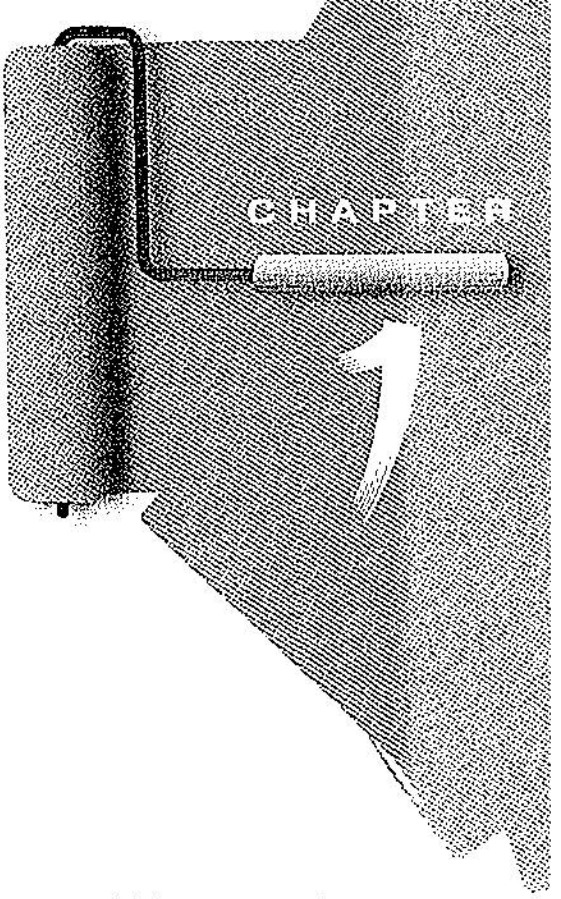


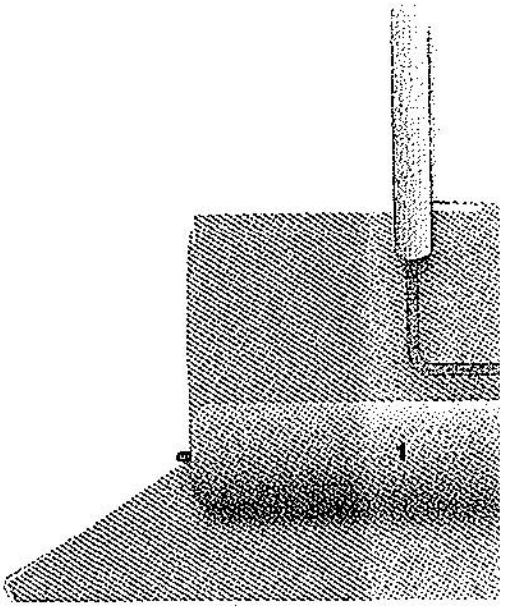
Introduction to the Internet and World Wide Web

Chapter Objectives

In this chapter, you will learn about ...

- The evolution of the Internet, Internet standards organizations, and the difference between the Internet, intranets, and extranets
- The beginning of the World Wide Web, ethical use of information on the Web, Web accessibility, and future Internet trends
- The client/server model, Internet protocols, networks, URLs and domain names, and markup languages

The Internet and the Web are parts of our daily lives. How did they begin? What networking protocols and programming languages work behind the scenes to display a Web page? This chapter provides an introduction to some of these topics and is a foundation for the information that Web developers need to know. Some material in this chapter may be a review from your life experience or earlier studies.



1.1 Evolution of the Internet

The Internet, the interconnected network of computer networks, seems to be everywhere today. It has become part of our lives. You can't watch television or listen to the radio without being urged to visit a Web site. Even newspapers have their place on the Net.

The Internet began as a network to connect computers at research facilities and universities. Messages in this network would travel to their destination by multiple routes or paths. This would allow the network to function even if parts of it were broken or destroyed. The message would be rerouted through a functioning portion of the network while traveling to its destination. This network was proposed to the Advanced Research Projects Agency (ARPA)—and the ARPAnet was born. Four computers (located at UCLA, Stanford Research Institute, University of California Santa Barbara, and the University of Utah) were connected by the end of 1969.

As time went on, other networks, such as the National Science Foundation's NSFnet, were created and connected with the ARPAnet. Use of this interconnected network, or Internet, was originally limited to government, research, and educational purposes. Even with this restriction, by 1989 there were over 100,000 hosts on the Internet. The ban on commercial use was lifted in 1991, and by the end of 1992 there were over 1 million hosts connected. Hobbes' Internet Timeline reports that as of 2006, there were over 439 million host computers on the Internet. The communications protocol that enabled all this to happen is the Transmission Control Protocol/Internet Protocol (TCP/IP), proposed by Vinton Cerf and Robert Kahn.

If you are interested in the history of the Internet, visit either of the following links for more information.

- A brief history of the Internet written by the people who created it can be found at http://www.isoc.org/internet/history/brief.shtml.
- For a classic treatment of the Internet's history, visit Hobbes' Internet Timeline at http://www.zakon.org/robert/internet/timeline/.

FAQ

How can I tell whether a Web page is a reliable source of information?

There are many Web sites—but which ones are reliable sources of information? When visiting Web sites to find information it is important not to take everything at face value.

First, evaluate the credibility of the Web site itself. Does it have its own domain name, such as http://mywebsite.com, or is it a free Web site consisting of just a folder of files hosted on a free Web server? The URL of a site hosted on a free Web server usually includes part of the free Web server's name and might begin with something such as http://mysite.tripod.com or http://www.angelfire.com/foldername/mysite. Information obtained from a Web site that has its own domain name will usually (but not always) be more reliable than information obtained from a free Web site.

Evaluate the type of domain name—is it a nonprofit organization (.org), a business (.com or .biz), an educational institution (.edu)? Businesses may provide information in a way that gives them an advantage, so be careful. Nonprofit organizations or schools will sometimes treat a subject more objectively.

Another item to look at is the date the Web page was created or last updated. Although some information is timeless, very often a Web page that has not been updated for several years is outdated and not the best source of information.

1.2 Internet, Intranets, and Extranets

The Internet is an interconnected network of computer networks that is globally accessible. When an organization needs the communication capabilities of the Internet but doesn't want its information to be accessible to everyone, either an intranet or extranet is appropriate.

An **intranet** is a private network that is contained within an organization or business. Its purpose is to share organizational information and resources among coworkers. When an intranet connects to the outside Internet, usually a gateway or firewall protects the intranet from unauthorized access.

An **extranet** is a private network that securely shares part of an organization's information or operations with external partners such as suppliers, vendors, and customers. Extranets can be used to exchange data, share information exclusively with business partners, and collaborate with other organizations. Privacy and security are important issues in extranet use. Digital certificates, encryption of messages, and virtual private networks (VPNs) are some technologies used to provide privacy and security for an extranet. Digital certificates and encryption used in e-commerce are discussed in Chapter 12.

The Evolution of the World Wide Web

Recall that the original Internet—the ARPAnet—began with four hosts. The number of host computers connected to the Internet grew each year. However, the communication was text-based and the information stored on computers connected to the Internet was not easy to obtain. Initially, the use of the Internet was limited to academics, researchers, students, and government employees. Even with these restrictions there were over 300,000 hosts in 1990.

Why did the Internet grow from 300,000 hosts in 1990 to over 109 million in just over a decade? In the early 1990s, the convergence of three events occurred to cause explosive growth of the Internet.

In 1991, the NSFnet removed the restriction on commercial use of the Internet, setting the stage for future electronic commerce. Businesses were now welcome on the Internet. However, while businesses were no longer banned, the Internet was still text-based and not easy to use. The next developments solved this issue.

While working at CERN, a research facility in Switzerland, Tim Berners-Lee envisioned a means of communication for scientists where they could easily "hyperlink" to another research paper or article and immediately view it. Berners-Lee created the **World Wide Web** to fulfill this need and in 1991 he posted the code in a newsgroup. This version of the World Wide Web used **Hypertext Transfer Protocol** (HTTP) to communicate between the client computer and the Web server, used **Hypertext Markup Language** (HTML) to format the documents, and was text-based.

In 1993, the first graphics-based Web browser, Mosaic, became available. Marc Andreessen and graduate students working at the National Center for Supercomputing Applications (NCSA) at the University of Illinois Urbana-Champaign developed Mosaic. Some individuals in this group later created another well-known Web browser—Netscape Navigator.

The combination of commercial use, HTTP, and a graphical user interface made the information on the Internet much easier to access. The World Wide Web—the graphical user interface to the information stored on computers connected to the Internet—had arrived!

1.3 Standards and Coordination

You are probably aware that no single person or group runs the entire Internet. Each separate network is managed individually. However, there are a number of groups that develop standards and guidelines. These groups are a driving force in the growth and evolution of the Internet.

The **Internet Society**, http://www.isoc.org, is a professional organization that provides leadership in issues related to the future of the Internet. The Internet Society is the organizational home for the groups responsible for Internet infrastructure standards, including the **Internet Engineering Task Force (IETF)** and the **Internet Architecture Board (IAB)**.

You can think of the IETF as the protocol engineering and development arm of the Internet. It is the principal body engaged in the development of new Internet standard specifications. The IETF is an open international community of network designers, operators, vendors, and researchers concerned with the evolution of Internet architecture and the smooth operation of the Internet. The actual technical work of the IETF is completed in its working groups. These working groups are organized into areas by topic, such as security and routing.

The IAB is responsible for defining the overall architecture of the Internet, by providing guidance and broad direction to the IETF. As a function of this purpose, the IAB is responsible for the publication of the **Request for Comments (RFC)** document series.

An RFC is a formal document from the IETF that is drafted by a committee and subsequently reviewed by interested parties. RFCs are available for online review at http://www.ietf.org/rfc.html. Some RFCs are informational in nature, while others are meant to become Internet standards. In the latter case, the final version of the RFC becomes a new standard. Future changes to the standard must be made through subsequent RFCs.

The **Internet Corporation for Assigned Numbers and Names (ICANN)**, http://www.icann.org, was created in 1998 and is a nonprofit organization. Its main function is to coordinate the assignment of Internet domain names, IP address numbers, protocol parameters, and protocol port numbers. Prior to 1998, the **Internet Assigned Numbers Authority (IANA)** coordinated these functions. IANA still performs certain functions under the guidance of ICANN and maintains a Web site at http://www.iana.org.

1.4 Standards and the World Wide Web Consortium

As with the Internet in general, no one person or group runs the World Wide Web. However, the **World Wide Web Consortium (W3C)**, http://www.w3.org, takes a proac-

tive role in developing recommendations and prototype technologies related to the Web. Four major areas that the W3C addresses are Web architecture, user interface, technology and society, and the Web Accessibility Initiative (WAI). In an effort to standardize Web technologies, the W3C produces specifications called recommendations.

The W3C Recommendations are created in working groups with input from many major corporations involved in building Web technologies. These recommendations are not rules; they are guidelines. Major software companies that build Web browsers, such as Microsoft and Netscape, do not always follow the W3C Recommendations. This makes life difficult for Web developers because not all browsers will display a Web page in exactly the same way.

The good news is that there is a convergence toward the W3C Recommendations in new versions of major browsers. There are even organized groups such as The Web Standards Project, http://webstandards.org, whose mission is to promote W3C Recommendations (often called Web standards) not only to the creators of browsers but also to Web developers and designers.

Accessibility and the Web

Focus on Accessibility

The Web Accessibility Initiative (WAI), http://www.w3.org/WAI/, is a major area of work by the W3C. Since the Web has become an integral part of daily life, there is a need for all individuals to be able to access it. According to Tim Berners-Lee at http://www.w3.org/WAI/, "The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect."

The Web can present barriers to individuals with visual, auditory, physical, and neurological disabilities. The WAI has developed recommendations for Web content developers, Web authoring tool developers, Web browser developers, and developers of other user agents to facilitate use of the Web by those with special needs. See the WAI's Web Content Accessibility Guidelines (WCAG) at http://www.w3.org/WAI/WCAG20/quickref/ for a list of these recommendations.

The Americans with Disabilities Act (ADA) of 1990 is a Federal civil rights law that prohibits discrimination against people with disabilities. The ADA requires that business, federal, and state services are accessible to individuals with disabilities. A 1996 Department of Justice ruling, http://www.usdoj.gov/crt/foia/cltr204.txt, indicated that ADA accessibility requirements apply to Internet resources.

Section 508 of the Federal Rehabilitation Act was amended in 1998 to require that U.S. government agencies give individuals with disabilities access to information technology that is comparable to the access available to others. This law requires developers creating information technology (including Web pages) for use by the federal government to provide for accessibility. The Federal IT Accessibility Initiative, http://www.section508.gov, provides accessibility requirement resources for information technology developers. In recent years, state governments have also begun to encourage and promote Web accessibility. The Illinois Web Accessibility Standards, http://www100.state.il.us/ito/iwas1_2.cfm, are an example of this trend.

Forward-thinking Web developers design with accessibility in mind. Providing access for visitors with visual, auditory, and other challenges should be an integral part of Web design rather than an afterthought.

A person with visual difficulties may not be able to use graphical navigation buttons and may use a screen reader device to provide an audible description of the Web page. By making a few simple changes, such as providing text descriptions for the images and perhaps providing a text navigation area at the bottom of the page, Web developers can make the page accessible. Often, providing for accessibility increases the usability of the Web site for all visitors. For example, text in high contrast to the background is easier for everyone to read. As this text introduces Web development and design techniques, corresponding Web accessibility and usability issues are discussed.

Ethical Use of Information on the Web


Focus on Ethics

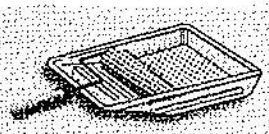
This wonderful technology called the World Wide Web provides us with information, graphics, and music—all virtually free (after you pay your Internet service provider, of course). Let's consider the following issues relating to the ethical use of this information:

- Is it acceptable to copy someone's graphic to use on your own Web site?
- Is it acceptable to copy someone's Web site design to use on your own site or on a client's site?
- Is it acceptable to copy an essay that appears on a Web page and use it or parts of it as your own writing?
- Is it acceptable to insult someone on your Web site or link to their site in a derogatory manner?

The answer to all these questions is no. Using someone's graphic without permission is the same as stealing it. In fact, if you link to it you are actually using up some of their bandwidth and may be costing them money. Copying the Web site design of another person or company is also a form of stealing. The Web site http://pirated-sites.com presents a somewhat quirky look at this issue. Any text or graphic on a Web site is automatically copyrighted in the United States whether or not a copyright symbol appears on the site. Insulting a person or company on your Web site or linking to them in a derogatory manner could be considered a form of defamation.

Issues like these, related to intellectual property, copyright, and freedom of speech are regularly discussed and decided in courts of law. Good Web etiquette requires that you ask permission before using others' work, give credit for what you use ("fair use" in the U.S. copyright law), and exercise your freedom of speech in a manner that is not harmful to others. The World Intellectual Property Organization (WIPO), http://wipo.int, is dedicated to protecting intellectual property rights internationally.

What if you'd like to retain ownership but make it easy for others to use or adapt your work? Creative Commons, http://creativecommons.org, is a nonprofit organization that provides free services that allow authors and artists to register a type of a copyright license called a Creative Commons license. There are several licenses to choose from—depending on the rights you wish to grant. The Creative Commons license informs others exactly what they can and cannot do with your creative work. See http://meyerweb.com/eric/tools/color-blend for a Web page licensed under a Creative Commons Attribution-ShareAlike 1.0 License with "Some Rights Reserved."



CHECKPOINT 1.1

1. Describe the difference between the Internet and an intranet.
2. Explain three events that contributed to the commercialization and exponential growth of the Internet.
3. Describe the difference between the Internet and the Web.

1.5 Network Overview

A network consists of two or more computers connected for the purpose of communicating and sharing resources. Common components of a network are shown in Figure 1.1 and include the following:

- Server computer(s)
- Client workstation computer(s)
- Shared devices such as printers
- Networking devices (hub) and the media that connect them

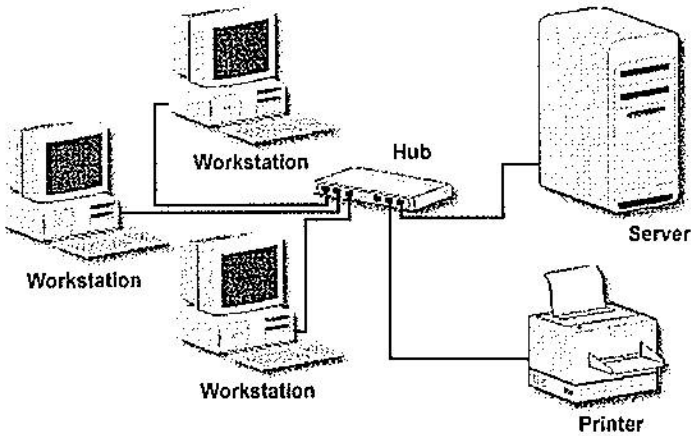


Figure 1.1
Common components of a network

The clients are the computers used by individuals, such as a PC on a desk. The server receives requests from client computers for resources such as files. Computers used as servers are usually kept in a protected, secure area and are only accessed by network administrators. Networking devices such as hubs and switches provide network connections for computers, and routers direct information from one network to another. The media connecting the clients, servers, peripherals, and networking devices may consist of copper cables, fiber optic cables, or wireless technologies.

Networks vary in scale. A Local Area Network (LAN) is usually confined to a single building or group of connected buildings. Your school computer lab may use a LAN. If

you work in an office, you probably use a computer connected to a LAN. Recently, many people have begun to set up LANs in their homes to share resources among computers. A **Metropolitan Area Network (MAN)** connects users with computer resources in a geographical area. It also can be used to connect two or more LANs. A **Wide Area Network (WAN)** is geographically dispersed and usually uses some form of public or commercial communications network. For example, an organization with offices on both the East and West Coasts of the United States probably uses a WAN to provide a link between the LANs at each of the offices. See Figure 1.2 for a diagram of this connectivity.

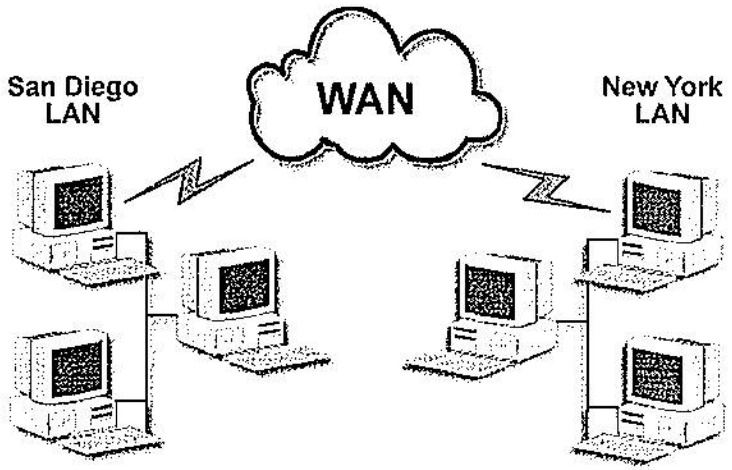


Figure 1.2
WAN connecting two LANs

A **backbone** is a high-capacity communication link that carries data gathered from smaller links that interconnect with it. On the Internet, a backbone is a set of paths that local or regional networks (MANs) connect to for long-distance interconnection. The Internet is a group of interconnected networks with very high-speed connectivity provided by the Internet backbones. Figure 1.3 shows a commercial backbone network map generated by http://www.caida.org/tools/visualization/mapnet/Backbones.

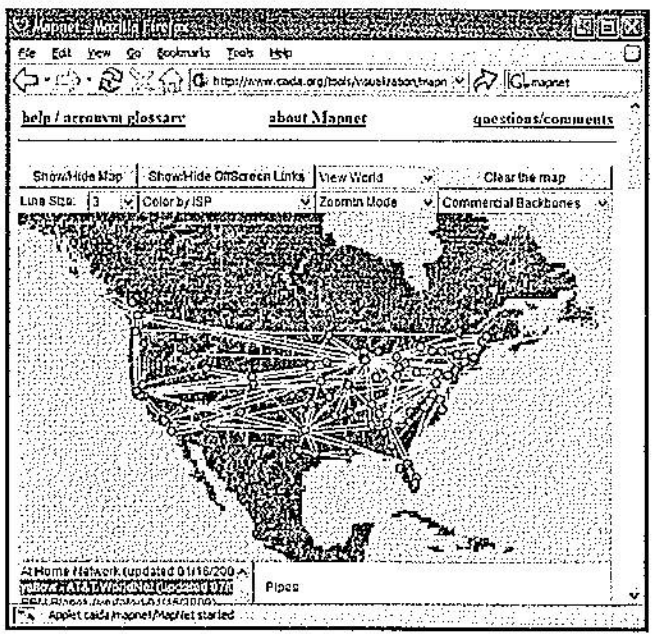


Figure 1.3
A commercial backbone network

Access points or junctions to the Internet backbone in major cities are called **Network Access Points (NAPs)**. Chicago, New York, and San Francisco are three key NAPs in the United States.

1.6 The Client/Server Model

The term client/server dates from the last millennium (the 1980s) and refers to personal computers joined by a network. Client/server can also describe a relationship between two computer programs—the client and the server. The client requests some type of service (such as a file or database access) from the server. The server fulfills the request and transmits the results to the client over a network. While both the client and the server programs can reside on the same computer, typically they run on different computers. It is common for a server to handle requests from multiple clients.

The Internet is a great example of client/server architecture at work. Consider the following scenario: An individual is at a computer using a Web browser client to access the Internet. The individual uses the Web browser to visit a Web site, let's say http://www.yahoo.com. The server is the Web server program running on the computer with an IP address that corresponds to yahoo.com. It is contacted, locates the Web page and related resources that were requested, and responds by sending them to the individual.

In short, here's how to distinguish between Web clients and Web servers:

Web Client

- Connected to the Internet when needed
- Usually runs Web browser (client) software such as Internet Explorer or Netscape
- Uses HTTP
- Requests Web pages from a server
- Receives Web pages and files from a server

Web Server

- Continually connected to the Internet
- Runs Web server software (such as Apache or Internet Information Server)
- Uses HTTP
- Receives a request for the Web page
- Responds to the request and transmits the status code, Web page, and associated files

When clients and servers exchange files, they often need to indicate the type of file that is being transferred; this is done through the use of a MIME type. **Multi-Purpose Internet Mail Extensions (MIME)** are rules that allow multimedia documents to be exchanged among many different computer systems. MIME was initially intended to extend the original Internet e-mail protocol, but it is also used by HTTP. MIME provides for the exchange of seven different media types on the Internet: audio, video, image, application, message, multipart, and text. MIME also uses subtypes to further describe the data. The MIME type of a Web page is text/html. MIME types of gif and jpeg images are image/gif and image/jpeg respectively.

A Web server determines the MIME type of a file before it is transmitted to the Web browser. The MIME type is sent along with the document. The Web browser uses the MIME type to determine how to display the document.

How does information get transferred from the Web server to the Web browser? Clients (such as Web browsers) and servers (such as a Web server) exchange information through the use of communication protocols such as HTTP, TCP, and IP.

1.7 Internet Protocols

Protocols are rules that describe how clients and servers communicate with each other over a network. There is no single protocol that makes the Internet and Web work—a number of protocols with specific functions are needed.

File Transfer Protocol (FTP)

File Transfer Protocol (FTP) is a set of rules that allow files to be exchanged between computers on the Internet. Unlike HTTP, which is used by Web browsers to request Web pages and their associated files in order to display a Web page, FTP is used simply to move files from one computer to another. Web developers commonly use FTP to transfer Web page files from their computers to Web servers. FTP is also commonly used to download programs and files from other servers to individual computers.

E-mail Protocols

Most of us take e-mail for granted, but there are two servers involved in its smooth functioning—an incoming mail server and an outgoing mail server. When you send e-mail to others, **Simple Mail Transfer Protocol (SMTP)** is used. When you receive e-mail, **Post Office Protocol** (POP; currently POP3) and **Internet Message Access Protocol (IMAP)** can be used.

Hypertext Transfer Protocol (HTTP)

HTTP is a set of rules for exchanging files such as text, graphic images, sound, video, and other multimedia files on the Web. Web browsers and Web servers usually use this protocol. When the user of a Web browser requests a file by typing a Web site address or clicking a hyperlink, the browser builds an HTTP request and sends it to the server. The Web server in the destination machine receives the request, does any necessary processing, and responds with the requested file and any associated media files.

Transmission Control Protocol/Internet Protocol (TCP/IP)

Transmission Control Protocol/Internet Protocol (TCP/IP) has been adopted as the official communication protocol of the Internet. TCP and IP have different functions that work together to ensure reliable communication over the Internet.

TCP. The purpose of TCP is to ensure the integrity of network communication. TCP starts by breaking files and messages into individual units called packets. These packets (see Figure 1.4) contain information such as the destination, source, sequence number, and checksum values used to verify the integrity of the data.

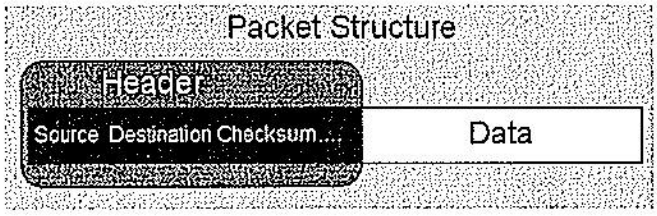


Figure 1.4
TCP packet

TCP is used together with IP to transmit files efficiently over the Internet. IP takes over after TCP creates the packets, using IP addressing to send each packet over the Internet using the best path at the particular time. When the destination address is reached, TCP verifies the integrity of each packet using the checksum, requests a resend if a packet is damaged, and reassembles the file or message from the multiple packets.

IP. Working in harmony with TCP, IP is a set of rules that controls how data is sent between computers on the Internet. IP routes a packet to the correct destination address. Once sent, the packet gets successively forwarded to the next closest router (a hardware device designed to move network traffic) until it reaches its destination.

Each device connected to the Internet has a unique numeric IP address. These addresses consist of a set of four groups of numbers, called octets. The current version of IP, IPv4, uses 32-bit (binary digit) addressing. This results in a decimal number in the format of xxx.xxx.xxx.xxx, where each xxx is a value from 0 to 255. The IP address may correspond to a domain name. The Domain Name System (DNS) associates these IP addresses with the text-based URLs and domain names you type into a Web browser address box (more on this later). For example, at the time this was written the IP address of Google was 64.233.187.99.

You can enter this number in the address text box in a Web browser (as shown in Figure 1.5), press Enter, and the Google home page will display. Of course, it's much

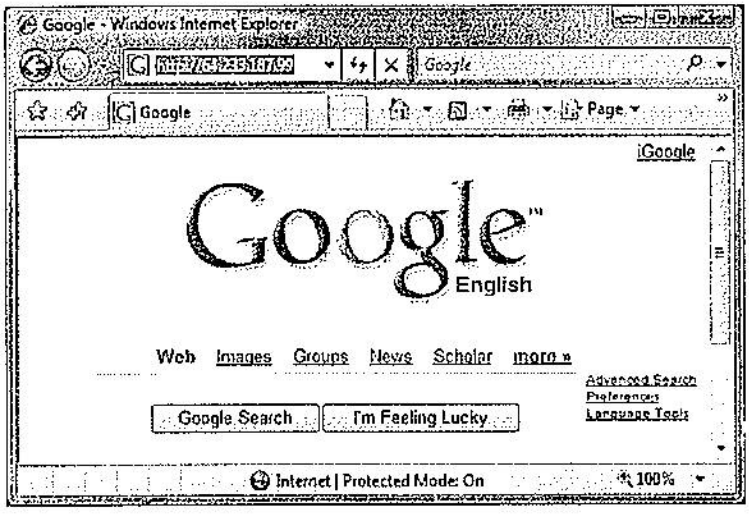


Figure 1.5
Entering an IP address in a Web browser

easier to type "google.com," which is why domain names such as google.com were created in the first place!

Since long strings of numbers are difficult for humans to remember, the Domain Name System was introduced as a way to associate text-based names with numeric IP addresses.

1.8 URLs and Domain Names

URLs

The Uniform Resource Locator (URL) represents the address of a resource that is available on the Internet. This resource could be for example, a Web page, a graphic file, or a Java applet. The URL consists of the protocol, the domain name, and the hierarchical location of the file on the Web server.

The URL http://www.webdevfoundations.net, shown in Figure 1.6, denotes the use of HTTP protocol and the Web server named www at the domain name of webdevfoundations.net. In this case, the root file (usually index.html or index.htm) will be displayed.

Figure 1.6
Parts of a URL

If the URL was of the form http://www.webdevfoundations.net/chapter1/links.html, as shown in Figure 1.7, it would denote the use of HTTP protocol and a Web server named www at the domain name of webdevfoundations.net. The resource to be displayed is the Web page named links.html in the chapter1 folder.

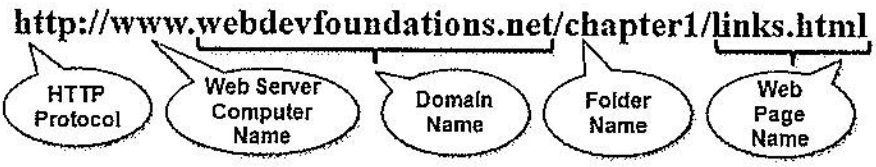


Figure 1.7
URL describing a file within a folder

FAQ What is a Universal Resource Locator?

Officially, URL stands for Uniform Resource Locator; but originally, Tim Berners-Lee (the inventor of the Web) envisioned a Universal Resource Locator. That is why some texts or Web pages refer to the URL in that manner. Read Tim Berners-Lee's book *Weaving the Web* for an interesting view of the creation of the Web.

Figure 1.8 shows a URL used to display files available for FTP download in the format of ftp://ftp.microsoft.com. This denotes the use of the FTP protocol, the server named ftp, and the domain name of microsoft.com.

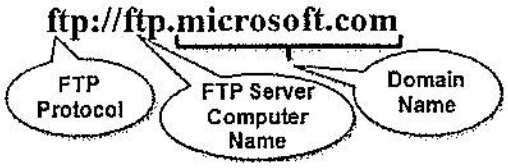


Figure 1.8
URL using FTP

Domain Names

A domain name locates an organization or other entity on the Internet. The purpose of the Domain Name System (DNS) is to divide the Internet into logical groups and understandable names by identifying the exact address and type of the organization. The DNS associates the text-based domain names with the unique numeric IP address assigned to a device.

Let's consider the domain name www.yahoo.com. The .com is the top-level domain name. The portion yahoo.com is the domain name that is registered to Yahoo! and is considered a second-level domain name. The www is the name of the Web server (sometimes called Web host server) at the yahoo.com domain. Taken all together, www.yahoo.com is considered to be a Fully-Qualified Domain Name (FQDN).

Top-Level Domain Names (TLDs). A top-level domain (TLD) identifies the rightmost part of the domain name. A TLD is either a generic top-level domain, such as com for commercial, or a country code top-level domain, such as fr for France. ICANN administers the generic top-level domains shown in Table 1.1.

Table 1.1 Top-level domains

Generic TLD	Used By
.aero	Air-transport industry
.asia	Pan-Asia and Asia Pacific community
.biz	Businesses
.cat	Catalan linguistic and cultural community
.com	Commercial entities
.coop	Cooperative
.edu	Restricted to accredited degree-granting institutions of higher education
.gov	Restricted to government use
.info	Unrestricted use
.int	International organization (rarely used)
.jobs	Human resource management community
.mil	Restricted to military use
.mobi	Corresponds to a .com Web site—the .mobi site is designed for easy access by mobile devices
.museum	Museums
.name	Individuals
.net	Entities associated with network support of the Internet, usually Internet service providers or telecommunication companies
.org	Nonprofit entities
.pro	Accountants, physicians, and lawyers
.tel	Contact information for individuals and businesses
.travel	Travel industry

The .com, .org, and .net TLD designations are currently used on the honor system, which means that an individual who owns a shoe store (not related to networking) can register shoes.net.

Country Code Top-Level Domain Names. Two-character country codes have also been assigned as top-level domain names. These were originally intended to be meaningful and relate the domain name country code to the geographical location of the individual or organization that registered the name. In practice, it is fairly easy to obtain a domain name with a country code TLD that is not local to the registrant. See http://register.com and many other domain name registration companies for examples. Table 1.2 lists some popular country codes used on the Web.

Table 1.2 Country codes

Country Code TLD	Country
.au	Australia
.de	Germany
.in	India
.jp	Japan
.nl	The Netherlands
.us	United States
.eu	European Union (a group of countries rather than a single country)

The IANA Web site at http://www.iana.org/cctld/cctld-whois.htm has a complete list. Domain names with country codes are often used for municipalities, schools, and community colleges in the United States. The domain name www.harper.cc.il.us denotes the United States, Illinois, community college, Harper, and the Web server named www as the site for William Rainey Harper College in Illinois.

The DNS associates domain names with IP addresses. The following happens each time a new URL is typed into a Web browser:

1. The DNS is accessed.
2. The corresponding IP address is obtained and returned to the Web browser.
3. The Web browser sends an HTTP request to the destination computer with the corresponding IP address.
4. The HTTP request is received by the Web server.
5. The necessary files are located and sent by HTTP responses to the Web browser.
6. The Web browser renders and displays the Web page and associated files.

The next time you wonder why it's taking so long to display a Web page, think about all of the processing that goes on behind the scenes.

1.9 Markup Languages

Markup languages consist of sets of directions that tell the browser software (and other user agents such as mobile phones) how to display and manage a Web document. These directions are usually called tags and perform functions such as displaying graphics, formatting text, and referencing hyperlinks.

Standard Generalized Markup Language (SGML)

Standard Generalized Markup Language (SGML) is a standard for specifying a markup language or tag set. SGML in itself is not a document language, but a description of how to specify one and create a document type definition (DTD). When Tim Berners-Lee created HTML, he used SGML to create the specification.

Hypertext Markup Language (HTML)

HTML is the set of markup symbols or codes placed in a file intended for display on a Web browser. The Web browser renders the code in the HTML file and displays the Web page document and associated files. The W3C (http://www.w3.org) sets the standards for HTML. Although the most recent version of HTML is called **XHTML** 1.1, this text uses XHTML 1.0 because it is less strict and is well-supported by popular browsers.

Extensible Markup Language (XML)

XML was developed by the W3C as a flexible method to create common information formats and share the format and the information on the Web. It is a text-based syntax designed to describe, deliver, and exchange structured information. It is not intended to replace HTML, but to extend the power of HTML by separating data from presentation. Using XML, developers can create whatever tags they need to describe their information.

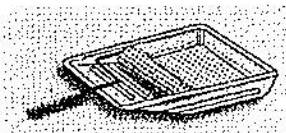
Extensible Hypertext Markup Language (XHTML)

XHTML was developed by the W3C to reformulate HTML 4.01 as an application of XML. It combines the formatting strengths of HTML 4.01 *and* the data structure and extensibility strengths of XML.

The primary advantages of XHTML include the ability to extend the language by creating new tags and the promise of increased platform interoperability as mobile devices are used more frequently to access the Web.

The Next Version of (X)HTML

As this was written, the W3C's HTML Working Group (HTML WG) was busy creating a draft recommendation for HTML 5—which is intended to be the next version of HTML 4 and XHTML 1. Check the blog on the textbook's Web site, http://webdevfoundations.net, for new developments.



CHECKPOINT 1.2

1. Describe the components of the client/server model as applied to the Internet.
2. Identify two protocols used on the Internet to convey information that use the Internet but do not use the Web.
3. Explain the similarities and differences between a URL and a domain name.

1.10 Internet and Web Trends

E-commerce, the buying and selling of goods on the Internet, is already an important part of the Web. According to a recent study by Jupiter Research (http://www.jupitermedia.com/corporate/releases/06.02.06-newjupresearch.html), revenue generated by e-commerce will continue to grow. By 2010, $144 million in online retail sales are projected. With 694 million people online worldwide (http://www.comscore.com/press/release.asp?press=849), that's quite a few potential shoppers!

As wireless Web access becomes more commonplace, e-commerce and Internet access not only will be regularly done from stationary computers but also from mobile devices—Palm Pilots, Pocket PCs, personal digital assistants (PDAs), cell phones, and Internet appliances we haven't even imagined yet.

As wireless access grows, so will the need for skilled technical workers. Expect to see a demand for network engineers familiar with wireless network technologies, such as Wireless Access Protocol (WAP), Bluetooth, and WiMAX. Web developers who are knowledgeable about markup languages such as Wireless Markup Language (WML), XML, and XHTML Basic will be sought as employees and consultants.

FAQ

What is an Internet appliance?

An Internet appliance is a device that is designed to access the Internet. It is different from PCs and Palm Pilots in that they are multipurpose devices, while an Internet appliance is a single-purpose device. As you read this, companies are working to develop new Internet-ready devices, ranging from digital cameras that instantly post photos on the Web, to Internet-connected wearable computers, Internet-enabled printers, and Internet-enabled point-of-sale (POS) terminals.

How do we keep track of all the devices (wireless and otherwise) that are connected to the Internet? You are already aware that each device on the Internet is assigned a unique number called an IP address. Currently, IPv4 is being used. Theoretically, this allows for at most 4 billion possible IP addresses (although many potential addresses are reserved for special uses). With the proliferation of mobile devices, even this many addresses may not be enough. IP Version 6 (IPv6) will provide a huge increase in the number of possible addresses and many technological advances.

FAQ

What is IPv6?

IPv6, Internet Protocol Version 6, is the most recent version of the Internet Protocol. IPv6 was designed as an evolutionary set of improvements to the current IPv4 and is backwardly compatible with it. Service providers and Internet users can update to IPv6 independently without having to coordinate with each other.

IPv6 provides for more Internet addresses because the IP address is lengthened from 32 bits to 128 bits. This means that there are potentially 2,128 unique IP addresses possible, or 340,282,366,920,938,463,463,347,607,431,768,211,456. (Now there will be enough IP addresses for everyone's PC, notebook, cell phone, pager, PDA, automobile, toaster, and so on!)

The development of the Internet2 is another effort in advancing Internet technology. The Internet2 consortium comprises more than one hundred U.S. universities in partnership with industry and government. Their mission is to develop and deploy advanced network applications and technologies, focusing on applications related to learning and research such as telemedicine, digital libraries, and virtual laboratories. Visit the Internet2 Web site at http://www.internet2.edu for information on this initiative.

Another area to watch is Web services. A **Web service** is a self-describing, self-contained application that provides some business functionality through an Internet connection. For example, an organization could create a Web service to facilitate information exchange with its partners or vendors. The Universal Discovery, Description, and Integration (UDDI) standard, http://uddi.xml.org, is backed by a number of technology companies, including IBM, Microsoft, and Sun Microsystems. Essentially, UDDI provides a method of describing a service, invoking a service, and locating available services. Microsoft's .NET platform supports Web services. Microsoft and IBM jointly developed Web Services Description Language (WSDL) to facilitate the use of Web services.

While the Web service initiative is driven by large corporations, the trend of keeping a Web log, or blog, has been driven by individuals as a forum for personal expression. A **blog** is a journal that is available on the Web—it's a frequently updated page with a chronological list of ideas and links. Blog topics range from political journals to technical information to personal diaries. Blogs can focus on one subject or range across a diverse group of topics—it's up to the person, called a blogger, who creates and maintains the blog. Bloggers usually update their blogs daily with easy-to-use software designed to allow people with little or no technical background to update and maintain the blog. The PEW Internet & American Life Project (http://www.pewinternet.org) reports that 39 percent of American adults read blogs daily and about 8 percent of American adults keep a blog. Many blogs are hosted at blog communities such as http://blogspot.com, http://www.diaryland.com, or http://www.xanga.com. Others are hosted at individual Web sites, such as the blog kept by the CSS expert Eric Meyer at http://meyerweb.com. Businesses have noted the value of blogs as communication and customer relationship tools. Companies such as IBM, http://www.ibm.com/developerworks/blogs/, and Adobe, http://weblogs.macromedia.com/mxna/, utilize blogs in this manner.

Really Simple Syndication or **Rich Site Summary (RSS)** is commonly used to create newsfeeds from blog postings and other Web sites. The RSS feeds contain a summary of new items posted to the site. The URL to the RSS feed is usually indicated by the letters XML or RSS in white text within an orange rectangle. A **newsreader** is needed to access the information. Some browsers, such as Firefox, Safari, and Internet Explorer 7 can display RSS feeds. Commercial and shareware newsreader applications are also available. The newsreader will poll the feed URL at intervals and display the new headlines when requested. RSS provides Web site developers with a method to push new content to interested parties and (hopefully) generate return visits to the site.

Podcasts are audio files on the Web—they may take the format of an audio blog, radio show, or interview. Podcasts are typically delivered by an RSS feed but can also be made available by recording an MP3 file and providing a link on a Web page. These files can be saved to your computer or to an MP3 player (such as an iPod) for later listening. Forrester Research (http://forrester.com/Research/Document/Excerpt/0,7211,36428,00.html) predicts that by 2010 more than 12.3 million households will use podcasting.

A wiki is a Web site that can be updated immediately at any time by visitors using a simple form on a Web page. Some wikis are intended for a small group of people, such as the members of an organization. The most powerful wiki is Wikipedia, http://wikipedia.org, an online encyclopedia, which can be updated by anyone at anytime. This is a form of social software in action—visitors sharing their collective knowledge to create a resource freely used by all. While there have been isolated incidents of practical jokes and occasionally inaccurate information posted at Wikipedia, the information and resource links are a good starting point when exploring a topic.

Blogs and wikis have provided Web visitors new methods to utilize and interact with Web sites and other people—referred to as social computing or social networking. Flickr (http://www.flickr.com/) and del.icio.us (http://del.icio.us/) are two social software sites that provide information-sharing opportunities. Flickr, a photo sharing site, calls itself the "best way to store, search, sort, and share your photos." Recently acquired by Yahoo!, del.icio.us is a collection of favorite sites—allowing registered users to post lists of favorites, share their favorites with others, and discover new sites. Wikipedia, Flickr, and del.icio.us are examples of what is called **Web 2.0**. While a consensus on the definition of Web 2.0 still needs to be reached, think of it as the next step in the transition of the Web from isolated static Web sites to a platform that utilizes technology to provide rich interfaces and social networking opportunities. Read Tim O'Reilly's informative Web 2.0 essay at http://oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html for more information on this developing topic.

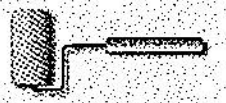
The single future trend that you can expect to remain the same is the trend of constant change. Internet- and Web-related technologies are in a constant state of development and improvement. If constant change and the need to learn something new excites you, Web development is a fascinating field. The skills and knowledge you gain in this book should provide a solid foundation for your future learning.

FAQ

What is the next big thing on the Web?

The Web is changing by the minute. Check the textbook's companion Web site at http://webdevfoundations.net for a blog that will help you stay current about Web trends.

CHAPTER SUMMARY



This chapter provided a brief overview of Internet, Web, and introductory networking concepts. Much of this information may be familiar to you. Visit the textbook Web site at http://www.webdevfoundations.net for the links listed in this chapter and for updated information.

Key Terms

accessibility
backbone
blog
client/server
clients
domain name
Domain Name System (DNS)
extranet
File Transfer Protocol (FTP)
Fully-Qualified Domain Name (FQDN)
Hypertext Markup Language (HTML)
Hypertext Transfer Protocol (HTTP)
Internet
Internet Architecture Board (IAB)
Internet Assigned Numbers Authority (IANA)
Internet Corporation for Assigned Numbers and Names (ICANN)
Internet Engineering Task Force (IETF)
Internet Message Access Protocol (IMAP)
Internet Society
intranet
IP
IP address
IP Version 6 (IPv6)
Local Area Network (LAN)
markup languages
media
Metropolitan Area Network (MAN)
Multi-Purpose Internet Mail Extensions (MIME)
network
Network Access Points (NAPs)
newsreader
packets
podcasting
Post Office Protocol (POP3)
protocols
Really Simple Syndication or Rich Site Summary (RSS)
Request for Comments (RFC)
server
Simple Mail Transfer Protocol (SMTP)
social computing
Standard Generalized Markup Language (SGML)
TCP
Top-level domain (TLD)
Transmission Control Protocol/Internet Protocol (TCP/IP)
Uniform Resource Locator (URL)
Web 2.0
Web Accessibility Initiative (WAI)
Web host server
Web service
Wide Area Network (WAN)
wiki
World Intellectual Property Organization (WIPO)
World Wide Web
World Wide Web Consortium (W3C)
XHTML
XML

Review Questions

Multiple Choice

1. Of the following organizations, which one studies technical problems of the Internet and proposes solutions?

a. Assigned Numbers Authority (IANA)
b. Engineering Task Force (IETF)
c. Corporation for Assigned Numbers and Names (ICANN)
d. World Wide Web Consortium (W3C)

2. Which of the following is a network that covers a small area, such as a group of buildings or campus?

a. LAN
b. WAN
c. Internet
d. WWW

3. At which of the following organizations did individuals develop the World Wide Web?
   a. CERN
   b. NCSA
   c. NSF
   d. ARPA

4. What is a unique text-based Internet address corresponding to a computer's unique numeric IP address called?
   a. IP address
   b. domain name
   c. URL
   d. user name

5. New Top-Level Domains (TLDs) are coordinated by which of the following?
   a. ICANN
   b. no one, because anyone can add a TLD to the DNS
   c. W3C
   d. TCP

True or False

6. ____ Markup languages contain sets of directions that tell the browser software (and other user-agents such as cell phones) how to display and manage a Web document.

7. ____ The World Wide Web was developed to allow companies to advertise over the Internet.

8. ____ A numerical Internet address used to identify computers is called an IP address.

9. ____ A domain name that ends in .com indicates that it is a computer company.

Fill in the Blank

10. _____ combines the formatting strengths of HTML 4.0 and the data structure and extensibility strengths of XML.

11. A standard language used for specifying a markup language or tag set is _____.

12. _____ is the set of markup symbols or codes placed in a file intended for display on a Web browser.

13. A language using a text-based syntax intended to extend the power of HTML by separating data from presentation is called _____.

14. Access points or junctions to the Internet backbone are called _____.

15. The purpose of _____ is to ensure the integrity of the communication.

Hands-On Exercise

1. Create a blog to document your learning experiences as you study Web development. Visit one of the many sites that offer free blogs, such as http://blogspot.com, http://www.diaryland.com, or http://www.xanga.com. Follow their instructions to establish your own blog. Your blog could be a place to note Web sites that you find useful or interesting. You might report on sites that contain useful Web design resources. You might describe sites that have interesting features, such as compelling graphics or easy to use navigation. Write a few sentences about the site that you find intriguing. After you begin to develop your own sites, you could include the URLs and reasons for your design decisions. Share this blog with your fellow students and friends. Display your page in a browser and print the page. Hand in the printout to your instructor.

Web Research

1. The World Wide Web Consortium creates standards for the Web. Visit its site at http://www.w3c.org and then answer the following questions:
   a. How did the W3C get started?
   b. Who can join the W3C? What does it cost to join?
   c. The W3C home page lists a number of technologies. Choose one that interests you, click its link, and read the associated pages. List three facts or issues you discover.
2. The Internet Society takes an active leadership role in issues related to the Internet. Visit its site at http://www.isoc.org and answer the following questions:
   a. Why was the Internet Society created?
   b. Determine the local chapter closest to you. Visit its Web site. List the Web site URL and an activity or service that the chapter provides.
   c. How can you join the Internet Society? What does it cost to join? Would you recommend that a beginning Web developer join the Internet Society? Why or why not?
3. The World Organization of Webmasters (WOW) is a professional association dedicated to the support of individuals and organizations that create and manage Web sites. Visit its site at http://www.joinwow.org and answer the following questions:
   a. How can you join WOW? What does it cost to join?
   b. List one of the events that WOW participates in. Would you like to attend this event? Why or why not?
   c. List three ways that WOW can help you in your future career as a Web developer.

Focus on Web Design

1. Visit a Web site referenced in this chapter that interests you. Print the home page or one other pertinent page from the site. Write a one-page summary and your reaction to the site. Address the following topics:
   a. What is the purpose of the site?
   b. Who is the intended audience?
   c. Do you think that the site reaches its intended audience? Why or why not?
   d. Is the site useful to you? Why or why not?
   e. List one interesting fact or issue that this site addresses.
   f. Would you encourage others to visit this site?
   g. How could this site be improved?