An Introduction to the Internet, Part 1: History, Organization and Function

Jerry V. Glowniak

Nuclear Medicine Service, VA Medical Center, Portland, Oregon

Key Words: Internet; computers; networks; protocols

J Nucl Med Technol 1995; 23:56–61

The Internet is a rapidly growing and changing entity that allows access to a great deal of medical information. This is the first article in a four-part series on the Internet. Upon completion, the technologist should be able to describe the basic organization of the Internet, understand how information is sent over the Internet, and give a general description of how computers and networks are identified on the Internet.

INTRODUCTION

The 1990s have witnessed an explosive growth in telecommunications, a phenomenon that will undoubtedly continue into the next century. At the forefront of this growth is the Internet, a worldwide communication network that links together millions of computers. The amount and diversity of information on the Internet is staggering, ranging from the amusing and trivial to the technical and scientific. The increasing use of the Internet by the general public over the last two years can be attributed to two major developments: the emergence of commercial Internet access providers that offer low-priced Internet connectivity to individual users; and the availability of user-friendly software for exploring the Internet. This article describes the basic structure of the Internet and how it functions.

HISTORICAL BACKGROUND

To the average user who first approaches the Internet, its organization appears chaotic. Information and resources are scattered among thousands of computers, with no logical connection between these computers. Information may be redundant or nonexistent in certain areas, and methods for accessing resources vary between computers. Much of this disorganization is the result of the origins of Internet. In its early development, the Internet was designed to be a small-scale networking experiment between computers at large research institutions. It was not meant to be user-friendly since only individuals with a high level of computer knowledge formed its user base. The growth of the Internet occurred as a series of ad hoc measures to accommodate an increasing number of computers, with no clear objective that a worldwide system of computers would become the ultimate goal.

The beginning of the Internet is usually ascribed to a networking experiment between computers at Stanford, UCLA, the University of California at Santa Barbara and the University of Utah (*1*). This experiment, begun in 1969, was funded by ARPA, the Advanced Research Project Agency of the Department of Defense (now renamed DARPA, the Defense Advanced Research Project Agency). The network was called the ARPANET and the project had two main goals: to develop the technology for sharing resources between different types of computers with different operating systems; and to develop methods of ensuring reliable computer communications even when part of the network was disabled or communications became unreliable, as might occur in a nuclear war. The result of these efforts was the development of the TCP/IP protocol suite which forms the basis for all communications on the Internet and many other networks.

ARPANET was very successful, so much so that many universities and academic institutions started connecting their networks to ARPANET which itself was continually being expanded by the Department of Defense. Because the main method of connecting computers together was by linking computer networks to the ARPANET, the concept of the Internet arose as a network of computer networks. By the early 1980s, about 20 universities and 1,000 computers were on the Internet. DARPA, however, had neither the funds nor the mandate to build a larger network, and the Internet almost came to an end at this point.

About this time, the National Science Foundation (NSF) decided that linking computer networks together at major universities and research centers would further its goals of advancing research and education in the U.S. NSF created an administrative and funding agency called NSFNET

For correspondence and reprints contact: Jerry Glowniak, MD, VA Medical Center (115), PO Box 1034, Portland, OR 97207.

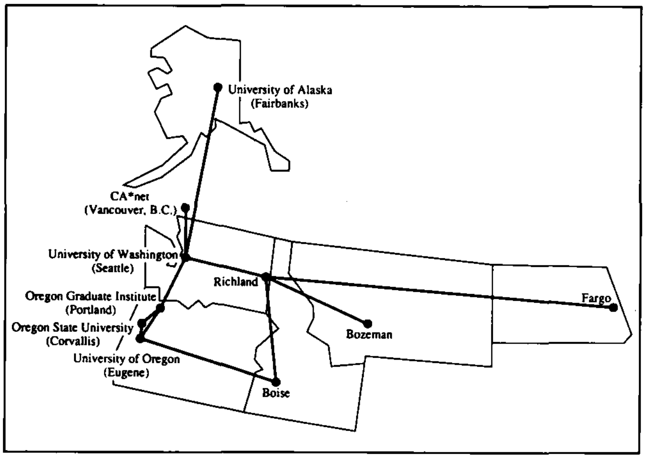

FIGURE 1. An example of a regional network, NorthWestNet, supported by NSFNET. The major connecting links and hubs in this network are shown. Over 200 academic, research and commercial networks, primarily in the Portland and Seattle areas, are connected to NorthWestNet. NorthWestNet connects to a U.S. Internet backbone network, ANSNET (NSFNET), at Seattle (see Fig. 2). CA*net is the major national Canadian backbone network on the Internet. It connects to the U.S. backbone network in western Canada through NorthWestNet.

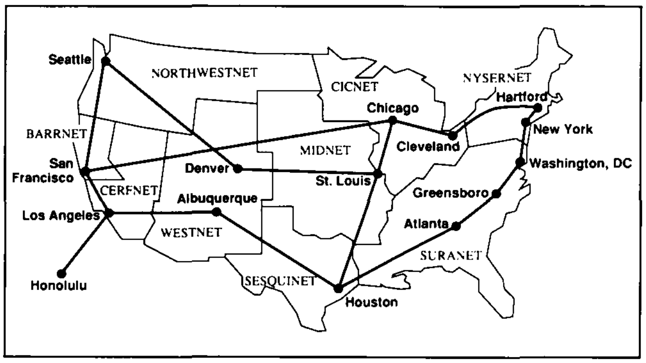

FIGURE 2. The structure of the major Internet backbone network, ANSNET, in the U.S. near the end of 1994. The 15 nodes indicated by city names are the points at which the regional networks connect to the backbone. Links between the nodes are fiber optic cables owned by MCI. Some of the major regional networks and their boundaries are shown.

whose purpose was to develop a nationwide system of computers linking major academic institutions. In 1985, NSFNET funded a network linking the six NSF sponsored supercomputer centers in the U.S. Links between these computers were leased telephone lines that operated at a maximum speed of 56,000 bits per second (56 Kbps), an exceedingly fast rate at the time. As with ARPANET, this network became overloaded as academic institutions began connecting their computer networks to it. In order to accommodate this growth, NSF funded the creation of a new, larger network that operated at 1.5 million bits per second (1.5 Mbps) and had 14 connecting points, or nodes, located at strategic points throughout the U.S. This network, referred to as the backbone, was built by a partnership between MCI, IBM and MERIT (a Michigan-based education and research network) and provided high-speed long-distance connectivity between widely separated computer networks throughout the U.S. MCI provided the leased lines between the nodes; IBM developed the equipment, called routers, for sending, receiving and switching traffic between the nodes; and MERIT supervised network operations. The actual day-to-day operations of the backbone was handled by Advanced Network and Services, Inc. (ANS), an independent corporation established by the three partners, IBM, MCI and MERIT. NSFNET also funded the creation of 19 regional networks that connected universities and research centers in specific geographic areas (Fig. 1, 2). NSFNET paid ANS to provide connectivity of these regional networks to the backbone network.

The new NSF network became operational in 1988 and again was a huge success. By October 1990, the Internet consisted of approximately 300,000 computers on 2,000 networks. The Internet was also becoming more international with 25% of its networks being foreign, nearly all of which were in Europe, Australia, Japan and Canada. The new backbone network was so popular that by the end of 1991, the number of networks and computers on the Internet had doubled, and the backbone was again becoming overloaded with traffic. In response ANS, with NSFNET funding, built a new backbone network, ANSNET, with 15 nodes and a capacity for transmitting 45 Mbps across any of its links (Fig. 2). At this rate, nearly 1,000 pages of text could transmitted in one second. This network became operational in 1992 (*2*). Since ANSNET was funded principally by NSFNET, the backbone network was usually referred to as NSFNET.

With the establishment of a viable nationwide computer network, NSFNET decided to stop funding ANS for providing backbone connectivity to the regional networks in order to foster competition between other providers of national and international network services. Companies that provide long-distance high-speed Internet connections (backbone service) are referred to as network service providers (NSPs). Well-established telecommunication companies, such as Sprint and MCI, already have high-speed communication networks that can provide backbone connectivity and, by early 1995, MCI and Sprint had captured most of the backbone traffic on their own networks. Other NSPs that are entering the field include PSI (Performance Systems International), AlterNet and Netcom. More are expected to emerge in the coming years. By mid-1995, NSFNET will have terminated payments to ANS for providing backbone connectivity of the regional networks to ANSNET. Instead, NSFNET will give direct grants to the regional networks who will then choose an NSP. This direct funding of the regional networks will end after four years, at which time they will have to find other sources of funding (*3*).

While the NSF was the major force in creating the Internet in the U.S., other networks were built by government, commercial and private groups that had varying degrees of connection to the Internet. Major networks such as the Energy Science Network (ESNET), of the Department of Energy,

TABLE 1
Top-level Domains for Computer Networks in the U.S.

edu	Educational
com	Commercial
gov	Government
mil	Military
net	Network
org	Organization
int	International

readily connected to the Internet and often provided connectivity to various smaller networks. Commercial activity was prohibited on NSF-sponsored networks and for-profit organizations had a relatively small presence on the Internet until recently. Now that government funding of most Internet network operations is being withdrawn, commercial activity is growing very rapidly on the Internet.

While this discussion has focused on Internet development within the U.S., similar developments were occurring in other regions, primarily Europe. Backbone networks in these countries were connected to each other and to the U.S. backbone to create the worldwide Internet.

ORGANIZATION OF THE INTERNET

The basic organizational unit on the Internet is the network. Every computer on the Internet must belong to a computer network which itself may be part of a larger network, such as a regional network, or the network of a commercial Internet access provider. These larger networks are joined to the Internet by connecting to an NSP. In order to create some degree of order on the Internet, computer networks are designated by types called top-level domains. Seven top-level domains, specified by three-letter codes, are recognized in the U.S. and in some foreign countries (*3*) (Table 1 and Fig. 3). For example, the computer network at the University of North Carolina is named unc.edu. Outside the U.S., top-level domains are the countries in which the networks reside and are designated by two-letter country codes. Computer networks and computer names in France, for example, end with the designation "fr".

Computers on the Internet are identified by numbers and usually by names. A particular name or number can refer to only a single computer, but a computer can have more than one name or number. Either designation consists of two parts. One portion identifies the computer or host; the other portion identifies the network and, optionally, any subnetwork on which the host is located. Of the two naming schemes, computer numbers, called Internet Protocol (IP) numbers or IP addresses, are the more fundamental.

All computer communications on the Internet use IP numbers. An IP number is a 32-digit binary number. Thus there are $2^{32} = 4{,}294{,}967{,}296$ possible IP numbers which can be used as computer addresses. IP numbers are written as four 8-bit segments called bytes that are expressed in decimal

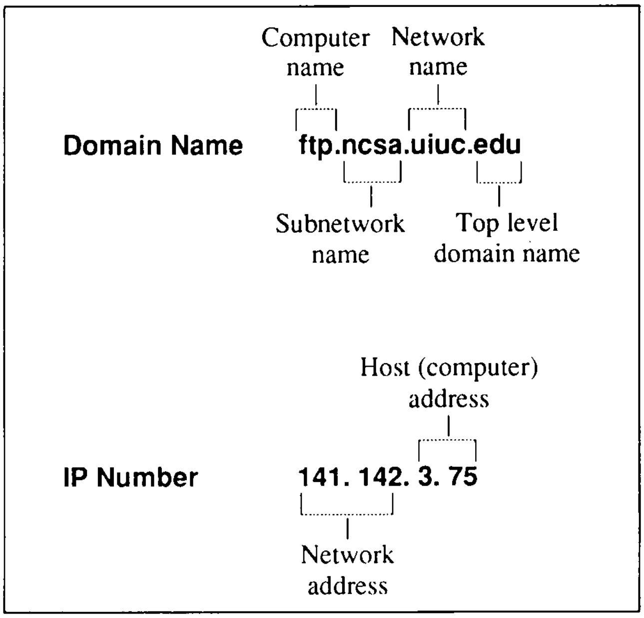

FIGURE 3. Naming and numbering scheme for computers on the Internet. Domain name and IP number (IP address) for a computer at the National Center for Supercomputing Applications at the University of Illinois, Urbana-Champaign. The leftmost number in the IP number, 141 in this example, determines what portions of the IP number constitute the network address and the host address. Since the leftmost number is between 128 and 191, the network address is the leftmost two numbers (141 is the first number, 142 is the second). The remaining numbers, 3 and 75, form the host portion of the IP number.

form and separated by periods (Fig. 3). The four numbers in an IP address can range from 0 to 255, although the numbers 0 and 255 are usually not used in actual addresses. The network portion of the IP number varies according to the value of the leftmost number. If the leftmost number is in the range of 1 through 127, this number constitutes the network address. If the leftmost number is in the range 128 through 191, the leftmost two numbers are the network address. If the leftmost number is 192 through 223, the leftmost three numbers are the network address. The remaining numbers are the host portion of the address.

Although computers use IP numbers to communicate, humans find it difficult to remember numbers. Therefore, names are usually also assigned to computers and networks. The combination of a computer name and a network name is called a domain name and this system of naming is referred to as the domain name system (DNS). Computer and network names consist of groups of characters separated by periods. The computer name is the leftmost group of characters, followed by zero to three subnetwork names, a network name, and finally the top level domain name (Fig. 3). Whenever a computer on the Internet is contacted, the domain name or IP address of the computer must be given. If the domain name is given, it must be converted to the corresponding IP address. It would be impossible for any computer to keep a list of the domain names and IP addresses of all the computers on the Internet. Even a list of several thousand names is difficult to keep up-to-date, let alone a list of several million. Rather, each network on the Internet is required to have access to at least two computers which keep lists of a portion of the domain names and their IP addresses. These computers are called nameservers and are contacted whenever a computer program specifies a domain name. If the primary nameserver does not have the IP address for the domain name, it will contact other nameservers until one is found that has the IP address for the domain name. This IP address is then returned to the local computer and substituted for the domain name in the computer program.

Since the proper functioning of the Internet requires that each IP address and domain name be associated with only one computer, there must be a method of making sure that these assignments are unique. Given the size of the Internet, it would be impossible to maintain a list of all the computers on the Internet. Instead, an organization funded by NSFNET, the InterNIC (Internet Network Information Center) Registration Services Center registers all computer networks on the Internet and assigns them unique network names and numbers. The individual networks then assign the host portions (and subnetwork names if necessary) to the computers on their networks. As of February, 1995, the InterNIC had approximately 50,000 registered networks of which 58% were in the U.S. Since neither the InterNIC nor any other agency registers individual computers, the number of computers on the Internet cannot be known precisely. Using various search techniques, it was estimated that as of January, 1995 there were 4.8 million computers on the Internet. It is even more difficult to estimate the number of people who actually use the Internet. The general rule of thumb is that there are 10 users per computer which means that as of January, 1995 there were 48 million users.

COMMUNICATION ON THE INTERNET

One of the purposes of ARPANET was to develop a means of allowing different types of computers with different operating systems to communicate seamlessly. The way this was accomplished was by creating an extensive set of rules that all computers on the Internet had to follow. Rules on a particular aspect of computer communication are grouped together to form protocols. These protocols as well as a great deal of informational material about the Internet are published at regular intervals as documents called RFCs (Request for Comments). RFCs are numbered consecutively and approximately 1,700 have been published to date. They are available on-line from the InterNIC. RFCs which specify protocols that have become widely used are given the special designation of Internet standards. These standards, about 50 in number, describe the most important rules for Internet communications. The entire set of standards is often referred as the TCP/IP protocol suite, named after the two most important protocols: the transmission control protocol (RFC793) and Internet Protocol (RFC791). TCP ensures that messages sent between computers are reliably received while IP is responsible for the routing of messages between

ISO/OSI Model
Application
Presentation
Session
Transport
Network
Data Link
Physical

Internet Model	Protocol
Application	ftp, telnet, gopher
Transport	TCP, UDP
Internet	IP
Subnet	Ethernet
Link	SLIP, PPP
Physical	RS 232

FIGURE 4. Protocol layers in the ISO/OSI and Internet models. Typical protocols for the Internet model are shown at the right. SLIP = Serial Line Internet Protocol; PPP = Point to Point Protocol. Adapted from Lynch (*1*).

computers (*3*). From a user's standpoint, the most important protocols are the application protocols which describe Internet services such as the telnet protocol, which describes rules for logging on to a remote computer, and ftp, the file transfer protocol which delineates a method for transferring files between computers. Internet services will be described in more detail in the third article in this series.

Most of the protocols for computer communications on the Internet were developed in the U.S. for the ARPANET. Other countries, however, were developing communications protocols that were not always compatible with those used on the Internet. In response to this situation, the International Standards Organization (ISO), an international body charged with setting communication standards, began developing a set of protocols that was meant to form the basis for computer communications worldwide. A model of computer networking was created in which the entire process from user input at the application program level to the sending and receiving of messages on the Internet was broken up into a series of steps with protocols describing how a message was to be handled at each step. This model was officially called the Open Systems Interconnection (OSI) model.

Figure 4 shows the components of this model. In the OSI model, user input is handled by application level protocols. The resulting output is passed on to the presentation level where it is modified according to protocols at this level. The message is consecutively handled by each lower level until it reaches the physical level where it is placed on the transmission medium of the local network and sent to the remote computer. When the message arrives at the remote computer, it is successively passed up to the application level where it's then acted upon by the remote application program. The ISO, however, was rather slow in developing the protocols for this model. The Internet community could not wait for this formal set of standards and, in an ad hoc manner, developed a similar set of protocols called the Internet model that became widespread throughout the U.S. and parts of Europe (Fig. 4). In both models, these networking standards are called protocol stacks or layers. The major

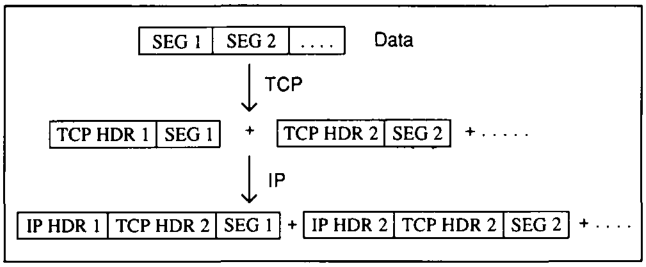

FIGURE 5. Packet formation by the TCP/IP protocol. A file or other type of message, represented by DATA in the figure, is broken up into consecutive segments (SEG1, SEG2, etc.) by the TCP protocol. A TCP header (TCP HDR) and IP header (IP HDR) are added to each segment to form packets which are transmitted over networks on the Internet.

protocols in the Internet model form the TCP/IP protocol suite. For particular applications, not all levels of these models need to be used. Some applications do not need to use the TCP protocol and substitute a simpler protocol called UDP (User Datagram Protocol).

The following is a brief description of how the protocol stacks create a message that can be placed on the Internet (Fig. 5, 6). Data from an application program, for example a file which is to be transferred to a remote computer, is passed to software that implements the TCP protocol. TCP breaks the data up into segments and attaches an additional piece of data, called a header, to each segment. The header contains various kinds of information such as the type of application program, the segment number and instructions to ensure that the data is delivered correctly. This packet is then passed on to the IP protocol software which adds another header containing the IP address of the source and destination computers and instructions for routing the packet through the various networks on the Internet. Finally the packet is handed over to software at the lowest levels of the protocol stack that attaches a third header that directs the packet to a router on the local network. Routers are communication devices that connect networks and direct packets to their destinations. When the packet arrives at the router, the network header is stripped off, the destination IP address is read in the IP header, and a new network header is generated directing the packet to a router on the next network. This process is continued until the destination computer is reached. The networking software on the destination computer successively strips off the headers, reassembles the data into the original message and passes it on to the application program.

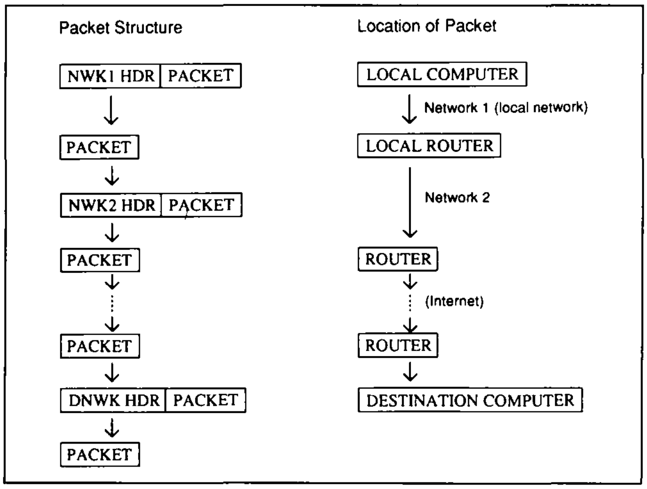

FIGURE 6. Packet transmission on the Internet. A packet, consisting of an IP header, a TCP header and a data segment, has a network header added to it (NWK1 HDR) by the local computer that directs it to the local router on the local network (network 1). The local router removes the network header and adds a new network header (NWK 2 HDR) that directs the packet through the next network (network 2) to another router. As the message traverses the Internet, each router removes the old network header and adds a new one until the destination network is reached. The last router adds a destination network header (DNWK HDR) to the packet which delivers it to the destination computer.

The above description demonstrates a cardinal feature of the Internet—packet switching (*4*). Each packet contains all the information necessary to deliver it to its final destination, but the route to that destination is not predetermined. When it receives a packet, each router decides the best route the packet should take to reach its final destination. This route can change from packet to packet depending on network conditions, such as network congestion or whether a certain segment of a route has gone down. Because every packet carries its own identification, a given network can carry many different packets from different sources at the same time. This method of data transmission contrasts with that used in telephone networks which is called circuit switching. In circuit switching, a single circuit is created and maintained during the entire communication process. Data segmentation and headers are not needed for the transmitted information because only one route is used and only one conversation can use the circuit at a time. This method of communication, however, would be extremely inefficient for computer networks. Because message transmission is very much more rapid on computer networks than user input, the circuit would be idle most of the time if only a single user had access to the circuit. Packet switching allows many messages to be transmitted on a circuit at the same time.

SUMMARY

The Internet is a worldwide system of computer networks that allows millions of computers to communicate. Computer communication takes place by means of the TCP/IP protocols. TCP breaks up a message on a local computer into segments. The TCP and IP protocols then add headers to the segments to form packets. The TCP and IP headers contain information that controls the transmission of the packets on the Internet and the reassembly of the original message on the remote computer. A packet is placed on the local network and sent to a router. The router reads the destination IP address in the IP header and sends the packet to another router that is connected to another network. Packets travel through a series of routers and networks until they reach the destination network where they are delivered to the remote computer. The remote computer strips off the headers and reassembles the original message.

The next article in this series will cover how to access the Internet.

REFERENCES

1. Lynch DC. Historical Evolution. In: Lynch DC, Rose MT, eds. *Internet System Handbook.* Reading, MA: Addison-Wesley Publishing Company, Inc.; 1993:3–14.
2. Catlett CE. Internet evolution and future directions. In: Lynch DC, Rose MT, eds. *Internet Systems Handbook.* Reading, MA: Addison-Wesley Publishing Company, Inc.; 1993:717–749.
3. Lawler A. NSF hands over the Internet. *Science* 1995; 267:1584–1585.
4. Krol E. *The Whole Internet User's Guide and Catalog.* 2nd ed. Sebastopol, CA: O'Reilly and Associates, Inc.; 1994:23–34.