# TCP/IP: The next generation

by E. G. Britton
J. Tavs
R. Bournas

***After tracing the evolution of Transmission Control Protocol/Internet Protocol (TCP/IP) from its academic and government research heritage to its current commercial use, we survey new directions in TCP/IP. We explore the issues that derive from growth in scale and function, which are prompting the Internet community to assess significant changes in the protocol suite. We relate these issues to the IBM Open Blueprint™. Much as the IBM Personal Computer brought computing to millions of people, the Internet service of the IBM Global Network and IBM's Internet Connection products are bringing networking to millions of people and are changing how customers are transacting business among themselves and with IBM.***

The Internet protocol suite has evolved over the years as its set of users has expanded to include groups with many needs not originally considered, as new applications have emerged, and as new kinds of link facilities have become available. It is now undergoing yet another quantum jump in capabilities to satisfy new requirements, most of which can be grouped according to the four layers of the Internet protocol suite (Figure 1): application, transport, internetwork, and network interface. One network-layer protocol, Internet Protocol (IP), and two transport-layer protocols, Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), let many application-layer protocols run over many kinds of network interfaces. Some other new requirements concern the entire system because they span multiple layers. For each of these sets of new requirements, we briefly summarize the currently deployed technology and the extensions to it that are proposed for satisfying those new requirements. For background on subjects not discussed in this paper, such as network management, please see one of the tutorials cited in the references.[1,2]

## Evolution of the TCP/IP user community

The Internet is a very successful worldwide set of interconnected networks that can communicate with one another. Currently it comprises over 50 700 networks, about 44 percent of which are outside the United States. Millions of users on more than 4.8 million computers in over 89 countries transmit more than 16 million million bytes through the Internet each month. Although the Internet protocol suite includes many protocols, people often refer to it as TCP/IP, from the Transmission Control Protocol and the Internet Protocol. TCP/IP implementations are available for almost every operating system and hardware platform.

Although many of the networks to which the Network Information Center of the Internet has assigned network numbers do not connect to the Internet, its history provides a useful structure for characterizing the expansion of requirements and concomitant expansion of the TCP/IP protocol suite.

**Original ARPAnet for defense research.** In 1968 the Advanced Research Projects Agency (ARPA) of the U.S. Department of Defense (DoD) proposed that




0018-8670/95/$3.00 © 1995 IBM

computers be interconnected by telecommunications links to allow researchers under ARPA contract to share resources on ARPA-funded computers at diverse sites.[3] As a result, the original ARPAnet, which included a System/360* at the University of California at Santa Barbara, began operation in 1969. It quickly developed protocols for remote log-in, file transfer, and electronic mail.

By 1973 the original ARPAnet protocols began to seem inadequate for transmission over a sequence of different kinds of physical connections (terrestrial teleprocessing links, packet satellite, packet radio, etc.) with different addressing schemes, maximum frame sizes, etc.—some of which could not guarantee error-free delivery. In 1974 Cerf and Kahn[4] proposed the Internet Protocol for interconnecting and routing among dissimilar networks, and the Transmission Control Protocol for ensuring the end-to-end interprocess communication above the potentially unreliable links.

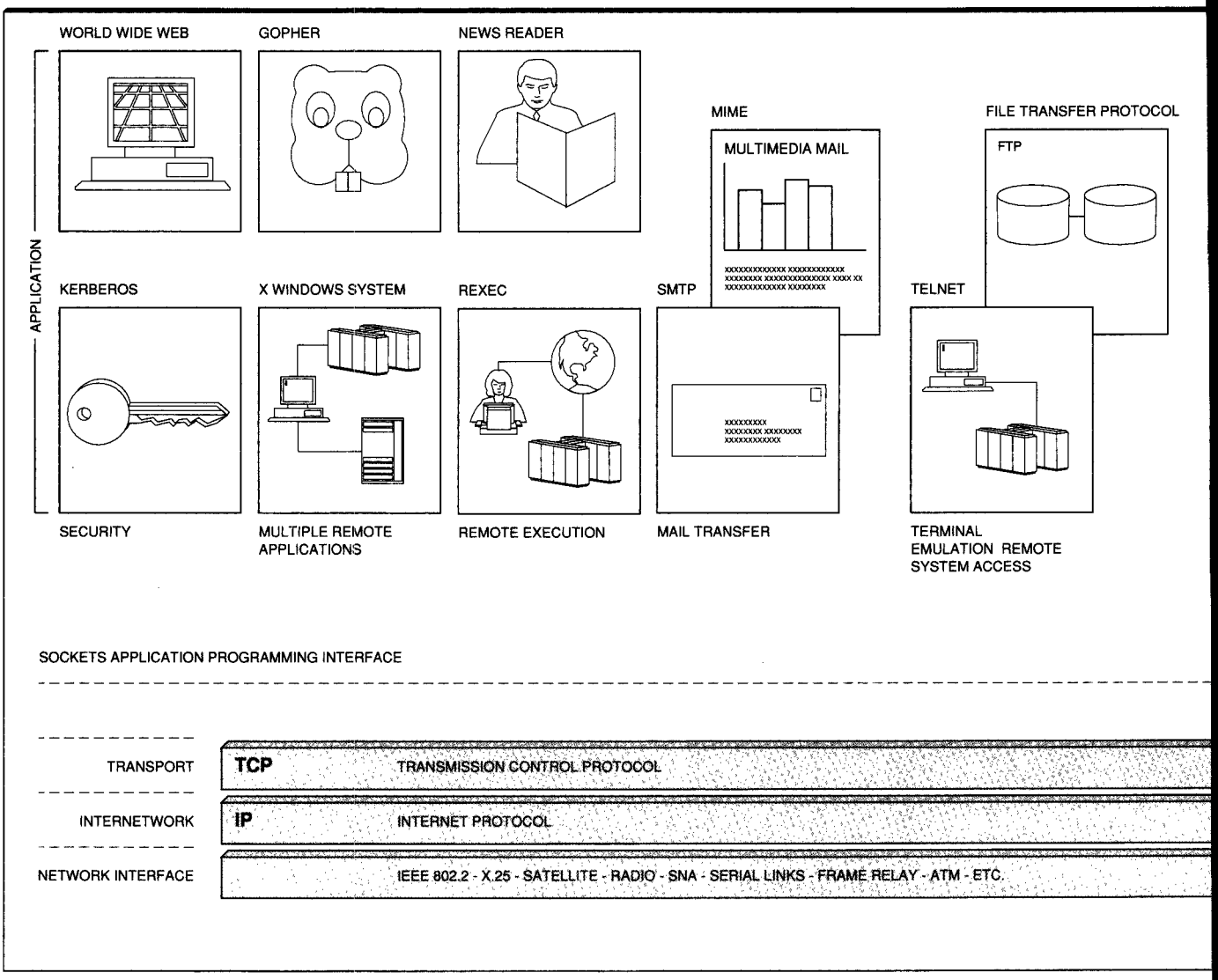
**Proliferation after Department of Defense standardization.** In 1978, after several years of experimentation with TCP/IP, the DoD declared these protocols to be the standards for its data communication networks. To facilitate dissemination of these protocols throughout its user community, the DoD contracted with Mitre Corporation to have TCP/IP included in the Berkeley Software Distribution** (BSD**) version of the UNIX** operating system.[5] Since many universities were under contract to do research for the DoD, they could get copies of the BSD UNIX operating system with TCP/IP essentially for free. Soon TCP/IP became the common language of academic networking in the United States. In 1983 the ARPAnet finished replacing its original protocol with TCP/IP.

**From ARPAnet to the Internet.** In 1986 the National Science Foundation (NSF) decided to proliferate Internet attachment throughout the academic community. They proposed the NSFnet, a 56-kilobit-per-second backbone network to interconnect supercomputer sites in the United States. Regional networks attached to this backbone, and campus local area networks (LANs) attached to the regionals. By 1987 the NSF contracted with Merit Network, Inc., MCI, and IBM to design, build, and operate a 1.544-million-bit-per-second (T1) backbone. IBM developed special-purpose routing software executing on multiple IBM RT* systems loosely coupled by token rings, and helped design the Border Gateway Protocol (RFCs 1105, 1163, 1164, 1265, 1266, 1267, 1268, 1654, 1655, 1745, 1771, and 1772) to handle topology exchanges in the backbone and with the regional networks. (An RFC, or a Request for Comments, is an Internet document available on line from ds.internic.net and many other repositories.) In 1990 the backbone started upgrading to 45-million-bits-per-second (T3) links.

In 1990 IBM, Merit Network, and MCI formed Advanced Network Services (ANS) to set up and operate the NSFnet. By 1991 several providers were offering TCP/IP service to commercial users, but NSF's acceptable use policy would not allow commercial traffic across NSFnet, which offered the only interconnections between many pairs of other service providers. ANS solved this problem by creating a wholly-owned taxable subsidiary, ANS CO+RE Systems, Inc., to offer Internet service to both commercial and research or educational enterprises. In a sense, ANS CO+RE leased to commercial users the portion of the backbone capacity not paid for by NSFnet. ANS CO+RE further agreed to participate in the Commercial Internet Exchange, a cooperative venture among commercial Internet service providers founded in 1991 to interchange traffic. A mesh of interconnected service providers has replaced the backbone of the Internet. IBM contributed extensively (RFCs 1222, 1322, 1560, 1597, 1620, 1629, and 1787) to the design and development of the interdomain routing of the Internet and is currently working on a Routing Arbiter[6] that will enhance routing among the many Internet service providers.

**Globalization of the Internet.** ARPAnet growth outside the United States started in 1973 with satellite links to England and Norway and grew rapidly after Reseaux IP Europeens (RIPE) started orchestrating TCP/IP activities in Europe in 1989. RIPE coordinates activities of its member organizations, which are European Internet service providers, but it does not operate a network itself. The more than 60 organizations that collaborate via RIPE account for over 650000 interconnected hosts. The EBONE European Backbone operates a high-speed backbone network across Europe to which over 21 national and regional networks connect. It started in 1992 and carries both commercial and research traffic. It provides several high-speed links to the United States and several lower speed links to other non-European countries. The spread of Internet connectivity through Eastern Europe and the former Soviet Union in the 1990s has been especially exciting.

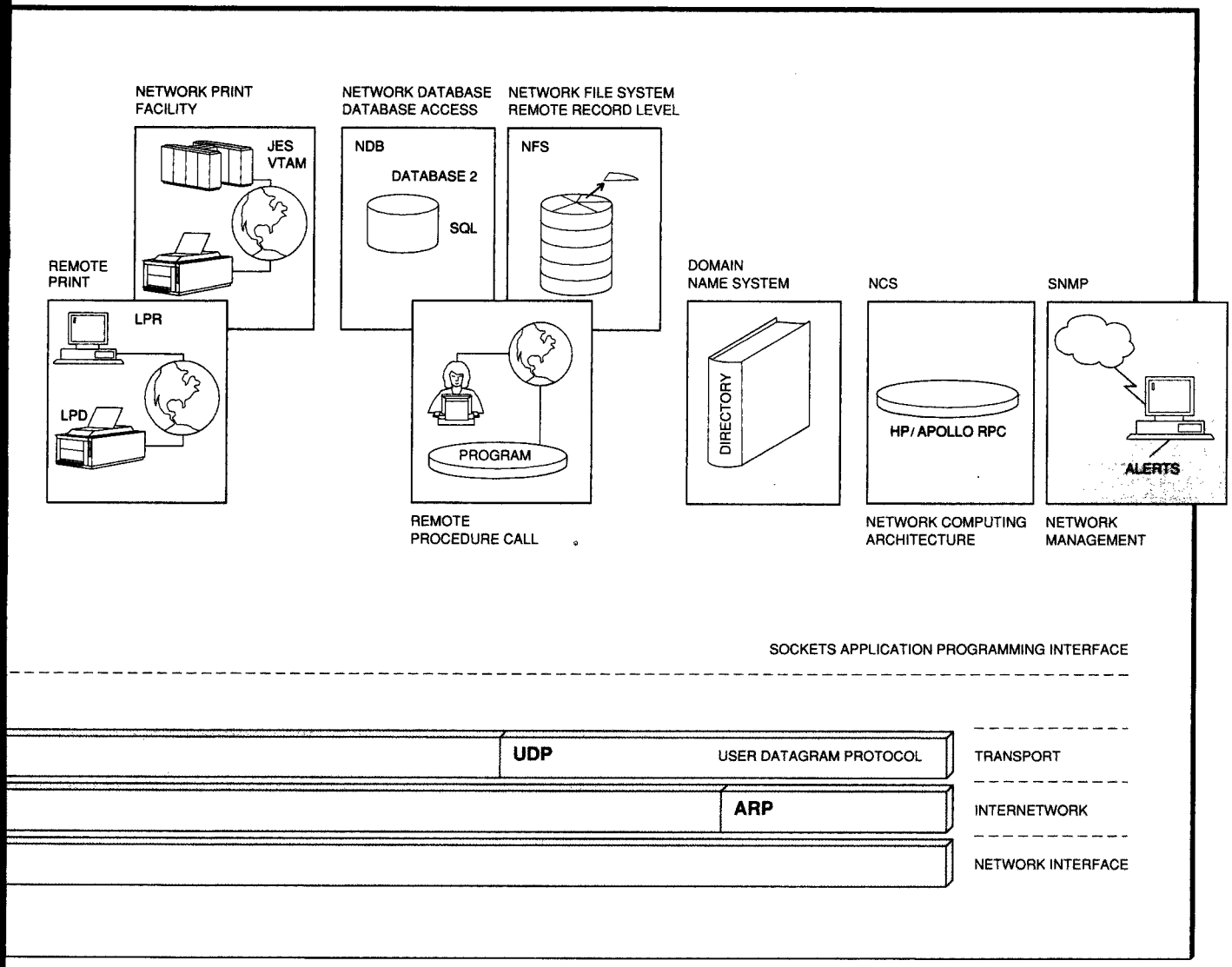
**Figure 1 TCP/IP protocol suite**




Australia has been active in the Internet for many years and has an extensive internal network. Asian and Latin American activity has picked up dramatically since 1993, but connectivity into Africa is still quite limited. A satellite link provides connectivity to Antarctica.

**The National Information Infrastructure (NII).** In 1993 Vice President Gore released a document, *The National Information Infrastructure: Agenda for Action*, that called for the "construction of a seamless web of communications networks, computers, databases, and consumer electronics that will put vast amounts of information at users' fingertips. Development of the NII can help unleash an information revolution that will change forever the way people live, work, and interact with one another" by facilitating advances such as telecommuting, remote education, and remote health care.[7]

The mesh of diverse interoperating NII service providers will bring together users and information providers in an information marketplace that will stimulate the provision of new applications. De-




rived from the Internet, the telephony network, and cable television, it will be a multimedia—data, voice, video, and image—information utility available to everyone at affordable prices. The NII must be robust, dependable, reliable, cost-effective, secure, scalable, and easy to use.

The Cross-Industry Work Team (XIWT), consisting of IBM and about 27 other companies, explores ways to encourage and accelerate deployment of the NII as a data superhighway.[8] They hope to recommend architectural directions that will lead to a secure multimedia network that will satisfy commercial as well as home and research users.

The National Research and Education Network (NREN) is part of the U.S. Government's High-Performance Computing and Communications program studying technology for the NII. The Corporation for National Research Initiatives (CNRI) is coordinating five NREN testbeds for experiments in very-high-speed networking. IBM has been working with the University of Pennsylvania, Bellcore, and the Massachusetts Institute of Technology

(MIT) on the AURORA testbed to experiment on 622-million-bits-per-second (Mbps) facilities with various switching technologies, including IBM's Prizma variable-length frame switch for Packet Transfer Mode.[9]

Many other countries propose to upgrade their public TCP/IP networks in a fashion similar to the NII and to enhance their international connectivity.

**Future users.** As the Internet evolves into the NII, it is attracting new kinds of users—from businesses, homes, and primary and secondary schools—who are placing new kinds of requirements on TCP/IP. Businesses like the way networking allows customers to communicate with them at any time from anywhere. They also find a great deal of information about customer wants and needs from monitoring on-line bulletin boards. IBM and other companies read postings about their own products so that they can answer customer questions on line.

In 1989 the first electronic mail relay was established between the Internet and a commercial electronic-mail carrier. The first commercial Internet services were provided in the United States and Europe in 1991; now dozens of companies around the world provide commercial Internet service. More Internet sites are industrial than academic, and commercial sites are attaching to the Internet faster than academic ones are. IBM's corporate gateway to the Internet passes over 50000 messages per month. Several companies are experimenting with various forms of shopping on the Internet; they let prospective customers access product information across the Internet but typically require some extra security protection, such as a telephone call-back, to confirm purchases made on line. Several techniques are being explored for securing financial transactions across the Internet.

IBM is a member of the CommerceNet Consortium,[10] an industry group that is developing a better model for commercial use of the Internet. Their initial direction is to let companies put catalogs on line from which customers can place orders on line. Electronic catalogs promise to avoid both the cost of printing and mailing paper catalogs and the environmental impact of disposing of them.

Small businesses find that the Internet gives them the global connectivity they could not afford to build for themselves. Consultants in many fields access databases around the world for information relevant to their clients. The home office is especially expanded by Internet access. IBM's Internet Access Kits for Operating System/2* (OS/2*) and for Microsoft Windows** make it easy for business and home users to sign up for Internet service from the IBM Global Network and other service providers around the world.

The first activity of the Global Schoolhouse[11,12] project had students in the fifth through eighth grades in several schools in the United States and England read Vice President Gore's book *Earth in the Balance*, study environmental databases around the Internet, exchange their views via electronic mail, and then hold a videoconference across the Internet to present their conclusions about the Earth's ecology. This project applied to primary and secondary (also known as K–12) education the kind of collaborative on-line work that is so typical of the world's collegiate community today. *K–12 Internetworking Guidelines* (RFC 1709) explains how "a great deal can be done immediately, with relatively few dollars, to provide modern communications systems in and between all schools around the nation." IBM's K–12 Services Offering includes workstation applications with an interface designed for school children and server software for connecting the LAN of a school to the Internet.

Home and K–12 users require especially easy-to-install and easy-to-use systems. Future school children may carry portable units from home to class and then back, all the while maintaining network connectivity. Many home users also would like to take their portable computers shopping, so that they can access databases back home and even search for product reviews or recipes while shopping.

## Requirements for the next generation of TCP/IP

Before discussing work in progress, any of which may change as it evolves toward standardization, we need to explain a few things about the Internet Engineering Task Force (IETF). *IETF Working Group Guidelines and Procedures* (RFC 1603) describes how the IETF, a group open to anyone who wishes to participate in person or on line, organizes working groups to specify new Internet protocols. *Internet Standards Process* (RFC 1602) explains how the specification of a working group matures

from Proposed Standard, to Draft Standard, and finally to Internet Standard. Dr. David Clark has said the process is based on "rough consensus and running code," because multiple interoperating implementations of a specification are required for standardization.

**New applications.** TCP/IP applications already widely available include Simple Mail Transfer Protocol, File Transfer Protocol, Telnet and X Windows System** for remote log-in, bulletin-board news readers, Network File System**, Remote Procedure Call (RPC), Remote Execution Protocol (REXEC), remote printing via the line printer requester and daemon (LPR/LPD) and Network Print Facility, Motif**, Network Database, Simple Network Management Protocol, and many more. Since programmers find the TCP/IP sockets interface especially easy to use, new applications are being developed continuously to satisfy new requirements, for example, networked information retrieval, multimedia, and directory services.

*Networked information retrieval.* As the volume of information available on the Internet has grown to over $10^{12}$ bytes on over 50000 databases, two questions arise: How can a publisher best make publications available, and how can a consumer of information best find useful information?[13]

By convention, to make a file publicly available on a TCP/IP network, it is put in a system that accepts File Transfer Protocol (FTP) requests from anyone who logs in with *anonymous* as the parameter on his or her *user* command (RFC 1635). Thousands of anonymous FTP servers around the world publish millions of files. FTP is available on all IBM platforms, and IBM maintains several anonymous FTP servers (e.g., software.watson.ibm.com). Anonymous FTP is extremely easy to use if the user knows the computer, directory, and file name to access, and many enterprises make such information readily available. However, a user who does not know all of that may find it tiresome to search through many sites and directories to find a likely looking file name.

Archie[14] automates such searching by regularly indexing anonymous FTP archives around the world. Thus it is a guide to anonymous FTP databases. Archie returns a list of sites with directories or file names that match a search string. Then the anonymous FTP can be used to access the files. Although it works, this method leaves the user dealing with directories and file names and depends on the information publisher having selected useful directory and file names. If the user knows what to look for, Archie can tell where to find it.

Gopher[15] helps when the user is not sure in what file the desired information resides. With Gopher, it is not necessary to know the location or file name of an information resource. Information is organized in a hierarchy of nested menus that are much easier and more meaningful to browse than file directories. All the titles in a Gopher server can be searched for a text string, with the resulting hits gathered together in a new menu. One Gopher server can link to another; the set of linked Gophers, called Gopherspace, includes over 5000 servers. Gopher client and server programs are available for most IBM platforms.

**Publication on a network allows faster and less cumbersome distribution and update of information.**

Veronica[16] extends the search throughout Gopherspace, returning a menu of documents from all around the world that satisfy the query. The eight Veronica servers of the Internet index over 15 million titles from more than 5000 Gopher servers and resolve over a million queries per month. Veronica is accessible through most Gopher clients.

World Wide Web (WWW)[17] bases its network information retrieval on hypertext, a technique for linking documents that can be on different computers. The web comprises documents and hypertext links between them; some documents are indexes that can be searched for key words. The home page of a site is the default screen that a web browser displays when it connects to that site. WWW documents can link to Gopher and anonymous FTP files. As with Veronica, the user can get information from many different computers without having to log on to them explicitly. For example, someone using an easy-to-use graphic WWW

browser such as IBM's WebExplorer* for OS/2 or IBM's WebExplorer Mosaic** for Windows to read a WWW publication on Beethoven could use a mouse to click on a hypertext link to hear one of his compositions, or click on another link to see a video clip from a movie about him, each coming from a different computer. WWW software is available for the AIX/6000*, DOS, OS/2, VM, and Windows operating system environments.

The Wide Area Information Server (WAIS) indexes the full text of documents rather than just their file names or titles. *WAIS over Z39.50-1988* (RFC 1625) explains how WAIS uses the American National Standards Institute (ANSI) standard Z39.50 Information Retrieval Service for library applications to allow efficient keyword searches of files hundreds-of-megabytes large. WAIS returns excerpts of the file that include the search string and can select additional documents based on user feedback that indicates which selections were most useful. The selected files can include image, audio, and video as well as text. *Using the Z39.50 Information Retrieval Protocol in the Internet Environment* (RFC 1729) specifies how to implement Z39.50 over TCP/IP.

Many libraries make their catalogs available on the Internet, and Archie and Veronica automate the classical library function of building indexed catalogs. The library and networking communities are engaged in cross-disciplinary research to integrate these different information services, some of which already can interact with others to some extent; e.g., WebExplorer can be an interface to FTP and Gopher as well as to hypertext documents. *A Vision of Integrated Internet Information Service* (RFC 1727) proposes a distributed catalog of information resources that tools such as Archie and WWW can use to find information relevant to a query, and thus entirely hide from the user the process of navigating through the network. A directory service will resolve *Uniform Resource Names* (RFC 1737), which uniquely identify files that can be accessed across the Internet, into *Uniform Resource Locators* (RFC 1738), which tell what application (e.g., FTP or Gopher) can retrieve the file, on which host the file is located, and the name of the file. The goal is to put a global virtual library at your fingertips.

Publication on a network allows faster and less cumbersome distribution and update of information than publication on hard copy or on machine-readable media such as magnetic disks or CD-ROM allows. Already several free academic journals and nonacademic magazines are published on the Internet. However, charging, and therefore security, mechanisms must be deployed for practical commercial publication over networks.

*Multimedia.* The TCP/IP protocol suite has been enhanced to take advantage of the audio and video capabilities that are becoming more common on personal computers and workstations. *Multipurpose Internet Mail Extensions* (MIME) (RFC 1521) specifies how to include nontextual data in Simple Mail Transfer Protocol for playback by the receiver. IBM ships MIME support for OS/2, which can work with additional IBM hardware and software for audio and video capture and playback, and for conversion of audio to text.

As *Network Access to Multimedia Information* (RFC 1614) notes, the hypertext links of the WWW facilitate handling multiple media (hypermedia), since links can point to files that use different encoding techniques. The Library of Congress's WWW exhibition, "Rome Reborn: The Vatican Library and Renaissance Culture,"[18] dramatically demonstrates the power of such hypermedia. Researchers are exploring various ways of indexing nontextual documents, which today are indexed by their titles.

Present use of audio and video files usually requires downloading them in nonreal time from the server to the client, which plays them back locally. Because video files tend to be very large relative to the disk space available on typical client workstations, it would be economical to let many workstations access a video that is stored only on a server. However, playback of audio or video across the network is complicated by the sensitivity of humans to the variation in delay, or jitter, between samples. People are not satisfied with audio or video in which the delay between playback of samples varies significantly from the delay between when those samples were taken. Flows in which the delay between playback of successive samples is satisfactorily uniform are said to be isochronous.

Since TCP makes no guarantee about when it will transmit datagrams passed to it from an application, and since IP makes no guarantees about maintaining a uniform delay between datagrams as it routes them, TCP/IP cannot guarantee isochronous

audio or video transmission. However, the experimental MBONE (multicast backbone) network regularly multicasts audio and video of the IETF meetings around the world satisfactorily by tunneling the multicast packets inside IP through regions where multicast is not yet supported. Over 300 viewers have participated in some multicasts. Although not a practical solution for large-scale use, MBONE research has motivated the Internet community to start defining protocols appropriate for production deployment of isochronous audio and video across the Internet. Videoconferencing among many sites could save the cost and disruption of traveling to meetings but is sensitive to delay as well as to jitter. IBM's Person to Person* product transmits images and video over TCP/IP for shared chalkboard and desktop videoconferencing.

**The parameters of the flow specification for isochronous playback are fundamentally fidelity and latency.**

*Integrated Services Architecture* (ISA) (RFC 1633) proposes extensions that will let the Internet support real-time flows as well as the current best-effort traffic. It also addresses the desire of network operators to control the sharing of bandwidth on a busy link by different kinds of traffic (various applications, protocols, etc.). For real-time flows they propose predictive and guaranteed service levels, where the predictive level provides satisfactory service an adequate amount of the time for less cost than the guaranteed level, which provides the requested service all of the time. Either level requires reservation of resources in IP routers and control of admission to the network. Since this requirement implies special privileges, and possibly costs, for some traffic, authentication is necessary to ensure that only authorized traffic uses the reserved resources. ISA proposes to use the same network-layer protocol for both best-effort and real-time traffic.

A multimedia application can specify its desired quality of service via a proposed new control protocol, Resource Reservation Protocol (RSVP),[19] which would convey the corresponding flow specification (RFC 1363) to the intermediate and end systems. If the capacity necessary to sustain the requested quality of service is available, both in those systems and in the underlying transmission facilities, admission control functions reserve that capacity for this connection. The parameters of the flow specification for isochronous playback are fundamentally fidelity and latency. People are sensitive to jitter in both audio and video streams, both in real time and in playback of recordings. Two-way interactive communication between people, such as in a telephone call, is quite sensitive to delay, whereas one-way communication, such as playing back a video, is less so. People usually can tolerate more loss of fidelity (i.e., a few dropped frames) in video than in audio. A prototype implementation of RSVP demonstrated at the ACM Multimedia '94 Conference provided isochronous transmission of voice traffic despite heavy data traffic.

ISA also proposes ways to let network providers control how traffic from different enterprises, different protocol stacks, and different applications can share busy links. Isochronous IP will retain the model of encapsulating IP packets in the data link control frames of the underlying media in order to remain independent of the link technology. Thus, isochronous IP will be able to take advantage of a variety of isochronous technologies, including future ones as they arise.

To reduce the overhead to determine the service class of a datagram, each datagram could carry an identifier of the flow to which it belongs, and each router on its path could remember the service class associated with that flow identifier. As we describe below, the next-generation IP proposal includes flow identification.

The Audio/Video Transport Protocol discussed later is being defined to specify the encoding technique, timing, and sequence of real-time datagrams, such as audio and video samples. Techniques have been proposed for packetization of H.261, the CCITT (International Telegraph and Telephone Consultative Committee) standard for videoconference codes.

*Directory.* The TCP/IP protocol suite includes several directory applications that permit finding details about something, given a partial description

of it. For example, the *Domain Name System* (DNS) (RFCs 1032–1035) returns the address of a computer (by which IP identifies it), given its name (by which people identify it). Current Internet research on directories includes extensions to DNS, "white pages" for locating a computer that a given person uses, and "yellow pages" for locating a computer on which a given service is available. RFCs 1758, 1712, 1684, 1632, 1609, 1608, 1591, and 1464 contain additional details.

DNS can contain more information than IP address and domain name, such as the descriptor of a well-known service or a mail-exchange record, which gives the address of a computer serving as the "post office" for the electronic mail of the named destination. There are proposals for including public encryption keys and Open Systems Interconnection (OSI) Network Service Access Point (NSAP) addresses in DNS. Of course, DNS must be modified to include the addresses that have been proposed for the next version of IP, discussed below.

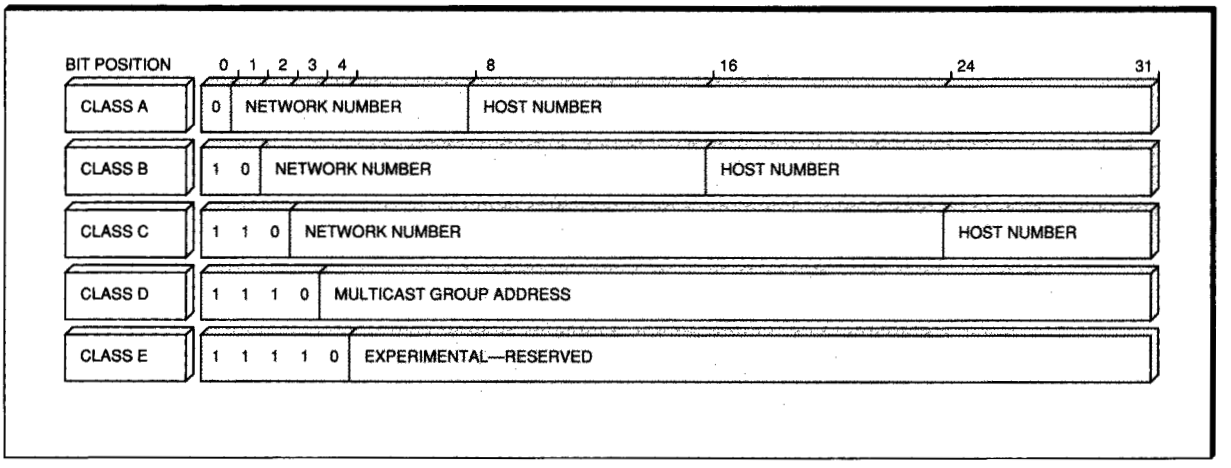
The TCP/IP community has experimented with several forms of white pages directory services, which return information about a person, such as electronic mail address, given a possibly incomplete description of the person (e.g., a partial name, some location information, and maybe some organizational affinity). The *WHOIS* (RFC 954) and *FINGER* (RFC 1288) protocols are useful for finding people, but databases are spotty, and FINGER has security problems.[20] The X.500 directory standard protocol of OSI, which *Technical Overview of Directory Services Using the X.500 Protocol* (RFC 1309) discussed from a TCP/IP perspective, has been used in some administrative domains, such as Paradise with over 1.5 million entries. Its inherently distributed structure makes it a strong candidate for a worldwide directory, but the cost of the clients and the complexity of administering a server have hampered deployment throughout the Internet. The *White Pages Meeting Report* (RFC 1588) concluded that directory services should take advantage of the deployed X.500 directories while facilitating interoperability among various clients and servers. Such a solution should be able to return the individual's security certificate as well, and must deal with the different laws of different countries regarding personal data. To facilitate deployment, a white pages directory should try to satisfy most of the requirements with a simple solution, rather than developing a complex solution to all problems, and make it easy to gather the data for the database, e.g., by building on local directories.

**New transport layers.** Transmission Control Protocol is a connection-oriented, full-duplex, peer-to-peer transport protocol that guarantees error-free, duplicate-free, in-sequence delivery of data given to it. Sliding-window flow control guarantees that the sender never transmits more traffic than the receiver has buffers to hold. The receiving TCP acknowledges error-free segments; the sending TCP retransmits unacknowledged segments after waiting a period based on round-trip time (RTT) and variance of RTT, both of which it constantly re-estimates. TCP throughput of over 700 Mbps has been achieved between supercomputers over a fast channel. *User Datagram Protocol* (UDP) (RFC 768) is a connectionless alternative to TCP useful for certain kinds of applications, such as query response with a file server. UDP was invented to transport packetized voice without the error checking or retransmissions of TCP.[21]

*TCP enhancements.* The TCP window size is the maximum number of bytes of unacknowledged data that TCP can have outstanding on one connection at a time. If the window size is less than the product of the RTT and bandwidth of the connection, that connection will have to quit transmitting while it waits for acknowledgments, and thus will not be able to utilize the medium fully (although several such connections together could utilize the medium fully). On today's typical facilities, the 64-kilobyte window of TCP is large enough to avoid such waits, but a single application-to-application connection with TCP as originally defined might not be able to fully utilize some facilities proposed for the future; for example, a gigabit-per-second link over 47 kilometers long would have a round-trip delay-bandwidth product greater than 64 kilobytes.[22] Therefore, *TCP Extensions for Long, Fat Networks* (RFC 1323) defines a backwardly compatible, negotiable option for larger windows that lets even a single TCP connection fully utilize an around-the-world link at a gigabit-per-second rate or faster. AIX/6000 4.1 includes RFC 1323 support for large windows.

*Audio/video isochronous transport.* An IETF work group is defining a real-time transport protocol for isochronous connections that will ensure that the sender does not overrun or starve the receiver. For example, the receiver of video must obtain the next screen image every refresh cycle (typically every

**Figure 2 IP Version 4 class-based address structure**




1/30th of a second). If many screen images arrive too much sooner than that, the receiver may lack buffers for them. If frames arrive too late, the movement in the image will not be smooth. The sender must transmit pictures at almost exactly the rate the receiver should display them. Rate-based flow control, rather than the sliding-window flow control of TCP, smooths the entry of traffic into the network to match the capacity of the receiver. On connection setup, the sender and receiver negotiate how much bandwidth the connection will use. Techniques such as leaky bucket can regulate the access to that bandwidth and may do other things to smooth the input. Later we discuss how the next generation of IP will handle audio and video isochronously.

**The new Internet Protocol.** The Internet Protocol is a network-layer protocol that makes a best effort to deliver datagrams, relying on TCP to guarantee delivery with transport-layer recovery. An IP router is a computer that interconnects multiple physical links and decides on which ones to send, or forward, IP datagrams. The routing tables of IP are updated dynamically; in conjunction with TCP, IP can nondisruptively switch a connection from a broken to a working link. IP forwarding rates in the tens of thousands of packets per second are common.

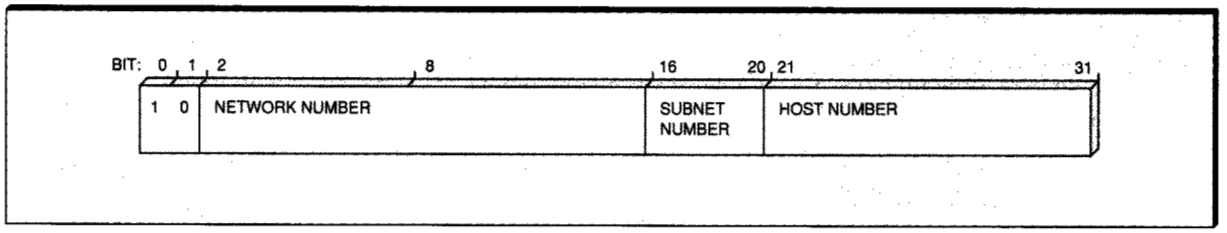
A network is said to be congested when the offered load so far exceeds the engineered load that the carried load decreases. Before 1989 the Internet sometimes experienced congestion; then congestion control algorithms that all hosts are now required to implement were designed. These algorithms have alleviated Internet congestion. The TCP/IP congestion control algorithm can be summarized as follows:

- IP discards packets for which it has no buffers.
- When the retransmit timer for the discarded data expires, TCP reduces its window size to one packet, retransmits that one packet, and increases the retransmission timer to avoid further congestion.
- When data again are acknowledged, TCP increases the window size quickly until it reaches half the size of when a packet last had to be retransmitted, and then increases it slowly.

Additional refinements, such as fast retransmit and fast recovery algorithms, have further improved the bandwidth utilization of IP.

Each adapter in an IP network has a 32-bit address. In the most commonly used three classes (A, B, and C), the address consists of a network number and a host number; all end systems, whether personal computers (PCs) or mainframe computers, are called hosts in a TCP/IP network. As Figure 2 illustrates, the lengths of the network number and host number depend on the settings of the high-order bits. Thus, IP addresses can accommodate a few huge networks, many small networks, and a moderate number of moderately large networks (for a

**Figure 3 Subnet addresses—this class B address has five bits of subnet address**




theoretical total of over 3.7 thousand million end systems on over 2.1 million networks).

What do we mean by a network in this context? With respect to IP network address classes, a network is a set of devices that can communicate with one another directly via a data link control or logical link control protocol, i.e., without going through an IP router. Thus, an unbridged LAN is a network, and a set of bridged LANs is a network. For example, suppose an organization has 1000 PCs and 11 unbridged LANs arranged as 10 LANs each with 100 PCs, and each is attached to the eleventh backbone LAN via one of 10 IP routers. Then the organization could use 11 class C network numbers. Even though there are more PCs than one class C network can address (i.e., 254), the organization need not have a class B address.

However, many organizations with many LANs have one class B network number for their entire organization, instead of using a different class C network number for each LAN. If they advertise only this one class B address to other enterprises, they can avoid carrying transit traffic from those enterprises across their networks. To route among their own LANs, they implement subnetting (RFCs 950 and 1219), which allows routing among different LANs of the one network number to be based on some bits of the host number (Figure 3). Furthermore, routers outside this organization need only one routing table entry for the class B network number, rather than an entry for each subnet, which significantly reduces the size of routing tables in the heart of the network.

*New requirements.* The number of networks in the Internet has been approximately doubling annually. In 1990 some members of the Internet community started to think about the effect of this rapid growth on the IP address space. By April 1992 less than 3 percent of the IP network numbers had been assigned, but over 40 percent of the popular class B network numbers had been assigned. It appeared that the class B address space would be completely consumed by 1994. Furthermore, the size of the address tables in the routers in the heart of the Internet had grown dramatically. Whereas IP routers of campus and regional networks can keep their tables small by sending all traffic destined for outside their enterprise to a default router of their service provider, the routing tables in the heart of the network must have entries for all destination networks (but not subnets). By 1993 these tables had grown to have entries for about 25000 networks, and network administrators were concerned that further growth could impact the network badly.

In RFCs 1518–1520, Yakov Rekhter of IBM Research and others proposed Classless Inter-Domain Routing (CIDR), which allowed assignment of blocks of contiguous class C addresses in a way that made them more attractive to large enterprises and also reduced the growth of the routing tables. CIDR lets a power-of-two-sized cluster of contiguous class C addresses be assigned as one network number by associating with each address a bit map that says which bits represent network number and which represent host number. It lets a more flexible granularity be applied to the address hierarchy than the class-based model did. Furthermore, this aggregation lets a single entry in each routing table advertise reachability to many networks. CIDR has been deployed extensively already and has reduced the depletion of class B addresses and the growth of routing tables; 2246 CIDR-based routes had replaced 9671 class-based routes as of December 1, 1994. IBM helped update the Border Gateway Protocol to handle CIDR.

As described in *The Recommendation for the IP Next Generation Protocol* (RFC 1752), estimates of

the life expectancy of the current IP address space, after deployment of CIDR, range between the years 2005 and 2011. Even if some shift in paradigm reduces this to the year 2000, there is adequate time to develop, test, and deploy a replacement for the current IP, albeit with little available slack time.

In late 1993 the IETF chartered the IP: Next Generation (IPng) Directorate to assess the situation and recommend a direction for updating the Internet Protocol. In December 1993 the directorate solicited requirement statements from the Internet community (RFC 1550). IBM and 20 other enterprises, spanning a wide range of interests beyond the normal IETF constituency, submitted requirements, which were published on line (RFCs 1653, 1667–1683, and 1686–1688).

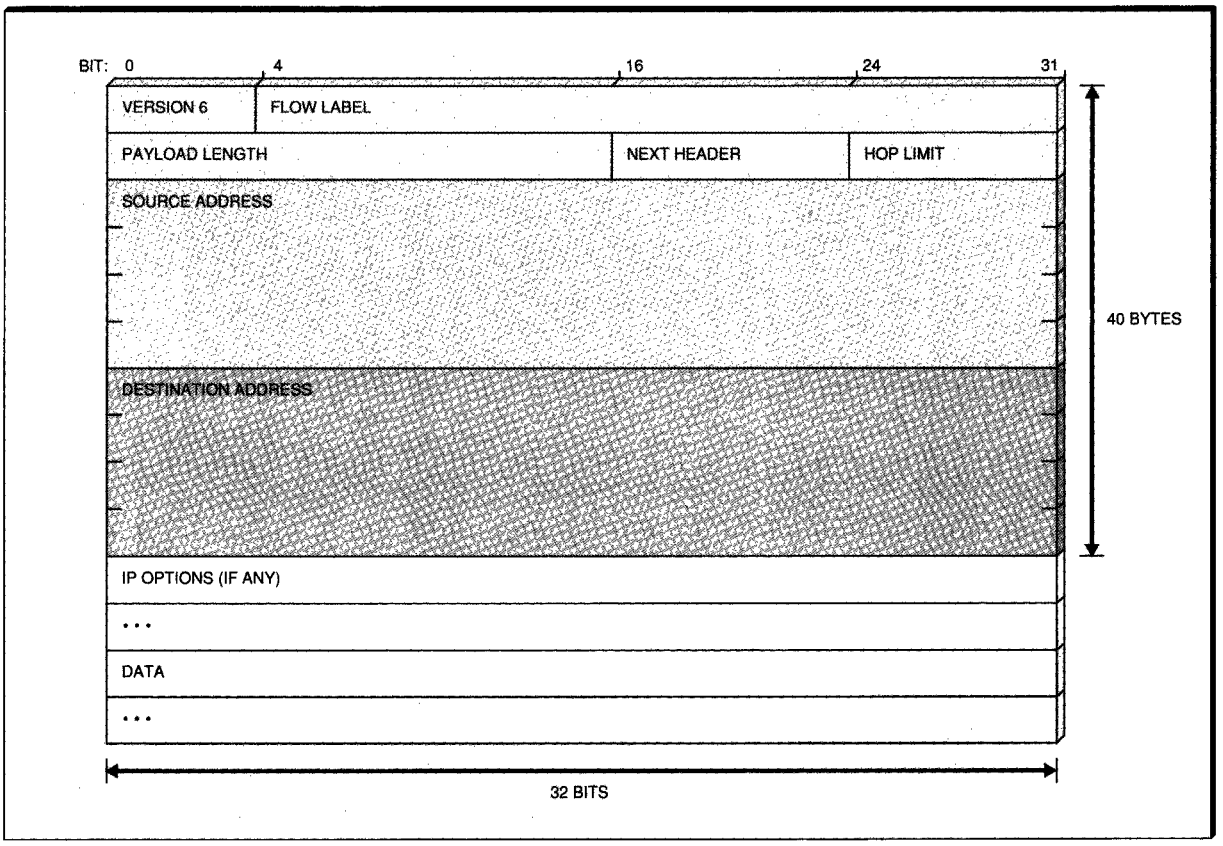
The most frequently cited requirement for IPng has been expansion of the address space. However, we and others observed that end users whose computers execute the current IP (version 4) will not be adequately motivated to upgrade to IPng unless it adds more value than merely enlarging the address space. After a period of on-line debate by anybody who cared to speak up, the directorate settled on the *Technical Criteria for Choosing IP the Next Generation* (IPng) (RFC 1726). Proposed shifts in the use of IP addresses for such things as electric meters, plus the desire for greater structure in, and concomitant sparse allocation of, the addresses called for a dramatically larger address space: at least $10^9$ networks, preferably $10^{12}$, and at least $10^{12}$ end systems, preferably $10^{15}$. IPng should allow encapsulation of its own packets or the packets of other protocols. IPng should have service classes that distinguish a variety of flows, such as isochronous audio and video. Multicast addressing and automatic address assignment, particularly for mobile use, are required. IPng must provide authentication and encryption. Encapsulation, multimedia, multicast, mobility, type-of-service, configuration automation, and security have been used, to some extent, with the current IP. For example, AIX/6000 4.1 supports multicast addresses. The requirement is to integrate them more fully with one another and with the rest of the protocol suite, and in some cases to improve them based on observations of experimental prototypes. Furthermore, IPng should continue to provide the virtues of the current IP: robustness, independence from transmission media, high performance, topological flexibility, extensibility, datagram service, globally unique names, a built-in control protocol, and freely available standards. Moreover, a simple transition plan must allow easy migration from, and coexistence with, the current IP.

*The IPv6 recommendation.* Several work groups proposed new protocols for IPng. The IPng directorate evaluated them against the list of accepted requirements and concluded that all were lacking, but that *Simple Internet Protocol Plus* (SIPP) (RFC 1710) came closest. After the SIPP work group made several changes to their proposal, such as increasing the address length to 128 bits, the directorate decided SIPP should be the base to which would be added ideas from the other proposals and from further work. The proposed solution is known as IP version 6 (IPv6).

The basic IPv6 protocol is a straightforward extension of IPv4, designed to keep the best parts of IPv4 and improve the rest. It uses 16-byte addresses, allowing both more addresses and more structure in the addresses. To allow faster forwarding by IP routers, the basic 40-byte header of IPv6 (Figure 4) is simpler than the 20-byte header of IPv4 (Figure 5). Several infrequently used fields of the IPv4 header are moved to optional extension headers to speed up the normal-case processing. As seen in Figure 5, each IPv4 header is 20 bytes, including two 4-byte addresses. In IPv6, payload length replaces header length (HLEN) and total length; flow label replaces type of service (TOS); fragment header replaces identification, flags, and fragment offset; hop count replaces time-to-live; next header replaces protocol; and header checksum is removed. To facilitate adding options in the future, the option header can specify what a router that does not support the option should do with the datagram, e.g., ignore the option and continue processing, silently discard, or discard and return an error message. Header compression lessens bandwidth cost of small packets on expensive serial links. Modifications are necessary for some other protocols in the TCP/IP family that use IP addresses; for example, Thomas Narten of IBM and others are defining a Neighbor Discovery Protocol that, among other things, provides the mapping between IP addresses and physical network addresses for IPv6 that Address Resolution Protocol (ARP) provides for IPv4.

All IPv6 nodes are expected to perform *Path MTU (Maximum Transmission Unit) Discovery* (RFC 1191) to dynamically find the largest physical frame size supported by all links along the route.

**Figure 4 IPv6 header—each IPv6 header is 40 bytes, including two 16-byte addresses**



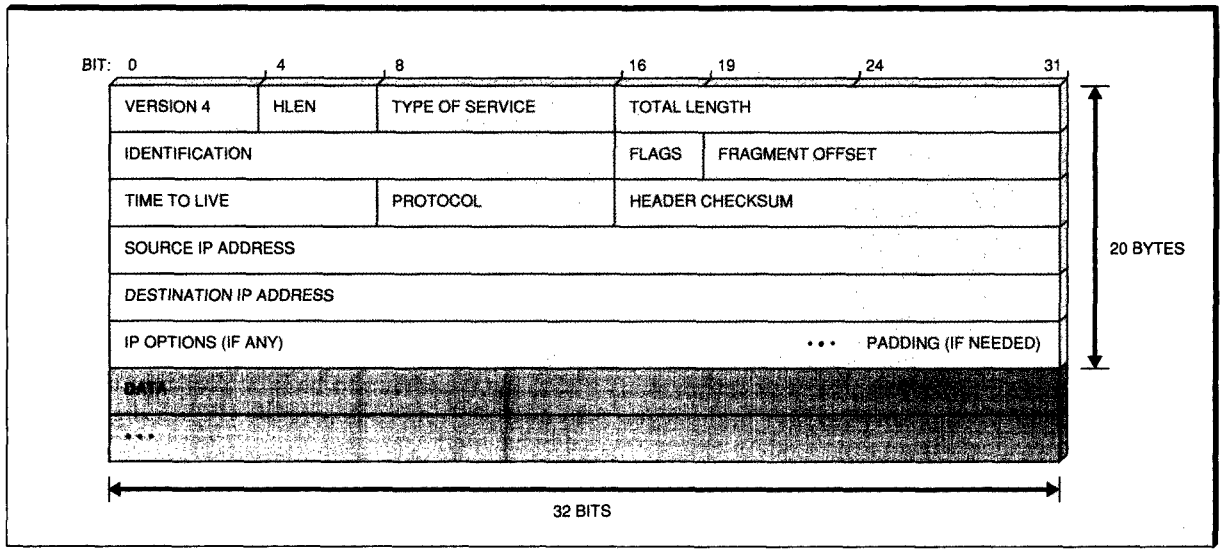

IPv6 routers no longer will fragment datagrams; only source end systems will fragment datagrams, based on Path MTU Discovery feedback. This should allow much more efficient utilization along multihop routes and let fragmentation become a rarely used option. IPv6 routers need to examine only the hop-by-hop options such as source routing.

If the information encapsulated by the IPv6 datagram exceeds 64 kilobytes, the payload length field of the datagram is set to zero, and the true length is specified in an option header. IPv6 renames the IPv4 time-to-live (TTL) field the hop limit, because in practice it is the number of routers through which a datagram can pass, rather than a measure of time, before the datagram is assumed to be in a routing loop and discarded.

Rekhter and Li[23] have proposed a structure similar to CIDR for unified management of the IPv6 address space. The high-order bits of the IPv6 address are a Format Prefix that specifies the address type. Types defined so far include IPv4, provider-based, OSI NSAPs, Novell IPX** (Internet Packet Exchange), local use, and multicast. The addressing structure balances the conflicting goals of route optimization, routing-algorithm efficiency, ease and administrative efficiency of address assignment, and considerations for which addresses are assigned by which addressing authority. Rekhter and Traina[24] have proposed Inter-Domain Routing Protocol for IPv6 as an extension to Border Gateway Protocol that better supports CIDR-based aggregation, policy-based routing, and reachability information for multiple protocols.

The hierarchical address structure allows route aggregation to control the size of routing tables and to limit the amount of topology information percolated through the entire network. The aggregation cannot be achieved significantly unless the

**Figure 5 IPv4 header**




addresses are assigned to sensibly reflect the underlying topology, e.g., to reflect the relatively few transoceanic links. If addresses are associated with service providers, it must be easy for users to change addresses so that they will not be deterred from changing service providers whenever they wish.

When an IPv6 host changes service providers or otherwise attaches to a network, it will be able to dynamically request that autoconfiguration dynamically assign it an IP address. The larger address space makes it more practical to reserve a pool of addresses for temporary assignment. Although addresses are leased for a limited time, a host can renew its lease on an address. It can use *Dynamic Host Configuration Protocol* (RFC 1541) to obtain all the other configuration parameters, if it wishes, and will be able to automatically update the Domain Name System. This plug-and-play characteristic can be made secure by the IPv6 security headers.

IPv6 provides optional packet-level authentication and integrity protection to all users. A standard encryption method will be recommended for confidentiality, but to avoid conflict with the different laws around the world, alternative encryption methods may be used instead. Key management and security techniques are both independent of the IPv6 header, so other techniques can be substituted easily in the future. Experimental implementations exist. A new key management work group has been chartered to define the infrastructure that will make use of IPv6 security practical across the Internet. An authentication header assures authenticity of the source and integrity of the content of a datagram. Some, but not all, authentication algorithms that could be used with this header also can provide nonrepudiation. A privacy header puts encrypted data in its payload field. Either the entire datagram or just the transport-layer protocol-data unit can be encrypted. The privacy header can be used end to end, or merely over part of the route, e.g., between barriers, or so-called firewalls, which we discuss later.

IPv6 datagrams can include a flow label that identifies a sequence of datagrams for which special handling by IP routers is requested. Information either carried in the datagram header or conveyed to the IP routers by a resource-reservation protocol, such as RSVP, can specify the requested handling to the routers. It has been proposed that IP datagrams should be able to include a traffic class that specifies either the time-sensitivity (priority) of flow-controlled traffic (such as file transfer), or the drop-priority of nonflow-controlled traffic (such as isochronous video). If an audio or video packet is on a flow for which enough resources have been

reserved to probably provide, but not absolutely guarantee, acceptably low jitter, any packet that has been delayed long enough to disrupt the smooth playback should be discarded. Some datagrams can be marked as more discardable than others; e.g., throwing away a few frames may disrupt some video less than it does some audio. In contrast, flow-controlled traffic never should be discarded merely because it has sat in a forwarding queue for a long time, but some kinds suffer from such delay more than other kinds. For example, interactive Telnet traffic can be marked as being more time-sensitive than file-transfer traffic.

IPv6 autoconfiguration allows mobile hosts to be assigned new temporary addresses when they roam from one network to another. The IPv6 authentication header is especially important for mobility, because the attachment lacks physical security. By streamlining the IPv6 header, and assuring fixed offsets for its fields, header processing will be simple enough for even portable units with limited capabilities.

The transition from IPv4 to IPv6 can proceed one node at a time because IPv6 has a Format Prefix for IPv4 addresses. During the transition, all IPv6 nodes also will be assigned IPv4 addresses. After the Domain Name System and other necessary services have been upgraded to support IPv6 addresses, and until all IPv4-compatible addresses have been assigned, IP end systems and routers can migrate to IPv6 one step at a time by installing software that supports both IPv4 and IPv6. Existing IPv4 hosts and routers can be upgraded to be dual-IPv4/IP46 nodes independently of all other hosts and routers, and can continue to use their already assigned IPv4 address. Likewise, new IPv4/IPv6 hosts and routers can be installed independently of all other hosts and routers. IPv4/IPv6 routers on the edges of an IPv4-only routing infrastructure can tunnel IPv6 traffic through it by encapsulating IPv6 datagrams inside IPv4 datagrams.

If any IPv4-only nodes remain after all of the IPv4-compatible addresses have been assigned, they will not be able to communicate with nodes that are assigned IPv4-incompatible addresses. By the time all IPv4 addresses have been assigned, each node will need to have upgraded to IPv6 in order to communicate with nodes coming on line thereafter without IPv4-compatible addresses. The capabilities that IPv6 adds will likely motivate people to upgrade to it before the IPv4 address space is exhausted.

Many details must be ironed out to turn the IPv6 direction into a set of complete protocol specifications. The IPng directorate proposes to bring the IPv6 RFCs to proposed standard status with all deliberate speed. However, several other protocols (such as Domain Name System and File Transfer Protocol) must be changed to work with IPv6 addresses, and an updated network management and administration infrastructure (address autoconfiguration servers, security key servers, etc.) must be deployed before it will be practical to use all IPv6 capabilities in a production mode. Nevertheless, there appears to be plenty of time for users to migrate to IPv6 before the IPv4 address space is exhausted.

**New network interfaces.** The transit facilities of the Internet include wide-area links as fast as 45 Mbps (T3), as well as many slower links. Just about every kind of access link is supported somewhere in the Internet: Fiber Distributed Data Interface (FDDI), frame relay, switched multimegabit data service (SMDS), all the 802.2 LANs, X.25, mobile radio, satellite links, serial links with the serial line Internet Protocol (SLIP) or the Point-to-Point Protocol (PPP), and many more.

*Mobility.* TCP/IP has been used over mobile connections since 1977 on the Bay Area Packet Radio Network. More recently KA9Q[25] software has allowed thousands of amateur radio enthusiasts to connect to the Internet, and several vendors provide wireless TCP/IP service. The next step is to let mobile TCP/IP users roam from one wireless network to another.

Because the IP address includes a network number that identifies the physical network to which the host attaches, a host must change its IP address if it moves to another network. Other hosts cannot communicate with the moved unit until they learn its new address, typically after the DNS has been administratively updated. A TCP connection between the hosts is lost if one of them moves to a new network and changes its IP address.

*IPng Mobility Considerations* (RFC 1688) proposes to maintain communication with a host as it roams among wireless networks by assigning the mobile host an unchanging home address to which are

dynamically associated transient "care-of" addresses that reflect its current point of attachment. Datagrams addressed to the home address are intercepted by a home agent, which forwards them to the current care-of address. IPv6 facilitates this by providing secure, dynamic address assignment from a pool of addresses.

*Asynchronous transfer mode (ATM).* Increasing availability of high-bandwidth, low-error fiber optic transmission facilities is making gigabit-per-second networks conceivable. Economies of scale may make it economical to replace several fast links between a pair of routers with one fiber optic link that offers greater bandwidth than the replaced links in aggregate.

For today's nonisochronous applications, it is adequate to let IP encapsulate the datagrams in ATM headers and pass them down through an ATM adapter, from which they will be switched through the ATM network. *Classical IP and ARP Over ATM* (RFC 1577) specifies how to do this. IBM has shipped RFC 1577 support in AIX/6000.

Although fast and useful for many applications, RFC 1577 support does not exploit the isochronous capabilities of ATM. For that, the isochronous applications will need to specify their flow requirements, such as bandwidth and jitter values, to isochronous transport and network layers, which will translate those requirements into ATM quality-of-service requests. The Resource Reservation Protocol conveys the flow specification to the IP routers, which could use these parameters to request appropriate levels of service from the underlying ATM network.

If a pair of end systems could communicate across the Internet at a gigabit per second, they could possibly support some new categories of applications that would require sustained bandwidth in excess of today's 45 Mbps. Interprocess communication between remote supercomputers, called a global backplane, may require 100 Mbps bandwidth and under 100 millisecond delay. Isochronous high-quality voice and video multimedia conferencing could require 100 Mbps bandwidth and under 100 millisecond delay with no perceptible jitter.[26] Remote medical image viewing and remote graphic visualization of supercomputer results could require 30 updates per second of a million pixels of 24 bits each, or 720 Mbps, again, isochronously. To be interactive, such applications also would require low delay.

*Internet Architecture Extensions for Shared Media* (RFC 1620) discusses alternatives for optimizing paths when multiple IP subnets are defined on one shared medium, such as a large frame relay or ATM network. IBM has proposed extensions to IP that enable full utilization of ATM switching.[27]

**Network security threats often are categorized as attacks on confidentiality, integrity, or service availability.**

**System requirements.** Users have certain expectations that cut across protocol layers, such as administration, security, and performance. We already have discussed changes in address administration that IPv6 will bring about.

*Security.* An enterprise can run a private TCP/IP network quite securely by disallowing links to untrusted sites, encrypting packets that leave the premises of the enterprise, and physically securing the premises against untrusted people. The questions being addressed by the Internet community involve how to maintain security without these restrictions.

Security support should address the threats to be defended against in the least expensive way that provides adequate protection, where adequacy is measured in terms of the likelihood and potential cost of compromise. Network security threats often are categorized as attacks on confidentiality, integrity, or service availability. To support confidentiality, a system must not disclose sensitive data to anyone unauthorized to see it. To support integrity, a system must not allow any unauthorized change to the data. To protect against denial of service, the system must keep anyone from inappropriately reducing the service level to other users.

*Report of IAB Workshop on Security in the Internet Architecture* (RFC 1636) summarizes discussions of experts invited by the Internet Architecture Board to recommend future directions for Internet security. To allow open access to a secure network, the network must first authenticate that a packet comes from whom it claims to come, and then ensure that the authenticated user does only what he or she is authorized to do. The workshop concluded that existing authentication techniques, such as Kerberos,[28] are satisfactory but need to be applied more (for example, in routing and management flows) to defend against attacks on the network infrastructure. Kerberos was one of the many distributed computing innovations designed as part of Project Athena**, an MIT project done in partnership with IBM and Digital Equipment Corporation. Since the new quality-of-service categories for IPv6, especially real-time flows that consume expensive network resources, could be used to deny service to others, their invocation should be authenticated and authorized. Experience with privacy-enhanced mail suggests that slow adoption of security may result from problems in procurement and deployment, as opposed to technology: lack of a network-wide infrastructure for authentication, and export restrictions on encryption technology. The workshop concluded that security would be deployed more easily throughout the Internet if built into the IP layer. The workshop proposed separating authentication from encryption, and possibly using a symmetric key technique.[29] Key management must be deployed across the Internet to make security practical.

Firewalls are a pragmatic way for an enterprise to build a defensible security perimeter around unsecured end systems. The Socks[30] protocol lets applications such as Telnet, FTP, and World Wide Web send traffic through a firewall securely. IBM's NetSP Secured Network Gateway and the Firewall Services of the IBM Global Network let customers have secure access to the Internet.

*Performance considerations.* IBM has been closely studying the performance of its TCP/IP products and developing models to analyze product performance and predict the performance of network configurations. Performance is especially important with the emergence of high-speed networks, multimedia applications, and distributed client/server applications. With so many components, their complexity, the increasing size of communication networks, and the interoperability among multivendor hardware and software, performance modeling is needed to analyze network performance and to do capacity planning. Modeling also can be used to design and implement high-performance products to keep up with the emerging network technologies. Performance modeling work may lead to tremendous savings in resources and may shorten development cycles.

IBM has developed various advanced mathematical techniques to model network performance. These techniques are implemented in a tool, IBM Network Performance Modeler, that runs under OS/2. Unlike discrete-event simulation techniques, the tool takes little time to calculate the performance statistics of complex network configurations. The tool has an easy-to-use graphical interface that allows fast modeling of a product or network. By developing a model for specific network nodes, a library of basic models is built from which network models can then be constructed. The portability and integration of models may help solve the problem of interoperability. As expected, the performance statistics are not as accurate as those of simulation; however, this limitation may be acceptable to most users, because the savings in execution time could be significant.

The models being developed are portable to various platforms with minor modifications. Performance models exist for the following: TCP/IP stack, File Transfer Protocol, Telnet, TCP sockets, UDP sockets, and remote procedure call—all on the MVS and OS/2 platforms. The implementation of the TCP/IP stack and each of the above applications on each platform has been carefully studied to capture the main elements that contribute to the overall system performance. Every model has its own set of parameters that are initialized before execution. Such parameters include the network frame size, the maximum TCP window size, and buffer sizes. The tool provides performance models for network components such as token ring, token-ring adapters, Ethernet, Ethernet adapters, System/390* parallel and serial channels, 3172 I/O controller, System/390 file server, OS/2 client workstation, and others. A network configuration is built using these components. The input to this configuration is a workload consisting of the type of transaction used (Telnet, File Transfer Protocol, etc.), the amount of data to be transferred between the source and the destination, the distribution of interarrival times of transactions, and the mean number of arrivals per unit of time. For each workload

the tool calculates the mean transaction time and the utilization of each network resource.[31]

IBM also has developed a performance tuning guide[32] to help customers optimize their TCP/IP networks. The guide explains performance objective prioritization, available monitoring tools, and key tuning parameters. It is designed for system programmers installing TCP/IP, performance analysts tuning system parameters, application programmers seeking the available options and their tradeoffs for optimal performance, system planners looking for information on capacity planning, and network specialists needing insight into how the IBM products work and how different parameters affect their performance.

IBM has been enhancing OS/2 and MVS TCP/IP products for better performance. Clark et al.[33] showed that TCP is not the source of the packet overhead processing, and that it could support very high speeds if properly implemented. Indeed, IBM improved OS/2 TCP/IP performance significantly by enhancing use of operating system services such as memory allocation, interrupt handling, and scheduling. At the application and stack layers, performance improved even more by reducing the number of data moves, optimizing the checksum calculation, and implementing TCP header prediction.

*Multiprotocol support.* As indicated in IBM's Open Blueprint*,[34] applications from other protocol stacks can run over TCP/IP; for example, NetBIOS, Systems Network Architecture (SNA), and OSI applications. Likewise, SNA can transport sockets applications. Data link switching lets a router encapsulate SNA and NetBIOS traffic in TCP/IP packets. Furthermore, data link layers such as IEEE 802.2, frame relay, point-to-point protocol, and ATM can carry multiple protocol stacks simultaneously (TCP/IP, OSI, SNA, etc.).

## IBM directions for TCP/IP

IBM intends to expand further on the significant role it has had in the Internet by participating in the IETF, by developing TCP/IP products, and by providing TCP/IP service. IBM will continue to support a broad spectrum of TCP/IP users, from large commercial businesses to home users. Although some requirements—such as security, reliability, availability, serviceability, and high performance—seem especially important for businesses, noncommercial users want them, too. Likewise, ease of use and ease of installability are especially important for potentially less experienced home users but are also valuable to the largest corporations.

IBM intends to provide TCP/IP with appropriate consistency across a broad spectrum of platforms, both end systems (from PCs through mainframes), and intermediate systems (such as hubs and IP routers). Furthermore, the IBM Global Network provides Internet service in many countries. As John R. Patrick, vice president of communications for IBM Networked Application Services, said, "This is a tremendous opportunity for IBM because IBM had, in fact, created much of the technology behind the Internet."[35]

## Conclusion

TCP/IP already supports over 4.8 million hosts with up to 45-Mbps transit facilities and a wide variety of access links. Its architecture is evolving to support many tens of millions of end users, gigabit-per-second backbone links, and up to gigabit-per-second communication between individual pairs of end users. These three aspects of growth impose different requirements, but probably will evolve simultaneously. IBM is actively involved in the efforts of the Internet community to enhance TCP/IP to satisfy these requirements.

In the 21st century Internet service may be nearly as pervasive, useful, and easy to use as telephone service, enabling people to work with information resources and processors regardless of their locations. Enterprises and individuals can feel confident embracing TCP/IP now, because TCP/IP architecture is evolving to handle larger, faster networks that will continue to support today's applications, transmission media, and protocols, as well as new applications, transmission media, and protocols.

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of Berkeley Software Distribution, X/Open Co., Ltd., Microsoft Corporation, Massachusetts Institute of Technology, Sun Microsystems, Inc., Open Software Foundation, Inc., Netscape Communications Corp., or Novell Corp.

## Cited references and notes

1. *TCP/IP Tutorial and Technical Overview,* GG24-3376, IBM Corporation (December 1992); available through IBM branch offices.

2. D. E. Comer, *Internetworking with TCP/IP Vol. 1: Principles, Protocols, and Architecture*, 2nd edition, Prentice Hall, Englewood Cliffs, NJ (1991).
3. L. G. Roberts and B. D. Wessler, "Computer Network Development to Achieve Resource Sharing," *Proceedings of the Spring Joint Computer Conference* **36**, AFIPS Press, Montvale, NJ (1970), pp. 543–549.
4. V. G. Cerf and R. E. Kahn, "A Protocol for Packet Network Interconnection," *IEEE Transactions on Communications* **COM-22,** No. 5, 637–647 (May 1974).
5. S. J. Leffler, M. K. McKusick, M. J. Karels, and J. S. Quarterman, *The Design and Implementation of the 4.3BSD UNIX Operating System*, Addison-Wesley Publishing Co., Reading, MA (1989).
6. S. R. Harris, "A New Architecture for NSFnet," *Link Letter* **7**, No. 1, 14–16 (July 1994).
7. National Telecommunications and Infrastructure Administration, *The National Information Infrastructure: Agenda for Action*, U.S. Department of Commerce, Washington, DC (1993).
8. M. Dziatkiewicz, "XIWT: Can It Make Convergence as Simple as ABC?" *America's Network* **1**, No. 1, 34–38 (February 15, 1994).
9. D. D. Clark, B. S. Davie, D. J. Farber, I. S. Gopal, B. K. Kadaba, W. D. Sincoskie, J. M. Smith, and D. L. Tennenhouse, "An Overview of the AURORA Gigabit Testbed," *Proceedings of INFOCOM'92*, IEEE Communications Society, New York (1992), pp. 569–581.
10. B. Metcalfe, "Internet Goes Commercial with CommerceNet," *InfoWorld* **16**, No. 16, 62 (April 18, 1994).
11. L. Nicastro, "The Internet Goes to the Head of the Class," *Network Computing* **4**, No. 8, 28–30 (August 1993).
12. R. Taylor, "Brave New Internet," *Internet World* **5**, No. 6, 36–42 (September 1994).
13. R. Pool, "Beyond Databases and E-Mail," *Science* **261**, No. 5123, 841–843 (August 13, 1993).
14. A. Emtage and P. Deutsch, "Archie—An Electronic Directory Service for the Internet," *USENIX*, USENIX Association, Berkeley, CA (Winter 1992), pp. 93–110.
15. J. Rosenfeld, "If You See It and Want It, How Do You Get It? Help for Hunters on the Internet," *Computer Librarian* **14**, No. 6, 46–48 (June 1994).
16. S. Foster, *Frequently-Asked Questions (FAQ) About Veronica*, Uniform Resource Locator gopher://veronica.scs.unr.edu:70/11/veronica (January 13, 1995).
17. T. Berners-Lee, R. Cailliau, J.-F. Groff, and B. Pollermann, "World-Wide Web: The Information Universe," *Electronic Networking: Research, Applications, and Technology* **1**, No. 2, 52–28 (Spring 1992).
18. *Rome Reborn: The Vatican Library and Renaissance Culture*, Uniform Resource Locator http://lcweb.loc.gov/homepage/exhibits.html (1994).
19. L. Zhang, S. E. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A New Resource ReSerVation Protocol," *IEEE Network* **7**, No. 5, 8–18 (September 1993).
20. W. R. Cheswick and S. M. Bellovin, *Firewalls and Internet Security*, Addison-Wesley Publishing Co., Reading, MA (1994).
21. B. Aboda, *The Online User's Encyclopedia*, Addison-Wesley Publishing Co., Reading, MA (1993).
22. (64 kilobytes) * (8 bits/byte) * (.6 * 300,000 km/sec) * (1 sec/10**9 bits) = 94 km round-trip, since light travels through an optical fiber at 0.6 the speed of light.
23. Y. Rekhter and T. Li, working draft (1995).
24. Y. Rekhter and P. Traina, working draft (1995).
25. P. Karn, "The KA9Q Internet (TCP/IP) Package: A Progress Report," *ARRL Amateur Radio Sixth Computer Networking Conference*, American Radio Relay League, Newington, CT (August 29, 1987), pp. 90–94.
26. B. K. Aldred, G. W. Bonsall, H. S. Lambert, and H. D. Mitchell, "An Architecture for Multimedia Communication and Real-Time Collaboration," *IBM Systems Journal* **34**, No. 3, 519–543 (1995, this issue).
27. Y. Rekhter and D. Kandlur, working draft (1995).
28. J. Steiner, B. C. Neuman, and J. I. Schiller, "Kerberos: An Authentication Service for Open Network Systems," *USENIX*, USENIX Association, Berkeley, CA (Winter 1988), pp. 191–202.
29. *Data Encryption Standard*, Federal Information Processing Standards Publication No. 46, National Bureau of Standards, Gaithersburg, MD (15 January 1977).
30. D. Koblas and M. R. Koblas, "Socks," *UNIX Security III Symposium*, USENIX Association, Berkeley, CA (September 1992), pp. 77–83.
31. A detailed paper on TCP/IP performance modeling using this newly developed IBM tool is a subject for future publication.
32. *TCP/IP for MVS V3R1 Performance Tuning Guide*, SC31-7188-00, IBM Corporation (September 1994); available through IBM branch offices.
33. D. D. Clark, V. Jacobson, J. Romkey, and H. Salwen, "An Analysis of TCP Processing Overhead," *IEEE Communications Magazine* **27**, No. 6, 23–29 (June 1989).
34. M. L. Hess, J. A. Lorrain, and G. R. McGee, "Multiprotocol Networking—a Blueprint," *IBM Systems Journal* **34**, No. 3, 330–346 (1995, this issue).
35. J. Ubois, "New Shades of Blue," *Internet World* **6,** No. 3, 62–64 (March 1995).

*Accepted for publication April 10, 1995.*

**Edward G. Britton** *IBM Networking Software Division, P.O. Box 12195, Research Triangle Park, NC 27709 (electronic mail: brittone@vnet.ibm.com).* Dr. Britton earned his B.S. in mathematics at Duke University and his Ph.D. in computer science at the University of North Carolina at Chapel Hill. In 1976 he started working on TCP/IP development for the U.S. Defense Communications Agency. Since joining IBM in 1981, he has contributed to SNA architecture, OSI system design, telephony, and telecommunications cross-product design. As a senior programmer in the TCP/IP Strategy and Design department, his responsibilities include IPng, security, performance, wireless and asynchronous transfer mode communications, the National Information Infrastructure forum, and the relationship of TCP/IP with other protocol suites.

**John Tavs** *IBM Networking Software Division, P.O. Box 12195, Research Triangle Park, NC 27709 (electronic mail: tavs@vnet.ibm.com).* Mr. Tavs earned his M.B.A. at Duke University and his B.S.E.E. at the University of Florida. In 1982 he joined IBM to work on the architecture and development of the I/O architecture for the IBM 9370, AS/400, and 3745-900. In 1988 he moved to Research Triangle Park, where he has worked on 3172 Network Management, frame relay and TCP/IP of NCP, 6611 TCP/IP, and IBM's TCP/IP host products. He manages IBM's TCP/IP Technical Strategy group, evangelizes the use of TCP/IP in IBM products, and drives IBM's participation in the IETF.

**Redha Bournas** *IBM Networking Software Division, P.O. Box 12195, Research Triangle Park, NC 27709 (electronic mail: bournas@vnet.ibm.com).* Dr. Bournas received the B.S. degree with honors in computer science and mathematics in 1980, and the M.S. degree in electrical engineering in 1981, both from the University of Pittsburgh. He earned the Ph.D. degree in electrical engineering systems in 1990 from the University of Michigan, Ann Arbor, where he received the graduate distinguished achievement award in electrical engineering systems. He joined IBM in 1982 and was awarded a three-year IBM graduate fellowship from 1987 to 1990. He contributed to the design and development of the IBM 4381 and 9370 processors. As an advisory engineer in the TCP/IP Performance department, he works on performance design, analysis, and modeling. His research interests include multidimensional queuing systems, performance modeling, and analysis of communication protocols and networks.

Reprint Order No. G321-5577.