4 Internet and Web Technology

In this chapter …

- From Soap Operas to Tidal Waves 70
- Computer Network Basics 71
- Packet Switching and TCP/IP 73
- The Web 76
- Basic Internet Tools 79
- The Domain Name System 83
- Advanced Searches on the Web 86
- Future Trends and Implications 88
The Internet is the core of Internet business and the Web is the core of the Internet. What exactly is the Internet? What is the Web? How do they differ? First the Internet. The Internet can trace its roots back to the Soviet launch of the Sputnik satellite in October, 1957. This Soviet achievement created quite a crisis in the United States. If people were not worried about the Soviets launching missiles from outer space at the United States, they were at least disturbed that the Soviets had leapt ahead in science and technology. At the time, the newly appointed Secretary of Defense was Neil McElroy, the former President and CEO of the Procter & Gamble Company. While he was at the helm of the company, he had a brainchild (in addition to the concept of brand management), which was to sell soap during radio and television dramas; hence, the name “soap opera.” It was clear that the United States needed to regain the upper hand in science and technology. As a means to do so, McElroy, a strong believer in research while running Procter & Gamble, proposed creation of the Advanced Research Projects Agency (ARPA). Funding was approved by Congress in 1958 and the agency was established.

Fast forward to 1966. Bob Taylor was the Director of the Information Processing Techniques Office (IPTO) at ARPA. He had three computer terminals in his office (we are at 2 o’clock in Figure 2–1, our cycle of computing). One terminal was connected to a computer in Boston, one to a computer in San Francisco, and one to a computer in Santa Monica. Each computer had its own set of commands and programs and was unable to “talk” to the others. Taylor was frustrated by this, and his frustration led to the creation of what was known as ARPANET.

The ARPA network was based on a technology called packet switching and led to the development of the networking protocol called TCP/IP (transmission control protocol/Internet protocol). For most people, the word protocol is usually associated with diplomatic relations. The use of the word protocol in networking is probably due to Tom Marill and a project he did for ARPA in 1966 [66]. In the world of networking, the word protocol refers to a set of rules for exchanging messages over a network. So think of a protocol as a set of rules computers use on a network in order to talk to each other. A network protocol would include rules for acknowledging message receipt, error checking, and data formatting. In the next section, “Computer Network Basics,” we provide a very brief introduction to computer network technology. Most computer networks communicate using packet switching, and this is the focus of the section, “Packet Switching and TCP/IP.” The ARPANET evolved into the Internet. In fact, the Internet can be defined as a network of computers using TCP/IP.

In 1995, Bill Gates said, “The Internet is the most important single development to come along since the IBM PC was introduced in 1981. It is even more important than the arrival of graphical user interface (GUI). The Internet is a tidal wave. It changes the rules.” For a number of years the Internet was used mainly by universities. It was a research tool, not a tool for commerce. Email was the most commonly used Internet tool. This changed when Tim Berners-Lee created the World Wide Web. In the section “The Web” we learn what makes the Web unique and how it uses the Internet, but is not the
same thing as the Internet. The Web became popular with the common user when Marc Andreesen, then a student at the University of Illinois at Champaign-Urbana, developed a graphic user interface (GUI) for the Web called Mosaic. Microsoft’s realization of the importance of the browser and the Web led to the famous and well-documented “browser war” between Microsoft and Netscape. The competitive (or anticompetitive) strategies used by Microsoft in this war led to its Department of Justice lawsuit. The browser has become a universal interface, and more than any other software tool, led to the Internet business revolution. The browser is most certainly the killer app of the World Wide Web.

In the section “The Web” we introduce the three key components of the Web—the URL, HTML, and HTTP. These three things define the Web. Of course, the Web is just one Internet technology or tool. In the section “Basic Internet Tools” we learn about other tools such as FTP and Telnet that are also important for Internet business. Part of the URL is based on a naming (and underlying numbering) scheme for computers. The naming and numbering scheme has important implications for Internet business. The process of resolving the name of a computer into its address is covered in the section “The Domain Name System.” Also important for Internet business is to know where to go to get what you need on the Web, whether you’re looking for information, products, or services. How to conduct more directed and time-saving searches is covered in “Advanced Searches on the Web.” The chapter concludes with the section, “Future Trends and Implications.”

COMPUTER NETWORK BASICS ........................

Internet business would not exist without a computer network. In fact, it was the ability to network personal computers that made the computer an incredibly powerful tool for business. Imagine having to send someone a floppy disk or Zip disk by mail or courier every time you wanted to send information in digital form. Imagine not being able to submit an order using the Internet. In this section we look at some basic network concepts. A very simple computer network is illustrated in Figure 4–1. For obvious reasons this is called a star network. The hub at the center of this network is a very simple piece of networking hardware that takes the data from one computer in the network and passes them on to the other computers in the network. The network illustrated in Figure 4–1 is known as a local area network or LAN. A LAN is a communications network consisting of cables, computers, and network devices confined to a very small geographic region such as a single building, or floor of a building.

Tech Talk

Packet: When sending a message over a network (e.g., an email), the message is usually broken up into a smaller set of messages called packets.

A very important LAN technology is ethernet. This technology was developed by Robert Metcalfe and David Boggs at Xerox PARC. With ethernet, any computer on the network can send data packets to any other computer on the network. However, no two computers can “talk on the line” at the same time. If two computers send a packet at the
same time there is a “collision.” Ethernet is a technology for handling these collisions and retransmitting the packets. When a collision occurs the packets are retransmitted after a very small random interval. Ethernet is a very effective protocol and it is easy to connect a computer into the network.

**Tech Talk**

**Bandwidth:** When talking about computer networks, bandwidth refers to the capacity of a network. Bandwidth is often measured in kilobits per second (Kbps), megabits per second (Mbps), or gigabits per second (Gbps). The terms broadband, narrowband, and midband are often used to describe the bandwidth. These are not precisely defined, except in an ordinal sense, and the actual bandwidth associated with these terms is changing rapidly over time as network capacity increases. We will define broadband as at least 10 Mbps, midband as 1–10 Mbps, and narrowband as less than 1 Mbps.

**ALOHAnet**

ALOHAnet was a radio network designed to allow computers in the Hawaiian Islands to communicate with each other. It was developed by Norman Abramson, a professor at the University of Hawaii and a surfing (on real waves, not the Internet) enthusiast. ALOHAnet was built on the idea of retransmitting packets after a small random interval if there is a collision on the network. This is the basic idea used in ethernet.

When an organization’s local area networks are connected together over a larger geographical region, perhaps the world, they are called a *wide area network* or WAN. This
is illustrated in Figure 4–2. A router is a network device used to send or route packets from one LAN to another LAN. The LANs may be connected by dedicated leased phone lines, fiber optic cable, or through various wireless technologies.

![Figure 4–2 A basic Wide Area Network (WAN).](image)

If you are planning to install or upgrade your company’s network, you should be concerned about the following:

- Which kind of data are transmitted—voice, text, video?
- What is the network bandwidth? Is it narrowband, midband, or broadband?
- How reliable is the network?
- Will it provide the bandwidth to meet future needs?
- How is speed degraded by traffic load?
- How much does it cost to operate?
- How secure is it?
- What protocols are used? Are they compatible with other networking protocols?

We will touch on a number of these issues throughout the book. The emphasis of this chapter is the Internet and we want to concentrate on TCP/IP, which is the networking protocol used by the Internet.

**Packet Switching and TCP/IP**

Consider the computer network illustrated in Figure 4–3. Assume computer A in Chicago wants to communicate with computer B in San Antonio. There are different paths on this network between Chicago and San Antonio. One potential way to communicate is to select one of these paths and then establish a connection between computers A and B on this path. This path would be dedicated to A and B for the duration of their message exchange. Establishing a connection between two points using a specific path on a network
for the duration of a message exchange is called \textit{circuit switching}. This is exactly how the public switched telephone network (PSTN) works. It is also referred to as POTS, for plain old telephone service. The public telephone system is also analog, which means that information travels over it in wave form.

The problem with circuit switching is that the line is tied up regardless of how much information is exchanged. To see this, consider circuit switching applied to a highway system. Assume Figure 4–3 now represents a highway system and a person wants to drive from Chicago to San Antonio. With the circuit switching philosophy, you would select a path, say, Chicago to Memphis to San Antonio, on the highway system between these two cities and then reserve all of the highways on this path for the period of time it takes the person to drive from Chicago to San Antonio. This means that if a second person wanted to drive from Chicago to Nashville he would be blocked from using the Chicago to Memphis link until the first person arrived in San Antonio.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{packet_switching_network.png}
\caption{Packet switching network.}
\end{figure}

Circuit switching is a big problem for computers because their information exchange is typically “bursty” rather than even and constant. Two computers might want to exchange a file, but after that file is exchanged the computers may not engage in communication again for quite some time. Thus, we call the need to exchange data between computers “discrete,” not continuous. An alternative to circuit switching is \textit{packet switching}. The concept of packet switching was developed independently by Paul Baron and Donald Watts Davies in the 1960s. The interested reader may wish to see one of Paul Baron’s original memos [9] at the RAND Corporation describing packet switching. This work was done during the time of the Cold War and part of his motivation for developing packet switching was the design of a network that could withstand a nuclear attack.

To understand packet switching, now consider sending a letter from Chicago to San Antonio. Rather than sending the entire letter over a particular route, we could cut the letter into a set of packets. The packets are cut, addressed, and sent in a particular way, that is:
The packets are numbered so they can be reassembled in the correct sequence at the destination.

Each packet contains destination and return addresses.

The packets are transmitted over the network as capacity becomes available.

The packets are forwarded across the network separately and do not necessarily follow the same route; if a particular link of a given path is busy, some packets might take an alternate route.

Packet switching was radical thinking at the time. AT&T stated flatly that packet switching would never work. However, the work on packet switching very much influenced Larry Roberts at ARPA, who had been hired by Bob Taylor. He opened bidding to build a packet switched network for ARPA. Two important companies, IBM and AT&T, did not even bid. The contract was awarded to the Boston consulting firm Bolt Beranek and Newman, known as BBN. The project was led by Frank Heart.

Packet switching is a philosophy of network communication, not a specific protocol. The protocol used by the Internet is called TCP/IP. The TCP/IP protocol is due to Robert Kahn and Vint Cerf. The IP in TCP/IP stands for Internet protocol and is the protocol used by computers to communicate with each other on the Internet. Computers using the Internet must have IP software. Packets that follow the IP specification are called IP datagrams. These datagrams have two parts: header information and data. To continue with the letter analogy, think of the header as the information that would go on an envelope and the data as the letter that goes inside the envelope. The header information includes such things as:

- Total length of the packet
- Destination IP address
- Source IP address
- Time to live—the time to live is decremented by routers as the packet passes through them; when it hits zero, the packet is discarded; this prevents packets from getting into an “infinite loop” and tying up the network
- Error checking information

The IP packets are independent of the underlying hardware structure. In order to travel across different types of networks, the packets are encapsulated into frames. The underlying hardware understands the particular frame format and can deliver the encapsulated packet.

The TCP in TCP/IP stands for transmission control protocol. This is software that, as the name implies, is responsible for assembling the packets in the correct order and checking for missing packets. If packets are lost, the TCP software requests new ones. It also checks for duplicate packets. The TCP software is responsible for establishing the session between two computers on a network. The TCP and IP software work together.
An important aspect of packet switching is that the packets, just like a letter, have forwarding and return addresses. What should an address for a computer look like? Since it is a computer, and computers only understand binary (0/1) information, the most sensible addressing scheme is one based on binary numbers. Indeed, this is the case and the addressing system used by IP software is based on a 32-bit IP address. An example IP address is:

128.135.130.201

All IP addresses have this format: four sets of numbers separated by three periods. This is called *dotted-decimal notation*. Each of the four sets of numbers requires one byte (eight bits), for a total of 32 bits. By convention, we are stating the IP address in decimal rather than binary format. For more details on the IP addressing scheme we refer you to the Technical Appendix.

**Tech Talk**

*NAP: These are switching points in large cities where the very high bandwidth network backbones of major telecommunications companies meet.*

We have defined the Internet as a network of computers using the TCP/IP. Who built the network and how do you get connected? The Internet backbones, the high bandwidth fiber-optic cable connecting major cities in the United States, are controlled by major telecommunications companies like MCI WorldCom, Sprint, AGIS, PSINet, and BBN. Large Internet service providers are connected directly into the backbone at NAPs like San Francisco, Chicago, New York, and Washington D.C. A person or company connects to the Internet using an Internet service provider. This is discussed in more detail in Chapter 7, “Getting Your Business Online.”

You should be aware that there are many networking protocols other than TCP/IP. One of the great innovations of the Macintosh, in addition to the GUI, was easy networking. At one point it was much easier to network Macintosh computers than IBM PCs. The Macintosh networking protocol was called AppleTalk. Other well-known protocols include Token Ring and Novell NetWare.

**THE WEB..........................**

The Internet is a network of computers using TCP/IP. What is the Web? It is the creation of Tim Berners-Lee and is based on his major insight to combine hypertext with the already existing Internet.

First, the idea of hypertext. This term is due to Ted Nelson. The standard way of reading a book is in a linear fashion, starting with page one. The concept of hypertext is to allow a person to read or explore in a nonlinear fashion. The key concept is that hypertext contains “links” to other text. By following the links the reader is not constrained to follow any particular order. Hypertext may contain links that do not necessarily lead to other text, but to sound or video files. Before the Web came into being there were hypertext products in the marketplace. One such commercial product was Guide, distributed by Owl
The Web

If you clicked on a link in Guide, a new document would be inserted in place of the link. Also, in keeping with its tradition of being a great innovator, Apple Computer had a product called Hypercard that implemented hypertext. However, these products did not use the Internet.

Tim Berners-Lee was working at CERN (a European particle physics laboratory located near Geneva, Switzerland) in a department charged with processing and recording the results of the scientific experimental work being done there. At CERN there were scientists from many different countries, so there were many different computer operating systems and document formats in use. It was difficult for a scientist working with one computer system to obtain information from a colleague using a different computer system. This is the same problem that faced Bob Taylor at ARPA. Berners-Lee realized that it would not be feasible to force the wide mix of researchers at CERN to reorganize their way of doing things to fit a new system. It was crucial that everyone work with his or her own operating system, but still easily share information. Berners-Lee’s solution was to marry hypertext with the Internet. This marriage is the World Wide Web and consists of three key components, HTTP, HTML, and URL, all developed by Berners-Lee.

1. HTTP (hypertext transfer protocol): Recall that a protocol is a set of rules for exchanging information on a network. HTTP is a high-level protocol used to exchange information between a browser and a server. The HTTP protocol uses TCP/IP to locate and make a connection between the browser and the server. The messages sent between the browser and server are either request or response messages. The request message contains 1) a request line containing the name of the requested file and whether the request is a GET or POST (see Tech Talk in this section), 2) a header containing information such as the type of browser and operating system, and 3) a body containing data, for example, data entered into a form. The response from the server will contain 1) a response line with a code indicating that the requested file was found or an error code (almost everyone has had to deal with the dreaded HTTP 404 Error - file not found) if there was a problem, 2) header information such as the type of server software, and 3) a body containing the HTML of the requested file. An HTTP request and response is illustrated in Figure 4–4.

Tech Talk

GET and POST: In the request line sent from the browser to the server is an HTTP command called the method. The method is usually a GET or a POST. The GET method is a request for a specific URL. With a GET request, the body is empty. The POST method tells the server that data will be sent in the body of the request. The POST method is used when you submit forms.

2. HTML (hypertext markup language): This is the language used by the browser to display the text and graphics on a Web page. In the next chapter we describe what a markup language is, and how to create a Web page using HTML.

3. URL (uniform resource locator): This is the “address” of a Web page. When you click on a link in a Web page, you are taken to a new location. The link contains the URL for your destination and the URL must follow a very specific syntax used in naming the destination.
There are three parts to a URL. They are:

- The Internet protocol used, e.g., HTTP (or FTP or Telnet as discussed later)
- The address or name of the server
- The location and name of the file on the server

Consider the URL in Figure 4–5. In this example, the protocol is HTTP. The name of the server or host machine is gsbkip.uchicago.edu. The target file being requested by the browser is named foo.html. It is located in the directory tmp which is a subdirectory of htmls. Thus, the URL specifies the directory path for the requested file.

![Diagram of an HTTP request](image)

**Figure 4–4** An HTTP request.

![Diagram of the parts of a URL](image)

**Figure 4–5** The parts of a URL.
Suppose a user with a browser views a Web page that has a link in its hypertext to the file foo.html on the machine gsbkip.uchicago.edu. The text and graphics of the Web page are displayed according to the underlying HTML. The link contains the URL given above in the example so the packets know what server to go to and what file to retrieve. When the user clicks on that link an HTTP request for the file foo.html is sent over the Internet to the server machine, gsbkip.uchicago.edu. The server machine then returns the requested file. The beauty of this process is that the operating systems used by the desktop machine and by the server machine are irrelevant. They do not have to be compatible.

There are two pieces of software required for this process to take place. The desktop PC must have a browser such as Netscape Navigator or Internet Explorer. The server machine must have an HTTP server. The HTTP server software is “listening” for packets addressed to it. When we use the term server we are referring to two things. Server refers to both the physical machine as well as the software on the machine that is serving up the files. When the server software receives packets requesting files, it sends the requested file back to the desktop PC.

There are a number of server software packages on the market. The current leader is Apache with almost 60 percent of the market [122]. Apache is open source software. It runs in the Linux, Unix, and Windows environments. The name Apache came about because the software is “a patch” work of code from the numerous independent coders who worked on it. The Windows 2000 operating system comes bundled with Internet Information Server, which is Microsoft’s HTTP server. It has about 20 percent of market share. Sun Microsystems’ iPlanet is a distant third, with about 6.5 percent of the market.

BASIC INTERNET TOOLS ..........................

It is important to understand that there are important Internet protocols other than HTTP. The other protocols work in much the same way as HTTP in that there is a piece of client software making requests to a server machine running the appropriate server software. Here are some of the most common and useful protocols.

FTP (File Transfer Protocol): This existed before HTTP and as the name implies is a protocol for exchanging files over the Internet. There are two types of FTP: nonanonymous and anonymous. With nonanonymous FTP, you are required to have an account name and password on the server in order to access files on the server. When using nonanonymous FTP client software you must provide three things: 1) the name of the server, 2) the account name, and 3) your password.

When you use nonanonymous FTP you are connecting to a server and have access to certain files on the server for which your account has permission. Figure 4-6 shows the screen when you FTP to the server gsbkip.uchicago.edu using the account kmartin (that is, when you enter ftp://kmartin@gsbkip.uchicago.edu on your browser). As the figure indicates, you can click on file folder icons and “drill down” through the file structure in a hierarchical fashion just as though the server were your own machine. Note the contrasts between HTTP and FTP. First, when you FTP to a server and click on folder or file icons, you remain connected to that server machine. This is
in contrast to HTTP, where a click on a hypertext link may take you to a different server altogether. The second difference between HTTP and FTP is that with HTTP you can see the actual content of the files you request displayed using HTML, whereas with FTP you can see only the folders and directories where your requested file is housed. With FTP you must first download a file from the server onto your computer and then open that file, assuming you have the required software to read it. Note also that when using FTP, the only service we require from the server is to serve up files. We are not using the server CPU for computing.

FTP is an important protocol for Internet business. It is an excellent way for a company to exchange files quickly and securely with customers or partner companies. Perhaps the most common example is in distributing software. Many software companies offer their customers the option of using FTP to download software after making a purchase rather than sending a CD-ROM via the postal service. Another typical example is provided by a company like Hewlett Packard. You can go to the HP Web site and download via FTP the most current version of their printer drivers (software used by an application such as a word processor to communicate with a specific type of printer). Netscape has always used FTP as a means to distribute its browser. Of course, FTP use is not restricted to software companies. Companies might also use FTP to exchange purchase orders. It is also quite common in the Web development process to create and design Web pages on a client machine and then FTP the resulting work to the server machine of the Web-hosting company.
If you download software over the Internet using FTP, it is important to understand the following distinctions:

- **Public domain software**—this is software that is free, carries no copyright, and there are no restrictions on redistribution of the software. In this context, free refers to price, not free in the sense of the GNU GPL as developed by Richard Stallman and described on page 65. Public domain software may well be an executable binary file with no source code provided.

- **Freeware**—this is software that is free, but is copyrighted. There may be restrictions on the redistribution of freeware. Again, in this context, free refers to price, not free in the sense of the GNU GPL as developed by Richard Stallman and described on page 65. Freeware software may well be an executable binary file with no source code provided. Also, the policy on use and redistribution may be considerably more restrictive than in the GNU GPL.

- **Shareware**—this is copyrighted software that is distributed on a “try-before-you-buy” basis. You download the software using FTP. The software will often have an expiration date. You use the software until the expiration date. If you like the software, you purchase it after the trial period ends. Often there is no trial period and the users are on their honor to purchase this software if they decide to keep it. This is a surprisingly effective business model. John McAfee used this shareware model to build a very successful software company based upon his anti-virus software. Shareware software often requires a purchase of the software in order to get technical support. The technical support appeals to large corporations, which motivates them to pay for the software.

Shareware and freeware are often distributed using anonymous FTP. With anonymous FTP the user does not need a password. When using FTP client software, the word anonymous is used as the account name and any string of characters can be used for the password. Distributing files with anonymous FTP is a good way to maximize their availability.

Client software is necessary in order to FTP files from a server. There is much dedicated FTP client software available as shareware or freeware. You can also use a browser as an FTP client. When Tim Berners-Lee was developing the Web protocols he made a conscious decision to incorporate the already existing FTP protocol into the Web. This allows users of the Web to easily access the numerous documents available on FTP servers. The URL syntax for using anonymous FTP with a browser is:

\[
\text{ftp://name\_of\_server}
\]

and for nonanonymous FTP it is

\[
\text{ftp://account\_name@name\_of\_server}
\]

In the latter case you will be prompted for the password associated with account\_name.

In Chapter 2, “What Do I Need to Know about Hardware?,” we discussed file compression algorithms. Many files stored on the Internet for download are compressed. A
summary of the most common file extensions and associated compression programs is given in Table 4–1. Shareware or freeware programs are available on the Internet that allow a user to uncompress compressed files. Some programs provide the capability to uncompress without charge, but you are required to pay for a version of the program if you want to compress a file. This is yet another example of the classic Internet business model to give a useful product away free but charge for enhanced versions of the product.

Table 4–1 Compression Programs and Extension

<table>
<thead>
<tr>
<th>Extension</th>
<th>Compression Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>.Z</td>
<td>Compress (Unix)</td>
</tr>
<tr>
<td>.sit</td>
<td>Stuffit Deluxe</td>
</tr>
<tr>
<td>.sea</td>
<td>Self-extracting archive (Stuffit Deluxe)</td>
</tr>
<tr>
<td>.zip</td>
<td>Pkzip or Winzip - Windows and DOS</td>
</tr>
<tr>
<td>.exe</td>
<td>Windows/DOS self-extracting archive</td>
</tr>
</tbody>
</table>

Just as with the HTTP protocol, special FTP server software must be running on the server computer. A server computer may run more than one piece of server software simultaneously. However, for a company serving many users, there may be separate computers that are dedicated FTP servers. The Microsoft Internet Information Server comes bundled with both an HTTP and FTP server.

**Telnet:** In a sense, Telnet is the converse of FTP. With FTP you are using a dumb server—the server is only providing files and all the computing is done on the client machine. With Telnet, the client machine is nothing more than a dumb terminal. You connect to the server machine running Telnet server software and all the computing is done on the server machine. The Telnet URL has the form

telnet://name_of_server

In order to use the Telnet protocol you need Telnet client software. You cannot Telnet using a browser without this additional client software. When you Telnet to a server machine, you are given a command line (not GUI) interface to the server machine. This is akin to being at the 2 o’clock position in Figure 2–1, the old time-sharing mode of the late 1960s and early 1970s. Telnet is still widely used, especially in Unix environments. It is a great way to access the CPU of another machine and test software. Telnet is often used by systems administrators to connect remotely to computers they are managing.

**News:** Internet newsgroups are a very common way of exchanging information. There are newsgroups on everything from archery to zen. An example news URL is

news://uchinews.uchicago.edu/rec.backcountry

In this URL, news is the service required, uchinews.uchicago.edu is the news server, and rec.backcountry (a newsgroup devoted to backpacking) is the target newsgroup desired. In order to access a newsgroup you need client software, such as Microsoft Outlook, and the server needs to be running the appropriate server software.
Newsgroups are asynchronous. That is, you post a message, someone responds at a later date, and you go back to the newsgroup and read the response to your message. Chat rooms and instant messaging are becoming increasingly popular because they are synchronous.

Mailto: Electronic mail, better known as email, is perhaps the most popular Internet protocol. As with all Internet software tools, you need client software and server software to use email. Indeed, when configuring your client email software to use programs like Eudora or Microsoft Outlook, you have to provide the name of the server where you send and receive mail. You can also send email using your browser and the mailto protocol. There is a mail URL for the browser that has the format

mailto:johndoe@company.com

where mailto is the service required, johndoe is the name of the person (or whatever name the person chooses to go by) to whom you’re sending email, and company.com is the mail address. The mailto tool is similar to Telnet and news, in that you can execute the URL in a browser, but additional client software is needed in order to actually use the tool.

File or Path: Unlike the other URLs, this URL is used on the client machine only. It has a very useful function, which is to let a developer test a Web page before putting it on a server. Using this URL allows the developer to view a Web page through a browser where both the browser and Web page are on the client machine. For example, assume you have a file test.html that you wish to view in a browser. Assume this file is in the directory temp on the C drive of a Windows machine. Then the URL

C:\temp\test.html

will allow the user to view test.html in the browser. The client machine does not need to be connected to the Internet in order to use the file URL.

THE DOMAIN NAME SYSTEM .....................

The URL naming scheme developed by Tim Berners-Lee is very powerful. Recall that a URL has three parts, one of which is the name of the server. This name is called a domain name. An example of a domain name is gsbkip.uchicago.edu. The domain name is ordered right to left. The top level domain in this name is .edu, the domain for educational institutions. The current top level domains are given in Table 4–2. Besides the domains listed in Table 4–2 there are extensions that denote a foreign country. For example, .jp is the extension for Japan and .de for Germany. Reading right to left, the next part of the domain name in our example is uchicago (note that dots separate levels of the domain name). This is a second level domain, in this case, the University of Chicago. Finally, gsbkip represents a host machine within the University of Chicago.

When you click on a link with an underlying URL, you are taken to the computer named in the URL. Having a name for a server is important for Internet business. For example, if you wanted to go to the IBM Web site, you would naturally guess the name to be www.ibm.com. Certainly names like amazon.com have become valuable trademarks. Indeed, there have been numerous law suits over domain names. Many businesses invest a
great deal of time and money in selecting an appropriate domain name. It is important to have a name that is easy to remember. A good domain name can also be a marketing tool.

In an earlier section we saw that TCP/IP software uses IP addresses, not names, for addressing packets. This means that somehow the domain names must get converted or resolved into IP addresses. This is done through an Internet service called the domain name system (DNS). This system makes use of special servers called domain name servers. The process is illustrated in Figure 4–7.

For example, assume an employee sitting at a desktop PC makes a request for a file to the machine with domain name gsbkip.uchicago.edu. The domain name gets resolved from right to left. The first thing that would happen is that the local DNS server would make a request to a root server for addresses in the .edu domain. There are multiple copies of root servers throughout the world that do nothing but serve up addresses in top level domains. This root server would then provide the address for a DNS server that had the address for the second level domain uchicago.

In Figure 4–7 this is the server with the address 128.135.4.2. This server knows all of the addresses of host machines for the uchicago domain including the host machine gsbkip. It then responds to the local DNS, sending it the IP number 128.135.130.201, which is the IP address for the domain name gsbkip.uchicago.edu. The local DNS server then passes that IP number to the desktop PC that originally made the request. It now has an IP address associated with the domain name gsbkip.uchicago.edu and can send packets out over the Internet requesting files from the server (labeled enterprise Web server in the figure) with this domain name. All of this happens in a matter of seconds or even less. Simple!

### Table 4–2 Top Level Domains

<table>
<thead>
<tr>
<th>Domain Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>commercial users</td>
</tr>
<tr>
<td>.net</td>
<td>network providers</td>
</tr>
<tr>
<td>.edu</td>
<td>educational institutions</td>
</tr>
<tr>
<td>.org</td>
<td>nonprofit organization</td>
</tr>
<tr>
<td>.gov</td>
<td>United States government</td>
</tr>
<tr>
<td>.mil</td>
<td>United States military</td>
</tr>
<tr>
<td>.aero</td>
<td>(new)</td>
</tr>
<tr>
<td>.biz</td>
<td>commercial use (new)</td>
</tr>
<tr>
<td>.coop</td>
<td>(new)</td>
</tr>
<tr>
<td>.info</td>
<td>all uses (new)</td>
</tr>
<tr>
<td>.museum</td>
<td>(new)</td>
</tr>
<tr>
<td>.name</td>
<td>for individuals (new)</td>
</tr>
<tr>
<td>.pro</td>
<td>(new)</td>
</tr>
</tbody>
</table>
Clearly, in order for the process outlined in Figure 4–7 to work properly, the IP addresses and domain names must be globally unique. The organization entrusted with ensuring this is the Internet Corporation for Assigned Names and Numbers (ICANN). This organization was established by the United States Commerce Department for the purpose of making the Internet run smoothly. At one time, Network Solutions Inc., which is now owned by VeriSign, had a monopoly on assigning domain names. They no longer have a monopoly on assigning domain names, but they still have control of the domain name database for .com and .net and get $6 per year per name for maintaining this database. Now, if a company wants to register a domain name they can contact one of the ICANN Accredited Registrars. Many Internet service providers will, for a fee, take care of the process of getting an IP address and registering a domain name.

Many companies are discovering that if they want to register a domain name, the name is already taken. To help alleviate this problem, ICANN has approved seven new top level domain names. They are .aero, .biz, .coop, .info, .museum, .name, and .pro. Not surprisingly, there are companies trying to usurp the power of ICANN. One such company is New.net. They have come up with their own top level domain names like .family, .tech, and .xxx. However, because they are not approved by ICANN, their domain names are not part of the root server system. In order to reach host computers using the New.net domain names, a user, or the user’s Internet service provider, must use special software that adds the extension new.net onto every address [29].
Cybersquatting

Cybersquatting is the practice of registering a domain name for the purpose of reselling it at a later date. Early in the domain name registration process, clever individuals registered domain names related to well-known products or company names, e.g., McDonalds.com, with the express purpose of charging a large ransom fee on the name when the company finally decided to use it. This is also called domain name hijacking. This is now much more difficult to do. In October of 1999, ICANN approved a set of rules for the Uniform Domain Name Dispute Resolution Policy. This policy is followed by all domain name registrars for the top level domains of .com, .org, and .net. Victims of cybersquatting can make a formal complaint under this policy and be heard by an ICANN-approved dispute-resolution service provider. If someone has registered a domain name in "bad faith," there is high probability he will be forced to give it up. We refer the reader to the ICANN Web site for what constitutes "bad faith."

How can New.net establish new top level domain names without being given the authority to do so? No government or company owns the Internet. Organizations such as the W3C try to provide standards on HTML and HTTP, and ICANN tries to control domain names and numbers, but there is no central worldwide authority. Perhaps that is why the Internet has worked so well to date.

Advanced Searches on the Web

In addition to using Internet tools, you may wish to conduct advanced searches on the Web for your Internet business as another way to take advantage of what the Web has to offer. You may have already conducted basic searches using one of the search engines like Yahoo!, Google, Lycos, or Altavista. Yahoo! offers a "directory" form of search where users "drill down" through increasingly specific topic areas to arrive at the subject they are seeking. The other search engines offer an "index" form of search where users enter a keyword and the search engine returns a list of Web pages in which the keyword is used. See the section, "How to Get a High Ranking for Your Web Site" in Chapter 6, "Web Site Design and Content," for information on how search engines find Web pages and determine their ranking.

Some of the newer search engines even have features like categorizing the entries for you to help narrow your search. Vivisimo, for example, not only returns standard entries when the word "books" is entered, but also categories such as "music," "children," "rare-print," and "art" that can be clicked on for more efficient searches. Wisenut is another categorizable search engine that offers a feature called "Sneak-a-Peek" that allows users to
catch glimpses of listed sites without ever leaving the Wisenut site. This allows for faster searching without having to wait for the browser to switch between the listings and the search engine site [172].

For more sophisticated and directed searches that can help save time and money, more than one keyword with advanced search code words and syntax can be entered at a time. Say for example that you are interested in learning about different ways to promote your company’s Web site. It is difficult to think of a single keyword that would provide you with the information you need. However, once you have mastered the advanced search techniques listed here, you will find whole new worlds of information opening up for you. Advanced search code words and syntax differ from search engine to search engine, so check with each before using them. For more information on search engines see [13] and [57]. Here are a few of the more commonly used advanced search techniques:

- **“”** Quotation marks are used around a multiword phrase that you want searched verbatim. Using the example of wanting to learn about different ways to promote your Web site, you might enter “Web site promotion” as is, in quotations. The search engine would return a list of Web pages with the phrase Web site promotion in them. Without quotes around the phrase, each word in the phrase may be searched for separately (depending on the search engine). For example, if you enter the phrase Web site promotion without quotes, your search engine may do a search for pages with the word Web, or the word site, or the word promotion in it.

- **+** Inserting a plus sign before a keyword indicates that the keyword must be on a page returned. Just entering a keyword by itself may not necessarily return a page with that word in it. For example, if you’re searching for a consultant who might help you with your Web site promotion, if you enter +“Web site promotion” consultant some search engines will return pages with Web site promotion in it, but not necessarily the word consultant (some search engines rank pages with both phrases Web site promotion and consultant in it higher than pages with Web site promotion alone). In order to have pages returned with both entries in it, you would have to enter both terms with plus signs before each term, i.e., +“Web site promotion” +consultant.

- **AND** The word AND entered between keywords will find pages that have all of those keywords in a page. For example, if you are searching for software that will help you with Web site promotion, a search on “Web site promotion” AND software will find pages with both Web site promotion and software in it. Note that entering “Web site promotion” AND software is equivalent to entering +“Web site promotion” +software.

- **OR** The word OR entered between keywords will find pages that have at least one of the keywords in it, but not necessarily all keywords on a page. For example, a search on “Web site promotion” OR software finds pages with the phrase Web site promotion in it as well as pages with the word software in it, but not necessarily both entries in a single page.
• **NEAR** The word NEAR entered between keywords will find pages with the keywords in close proximity to one another in the page. For example, suppose you wanted to search Web pages that not only have the phrase Web site promotion in it, but also phrases like promotion of Web site. Then you would want the phrase Web site in close proximity to the word promotion in a page, and you would enter “Web site” NEAR promotion.

• **NOT** The word NOT entered before a keyword will find pages that do not contain the keyword. For example, suppose you’re interested in Web site promotion, but not in banner ads. Entering “Web site promotion” NOT “banner ads” will return pages with the phrase Web site promotion in it, but not the phrase banner ads. Note that a hyphen works in the same way as NOT; that is, entering “Web site promotion” NOT “banner ads” is equivalent to entering “Web site promotion” - “banner ads”.

• () Parentheses used around keywords can create a much more sophisticated search. If you enter “Web site promotion” AND (consultant OR software) the response will have the entries Web site promotion and consultant in it, or the entries Web site promotion and software in it.

**FUTURE TRENDS AND IMPLICATIONS**

Understand the power of a network and the effect of connecting people to the network. What makes the telephone network so useful is that virtually everyone is connected to it. It has become a standard by sheer number of the people using it. By Metcalfe’s Law (the same Robert Metcalfe who developed ethernet), the power or utility of a network does not grow as a simple linear function of the number of people in the network; rather, it grows as the square of the number of people connected. This growth is illustrated by the graph in Figure 4–8. As a consequence of Metcalfe’s Law, you want to use open, not proprietary, networking protocols. Make it easy to use your company network.

A critical mistake by management is to underestimate bandwidth availability when planning new business ventures. Do not kill a project just because it will require a great deal more bandwidth than you currently have. There is not a Moore’s Law for bandwidth, but there should be. Bandwidth is growing exponentially and becoming cheaper.

The networks discussed in this chapter (for example, the network in Figure 4–1) are client/server networks. There is a central server machine that serves the client machines on the network. When you use a browser on the Internet, you are working in client/server mode. The browser is downloading files from an HTTP server. This is currently the main paradigm in networking. However, peer-to-peer (P2P) networks in which each machine is both client and server are growing in popularity. An example of a P2P network is Gnutella. This is a file-sharing system on the Internet that allows you to directly search other computers for files and software. In order to do this you need special software that allows your machine to be both a client and a server. When using Gnutella you connect directly to other computers using an IP address (not DNS).
The browser has had an amazing effect on computing. By providing the user with a universal interface to the Internet, the browser is replacing the operating system shell as the user’s interface to the computer. In addition, the Internet is replacing the computer. The Internet is making Sun Microsystem’s saying, “the network is the computer” look more and more prophetic. Plan your business model around your employees’ and customers’ having low-cost client machines with high bandwidth connections to the Internet.

**Tech Talk**

**GigaPOP: Gigabit Point of Presence.** Very high-speed switching points that connect the backbone of the new Internet2.

The government sponsored ARPANET project was initially used for connecting university computers. It eventually morphed into the Internet and is currently overloaded. The Internet2 is a joint government, university, and industry project to accelerate the creation of tomorrow’s Internet. The Internet2 is being administered by UCAID (University Corporation for Advanced Internet Development). Currently, the network is designed to be a leading edge network for the national research community and to foster the development of new Internet technologies. Eventually, these services and technologies will be transferred to the broader community. The original intent of the Internet was to exchange text; the Internet2 is being developed to exchange multimedia data at high-speed. The Internet2 is based upon two very high-speed optical backbones: one is MCI Worldcom’s vBNS (very high-speed Bandwidth Network Service), the other is Abilene, also an optical network developed specifically for Internet2 by UCAID. Abilene is initially designed to run at 2.4 Gbps. Connections to the backbone are through GigaPOPs located throughout the country. Internet2 uses the IPv6 addressing scheme (see the Technical Appendix).