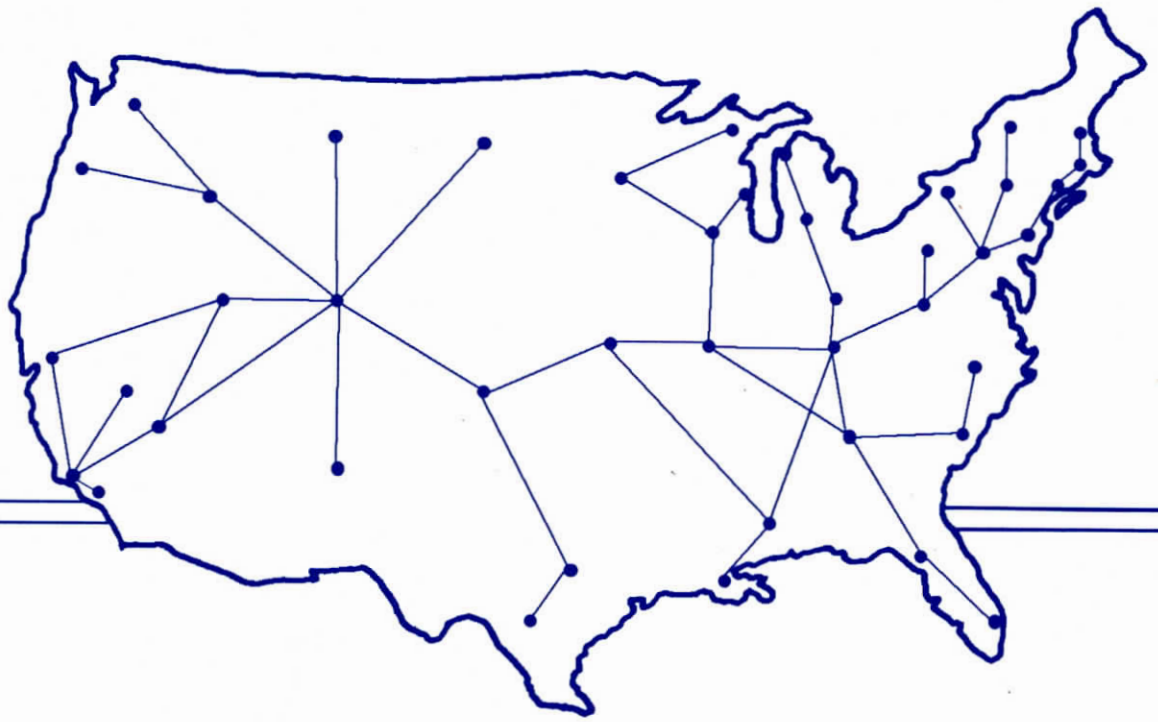
The Easy Internet Handbook



Javed Mostafa
Thomas Newell
Richard Trenthem

Hi Willow Research and Publishing

The Easy Internet Handbook

Javed Mostafa
Thomas Newell
Richard Trenthem

Hi Willow Research & Publishing
Castle Rock, Colorado
1994

Suggested Cataloging:

Mostafa, Javed
The Easy Internet handbook / Javed Mostafa, Thomas Newell, Richard Trenthem. -- Castle Rock, Colo.: Hi Willow Research and Publishing, 1994
iii, 150 p. 22x28 cm.

1. Internet (computer network). 2. Communication networks - Handbooks I. Newell, Thomas. II. Trenthem, Richard. III. Title.

004.6 M916 1994 TK5105.875.M68 1994

ISBN 0-931510-50-3

Publisher: Hi Willow Research and Publishing
P.O. Box 266
Castle Rock, CO 80104-0266

Can be ordered from the publisher or from its distributor:
Libraries Unlimited
P.O. Box 6633
Englewood, Colorado 80155-6633
800-237-6124

Table of Contents

Preface

Few sources today provide coverage of basic Internet resources and tools with an eye toward both new and regular users. It is essential to provide sufficient background information so that new users can place the functions of tools and resources in their proper contexts. It is also necessary that information be organized in a fashion that allows regular users to quickly find specific facts without getting bogged down in details. A new user may eventually become a regular user; hence, the structure of the document should also support the changing demands of the same user. We were inspired to write this book to meet the demands of different Internet user groups and the changing demands of individual users.

The book is divided into two parts. Part I, Internet: Introduction and Background, is aimed at new users. It explains concepts, provides information on devices and services related to the Internet, and compares and contrasts the major tools and resources. It also includes a comprehensive networking glossary and a bibliography. Part II, Internet Resources and Tools, is aimed at regular users. It is designed to provide instructional information, presented in a step-by-step manner, to assist the user to quickly locate and understand information on using basic tools and resources.

We would like to thank the many individuals who helped us in the compilation of this book. Unfortunately, space does not permit us to list all their names. Four individuals, however, deserve special mention. Tom Miller, director of the Information Processing Laboratory in our department, helped us locate a suitable publisher for the book. Dr. Brooke Sheldon, dean of the department, has been very supportive of this project and helped in publicizing it. Philip Doty and Dr. Glynn Harmon, both professors in our department, provided valuable suggestions for improving the book. We are grateful to them all.

Dedications

Javed dedicates this book to Charles Zakrzwski and Esmaeil Samimy-- two of his best friends.

Tom dedicates this work to his wife Janice and his children TK and Brittany. They are his inspiration.

Richard dedicates this book to his wife Wendy, and appreciates her constant support.

Addresses

Those with Internet access can reach the authors directly at the e-mail addresses below:

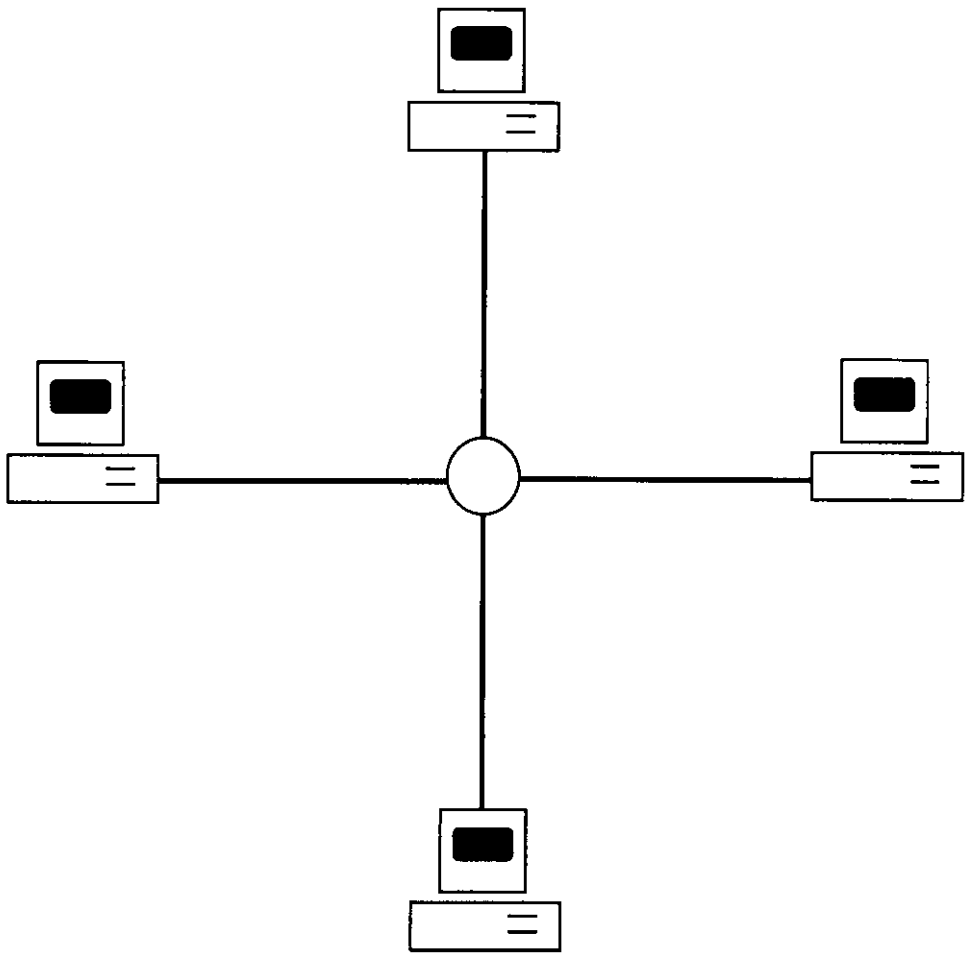
Javed Mostafa	jm@bongo.cc.utexas.edu
Thomas Newell	tnewell@bongo.cc.utexas.edu
Richard Trenthem	richie@bongo.cc.utexas.edu

For the less fortunate readers without Internet access (yet) we provide the mail address below:

Thomas Newell or Javed Mostafa
2021 Guadalupe Street
Suite 100-40
Austin, TX 78705

PART I

Internet: Introduction and Background



Chapter 1

What Is a Computer Network?

Introduction

In society the exchange of information touches all facets of human activity and is the impetus for the establishment of networks of all kinds. Networks facilitate that exchange and are nothing more than a collection of entities that communicate or exchange information across established channels using recognized standards. A group of people can make up a network, whether their communications are formal or informal, regular or sporadic.

A computer network is not much different. It consists of computers and other computational devices that exchange information over physical links using conventions recognizable by all nodes. In figure 1.1, a computer and its printer represent a network. The computer communicates with the printer via cabling, sending it bits of recognizable information. The stand-alone computer controls and manages the network and could be called a "hardware-free" network because it requires no other physical devices to manage network activity.

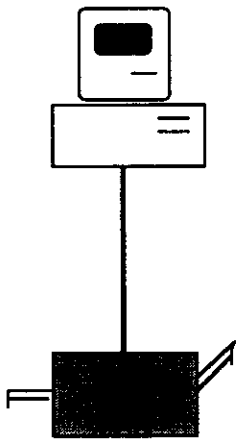


Fig. 1.1. A Network Consisting of a Computer and a Printer.

More complicated computer networks aren't much different. They come in many forms and vary in how they communicate and the types of devices necessary to manage them. But they are all essentially just devices communicating over common links. They may consist of one or a few machines sharing a common printer or file storage device within an office, or they may cross international boundaries and incorporate shared computational power and file storage devices distributed over geographically dispersed locations and connect hundreds or thousands of computers. They may be physically connected through network cabling, or they may rely upon satellite communication. They may be a single network, or they may be several interconnected networks (an internet).

What are the advantages to networking?

Whether consisting of a few machines or a vast, complicated labyrinth of sophisticated hardware and communication schemes, all networks have similar benefits and problems. Perhaps the principal reason for establishing a network is a matter of economy. A group of people (or organizations) perceive that by establishing a network, their individual goals will be more economically achieved or achieved more effectively. Specifically, networks offer:

1. sharing of expensive computational resources,
2. sharing of peripheral devices such as printers or disk drives,
3. sharing of information resources, and
4. channeling of information efficiently.

Access to powerful computers across networks was used as a justification for proposing a National Research and Education Network (NREN) to succeed the Internet. Proponents argued that a network of this order would give researchers the computational power necessary to solve complex, high-order mathematical and simulation problems. By distributing the computers and creating means for researchers to communicate with them over this proposed national information highway, the high cost of fast processors is spread among many organizations. The Internet today is used to this end, but NREN will provide faster information interchange. Many university and business networks

are organized along similar lines. By purchasing one powerful system and establishing and controlling access to its tools, the entire organization benefits.

Typical office networks connect workstations to shared printers and in some cases distributed file storage devices. Such connections allow an office manager to provide the use of an expensive tool to a wider range of users. Shared file storage provides a means for maintaining large files of information that may be beyond the storage capacity of individual workstations.

Computer networks provide means for groups of people to share common information resources. An office work group or a team of research scientists working at separate sites can make use of a network to share a common set of data. Consider the example of a business that sells widgets through a national distribution network. Distribution sites take customer orders for widgets and input those orders into a common database. Using a shared network, the distribution sites can track the manufacturing sites' scheduling information and determine delivery dates. The manufacturing sites can stay abreast of sales data and more efficiently schedule production. Managers can follow sales and distribution processes to make informed marketing decisions (See figure 1.2).

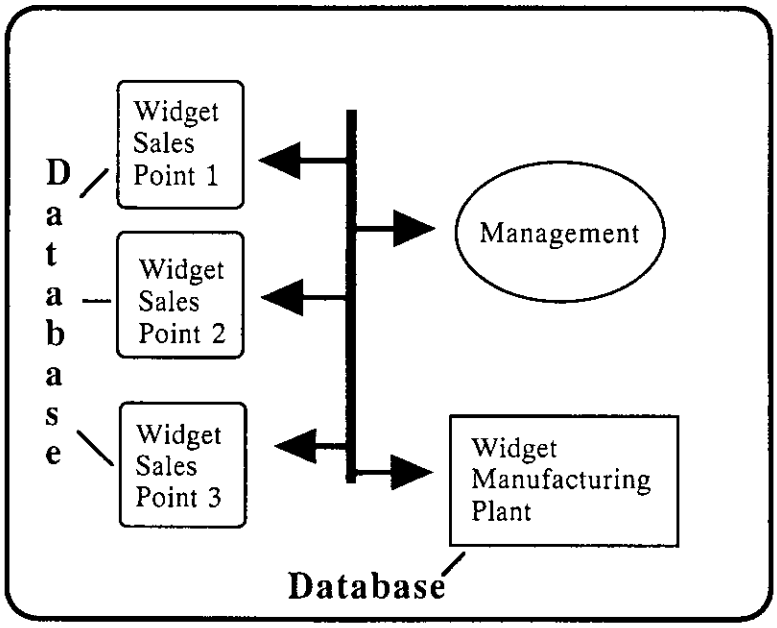


Fig. 1.2. A Shared Business Network.

Computer networks give organizations the ability to apply economy of force to solving problems and achieving goals. The traditional linear and hierarchical organizational model can be modified quickly and temporarily to foster efficient work group relationships geared to solving immediate problems. Networks allow for assigning geographically dispersed people (and even organizations) to the same project using telecommunication rather than collocation. This capability has generated much discussion centered on what has come to be known as a "virtual corporation."

Many software tools common in the modern office incorporate features that allow many people to edit and contribute to joint projects. Consider a group presentation that merges images, spreadsheets, and text. One person may edit images, another may perform spreadsheet analysis, and yet another may edit text. Despite the fact that the project contains three distinct groups of data stored in three different locations, the end product appears as a whole. Software tools recognize and assemble all the pieces of the project across the network.

Networked communication tools have made the process of seeking fast, interactive consultation or advice an easy and inexpensive alternative to telephone, fax, or mail interchange. Other examples of shared information resources include user communication such as e-mail programs, electronic discussion groups (LISTSERVs), and electronic bulletin boards. Networks provide people access to capabilities such as these, and their impact on our lives is growing daily.

What are some disadvantages to networking?

Ultimately, the purpose of a network is to empower its users. Information is power, and computer networks deliver it into the hands of its users. As people are the first and most important component of an organization, an organization is only as talented and capable as its people. Networks are a tool that people can use to improve productivity through rapid information interchange. Computer networks, however, are not without their disadvantages. Common problems with computer networks include:

1. implementation and maintenance costs,
2. security and privacy,
3. accuracy and reliability of data,

4. disparities between individuals and groups with regard to wealth and information access,
5. varied technical standards, and
6. an accelerating pace of technological change.

The physical components of a network can be expensive to purchase, install, and maintain. Information managers must weigh these costs with the value gained in order to determine if a network is necessary or beneficial to solving the problem at hand. This consideration can often be a question of scale. "We need a network, but not a large one." "We need a way to communicate and share information with an office overseas." "We just want to share a printer between offices." These are all examples where scale is the most important factor to consider for network implementation.

All networks face security problems. Because all machines on a network are connected (and certain machines may share storage devices), the potential exists for unauthorized access to sensitive or private information. Network tools exist for system administrators to minimize and manage risk, but risk cannot ever be completely eliminated. Certain files that may contain privileged information will have to be reserved for use by select individuals. Part of security revolves around privacy related to electronic means of data and information transfer. Procedures must be in place to ensure network communications reach their intended recipients. Managers must give thought to the protection of the system from computer viruses.

Information obtained from a network is not necessarily accurate or reliable. Information such as that in distributed databases or research data used exclusively within the bounds of a particular group is only as accurate as that group wants to make it. It is often difficult to gauge the accuracy of data gathered across network ties. The inaccuracies of network information are especially evident in information gathered from electronic forums such as discussion groups or news groups. No institution ensures the reliability of information disseminated through these channels. Thus, users should consider the accuracy of data before making decisions based upon what they have found.

In the growing debate about a national information infrastructure (NII), there are those who argue that information and access to it are fundamental rights and that communications networks implemented as part of the NII must accommodate all people. Some researchers have concluded that an information

gap exists between the wealthy and the poor, and that the gap is growing. Access to electronic information is obviously limited to those with computers and a means to link with a network. Certainly some organizational networks will remain closed to public access, such as those operated by businesses or military organizations, but what about public access to the NII? Much debate is centered on how to ensure that all people will have a means to tap into this information superhighway, but no clear answer has yet emerged as to how this will happen.

Network managers often must coordinate the connection of dissimilar hardware and software. Not all users employ the same software, use the same type of computer, or even the same network communications scheme. This causes translation and protocol problems that must be resolved in order for a network to realize its potential. The wide-scale adoption of TCP/IP as the network protocol has encouraged and eased the connection between dissimilar networks and machines and helped to accelerate the growth of the Internet.

The rapid growth of the Internet highlights a phenomenon of the computer industry; the rate of change is accelerating. Machines purchased as few as two years ago can be considered obsolete today. Network communication technologies are changing the nature of how people communicate and share ideas and changes in technology force changes in these relationships. Network managers must consider and plan for the pace with which change occurs in order to best support the demands and needs of the people whom they support.

What are the pieces of a network?

Now that we've defined networking and the role of computers, let's turn to an examination of the physical components of a computer network. Any discussion of the pieces of a network must center around its layout (topology) and its components. The basic components of a computer network consist of a server, workstations, a network operating system (NOS), communications links, cables, and interface cards. Some networks are connected to other networks. In order to internetwork, devices such as modems, routers, gateways, brouters, and bridges may be required.

Topology

The design of a computer network begins with what is called topology. A network topology describes the physical layout of a network in terms of how the nodes are linked to each other and how data transfer occurs between the nodes. Topologies can be described as centralized or decentralized. In centralized topologies, a central machine serves as an access point to network resources. In a decentralized topology, all machines on the network have access to all others. The most common network topologies are the bus, ring, and star. The importance of topologies is that different network operating systems require different network configurations. For instance, an IBM Token Ring network requires a star topology while most Ethernet networks are configured as a bus layout.

The bus topology has all devices connected in a line, as shown in figure 1.3.

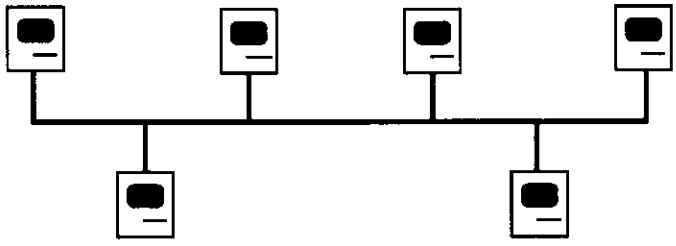


Fig. 1.3. A Bus Topology.

A ring topology has all devices connected in a circular loop, as shown in figure 1.4.

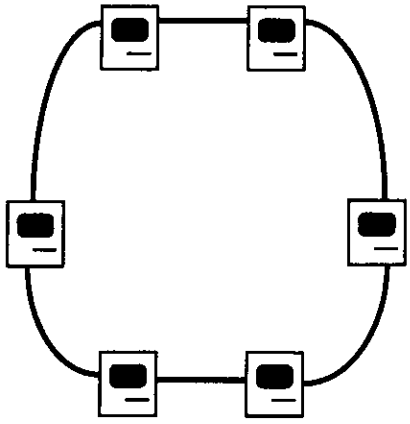


Fig. 1.4. A Ring Topology.

The star topology has devices linked to a central machine, as shown in figure 1.5.

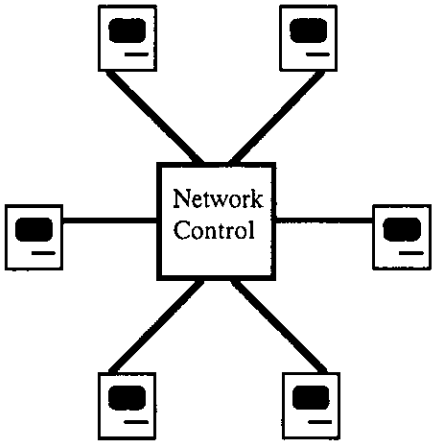


Fig. 1.5. A Star Topology.

Physical components of a network

A server is the computer that controls access to network resources. A file server maintains data storage for a collection of machines. When a workstation on the network needs access to a file, the file server provides the link to the requested information. The information may be data, text, or

program files. The server also performs the function of managing communication between nodes and scheduling tasks for certain devices such as printers.

A workstation is a computer connected to a network. Not only can it access the central processing unit (CPU) of the network server, but it has a CPU of its own. Workstations can be contrasted with "dumb terminals," which have no CPUs and act as unsophisticated user interfaces to a network computer.

A network operating system (NOS) is system software that controls the hardware of a computer network. It establishes and maintains the network connection between server and workstations. This software consists of a part that runs the server and a part that takes up residence in each workstation's memory, allowing it to function as part of the network. A typical NOS consists of utilities for network administrators to manage the efficiency of the network, establish security procedures, control access to shared peripheral devices, and schedule network maintenance tasks.

A communications link is a physical and a functional connection between a device and a computer. Physically, this is the cable or the wiring that connects a computer with another device; functionally, it is the capacity for the exchange of data between a computer and another device. In order for data to be exchanged across a network, a reliable link must be established and maintained. Nodes (specialized devices or computers) on a network communicate with one another either through direct physical links or through intermediate devices such as modems, routers, bridges, and brouters.

Networks are connected by a variety of cables. Cables can range from twisted-pair (similar to telephone wiring), to coaxial, to fiber-optic. In fact, newer networking technology includes wireless network connections that rely upon radio or infrared frequency transmissions, as shown in figure 1.6.

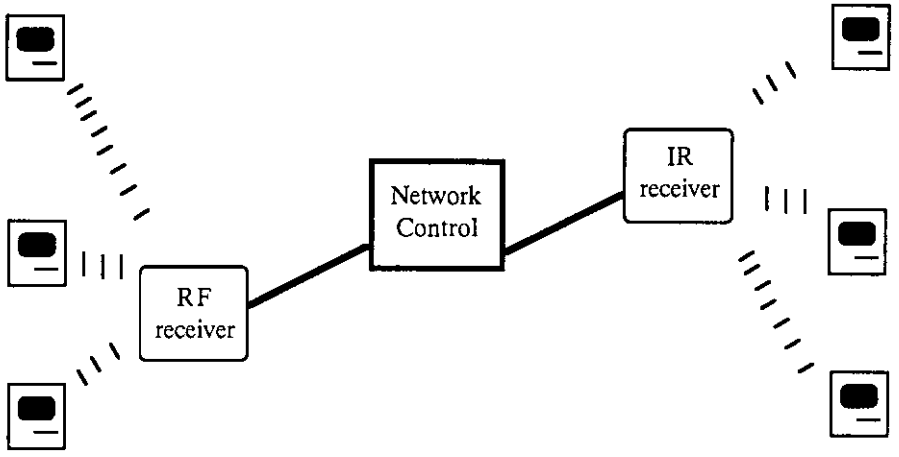


Fig. 1.6. A Wireless Network.

Interface cards are installed in network workstations and servers to allow them to exchange information according to established protocols. They allow network communication to bypass a computer's existing hardware and communicate at a greater speed. Networks such as AppleTalk that use the computer serial port to communicate are slower than card-based networks such as Ethernet or ARCnet. Interface cards organize data into packets that contain data and address information and then transmit them across the network. They also handle incoming packets addressed to that machine, stripping them of addressing data in order for them to be understandable by the workstation processor.

The utility of networks increases dramatically once they can be connected with other networks. This process is called internetworking (from which the term "Internet" is derived). Internetworking requires special protocols (which are the mutually acceptable conventions between two computers on different networks) to structure and exchange data. For example, Internet uses the transmission control and internet protocol (TCP/IP). Internetworking also requires devices that control, route, and manage information exchange.

Modems translate digital signals to analog signals for transmission over phone lines and then back again. They are used for establishing a "dial-up" connection to a network.

The function of bridges is to add routing data to information packets and ensure that the packets are transmitted to their ultimate destination. Bridges are devices that connect similar networks. When networks communicate with one another, they exchange information in the form of packets similar to those exchanged through network interface cards. A bridge uses protocols to add address information. The connection is transparent to a network user in that the user can communicate and exchange information with a user on another network as if it were a node of that local network.

Routers are special-purpose devices placed between two computers on a network. They examine the destination data of an information packet and make a decision regarding the optimal path over which to forward the packet. They do not, however, perform any translation of protocols required between networks.

Brouters are a cross between bridges and routers. They perform both the functions of connecting networks and routing information packets. A brouter adds addressing information to an information packet and selects a route along which the data will be transferred.

Gateways are devices similar to bridges; however, they connect dissimilar networks. For example, gateways allow microcomputers to connect to mainframes. Gateways are usually dedicated computers that run specialized software used for protocol translation between different networks.

Types of Networks

Local Area Network (LAN)

A LAN connects a series of workstations over a limited geographic range, typically an office or building, so that users may exchange information

and access shared devices such as a printer or disk storage space. Smaller LANs may only connect a few computers with a printer, while larger and more complex networks may rely upon a file server to control access to shared devices. The file server provides storage space for files and programs, which are accessible by network users, and it is where the network operating system is installed and managed. A node or workstation can function as the file server on smaller LANs, but a separate computer is necessary to perform this role in more complex networks.

LANs may be configured with any of the topologies discussed on pages 8-9. A LAN in bus configuration is shown in figure 1.7.

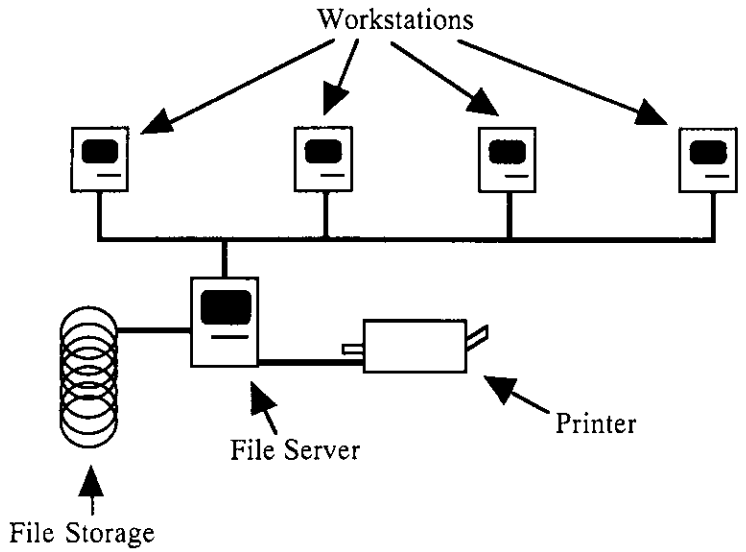


Fig. 1.7. A LAN in Bus Configuration.

Larger computers (such as mainframes) are designed to interact with special terminals, and data are transmitted with the capabilities of these terminals in mind. These terminals are strictly interactive devices and perform no computations. LAN workstations differ from these mainframe terminals in that they rely upon their own processing power to run programs. The file server network operating system may store programs and offer access through a menu system. When requested, the server loads that program into the memory of the workstation processor. "Dumb terminals" connected to mainframes rely upon the remote computer's memory to process instructions. The result is merely "echoed" to the local terminal. When communicating

with remote computers, LAN workstations must emulate the characteristics of a terminal that the remote machine recognizes.

Metropolitan Area Network (MAN)

A MAN is a network that spans a limited geographic area such as a city or suburb and tends to have a community focus. A MAN may be a network of community business or government agencies sharing a common information resource or even a community Freenet. It offers access to shared information and provides computing capabilities to members of a community. A MAN may consist of a combination of dedicated physical cabling connections and dial-up, phone line connections.

Wide Area Network (WAN)

A WAN is a network that spans a wide geographic area such as a county or state. WANs typically employ fiber optics or coaxial cables to establish and maintain long-distance data links. An example of a WAN would be a state agency which has established links between its main office and its subordinate locations. For instance, a state police headquarters may have a central repository of data which local police units contact to retrieve information. A subordinate office may have a LAN that is connected to the WAN through an intelligent bridge device or a router.

The Texas Higher Education Network (THEnet) which connects Texas universities, colleges, and research centers and provides its members access to the Internet, Bitnet, and other networks, is an example of a WAN. Because of the dissimilarity among networks, it must rely upon internetworking devices and protocols in order to exchange information.

So how do networks communicate?

LANs, MANs, and WANs must have established protocols for exchanging information. These protocols describe hardware and software standards to which people comply in developing networks and they govern how nodes on a network communicate.

In a distributed computing environment such as a network, the phrase "client/server" is used to describe how a workstation and a network server interact with one another. A workstation is considered a client of the network communicating with another computer (server). The client computer provides the interface and performs some processes. It accesses data on the server by sending queries which the server processes and answers. Networks using client/server conventions employ what is referred to as a layered protocol structure for communication, which assumes application, data, and physical hardware links are separated into individual layers. Open System Interconnect (OSI) is an international standard used in networking that is based on a layered structure.

Broadband versus baseband transmission

Two transmission methods exist for communicating between clients and servers: broadband and baseband. Broadband transmissions are analog communications using multiplexed signals to provide multiple communications channels. In this mode, links are established between devices and digital data are modulated into analog signals, which are transmitted over network cables to other nodes. The technique employed is called frequency division multiplexing (FDM). Each signal is assigned a discrete frequency over which it is transmitted. The digital signal is modulated and demodulated by a modem. The benefits of broadband transmission include data transfer at long distances and high speed. Also, the use of devices called multiplexers allows for data to be transmitted across many channels simultaneously.

In baseband transmission the data remain in digital form, making modulation and demodulation unnecessary. The signal is varied between a high and a low state (i.e., on or off, zero or one). Most personal computer networks use baseband transmission because it includes less circuitry and is cheaper to install (its cabling is often cheaper). Its disadvantages include limited transmission range and the fact that data flow is restricted to one channel; the whole channel is used by a single information packet. In order to differentiate between data packets, a technique called time division multiplexing (TDM) is used. In this scheme, two or more signals can be sent over the same channel by dividing them into packets and alternating the transmission of the separate signal packets.

Both broadband and baseband transmissions are processed by a workstation's interface to the network, its network interface card. The card varies, depending upon what network topology and protocol are being used to communicate. In the following sections we will discuss how data transmission is influenced by network topologies and protocols using Token Ring and Ethernet as examples.

Token-passing and Ethernet

In a token-passing network, the server transmits a packet called a token to each node. A node can communicate or transmit data only when it is in receipt of an empty token. The node attaches its data to the token as well as a destination (address) and transmits the token across the network. All nodes on the network constantly monitor the data stream to determine if they are the intended recipient of a packet. If a node is the intended recipient, it "grabs" the token and reads the information. A token-passing network uses a bus or ring topology. A Token Ring network is an IBM product that combines token passing with a hybrid star/ring topology and uses devices called media access units (MAU) to interconnect stars. An advantage of token-passing networks is that they preclude data collisions. A node can transmit data only if it receives permission to do so (an empty token). This makes it a strategy suitable for large networks that transmit a high volume of data. Its disadvantage is that as the network grows, it slows down because a token must pass through all nodes of the network along a circular route on its journey toward its destination.

Ethernet networks are configured as a bus topology. In an Ethernet network, all nodes listen before they transmit any data. If a transmission from another node is detected, the listening node ceases its attempts to send data and waits a random interval before trying again. To avoid data collisions, nodes employ a technique called Carrier Multiple Access with Collision Detection (CMA/CD) to check if any other node is currently transmitting. Because larger networks are more prone to data collisions, nodes must also listen for collisions while transmitting packets and cease if a collision is detected. Depending upon the number of nodes, cabling requirements and transmission lengths vary.

Circuit switching and packet switching

Networks pass information packets back and forth using two common strategies: circuit switching and packet switching.

Circuit switching is the familiar strategy employed in telephone communications. In a circuit switched network, a dedicated end-to-end path is established between sending and receiving nodes, which is maintained for as long as data must be exchanged. Circuit switching simplifies data transmission because all information is transmitted at once. It can be inefficient, however, because it monopolizes network channels for periods of time during which other nodes may not make use of that channel. It is useful for applications that must exchange large amounts of data for a long time but isn't appropriate for applications that must share resources or communicate with other machines during a process. A further drawback with circuit switching is the time necessary to establish the circuit. Although, circuits are established in fractions of a second, that time represents a significant period during which data may not be exchanged.

In figure 1.8, a circuit switched link has been established between nodes A, B, and C. As long as the switch is maintained, data transfer is limited to that channel. Nodes F and G can communicate with each other as can D and E, but the switch has cut connections from one side of the net to the other.

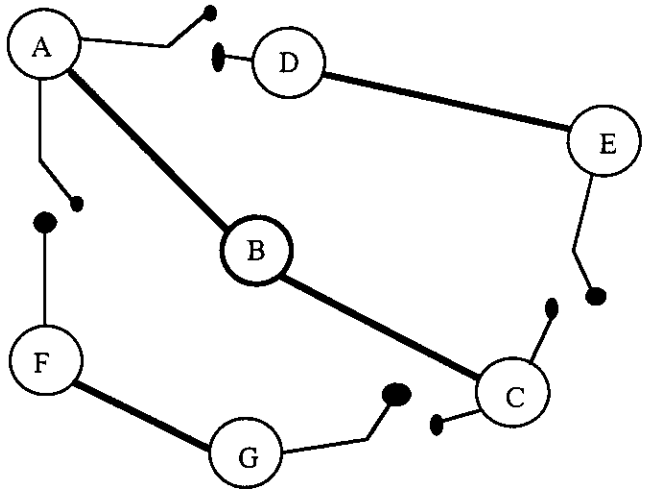


Fig. 1.8. An Example of Circuit Switching.

In a packet-switched network as shown in figure 1.9, lines of communication between all nodes are never severed. In this scheme, data to be exchanged between nodes are broken into packets that contain addressing information (origin and destination), synchronization data, length of packet information, the actual data, and a flag that the packet has ended. Packets are transferred back and forth with each node examining the destination information and forwarding it along the path to the intended recipient. Most WANs employ packet switching for communication.

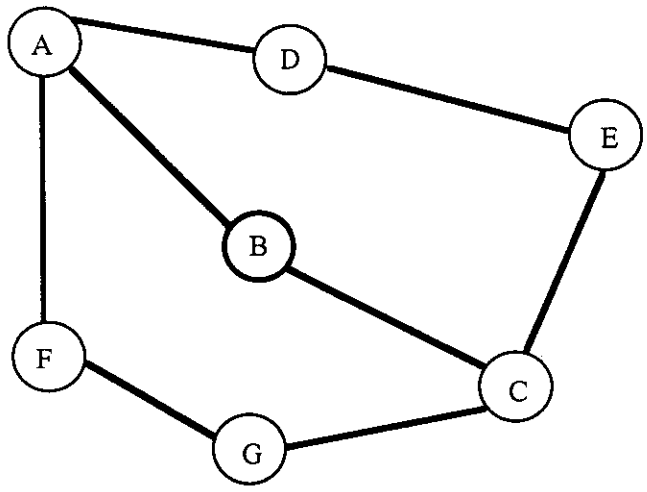


Fig. 1.9 An Example of Packet Switching.

Although congestion may be a problem, all nodes can communicate at the same time. Nodes employ a strategy called store and forward to deal with congestion. A packet may be stored temporarily before it is forwarded to its destination. In figure 1.9, all nodes may exchange packets with their connected neighbors, using them to relay packets to other nodes. Because packets may travel different paths, the receiving end must know how to reassemble the packets into the correct sequence.

Where do we go from here?

Networking continues to evolve, but it certainly appears to be moving toward community-based information delivery systems. Networking

technology is constantly improving the pace and volume with which we exchange information. Less expensive communication links and computers continue to make information available to more and more people, but in order for information to improve the quality of life, networks must provide tangible and measurable benefits.

As networks expand to incorporate more resources, it is becoming increasingly difficult to organize them. A resource may exist which answers an information need of a network user, but if the user is unaware of its existence, the question may go unanswered. Additionally, different information resources may use disparate searching and retrieval schemes, which require an information seeker to master a variety of dissimilar search skills. Work continues on standardizing protocols used for searching and retrieval across diverse computerized platforms. Such a standard protocol is Z39.50, developed by members of the library community and now a U.S. national standard. Z39.50-compliant systems can describe query and retrieval information in a standard manner mutually understandable (and acceptable) by both the client and the server computers, making it possible for different computers to search the same database server.

The rapid growth of the volume of work conducted across electronic media has opened a Pandora's box of complexities associated with intellectual property rights. Who is the intellectual or artistic owner of a work that exists in intangible electronic form? How is ownership defined? Organizations such as the Electronic Frontier Foundation are dedicated to encouraging legal and social institutions to adapt to new technologies while preserving individual constitutional freedoms. They argue that, in the emerging global village, access to information should not be limited to only those who can afford it. Vice President Al Gore describes the global village as benefiting from computer networks by increasing participation in government. This goal presupposes the accuracy of and access to communicated messages across network media. In its present form the network does not verify or lend accuracy to the transmitted materials. Similarly, without changes in entities responsible for the administration of the network, universal access to the network will remain a concept and not a reality.

Perhaps of all the issues surrounding the nature of networking and the creation of any national network, the most pressing concern is that of who will fund its implementation. The proposed infrastructure is very costly. In light of

the pressing concern for curtailing congressional spending, it is highly unlikely that the federal government will absorb the costs. In fact, some argue that to do so would be inappropriate for a government in a free-market economy. Networks do not fund themselves, however, and as the network evolves toward a more commercial enterprise, the issue of cost recovery will continue to develop. If and how fees are to be assessed is a matter of debate in light of the fact that many strongly feel access to information is a fundamental right.

Chapter 2

Evolution of the Internet

As early as 1969, the Defense Advanced Research Projects Agency (then called only ARPA) began investigating the feasibility of connecting heterogeneous computers over a wide geographic area. From the beginning, the goal was to use a packet-switched network, which was thought to be more efficient than a circuit-switched network. It is instructive to review the advantage of one type of network over another. In circuit-switched based networks, an end-to-end data path between two computers must be established which is then totally dedicated to the communication process between two machines. Establishing such an end-to-end path is often a slow process, and because paths have to be dedicated, other machines cannot simultaneously use the same paths. As a result the efficiency of the network is reduced. In a packet-switched network, however, data sets are transformed into individual units of information called packets. A header is then attached to individual packets describing their destination. In this way multiple packets originating from different machines can use the same data paths, helping to increase the efficiency of the network. Furthermore, using the destination information found in each packet, intelligent routing computers placed in between data paths can reroute data to alternative paths when traffic becomes unduly heavy or when path corruption occurs.

The first large-scale packet-switched network, called the ARPANET, connected four sites and was developed by ARPA in the early 1970s, which connected four sites. By the mid 1970's this network had evolved into a fully operational Department of Defense (DoD) computer network. To manage communication between diverse computer platforms with different operating systems and to execute various routing operations between a quickly growing number of nodes, a common command language, mutually usable in different environments, was developed in 1973-74. The Transmission Control Protocols and Internet Protocols (jointly called TCP/IP) are important because

they have, to this day, remained the standard Internet protocols and support such common functions as file transfer (File Transfer Protocols), electronic mail (Small Mail Transfer Protocols), and remote login (Telnet). In the early 1980s two very important events took place: DoD adopted TCP/IP as a standard network protocol, and it funded the Berkeley UNIX development group to incorporate TCP/IP as part of the UNIX environment. This initial partnership between TCP/IP and UNIX meant UNIX would always have an edge over other operating systems in being better able to accommodate Internet-based functions and resources.

Observing the success of ARPANET in linking defense researchers over a wide area, the National Science Foundation (NSF) in 1981 funded the creation of CSNET, a network that would allow connection between scientists, engineers, and researchers. CSNET was primarily based on TCP/IP, and an agreement between NSF and DARPA ensured that users on both networks could share and access resources across the two networks.

In 1984, NSF began to invest in a more grand-scale network project, which evolved into what is today called the NSFNET. NSFNET, from its early inception to its current state, has significantly influenced the structure and function of the Internet. At a topological level, NSFNET was to cover a much broader area than the CSNET (connecting supercomputer centers and NSF-funded institutions across the country) and would explicitly incorporate a linked hierarchy of networks at the national, regional, and campus levels. At the functional level, the NSFNET adopted TCP/IP as its underlying protocol, and the minimum bandwidth was chosen to be 1.5 megabits per second (a T1 network).

The linked hierarchy topology ensured that large, campuswide networks such as those present at UC-Berkeley, MIT, and UT-Austin could connect to larger mid-level networks, such as Barrnet (U.S. Western region), NearNet (U.S. Eastern region) and THEnet (U.S. Southwest region) respectively, and that these mid-level networks could then be connected to even larger networks such as the NSFNET. In this topology, CSNET became a mid-level network and was linked to the NSFNET.

In 1990, DoD decided to discontinue the ARPANET. Consequently NSFNET and a number of other agencywide networks (for example, ESNET - Energy Science Network - managed by the U.S. Department of Energy) became the primary backbone networks (usually at the highest level of the linked

hierarchy) supporting the functions of the Internet. In late 1992 the NSFNET backbone was upgraded to support 45 megabits per second of data transmission (T3 capabilities) to keep up with increased traffic on the Internet.

From the brief review above, two characteristics of the Internet should stand out. First, the Internet is not a single network. It comprises many networks of various sizes with many nodes dispersed geographically. Second, the Internet is massive in terms of the number of networks, computers, and people it connects. There are today more than 15,000 large wide-area networks officially linked to the Internet. Of the total, more than 6,000 are foreign networks. The Internet is estimated to connect nearly a million computers worldwide, and the volume of data traffic, according to a March 1993 statistic[1], can reach as high as 6 trillion bits of information per month. According to John Quarterman, an Austin-based Internet consultant, the size of the Internet has been doubling every year since 1988. Today, the total number of Internet users is approximately 8 million, and of these, approximately a million use the Internet every day.

[1]Statistic collected from nic.merit.edu anonymous ftp site. The file containing the statistical data is named history.bites, to be found under the directory /statistics/nsfnet/.

Chapter 3

A Model Internet Connection

There is no such thing as a typical connection to the Internet. Every user and every microcomputer that can reach Internet resources does so over different channels and through varying types of services. However, the general process for all users is similar.

Most users connect to the Internet through their work or campus network. More recently, additional independent service providers have offered access to individual home computers through dial-up facilities. A generic network model is shown in figure 3.1.

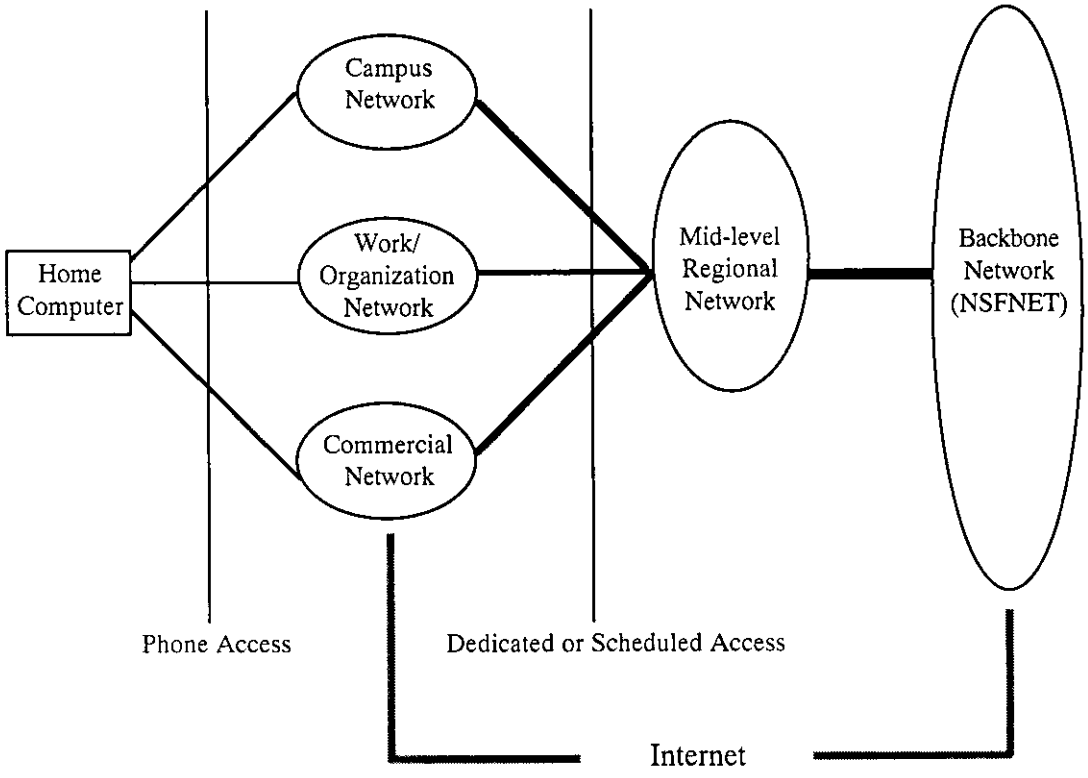


Fig. 3.1. A Generic Wide-Area Computer Network Model.

As figure 3.1 shows, the Internet is not a single network. Rather, it is a collection of networks. Thus the campus network, the mid-level regional network, and the backbone all together are part of the Internet. Usually, data transmission speeds vary between different networks. In figure 3.1 this is indicated by different thicknesses in the links (thicker lines represent faster speeds).

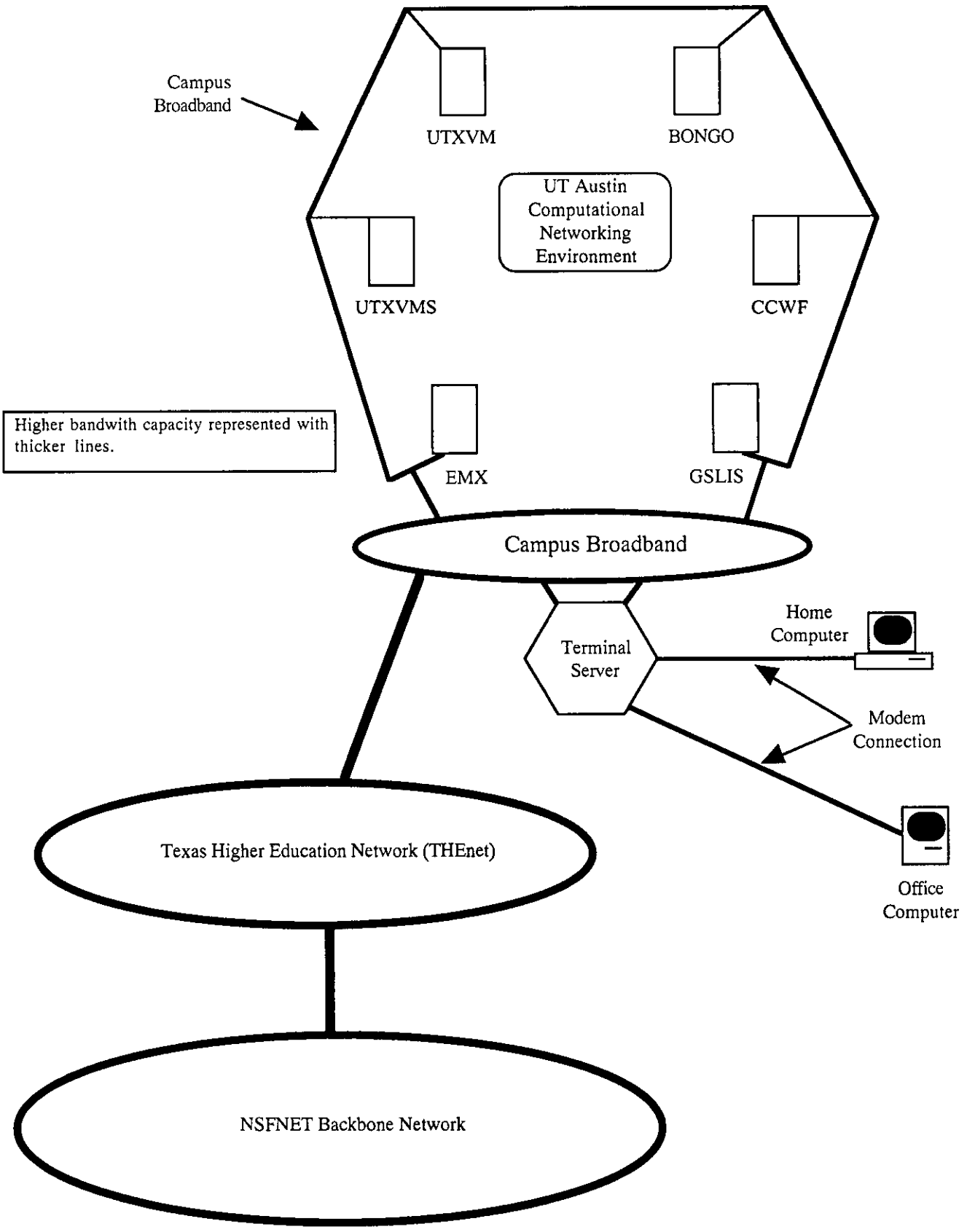
Many typical connections from the home to the Internet make use of university networks. The University of Texas at Austin's campuswide network is a good example of a wide-area network (WAN) and a connection to the Internet. It is composed of several interconnected local area networks (LANs) throughout the campus. A campuswide Ethernet system joins many of these smaller local networks, which include buildingwide or departmental groupings of computer systems. All of these interconnected systems are known collectively at UT-Austin as UTnet.

The most important components of UTnet for Internet access are the mainframes, minicomputers, and workstations of the campus computation center. These systems have a dedicated Internet connection. Through a personal or public-use account on these systems, an individual gains access to Internet resources. The primary computers of UTnet are named EMX, UTXVM, UTXVMS, BONGO, and CCWF (itself a series of workstations). These systems all offer access to basic Internet applications such as FTP and Telnet. Many users choose a certain machine for Internet access because of its particular operating system and user interface. For instance, many people choose to use a UNIX computer because of the compatibility of Internet resources with the UNIX environment and the popularity of UNIX among Internet users.

Access to these computers requires either a microcomputer and a modem or an on-campus terminal with a direct connection to the computation center facilities. A user's terminal or microcomputer connects to the mainframe via an intermediary system such as Micom or Telesys. Micom is a terminal server used to connect on-campus terminals to the various computer systems, and Telesys is used for dial-up modem connections.

The network hierarchies depicted on the following pages chart a connection from a home computer through a university network to a mid-level regional network and onto the NSFNET (considered to be the backbone for Internet traffic within the U.S.). Figure 3.2 portrays the University of Texas

Model Internet Connection



Model Internet Connection

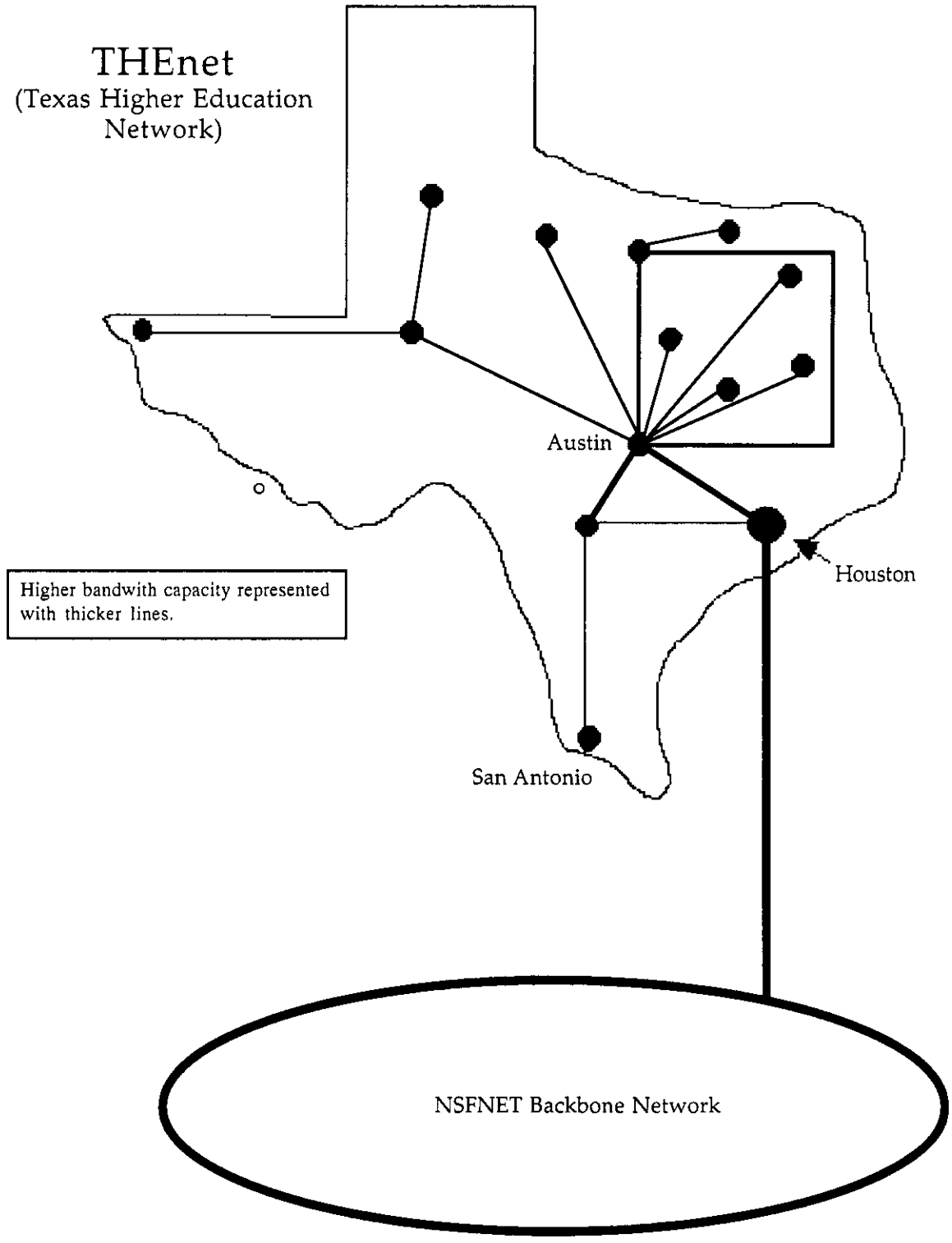


Fig. 3.2. The University of Texas at Austin Computing Environment and Its Connection to the Texas Higher Education Network (THEnet) and NSFNET.

at Austin computing environment and its connection to the Texas Higher Education Network (THEnet, a mid-level regional network).

Another example of a mid-level regional network is the California Education and Research Federation Network (CERFnet), shown in figure 3.3. It is a high-speed data communications network that connects educational and research institutions in California. Like THEnet users California users can connect to CERFnet in several ways. They may have an account on one of the several university networks that are members of CERFnet and access the Internet through a university computer facility. They might access the Internet from a home computer by using a university service to dial into the campus

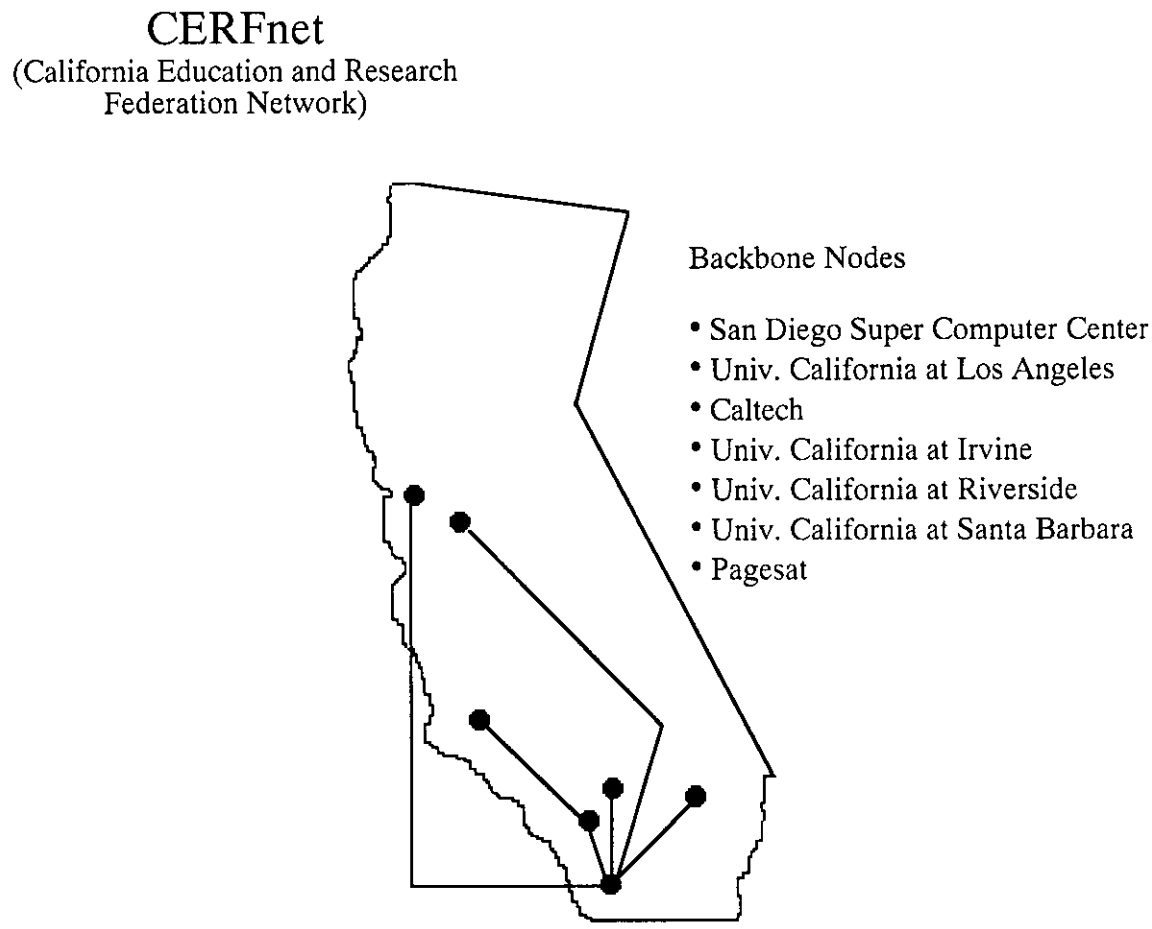


Fig. 3.3. The California Education & Research Federation Network (CERFnet).

network. They might work for an organization that is connected to CERFnet and reach Internet facilities over a work computer. Or finally, they may work for an organization that allows employees to dial into organizational computers from a home computer over phone lines.

The relationship between the previously discussed mid-level regional networks (THEnet and CERFnet) to the U.S. Internet backbone (NSFNET) is shown in figure 3.4.

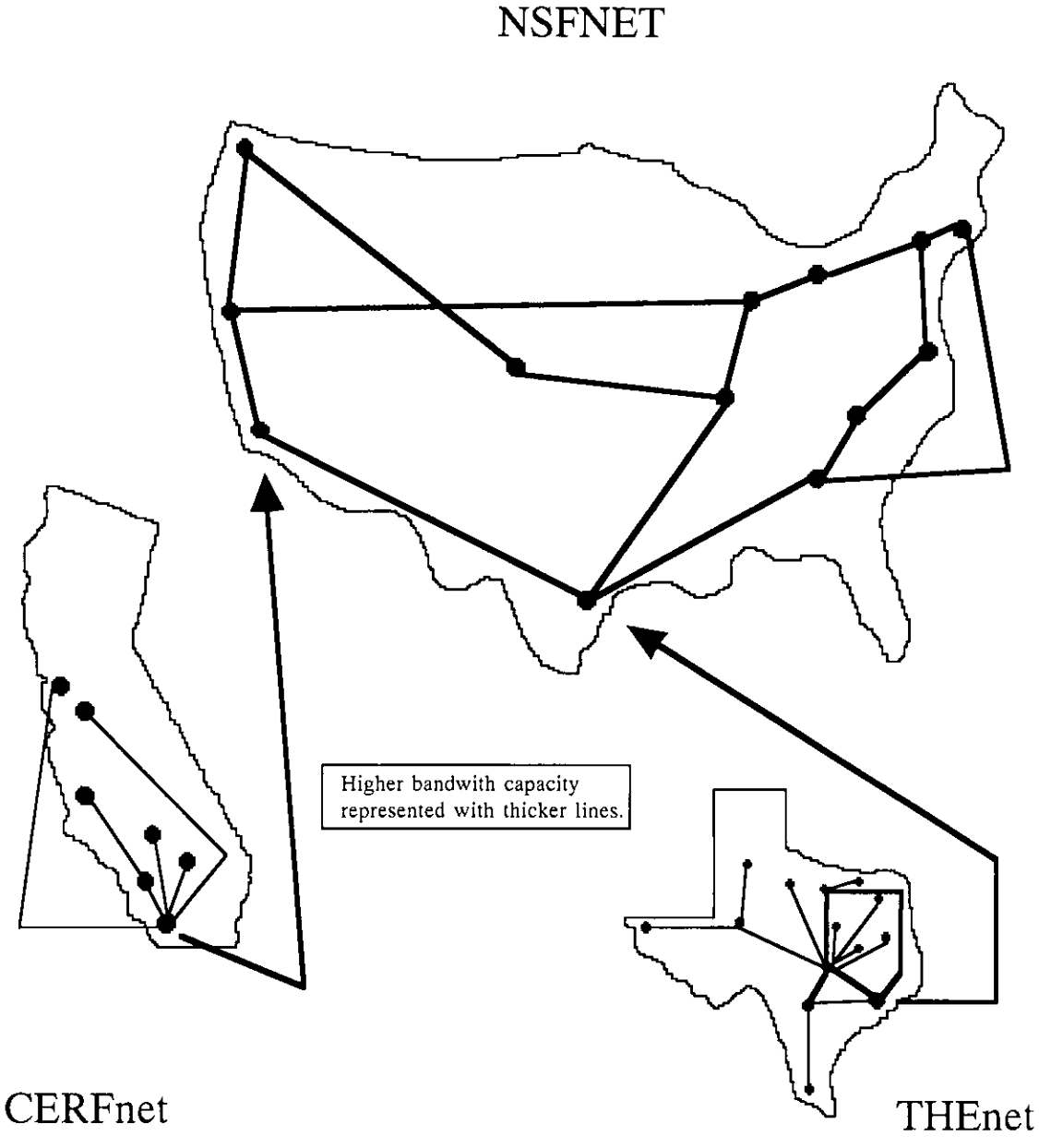


Fig. 3.4. Relationship of CERFnet and THEnet to NSFNET.

Model Internet Connection

Houston is a node of THEnet and also a node of the NSFNET backbone. NSFNET backbone carries the bulk of Internet traffic within the U.S. It provides the link between the regional network (THEnet) and the Internet. The San Diego Supercomputer Center is also a NSFNET node and connects CERFnet and its members to the Internet. Each of the other NSFNET sites depicted performs similar functions for other regional networks.

Chapter 4

Major Internet Resources

This chapter will introduce various information resources that exist on the Internet. To take maximum advantage of information resources, some background knowledge about data sources is necessary. In this section we will attempt to provide that background knowledge, in a general fashion, by covering the major data types, relevant information resources, and various information transfer and searching tools that currently exist in the Internet. More detailed information about how specific information resources can be accessed is provided at the end of this chapter and in Part II of this book (reference section).

Data Types and File Formats

The most prevalent type of data that you will come across inthe Internet is in ASCII (American Standard Code for InformationInterchange) format. The term text is sometimes used to refer to this format to signify the fact that ASCII documents primarily contain textual information (alphabets, numbers, and other commonly used written symbols) and are absent of any special format information (e.g., underline, fonts, and superscript). The absence of format information endows ASCII documents with certain advantages over other types of documents. ASCII-formatted information can be transferred easily from one machine to another with little or no loss in information. Also, ASCII documents usually take up less disk space than documents that contain special format information. In creating a document that will be transferred over the network, strong consideration should be given to ASCII format. Most word processors offer the option of saving documents in ASCII format.

There exist other formats for documents to serve the needs of special use and users. For starters, virtually all word processors have their own formats.

For example, many documents created by government agencies contain statistical information in tabular form. If such documents are created in WordPerfect (word processing software), they would be saved in WordPerfect's own format, not in ASCII, to maintain the format information associated with tables.

Many scientists create documents that contain formulas and illustrations. For these documents ASCII format is inappropriate. The common formats used for such technical documents are usually PostScript, TeX (pronounced "tekh"), and LaTeX. Certain documents contain only graphical or image information, for example, maps, charts, and art materials. Common formats used for such documents include GIF (General Image Format), TIFF (Tagged Image Format), PICT (Internal Macintosh format for images), and the more recently available JPEG (Joint Photographic Experts Group).

Special steps have to be taken to identify a document's format. Sometimes when you access the directory or see a list of files in a remote computer, the suffix information attached to a file name will provide hints about the format of the file: "file.wp" would normally indicate that the file is in WordPerfect format, "file.ps" would indicate the file is in PostScript format, "file.gif" would indicate the file is in GIF format, and so on. However, sometimes such information is missing or is not reliable. In such cases, look for a README file or an INDEX file in the list, whose content may provide the necessary format information about individual files. Another useful method is to use the UNIX "file" command, which when executed on a file name (e.g., file filename) displays the specific format of the file. In all cases, when a file is not in ASCII format you must use the binary mode to transfer a file from a remote source to your own directory.

Disk space is a premium resource in computing. Therefore, various file compression techniques have been developed to economize on storage space. Unfortunately, no standards exist in compression methods, and certain methods are dependent on hardware and operating systems. Some of the common compression formats are ARC (DOS), BINHEX (Macintosh), COMPACTOR (Macintosh), SEA (Macintosh), Compress (UNIX), StuffIt (Macintosh), and PKZIP/UNZIP (DOS). Files found in remote data sources are often compressed using one of these methods. Look for file names such as "file.Z," "file.sit," "file.arc," and "file.zip" to determine if a file is in compressed

form. As in the case of non-ASCII files, the binary mode needs to be used during the utilization of File Transfer Protocol to transfer compressed files from a remote source to your directory.

Data Sources

At least four major types of data sources exist on the Internet. These are Bulletin Boards (BBs), FTP sites, databases, and LISTSERVs.

Electronic BBs

In many respects electronic bulletin boards are similar to actual bulletin boards. Users can "post" messages to a BB which other users can read and respond to by posting their own messages. Many BB's have a moderator (often the person responsible for the creation of a BB) who, based on certain pre established policies, can filter messages that eventually get posted. Any regular user of the Internet can start a BB if enough interest among users can be demonstrated.

USENET is the largest and the most frequently used BB system available through the Internet. It is estimated that approximately 37,000 organizations subscribe to and support USENET. On USENET a user can access approximately 2,000 interest groups (specific BBs devoted to a particular topic and referred to as "newsgroups"), covering such topics as neuroscience and numerical analysis as well as gardening and bicycling. About 1.5 million Internet users subscribe to at least one BB interest group. Depending on the particular organization that provides the Internet account to users, the number and type of USENET newsgroups available at a particular site may vary.

Various functions are available to users of USENET. The major functions are automatic subscribing or unsubscribing, posting replies to messages, and exporting messages to be saved to user's own disk. The rn (readnews) program is used in most UNIX-based computers to access USENET. After accessing USENET, use the h (help) command to view the list of available commands.

FTP Sites

Numerous FTP (File Transfer Protocol) sites are accessible by the users of Internet. An Internet FTP site can include software, documentation,

bibliography, announcements, images, directories, and other types of information. If a file is compressed or is not in ASCII format, you have to set the transfer mode to binary before proceeding with file transfer. Some of the most useful and popular FTP sites are listed at the end of this section. For detailed step-by-step guidelines on using FTP and accessing these sites see Part II of this book.

Databases

Various types of databases are accessible via Internet. Many of these databases do not have any line-charges built into them; in other words, you can search them for as long as you want without incurring any time-based cost. The most prevalent among these are directories, library catalogs, reference sources, and various government databases.

Most of the directories available via the Internet contain addresses and phone numbers of people affiliated with U.S. universities and colleges. These are usually part of Campus-Wide Information Systems (CWIS), which may also include other databases that contain jobs and rules/policy information. Many CWIS can be accessed directly, if you know their Internet address. For example, the address for the University of Texas at Austin CWIS is utinfo.cc.utexas.edu and it is open to public use. To access it, use the Telnet command and login with the user id "UTINFO" (see the reference section for directions on using the Telnet command). Some directories contain information on Internet addresses of individuals who have Internet accounts. There are two such databases that are considered to be among the most comprehensive. One is located in Colorado, accessible by Telnet at: bruno.cs.colorado.edu (login as "netfind"). The other is located at the Network Information Center (NIC), and can be reached at the address: rs.internic.net (type "whois" when prompted).

Hundreds of online library catalogs in the U.S. and 14 other countries are accessible via the Internet. The three most important factors to remember about these catalogs are the following:

- Each of these systems has its own search commands which may not be the same as other systems.
- Special passwords or a user id may be needed to enter some of these databases.

- Certain systems require you to enter cryptic command sequences to exit their systems which, if mis-entered, can prevent you from exiting or even continuing with a new session.

If you are careful in writing down instructions presented in the initial screens, which generally describe how to enter and exit the system, the experience of searching these sources can be highly rewarding.

You may wonder how searching a library that is hundreds or even thousands of miles away can be of any benefit. Actually, several distinct benefits can be achieved from such searching. A library catalog search can confirm the existence of a particular item, it can help you assess the depth and breadth of coverage of a particular topic, and, if you wish, you can use interlibrary loan service to request items to be loaned to you from many libraries.

Government Databases

Two data sources in this category are worth mentioning: FEDIX and STIS. The FEDIX database will be of particular value to university faculty members and academic advisors because it contains many sources of funding and professional development for both undergraduate and graduate students. FEDIX is especially strong in minority programs. It contains government support programs for students divided into the following subject disciplines: architecture, arts, business/management, education, engineering, graphics arts/printing, health/medicine, journalism, military, music, and science. The STIS database is especially useful to researchers in academia, government, and nonprofit institutions who wish to seek National Science Foundation funding information. STIS contains news about NSF committees, NSF announcements, abstracts of NSF programs, and information on awards provided by NSF. Login information for accessing FEDIX and STIS is provided at the end of this chapter.

Charge-based Database Sources

Two very well-regarded commercial online database sources are now available via the Internet. The DIALOG system provides access to more than 400 different databases covering all the major disciplines. Areas of focus include agriculture, biosciences, business, chemistry, computers and software,

engineering, medicine/health care, and physical sciences. DIALOG uses a very simple command structure (including Boolean combinations). A preselected subset of databases can be searched at once, making it unnecessary for the user to reenter searches. Search strategies can also be saved for later execution. The DIALOG marketing department can be reached at (800) 334-2564.

The LEXIS/NEXIS system is a service of MEAD Data Central. This source is particularly strong in legal, government, and news information. It is not an exaggeration to say it is perhaps the world's most comprehensive electronic source for newsprint information. The LEXIS component includes various legal documents (case laws, legislation, regulations, etc.) from all 50 states, the U.K., France, Australia, Japan, and other countries. The NEXIS news information component includes company and industry profiles, product evaluations, individuals in the news, and patent information. Under NEXIS many international, national, and regional newspapers can be searched and accessed within 24 hours of their publication. Wire services from more than 650 worldwide news bureaus are available -- of these, two major sources are updated every 15 minutes. For subscription information call (513) 865-6800. Note that for both DIALOG and LEXIS/NEXIS substantial discounts are available for educational institutions.

LISTSERVs

LISTSERV is a special messaging program created around a particular theme or topic. A LISTSERV can maintain an up-to-date list of subscribers and automatically distribute information to users based on that list. There are many LISTSERVs covering diverse subject fields and interest areas that can be accessed via the Internet. Although many LISTSERVs are now accessible via the USENET system, in actuality LISTSERVs are managed by different software and have unique functions built into them. A user can directly subscribe to a LISTSERV such as the CWIS-L (Campus-Wide Information Systems discussion group) by sending e-mail to that group's LISTSERV address. For example, in subscribing to CWIS-L, a user sends to LISTSERV@WUVMD.BITNET. The subject line of the message would be left empty and the body of the e-mail would contain: SUB CWIS-L *your name*. The name of the user then is added to the user list maintained by the CWIS-L LISTSERV. The user can also send a message to the CWIS-L LISTSERV with the intention of "broadcasting" the message to all the subscribers by sending a regular e-mail to the address CWIS-

L@WUVMD.BITNET. The user, as well as the rest of the subscribers, would then receive an e-mail on their account that would contain the new message. A user can unsubscribe to a LISTSERV by sending an e-mail similar to the subscription message replacing the term "SUB" with the term "UNSUB." Some LISTSERVs can be also be searched for various types of information by sending an e-mail query to the LISTSERV. Using such queries a list of subscribers to a LISTSERV or past messages containing particular keywords can be retrieved. For a list of possible search queries and also examples specific to a LISTSERV, send an e-mail to the LISTSERV of your interest with the words INFO DATABASE in the body of the message. Address the message to the LISTSERV software not the LISTSERV group. For example, to retrieve searching information on CWIS-L address the e-mail to:

LISTSERV@WUVMD.BITNET.

Data Transactions and Transfer

There are three major ways an Internet user can transfer information from one location to another: e-mail, FTP, and BB posting. One should be aware of the specific procedures and outcomes associated with each of these in order to use them successfully.

E-mail & Internet Addresses

E-mail is a limited means of conducting information transactions between one-to-one or one-to-many users. It is limited because the underlying protocol (SMTP) that manages mail traffic is designed to handle relatively smaller data volumes as compared to FTP. That is why e-mail should not be used to send lengthy information files to other users. The PINE system in the UNIX environment has a limitation of about 2,000 lines of text that can be sent using e-mail. A protocol called MIME allows for a user to attach a file to an e-mail message (either textual or binary), facilitating the transfer of larger files of varying types. Not all systems have support for MIME messages, however; you should check with your system administrator for more information.

An e-mail always has a header that contains information on who sent the message, to whom it was sent, and subject information. To send e-mail to someone you must have his or her Internet address. If the person to whom you are sending a message has previously sent you an e-mail, look in the

header of the message to find the address. The address of the sender of a message immediately follows the word FROM: in the header. If you are initiating an e-mail communication with another Internet user for the first time and you do not know his or her e-mail address, then you have two options. The first option is to conduct a search using an Internet address directory. To effectively search the directories, you will need a basic understanding of the Internet addressing scheme. An Internet address usually consists of several distinct parts. For example, the Internet address of the first author of this book is: jm@bongo.cc.utexas.edu. This is based on a domain-level addressing scheme. In the Internet, computers are identified by domain names, subdomain names, and a name for the computer (host). The last part, "edu," is an Internet-wide domain name. The common domain suffixes used in the Internet are listed with explanations in Table 4.1.

Table 4.1 Internet Top Level Domains.

Domain Names	Explanation
edu	Educational institution
com	Commercial institution
gov	Governmental institution
mil	Military institution
org	Other institutions

In the address jm@bongo.cc.utexas.edu, "utexas" stands for a domain name associated with the University of Texas. The "cc" designates a subdomain name, in this case the UT Computation Center. The name of the particular computer in the subdomain is designated by "bongo." Finally, the login id of a particular user is represented by "jm." The directory databases can be searched by domain, subdomain, computer (host), and user ids. The more information you can provide, the more specific the search and the greater the probability of finding the right address. The second option available to you is really rather elementary. It involves calling the person you are attempting to send the message to and asking for the e-mail address or providing your e-mail address to that person and requesting that the person send you a test message. For discussion on a specific e-mail interface known as PINE, please see Part II of this book.

FTP

To send or receive large files over the Internet, FTP should be used. FTP is most frequently used to retrieve useful information available at numerous locations in the Internet, hence, the receive mode of FTP will be covered first.

To transfer files from a particular site, you must have three pieces of information: (1) an Internet address for the site (IP address or domain address), (2) a user id number, and (3) a password. Some FTP sites require you to use a specific login id, while many accept "anonymous" as the login id. Sites that accept ids of the latter type are known as Anonymous FTP sites. Before you FTP to a site, make sure you know where in the directory tree (in the remote site) the file you plan to transfer resides. After logging in to the site, you are frequently required to use directory commands (cd, cdup, cd .. etc.) to move up or down the remote directory to find a file. If you run into trouble locating a file, move back up to the topmost level of the directory and transfer the "README" or "INDEX" file to your account. This file often includes names and locations of individual files. To transfer files to your account you will need to use the get (single file) or the mget (multiple files) command. If you know a file is not in ASCII format (as is the case with software, compressed files, images, formatted text, etc.), you must set the transfer mode to binary.

To send a file to a particular site using FTP, you will need three pieces of information: (1) the Internet address for the site, (2) a user id for a particular account and (3) a password for that account. After logging into that account, you should issue the command: put filename (filename is the name of the file that you wish to send). If you want to send multiple files, you can use the command mput instead of put. Note: the file name can be truncated, so if you wish to send only those files that begin with the term "DOC," you should issue the command: mput DOC*.

Bulletin Boards

A very common way to share information on the Internet is to use the bulletin board (BB) facilities. The most commonly used and the largest among BBs is the USENET system. In UNIX-based environments the rn command initiates a news-reader program that provides direct access to interest groups available via the USENET. However, some environments may provide a different news-reader program. For example, the trn program is a more up-to-date version of rn, which can display a "tree" representation at the top right-

hand corner of the message depicting the origin and the subsequent postings associated with a message. This representation is sometimes called a "thread," referring to the fact that often a BB message is followed by other responses that are connected or linked to each other. Ask your network administrator or just type in trn to see if you can use this news-reader.

The Pnews command (a UNIX command) is used to post messages to a USENET newsgroup. Executing this command is similar to sending e-mail. In the header, you must enter the address of the interest group, provide your address, and include a descriptive subject line. Posting messages to newsgroups is a one-to-many operation. Your message may be seen and read by hundreds or even thousands of users. So certain precautions are in order. Do not use this forum to distribute lengthy documents. The best way to distribute a lengthy document is to upload it to an Anonymous FTP site. "KISS" (Keep It Simple Stupid) is an acronym that is instructive and often referred to when describing network posting guidelines. Keep the prose direct and concise. Also do not assume everybody in the Internet world must think or act exactly like you. If certain postings bother you and you wish to respond in a negative fashion, take some time to think about the message before posting a reply. It is possible to write your message using your own word processor and to include that message in a Pnews document. To do this, you must upload the message to your own account (use ASCII format to create this document) and in the body of the Pnews document use ~r filename (where filename is the name of your file) to include the file in the message. Upload and download operations vary depending on the communication software you use. Consult the user manual or online documentation of the particular software for guidelines.

Data Searching and Access

By now you have probably realized there are numerous sources and types of information to be found on the Internet... maybe too numerous for your liking! The distributed nature of sources and the wide variety of information types can make it quite frustrating to look for material on the Internet. Fortunately, much attention has recently been given to "resource discovery" tools for the Internet. Five such tools deserve attention here: (1) Archie; (2) Wide Area Information Server (WAIS, pronounced "ways");

(3) World-Wide-Web (also known as WWW or W3); (4) Gopher; and (5) Very Easy Rodent-Oriented Net-wide Index to Computerized Archives (VERONICA). Each of these sources is covered in more detail in Part II of this book (reference section). The goal in this section is to provide general background information about these sources, compare them with each other, and point out their weaknesses and strengths.

Archie

This is one of the earliest data searching tools developed for the Internet. Archie was developed by students and network experts at McGill University, Canada. The original purpose of developing this system was to provide a means to search Anonymous FTP sites for software, documentation, and other related information. There are now approximately one million or so computers (hosts) connected to the Internet, and many of these computers provide Anonymous FTP service to distribute files containing numerous types of information. Such information may include satellite imagery, weather data, statistical data, bibliographical data, higher-level data called metadata, purely textual information, and software. Archie is an attempt at an "umbrella" searching tool that covers all these distributed and diverse resources. Once a month, an automated process is run by a computer in Montreal, which sequentially goes through a list of host addresses (a host owner has to register with Montreal to get its host address on the list), collecting from each host directories of files. These directories are then merged into a single list. Archie uses this list for its main searching operation, which involves matching a file name entered by the user against the list; however, if a file name is not sufficiently descriptive or is ambiguous, the list fails to be helpful. To eliminate or alleviate this problem, Archie also provides a "whatis" search. Terms provided in a whatis search are matched against a different list that includes alternative keywords or descriptions of files available in FTP sites. The latter list is manually created by local administrators, usually covering major information resources or tools in a site. However, it is a much more limited data set; thus, it may not include as many sources as the main Archie list.

There are many different sites around the world where Archie lists are maintained and can be searched. These are called Archie hosts or servers. Users are strongly encouraged to choose a site that is closest to their own

geographic location (a list of site addresses is provided at the end of this chapter).

Two very important pieces of information are displayed as output of searches conducted on Archie: (1) the address of the site where the matched file name exists and (2) the directory path under which the file is actually stored. Using this information, one can then perform an FTP session to retrieve the desired file.

Wide Area Information Server

You might wonder, is it possible for someone to search the actual contents of files, not just names of files, available via the Internet? After all, searching the names can only tell you if a file exists or not; it does not guarantee that the file contains what you are actually seeking. WAIS was originally created by Thinking Machines Corporation of Massachusetts to index the contents of files and to provide a retrieval engine based on a massively parallel machine (many processors working at the same time) to search those indexes. However, today there are numerous WAIS hosts around the world (500 or so). These hosts maintain both the documents and the indexes associated with the documents. To search with WAIS you will need access to a WAIS client.

You may run a client on your local machine or you may access clients available on remote machines. In either case, your local machine or the remote machines must be connected to the Internet to access all the databases distributed around the world. Screen-oriented clients (for VT100 or similar terminals) are called swais, and there is a client available for the X-Windows environment called xwais. If you don't have access to a local WAIS client, the best and perhaps the most up-to-date client is available in the Thinking Machines Corporation's host, at the Internet address quake think.com. You must use Telnet protocol to connect and use the password wais to login.

Because contents of files cover a much larger and complex set of information, searching WAIS is slightly more complicated than searching through some of the other tools. When you start up a WAIS client, the first thing you will have to do is select one or more prospective database sources (that may reside anywhere in the Internet) that you think will contain the documents you are looking for. There is a brief phrase describing each database available through the Thinking Machines' swais client. Frequently though,

you will find this brief description of the sources or the names of the sources is not meaningful enough to serve your information need. Fortunately, there is a master index of all the indexes called a Directory of Servers, which is also searchable. Thus, a good approach is to first search this database using broad search terms. Doing so will retrieve a set of databases that WAIS will rank on a numerical scale ranging from 0 to 1000 designating the likelihood that individual databases contain relevant documents (1000 means highly likely). Based on this list and the scores, you then should select a small set of databases that you think will contain the files you are looking for and run a new search based on more specific terms. Avoid using words like and, or, and not which do not contain specific or concrete meaning as search terms. You can reuse the documents that you find by selecting them as relevant documents. WAIS will then use information (terms) from them in the subsequent search cycles to narrow the final retrieved set. The swais client at the Thinking Machine Corporation site can display the documents it finds, which you can then transfer to your own account by executing the mail command and entering your e-mail address.

World-Wide-Web

The W3 system was created with a goal similar to that of the WAIS system - to allow information searching (and access) based on the contents of files, not just their names. However, there are some important differences between W3 and WAIS in their origin and overall functionality. W3 was developed as an information retrieval tool in the High Energy Physics Lab at Cern, Switzerland, to provide access to physics literature. The design was based on the hypertext model. Simply put, in this model, documents in the database are linked with each other based on content relationships, which allows the user to traverse the document space in any order or direction based on those links. In WAIS, there are no explicit links between documents; hence, once you find a document, you don't have the capability to quickly jump to another, related document. In WAIS, the only way to retrieve other documents with some content relationship is to select the current document as a relevant document and to perform a new subject search. In W3, however, starting from the initial screen (which is a document by itself), everything you see or retrieve is linked to each other. On graphical W3 X-clients, linked terms are

underlined, whereas in screen-oriented W3 clients the linked terms are numbered and enclosed in square brackets as [number].

As is the case for WAIS, several client interfaces exist for W3. X-Window clients as well as screen-oriented clients are readily available (can be FTPed) from the anonymous FTP Cern site (info.cern.ch). If you do not wish to or cannot run a W3 client locally on your machine, there is a screen-oriented client available at the Cern site. To access it you will have to use the Telnet protocol to login to the Cern host at NXOC01.CERN.CH (no password necessary). Once you successfully login, W3 allows you to start traversing the document space through several useful webs. The broadest of them are by subject and by type of document. In the Cern site, the first screen will show these two links marked as link [1] and link [2]. If you enter 1, you will then have the option to select a particular subject from several topical links. Just entering the link number and pressing the enter key would then lead you to related documents under each topic. W3 also allows you to link to documents by their format type (indirectly their content type). Table 4.2 shows some of the possible document types searchable by W3.

Table 4.2 W3 Searchable Links by Data Type.

[2] World-Wide-Web	List of W3 servers - by location
[3] WWW initiative	
[4] WAIS	Addresses of WAIS hosts
[5] WAIS Directory of Servers	Searching key words
[6] WAIS Directory of Servers by name	List of servers by name only
[7]WAIS Directory of Servers by domain	List of servers by domain
[8] About WAIS	
[9] Network News	Available directly in all www browsers
[10] Gopher	Campuswide information systems
[11] About Gopher	
[12] Telnet access	List by Peter Scott
[13] Telnet access	List by Scott Yanoff
[14] Telnet access	List by Art St. George
[15] VAX/VMS HELP	
[16] VAX/VMS help gateway	
[17] Anonymous FTP	Tom Czarnik's list of (almost) all sites
[20] TechInfo	A CWIS system from MIT
[22] X.500	Directory originally for electronic mail
[23] WHOIS	A simple Internet phone-book system
[24] Other protocols	Other forms of online data

The W3 way of linking to related documents as a way to access and retrieve them can become onerous when all you want to do is quickly find out if certain documents are even related to a particular topic. Hidden deep in Table 4.2 is a clue related to searching W3 using subject terms. This clue is option [5], WAIS Directory of Servers. The W3 system and WAIS can actually be used together to locate information. When you select [5], you will then be prompted for keywords. W3 takes these key words and automatically runs a search against the Directory of Servers (in a WAIS site) and returns a listing of sources that contain the keywords that you entered. In addition, many W3 sources linked via topical headings (see the resources section at end of this chapter) have a find option built into them allowing you to search these individual sources.

Gopher

This Internet resource discovery tool is the friendliest and the broadest in scope in terms of the number of different functions it covers. Gopher was developed at the University of Minnesota (home of the "Golden Gophers") to provide access to distributed campus information resources. In the Minnesota environment, the computing staff found that diverse resources (online

catalogs, directories, software, and other files) existed in several different locations on campus. They thought that, by organizing these resources in a fashion that allowed users to access them under a uniform menu interface, their use could be enhanced. The Gopher system was developed using the client-server model, so that the task of the interface and the task of data storage-retrieval could be shared among different computers. The client-server design allows Gopher to be used on one computer while the data are stored and searched on a different computer. Soon people realized this underlying model allowed Gopher to be applicable to the Internet information environment in general, and this has led to the swift growth in Gopher use in the Internet.

Various versions of Gopher client exist for different computer platforms. There is a Macintosh, PC, and a X-Windows client available that can be received for free by Anonymous FTP from boombox.micro.umn.edu under the directory pub/gopher. For the purpose of description here, we are going to refer to the client that exists in a University of Minnesota host, but many other publicly accessible Gopher hosts are known to exist around the world.

It is highly recommended that you check with your local computer administrator first to see if a Gopher client is available at your site. If such a client is available then all you have to do is type "gopher" at the UNIX prompt to start it (remember, UNIX is case-sensitive and the command must be in lowercase). Otherwise, to access a client available through the sites listed at the end of the chapter you have to use the Internet address of that Gopher client with the command "gopher."

Every Gopher site maintains menus that point to various resources available at that site or other known sites. In "Gopherspace" (all of the resources known to the Gopher client), a resource can be a file, an FTP site, a directory, an index with a simple query interface, or an interactive tool accessible via the Telnet protocol. When you first connect to the University of Minnesota site you will see a home menu that lists various resources using topical or descriptive terms in a menu format. At the topmost level menu lines end with a period "." or with the symbol "/". A "." indicates that the item represented is a file, whereas a "/" indicates there are other menus under that option. Gopher also provides more specific information about each of the menu items. To see this information you will have to use the arrow keys to move to an item and press the "=" symbol on your keyboard. This will result in a display that tells you which host computer contains the item, the directory

path, and also the name of the item. You will also see a field named "Type" with a coded number or an alphabet. Table 4.3 explains the codes.

Table 4.3. Gopher Codes.

Type =	0	means	Item is a file
	1	"	Item is a directory
	2	"	Item is a CSO phone-book server
	3	"	Item contains Error
	4	"	Item is a Binhexed Macintosh file
	5	"	Item is a DOS binary archive
	6	"	Item is a UNIX uuencoded file
	7	"	Item is an Index-Search server
	8	"	Item points to text-based Telnet tool
	9	"	Item is binary file that the client must read until end
	T	"	Item points to TN3270 tool
	s	"	Item is a sound file
	g	"	Item is a GIF file
	M	"	Item is a MIME file
	h	"	Item is of html type
	I	"	Item is of image type
	i	"	Item is of "inline" text type

Based on the specific type of information associated with a menu option (found by using "=" as described above) or information represented in the option itself you can then make a choice as to what option is worthwhile to use. To select an option all you have to do is move to the right option and press enter. The Minnesota Gopher client also knows about other clients around the world. If you wish to use a different client, select the option titled "Other Gopher Servers," and this will result in a display of options, arranged under geographical categories, pointing to other Gopher clients worldwide.

Gopher is not just a pointing device. It actually automates (and hides) a lot of underlying processes necessary in interaction and retrieval. For example, when you find a file that you want you can ask Gopher to mail it to you or to save it under a file name that you specify. What takes place in such a situation is an FTP session, the details of which you don't see. Gopher can actually run searches against FTP lists (like Archie) or use WAIS to find files for you (these are the type 7 options), with keywords entered by you. You don't have to logon to an Archie server or WAIS server to do the searches yourself. Furthermore, once it finds sources it lists them for you, as options on a menu, and you can then request Gopher to save the related files or to mail the files to you. Certain client software interfaces are smart enough to recognize the file type (binary or

ASCII), and they indicate the file type in the option. Then you can decide what to do with the file once you download it.

Gopher also eases the pain of performing Telnet sessions with interactive tools by organizing them under descriptive terms and automatically connecting to them any time you select them. Among the various interactive tools available via Gopher, the directory services are especially useful. The directories are sometimes listed as phone-books in a Gopher menu. Many other interactive tools such as library catalogs, various government and educational databases, and games are accessible via Gopher (although theoretically virtually anything that can be stored and executed on a host computer and has a Gopher link is open for use).

Very Easy Rodent-Oriented Net-Wide Index to Computerized Archives (VERONICA)

You will recall that we discussed Archie as our first data searching tool. You know that Archie is a tool that allows you to find out what files are available at FTP sites. Now, here is an Internet trivia question: What is an Archie for Gophers? (Clue: It is related to Archie in more than one way!). The answer is: VERONICA. It is the newest and the simplest of the Internet data searching tools. The analogy to Archie is instructive. VERONICA regularly collects all of the Gopher menus from hosts it knows about and merges them into a searchable list. This list is kept in a server, which is then accessed and used by VERONICA every time a user requests a VERONICA search. VERONICA was originally developed at the University of Nevada, but now is available to many Internet users through three servers: one at the University of Nevada- Reno, one at Coalition for Networked Information (CNI), and one at New York State Education and Research Network (NYSERNet).

At the moment, VERONICA is not available as an independent client software; in other words, you cannot simply use the Telnet protocol to connect to a VERONICA server and start a VERONICA session. VERONICA is currently only available through a Gopher client. To use it you must find the nearest Gopher client you can access and look for a menu option for VERONICA. If your Gopher client does not have VERONICA as an option, then you must find another Gopher client that has it. For instance, the Gopher client at the University of Nevada has a VERONICA option. VERONICA prompts the user for keywords and then performs a simple Boolean search

against menu entries in the merged list. Both the Boolean "and" and "or" searches are possible. For example, to find menu items that contain the words gopher+ and pictures, the user should enter:

gopher+ and pictures

To search for items that can have either the word WWW or the word W3, the user should enter:

WWW or W3

If the word and is left out, then VERONICA assumes, by default, it is a Boolean "and" search. Currently, truncated word searching is not possible. The result of a successful VERONICA search is always a list of Gopher menu options. You can traverse these options, find information about the resources described by each option, and connect to the resources the options represent just as you can in any Gopher menu page. Once VERONICA retrieves the matched menu options, you can proceed as if you had just moved from one Gopher menu page to another.

Some Examples of Sources

Now that you are familiar with the basic types of Internet resources, you may be wondering about some specific information sources that are currently available. Following are some examples of useful resources that may be accessed through various Internet channels. Most of the information sources will be of interest to academic users wishing to conduct research with the aid of Internet. However, sources aimed at more general interests are also included. For step-by-step instruction and more detailed guidance on accessing the individual types of resources, please consult Part II of this book (reference section).

USENET Groups

Here is a sample of the variety of USENET groups available.

Name of Group	Purpose
General	
misc.jobs.offered	Job notices
rec.humor.funny	Humor
misc.forsale	Items for sale
misc.forsale.computers	Computers for sale
alt.sources	Share and inquire about software codes
news.announce.newusers	Announcements of new users
news.announce.conferences	Conference announcements
news.announce.important	Important news announcements
Subject Oriented	
sci.aeronautics	Aeronautics
sci.anthropology	Anthropology
sci.classics	Classic literature
sci.edu	Science education discussions
sci.math	Math topics
sci.physics	Physics
sci.psychology	Psychology
sci.research.careers	Discussions on scientific research careers
sci.space	Space
sci.bio.technology	Biotechnology
sci.environment	Environmental issues

FTP Sites

Some of the most useful FTP site addresses are listed below, along with their purpose.

Address	Purpose
ftp.nisc.sri.com	Network-related educational information
rs.internic.net	Network-related educational information
wuarchive.wustl.edu	DOS and Mac software
simtel20.army.mil	DOS software
sumex-aim.stanford.edu	Mac software
mrcnext.cso.uiuc.edu	Electronic books

LISTSERVs

A comprehensive list of LISTSERVs can be created by accessing the following list of LISTSERVs available via BITNET. Apply the steps below.

1. ftp ftp.cc.berkeley.edu
2. Login as anonymous
3. Change directory to /netinfo/bitnet
4. Get the file listserv.groups

Following is a small sample of some academic LISTSERVs. Although many of these LISTSERVs cover a broad topic or a general academic discipline, this is not true of all LISTSERVs; many are very specific and often represent focused subsets of the more general discussion groups.

List Name	Topic Covered
ANTHRO-L@UBVM.BITNET	General topics in anthropology
ARCH-L@DGOGWDG1.BITNET	General topics in archaeology
FOLKLORE@TAMVM1.BITNET	General topics in folklore
DESIGN-L@PSUVM .BITNET	Design issues in art and architecture
FINEART@RUTVM1.BITNET	Discussion of the fine arts
HUMANIST@BROWNVM.BITNET	Humanities and computer study
CJUST-L@IUBVM.BITNET	Criminal Justice
LAWSCH-L@AUVM.BITNET	Issues affecting law school students
MACPSYCH@STOLAF.EDU	Macintosh use in psychology research
IPE@CSF.COLORADO.EDU	International political economy

List Name	Topic Covered
AFRICA-L@BRUFPB.BITNET	African cultural and political topics
AMERSTDY@MIAMIU.BITNET	American studies
CHINA@PUCC.BITNET	Broad discussions about China
JPINFO-L@JPNSUT00.BITNET	General discussions about Japan
HISTORYA@UWAVM.BITNET	General discussions on study of history
VWAR-L@UBVM.BITNET	Discussions about the Vietnam War
RUSSIAN@ASUACAD.BITNET	Topics related to the Russian language
CLASSM-L@BROWNVM.BITNET	Classical music discussions
POLI-SCI@RUTVM1.BITNET	Current American and world politics
STATEPOL@UMAB.BITNET	Politics of the American states
BEHAVIOR@ASUACAD.BITNET	Child psychology and behavior disorder
HEGEL@VILLVM.BITNET	The philosophy of Hegel
RELIGION@HARVARDA.BITNET	Religious studies
COMPOS01@ULKYVX.BITNET	Computers and writing composition
ASTR-O@BRFAPESP.BITNET	Astronomical observations
CHEME-L@PSUVM.BITNET	General topics in chemical engineering
ORGCHE-L@RPIECS.BITNET	Topics in organic chemistry
MECH-L@UTARLVM1.BITNET	Topics in mechanical engineering
QUAKE-L@NDSUVM1.BITNET	Earthquakes
PHYS-STU@UWF.BITNET	Issues affecting physics students

A comprehensive list of all BITNET LISTSERVs is available through any LISTSERV program. The procedure is similar to the subscription process. Simply send a e-mail message to a LISTSERV address (such as LISTSERV@PSUVM.BITNET) with a blank subject line and the command LIST GLOBAL in the body of the text. The complete list will be sent via e-mail.

There are countless other "lists of lists" available on the Internet. Individuals or organizations will often compile a short list of discussion groups related to their discipline and distribute it to their colleagues. An Archie substring search for "lists" or "listserv" will find dozens of these.

An especially ambitious subject-oriented list is maintained at Kent State University. The ACADLIST series of files catalogs hundreds of academic LISTSERVs, and arranges them by the various academic disciplines. A user guide, index, and the files themselves are available via Anonymous FTP from KSUVXM.KENT.EDU or for BITNET users via e-mail using the following procedure:

1. Send e-mail to LISTSERV@KENTVM.KENT.EDU
2. Leave the subject line blank
3. Input the message: GET *File name File type*

Where,

File name	*File type* is:
ACADLIST	FILE1 (Anthropology and Education)
ACADLIST	FILE2 (Geography and Library and Information Science)
ACADLIST	FILE3 (Linguistics and Political Science)
ACADLIST	FILE4 (Psychology and Writing)
ACADLIST	FILE5 (Biological Sciences)
ACADLIST	FILE6 (Physical and Computer Sciences)
ACADLIST	FILE7 (Business and Academia news)
ACADLIST	Changes (Major additions and deletions)

For all new LISTSERV sources:

Send an e-mail with the words SUB NEW-LIST in the body of the message to LISTSERV@NDSUVM1.BITNET

Online Library Catalogs

Some of the most significant library catalogs are listed below along with annotations about their strengths and their Internet addresses.

Library of Congress. The largest library in the world. By law, it receives a copy of every book published in the U.S. (enter telnet dra.com).

Harvard University's HOLLIS system. The largest university library in the world (enter telnet harvardc.harvard.edu).

University of California's MELVYL system. Contains records of all the institutions in the University of California system, including Berkeley, UCLA, UCSD, and others (enter telnet melvyl.ucop.edu).

Dartmouth College. Provides access to "nonconventional" electronic information, including the full text of Shakespeare's plays, Shakespeare's sonnets, the Bible, and the Koran (enter telnet lib.dartmouth.edu).

Colorado Alliance of Research Libraries (CARL). Contains library records of all the large research libraries in Colorado. CARL also provides access to book reviews and Internet guides (enter telnet pac.carl.org).

Gophers

Listed are six major Gopher clients with addresses, login IDs, and locations.

Host	IP address	Login IDs	Location
consultant.micro.umn.edu	134.84.132.4	gopher	North America
ux1.cso.uiuc.edu	128.174.5.59	gopher	North America
info.anu.edu.au	150.203.84.20	info	Australia
gopher.chalmers.se	129.16.221.40	gopher	Sweden
tolten.puc.cl	146.155.1.16	gopher	South America
ecnet.ec	157.100.45.2	gopher	Ecuador

The University of Minnesota Gopher (consultant.micro.umn.edu) has one of the best selections of reference sources among Internet Gopher servers. One of these sources is the periodic table of the elements. The interface is a simple menu system: Choose an element and view a short table of its basic properties and characteristics. The table is in the "Libraries" section under "Reference Works."

Also in this section of the UM Gopher are The Hacker's Dictionary (very useful for those trying to become familiar with computer jargon) and The American English Dictionary.

WAIS Databases

The online CIA World Factbook is available as world-factbook via WAIS. It provides basic information (description, statistics, etc.) about the world's nations. Also, full-text listings of U.S. Supreme Court decisions are available as supreme-court via WAIS.

W3 Sources

The following list shows the W3 topic links available via Internet.

W3 Searchable Links by Topics

[2] Aeronautics Mailing list archive index
[3] Agriculture
[4] Separate list, see also Almanac mailservers
[5] Astronomy and Astrophysics - Abstract Indexes
[6] Astrophysics work at FNAL
[7] Bio Sciences
[8] Computing
[9] Engineering
[10] Environment
[11] Geography - CIA World Fact Book
[12] India: Miscellaneous information
[13] Thai-Yunnan: Davis collection
[14] Law
[15] U.S. Copyright law
[16] Uniform Commercial Code
[17] Libraries - Lists of online catalogues, etc.
[18] Literature & Art
[19] Mathematics CIRM library (French)
[20] The International Journal of Analytical and Experimental Modal Analysis
[21] Meteorology - U.S. weather, state by state
[22] WAIS weather (around MIT)
[23] Music - MIDI interfacing
[24] Song lyrics (apparently disabled for copyright reasons)
[25] Philosophy - American Philosophical Association
[26] Psychology - " Psycoloquy " electronic journal
[27] Physics - High Energy Physics
[28] Astrophysics abstracts
[29] Space Science
[30] Politics and Economics U.S. politics
[31] Reference Roget's Thesaurus
[32] Experimental English dictionary
[33] Religion The Bible (King James version)
[34] The Book of Mormon
[35] The Holy Qur'an
[36] Social Sciences - Coombs papers archive
[37] Journals of the scholarly communications

Archie

The following list shows Internet addresses and geographic locations of Archie hosts. After connecting, login as Archie.

Archie Host Addresses	Region
archie.rutgers.edu	Northeastern U.S.
archie.sura.net	Southeastern U.S.
archie.unl.edu	Western U.S.
archie.ans.net	ANS sites only
archie.au	Australia and Pacific
archie.funnet.fi	Europe
archie.doc.ic.ac.uk	United Kingdom

Freenet

Here are some existing Freenet systems:

System	Area Served	IP Address	Login
Cleveland Freenet	Cleveland, OH	freenet-in-a.cwru.edu	guest
Heartland Freenet	Peoria, IL	heartland.bradley.edu	bbguest
Lorain County Freenet	Elyria, OH	freenet.lorain.oberlin.edu	guest

A common menu item on Freenets is the "Freedom Shrine" or "Freedom documents." This is generally a collection of important documents in the development of western democracy, with an emphasis on American government documents. This is a useful tool for those studying American history or political science.

DIALOG and LEXIS/NEXIS

If you already have DIALOG and LEXIS/NEXIS accounts use the following information to connect to these sources over the Internet.

DIALOG
Enter: telnet dialog com

LEXIS/NEXIS
Enter: telnet meaddata.com

Another way to access DIALOG and LEXIS/NEXIS as well as other charge-based database sources is through MichNet. MichNet is an Internet service that provides a gateway to many data sources, several of which are commercial, fee-based systems. In addition to DIALOG and LEXIS/NEXIS, MichNet offers connections to CompuServe, BRS, and OAG (Official Airline Guide). BRS is a database provider similar to DIALOG. OAG is a service that allows users to access worldwide flight schedules and check arrival and departure times of flights. CompuServe is a much broader service, aimed at the home user, which offers several information resources, many of which are the very sources listed above. For example, OAG is available to CompuServe

subscribers, as well as Knowledge Index, which is a limited, after-hours version of DIALOG. Remember, these sources are not like the free services discussed elsewhere in this manual. To use commercial systems you must be a registered user with an official account.

To access these and other services through MichNet:

1. Telnet to hermes.merit.edu
2. Enter the name of the desired service at the Which Host? prompt.

MichNet has an extensive help system, accessed by typing HELP at the Which Host? prompt. The help system will provide a complete list of known hosts if you do not remember the exact name of a service.

Other Resources by Telnet

Some databases may be accessed by a direct Telnet to the database site, without using WAIS, Gopher, or other tools. One such database is the Dante Project at Dartmouth University. This database is a collection of full-text reviews of Dante's writings. This database is particularly useful because it contains reviews from several different periods in history; many of the reviews are historical documents in themselves. To access the Dante project, Telnet to library.dartmouth.edu and at the first prompt enter "connect dante."

Another database accessible via Telnet is made available by the Environmental Protection Agency. The EPA has databases on Clean Lakes, Hazardous Waste, and Chemical Collection Systems, which may be accessed at epaibm.rtpnc.epa.gov (menu item 6, "Public Applications").

These are just a few examples of the types of services available on the Internet. Keep in mind that these information sources are added and removed frequently from Internet sites. The best strategy for dealing with this impermanence is to familiarize yourself quickly with broad-based searching tools such as Archie and VERONICA, which should allow you to track the disappearance and emergence of the various Internet resources.

Chapter 5

Gaining Access to the Internet

There are several different methods that can be used to connect to the Internet, each providing a different level of access and varying degrees of functionality. These range from a simple electronic mail gateway for personal use to a high-capacity dedicated connection for an institution, providing full access to the network as an Internet host. Equipment and service costs vary accordingly. Because of the wide range of options available for Internet connection, you must not decide on a service until you ask the question: "What exactly do I need from the Internet?" If you can answer this question specifically, then you have completed the most important step in the decision making process. The rest of the process is simply a matter of researching appropriate providers to determine the most reliable and cost-effective service.

INDIVIDUAL ACCESS

As an individual, your access will probably involve a dial-up connection from your home or office. There are many different levels of dial-up service, but all require two essential pieces of equipment: a personal computer and a modem.

What kind of computer do I need?

Any computer that supports a modem connection and runs telecommunications software will be sufficient. However, before investing in a new computer, you must consider all possible applications you may have for the equipment, and decide on specifications such as memory and disk storage accordingly. Internet users connect from home with a variety of computers, from state of the art Macintosh and IBM-compatible systems to "ancient" computers such as Apple IIs or IBM XTs.

Although these computers have different operating systems and run different software programs, most Internet applications are not affected. They need only conform to a few basic standards, and have a modem connected.

Why do I need a modem?

A modem is a device that allows your computer to use standard voice phone lines to communicate with remote computers. Modems are relatively inexpensive devices that have one connector for your computer and a regular phone jack to accept a telephone cord. The modem can share the same line used for standard phone calls, coexisting with the phone system in much the same way as an answering machine or fax machine.

The main factor to consider in purchasing a modem is the speed (measured in bits per second). Although there are other factors such as error correction and data compression capabilities, the overriding factor affecting the price of the modem is its speed. Modem speeds range from 300-19,200 baud and range in price from $40 up to $1,000. Although many dial-up networks support baud rates as low as 300, it would not be wise to invest in a modem slower than 2400 baud. The growing standard for home modems is actually 9600 baud, but the price gap between 2400 and 9600 baud modems is still large enough to seriously consider both options before making a purchase.

I have the equipment. What comes next?

Once you have the proper equipment set up at home, you must arrange for service with an Internet provider. The most convenient access is usually obtained through some kind of institutional affiliation, such as school or work. For instance, if you are a student at a university, you may be able to obtain an account on an Internet-connected machine. If your college, university, or private company has an Internet-connected computer or network, then gaining personal access is as easy as obtaining a personal account on the network.

If you have no such affiliation, then you must shop around for an independent Internet service provider. In making this decision, you must outline exactly what kind of service you need. For example, if your only need is to exchange e-mail with selected people on the Internet, then you may consider getting an account with a commercial information service such as

CompuServe, America Online, or MCI Mail, all of which provide e-mail gateways to the Internet.

If you need more connectivity, such as file transfer and remote login, then you will need terminal access to an Internet-connected computer. This is usually the level of service provided through an institutional affiliation. Several independent organizations also provide this service, with varying fee schedules and access procedures. Which one you choose depends on where you live, what level of service you desire, and your predicted usage patterns. Table 5.1 lists access providers, including contact and service information.

Table 5.1 Internet Access Providers.

National

Service Provider	Service
Advanced Networks and Services (ANS) 2901 Hubbard Rd. Ann Arbor, Michigan 48105 (313) 663-7610	Dedicated service; 1.5 to 45 Mb Worldwide
Performance Systems International (PSI) 1180 Sunrise Valley Dr. Suite 1100 Reston, Virginia 22091 (703) 620-6651	Dedicated service; 9.6 Kb to 1.5 Mb SLIP, UUCP, Worldwide
Sprint International 13221 Woodland Park Dr. Herndon, Virginia 22071 (703) 904-2156	Dedicated; 9.6 Kb to 1.5 Mb Worldwide
UUNET Suite 570 3110 Fairview Park Dr. Falls Church, Virginia 22042 (703) 204-8000	Dial-up, dedicated; 9.6 Kb to 1.5 Mb SLIP, UUCP

Regional Service Providers	Services
West	
BARRNET Pine Hall Room 1105 Stanford, California 94305-4122 (415) 723-3104	Dedicated; Dial-up, SLIP
CERFnet P.O. Box 85608 San Diego, California 92186-9784 (619)-455-3990	Dedicated; 14.4 Kb to 1.5 Mb Dial-up (local and 800 number), SLIP
Northwest	
NorthWestNet 2435 233rd Place, N.E. Redmond, Washington 98053 (206) 562-3000	Dedicated; 56 Kb to 1.5 Mb
Southwest	
THEnet Texas Higher Education Network Information Center Austin, Texas 78712 (512) 471-2444	Dedicated; 1.5 Mb Dial-up, SLIP
Sesquinet Office of Networking and Computing Rice University Houston, Texas 77251 (713) 527-4988	Dedicated; 9.6 Kb to 1.5 Mb SLIP

Central

MIDnet
29 WESC
University of Nebraska
Lincoln, Nebraska 68588
(402) 472-5032

Dedicated; 56 Kb to 1.5 Mb

Westnet
601 S. Howes, 6th Floor South
Colorado State University
Fort Collins, Colorado 80523
(303) 491-7260

Dedicated

Midwest

OARnet
1224 Kinnear Rd.
Columbus, Ohio 43085
(614) 292-9248

Dedicated
SLIP

netIllinois
Bradley University
1501 W. Bradley Ave.
Peoria, Illinois 61625
(309) 677-3100

Dedicated; 9.6 Kb to 1.5 Mb

Southeast

VERnet
Academic Computing Center
University of Virginia
Charlottesville, Virginia 22903
(804) 924-0616

Dedicated
Dial-up, SLIP

SURAnet
1353 Computer Science Center
8400 Baltimore Blvd.
College Park, Maryland 20740-2498
(301) 982-4600

Dedicated; 56 Kb to 45 Mb

Northeast

NYSERNet
111 College Place
Room 3-211
Syracuse, New York 13244
(315) 443-4120

Dedicated; 9.6 Kb to 1.5 Mb
Dial-up, SLIP

NEARnet
BBN Systems and Technologies
10 Moulton St.
Cambridge, Massachusetts 02138
(617) 873-8730

Dedicated; 9.6 Kb to 10 Mb
SLIP

Lists of providers can also be found in several books, including *The Whole Internet User's Guide and Catalog*, by Ed Krol, and *Internet: Getting Started*, published by SRI International. Additionally, a list is available through the Internet using Anonymous FTP at the site "ftp.nisc.sri.com" in the directory "netinfo" under the file name internet-access-providers-u.s.txt.

Another less common option for individual Internet access is through a Freenet. A Freenet is a community service, offering a dial-in information system to the local community. Much of the system consists of unique databases and e-mail systems, but Internet access is always offered to some degree.

Freenets are not widespread, although they are gaining in popularity. If you are lucky enough to live in an area served by a Freenet, then your access problem is solved! (See Chapter 4 for Internet addresses of major Freenets.)

Future trends

There is a growing movement toward free dial-up Internet access. This is not to be confused with a Freenet, which is an information service in itself and offers Internet connectivity. This new type of access merely provides a point of access to the Internet, and offers no user support or value-added services. The providers are often public universities, and sometimes local computer user groups. The resulting service is usually similar to the "institutional affiliation" mentioned above, but without the training and support. This is a recent development, but it is worth investigating in your area as a possible option for Internet access.

Institutional Access

Connecting an institution to the Internet is more complicated than connecting an individual. As with individuals, there are different levels of connectivity to be considered, with similar price/performance trade-offs. The same question an individual asks concerning connecting procedure must be asked by the institution: "What do we need from the Internet?" A manager or system administrator considering an Internet connection will have to weigh such things as existing LAN traffic, number of users, and support services needed. Many of the complications of an institutional connection are related to the nature of connecting a network of computers, rather than a single computer, to the Internet. Problems often arise in trying to reconcile existing network protocols with the TCP/IP standard of the Internet.

Where can I start?

Some organizations contemplating an Internet connection will often test the waters before implementing a costly overhaul of their network or investing in an access provider. For example, a community library may dedicate an existing personal computer for simple dial-up sessions to a relatively inexpensive public access system. Patrons and staff may be introduced to Internet resources, and the level of interest can be measured before making a substantial investment in a direct connection.

We definitely need a connection. What are some options?

UUCP (UNIX to UNIX Copy Program). UUCP is a very basic, but inexpensive means to access some Internet resources. Through a UUCP arrangement, your local computer temporarily connects to an Internet host and transfers information such as USENET news and electronic mail. The connection is made periodically and the frequency of the connection largely determines the usefulness of the service. For instance, if your system only downloads mail every other day, it is not much better than the postal service. If, however, the UUCP connection is made every three hours, it will provide much more timely information.

SLIP (Serial Link Internet Protocol). SLIP is a unique service offered by many full-access providers. Through a SLIP service, a personal computer may become an Internet host, without investing in a direct connection. The connection is handled over standard phone lines, with a high-speed modem. But don't confuse this service with the standard dial-up services discussed earlier for individual access. The earlier examples referred to a simple terminal emulation connection to an actual Internet-connected computer. With a SLIP connection, your computer actually becomes a network host with its own temporary IP address. All file transfers arrive at your local machine, and Telnet sessions originate from your site as well. One of the benefits of this arrangement is that you can continue to operate with your personal computer's standard interface, without the added complications that terminal emulation causes in less direct connections.

The SLIP connection is designed for a single personal computer or workstation or a very small network. A larger network would require the next level of connectivity, a direct Internet connection.

Direct connection. A direct Internet connection is the best method for an organization with a medium or large-sized LAN that needs full Internet access. It will involve operating through an independent service provider. A main connection is made (with a device called a *router*) to the LAN, giving every machine on the network access. At this level, the decision to be made about the degree of service relates to the bandwidth of the physical connection. This decision is similar to an individual choosing a certain speed of modem. A

network administrator will want to arrange a connection with enough bandwidth to support the traffic of his or her site. At the low end are leased phone lines, for networks with minimal traffic, and at the high end are lines classed as T3, which transmit data at up to 45 Mb/sec. These are used for network backbones and connections to very large corporations and universities.

The process of obtaining a direct Internet connection also requires that the site register for a unique domain name and IP number. These unique identifiers control and direct the overwhelming amount of network traffic. Similar to a postal address and zip code, this information ensures that every program, document, or electronic mail message is sent to the proper destination.

Finally, the network administrator must also develop a clear network usage policy before finalizing the Internet connection. Use and misuse of the Internet raises many questions about social and legal responsibility, and increased conflicts related to privacy issues encountered on the existing LAN. These concerns should be addressed as part of the serious planning required of organizations wishing to take the plunge and join the worldwide Internet community.

Glossary of Networking Terms

Acceptable use policy.

A network policy that states that the transmission of commercial traffic across its links is prohibited; information in support of academic and research activities is acceptable. The rapid growth and evolution of the Internet is beginning to test this policy.

ANSI.

American National Standards Institute. A standards body and the U.S. agent for the International StandardsOrganization.

Application Layer.

Layer 7 of the OSI model. This layer covers the the applications (software) with which the user interacts. Information about data formats are not provided in this layer.

ARPANET.

DARPA's packet-switched digital communications network, which linked a wide variety of Department of Defense sponsored computers at research centers around the world. Founded in 1969 and decommissioned in 1989.

ASCII.

American Standard Code for Information Interchange. The ASCIIcharacter set consists of 128 characters, 96 for the uppercase and lowercase letters and 32 nondisplayed control characters. ASCII text is often referred to as "plain vanilla" text because it has no special formatting applied through a word processor such as font attributes, boldface, etc. Most computers use this and an extended character set to standardize the interchange of information.

Asynchronous transmission.

A data communication method in which start and stop bits are used to mark the beginning and end of a data segment and a clock signal is used to synchronize the sending of data bits.

Backbone.

That part of the network that handles the major traffic flow and generally at the greatest speed. Within the U.S., the NSFNET is often referred to as the Internet backbone.

Bandwidth.

The difference between the highest and lowest frequencies of a spectrum. This information is used to express the capacity of a data link and is measured in bits per second. It is more practically the physical capacity of a network to exchange information. The NSFNET backbone is connected over T3 lines that are capable of exchanging data at 45 Mbps. When someone transfers a file or sends e-mail, he or she monopolizes the transmission medium for a finite period of time and, in essence, consumes a portion of the network's bandwidth. The larger the file, the more time necessary to transfer it and the more bandwidth consumed.

Baud

Number of bits per second (bps). Baud is a unit of data transfer rate.

Baseband.

Transmission of signals without modulation at low frequencies across coaxial cables for short distances in local area networks.

Binary transfer.

A file transfer of binary data. Used in the process of downloading or uploading files from one computer to another in order to retain special characteristics of the data. This transfer mode preserves the nature of formatted text (text that has been word processed), executable code (computer programs), images, etc.

BITNET.

Because It's Time Network. A store and forward network of computers worldwide supporting mail, file, and interactive transmission among academic and research organizations.

Bridge.

A device that links similar networks together. It accepts packets from each network addressed to devices on the other and retransmits them. It functions at the data link layer of the OSI model.

Broadband.

The use of coaxial cable for providing data transfer by means of analog (radio-frequency) signals. Digital signals are sent through a modem and transmitted using frequency division multiplexing.

CCITT.

International Telegraph and Telephone Consultative Committee. An organization that develops standards for international telecommunications. CCITT is responsible for X.25, a standard for packet-switched networking.

Circuit switching.

A method of communicating in which a dedicated communications path is established between two devices through one or more intermediate switching nodes. Unlike packet switching, circuit switching sends digital data continuous streams of bits. In terms of bandwidth, this method can be costly in that it requires the establishment of a dedicated circuit until the transmission is complete. It contrasts with message switching and packet switching.

Client-server.

A design in which a local computer provides an interface to a remote computer. The local machine is called the client and the remote machine a server. The client sends queries to the server which in turn sends back information. The client-server model is common in computer-to-computer communication and even in software that communicates with other applications. It is a communication model and part of layer 7 in the OSI model.

CWIS.

Campus-wide Information Systems. A menu-based computer information service that gives information about an academic institution and which is accessible over a computer network. Information typically includes campus and community topics, class information, directories, calendars, and other topics of interest to students and the academic community.

Cyberspace.

A term first used by William Gibson to describe a futuristic computer network into which people could connect their brains. It is used popularly today to refer to the network and its intangible nature.

DARPA.

Defense Advanced Research Projects Agency. The funding agency for ARPANET, the forerunner of today's Internet.

Data compression.

A technique used to reduce the size of files while maintaining the integrity of the data. Compression is done to reduce the consumption of space on a storage device as well as to reduce the consumption of bandwidth during file transfer.

Data link Layer.

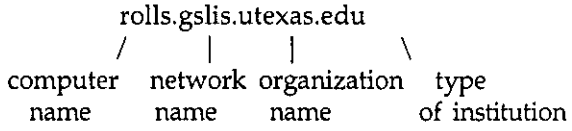
Layer 2 of the OSI model. It is responsible for the transmission of one unit of data between two nodes in a fashion that protects the integrity of the data.

Domain.

In a hierarchical structure, a domain refers to a node which in a sense is superior to its lower or connected nodes. A domain is the plain English version of an IP address. In addressing and routing data, domain names provide convenient schemes to ensure the accurate delivery of information to the intended recipient.

Domain name system.

The domain name system is a worldwide distributed database of computer names and their IP addresses (that is, their plain language name and associated numeric equivalent such as rolls.gslis.utexas.edu and 128.83.248.12). This system is used so that routers on the network can make the correct network connections. The names are hierarchical from right to left:



Ethernet.

A local area network data transmission standard capable of transmissions up to 10 Mbps and connecting up to 1,024 nodes.

FAQ.

Frequently Asked Questions. Many USENET newsgroups, LISTSERVs, BBSs, etc., prepare a FAQ to address commonly asked questions and their answers.

File.

Any collection of electronic data treated and named as a single unit. It can include plain text data, formatted text, programs, and images.

Flame.
A slang term to describe a derogatory or insulting remark passed electronically between participants of a discussion (such as USENET or a LISTSERV group).

Freenet.
A grass-roots, community-oriented effort to provide public access to electronic information. Services commonly include Internet gateways, e-mail, information services, and file retrieval. Freenets are usually funded and operated by volunteers. Access is typically over a dial-up modem line from a home computer and some public library computer terminals. The most well-known Freenet is the Cleveland Freenet operated by the Case Western Reserve University Community Research Laboratory (try telnet freenet-in-a.cwru.edu or use a modem and dial 216-368-3888).

FTP.
File Transfer Protocol; the TCP/IP network protocol that allows file transfer across the network. Also, the software used to transfer files.

Full-duplex transmission.
Data transmission in both directions at the same time.

Gateway.
A computer that connects two networks and performs any needed protocol conversion or address translation.

Gopherspace.
A popular term used to describe the information world accessible and searchable using the Gopher software. Derived from the term cyberspace.

Half-duplex transmission.
Data transmission in either direction, one direction at a time.

Host.
In a distributed computing environment, the host computer is one that controls or manages access and schedules processing time. If you use your home computer to connect to a remote computer, the remote machine is generally the host and yours is the guest (or client).

Internet Architecture Board (IAB).
Part of a nonprofit organization that determines the needs of the Internet and develops solutions toward meeting those needs. The IAB oversees the work and direction of the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF).

Internet Engineering Task Force (IETF).
Part of a nonprofit group that promotes the development of the Internet. This group explores operational and technical issues.

Internet Research Task Force (IRTF).
Part of a nonprofit group that oversees the development of the Internet. This group explores research and developmental issues.

Internet-relay-chat.
A real-time, interactive discussion tool. It is a means for many people to assemble in one location (in a virtual sense) and talk simultaneously over a computer network.

Internetworking.
The process of two or more networks connecting.

IP address.
The numerical equivalent of the plain language address, the IP is part of the domain name system. It is used by routers to make proper connections.

ISDN.
Integrated Services Digital Network. The international standard for transmitting voice, video, and data over a digital communications link. A "narrowband" ISDN communicates at a rate of 64 Kbps while a "broadband" ISDN can approach 150 Mbps using fiber-optic cable.

LAN.
Local Area Network. A network that connects servers, workstations, and other devices in a limited geographic area such as an office or building.

Layer.
In reference to the OSI model, a group of functions that make up one level of a hierarchy of network functions.

LISTSERV.
A program that runs on a BITNET computer and forwards e-mail to subscribing members. LISTSERV groups are organized along subject-specific areas. Members submit postings to the LISTSERV and the program automatically forwards them to all subscribers. Unlike USENET, once a member subscribes to a LISTSERV, they receive all postings through e-mail whether they are wanted or not (unless they unsubscribe or take steps to suspend delivery).

MAN.
Metropolitan Area Network. A communications network which spans a geographic area such as a city or suburb.

Modem.
Modulator/Demodulator. Transforms a digital stream of information to an analog signal (modulator) and vice versa (demodulator) in order for data to be transmitted over phone lines.

Moderator.
A person within some LISTSERV or USENET groups who reviews all submissions to decide whether they will be redistributed to the group as a whole.

National Science Foundation.
A U.S. government institution that provides funding assistance for scientific and technological research. It was the funding agency for the development of the NSFNET. However, it is not a policy organization, and it does not control the Internet.

Net etiquette.
A way of describing courtesy practices in networked communications.

Network Layer.
Layer 3 of the OSI model, which is responsible for routing data through a communication network. It establishes the route between sending and receiving nodes.

Noise.
Unwanted signals that combine with data transmissions, resulting in unreliable data being received.

NREN.
The National Research and Education Network. This network grew out of Vice President (then Senator) Al Gore's High Performance Computing Act of 1991, which promoted the growth of a coordinated effort to develop a high-speed successor to the research and education portion of the Internet in the U.S. Research is aimed at developing a multi-gigabit per second transmission rate. The NSFNET is often referred to as the "interim NREN."

NSFNET.
A internetwork consisting of a 45 Mbps (T3) transcontinental backbone that interconnects a number of mid-level networks. Each of these mid-level networks interconnects a dozen or more campus-area networks.

OPAC.
Online Public Access Catalogs. Library catalogs that are mounted onto computers and made accessible for searching over network connections.

Operating systems.
A program that runs a computer. It schedules tasks, allocates resources (such as memory), and serves as the interface between computer hardware and high-level applications. Common operating systems include UNIX, DOS, OS/2, and Macintosh System 7.

Optical fiber.
A thin filament of glass or other transparent material through which a light beam that has been encoded with data is transmitted at high speed (billions of bits per second).

OSI.
Open Systems Interconnect. A structure for network operations standardized within the International Standards Organization (ISO). It calls for a seven-layer model to define network communication protocols, enabling an OSI-compatible computer to communicate with any other OSI-compliant computer.

Packet switching.
In packet-switched networks, messages are split into fixed length pieces that are transmitted to their destination by the most expedient route; consequently, not all packets may travel by the same route. A packet includes data from the original message as well as addressing information. Packet switching is usually more efficient and rapid than circuit switching.

Physical layer.
Layer 1 of the OSI model. It defines the characteristics of hardware needed for passing information bits to and receiving them from a connecting medium.

Presentation layer.
Layer 6 of the OSI model. Provides a common means for the representation of data to the applications.

Protocol.
Standardized rules that govern the exchange of information between or among computers.

Routing.
The process for determining the route by which a packet of data is to be sent. A router examines the destination address of a packet and determines the most expedient path of delivery.

Session layer.
Layer 5 of the OSI model. Manages a communication session between two communicating processes, applications, or computers.

Smiley.
Characters appended to a message to indicate the mood of the sender or to display "emotion." Examples include :) (happy) :((sad) ;) (wink).

Store-and-forward.
The temporary storage of data before forwarding it to its ultimate destination. This technique allows for the transmission of information between networks that may not maintain a continuous connection. Fidonet, BITNET, and USENET are examples of store-and- forward networks.

Switch.
A mechanical, electrical, or software-driven device that controls the routing or path of data.

TCP/IP.
Transmission Control Protocol/Internet Protocol. The set of standards developed by a DARPA funded project in 1973-74 to allow the internetworking of dissimilar networks. There are implementations of TCP/IP for almost all types of computers.

Terminal emulation.
A means used to make a personal computer terminal emulate the video display terminal of a mainframe computer. The most prevalent emulation is that of the VT100 terminal made by Digital.

Thread.
In USENET newsgroups and LISTSERV discussions, conversation often proceeds along topical lines. A thread is a series of messages on a related topic.

Topology.
The physical structure of the network. If a person were to map the physical layout of a computer network, including computers, servers, switches, lines, and other devices, the resulting map would be referred to as a network topology.

Transmission channel.
The physical path between transmitters and receivers in a communications or computer network (i.e., the path between two nodes).

Transport layer.
Layer 4 of the OSI model. Responsible for the reliable and transparent travel of data between connections. Transport services include layers 1 to 4.

USENET.
A set of more than 2,000 discussion groups (newsgroups) which exchange messages centered on specific subjects. USENET messages are distributed worldwide and offer network users the ability to discuss subjects of interest to them with others around the world.

UUCP.
UNIX-to-UNIX CoPy is a UNIX utility that copies files from one UNIX system to another. It is often used to send USENET news to systems without a direct Internet connection.

VT100
A common terminal emulation mode used to display information on personal computers connected to remote computers using the ASCII character set.

Wide area networks.
Computer networks that cover wide geographic areas such as counties and states.

X.25.
A network access standard for the interface between a terminal and a switched network. Covers OSI layers 1-3.

X.500.
A CCITT standard for mail and messaging. It includes the capability of maintaining directories of users and is meant to simplify the process of locating e-mail addresses.

Z39.50.
An American national standard for search and retrieval protocol for databases. The maintaining agency is the Library of Congress.

Bibliography

This list contains both print and electronic sources. Electronic sources are accompanied by the corresponding Anonymous FTP site, or other appropriate access instructions.

Beginning User Guides

Dern, Daniel P. *The Internet Guide for New Users.* New York: McGraw-Hill, 1994.
One of the largest Internet Guides at more than 500 pages. It includes everything from Internet History to instructions on the use of specific tools. Network etiquette and convention are a common theme in each of the chapters.

Gilster, Paul. *The Internet Navigator.* New York: John Wiley and Sons, 1993.
Explains network tools by example, specifically examples involving dial-up terminal connections for individual Internet users.

Internet: Getting Started. April Marine, ed. Englewood Cliffs, NJ: Prentice Hall, 1992.
Contains some discussion of Internet history and development, explanations of network tools, but also has lengthy lists of service providers, policy makers, and other network organizations. There is a substantial section on Non-U.S. networks.

Kehoe, Brendan P. *Zen and the Art of the Internet: A Beginner's Guide.* Englewood Cliffs, NJ: Prentice-Hall, 1992.
Explains, in a straightforward and clear manner, basic Internet tools and gives examples of popular Internet resources.

LaQuey, Tracy, and Jeanne C. Ryer. *The Internet Companion: A Beginner's Guide to Global Networking.* Reading, MA: Addison-Wesley, 1992.
This small handbook includes interesting anecdotes involving actual Internet resources. Its audience is not limited to the academic community because it contains instructions on how any individual may access the Internet.

Tennant, Roy, John Ober and Anne G. Lipow. *Crossing the Internet Threshold: An Instructional Handbook.*
Uses a succinct and direct approach to describe Internet resources. It is especially helpful for trainers in computer or network centers.

Specialized User Guides

Farley, Laine, ed. *Library Resources on the Internet: Strategies for Selection and Use.* Chicago: Reference and Adult Services Section, American Library Association, 1991.
Although this directory includes lists of Internet-accessible online public access catalogs (OPACs), it is more than a directory. It provides instructions for searching many OPACs, as well as a general perspective on the use of remote OPACs for research.

Henry, Marcia Klinger. *Search Sheets for OPACs on the Internet: A Selective Guide to U.S. OPACs Utilizing VT100 Emulation.* Westport, CT: Meckler, 1991.
This work is not comprehensive, but several of the most popular OPAC software vendors are included.

Krol, Ed. *The Whole Internet User's Guide and Catalog.* Sebastopol, CA: O'Reilly & Associates, 1992.
Like the other guides, this book covers the basic Internet tools, but its tone and its examples of resources are geared toward postgraduate academic users.

Lane, Elizabeth S., and Craig A. Summerhill. *An Internet Primer for Information Professionals: A Basic Guide to Networking Technology.* Westport, CT: Meckler Corp., 1992.

Internet Standards

Lynch, Daniel C. and Marshall T. Rose. *The Internet System Handbook.* Reading, MA: Addison-Wesley, 1993.
This book provides very useful technical information. It is not aimed at beginning users, rather at network managers and engineers. It is full of helpful programming examples for customizing and configuring networks.

Malamud, Carl. *Exploring the Internet: A Technical Travelogue.* Englewood Cliffs, NJ: Prentice Hall, 1993.
This is a description of the current international implementation of Internet standards. Although the book has a highly technical subject, the narrative format is readable for those familiar with the major movements in network standardization.

Malamud, Carl. *Stacks: Interoperability in Today's Computer Networks.* Englewood Cliffs, NJ: Prentice Hall, 1992.
In this work, Malamud discusses the many emerging standards defining network interconnectivity.

Quarterman, John S., and Susanne Wilhelm. *UNIX, POSIX, and Open Systems: The Open Standards Puzzle.* Reading, MA: Addison-Wesley, 1993.
As the title suggests, this book explains the relationships between the countless and often conflicting standards that define modern computing environments.

Resource Catalogs and Directories

Barron, Billy. *UNT's Accessing On-Line Bibliographic Databases.* Denton, TX: University of North Texas, 1991.
This worldwide OPAC directory is arranged alphabetically by name of institution. It is also useful as a quick reference for common OPAC software searching commands. It is available in several formats via Anonymous FTP at ftp.unt.edu in the library directory.

Frey, Donnalyn, and Rick Adams. *A Directory of Electronic Mail Addressing and Networks.* Sebastopol, CA: O'Reilly & Associates, 1991.

Bibliography

Interest Groups. Menlo Park, CA: SRI International, Network Information Systems Center, 1992.
BITNET LISTSERVs are compiled in this work. One must keep in mind that these groups change frequently, and, as a result, these types of sources become outdated quickly. *Interest Groups* is arranged alphabetically by list name. Available via FTP at ftp.nisc.sri.com in netinfo/interest-groups.

Internet: Mailing Lists. Edward T.L. Hardie and Vivian Neou, eds. Englewood Cliffs, NJ: Prentice Hall, 1993.
This directory lists several LISTSERVs and USENET discussion groups found on the Internet. Each entry is accompanied by a helpful abstract describing the topic and scope of the discussion group.

LaQuey, Tracy. *User's Directory of Computer Networks*. Bedford, MA: Digital Press, 1990.
A directory of the various networks within the Internet, with an explanation of how they are related.

NorthWestNet User Services Internet Resource Guide. Bellevue, WA: NorthWestNet, 1992.

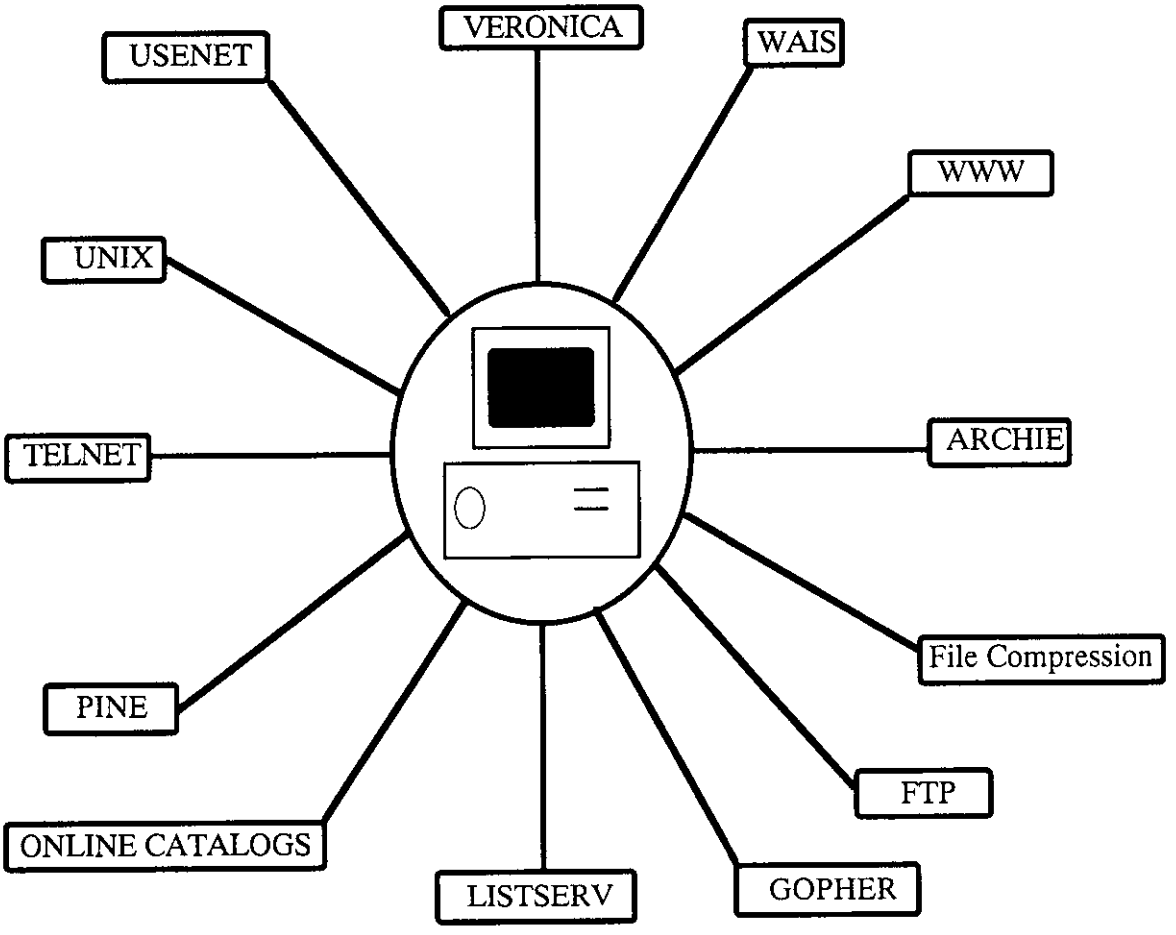
NSF Network Service Center (NNSC). *Internet Resource Guide*. Cambridge, MA: BBN, 1991.

Quarterman, John S. *The Matrix: Computer Networks and Conferencing Systems Worldwide*. Bedford, MA: Digital Press, 1990.
This work is similar in purpose to LaQuey's *UDCN*. It is one of the few books that describe the many lesser-known computer networks that help create the global connectivity of the Internet.

St. George, Art, and Ron Larsen. *Internet-Accessible Library Catalogs and Databases*. Albuquerque, NM: University of New Mexico, 1991.
This OPAC directory is organized by geographic region. It is available via AnonymousFTP in library/internet.library at ariel.unm.edu.

PART II

Internet Resources and Tools



Introduction

This part of the book may be used as a reference manual. Information is arranged so that important facts can be quickly found and understood. The major tools and resources are presented as individual sections, in alphabetical order. Each section begins with a discussion of the tool or resource covered, followed by step-by-step instructions for its use. Many sections also have a "Hints" or a "Helpful Commands" part that provides more information about an individual tool or resource.

Archie

What Is It?

Archie is a database program that allows a user to perform broad searches of the numerous Anonymous FTP sites on the Internet. The user supplies a search string and Archie queries a list of Anonymous FTP sites around the world for matching files and directories. Archie is an indispensable tool for those who know a program or text file exists but do not know where to obtain a copy.

How Do I Access It?

The Archie program is implemented according to the client-server model. The most common method for accessing the Archie program is to Telnet to an Archie server, login as 'archie', and then conduct the file search.

Here is an example of a connection to an Archie server:

```
emx.cc.utexas.edu% telnet archie.rutgers.edu
Trying...
Connected to dorm.Rutgers.EDU.
Escape character is '^]'.

SunOS UNIX (dorm.rutgers.edu) (ttypb)

login: archie

Last login: Thu Mar 25 00:31:28 from UNIXd2.cis.pitt.
SunOS Release 4.1.1 (TDSERVER-SUN4C) #3: Mon Nov 23 17:48:54 EST 1992

        Welcome to the Rutgers University Archie Server!

# Bunyip Information Systems, 1993
# Terminal type set to `vt100 24 80'.
# `erase' character is `^?'.

archie>
```

How Can I Use It?

Archie is commonly used to locate a text file, program, or graphic file which a user has learned of from some other network source, a printed citation, or another user. A user may search Archie with the name or partial name of the file. The command used to search for a particular file is **prog**.

Here is an example of a search for a game called "Robotron":

```
archie> prog robotron
working... \ /archie>

Host ftp.ms.uky.edu    (128.163.128.6)
Last updated 04:56 25 Mar 1993

   Location: /pub4/msdos/Games
     FILE    -rw-rw-r-- 26154 bytes 23:00 22 Oct 1990  robotron.arc

Host ftp.ms.uky.edu    (128.163.128.6)
Last updated 04:56 25 Mar 1993

   Location: /pub4/msdos/Games
     FILE    -rw-rw-r-- 18645 bytes 00:00  4 Jan 1980  robotron.zip
```

Archie returns with several locations of the requested file. The information also includes the exact directory location at the site, as well as file size and creation date.

Helpful Commands

The following commands are used to configure the Archie program environment:

• set search	<variable>	used with one of the following variables to define the type of search conducted by Archie.
	exact	search for files or directories that match the search string exactly
	regex	search for files or directories allowing wildcard characters in the search string
	sub	search for files or directories that include the search string, regardless of upper-case or lower-case letters
	subcase	like sub, except for the fact that the search string is case sensitive
• set maxhits		followed by an integer from 1-1,000, this limits the maximum number of file locations that Archie returns.

File Compression

What Is It?

A computer file is any collection of electronic data treated and named as a single unit. Files can be either text files or binary files. Plain text files contain only ASCII text. They have had no formatting applied using a word processor. All other files are treated as binary files. These include formatted text, programs and images. Files can be related to one another in that they are individual pieces of a larger entity. For example, most software applications consist of many files. Each piece has a distinct function or role in how the overall application works. Generally, all of these pieces are managed by a single file which is called the program file. It is a collection of instructions for the computer written in machine code (often referred to as executable code). When retrieving files using FTP or downloading files using communications software, files must be identified as either ASCII or binary. Compressed files are always in binary format. When transferring compressed files (as well as other binary files) it is important to tell the FTP session or communications software that the file contains binary information. If you download a compressed file using a text transfer mode, the resulting file will contain information of no use to you or the computer.

As you begin to retrieve files using Internet resources and tools, you will have to understand how to identify different types of files in order to make use of them. Some files are only usable under certain types of operating systems (for instance, some files that can be accommodated by your Macintosh are not recognized by your DOS machine or vice versa). By understanding file-naming conventions, you will be able to identify the type of file at hand and make decisions about whether or not to download it as well as how to handle it once you've retrieved it. This becomes especially important when dealing with files that have been compressed.

Files are typically given a name and an extension in order to identify their type. The most common file-name extensions are listed here.

Filename	File Type	Remarks
paper.txt	ASCII text	Plain ASCII text document.
paper.doc	ASCII text	Plain ASCII text document.
paper.exe	PC executable	Program that can run on a PC.
image.ps	PostScript	PostScript file
sound.wav	WAVE	A WAVE sound file
sound.voc	Soundblaster	A Soundblaster file
image.gif	GIF	Image file that requires a viewer software to see its content.
image.jpg	JPEG	Image file that requires JPEG viewer software.
paper.hqx	BinHex	An encoded ASCII file for the Mac.

Compression is a technique that reduces the size of a file while maintaining the integrity of its data. As you explore the realm of the Internet, you will certainly come across files that have been compressed. Compression provides two primary benefits. First, by reducing the size of a file to be transmitted over a network or downloaded to a local computer, you reduce the time necessary to make that transfer. This then reduces your consumption of bandwidth (capacity of communication links to transfer data). Second, file compression is useful in environments where disk storage space is limited. If you reduce the size of your files, you make room for more information on a storage device.

There are a variety of compression utilities on the market aimed at making compression of files transparent. These utilities are meant to economize the use of a hard disk and actually compress and uncompress files as they are needed. The compression schemes discussed here are used to compress individual files (or groups of files) in order to economize storage space or improve transmission rates over a network.

The most prevalent compression schemes used in this role are shown in the following list.

Compression Technique	Computer Type	File extension
Compact Pro	Mac	.cpt
InfoZip	PC, Mac, UNIX	.zip
PKZIP	PC	.zip
StuffIt	Mac	.sit
ARC	PC	.arc
ARJ	PC	.arj
compress & uncompress	UNIX	.Z
LHArc	PC	.lzh
self-extracting	Mac	.sea

How Do I Use It?

Before you use a compressed file, you must uncompress it. Here are a variety of examples of compressing and uncompressing files, depending on the environment or utility used.

.arc (a DOS compression utility. It includes features for archiving groups of files.)

uncompress

c:\> pkxarc filename.arc

compress

c:\> pkarc filename archive_name.arc

.arj (a DOS compression utility. It includes features for archiving groups of files.)

uncompress

c:\> arj e archive_name.arj

compress (the files in the current directory)

c:\> arj a archive_name

.cpt (a file compressed using Compact Pro)

uncompress/compress

Using Compact Pro or Aladdin Systems StuffIt tools, drag the icon of the compressed file onto the program icon.

.lzh (a DOS compression utility. It includes features for archiving groups of files.)

uncompress

c:\> lharc x archive_name.lzh

compress

c:\> lharc c archive_name filename

.sea (a self-extracting Macintosh utility)

uncompress

Double-click on the items icon, and the file will self-extract, resulting in a new icon.

.sit (an Aladdin Systems StuffIt file extension for Macintosh)

uncompress

Drag the icon for the compressed file onto the icon for the StuffIt program.

.Z (a UNIX compression utility)

uncompress

sleepy.cc.utexas.edu% uncompress internet.paper.Z

compress

sleepy.cc.utexas.edu% compress internet.paper

.zip (a DOS compression utility such as PKZIP or InfoZip)

uncompress

c:\> pkunzip filename.zip

compress

c:\> pkzip filename

FTP

What Is It?

File Transfer Protocol (FTP) is the most common software tool used to transfer files between computers on the Internet. It is designed to provide a transparent connection between the local and remote computer, allowing a user to access files and directory structures of other computers on the Internet.

The most common implementation of this network tool is Anonymous FTP which provides global access to specified files on a computer system. Not all sites allow for Anonymous FTP, but many have created directories of files which are available using this tool. To establish an Anonymous FTP connection to a remote computer, a user types ftp followed by the Internet address of the remote computer (i.e., ftp bongo.cc.utexas.edu). When prompted for a username, "anonymous" is entered, and the user's own Internet address is supplied as a password. Once the connection is established, the user may explore the directories and files on the remote computer using the same commands as on a local session. For instance, a user may choose specific directories; print directory listings with information about the names, sizes, and dates of creation of files; and copy files to his or her local computer account.

Here is an example of an initial FTP connection (remember, UNIX is case-sensitive and the command must be given in lowercase letters, that is, FTP):

```
emx.cc.utexas.edu% ftp bongo.cc.utexas.edu
Connected to bongo.cc.utexas.edu.
220 bongo.cc.utexas.edu FTP server (Version 6.18 Wed Mar 24 15:46:24 CST 1993) .
Name (bongo.cc.utexas.edu:richie): anonymous
331 Guest login ok, send e-mail address as password.
Password:  richie@bongo.cc.utexas.edu
```

FTP

How Can I Use It?

By using Anonymous FTP, an Internet user may obtain countless text files, executable programs, and even graphic images. Many of these files are in standard formats (especially text and graphics files), but many others are for specific microcomputer types (Macintosh, MS-DOS, etc.) or minicomputers such as UNIX, or VMS-based systems.

The most common method of accessing these file libraries is to connect to an FTP site, list one or more directories to locate the desired file, then transfer the file from the remote computer. The command used to obtain files from an FTP site is **get.** Here is an example of an FTP session conducted to acquire a "README" file:

```
ftp> dir
200 PORT command successful.
150 Opening ASCII mode data connection for /bin/ls.
total 1468
-rwxr-xr-x      1 root      staff      729 Mar 24 13:56    .cache
drwxrwxr-x      4 root      staff      512 May  1  1992    EMX-ftp
-rw-r--r--      1 root      staff     1471 Mar 10  1992    README
drwxr-xr-x     14 ifab734   ftp       1536 Mar 23 19:25    anime
dr-xr-xr-x      2 root      staff      512 Apr 28  1992    bin
drwxr-xr-x      2 root      staff      512 May  1  1992    bin.vax
226 Transfer complete.
1341 bytes received in 0.64 seconds (2.1 Kbytes/s)
ftp> get README
200 PORT command successful.
150 Opening ASCII mode data connection for README (1471 bytes).
226 Transfer complete.
local: README remote: README
1509 bytes received in 0.041 seconds (36 Kbytes/s)
ftp>
```

The abundance of Anonymous FTP sites can sometimes be a drawback. With the large number of sites, it is sometimes difficult to know where to begin searching for information. Two other network tools, the Archie and Gopher systems, help focus FTP searches and generally make their use more efficient.

Hints

When exploring an Anonymous FTP site, look for files with names such as "README" and "INDEX". These files will usually contain descriptions of the files available or other useful information such as general instructions or site policies.

Some helpful commands include:

ascii	Sets the file transfer type to ascii. This is the default type and is used for transferring text files.
binary	Sets the file transfer type to binary. This command must be issued before transferring executable, compressed, or graphics files.
dir	Lists a directory of files on the remote machine. A similar command is **ls**.
get	Used to obtain files from a remote FTP site, one at a time.
mget	Used to obtain multiple files, for example, mget doc*.
!	All commands typed after this symbol are sent to the local machine. This is useful for navigating the local directory structure and viewing local index and readme files.

Gopher

What Is It?

What is a Gopher? In the animal kingdom it's a rodent, but in the world of the Internet it's an information retrieval tool. Gopher is a software program that runs on your computer (or host computer) to provide a relatively uniform, simple menu interface to information resources. It provides directories of available files and then retrieves them to your machine (or host) using "hidden" FTP commands. The program "hides" the fact that it is connecting to remote machines (Telnet or FTP), moving around directories, and searching other information tools (WAIS, Archie, etc.) The original software was developed in 1991 by the University of Minnesota Microcomputer, Workstation Networks Center.

The Gopher software uses the client-server approach to information sharing. When running a Gopher session locally, your machine becomes the client. When your Gopher client makes contact with a Gopher server program on another machine, the remote Gopher is referred to as the server. The local machine (client) passes information queries to the server, and the server responds either by providing listings of available information or by sending files (text or executable) to the client (you). The purpose of the client-server model is to hide the details of the network connection (such as IP addresses, ports, FTP commands, etc.) from you.

How Do I Access It?

Although you need not have a **Gopher client** on your institution's machine to use Gopher tools, it is preferred. If not, you may Telnet to a remote Gopher host that offers public access and run a Gopher session as if the program were running on your own machine. Some popular locations of publicly accessible Gopher sites are listed here. It is best to login to the nearest site in order to reduce net congestion.

Hostname	IP Address	Login Name
consultant.micro.umn.edu	134.84.132.4	gopher
ux1.cso.uiuc.edu	128.174.5.59	gopher
info.anu.edu.au	150.203.84.20	info
gopher.chalmers.se	129.16.221.40	gopher
tolten.puc.cl	146.155.1.16	gopher
ecnet.ec	157.100.45.2	gopher

How Can I Use It?

Explore, explore, explore. We have found this to be the best method for tapping into the wealth of information resources connected to the Internet. Tools exist for searching Gopherspace (such as VERONICA), but an inquisitive mind is your best asset.

Below is the Gopher server menu at the University of Texas. Keep in mind that the Gopher menu at every site will vary depending upon what information, tools, and links to other resources (like other Gophers) the site manager has decided to include. Gopher sites are constantly evolving as new tools and resources are added or removed. Also, depending upon your interface environment (UNIX, XWindows, Macintosh, DOS, VMS, VM/CMS, NeXT, OS/2, etc.), the screens may vary.

```
        Internet Gopher Information Client v1.12S

       Root gopher server: gopherhost.cc.utexas.edu

    1. About this Gopher/
    2. UT-Austin/
    3. Library/
    4. Texas/
-->5. World/
    6. Jughead: Search gopher menus at UT Austin/
    7. VERONICA: Search gopher menus worldwide/
    8. Archie: Search FTP sites worldwide/

Press ? for Help, q to Quit, u to go up a menu          Page: 1/1
```

In the screen shown in the last example a user moves the arrow key to align with item number 5, World/. This item is selected by pushing the return or enter key. Because the item is followed by a "/", a user knows that it is a link to another menu.

The new menu screen then appears:

```
                Internet Gopher Information Client v1.12S

                             World

--> 1. First Gopher In World (U of Minnesota)/
    2. All Gopher Servers in World/
    3. Art and Images/
    4. Documents/
    5. E-Mail Searches/
    6. E-Mail and Phone Searches - X.500/
    7. Grants and Funding/
    8. Freenet Systems/
    9. Jobs/
    10. Library Catalogs/
    11. Online Phone Books/
    12. Music Archives/
    13. Software Archives/
    14. Subject Archives/
    15. USENET News/
    16. WAIS Searches/
    17. Weather Forecasts/
    18. Weather Underground <TEL>

Press ? for Help, q to Quit, u to go up a menu                Page: 1/1
```

Hints

Items on a Gopher menu are generally of three types. They either provide access to a submenu, to a remote database or site, or to actual files which may be viewed on screen, retrieved (downloaded), or mailed (e-mail). In the example above, item #6, "Libraries," is a submenu item (it has a "/" following its name). Item #10, "Search lots of places at the U of M <?>," is a link to a database that may be searched. Database links are followed by a <?>, <cso>, or <idx>. Items followed by <tel> are links to remote sites. Those items that have nothing following their name are files that may be viewed on screen (text files only) or retrieved by selecting them and pressing enter or return.

Files of this sort may be text files, programs, graphic images, sound recordings, and other types of files. Listed below are the most common suffixes found in Gopherspace.

Menu Item Extension	Explanation
No extension	Text file or item that can be retrieved
/	Access to a submenu
<?>	Searchable database/index
<idx>	Searchable index
<cso>	Searchable phone-book server
<tel>	Connection to another site via "telnet"

These menus are your navigational tools for traveling through what has come to be known as Gopherspace. Items are chosen by moving the arrow to line up with the selected item and then pressing the enter or return key. Depending on the type of item you have selected, this will either display a text file to your screen, provide yet another menu, or provide a link to some remote database.

Other Helpful Commands

Moving around

(Use the arrow keys or vi/emacs equivalent to move around.)

Up	:	Move to previous line.
Down	:	Move to next line.
Right, Return	:	"Enter"/Display current item.
Left, u	:	"Exit" item/Go up a level.
>, +, Pgdwn, space	:	View next page.
<, -, Pgup, b	:	View previous page.
0-9	:	Go to a specific line.
m	:	Go back to the main menu.

Additional Commands

s	:	Save current item to a file.
D	:	Download a file.
q	:	Quit with prompt.
Q	:	Quit unconditionally.
=	:	Describe current item and its location.
O	:	Change options.
/	:	Search for an item in the menu.
n	:	Find next search item.

Bookmark Commands

a	:	Add current item to bookmark list.
A	:	Add current directory to bookmark list.
v	:	View the bookmark list.
d	:	Delete a bookmark entry.

LISTSERV

What Is It?

LISTSERV is an electronic mail distribution system. The purpose of a LISTSERV program is to receive mail postings related to a specific topic and redistribute those messages to interested users who have *subscribed* to the service. The result of a successful LISTSERV program is an Internet-wide system, through which individual users may post messages or request feedback on particular topics which are quickly made available to other subscribers who share an interest in that particular LISTSERV topic.

How Do I Access It?

A LISTSERV list may be accessed by any e-mail system with access to the Internet or Bitnet. The Internet mail program an individual uses for daily correspondence with friends and colleagues is also capable of supporting LISTSERV subscriptions. To subscribe to a list, simply send an e-mail message to the address of the LISTSERV program; the text of your message is interpreted as a string of commands.

The address of a LISTSERV program is always in the format:

listserv@domain

The text of your message should be in the format:

SUB LISTNAME firstname lastname

Here is an example of a composed e-mail message initiating a subscription to the list PACS-L, which is located at the University of Houston, on the computer named uhupvm1. This list is devoted to the discussion of general library issues.

To : listserv@uhupvm1.uh.bitnet
Cc :
Attchmnt:
Subject :
----- Message Text -----

SUB PACS-L *your name*

To unsubscribe to the list, use the same procedure, using the command in the body of the message (do not mention your name):

UNSUB PACS-L

How Can I Use It?

Once you are subscribed to the list, LISTSERV mail is delivered to you automatically. Mail posted to the list is quickly distributed to the subscribed users via standard network e-mail channels. LISTSERV messages appear in your electronic mailbox just as other e-mail messages do.

As a LISTSERV subscriber, you can simply monitor the list traffic each day, browsing messages as they are delivered. Many people use LISTSERVs in this way, gleaning valuable information as it flows through their mail system.

However, you may wish to contribute to the discussion by asking or answering questions. To do this, you simply post a message to the list yourself. To do this, you send your message to the list address, which is in this form:

Listname@domain

It is very important not to confuse the address of the LISTSERV program and the address of the list itself. In our previous example, the address of the LISTSERV program is listserv@uhupvm1.uh.edu but the actual list address is pacs-l@uhupvm1.uh.edu

Here is an example of a composed message to be posted to the list PACS-L .

To : pacs-l@uhupvm1.uh.edu
Cc :
Attchmnt:
Subject : National Union Catalog
----- Message Text -----

Yes, I too have seen a remarkable decrease in the use of NUC
since we installed the OCLC terminal in the reading room.

Hints

- A comprehensive, up-to-date "list of lists" is available via Anonymous FTP at lilac.berkeley.edu. Login as anonymous. The file is in the directory "netinfo" (cd netinfo).

- A similar "list of lists," arranged by list topic, is available at listserv@kentvm or listserv@kentvm.kent.edu. Send the following commands in the body of an e-mail message to the program:

  get acadlist readme f=mail
  get acadlist file1 f=mail
  get acadlist file2 f=mail
  get acadlist file3 f=mail
  get acadlist file4 f=mail
  get acadlist file5 f=mail
  get acadlist file6 f=mail
  get acadlist file7 f=mail

- Keep in mind that LISTSERV mail is delivered automatically, so be prepared for a deluge of mail from the more popular lists.

- If you have access to USENET, you may access many LISTSERV lists as newsgroups that have the prefix bit.listserv in their names. For instance PACS-L is available under the newsgroup name bit.listserv.pacs-l.

Online Library Catalogs

What Are They?

A library catalog contains information such as author names, titles, and subject headings, mostly on books and periodicals that a library owns. Users can search an online catalog for a particular item, and display bibliographic information related to the item if the library owns it. Some libraries also provide access to the full text of items such as online journals and articles. Approximately 300 online library catalogs worldwide are accessible via the Internet.

How Do I Access Them?

There are several ways one can access an online library catalog. One such way is illustrated here.

Access the University of Texas at Austin campus information system (UTINFO). Telnet to the IP address 128.83.145.26 and enter the user name UTINFO (note the uppercase letters).

```
bongo.cc.utexas.edu> telnet 128.83.145.26
Trying 128.83.145.26...
Connected to 128.83.145.26.
Escape character is '^]'.

                UTXVMS

          UT-Austin Computation Center VMS Cluster
            VMS V6.0 on VAX node ORANGE

   Use of UT Austin computer and network facilities requires prior
   authorization. Unauthorized access is prohibited. Usage may be
   subject to security testing and monitoring. Abuse is subject to
   criminal prosecution.

Username: UTINFO
```

The following screen should appear.

```
Welcome to UTINFO, Thursday, 30-SEP-1993

      UTINFO - UT Austin Information Services

Please select an option number from the following list-

   1) UnCover              Index to 14,000 periodicals
   2) UTCAT PLUS           UT Austin library catalog, databases
   3) LIBS                 Other library catalogs, databases
   4) Logout

Please type the number of your choice, then press RETURN:
```

Choose option 3 above for the following screen.

```
      LIBS - Internet Access Software v2.0a
Mark Resmer, Sonoma State University, Mar 1993

On-line services available through the Internet:

1 United States Library Catalogs
2 Library Catalogs in other countries
3 Campus-wide Information Systems
4 Databases and Information Services
5 Wide-area Information Services
6 Information for first time users

Press RETURN alone to exit now or
press Control-C Q <return> to exit at any time

Enter the number of your choice:
```

Choose option 1 for United States library catalogs or option 2 for library catalogs in other countries.

Upon choosing option 1 you will see the following screen:

Libraries in these states are accessible

1. Alabama	2. Arizona	3. California
4. Colorado	5. Connecticut	6. Delaware
7. Florida	8. Georgia	9. Hawaii
10. Idaho	11. Illinois	12. Indiana
13. Iowa	14. Kansas	15. Kentucky
16. Maine	17. Maryland	18. Massachusetts
19. Michigan	20. Minnesota	21. Missouri
22. Mississippi	23. Montana	24. Nebraska
25. Nevada	26. New Hamp	27. New Jersey
28. New Mexico	29. New York	30. North Carolina
31. Ohio	32. Oklahoma	33. Oregon
34. Pennsylva.	35. Puerto Rico	36. Rhode Island
37. South Carol.	38. Tennessee	39. Texas
40. Utah	41. Vermont	42. Virginia
43. Washington	44. Wash. DC	45. Wisconsin

Press RETURN alone to see previous menu

Press Control-C Q <return> to exit at any time

Enter the number of your choice:

At this point if you enter a number corresponding to a state (or a district) of your liking, you will see a listing that typically looks like this:

```
Ohio Libraries:

    1 Air Force Inst of Technology
    2 Case Western Reserve University
    3 Cleveland Public Library
    4 Cleveland State University
    5 College of Wooster
    6 Kent State University
    7 Miami University
    8 NE Ohio U. College of Medicine
    9 Oberlin College
    10 Ohio State University
    11 University of Akron
    12 University of Cincinnati
    13 University of Dayton
    14 University of Toledo
    15 Youngstown State University

    Press RETURN alone to see previous menu
    Press Control-C Q <return> to exit at any time

    Enter the number of your choice:
```

To choose a library, type the corresponding number at the above prompt. If you would like to search the Ohio State library, you would enter 10. If you type that number and press enter, you will see a display that typically looks as the following.

Ohio State University

The catalog includes the holdings of all Libraries at The Ohio State University with the exception of the Law Library, which has only recently been adding new titles. Titles cataloged by OSU Libraries after September 1971 have full records and can be searched by author, title, subject, and series entry. Titles cataloged prior to September 1971 have short records and can be searched only by author and/or title. OSU has 2.2 million titles listed in LCS. LCS also contains the complete State Library of Ohio catalog, recently cataloged items available from the Center for Research Libraries, Chicago, and materials cataloged in the past ten years by the Ohio Historical Society Archives-Library Division.

Note the following instructions carefully

Once you are connected:

LCS is a command-driven system. The commands are usually three characters followed by a slash (/), e.g. aut/wolfe, tom. Basic search instructions are available by using the HELP commands on LCS. Press ESC and 1 for HELP or type HELP.

Press Control-C Q <return> to exit at any time

Do you want to connect now? (Y or N):

By entering Y and pressing return you will be able to connect to the Ohio State online library catalog.

How Can I Use Them?

Each library catalog has its own set of commands. Some typical commands are: (1) Author search, (2) Title search, and (3) Subject search. Some systems will allow you to perform many other types of searches. The format in which you enter an author's name is usually in phone-book style: last name, first name. When entering titles avoid entering words such as the, a, or an, as a first word. These words are called stop words and are usually not used in the index. In case of subject searches, if you are not familiar with Library of Congress subject headings, consult the Library of Congress Subject Heading manuals (found in the reference section of the library).

Hints

Once you reach the actual catalog source, read that initial screen very carefully. Pay special attention to the information about: (1) Terminal type, (2) Passwords, (3) Database name, and (4) Command to exit the catalog.

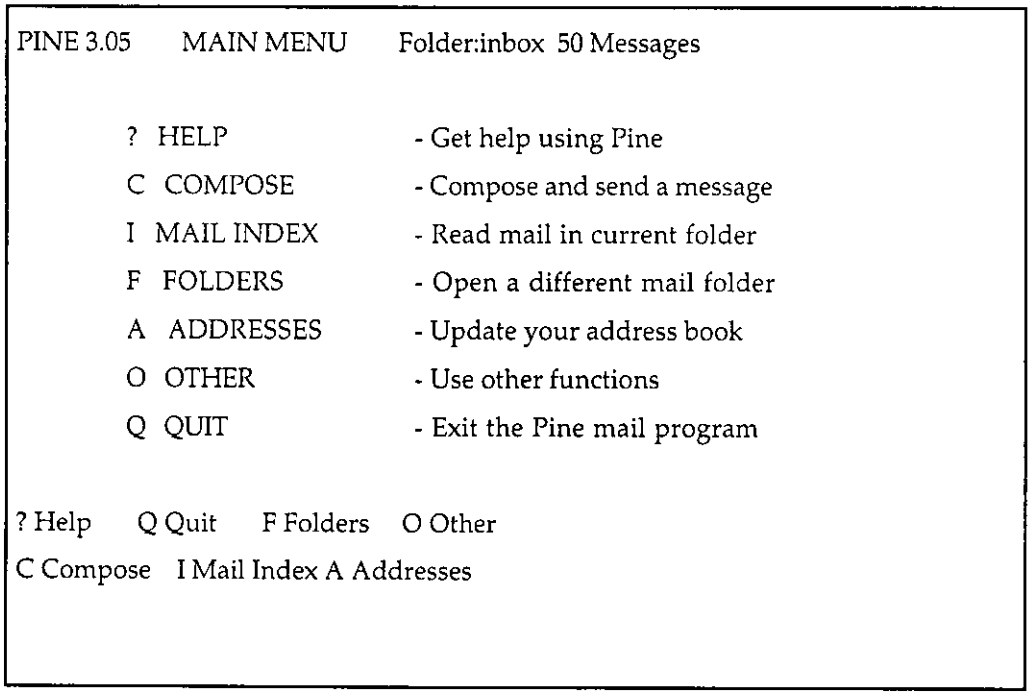
Electronic Mail (PINE)

What Is It?

PINE stands for "Pine is nearly ELM," signifying the fact that the mail software PINE is based on an older mail software called ELM. PINE is the most user-friendly mail software available to Internet users with UNIX-based computers. PINE offers a wide variety of useful commands prominently displayed in menus, and it incorporates a very easy-to-use editor known as PICO (also available separately outside PINE). Using PINE commands and the editor, one can quickly create, edit, send, display, and save electronic mail.

How Do I Access It?

After you login to a UNIX-based system, type in the command pine and press enter. You should then see the menu below.



How Can I Use It?

PINE has a relatively small command set. At every mode, the appropriate commands are displayed with sufficient information describing what they do. For every mode, there is also context-sensitive help available (user should enter "?") that provides more detailed narrative information describing the functions related to commands and the user actions necessary to successfully use each command. Certain commands require that the user press the control key and another key at the same time. For example the command to send a message is displayed as "^X" which means the user should press the control key and the X key at the same time to execute the send command.

Detailed help information covering the full PINE software is available at the topmost mode (as soon as one enters the program). Some of PINE's major commands are described below.

Creating and sending a letter (Compose command)

Choose the command "C" from the main menu. You should see the screen below.

```
PINE 3.05    COMPOSE MESSAGE    Folder:inbox  50 Messages

To    :
Cc    :
Attchmnt:
Subject :
----- Message Text -----



^G Get Help ^C Cancel  ^R Rich Hdr        ^K Del Line ^O Postpone
^X Send    ^D Del Char ^J Attach          ^U UnDel Lin^T To AddrBk
```

This is the PINE edit mode. The top three lines are called header information. Beside the word "To:" the electronic mail address of the person to whom the mail will go should be entered. "Cc:" stands for carbon copy. If any

address is written beside "Cc:," a copy of the letter will be sent to that address as well. "Attchmnt:" refers to a text file or a graphics file name, which can also be included with the mail. Beside the word "Subject:" a succinct phrase describing the content of the letter is appropriate.

The body of the letter is placed under the "--Message Text--" line. It is highly encouraged that the letter be short (five to six pages maximum). To transfer longer documents one should use the FTP (File Transfer Protocol command). Upon completion of the letter, the "^X" command should be used to send the letter. Note that the arrow keys are active in this mode, allowing flexible editing capabilities.

Viewing, saving, and forwarding a message (Index command)

As soon as one logs into a UNIX-based computer, if there is new mail for that account, the system displays the message "You have mail." To view new mail, the user should enter PINE (see the section "How Do I Access It?" above) and use the command "I" to view a listing of new mail. When "I" is selected from the main menu the screen should look similar to following.

```
PINE 3.05      MAIL INDEX      Folder:inbox  Message 39 of  41

39 Apr 1 SLIS                  (2,574) Update from SLIS at UC Berkeley
40  Apr  2 S_SWANK@cvu.edu     (964) Internet Workshop
41  Apr  2 JPISCITELLI%ewuvms  (20,238) INTERNET PROGRAM



? Help     M Main Menu  P Prev Msg   - Prev Page  F Forward   D Delete
O OTHER CMDS  V View Mail  N Next Msg  SPACE Next Page  R Reply      S Save
```

At this point, by pressing the enter key the message that is highlighted can be viewed. The highlight bar can be moved up or down by using the arrow keys. If message 41 were highlighted, it could also be saved by using the "S" command. If you wish to share your e-mail with someone else, you can forward that mail by highlighting it and then pressing the "F" key. Other

useful commands related to the view mode are displayed at the bottom of the screen.

Managing folders (Folders command)

PINE makes it very convenient to store and manage incoming and outgoing mail. One can create specific "folders" (actually directories) in which to store mail messages. PINE itself automatically creates and updates a "sent-mail" folder, which keeps copies of all mail you send. If the "F" command is selected from the main menu, a screen similar to the following is displayed.

```
PINE 3.05    FOLDER MAINTENANCE    Folder:inbox  50 Messages

inbox          sent-mail        saved-messages    geoscience
govbbs         internet         jobs              netdemo
sent-mail-dec-1992  sent-mail-feb-1993  sent-mail-mar-1993  serials
sigma



? Help     M Main Menu  G Go to Fldr  - Prev Page  A Add      D Delete
O Open     L Print   SPACE Next Page   R Rename     W Where is
```

The "Inbox" folder contains the most recent list of incoming mail. Folders such as "geoscience" and "sigma" were created by the user. By using the arrow keys the user can select particular folders and then, by pressing the return key, he or she can open each individual folder. To display and manipulate contents of a folder, go back to the main menu mode and select the "I" (Index) command.

Managing addresses (Addresses command)

Often electronic mail transfers take place between individuals who know each other and have communicated in the recent past. To relieve the user from having to type in the address of known individuals PINE provides the "A" (Addresses) command as a main menu option.

While viewing a mail (by using the "I" command), the user can subsequently select "T" to copy the address of the sender and save it in the address book. PINE prompts the user for a nickname and a full name under which to save the address. Then later, when sending a mail to that person, the user can retrieve that address by nickname or full name.

Hints

PINE offers detailed narrative help information for every mode. Use the "?" command to request help. Remember to regularly clean up your "Inbox" folder by marking messages as deleted and then issuing the "E" (Expunge) command.

Telnet

What Is It?

Telenet is one of the most fundamental software tools used to connect computers on the Internet. By using Telnet software, an individual may establish a virtually transparent connection to a remote computer and use all of the resources available to that computer. Through such a connection a user may, for example, access databases at other institutions, with the simplicity of standard local access.

How Do I Access It?

Telnet is invoked by typing "telnet" at your computer prompt, followed by the name or IP address of the remote site. After a connection is established, you will encounter the normal login screen of the remote computer or network that has been accessed.

Here is an example of an initial Telnet connection to the Cleveland Freenet, a community network in Cleveland, Ohio.

```
emx.cc.utexas.edu% telnet freenet-in-a.cwru.edu
Trying...
Connected to hela.INS.CWRU.Edu.
Escape character is '^]'.
```

How Can I Use It?

You can use Telnet to access library catalogs, campus information systems, and other special databases at other institutions. To access the services

```
emx.cc.utexas.edu% telnet freenet-in-a.cwru.edu
Trying...
Connected to hela.INS.CWRU.Edu.
Escape character is '^]'.

4.3 BSD UNIX (hela) (ttyrb)
                /\
 WELCOME TO THE...
              _| |_
             _|__ __|_
   __         |     |
 _|  |_       |  |  |  |
 |    |  /\   |  |  |  |
 |    |  |  |  |  |  |  |__
 |    |  |  |  |  |  |  |  |
 |    |_|_ |  |  |  |  |  |
 |    |  |  |  |  |  |  |  |
 _|    |  |_|_ |      |  |_
 |     |    |_|      |   |
 |                   |
 |      CLEVELAND FREE-NET       |
 |    COMMUNITY COMPUTER SYSTEM    |
 |_____________________________________|

brought to you by
Case Western Reserve University
Community Telecomputing Laboratory
Are you:
     1. A registered user
     2. A visitor
Please enter 1 or 2:
```

directly with Telnet, you will need to know the address of the site and also the login instructions for the particular service. Here is the complete connection procedure to the Cleveland Freenet.

You May Have Already Used Telnet!

Telnet is commonly implemented as part of other information access tools such as Gopher, WAIS, or Campus-Wide Information Systems. In fact, these tools frequently rely on Telnet to retrieve the information presented in their menus. The Telnet connection is hidden to the user when implemented in such a system, but the process is fundamentally the same. Many users will discover an information resource through a local Gopher system but later decide that it is faster and easier to bypass the Gopher's menus and Telnet directly to the site.

Hints

- TN3270 is a program very similar to Telnet but is designed specifically for connecting to systems requiring IBM 3270 terminal emulation. The University of Texas IBM VM system, UTXVM, requires a TN3270 connection. The command syntax is the same as Telnet.

- Rlogin is a program similar to Telnet but designed for connecting to another personal account of the user. Rlogin sends login information to the remote account automatically, speeding up the connection process.

UNIX

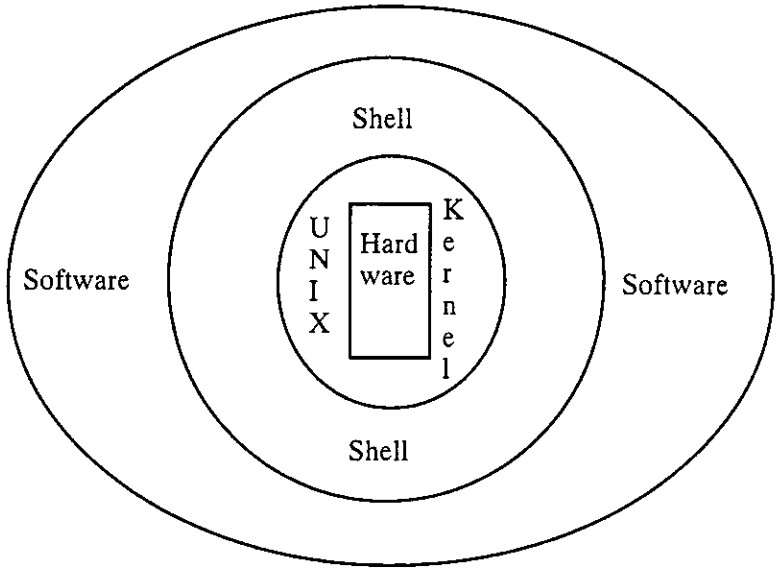
What Is It?

UNIX is the name of a computer operating system and its associated utility programs. A computer operating system controls computer resources and schedules tasks. UNIX is software that also controls the hardware. It was developed at the AT&T Bell Laboratories in Murray Hill, New Jersey, and has gained widespread acceptance at both commercial and academic institutions as the operating system of choice.

In the early days of computing, programmers submitted jobs to the processor in batch mode and the computer handled them one at a time. UNIX is a multi-user interactive system which allows many people to share the resources of the computer simultaneously. It is process oriented. When you logon to a multi-user system, through what is called the "shell," you begin a separate process which the operating system controls. UNIX is comfortable with many processes running together. The UNIX operating system which controls the hardware is referred to as the UNIX system kernel.

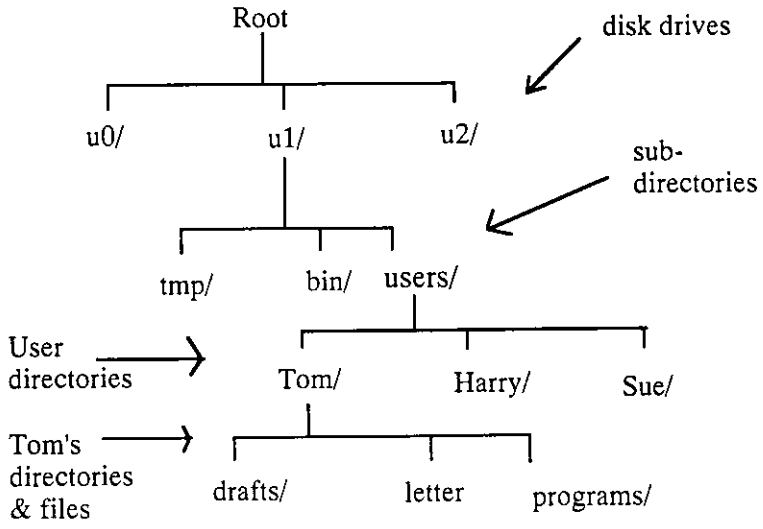
What is a shell?

The interface between you and the operating system kernel is called a shell. It is a command interpreter that translates your input to the system and executes the commands. There are many shells, but the most prevalent are the C Shell (csh), the Bourne Shell, and the Korn Shell. The nature of each of the shells is beyond the scope of this book, but many guides are available should you desire further explanation. The following diagram depicts the layers of a UNIX system.



What is a UNIX file?

UNIX organizes everything in files and directories of files. This structure is hierarchical and can be likened to the nature of a tree. At the base of the tree is the root. This is the foundation of the system. All the files of the system are contained within directories which spawn from this root. A file is a collection of information and can come in many forms. It can be a file of plain ASCII text, it can contain special formatted text that has been modified by a word processor; or it can be data that are interpreted by the system as executable commands (these types of files are called object code). In fact, UNIX even treats physical devices such as hard drives and printers as files. All files have unique names and are grouped within directories to make later retrieval or use more efficient. A typical directory structure is outlined here.



Whenever Tom logs in to his account, the computer orients him to his "home" directory, which in this case is called "Tom." Tom has two subdirectories ("drafts" and "programs"). Subdirectories in this model are indicated by the slash ("/") following the name. Those names with no suffix are files (such as the item labeled "letter"). In order to use files (remember a file can be text or executable code), you must be able to move about the directory structure and manipulate those files (actions such as copying, moving, and deleting). There are many commands available within the UNIX system, and they vary somewhat from shell to shell. If you can master the basic file manipulation commands and those required to navigate within the directory structure, you'll have the basic skills necessary to access those Internet information resources discussed in this book.

How Do I Access It?

If you have an account on a university computer, chances are good that you are using the UNIX operating system. We will assume that you are already comfortable with logging into your account. If not, your local computation center personnel probably have documentation to guide you through this process.

How Can I Use It?

There are many commands and utilities available to the UNIX user. The most common commands and their syntax are described below.

Command	Syntax	Use
cd	**cd** *[directory]*	To move between directories. Examples: cd ~ (return to your home directory) cd drafts (change to drafts directory) cd /bin (change to bin directory)
mv	**mv** *[options] old-name new-name*	**mv** is used to move a file from one directory to another. You can change its name in the process if desired. Examples: mv mom.txt ~/mom.txt Moves "mom.txt" from current location to the "home" directory while keeping the same name (C shell). mv mom.txt /drafts/mother.txt Moves "mom.txt" from the current location to the "drafts" directory and changes the name to "mother.txt"

Command	Syntax	Use
mv continued		
	Option: -i	Prompts you to ask if it is OK to overwrite an existing file.
rm	**rm** *[options] file-list*	Removes a file.
	Examples: rm mom.txt	Removes "mom.txt" if it exists in the current directory.
	rm /drafts/mom.txt	Removes "mom.txt" from the "drafts" directory if it exists.
	Option: -r	Deletes the contents of a named directory and the directory itself.
ls	**ls** *[options] file-list*	Displays the contents of the specified directory (file-list). If no file-list is specified the current directory is assumed.
	Examples: ls	Lists current directory contents.
	ls /drafts	Lists contents of "drafts."
	Option: -a	Displays all entries, including special files, that configure your environment.

Command	Syntax	Use
ls continued	Option:	
	-l	Displays directory contents in a long form that shows information about files including size and date.
more	**more** *[options] file-name*	Displays the contents of a plain ASCII text file one screen at a time.
	Example: more mom.txt	Displays mom.txt one screen at a time on the monitor.
	Option: -n	Specifies the number of lines displayed at one time.
mkdir	**mkdir** *directory-name*	Creates a directory with the name "directory-name."
	Examples: mkdir final	Creates a directory named "final"
	mkdir /drafts/papers	Creates a subdirectory called "papers" off of the "drafts" directory.
rmdir	**rmdir** *directory-name*	Removes named directory as long as it is empty.

Command Syntax		Use
rmdir continued		
	Examples: rmdir drafts	Removes the directory "drafts" if it contains no files.
	rmdir /drafts/letters.	Removes the directory "letters" if it contains no files.
who	**who**	Displays the current users logged into the system.
	Example: who	Responds with a list of current users arranged by login name.

Hints

- Many fine guides to UNIX are available. Your local computation center may have documentation available that covers more commands than are discussed here.

  Four recommended texts of the UNIX system are:

  Student's Guide to UNIX, Hahn, Harley, 1993.
  A concise guide aimed toward students.

  Introducing the UNIX System. McGraw-Hill, 1983.
  A good introductory text for new UNIX users.

  Peter Norton's Guide to UNIX. Bantam Books, 1990.
  Useful for someone with a DOS background.

Learning the UNIX Operating System. O'Reilly & Associates, 1989.

A brief guide aimed at a nontechnical audience.

- UNIX stores explanatory text of its various commands in what are referred to as "man" (for manual) pages. At the prompt, type "man" followed by the command in which you are interested. The system will then display more specific help regarding that command. For starters, type "man man."

- Experiment with the various Internet information tools that may be available on your system. Try Gopher for starters (type "gopher"). It should point you toward more resources.

- Find out from your system administrator which e-mail program is installed and how to use it. Ask for any available documentation. If there is none, try using the "man" command to gather more information (type "man" followed by the name of the program, e.g., "man pine").

USENET

What Is It?

USENET is a very large collection of computers that share data with each other. It is not an organization; it is not an academic network; it is not software; and it is not the Internet. USENET originated as a link between two universities in late 1979. Although its roots are in academia, most USENET sites today are commercial organizations. It originated in the United States but now extends throughout the world.

USENET is a cooperative agreement between people and organizations to exchange information and ideas in the form of "articles." The system administrators of computers at what are called USENET "feeds" have cooperative agreements to store and transport USENET articles. These articles are grouped into what are called "newsgroups." Users at particular sites can read articles using a variety of software. The process differs from an electronic discussion group that uses e-mail to communicate (such as a LISTSERV). With electronic discussion groups, if you are a subscribing member you receive all of the messages via e-mail whether you want them or not. The opposite is true of USENET articles. They are stored at participating sites, and you must use a newsreader program to access them when you need or want them.

The variety of discussion that occurs over USENET is staggering. The range of topics includes scientific, recreational, political, computer-related, and many, many others. There are more than 2,000 individual newsgroups, but they generally fall into one of a few dozen hierarchies. The most common among these appear below.

Newsgroup Hierarchy	Description	Example
alt	Alternative/popular	alt.books.review alt.fan.letterman
comp	Computer related	comp.UNIX.questions comp.sys.ibm.pc.games
misc	Miscellaneous topics	misc.invest misc.consumers.house
news	News about USENET	news.announce.newusers
rec	Recreational subjects	rec.autos, rec.arts.books rec.arts.tv.soaps
sci	Science subjects	sci.space, sci.bio
soc	College socializing Discussion of culture	soc.college soc.culture.bangladesh
talk	Lengthy debates on current issues	talk.politics.mideast

Newsgroup names are similar to IP addressing scheme in that from left to right each separate string runs from the general to the more specific (consider for example the group talk politics.mideast). Most USENET newsreader programs have the ability to search for newsgroups based upon a keyword, making it easier to find the group (or groups) that fit your interests or needs.

How Do I Access It?

If you have an account on a university mainframe computer, chances are that you have access to software to read news. There are many applications that can read USENET articles, including the following.

Software Name	Features
vnews	Screen-oriented
rn	Full-screen display; screen-oriented; user-definable patterns for reading and discarding; user-definable macros
trn	Similar to rn but allows the tracing of messages along subject headings

If you have a mainframe account either through a university or other organization, ask your system administrator if a "newsreader" has been installed. If so, request information regarding the particular software used and how to get more information.

If you are accessing USENET through a microcomputer-based application, consult your software documentation for specific guidance.

Once you have access to USENET articles, the first thing you should do is read the articles in news.announce.newusers. This newsgroup contains many valuable articles about the history of USENET, advice on net etiquette (netiquette), information on posting messages, and other information useful to new readers.

How Can I Use It?

USENET newsgroups are an excellent way to broaden your informational horizons. People use newsgroups to debate ideas, seek advice, post local and national news, ask for travel information, and for many other purposes. How you benefit from USENET is limited only by your imagination and inquisitiveness.

For specific procedures and commands, refer to the software documentation for your local system.

Hints

- Find someone who has experience with USENET and ask for a demonstration. At first, many of the newsreaders can be confusing; the help of an experienced user will speed you along the path toward becoming an expert.

- Keep in mind that in a societal sense, USENET is a culture and has developed its own customs.

- Never forget that you are communicating with another person, not a machine.

- For more information, use one of the information tools in this book (Gopher, VERONICA, etc.) to search for information regarding USENET. There are many documents describing USENET floating around cyberspace.

- Other good rules to live by: *

  Don't blame system administrators for their user's behavior.
  Be careful what you say about others.
  Be brief.
  Write postings you can be proud of.
  Use descriptive titles.
  Think about your audience.
  Be careful with humor and sarcasm.

(*From the USENET FAQ by Gene Spafford.)

VERONICA

What Is It?

Navigating through "Gopherspace" can open pathways to new and exciting information resources. By exploring these resources you will learn of new sites and servers. But what if you have an information need and haven't found a Gopher server that satisfies your requirements? Fortunately, a software utility exists that allows a person to search Gopherspace using keywords. This software is called VERONICA (for Very Easy Rodent-Oriented Net-wide Index to Computerized Archives). It was developed by Steve Foster and Fred Barrie at the University of Nevada and currently runs at a number of sites, including at the New York State Education and Research Network (NYSERNet).

VERONICA is a program that takes your keywords and searches all available Gopher server menus for matches. It then builds a menu of those locations on your screen which can be searched along the same lines as a Gopher session. VERONICA is not a full-text search device. It only searches the menu item names, not the contents of documents or files. VERONICA is typically an option on a Gopher client menu (such as item #7 of the University of Texas Gopher menu below) - a typical VERONICA menu follows the Gopher menu below.

How Do I Access It?

To search using VERONICA, you must generally use a Gopher server site. At the University of Texas, the Gopher server offers VERONICA searching as item #7 below.

```
                Internet Gopher Information Client v1.12S

               Root gopher server: gopherhost.cc.utexas.edu

    --> 1. About this Gopher/
        2. UT-Austin/
        3. Library/
        4. Texas/
        5. World/
        6. Jughead: Search gopher menus at UT Austin/
        7. VERONICA: Search gopher menus worldwide/
        8. Archie: Search FTP sites worldwide/

Press ? for Help, q to Quit, u to go up a menu                Page: 1/1
```

By using the arrow keys to select VERONICA and then pressing the return or enter key, the following screen appears.

```
                Internet Gopher Information Client v1.12S

                VERONICA: Search gopher menus worldwide

    -->1. Search gopherspace at NYSERNet <?>
       2. Search gopherspace at PSINet <?>
       3. Search gopherspace at University of Pisa <?>
       4. Search Gopher Directory Titles at NYSERNet <?>
       5. Search Gopher Directory Titles at PSINet <?>
       6. Search Gopher Directory Titles at University of Pisa <?>
       7. (empty presently)
       8. FAQ: Frequently-Asked Questions about veronica (1993/08/23).
       9. How to compose veronica queries (NEW June 24) READ ME!!.

Press ? for Help, q to Quit, u to go up a menu                Page: 1/1
```

The screen above provides a choice of three different VERONICA servers, and two search options for each server. The menu also offers information about frequently asked questions regarding VERONICA, as well as instructions on how to search VERONICA.

How Can I Use It?

By selecting item #1 "Search gopherspace at NYSERNet <?>", you are prompted to enter keywords. This screen and a search keyword entered by the user appear below.

```
            Internet Gopher Information Client v1.12S

           VERONICA: Search gopher menus worldwide

-->1.  Search gopherspace at NYSERNet <?>
   2.  Search gopherspace at PSINet <?>
   3.  Search gopherspace at University of Pisa <?>
   4.  Search Gopher Directory Titles at NYSERNet <?>


 +----------------------Search gopherspace at NYSERNet----------------------+

 Words to search for  NREN

                      [Cancel ^G] [Accept - Enter]       |

  +-------------------------------------------------------------------------+


Press ? for Help, q to Quit, u to go up a menu              Searching../
```

Upon entering keywords on which to search, VERONICA queries a database of Gopher menus for matches. Once the matches are found, VERONICA creates a new menu consisting of those sites that match the search term. Each item on this new menu is actually a Gopher link to the keyword item in the form of a retrievable file or submenu. The results of a keyword search appear below.

Veronica

```
                Internet Gopher Information Client v1.12S

                Search gopherspace at NYSERNet: NREN

-->1.  Interagency Interim NREN Implementation plan.
   2.  8.4.1 NREN.
   3.  Senate Passes Its NREN Version.
   4.  nren-discuss Mail List.
   5.  NREN.plan.
   6.  open-road-nren.
   7.  NREN -- Privacy -- Proposal -- CPSR.
   8.  Libraries and the NREN.
   9.  Kapor - Privatized NREN.
   10. "Building the Open Road: NREN as Test-Bed for the National Public.
   11. NREN/
   12. Internet/NREN Business Journal.
   13. 93-05-10-15: Tom Grunder Is Correct on Why NREN Should be Opposed.
   14. nsf9224     -'NSF 92-24 - Network Information Services Manager(s) ...
   15. nsf9352     - NSF 93-52 - NETWORK ACCESS POINT MANAGER, ROUTING AR...
   16. KQED radio panel on NREN proposal.
   17. A Voice in NREN at Last? (fwd).
   18. A Voice in NREN at Last? (fwd)/

Press ? for Help, q to Quit, u to go up a menu          Page: 1/12
```

Hints

- Remember that VERONICA only searches Gopher menu titles, not the contents of individual files.
- Once you have found a potentially valuable resource, you can mark your spot using a "bookmark." This saves connection and location information, which allows your local Gopher to reestablish the link when called upon to do so.

WAIS

(Wide Area Information Server)

What Is It?

WAIS (Wide Area Information Server) is a searching interface for approximately 500 public databases available over the Internet. These databases reside on computers at academic and commercial institutions worldwide, and cover various fields in the sciences, humanities, and other disciplines. Several of the databases are the result of research projects, and others represent volunteer efforts. Many databases are part of a local information system such as a Campus-Wide Information System (CWIS). Others may be accessed via Gopher menus or by a direct Telnet connection. The information is stored and accessed differently at these various sites; thus, the purpose of WAIS is to provide a consistent, uniform searching tool to access the different databases.

How Do I Access It?

There are several different versions of WAIS software designed for the different types of computers on the Internet. The two main categories are the text-based WAIS server and the graphically oriented WAIS software available for workstations and PCs. The graphic programs offer more flexibility in searching and the ability to customize the program environment. However, these programs are not as accessible as the text-based WAIS servers, which are available to any user with terminal-emulation access to the Internet. One of the main text-based servers resides at quake.think.com. A user may Telnet directly to this site and login as wais. A WAIS connection may also be offered as an option on a local Gopher menu or CWIS. Here is an example of an initial connection to a WAIS server.

```
emx.cc.utexas.edu% telnet quake.think.com
Trying...
Connected to quake.think.com.
Escape character is '^]'.



SunOS UNIX (quake)


login: wais
Last login: Sat May  8 23:45:28 from Violet.CCIT.Ariz
SunOS Release 4.1.1 (QUAKE) #3: Tue Jul 7 11:09:01 PDT 1992

Welcome to swais.
Please type user identifier (optional, i.e user@host): richie@emx.cc.utexas.edu
TERM = (vt100)
Starting swais (this may take a little while)...
```

How Can I Use It?

WAIS differs from many other network information tools in that it requires plenty of decision making and information filtering by the user. Unlike tools such as Archie and VERONICA, which automatically perform global searches, WAIS software requires the user to specify which databases are to be searched. The user must review the different databases beforehand to decide which ones are appropriate. The resources are then selected as "targets" for the search that follows. The WAIS server offers a list of the available resources, and the user simply scrolls down the list choosing appropriate sources. This narrows the search considerably. Here is a snapshot of the opening WAIS screen, showing the beginning of the long list of available databases.

```
SWAIS                Source Selection          Sources: 498
  #      Server                    Source                          Cost
001: [       archie.au]          aarnet-resource-guide           Free
002: [ndadsb.gsfc.nasa.gov]      AAS_jobs                        Free
003: [ndadsb.gsfc.nasa.gov]      AAS_meeting                     Free
004: [    munin.ub2.lu.se]       academic_email_conf             Free
005: [wraith.cs.uow.edu.au]      acronyms                        Free
006: [   archive.orst.edu]       aeronautics                     Free
007: [ ftp.cs.colorado.edu]      aftp-cs-colorado-edu            Free
008: [nostromo.oes.orst.ed]      agricultural-market-news        Free
009: [   archive.orst.edu]       alt.drugs                       Free
010: [   wais.oit.unc.edu]       alt.gopher                      Free
011: [sun-wais.oit.unc.edu]      alt.sys.sun                     Free
012: [   wais.oit.unc.edu]       alt.wais                        Free
013: [alfred.ccs.carleton.]      amiga-slip                      Free
014: [    munin.ub2.lu.se]       amiga_fish_contents             Free
015: [      150.203.76.2]        ANU-Aboriginal-EconPolicies     $0.00/minute
016: [  coombs.anu.edu.au]       ANU-Aboriginal-Studies          $0.00/minute
017: [  coombs.anu.edu.au]       ANU-Asian-Computing             $0.00/minute
018: [      150.203.76.2]        ANU-Asian-Religions             $0.00/minute

Keywords:

<space> selects, w for keywords, arrows move, <return> searches, q quits, or ?
```

After selecting the target databases, you are ready to perform the search. This process is simple. WAIS prompts you for one or more keywords relating to the topic at hand. The WAIS software then contacts the preselected databases and searches each one for occurrences of the keyword(s). WAIS then returns a list of the sources found. The information under a source may then be displayed on screen immediately.

Here is a sample WAIS search for Quranic verses about roses. Look closely and you will see the appropriate database selected (indicated by the asterisk to the left), the keyword entered, and the status message at the bottom indicating that the WAIS software is connecting to the selected database.

```
SWAIS                Source Selection          Sources: 498
 #       Server                      Source                          Cost
379:  [   quake.think.com]          quake.think.com-ftp             Free
380:  [ vector.intercon.com]        Queer-Resources                 Free
381:   [lurker.dfv.rwth-aach]       queuing-literature-database     Free
382: * [   wais.oit.unc.edu]        Quran                           Free
383:  [    zenon.inria.fr]          ra-mime-zenon-inria-fr          Free
384:  [    zenon.inria.fr]          ra-zenon-inria-fr               Free
385:  [   munin.ub2.lu.se]          rec.gardens                     Free
386:  [ wais.wu-wien.ac.at]        rec.music.early                 Free
387:  [   wais.oit.unc.edu]        rec.pets                        Free
388:  [   wais.oit.unc.edu]        recipes                         Free
389:  [bloch.informatik.uni]       reports-abstracts               Free
390:  [     gopher.uv.es]          Research-in-Surgery             Free
391:  [     wais.cic.net]          rfc-index                       Free
392:  [   ds.internic.net]         rfcs                            Free
393:  [     ns.ripe.net]           ripe-database                   Free
394:  [     ns.ripe.net]           ripe-internet-drafts            Free
395:  [     ns.ripe.net]           ripe-rfc                        Free
396:  [ cmns-moon.think.com]       risks-digest                    Free

Keywords: rose

Initializing connection...
```

Here is the list of items returned by WAIS. It is a list of three sections of the Quran that contain a reference to the word *rose*. Because this particular database is a full-text record of the Quran, these documents retrieved are actual Quranic text. Each one may be chosen from the list, and the section of text displayed on-screen.

```
SWAIS                Search Results          Items: 3
 #  Score      Source                Title                          Lines
001:  [1000] (   Quran)  018.TheCave   /usr4/wais/data/Quran/        441
002:  [1000] (   Quran)  033.TheAllies  /usr4/wais/data/Quran/       356
003:  [1000] ( Quran)    069.TheSureCalamity  /usr4/wais/data/Qu     132



<space> selects, arrows move, w for keywords, s for sources, ? for help
```

Helpful Commands

Here are some basic commands used by the text-based SWAIS software.

j, down arrow, ^N	:	Move Down one source
k, up arrow, ^P	:	Move Up one source
J, ^V, ^D	:	Move Down one screen
K, <esc> v, ^U	:	Move Up one screen
###	:	Position to source number ##
/sss	:	Search for source sss
<space>, <period>	:	Select current source
=	:	Deselect all sources
v, <comma>	:	View current source info
<ret>	:	Perform search
s	:	Select new sources (refresh sources list)
w	:	Select new keywords
X, -	:	Remove current source permanently
o	:	Set and show swais options
h, ?	:	Show this help display
H	:	Display program history
q	:	Leave this program

Advanced Topics

Choosing Databases

Many of the databases have cryptic names, and the ones that have useful names are often deceptive. How can you decide which databases are appropriate for your search? An easy way is to search the source *directory-of-servers*. This may be searched like any other WAIS database. Directory-of-servers is a master index of all indexes. It will help clarify what kind of information is available through WAIS.

Search Strategy

One of the main drawbacks to WAIS in its current form is its inability to perform Boolean searches. WAIS simply searches for all items containing the supplied keyword. This is yet another reason why WAIS requires more work from the user than other network tools. Boolean searches will be supported in future versions of WAIS software.

Interpreting Results

Each document listed on a WAIS search is assigned a score that indicates its usefulness. The score is related to the number of occurrences of the keyword in the document. The maximum score is 1000. Search results are sorted from highest to lowest score. Unfortunately, because of the current shortcomings of WAIS keyword searches mentioned above, this scoring device is not always accurate. This too will be improved in future software revisions.

World-Wide-Web

(W^3 or WWW)

What Is It?

The World-Wide-Web (W^3) is a wide-area hypermedia information retrieval initiative whose goal is to provide access to a large universe of documents. The project blends information retrieval and hypertext into an easy yet powerful to use global information system. W^3 is based on the philosophy that in general, academic information should be freely available to anyone.

This global information web was originally designed for high-energy physicists but now has a far broader utility to the Internet community. The web consists of documents and links to other documents that are actually nodes within the W^3. This linking of documents allows the searcher to navigate a weblike information structure from general to more specific or related topics. Some documents are text files that can be displayed on-screen while others are indexes that can be searched. Using search tools, W^3 creates virtual documents or indexes linked together ubiquitously. You may never know it, but the virtual document you are reading may be a compilation of bits and pieces retrieved from geographically remote computers (servers).

How Do I Access It?

W^3 as a searching tool is often accessible through campus information services such as Gopher. If your Internet access provider does not offer W^3 as a tool on a Gopher host menu, it can also be accessed through a remote Gopher. Additionally, if your system supports Telnet, you can Telnet to info.cern.ch (Geneva, Switzerland) or www.njit.edu (New Jersey Institute of Technology) to explore this powerful tool.

Below is an example of how W^3 can be reached through the University of Texas at Austin information server UTINFO.

```
Wide-area Information Access Tools
        1. Archie
        2. Gopher (University of Minnesota)
        3. Netfind
        4. Wide Area Information Server (WAIS)
        5. WorldWideWeb (W3)

      Press RETURN alone to see previous menu

      Press Control-C Q <return> to exit at any time

      Enter the number of your choice: 5
```

Selecting #5 above to begin W^3 will display the screen below. It is what is referred to as your home screen and is at the "virtual" top of your document. If you ever get lost moving through the web (and it's easy to do), type "home" at the prompt and you will be returned to the screen below.

```
                        GENERAL OVERVIEW
 There is no "top" to the World-Wide Web. You can look at it from many points of
view. If you have no other bias, here are some places to start:

 by Subject[1]      A classification by subject of interest. Incomplete
                    but easiest to use.

 by Type[2]         Looking by type of service (access protocol, etc.) may
                    allow you to find things if you know what you are looking
                    for.
 About WWW[3]       About the World-Wide Web global information
                     sharing project

  To use a different default page, perhaps one representing your field of
  interest, see "customizing your home page"[4].
1-6, <RETURN> for more, Quit, or Help:   1
```

How Can I Use It?

Once at the home screen you are at the beginning of your information quest. From this point you navigate along the strands of the web by entering the appropriate number of the item you want to explore. For instance, if you were interested in computer-related information you would enter "1" to begin navigating the web along subject lines. Notice at the bottom of the home screen (above) the available options. There are six links at this level (push return to see the next page of links), or you can quit, or ask for help. The other options on this screen include a search by "type of service" (number 2), information about W^3 (number 3 is highly recommended reading), and information about customizing your home screen (number 4).

Entering "1" to navigate by subject and search for computer-related information calls up the following screen.

```
                The World-Wide Web Virtual Library: Subject Catalogue
                              WWW VIRTUAL LIBRARY

This is the subject catalogue. See also arrangement by service type[1] .
Mail www-request@info.cern.ch to add pointers to this list, or if you are
prepared to take over administration of a subject area.

Aeronautics          Mailing list archive index[2] . See also NASA LaRC[3]

Agriculture[4]       Separate list, see also Almanac mail servers[5] .

Astronomy and Astrophysics
                     Abstract Indexes[6] at NASA,  Astrophysics work at
                     FNAL[7] , Princeton's[8] Sloane Digital Sky Survey,
                     the STELAR project, Space Telescope Electronic
                     Information System[9] , the Southampton University
                     Astronomy Group[10] , the National Solar
                     Observatory[11] , Astrophysics work at the AHPCRC[12]
                    See also: space[13] .

Bio Sciences[14]     Separate list .

Computing[15]        Separate list.
1-85, Back, Up, <RETURN> for more, Quit, or Help:
```

Every item followed by a number is a link along the web toward a more specific topic. Looking again at the bottom of the screen, observe that there are

85 "subject" links at this level. Let's say we want to explore the subject of computing -- we would enter the number "15" to travel farther along the web.

```
                    Information on Computing
COMPUTING

Mail additions to this list to www-request@info.cern.ch.

SPECIALIZED FIELDS

Algorithms[1]       MIT Algorithms indexes and Multigrid algorithms

Artificial Intelligence
            MIT AI Lab[2]; CTAN[3] (the Center for Applied and
            Theoretical Neuroscience)

Audio           The audio formats guide[4] .

Jargon[5]       Computer hacker's jargon index

Human-Computer Interaction
            HCI bibliography[6] ,

Hypertext and Information[7]
            Separate list, conferences[8] ,..

1-31, Back, Up, <RETURN> for more, Quit, or Help:
```

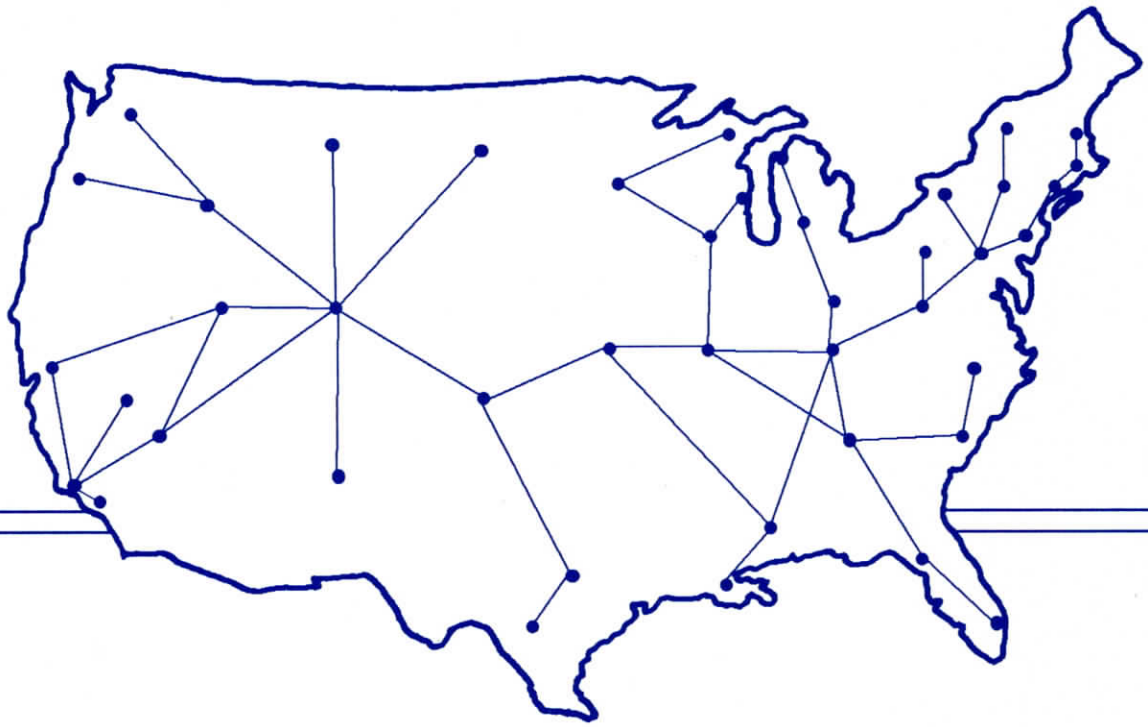
The screen above is only a partial listing of this level, but gives an idea of the nature of navigating along the web.

Hints

- Use "home" at any prompt to return to the top level/home screen.
- Explore the information about W^3 available at the home screen to expand your knowledge base and learn more about the project. This link includes information regarding the history of the project, mailing lists, FAQ lists, and pointers to W^3 servers.

- Use FTP to connect to info.cern.ch to retrieve text documents related to W^3. Login as "anonymous" and use your e-mail address as the password. The related files are in the directory pub/www/doc.

Index



ISBN 0-931510-50-3

Cover Design Joan Garner