TECHNOLOGY GUIDE

5

The Internet and the Web

T5.1 WHAT IS THE INTERNET?[1]

The **Internet ("the Net")** is a network that connects hundreds of thousands of internal organizational computer networks worldwide. Examples of internal organizational computer networks are a university computer system, the computer system of a corporation such as IBM or McDonald's, a hospital computer system, or a system used by a small business across the street from you. Participating computer systems, called **nodes,** include PCs, local area networks, database(s), and mainframes. A node may include several networks of an organization, possibly connected by a wide area network. The Internet connects to hundreds of thousands of computer networks in more than 200 countries so that people can access data in other organizations, and can communicate and collaborate around the globe, quickly and inexpensively. Thus, the Internet has become a necessity in the conduct of modern business.

Brief History

The Internet grew out of an experimental project of the Advanced Research Project Agency (ARPA) of the U.S. Department of Defense. The project was initiated in 1969 as ARPAnet to test the feasibility of a wide area computer network over which researchers, educators, military personnel, and government agencies could share data, exchange messages, and transfer files.

From four nodes at its initiation, the Internet has grown to millions of nodes today. The major growth occurred after commercial organizations were allowed to join ARPAnet, renamed the Internet in 1993. Each node may connect many individual users, sometimes tens of thousands of them. There are over 500 million Internet users today.

The computers and organizational nodes on the Internet can be of different types and makes. These are connected to each other by data communication lines of different speeds. The main network that links the nodes is referred to as the **backbone,** a fiber-optic network currently operated mainly by telecommunications companies such as MCI.

No central agency manages the Internet. The cost of its operation is shared among hundreds of thousands of nodes, and therefore the cost for any one organization is small. Organizations must pay a small fee if they wish to register their name, and they need to have their own hardware and software for the operation of their internal networks. The organizations are obliged to move any data or information that enters their organizational network, regardless of its source, to its internal destination, at no charge to the senders. The senders, of course, pay the telephone bills for using the backbone or regular telephone lines.

Next Generation Internet Services

Very-high-speed Backbone Network Service (vBNS) is a high-speed network designed to support the academic Internet2 (see below) and the government-sponsored Next-Generation Internet (NGI) initiative. The vBNS was first implemented as an OC-3 (155-Mbps) backbone but has been upgraded to OC-12 (622 Mbps). The goal is to increase the vBNS backbone to OC-48 (2.4 Gbps).

INTERNET2. **Internet2** is a collaborative effort by over 200 U.S. universities, working with over 70 companies and more than 40 affiliated organizations and government, to develop advanced Internet technologies and applications vital to

[1]See Internet 101 at *Internet-101.com, Internet101.org,* and *stimulus.com.*

support the research and education goals of higher education. Internet2 was announced in October 1996 with 34 universities, in October 1997 the University Corporation for Advanced Internet Development (UCAID) was formed to administer the Internet2 project. The primary goals of Internet2 are to create a leading-edge network capability for the national research community; enable revolutionary Internet applications; and ensure the rapid transfer of new network services and applications to the broader Internet community.

It focuses on advanced network applications, new network capabilities, middleware, and high-performance networks.

Internet2 will use part of vBNS as a backbone for providing high-speed connectivity to many of its member universities. The communications infrastructure designed to support Internet2 consists of 10 Mbs (nonshared) to the desktop, a 500 Mbs campus wide backbone, and at least 155 Mbs connectivity to GigaPoP that is a proposed national high-speed network. It supports Quality-of-Service (QoS) technology, which allows two hosts to establish a connection with a guaranteed bandwidth that is necessary for live video applications. It is expected to be 100 to 1000 times faster than the current Internet infrastructure. For details, see *internet2.edu*.

NEXT-GENERATION INTERNET. The **Next-Generation Internet (NGI)** initiative is a multiagency (with the NSF, NIE, etc.), U.S. federal government research and development program that is developing advanced networking technologies and revolutionary applications that require advanced networking. The broad goals of the NGI initiative are to research and develop advanced end-to-end networking technologies, focusing primarily on reliability, robustness, security, quality of service guarantees for multicasting and video, and bandwidth allocation. In other words, the NGI initiative aims to create an Internet that is fast, always on, everywhere, natural, intelligent, easy, and trusted. To this end, NGI is building two test beds, one at speeds at least 100 times faster than today's Internet, and the other 1000 times faster. The university-led Internet2 and the federally-led Next Generation Internet (NGI) are parallel and complementary initiatives. For details, see *ngi.gov*.

INTERNET GOVERNANCE. There is no CEO or board of directors of the Internet. However, the following organizations govern the various operations of the Internet:

- *Internet Society:* Enables global cooperation regarding the Internet.
- *World Wide Web Consortium (W3C):* Develops standards for the Web and promotes new Web technologies.
- *Internet Engineering Task Force (IETF):* Develops technical Internet standards and protocols.
- *Internet Architecture Board (IAB):* Oversees the Internet standards process.
- *Internet Engineering Steering Group (IESG).* Approves technical Internet standards.

T5.2 BASIC CHARACTERISTICS AND CAPABILITIES OF THE INTERNET

Accessing the Internet

There are several ways to access the Internet. From your place of work or your university, you can access an Internet-connected file server on your organization's LAN. A campus or company backbone connects all the various LANs and

servers in the organization to the Internet. You can also log onto the Internet from your home or on the road; wireless connections are also possible. Using a modem you can connect to your campus or company network from any place that offers a dial tone. Connection to the Internet is also available through commercial providers such as America Online (AOL), for which you pay a fee. Many telephone providers, such as AT&T and MCI, also sell Internet access, as do computer companies such as Microsoft. Such companies are called Internet service providers (ISPs). Following are the main Internet connection methods.

CONNECTING VIA LAN SERVER. This approach requires the user's computer to have specialized software called a *communications stack,* which provides a set of communications protocols that perform the complete functions of the seven layers of the OSI communications model. LAN servers are typically connected to the Internet at 56 Kbps or faster. This type of connection is expensive, but the cost can be spread over multiple LAN users.

CONNECTING VIA SERIAL LINE INTERNET PROTOCOL/POINT-TO-POINT PROTOCOL (SLIP/PPP). This approach requires that users have a modem and specialized software that allows them to dial into a SLIP/PPP server through a service provider at a cost of approximately $30 per month or less. This type of connection is advantageous, for example, for employees working at home who need to access the Internet or their own company's intranet.

CONNECTING VIA AN ONLINE SERVICE. This approach requires a modem, standard communications software, and an online information service account with an Internet service provider. The costs include the online service fee, a per-hour connect charge, and, where applicable, e-mail service charges.

One of the first **online services (OLS),** CompuServe, was launched in 1979. Prodigy and AOL (America Online) followed with offerings targeted toward home computer users. Services include, for example, access to specialized news sources (business, sports, etc.), market information, maps, and weather. This focus contrasted with the Internet's original orientation, which was geared to the academic and research community, and that of online research services such as Lexis-Nexis (a division of Reed Elsevier), which targets large corporations.

Originally, **Internet service providers (ISPs)** were established to provide connectivity, not content. Many ISPs now offering dial-in Internet access to consumers initially were set up to provide dedicated Internet connections to educational and commercial organizations. Others (such as NetCom) began by supplying dial-in access to Internet-connected time-sharing systems. UUNet started by providing dial-up connections for routing of e-mail and Usenet news (discussion groups devoted to specialized areas of interest) between non-Internet-connected sites.

Thanks to the growth of Internet usage (as well as deregulation of the telecommunications industry), the number of ISPs has grown rapidly. Gartner Group estimates that there were approximately 13,000 ISPs worldwide at the end of 2001. Leading ISPs, which also now provide online content as well, include America Online (*aoltimewarner.com*), MSN (*msn.com*), United Online (*unitedonline.net*), Earthlink (*earthlink.com*), and Prodigy (*prodigy.com*).

NSP, NAP, ROUTERS, AND PACKETS. The Internet is composed of thousands of interconnected networks, managed by multiple Internet service providers (ISPs) who control the routing of packets and their performance. Routing of packets between such networks is accomplished through the Border Gateway Protocol, which tells a receiving network how to reach customer addresses connected to the sending network. Inside this Internet connection, there are two features: *network service providers (NSPs)* and *network access points (NAPs)*.

NSPs, commonly known as tier-1 carriers, are responsible for the backbone's traffic connected by a high-speed connection (called *peering circuits*). Most ISPs operate regional networks and rely on the backbones of the NSPs to provide the required long-distance connectivity. NSPs can serve their ISP customers either through dedicated access circuits or through interconnections at public NAPs.

NAPs, which were created by the U.S. National Science Foundation (NSF), are the access points at which service providers can meet and exchange packets traffic. The NAP operator provides switching equipment to interconnect service providers present at the NAP. For more information on NSPs and NAPs, see *mitretek.org/pubs/sigma/summer2003/1.pdf.*

THE TV AS A CHANNEL TO THE INTERNET. For people who do not own a computer, television can also provide access to the Internet. Using a television set, a special connection device, and a telephone connection, viewers can surf the World Wide Web in their living rooms. The concept was pioneered by Web TV Networks, which manufactures an add-on device for the television, but several other companies are also developing similar devices. At the present time the quality of the Web image on a television screen is poor compared to what you see on a computer screen. It may take a few years before this situation improves.

An alternative is a combined PC and TV. This is a PC with a TV feature. It takes advantage of the convergence of telecommunications, television, and computer to deliver video and other multimedia content over the Internet at low cost. Technologies are now available to enable PCs to decode TV signals, and to receive full-motion video over the Internet. LCD TV monitors also enable access to the Internet through a $100 set-top box.

OTHER ACCESSES. There have been several attempts to make access to the Net cheaper, faster, and/or easier. Special Internet terminals called "Internet Lite" or *network computers (NCs)* (see Technology Guide 1) are used by some companies. Also, **Internet kiosks** are terminals placed in public places like libraries and coffee houses, and even in convenience stores in some countries, for use by those who do not have computers of their own. (For examples, see Payphone Cyberbooth at *atcominfo.com*.) Accessing the Internet from cell phones and pagers is growing very rapidly.

BROADBAND AND DSL. **Broadband** refers to telecommunication that provides multiple channels of data over a single communications medium, typically using some form of frequency or wave division multiplexing. There are several broadband technologies. These include Global System for Mobile communication (GSM), High-Speed Circuit-Switched Data (HSCSD), POTS, frame relay, GPRS (General Packet Radio Services), Integrated Services Digital Network (ISDN), xDSL, AppleTalk, Enhanced Data GSM Environment (EDGE), Satellite, and Fiber-Distributed Data Interface (FDDI).

There are several methods by which users can connect to the Internet—namely, through a dial-up modem, through a local area network, through a cable modem, or through a digital subscriber line (DSL) connection. **DSL** is a very high speed connection that uses the same wires as a regular telephone line. At the customer's location there is a DSL transceiver, which may also provide other services. At the DSL service provider's location, there is a DSL Access Multiplexer (DSLAM) to receive customer connections.

There are advantages and disadvantages of using DSL connection method. Advantages are:

- You can leave your Internet connection open and still use the phone line for voice calls.
- The speed is much higher than a regular dial-up modem (1.5 or even 10 Mbps versus 56 Kbps)
- DSL does not necessarily require new wiring; it can use the phone line you already have.
- The company that offers DSL will usually provide the modem as part of the installation.
- You can use a wireless system with the DSL.

Disadvantages and limitations of DSL are:

- A DSL connection works better when you are closer to the provider's central office. Otherwise, the connection may have interruptions, or it will slow down.
- The connection is faster for receiving data (i.e., downstream, usually at up to 10 Mbps) than it is for sending data (i.e., upstream, usually at 640 Kbps) over the Internet.
- The service is not available everywhere.

For further details, see *cis.state.mi.us/mpsc/comm/broadband/mediums/tech_def.htm.*

The TCP/IP Protocol

Procedures and rules for transferring data across the Internet are called telecommunications **protocols.** The original participants of the Internet, mostly universities, used *TCP/IP* (described in Technology Guide 4), which is now the **Internet Protocol (IP).** The information that passes through the Internet is divided into small portions, called **packets,** whose creation and transmission are governed by TCP/IP to provide for more consistent delivery and control.

One member of the TCP/IP family of protocols is **Telnet,** a service that allows users to connect to computers other than their own and interactively search for files, text, software, and so forth. This enables researchers who commute to conduct their research without interruption. You can also use Telnet to connect to your organization's computer system when you are traveling, so you can check your e-mail, send messages, or conduct other activities.

STANDARDS AND PROTOCOLS. *Standards* are a set of rules or specifications for the design or operation of a computing device. There are *proprietary standards,* which are those developed and promulgated by companies in the hope of assuring or increasing their market share, and *open standards,* which are published and available for use by anyone. Either type may become a *de facto* standard, a

set of rules or specifications that comes into such widespread use in the marketplace that it becomes normative, or a *de jure* standard, a standard given the endorsement of an official standards body such as the International Organization for Standardization (ISO).

Protocols are standard sets of rules that govern network communications functions by describing both the format that a message must take and the way in which messages are exchanged between computers.

Quite often the word *standard* is used to refer to so-called industry standards, i.e., relatively stable principles and practices applied by leading manufacturers. This is a very vague concept. Between official standards and "industry standards," there is a wide range of "standards" that have been agreed on in some forum. For more information, see *collectionscanada.ca/9/1/p1-253-e.html.*

IPv4 AND IPv6. *IPv4* is version 4 of the Internet Protocol (IP). It was the first version of the Internet Protocol to be widely deployed, and it forms the basis for most of the current Internet. Nearly 20 years old, IPv4 has been remarkably resilient in spite of its age, but it is beginning to have problems. Most significant of these problems is the growing shortage of IPv4 addresses, which are needed by all new machines added to the Internet.

IPv6 is version 6 of the Internet Protocol. It is the next-generation protocol designed by the Internet Engineering Task Force (IETF) to replace the current version (IPv4). IPv6 fixes a number of problems in IPv4, such as the limited number of available IPv4 addresses (128-bit instead of 32-bit). It also adds many improvements to IPv4 in areas such as routing (hierarchical routing, mobility) and network autoconfiguration to increase the security during the data transfer (IPSec: authentication and encryption). IPv6 is expected to gradually replace IPv4, with the two coexisting for a number of years during a transition period. For details, see *ipv6.org/.*

Internet Resources

Some information sources on the Internet are free. For example, using Telnet, you can access libraries and conduct research. Alternatively, the information may be owned by a commercial online service. As described above, an *online service* sells access to large (usually nationwide) databases. Such services can be used to add external data to a corporate information system in a timely manner and at a reasonable cost. Several thousand services are currently available. Representative services are Dow Jones Information Service, Mead Data Central, and Compustat.

Several magazines deal exclusively or mainly with the Internet, such as *Internet Business Advantage* (*zdjournals.com*), *Internet World* (*iw.com*), *NetGuide* (*netguide.com.au*), *IEEE Internet*, and *Wired*. There are also many online magazines such as *TechWeb* (*techweb.com*). Finally, there are several associations and societies through which members participate in activities related to the Internet, such as the *Internet Society* (*isoc.org*) and the *Electronic Frontier Foundation* (*eff.org*).

ADDRESSES ON THE INTERNET. Each computer on the Internet has an assigned address, called the **IP (Internet Protocol) address,** that uniquely identifies and distinguishes it from all other computers. The IP numbers have four parts, separated by dots. For example, the IP address of one computer might be 135.62.128.91.

DOMAIN NAMES. Most computers also have names, which are easier for people to remember than IP addresses. These names are derived from a naming system called the **domain name system (DNS).** Network Solutions Inc. (NSI) (now part of Verisign, Inc.) once had the exclusive authority to register addresses using .com,.net, or .org domain names. The company's contract ended in October 1998, as the U.S. government moved to turn management of the Web's address system over to the broader private sector. Currently, in November 2004, 376 companies, called registrars, are accredited to register domain names from the Internet Corporation for Assigned Names and Numbers (ICANN) (*internic.net*).

All domain names are administered by the Internet Corporation for Assigned Names and Numbers (ICANN). There are two types of top-level domains (TLDs): generic (gTLDs) and country code (ccTLDs). *Generic domains* were created for use by anyone on the Internet; *country code domains* were created for use by individual countries (*iana.org/cctld/cctld-whois.htm*). There is a special top-level domain (*.arpa*) for Internet infrastructure.

Domain names consist of multiple parts, separated by dots, and are translated from right to left. For example, consider the name *software.ibm.com*. The rightmost part of an Internet name, in this case "com," is its top-level specification, or the **zone.** The abbreviation "com" indicates that this is a commercial site. The 15 top-level specifications are listed below.

ZONE	USED FOR	ZONE	USED FOR
com	commercial sites	aero	air-transport industry
edu	educational sites	biz	businesses
mil	military sites	coop	cooperatives
gov	government sites	museum	museums
net	networking organizations	name	registration by individuals
org	organizations	pro	accountants, lawyers, physicians (professionals)
store	businesses offering goods for purchase	int	registering organizations
info	information service providers		

Finishing our example, "ibm" is the name of the company (IBM), and "software" is the name of the particular machine (computer) within the company to which the message is being sent. The rightmost two letters in a domain name, if present, represent the country of the Web site. For example, "us" stands for the United States, "de" for Germany, "it" for Italy, and "ru" for Russia. In the United States, the "us" is omitted. Many companies in other countries use ".com" without the country name also.

As of July 17, 2004, there were 39,363,493 domain names. Distribution of top-level domains was as follows:

TLD type	.com	.net	.org	.biz	.info	.edu
Total no.	29,321,771	4,791,592	3,029,358	1,001,991	1,211,384	7397

Source: zooknic.com/Domains/counts.html.

Most domain names have their own commercial value. This commercial value has led to the practice of *cybersquatting*—buying a potentially coveted domain name and hoping someone wants that name enough to pay for it. The practice of cybersquatting grew out of NSI's early policy of registering domain names on a first-come, first-served basis. This policy resulted in many companies or individuals registering a domain name associated with an established firm before the established firm did. The policy resulted in disputed names and legal actions. In response, Congress passed the Anti-Cybersquatting Consumer Protection Act in November 2000.

DNSs. A set of servers called **domain name servers (DNSs)** maps the human-readable names to the IP addresses. These servers are simple databases that map names to IP addresses, and they are distributed all over the Internet. Most individual companies, ISPs, and universities maintain small domain name servers to map host names to IP addresses. Inside the DNS system, there is a big database that has billions of IP addresses currently in use and the DNS requests made every day.

URLs. A **uniform resource locator (URL)** indicates the location (or address) of a Web site you want to visit. The address consists of several parts. For example, Yahoo's address—*http://www.yahoo.com*—tells us that to access the site, we use a **hypertext transfer protocol (http),** which is a procedure for retrieving hypermedia and hypertext documents. The next part of the URL, "www," indicates that you are making a request of a Web server. "Yahoo" is the name of the company, and ".com" a first-level domain name that refers to a commercial organization. As explained earlier, the URLs are classified into generic domains (e.g., .com, .edu, .net, and .gov). Addresses can be even more detailed, including the specific name of a document at a site.

To help minimize errors, most current browers allow you to type URLs without the "http://" and "www" portions of the URL. For example, you can type the text "yahoo.com" instead of "http://www.yahoo.com."

Wide Area Information Servers (WAIS)

Wide area information servers (WAIS) is an Internet directory designed to help end users find and retrieve information over the networks by providing efficient search methods. It is a distributed information retrieval system supported by Apple Computer, Thinking Machine, and Dow Jones. Users can ask for general topics, searching for key words and phrases. The search returns a list of documents, ranked according to the frequency of occurrence of the keyword(s) used in the search. The client can retrieve text or multimedia documents stored on the server. WAIS offers simple natural language input, indexed searching for fast retrieval, and a "relevance feedback" mechanism that allows the results of initial searches to influence future searches.

Mobile Internet

The **mobile Internet** refers to the use of wireless communications technologies to access network-based information and applications from mobile devices. The mobile Internet is also called the **wireless Web.**

Mobile Internet applications have several interesting applications. Users can access the mobile Internet anywhere, and the mobile Internet device can know where in particular it is located and use that knowledge to perform services that take advantage of spatial or geographic information. These applications are

called *location-based services*. The mobile Internet also provides services that are based on the type of location the mobile Internet device is in and that allow the user to act in ways that make sense only in that location. These applications are called *presence-based services*. Both types of services are examples of *location-based commerce* (*l-commerce*).

Internet Challenges

Challenges facing the Internet in the next few years include making the Internet more suitable for e-commerce transactions, incorporating rapidly evolving technologies, standards, and regulatory frameworks, responding to the growing need for additional bandwidth, and addressing privacy concerns. Another challenge for the Internet is providing the infrastructure for very large scale projects. We consider some of these challenges in the following paragraphs.

NEW TECHNOLOGIES. Vendors are adopting new technologies more rapidly than many users and customers can implement them. For example, the two most popular Web browsers are Netscape's Navigator and Microsoft's Internet Explorer. Many of the most innovative sites on the Web use Java applets, interactive three-dimensional graphics, and video and audio clips. To access these Web sites and take advantage of their innovative content and full functionality, users must have recent versions of Navigator and Explorer.

INTERNET REGULATION. Technical organizations, such as the Internet Engineering Task Force, the World Wide Web Consortium, and others, have played an important role in the evolution of the Internet and the Web. These organizations are not formally charged in any legal or operational sense with responsibility for the Internet. However, they perform the important task of defining the standards that govern the Internet's functionality. Hardware and software vendors also have been instrumental in submitting specifications for consideration to standard bodies and in creating de facto standards of their own.

Recent attempts by governments in the United States and elsewhere to regulate the content of Internet-connected computers have generated concerns about privacy, security, and the legal liability of service providers. Some content providers have addressed these issues with filters, ratings, and restricted access. However, it is difficult to regulate content across international borders. Regulation of services such as gambling also has been debated.

INTERNET EXPANSION. The Internet was not designed to provide a mass-market interchange of high-density information. As a result, the massive growth of Internet traffic has strained some elements of the network. The strains manifest themselves as slowdowns in retrieval time, unreliable transmission of streamed data, and denial of service by overloaded servers. The Internet's design, with many potential transmission paths, is in theory highly resistant to outages caused by failed links. In practice, however, the Internet often is affected by software problems.

A wide range of factors contributes to congestion or slowdowns. These problems include improperly configured networks, overloaded servers, rapidly changing Internet usage patterns, and too much traffic for available bandwidth. Approaches to solve these problems include installing high-speed transmission media to accommodate large amounts of data; bigger, faster routers and more sophisticated load balancing and management software to handle peak traffic

periods; local caching (storing) of frequently requested Web pages to improve response times; and more reliable tiers of service for those willing to pay for them.

INTERNET PRIVACY. Web sites collect information with and without consumers' knowledge. One way to collect information at Web sites is through registration (as on Amazon.com). Visitors to the site enter data about themselves, and obviously know that such information is available for future use by the company that collects the data.

The most common way Web sites collect information, though, is through "clickstream" data—that is, information about where people go within a Web site and the content they see. Clickstream data are most commonly collected by cookies. A **cookie** is a small data file placed on users' hard drives when they first visit a site. This software can be used to exchange information automatically between a server and a browser without a user seeing what is being transmitted.

Cookies are useful in tracking users' actions and preferences. When a user goes back to a site, the site's computer server can read the usage data from the cookies. This background information can then be used to customize the Web content that is given to the user. That information is stored in a database and can be used to target ads or content, based on the preferences tracked. Netscape and Internet Explorer browsers support cookie technology.

The Federal Trade Commission randomly checks Web sites to see if site operators are posting privacy notices that explain how personal information—such as e-mail addresses, shopping habits, and consumer financial data—is being used and whether that information is being protected from unwarranted intrusion. Privacy on the Internet at this point is not a sure thing.

There are bills in the U.S. Congress related to Web privacy, ranging from laws to regulate spamming (unsolicited e-mail) to legislation restricting disclosure of subscriber information by online services. Three possibilities exist:

- The government should let groups develop voluntary privacy standards but not take any action now unless real problems arise.
- The government should recommend privacy standards for the Internet but not pass laws at this time.
- The government should pass laws now for how personal information can be collected and used on the Internet.

Privacy is such an emerging important issue in the world of the Internet that it should be continuously considered in designing and using information systems.

INTERNET INVESTMENT. One final issue to consider is investment in Internet infrastructure. In the past, investments to improve the shortcomings of the Internet infrastructure were insufficient. As a result, equipment costs have outstripped revenue for the Web hosting, ASP, and outsourced computing companies as they have tried to catch up with the infrastructure needs. As entrepreneurs and financial backers develop products that will make data centers cheaper to equip and easier to manage, data centers will redirect innovation toward new technologies for truly distributed computing. According to Tier I Research, Web hosting is forecast to become a $28.5 billion business by 2005—a sevenfold increase. Companies like HP and IBM, in tandem with large telecoms, will offer computing-on-tap, so corporations will pay for computing

power much the same way they purchase electrical power—that is, they will pay for what they use. In other words, future deployment of hardware and software could become a service-centric computing model.

T5.3 Browsing and the World Wide Web

When you visit a library or a bookstore, you may browse through the books and magazines, reading a page in one and looking at a picture in another. Similarly, you can browse through the vast resources of the Internet. Initially, the amount of information on the Net was relatively small, so by going through catalogs and indices, it was possible to browse and find things quickly. But when the amount of information became very large, it was necessary to make the browsing more efficient. In 1991, a new way to organize information was introduced—the World Wide Web (WWW, or "the Web"). The **World Wide Web** (or "the Web")

The **Web** is a vast collection of interconnected pages of information that are stored on computers around the world that are connected to the Internet. The Web is a system with universally accepted standards for storing, retrieving, formatting, and displaying information (text, pictures, video, etc.) via a client/server architecture. The browser that runs on your PC is the client. When you search for information and services, Internet (or Web) servers process your request. The Web handles all types of digital information, including text, hypermedia, graphics, and sound. It uses graphical user interfaces, so it is very easy to use. The technology underlying the World Wide Web was created by Timothy Berners-Lee, who in 1989 proposed a global network of hypertext documents that would allow physics researchers to work together. This work was done at the European Laboratory for Particle Physics (known by its French acronym CERN) in Geneva, Switzerland.

Tools for Browsing the Web

Documents accessible on the Web contain **hyperlinks (links)** to other documents. Such links are used to connect documents, either internal or external. The links are basically an implementation of *hypertext*. When you click on a link, the browser will automatically display the associated document, called a **Web page,** on the topic you are exploring, regardless of where it is located on the Web. The **home page** is the starting point for your search. From a single Web page located on a computer attached to the Internet you can browse the Web by clicking on any interesting links you see. In most cases, the home page will lead users to other pages. All the pages of a particular company or individual are known as a **Web site.** Each Web site is a computer network, such as the one in your university, that has a connection to the Internet. Most Web pages provide a way to contact the organization or the individual. The person in charge of an organization's Web site is its **Webmaster.**

The most widely used browers are Netscape Navigator and Microsoft Explorer. These use GUI interfaces and require multimedia hardware and software. Similar capabilities, but without graphics, are provided by Lynx, a text-based browser developed by the University of Kansas.

NETSCAPE'S BROWSER SUITE. Netscape's browser suite includes Netscape Navigator, Netscape Mail, Netscape Instant Messenger, Netscape Composer, and Netscape Address Book. The suite provides functions for running Web applets, audio playback, streaming media, Web content, and Net2Phone for

free PC-to-phone calls anywhere within the United States. Netscape Communicator is a comprehensive set of components that integrates e-mail, Web-based word processing, and chat to allow users to easily communicate, share, and access information.

MICROSOFT INTERNET EXPLORER. Faced with the tremendous lead in the browser marketplace that Netscape established, Microsoft embarked on a strategy to gain market share and penetrate the installed base. It gave Internet Explorer away for free and bundled it with the Windows operating systems. This approach was successful commercially but also resulted in scrutiny by the U.S. government.

In October 1997 the U.S. Justice Department filed a petition in federal court to prevent Microsoft from requiring personal computer manufacturers to bundle Microsoft's Internet browser software with Microsoft's Windows operating systems. At issue was whether or not Microsoft tried to monopolize the Internet browser software business by refusing to let PC makers license the Windows operating systems unless they also shipped their PCs with Internet Explorer. Microsoft maintained that Internet Explorer was an enhancement of Windows, not a separate product, and that the company therefore was not violating its antitrust settlement. According to Microsoft, Internet Explorer's tight integration with Windows offers users the advantage of "one-stop computing."

In early November 2001, the Justice Department reached a settlement with Microsoft in the antitrust case. Under the settlement, a panel of three independent monitors will work onsite at Microsoft to oversee its conduct and review its accounts. Crucially, Microsoft must provide rival software firms with information to allow them to develop competing "middleware" products—in other words, software programs that interact with the computer operating system. Microsoft was prevented from retaliating against computer manufacturers and software rivals who brought out competing products, and had to deal with licensing partners on uniform terms.

METABROWSERS. Web surfing has typically been a one-page-at-a-time experience. Then Yahoo offered My Yahoo, which enabled personalization of news, sports, financial data, entertainment, and other topics. **Metabrowsers** automate the tasks of creating metadata, letting users concentrate on the task of cataloguing Web resources (*metabrowser.spirit.net.au/manual/Manual.1.10.doc*). Octopus is a leading example. Other metabrowsing services include CallTheShots (*calltheshots.com*), Quickbrowse (*quickbrowse.com*), Katiesoft (*katiesoft.com*), and Yodlee (*yodlee.com*).

Here is how the Octopus (*octopus.com*) provides portal personalization by giving consumers control over what they view: Users assemble customized pages, called "views" on Octopus, choosing to start with a blank view or to use one of the sample views arranged in a menu. To add elements, users grab items from the menu and drag them onto their pages. The views are dynamically linked to the Web sites providing the content. Users who click on any page element—for example, a news headline or a stock chart—are connected directly to that site. Users can store their views privately on Octopus, or they can "publish" them so other users can take a look. For users, the service is free. Octopus plans to make money by charging Web publishers a fee each time users click through to their sites.

NEW BROWSERS. There also are a couple of newcomers to the scene, Opera and NeoPlanet. Opera is renowned for being the world's fastest browser. It is much smaller than other major browsers. In its standard configuration, it almost fits on a floppy disk, yet it is all you need to surf the Web. NeoPlanet offers a "skinnable" browser. NeoPlanet *skins* are collections of graphics and sounds. It also utilizes personalization tools, allowing you to control your view of the online world. Designed to be small, full-featured, and easy to use, this free browser can look and function the way that best complements your needs.

OFFLINE BROWSERS. **Offline browsers** (or pull products) enable a user to retrieve pages automatically from Web sites at predetermined times, often during the night. WebWhacker (*bluesquirrel.com*) and WebCopier (*maximumsoft.com*) are offline browsers that allow users to define a group of sites by their URLs and then download text and images from those sites to their local storage. WebWhacker and WebCopier let users determine how much of a Web site to retrieve—title pages only, any linked pages, or all pages.

WEB SERVERS. A **Web server,** also called an *HTTP server,* is a computer (usually dedicated) storing all Web pages and/or Web content. Users can access the Web server using a Web browser with a unique URL or IP address (see Figure T5.1). There are many types of Web servers. The best known are Microsoft's Internet Information Systems (IIS) and Linux's Apache.

The Apache HTTP Server Project is an effort to develop and maintain an open-source HTTP server for modern operating systems, including UNIX and Windows NT. The goal of this project is to provide a secure, efficient, and extensible server that provides HTTP services in sync with the current HTTP standards. In October 2004, Netcraft Web Server Survey found that more than 67 percent of the Web sites on the Internet were using Apache.

Others Web servers, used in particular niche markets, include the following:

- *Jigsaw* is W3C's leading-edge Web server platform, providing a sample HTTP 1.1 implementation and a variety of other features on top of an advanced architecture implemented in Java.
- *WN* is a Web server that runs on a wide variety of UNIX platforms and is freely available at no cost for any use under the terms of the GNU General Public License.
- *Small HTTP Server,* version 3.05, created by Max Feoktistov, claims to be the smallest Web server while having numerous functionalities. For details, see *home.lanck.net/mf/srv/index.htm.*

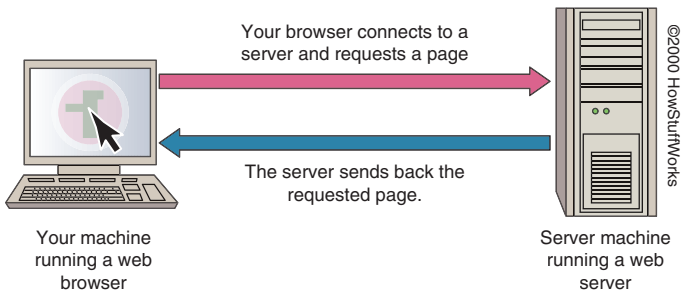

FIGURE T5.1 Relationship of a Web browser to a Web server.
Source: computer .howstuffworks.com/web-server.htm/printable