group or collaborative research.

**3. THE CHALLENGES AHEAD**

The strategic directions for future network evolution are very much related to some of the issues listed.

- the interplay of application requirements, which evolve rapidly in open computer networks, and the infrastructure that supports them;
- exploration of new models for layering and functional modularity;
- issues in distributed application design, such as multiparty applications, active objects within the net, and new paradigms for interaction among remote application elements;
- better models and approaches for designing networks, addressing issues such as reliability, performance, congestion, routing, network control, and security; and
- new media and network technology, such as wireless networks.

The following specific topics address one or more of these strategic considerations.

**Requirements and Services**

Although there is considerable understanding of how to design network services that are useful to application builders, there are a number of areas in which the specific requirements are not yet clear. An example is reliable multicast (multiparty) delivery of information. It is not clear exactly what is meant by reliable multiway delivery. If one recipient crashes, does the transfer

cease? Can a recipient join the transfer part way through? What happens if one recipient is much slower than the others? The research question here is not so much to design the mechanism, although that can be very challenging, but to argue what mechanisms are needed.

There is currently debate in a number of areas about the specific services that are actually needed.

- What models of remote procedure invocation represent the best combination of semantic clarity and performance?
- Are there general modes of partially reliable data delivery?
- What forms of real-time service are best for data networks?
- What is needed to better support authentication and security?
- How do we design with evolution in mind?

Answering these questions requires an interactive exchange between application designers and designers of the infrastructure on which the applications depend. It is an ongoing issue that will persist for at least another decade or more.

**Design and Validation of Network Services**

As the population of network users and applications grows larger, the range of network services is expanding dramatically. An increasing number of independent protocol functions must work simultaneously to provide reliable, efficient, and usable communications. The interactions are especially complex for protocols that involve more than two participants, for example, the functions of signaling, routing, congestion control, resource reservation, scheduling and admission control, reliable multicast, multicast flow control, adaptive applications with layered encodings, active networks, and web replication schemes. The resulting complexity may outstrip our ability to design and evaluate the

required protocols. New methods are required for design and evaluation of network protocols and services.

New design methods must address at least three evaluation criteria that are critical to protocol scalability: robustness, communication efficiency, and semantic clarity.

Robustness must be evaluated systematically in the context of a wide range of test scenarios, including large networks, multiple simultaneous network events, heterogeneous media characteristics, and interprotocol dependencies. Obviously it is impossible to explore the space of possibilities exhaustively; therefore, methods/heuristics are needed to identify the important parts of this space. We require tools and methods that support systematic study in the context of topologies (large numbers of nodes, connectivity degree, network diameter, numbers of alternate paths, link bandwidths, asymmetric links) and traffic inputs (session lifetimes, group sizes, arrival and duration of source bursts, etc.).

The anticipated everything-communicating-everywhere applications will stress the shared communication infrastructure. The possible advent of usage-sensitive pricing will favor efficient network usage. Consequently, overall communication efficiency will be an increasingly important performance metric for the design of higher-level protocols and distributed applications. Practical protocol design must recognize the real sources of inefficiency, and algorithmic evaluation methods should emphasize efficient average-case designs instead of guaranteed worst-case optimization.

The semantic clarity of the protocol interfaces will determine their usability by the very heterogeneous population of application developers. Application developers, who are not by training network experts, must be able to understand the features of the service that the network provides, so that

they can construct applications that work correctly and deal correctly with the various failures and performance degradations that may occur at the lower layers. This objective can sometimes be in conflict with the goal of communications efficiency, as simple abstractions sometimes do not perform as well as more complex ones. This suggests that there may need to be families of service interfaces that allow the user of that interface to trade off complexity against performance and flexibility.

We can think about this collection of requirements as the equivalent of software engineering (design and debugging tools and methods) for network protocols, including techniques for visualization, abstraction, and testing. To address the combined challenges of increasing scale, protocol interaction, and system details, use of analysis, simulation, and prototype implementation must be integrated.

### Protocol Specification

Although great strides have been made in writing practical protocol specifications that permit independent implementations to interoperate successfully, there has been much less success writing specifications that ensure predictable *performance* between conforming implementations. The problem of specifying performance requires real innovation, both at the practical level of specification techniques that ensure predictable performance between conforming implementations and formal approaches to specifying and reasoning about protocol performance.

### Economic and Social Implications for Protocol Design

In the design of protocols for the global network, social, policy, and economic issues often have as profound an effect on the protocol design as the traditional issues of performance, robustness, and scalability. However, neither the theo-

retical underpinnings, models, nor analysis tools are currently available to help deal with these issues. Social issues include privacy and public policy, as well as the motivational factors that encourage one kind of network usage over another. For example, in caching of Web pages, the client, the proxy cache, and the original server may be owned and operated by different organizations or individuals, with conflicting goals for minimizing use of the network, policies for access to material, or gathering demographic information.

Conversely, economic and policy mechanisms can be used to help achieve the goals of stability and availability. When a shared resource is subject to unconstrained use by individuals optimizing their local benefit, it is possible to experience “the tragedy of the commons”: individual activities, in concert, destroy the common good. To encourage efficient use of the network and prevent denial of service to legitimate users, mechanisms for network usage should be accompanied by an appropriate incentive structure based on market-economic principles, such that when each user optimizes his or her own use, the resulting overall behavior is good for the network. The design, implementation, and analysis of incentives and their effect is a complex problem, as illustrated by multicast, where large numbers of distributed users share in the costs and benefits.

### **Middleware—the Remodularization of Protocols**

Middleware describes a set of services and protocols that exist just below the applications themselves and provide a set of common application services. There is considerable speculation that these services actually exist but that we are still in the early stages of defining and implementing them. The range of services proposed for middleware is very diverse. They include security services, naming and directory services, and electronic commerce and payment

schemes. It is clear that work on all these services is very important; it is not clear what form the final result will take.

The mature form of middleware might be actual implementations of services, or just common concepts shared among applications. One of the intellectual exercises of network research is a reorganization of the concepts and functions between the protocol layers. For example, some of the functions traditionally performed at the transport layer (as in the TCP protocol), such as reordering of packets or recovery of lost packets, is sometimes now done as a part of application code, in order to provide application-specific variations of these functions. In time, these functions may migrate back again to a more complex and sophisticated transport layer, as we learn better how to abstract away from the application the needed functionality. “Middleware” is an hypothesis that there is a class of services for which this abstraction can now be done.

### **Multiparty Applications**

Increasing maturity of the network world has forced us, and will force us increasingly in the future, to understand the problems of highly distributed sets of cooperating programs. Today, most of the successful applications in use are, at their heart, two-party interactions. This includes email, retrieval of Web pages, file transfer, and remote login. Multiparty designs are much harder to realize. The complexity of multiparty systems arises from many causes: the heterogeneous nature of the different end nodes, which may have very different functional and performance characteristics; differences in performance in different parts of the network; and transient or persistent failures of some of the participants in the system. Designing robust and flexible approaches to multiparty systems is an important research topic for the future.

### **Objects in the Network**

Object-oriented approaches are becoming widespread in the software community. This shift is already having an impact on telecommunications and networks. Active objects as methods for remote execution of programs over the network (e.g., Java) are attracting wide interest. Moving objects rather than packets allows the possibility of a reprogrammable infrastructure by allowing the objects to carry with them the methods by which they should be processed by the network.

### **Modularity of Distributed Applications**

The physical separation of the parts of a distributed system forces a discipline on software construction. Within a single computer, the interaction among the parts of a system can occur in a number of ways: direct invocation of one part by another, indirect connection through the file system, global variables, and so on. These latter forms of connection are sometimes very powerful, but often the cause of poor modularity and obscure overall system behavior. In a system where the parts are separated across a network, these indirect links cannot exist unless they are explicitly designed and implemented, which makes the resulting modularity easier to understand and document. Although this may make the modularity cleaner, it also may inhibit certain forms of powerful system design. It may be necessary to come up with more sophisticated models of interaction among the parts of a distributed system than simple remote invocation, and research in this area is critical to the successful construction of next-generation distributed applications.

### **Network Computing**

Network computing allows the network to be viewed as a computer. The promise of a network computing system over more traditional ones is its increased scalable cumulative power, more efficient (and economical) use of existing

resources, and more effective system management. To support network computing on a large scale (such as the Internet), research is needed in at least three areas. To provide incentives for network resource sharing, one needs the ability to buy and sell resources, thus requiring flexible communication protocols for establishing contracts. The goal is to support the creation of a market economy of network-accessible resources. The second area is flexible communication (and computation) architectures where the basic type of message is not simple data but a program, for example, an agent programmed to buy computing and communication resources as it moves about. The third area is the provision of qualities of service. A market economy will not work if buyers do not know the quality of what they are buying, or sellers cannot ensure the quality of what they are selling. This points to two research directions: how to provide a range of qualities of service, including those that are at least predictable (but not necessarily guaranteed), and how to provide adaptivity tools for applications that require more than best-effort services in environments that support services whose qualities are less than guaranteed.

### **Reliability**

The issue of reliability or the mitigation of its alter ego, failure, will become increasingly important because of its societal impact as the community of participants becomes increasingly naive with respect to computers and communications. Failures will occur both in software and hardware, from a low-level network link or router to higher-level service elements such as a name server or web server. Although as technologists we have a set of tools that mitigate failures to some extent, neither is there a full understanding of how to address the problem of complex failures in distributed systems, nor have we done an adequate job of preparing programmers

and users for the fact that, despite our best efforts, failures will occur. The tools we do have include formal specification, verification, testing, simulation, and validation of functionality. In addition, certain design techniques can help limit the impact of failures, such as replication and distribution of functionality and information, as well as specification of what constitutes “adequate success” for a given application situation. We are left with two overarching issues, a framework for providing more consistent reliability and the need to further educate the public about the possibilities of residual failures.

### **Congestion Control**

Congestion control, the control of offered load when there is insufficient capacity in the network, is a persistent problem that is central to networking. Networks depend on the statistics of sharing in order to make efficient use of the underlying transmission facilities. However, any dependence on the statistics of offered load means that at some moments there will be traffic in excess of what can be carried. Further, there will always be parts of the network with relatively smaller capacity, and traffic entering these regions must adapt to the lower supportable rates. Congestion control has been studied for at least two decades, but is still a very difficult problem, only partially solved.

### **Routing**

Routing of traffic inside the network is another critical and only partially solved problem. As typified by the current Internet, practical solutions that scale to large networks are known that can construct reasonably efficient routes when measured by a single quality metric. However, the real requirements are much more complex than this. Today’s networks must support applications with very different service requirements, which can imply finding routes based on a number of service

metrics. The transition to the commercial world, in which providers compete to offer service, requires that routing respect the purchasing decisions of the consumers. Multicast and enhanced real-time services require the creation of routes “on demand” for individual applications. At the same time that we are dealing with increasing functional requirements, we are dealing with networks that are growing so large that we will require new approaches to scaling in route construction.

### **Signaling Protocols**

The control protocols used in networks, sometimes called “signaling” protocols, have to do with routing, allocation of resources, allocation of cost, collection of revenues, detection of and recovery from faults, addition of new equipment, security, and so on. Today there are several distinct approaches to the design of signaling protocols, as typified by the telephone system and by the Internet. The Internet has favored a more “lightweight” approach to signaling. However, it is not clear what approach will best scale to very large size, be robust in the face of failure, and implement the needed range of function. Research is required, both in these specific control aspects and in the more general issue of how large distributed systems should best be built.

### **Network Management**

Network management, an aspect of overall network control, is usually meant to describe the means by which communication is established with network elements so that they can be controlled. As networks have matured, the need for good management tools is clear, and a great deal of practical work is going on in this area today. However, from a research perspective, network management has been somewhat frustrating, because there seem to be few general principles or concepts that help to simplify the issue. What is needed is

a common framework for classes of managed objects and a recognition that objects themselves need to have a clearer articulation of what their expected behavior is. In fact, network management extends to software management in that the managed objects will no longer be limited to network elements, but will include software structures such as processes and files, databases, and the like.

### **Security**

Networks have altered the set of considerations about computer and information security. We have a large number of security mechanisms in place, including rather robust authentication tools, encryption for disclosure control and integrity of data on the network, firewalls for site security, and so on. What we lack are overall models that allow the user or system administrator to reason whether a system is actually secure against the relevant threats, given that some combination of these mechanisms has been deployed. We are facing a situation in which we have no effective overall models of computer security in a network context.

### **Wireless/Mobile**

Nowhere has the microsystems revolution had a larger impact than in the area of wireless communications and mobile computing. Capabilities such as cellular telephones, pagers, and notebook computers give but a glimpse of the changes to come. Integration of low-power, high-capability microprocessors and digital signal processing will provide the possibility of high-bandwidth communications on the move everywhere. This raises significant research issues ranging from the dynamic management of shared radio bandwidth and routing in wireless mobile environments, to the provision of services supporting applications having to operate in a mobile environment characterized by sporadic connectivity, changing

bandwidth, and rapid relocations. Mobility is not just a new network technology but the opportunity for a whole new class of objects to become resources in the network.

### **The Behavior of Complex Systems**

As networks grow to connect millions of nodes, and as these nodes all communicate in unpredictable patterns, the resulting overall behavior becomes very difficult to model or predict. Large highly connected systems can show aggregate behavior with complex characteristics: they can become chaotic, show self-organizing features, or oscillate. We see the possibility that large networks such as the Internet have these tendencies, but we lack the tools or methods to explore this eventuality, to model how this might happen, or to control the resulting behavior if necessary.

### **Preparing Society**

The process by which the Internet evolved to its present form, where it is seen by the public as essentially the Web, was one in which network technologists placed a number of options at the service of society's needs and, by iterating back and forth between needs and technical capabilities, the two communities arrived where we are today. The process has been very successful, but at the same time imperfect in both directions.

There are several examples of situations in which the networking community may not have done as good a job as it might in making the public aware of some of the societal implications of networked communications. Users do not completely understand yet that no matter how good a network is, it is capable of failing at times. A better job of outreach and education can help shape public policy debate as, for example, with cryptography. One of the future challenges to the network research community is to improve this outreach function.

Networks can change the nature of any highly connectable society, increasing communication among peers and reducing the power of central authority, but this seems to require relinquishing absolute control over the content and may require recasting some of the laws and regulations that have guided society's collective behavior. Examples include intellectual property rights, the rights of free speech in an international forum, and control of pornography. Better interaction between our field and the rest of society is critical to smooth the way to the future.

#### 4. CONCLUSIONS

##### The Forces that Shape the Future

The previous section listed a number of problems to be solved if we are to evolve the networks of today towards an even more ubiquitous network with global interconnectivity and advanced services. The specific technical issues—reliability, congestion control, routing, network management, and so on—are certainly critical. But these are not as important as the more architectural questions of overall design—how the network is modularized, how we can validate our designs before deploying them in millions of computers, or how we can understand and control the aggregate behavior of these very large systems. And finally, these issues in turn may not be as important as the issues of economics and utility. The principal driving force of network evolution today is not technical innovation, but demand from users. The principal barriers to further deployment of advanced networks today, for example, to the home, are issues of cost, not technical limitations. The principal issue in designing new network services today is not which options can be realized, but which would be useful. The network in five years will have a form that is bounded by the technical innovations of today, but it will be selected by utility and economics.

As a field matures, the nature of its

research changes. In the beginning, the field is concerned with the discovery of the new ideas that define how the field should evolve. Later, shaped by these new ideas, the research is also concerned with the removal of barriers to achieving the objectives that have already been articulated. At the same time, especially in computer science, new ideas seem to arise continuously, and deflect the trajectory of the field in unexpected directions.

The field of networking and telecommunications has now reached the point where removal of barriers is as important as the discovery of new concepts, and this has changed the nature of research. Barriers can be technical, but more often reflect a mix of technical, economic, policy, societal, and other issues. An instructive example is the design of packet switches. The first research issue was just to show that effective packet switches could be designed and constructed. Today the question is much more complex: what form of switches will be required in 5 or 10 years to meet the future needs of the telecommunications and data networking industry? This question explores the shape of the industry, the future of convergence, and the natural scale of the network of tomorrow, as much as it is a technology question. Future research, therefore, must increasingly take into account this broad range of issues.

##### Further Reading

Two other discussions of the future of networks and telecommunications, with specific suggestions for research, can be found in NIST [1994] and Computer Science and Telecommunications Board [1994]. The latter contains a discussion of the importance of layering and the economic issues central to shaping the future. For a review of current plans for network development, and how they might influence the next five to seven years' deployment, see Computer Science and Telecommunications Board [1996]. Two important network archi-

tectures today are the Internet, a datagram architecture, and asynchronous transfer mode, or ATM, a virtual circuit architecture. A collection of articles on the Internet can be found in Leiner [1994], and articles on ATM in Vetter and Du [1995]. For an introduction to technical issues such as layering, partitioning of functionality, and congestion control, using the Internet as an example, see Peterson and Davie [1996]. A discussion of protocol specification can be found in Holzmann [1991]. A seminal discussion of the issues in packet communications can be found in Kleinrock [1978].

#### REFERENCES

- COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD, NATIONAL RESEARCH COUNCIL 1994. *Realizing the Information Future: The Internet and Beyond*. National Academy Press, Washington DC.
- HOLZMANN, G. J. 1991. *Design and Validation of Computer Protocols*. Prentice Hall, Englewood Cliffs, NJ.
- KLEINROCK, L. 1978. Principles and lessons in packet communications. In *Proceedings of the IEEE*, Vol. 66, 1320–1329.
- LEINER, B. ED. 1994. Issue on Internet technology. *Commun. ACM* 37, 8 (Aug.).
- NIST 1994. *R&D for the NII: Technical Challenges*, Proceedings of a Workshop (Feb. 28–March 1). Educom, Washington, DC.
- PETERSON, L. L. AND DAVIE, B. S. 1996. *Computer Networks: A Systems Approach*. Morgan-Kaufmann, San Mateo, CA.
- SALTZER, J. H., REED, D. P., AND CLARK, D. D. 1984. End-to-end arguments in system design. *ACM Trans. Comput. Syst.* 2, 4 (Nov.).
- VETTER, R. J. AND DU, D. H. C., EDS. 1995. Issue on ATM networks. *Commun. ACM* 38 2 (Feb.).