```
<html>
<head>
<title>How The Internet Works</title>
<meta http-equiv="Content-Type" content="
</head>
<body bgcolor="#FFFFFF" link="#6666CC"
<div id="Layer1" style="position:absolu
<div id="Layer2" style="position:absolu
<div id="Layer3" style="position:absolu
<div id="Layer4" style="position:absolu
<div id="Layer5" style="position:absolu
<div id="Layer6" style="position:absolu
  <p><font size="3" face="Verdana, Arial
    what's happening behind the screens
  <p><font size="3" face="Verdana, Ari
    about creating your own Web site, bu
  <p><font size="3" face="Verdana, Arial
    all magic and madness to you?</
  <p><font size="3" face="Verdana, Arial
    have come to the right place! This
    out of the Internet and explaining
  <p> </p>
  <p><font face="Verdana, Arial, Helve
    about:</b></font></p>
  <p><b><font face="Verdana, Arial, Hel
    Anatomy of a Web Site</a></font><
  <p><b><font face="Verdana, Arial, Hel
    Architecture of the Web</font></a><
  <p><b><font face="Verdana, Arial, Hel
    Technologies on the Web</font></a><
  <p> </p>
  <p><font face="Verdana, Arial, Helveti
  <p><font face="Verdana, Arial, Helveti
    site you will also find lots of hyper
    Wide Web.</font></p>
</div>
<div id="Layer7" style="position:absolu
  <p><font color="#999999"><b><font size
  <p><font color="#9999CC"><a href="intr
  <p><font color="#999999"><b><font size
    I<br>
    <a href="pt01start.html">What is the
  <p><font color="#999999"><b><font size
    II<br>
    <a href="part02start.html">The Inter
  <p><font color="#999999"><b><font size
    III<br>
    <a href="part03start.html">Connecti
  <p><font color="#999999"><b><font size
    IV<br>
    <a href="part04start.html">Communic
  <p><font color="#999999"><b><font size
    V<br>
    <font color="#6666CC"><a href="part
  <p><font color="#999999"><b><font size
    VI<br>
    <font color="#6666CC"><a href="part0
  <p><font color="#999999"><b><font size
    VII<br>
    <font color="#6666CC"><a href="part
  <p><font color="#999999"><b><font size
    VIII<br>
    <font color="#6666CC"><a href="part
    the Internet</a></font></font></b>
  <p><font color="#999999"><b><font size
    IX<br>
    <font color="#6666CC"><a href="part
  <p><font color="#999999"><b><font size
    to <br>
    Table of Contents</font></a></
  <p> </p>
  <p> </p>
  </div>
<div id="Layer8" style="position:absolu
<img src="bot.gif" width="151" height="
<div id="Layer9" style="position:absolu
<div id="Layer10" style="position:absol
<div id="Layer11" style="position:absol
</body>
</html>
```

Page Title

Background Color

Layer 1

Hyperlink

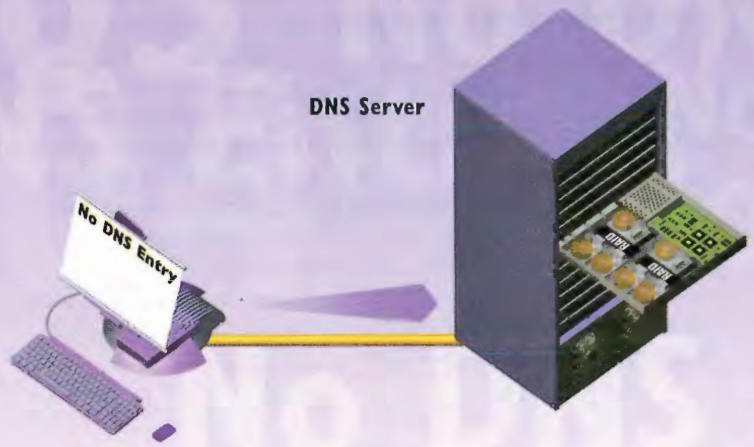
3 The coding in the HTML files tells your browser how to display the text, graphics, links, and multimedia files on the home page. The HTML file your browser loads to display the home page doesn't actually have the graphics, sound, multimedia files, and other resources on it. Instead, it contains HTML references to those graphics and files. Your browser uses those references to find the files on the server and then display them on the home page.

4 The Web browser also interprets HTML tags as links to other Web sites, or to other Web resources, such as graphics, multimedia files, newsgroups, or files to download. Depending on the link, it performs different actions. For example, if the HTML code specifies the link as another home page, the browser retrieves the URL specified in the HTML file when the user clicks the underlined link on the page. If the HTML code specifies a file to be downloaded, the browser downloads the file to your computer.

Many types of files exist on the Internet that Web browsers can't display. In particular, this pertains to many kinds of multimedia files, such as sound, video, and animation files. Still, there are often references to these types of files on Web pages. To view or play these files, you need what are called helper applications and plug-ins. You must configure your Web browser or operating system to launch these helper applications and plug-ins whenever you click an object that needs them in order to be viewed. Helper applications and plug-ins also can be used for displaying virtual reality pages, for chatting on the Internet, and for performing other Internet tasks.

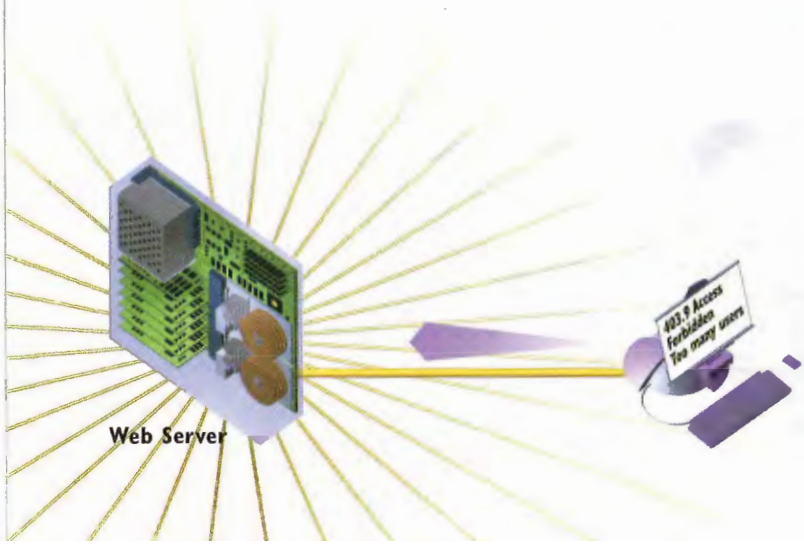
The meanings of tags are easily decipherable. Every HTML tag, or instruction, is surrounded by a less-than and a greater-than sign—<P>. Often tags appear in pairs, the beginning tag and the ending tag. They are identical except for a simple slash in the end tag. So, a paragraph of text frequently is surrounded by tags such as this: <P> Paragraph of text.</P>. Also, tags are not case sensitive. <P> equals <p>.

What Common Browser Error Messages Mean



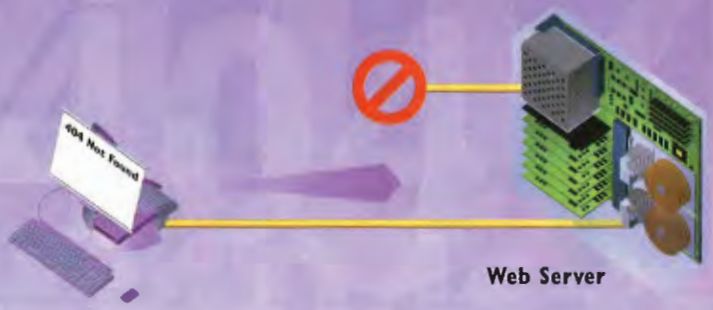
Server Does Not Have a DNS Entry When you type a URL in your browser to visit a site, your computer contacts a server called a Domain Name System (DNS) server. The DNS server translates the URL into the IP number that computers can understand—and after it does that, your browser can go to the site. (For more information about DNS servers, see Chapter 5, "How Internet Addresses and Domains Work.") If you get an error message telling you the server doesn't have a DNS entry, it means that the server doesn't have a listing for the URL you typed. This usually means either you typed the URL incorrectly or something is wrong with the DNS server. Check the URL and retype it.

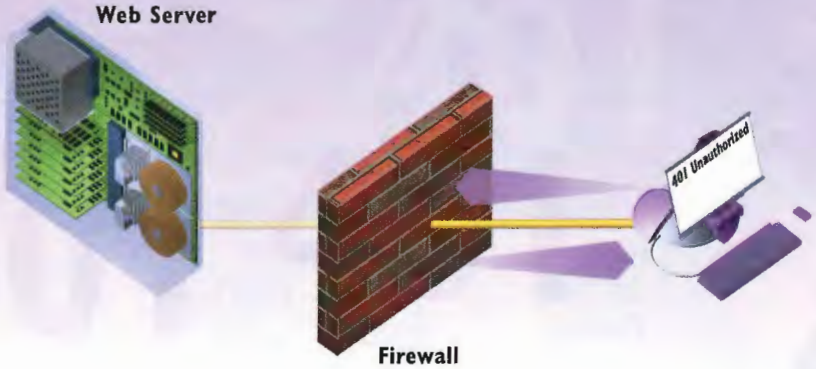
503 Service Unavailable This is a catch-all error message for a variety of problems, but all of them mean that the Web site has been incapable of being contacted by your browser. The problem might be that the site's server has crashed because of too much traffic or that there's network congestion.



403.9 Access Forbidden. Too Many Users Are Connected. Some Web sites recognize that if they get too much traffic at once, the entire site can be brought down and no one will be able to visit. Those sites put a limit on the number of people who can come to the site at once—that way, the site is always available, even if everyone who wants to visit can't get in. If you get this "Too Many Users Are Connected" message, it usually means that the Web site is up and running, but you can't get in because the maximum number of people are already on the site. Keep trying—when one person leaves, another can come in, and it might be you.

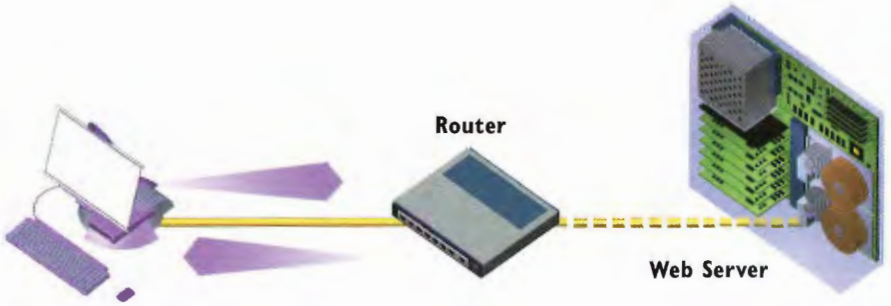
Spinning hourglass This isn't an error message your browser displays—instead, your Windows cursor turns into a spinning hourglass. The spinning hourglass tells you that your browser is trying to make a connection to a Web site. If it keeps spinning and a connection is never made, it can mean that an Internet router someplace between where you are and where you're trying to visit has crashed and you can't make the connection. It can also mean that you've lost your local Internet connection for some reason.





401 Unauthorized and 403 Forbidden If you get either of these error messages, you're trying to enter a Web site that allows only certain people in—and it's not allowing you to enter. Typically, these types of Web sites are password protected and also might allow only visitors who are from certain domains, such as zd.com. If you've entered a password, you might have entered it improperly, or you might not be in a domain that's allowed to enter the Web site.

404 Not Found When you get this message, you've arrived at the correct Web site, but the specific page you're looking for can't be found. That specific page might have been deleted from the site or moved—or you might have typed the location incorrectly.



CHAPTER 23

How Markup Languages Work

MARKUP languages are the road signs of a Web page. They are sets of directions that tell the browser software how to display and manage a Web document, much like written music scores are instructions that tell a musician how to play a particular song. These instructions (called tags or markups) are embedded in the source document that creates the Web page.

Tags reference graphic images located in separate files, and they instruct the browser to retrieve and display these images within the page. Tags can also tell a browser to connect a user to another file or URL when he clicks an active hyperlink. So each Web page has everything it needs to be displayed on any computer with a browser that can interpret the markup language.

Your original text will probably have headings, multiple paragraphs, and some simple formatting. A Web browser will not understand all these layout instructions because the original text isn't formatted with HTML, the language of the Web (discussed later in this chapter). Paragraphs, carriage returns, indents, and multiple spaces will be shown instead as a single space if no HTML markup is added.

Markup languages should not be confused with programming languages, such as C+ or Pascal. Programming languages are used to write complex applications, such as word processors or spreadsheets. Markup languages, in contrast, are much simpler and describe the way information should be displayed—for example, by defining when text should be boldface. In markup languages, tags are embedded within documents to describe how the documents should be formatted and displayed.

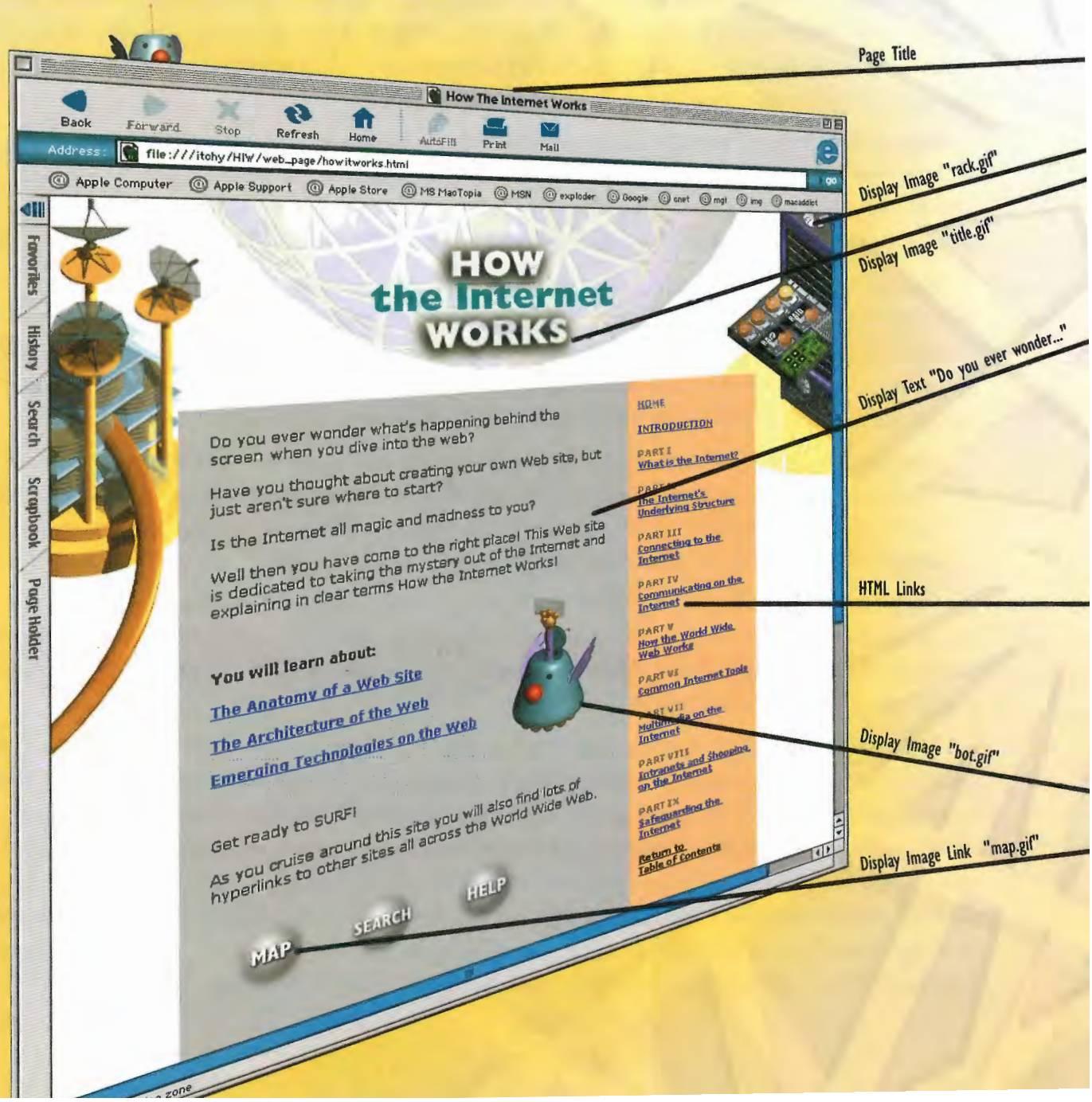
Hypertext Markup Language (HTML) is the markup language of the Web. It defines the format of a Web document and enables hypertext links to be embedded in the document. You can use any text editor or word processor to add HTML tags to an ASCII text document, although a number of shareware and commercially available HTML editors can assist Web page authors as well.

The Web evolves daily, and HTML also expands and changes along with it. The newest changes to HTML are a group of technologies that together are termed *Dynamic HTML (DHTML)*. These technologies allow HTML to be more than a static language, and they enable HTML to perform animations and become more interactive and flexible.

The eXtended Markup Language (XML) promises to bring even more significant changes to the Web. It's dramatically different from other markup languages because it separates the content of a page from its presentation. Rather than doing things such as giving instructions on text size, it tags different types of content and then has other technologies such as templates and style sheets determine how that content should look. For example, if a book were to be presented on the Web, it would tag chapter titles, chapter numbers, chapter text, and so on. It could then use style sheets to define how chapter titles, chapter numbers, and chapter text should look. Doing this means you can build the content of a page once and present it to many different devices, and in many different ways, without doing much extra work.

How HTML Works

To display Web pages in any browser, you must add HTML tags to your original text. This process is called *tagging*.



2 Use HTML to give your text structure. All HTML files begin and end with the HTML tags. Headings are marked as such, as are paragraphs, line breaks, block quotes, and special character emphasis. Any carriage returns or indentations within the source text do not affect the browser's display of the page. HTML tags need to be put in if they are to be displayed in a browser.

3 The finished HTML document will be the source page for any browser on any computer. This simplicity of HTML makes cross-platform compatibility easy and reliable. The more complex and specialized the HTML tagging, the longer it will take to download and display the document.

Title

Graphic

Graphic

Text

Text

Hyperlinks

Graphic

Graphic
Hyperlink

```
<html>
<head>
<title>How The Internet Works</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>

<body bgcolor="#FFFFFF" link="#6666CC" vlink="#666666">
<div id="Layer1" style="position:absolute; left:601px; top:-66px; width:180px; height:320px; z-index:1"><img src="rack.gif" width="217" height="372"></div>
<div id="Layer2" style="position:absolute; left:0; top:0; width:203px; height:352px; z-index:2"><img src="rtside.gif" width="125" height="450"></div>
<div id="Layer3" style="position:absolute; left:106px; top:0px; width:494px; height:162px; z-index:3"><img src="title.gif" width="545" height="173"></div>
<div id="Layer4" style="position:absolute; left:91px; top:180px; width:502px; height:875px; z-index:4; background-color: #CCCCCC; layer-background-color: #CCCCCC; border: 1px none #0
<div id="Layer5" style="position:absolute; left:506px; top:180px; width:155; height:877px; z-index:5; background-color: #FFCC99; layer-background-color: #FFCC99; border: 1px none #00
<div id="Layer6" style="position:absolute; left:115px; top:201px; width:369px; height:604px; z-index:6">
  <p><font size="3" face="Verdana, Arial, Helvetica, sans-serif">Do you ever wonder
    what's happening behind the screen when you dive into the web?</font></p>
  <p><font size="3" face="Verdana, Arial, Helvetica, sans-serif">Have you thought
    about creating your own Web site, but just aren't sure where to start?</font></p>
  <p><font size="3" face="Verdana, Arial, Helvetica, sans-serif">Is the Internet
    all magic and madness to you?</font></p>
  <p><font size="3" face="Verdana, Arial, Helvetica, sans-serif">Well then you
    have come to the right place! This Web site is dedicated to taking the mystery
    out of the Internet and explaining in clear terms How the Internet Works!</font></p>
  <p> </p>
  <p><font face="Verdana, Arial, Helvetica, sans-serif" size="3"><b>You will learn
    about:</b></font></p>
  <p><b><font face="Verdana, Arial, Helvetica, sans-serif" size="3"><a href="pt05_ch2102.html">The
    Anatomy of a Web Site</a></font></b></p>
  <p><b><font face="Verdana, Arial, Helvetica, sans-serif"><a href="pt01_ch0101.html"><font size="3">The
    Architecture of the Web</font></a></font></b></p>
  <p><b><font face="Verdana, Arial, Helvetica, sans-serif"><a href="pt07_ch3205.html"><font size="3">Emerging
    Technologies on the Web</font></a></font></b></p>
  <p> </p>
  <p><font face="Verdana, Arial, Helvetica, sans-serif">Get ready to SURF!</font></p>
  <p><font face="Verdana, Arial, Helvetica, sans-serif">As you cruise around this
    site you will also find lots of hyperlinks to other sites all across the World
    Wide Web.</font></p>
</div>
<div id="Layer7" style="position:absolute; left:517px; top:198px; width:139px; height:631px; z-index:7">
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif"><a href="index.html"><font color="#9999CC">HOME</font></a></font></b></font></p>
  <p><font color="#9999CC"><a href="intro.html"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">INTRODUCTION</font></b></a></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    I<br>
    <a href="pt01start.html">What is the Internet?</a></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    II<br>
    <a href="part02start.html">The Internet's Underlying Structure</a></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    III<br>
    <a href="part03start.html">Connecting to the Internet</a></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    IV<br>
    <a href="part04start.html">Communicating on the Internet</a></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    V<br>
    <font color="#6666CC"><a href="part05start.html">How the World Wide Web Works</a></font></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    VI<br>
    <font color="#6666CC"><a href="part06start.html">Common Internet Tools</a></font></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    VII<br>
    <font color="#6666CC"><a href="part07start.html">Multimedia on the Internet</a></font></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    VIII<br>
    <font color="#6666CC"><a href="part08start.html">Intranets and Shopping on
    the Internet</a></font></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif">PART
    IX<br>
    <font color="#6666CC"><a href="part09start.html">Safeguarding the Internet</a></font></font></b></font></p>
  <p><font color="#999999"><b><font size="2" face="Verdana, Arial, Helvetica, sans-serif"><a href="index.html"><font color="#000000">Return
    to <br>
    Table of Contents</font></a></font></b></font></p>
  <p> </p>
  <p> </p>
  </div>
<div id="Layer8" style="position:absolute; left:351px; top:380px; width:129px; height:132px; z-index:8"><img src="bot.gif" width="151" height="154"></div>
<div id="Layer9" style="position:absolute; left:128px; top:522px; width:77px; height:68px; z-index:9"><a href="map.html"><img src="map.gif" width="83" height="83" border="0"></a></di
<div id="Layer10" style="position:absolute; left:215px; top:623px; width:65px; height:76px; z-index:10"><a href="searchHLW.html"><img src="search.gif" width="83" height="83" border="
<div id="Layer11" style="position:absolute; left:312px; top:622px; width:67px; height:101px; z-index:11"><a href="helpme.html"><img src="help.gif" width="83" height="83" border="0"><
</body>
</html>
```

4 Most Web browsers enable your document to retain its structural integrity when you display, or *parse*, it. Headings will appear in a larger font size than text within paragraphs, for example, and block quotes will be uniformly indented. However, the look might vary from browser to browser. Note that browsers determine the exact font, size, and color. Also be aware that the relative importance of the elements is always kept intact.

How Dynamic HTML Works

1 Dynamic HTML (DHTML) differs from traditional HTML in that it enables Web pages to be changed on-the-fly, after they've been downloaded. In plain HTML, after a page is downloaded, it is static and can be changed only when a user takes an action of some kind. But DHTML, for example, could cause an animation of a rocket to fly across your browser window several seconds after the page has been downloaded—without your doing anything.

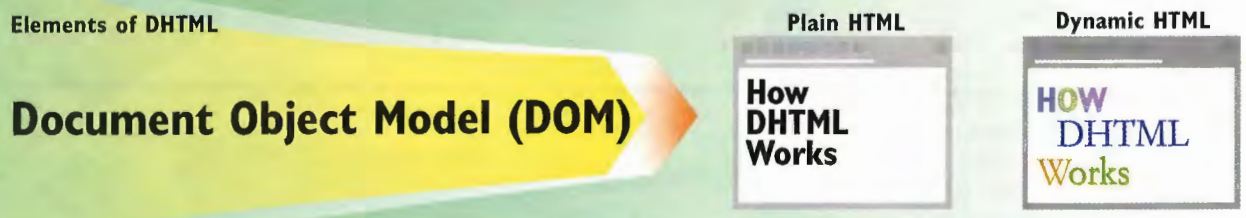
HTML

DHTML

Web Server

2 DHTML does its work without having to contact the server after the page downloads, so it can perform some interactive functions more quickly than other technologies that have to contact the server. The instructions for performing the commands are in the HTML commands that are in the page itself.

3 Although DHTML is often referred to as if it were a single technology, it is, in fact, a general term used for a group of technologies that can work together or by themselves to change a Web page after the page has been downloaded to your computer. These technologies are the Document Object Model (DOM), Cascading Style Sheets (CSS), and client-side scripting languages, such as JavaScript.



4 The DOM defines every object and element on a Web page and enables those objects to be manipulated or accessed. This includes fonts, graphics, tables, and visual elements, as well as elements you can't necessarily see, such as the browser's version number and the current date and time. Without DOM, all the elements on a page are static. So on the simplest level, DHTML could use the DOM to change the font of every letter, individually, on a Web page.

5 Cascading Style Sheets are, in essence, templates that apply formatting and style information to the elements of a Web page. They're called *cascading* because any single page can have more than one style sheet associated with it. Additionally, Cascading Style Sheets enable images to overlap one another. This enables animations to be created easily on a page.

6 Client-side scripting languages perform much of the work of DHTML. These languages access the DOM and manipulate its elements, and they do the same to Cascading Style Sheets. They perform the actions of DHTML. So a script, for example, could turn a word a different color when a mouse moves across it, or it could create easy-to-use collapsible navigation on every page on a Web site.

How XML Works

1 XML solves several a major problems for Web developers. Without it, to deliver Web pages to different devices such as computers, cell phones, and wireless Palm devices, a developer must create and maintain three separate Web sites, with special coding for each device—HTML for the computer, WAP for the cell phone, and what's called Web clipping for the wireless Palm devices. It's an expensive, difficult, and time-consuming proposition. And even if a developer is building a site only for computers, every time the design changes, all the pages must be recoded—again, an expensive and time-consuming proposition.

2 With XML, a developer can create the Web site just once. Then it can be automatically formatted to several types of devices, such as Internet-connected computers, wireless Palm devices, and cell phones using WAP. And even if the site is being built for only computers, when there is a redesign, with XML, all the pages need not be rebuilt.

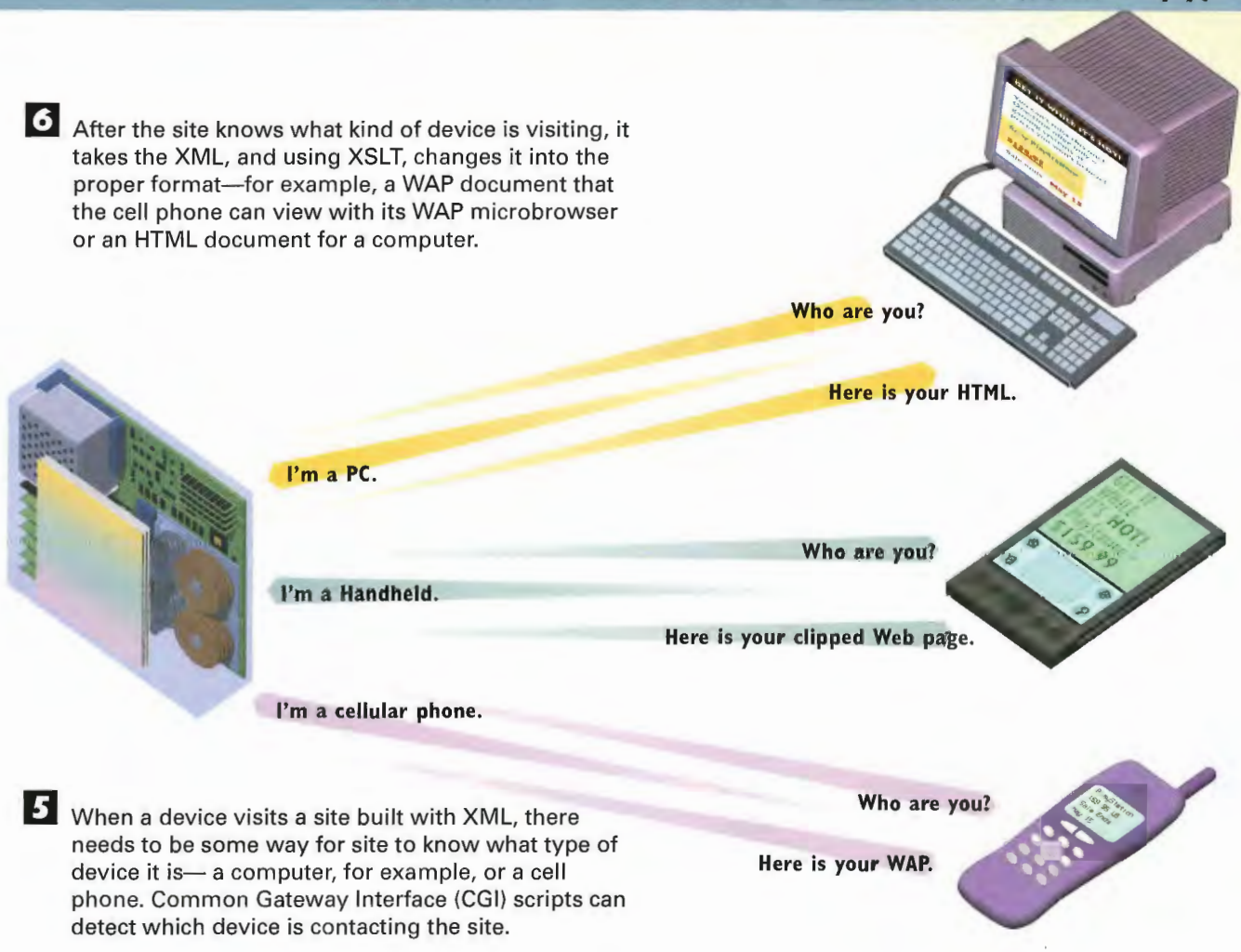
```
<Sale Flyer>
<Offer>Get It While It's Hot!</Offer>
<Promotional Copy>
You can't miss this one!
One-time offer only –
gaming systems at prices
you won't believe! </Promotional Copy>
<Product>Sony PlayStations</Product>
<Price> $159.95</Price>
Sale ends <End Date> May 15
</End Date>
</Sale Flyer>
```

XML

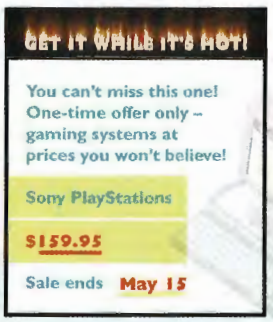
XSLT

3 The most important concept to understand about XML is that the language is used only to convey information about content, not about the presentation of the content. So, for example, it doesn't give instructions on what size text should be. But it uses tags to define the type of content on the page. Then it uses other techniques, as you'll see in the next steps, to display those pages. In that way, a single page can be displayed many different ways, without having to go back and alter the original page—only the designs, which are separate from the content, need to be changed.

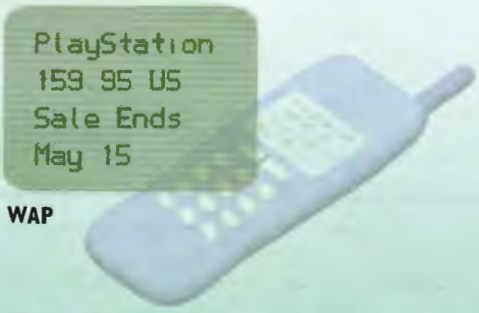
6 After the site knows what kind of device is visiting, it takes the XML, and using XSLT, changes it into the proper format—for example, a WAP document that the cell phone can view with its WAP microbrowser or an HTML document for a computer.



5 When a device visits a site built with XML, there needs to be some way for site to know what type of device it is— a computer, for example, or a cell phone. Common Gateway Interface (CGI) scripts can detect which device is contacting the site.



HTML

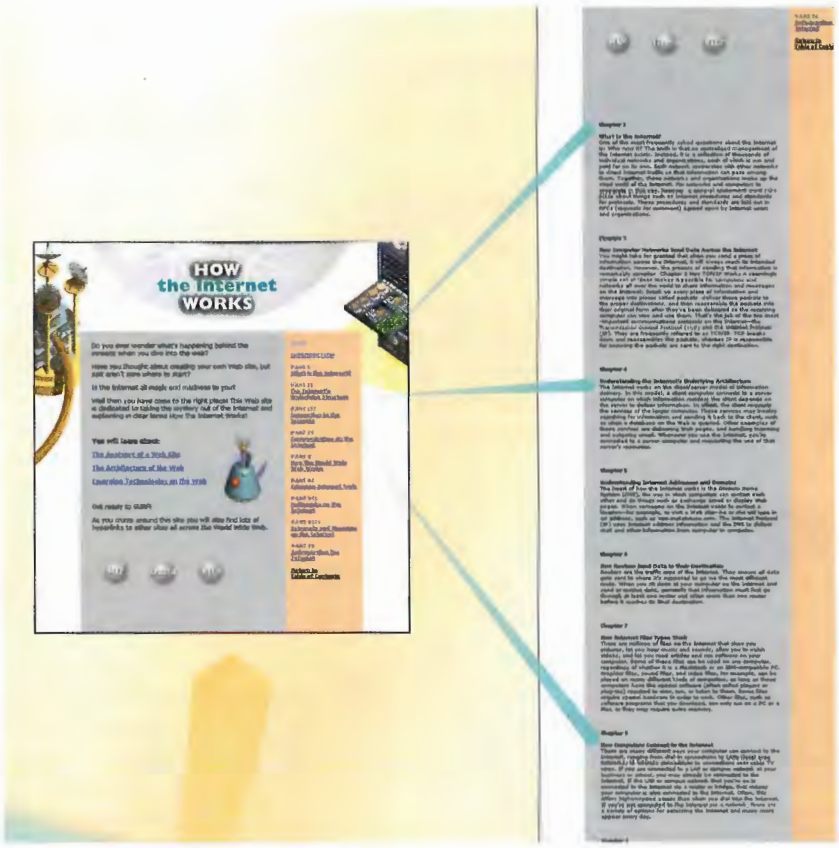


WAP

4 When XML content is posted on a Web site, different designs need to be applied to that content so that it can be viewable by devices connecting to it—for example, cell phones. eXtensible Style Language Transformations (XSLT) can be applied to the XML. XSLT can take XML and apply different designs to it or change it to other forms of XML—for example, it can take the XML and turn it into a WAP page that can be viewed by a cell phone and take the same XML and turn it into a HTML document with a different design.

CHAPTER

24 How Hypertext Works



IN the late 1960s, a computer scientist named Ted Nelson introduced hypertext, a concept that lays the foundation for the World Wide Web and its connections between documents, or pages.

Nelson wanted to create a new way of exploring information. He wanted to provide the reader with a spontaneous means of accessing more and more in-depth information about something that sparked the reader's interest when reading text on the page. Rather than reading a document from beginning to end, digesting the material in a sequential order, the reader could highlight a word and receive more information on the meaning of that word, for example.

Nelson imagined that you could read the United States Constitution, come across the term "electoral college," and then open another document that explains how the electoral college works. From that document, you might open another document that lists the votes of the electoral college from its founding to the present. From there, you might choose to open a document about William Jefferson Clinton, then another about the First Cat, Socks, then another about the care and feeding of felines, and so forth. Ultimately, you could find a subject that wasn't even remotely connected to the Constitution, but which you would find interesting or entertaining.

This hypertext concept was obviously on Tim Berners-Lee's mind when he began thinking about how researchers could share their work across the Internet. He envisioned a system in which a document could be linked to other documents, enabling researchers to easily find more and related information simply by following a link from one document on the network to another.

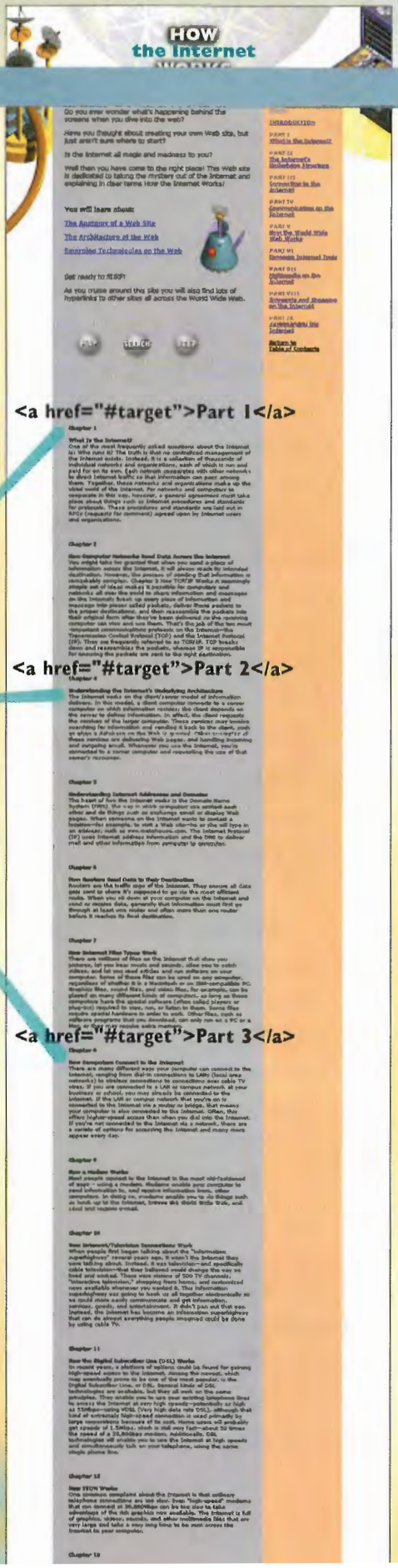
Typically, hypertext consists of a hyperlink that appears onscreen as a highlighted word, icon, or graphic. By moving a mouse cursor over the item, or object, and clicking it, you easily navigate to additional information. On the Web, that information can be located at any other place on the Web, be it on the same host server or one across the globe. A linked object can be various media, such as text (linking from one character to a whole document, for example), a graphical button (such as direction arrows that move from page to page), or still images (photos, icons, or a comic strip), for example. The documents and objects that are being linked to can be on the same site as the original document or on an entirely different site.

Hypertext links are embedded into a Web document using Hypertext Markup Language (HTML). A text link usually appears onscreen as an underlined word or phrase and is sometimes rendered in a different color from other text, depending on how your Web browser interprets the HTML codes. When you place the mouse cursor on this underlined text and click the mouse button, you initiate a request by the browser for a new Web page or—if the text references an internal link to information in the same document—direct your browser to scroll to another, specific point within the same document.

Images or icons can also act as hyperlinks. When you move the mouse cursor over the icon or graphic and click the mouse button, you launch the request to retrieve the linked information.

How Hyperlinks Work

1 The "hyperlinking" begins when you first retrieve a Web page from a remote Web server. Target links within the page move you quickly from one part of the page to the next.



Part 1

Part 2

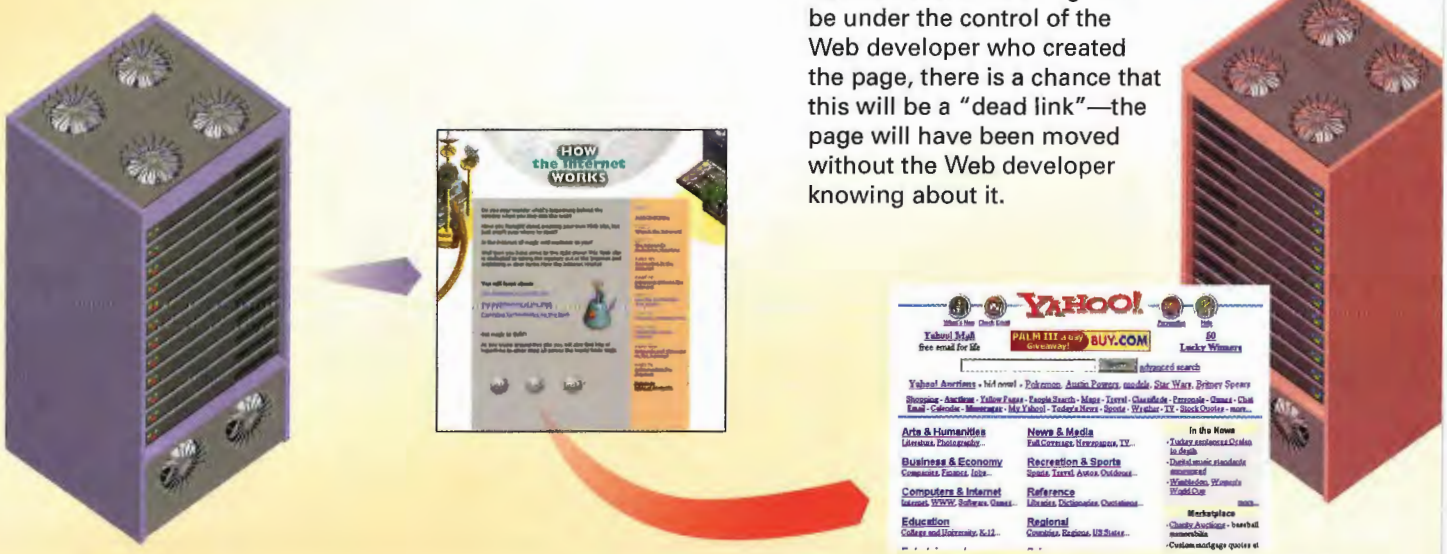
Part 3

2 A *relative* link is used to initiate a request by your Web browser to retrieve a page located on the same Web server as the page from which you are linking. Web developers use relative links when they're pointing to a page under their control because this enables the to more easily maintain their HTML coding if their server locations change.

Chapter 4

Chapter 7

3 A hyperlink that leads to a completely different Web server uses an *absolute* link. Because these links might not be under the control of the Web developer who created the page, there is a chance that this will be a "dead link"—the page will have been moved without the Web developer knowing about it.



<img src="searchyahoo.gif">

CHAPTER

25 How URLs Work

THE Web pages and the hosts that make up the World Wide Web must have unique locations so that your computer can locate and retrieve the pages. The unique identifier for a host is called the Internet Protocol (IP) address, and the unique identifier for a page is called the uniform resource locator (URL). A URL functions much like a postal or e-mail address. Just as postal and e-mail addresses list a name and specific location, a URL, or Web address, indicates where the host computer is located, the location of the Web site on the host, and the name of the Web page and the file type of each document, among other information.

A typical URL looks like this:

```
http://www.zdpress.com/internetworks/index.html/
```

If you were to interpret the instructions in this URL from left to right, it would translate to: "Go to the host computer called `zdpress` (a commercial business), in a directory called `internetworks`, and retrieve a hypertext document with the filename `index.html`." The URL, or address, tells the browser which document to fetch and exactly where to find it on a specific remote host computer somewhere on the Internet.

The first part of the URL indicates what type of transfer protocol will be used to retrieve the specified document. The most common request is for a hypertext document that uses Hypertext Transfer Protocol (HTTP).

The second portion of the URL refers to the specific host computer on which the document resides, which is to be contacted by the browser software. This part of the address is also called the *domain name*. See Chapter 5, "How Internet Addresses and Domains Work," for more information about domains.

The third part of the URL is the directory on the host computer that contains a specific Web site or multiple Web sites. This is always located after the first single slash in the URL and is essentially the subdirectory on the hard disk that houses the Web site. Subdirectories might also be indicated in this part of the address. For example, if the previous URL were changed to `http://www.zdpress.com/internetworks/partone/chapters/chapte.html` there would be two subdirectories—`part one` and `chapters`.

In the preceding example, the filename is `chapte.html`. This is always the last portion of the URL. If you see an address without a filename, it is assumed that the filename `index.html` contains the requested Web page. Therefore, the default document a Web server will deliver to the client when no other filename is listed is `index.html`. (Note that sometimes the last portion of the URL might not be a filename—it could be other types of information required by a Web server, such as codes required to log on to the Web server.)

The illustration in this chapter shows the process necessary to request and retrieve a Web document. When a request for a document occurs for the first time in a Web-browsing session, the host computer must first be located to find the file. After that, the specific subdirectory and document are retrieved.

How URLs Are Structured

1 The first part of the URL indicates which type of transfer protocol will be used to retrieve the specified document. The most common request is for a hypertext document that uses the HTTP protocol.

3 The third part of the URL is the directory on the host computer that contains a specific Web site. A host computer can house multiple Web sites. This third segment of the address is essentially the root directory that houses the HTML document, CGI, image, and so on that is being requested. Subdirectories might also be indicated in this part of the address.

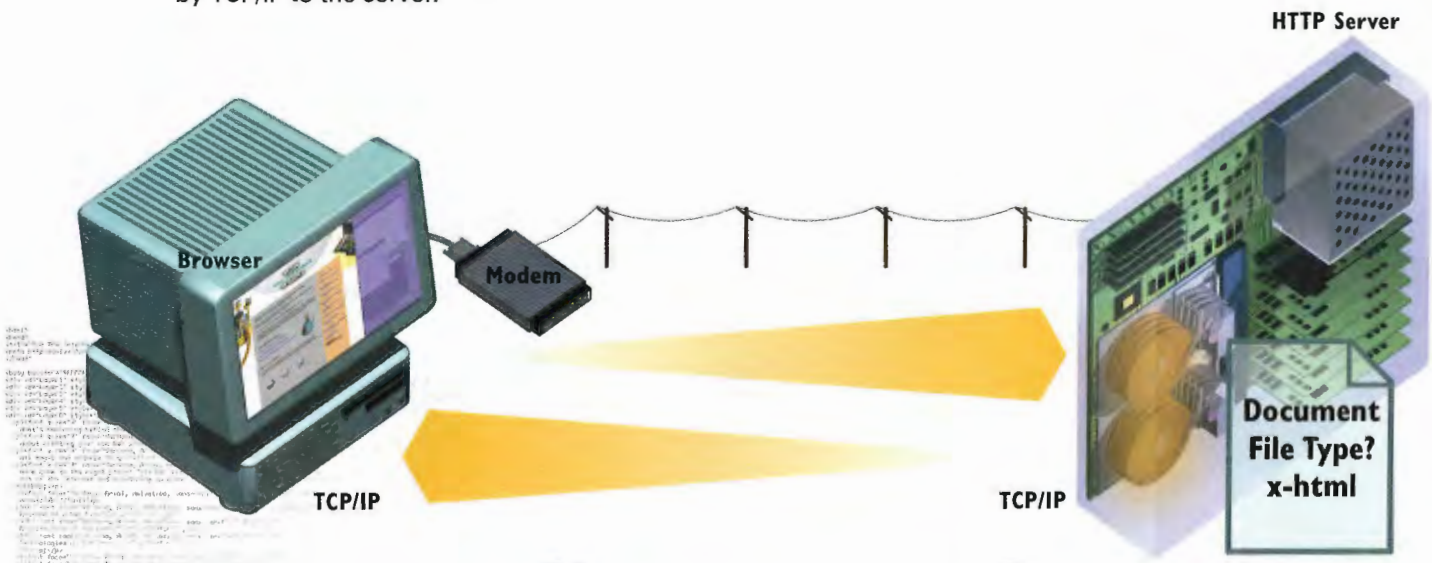
2 The second portion of the URL is the specific host computer on which the document resides, which is to be contacted by the browser software. This part of the address is also called the *domain*. Domain names end in a suffix that indicates which type of organization the domain is. For example, `.com` indicates a commercial business, `.edu` indicates a college or university, `.gov` indicates a government office, `.mil` indicates a military facility, and `.org` indicates a not-for-profit organization. The suffix also can indicate the country in which the host computer is located. For example, `.ca` is in Canada, and `.au` is in Australia.

4 The last segment of the URL is the filename of the specific Web page you are requesting. If no filename is indicated, the browser assumes a default page, usually called `index.html`.

How URLs Help Retrieve Web Documents

1 The Web browser installed on your local computer sends your TCP/IP software a signal that it is ready to request a document. TCP/IP makes a connection with the host TCP/IP software. After the connection is established, your browser makes a request for a document by sending its URL through the two-way connection maintained by TCP/IP to the server.

2 The HTTP server is the portion of the host computer that runs HTTP server software. TCP/IP makes and maintains the connection this way. The browser can use HTTP to send requests and receive pages through the host's Web server software. This software enables the host to communicate with the client browser, in HTTP, over TCP/IP.



5 The browser on your local computer reads the file type. If it is an HTML document, the browser examines the content, breaking it down into meaningful parts. Two general parts include text, which is displayed by the browser word for word. The other part consists of HTML markup information called *tags*, which are not displayed but display formatting information, such as normal text, bold headers, or colored hypertext. The results are displayed on your monitor.

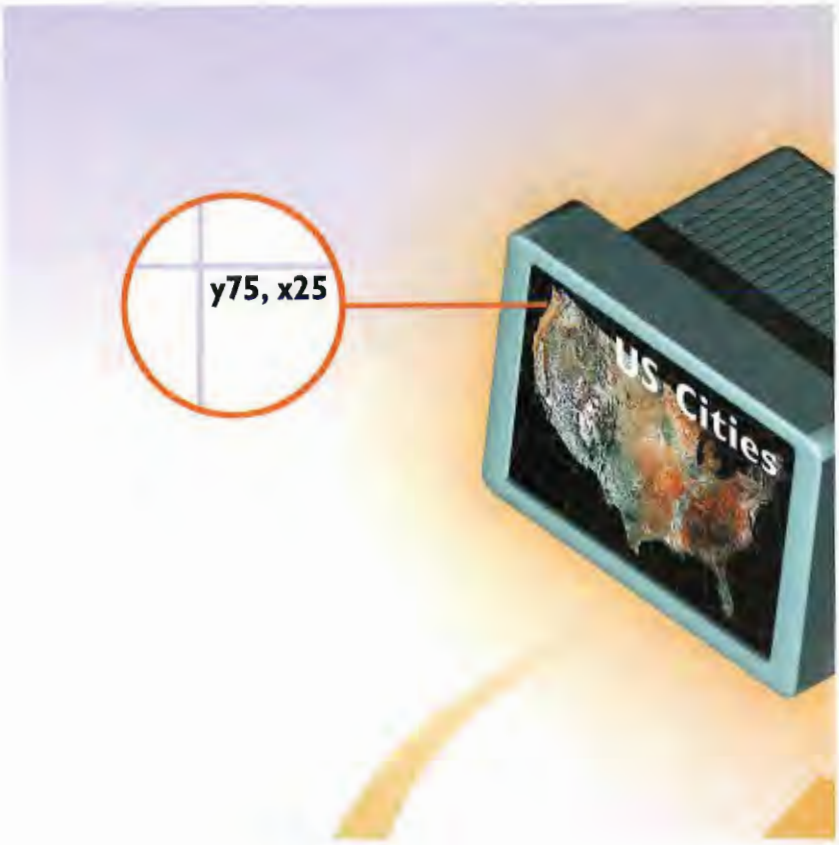
4 If the document is found, the host checks its file type (usually either x-html or x-text) and sends this information to the client with the requested page. When the client receives the page, it first checks the file type. If the type is one it can display, it does so; otherwise, it prompts the user to see whether she wants to save it to disk or open it using a helper application. The x-html file type is by far the most common one used when transmitting Web pages.

3 The server then receives the transmitted URL and responds in one of three ways. It follows the directory path given in the URL; the server finds the file on its local hard disk and opens it; the server runs a CGI script or detects an error (such as `file not found`) and generates an error document to be sent back to the client.

CHAPTER

26

How Image Maps and Interactive Forms Work



 called image maps and functions called interactive forms demonstrate two of the more common and helpful uses of HTML. *Image maps* are static images that have been turned into clickable images with various clickable parts. *Interactive forms* are HTML-based pages you fill out by providing information such as your name, e-mail address, and similar information. Both image maps and interactive forms are created using Common Gateway Interface (CGI), a communications protocol by which a Web server can communicate with other applications. (For more information about how CGI works, turn to Chapter 35, "How CGI Scripting Works.")

Image maps can be thought of as fancy hyperlinks. However, instead of a word, an entire icon, or an image linking to another page, an image is divided into different segments, or coordinates, that link to different HTML pages. That is, image maps link to another document through a predefined *hot* area within an image. As soon as you click your mouse on a hot spot, a CGI script and special image map coordinates file with the suffix .map go to work. A CGI application reads the map file to match coordinates of a mouse click with a corresponding URL. For instance, imagine an electronic map of the United States in which you click Washington, D.C. In the HTML code for that page, the electronic map is surrounded by a tag and an attribute called `ISMAP`. The code looks something like this:

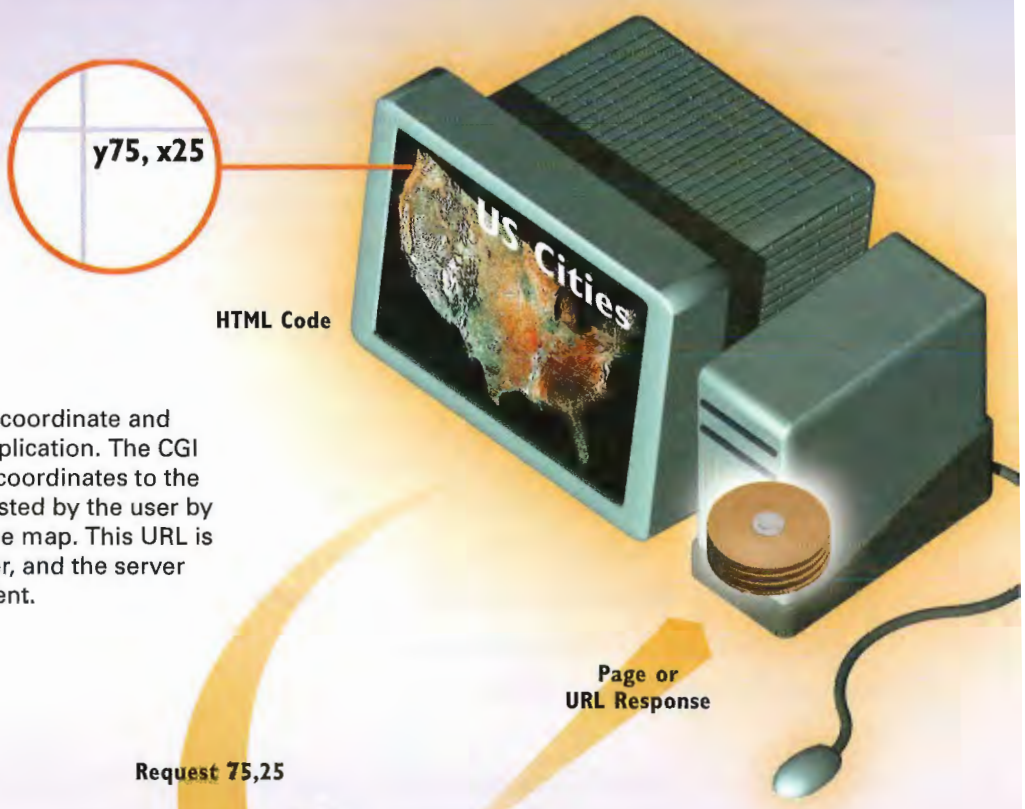
```
<A HREF="some.server/maps/clickable.map>
<IMG SRC+"clickable.map" ISMAP>
</A>
```

The x,y coordinate of your mouse click is sent to the server. The coordinates are received by the server and then redirected to a CGI application. The CGI application scans the file for matching coordinates and then forwards the corresponding URL to the server. Lastly, if the Web page resides on the same server, it delivers that Web page to the client browser. If not, the server returns the URL to the client browser, which in turn sends a request to the correct server for the page. You then see the page about Washington, D.C. begin to load on your browser. Behind the scenes, the server passed your mouse click coordinates to a CGI application via the CGI. Then, the CGI application matched those coordinates to its URL in a .map file. Finally, the URL sent the URL back to the server, which redirected the client browser to the new Web page.

Forms work differently, although they also use CGI. In a form, when you fill in information on a Web page, that information goes to the server for processing. Next, the server redirects the information to a CGI application that is called by the form *submit*. (CGI scripts are activated by the server in response to an HTTP request from the client.) Lastly, a CGI application might send form data to another computer program, such as a database; save it to a file; or even generate a unique HTML document in response to the user's request.

How Image Maps Work

1 In this map example, the user clicks Seattle. The x,y grid coordinate is 75, 25. In the HTML code, the browser recognizes the ISMAP image tag attribute. The mouse click activates the browser to send the x,y coordinate of the click to the server. The location of the `National.map` file is also sent to the server.



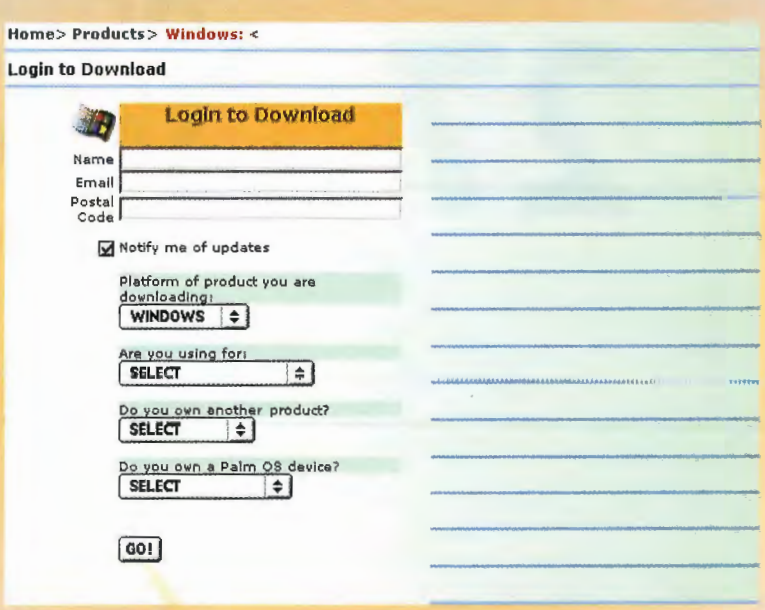
2 The server hands off the coordinate and map file data to a CGI application. The CGI application matches the coordinates to the URL that has been requested by the user by clicking that portion of the map. This URL is handed back to the server, and the server sends the page to the client.

3 The Web document is either served up (if it resides on the same server) or the client browser if forwarded to the new URL.

4 The client browser either displays the returned page or (based on the returned URL) sends a request to the correct server for the page.

How Interactive Forms Work

1 In working with an interactive form, the user clicks a data-entry submit button. The data in the data fields is sent to the server with the request.



2 When a form submission is received by the server, it activates a CGI application, a program, or a script that interacts with a Web server; it then passes the resulting information to a Web form. (The application could add the form data to a database or compare it to a password list of eligible users, among other tasks.) The program's output goes either to another program, such as a database, or into a unique HTML document, or both.

CHAPTER

27 How Web Host Servers Work

TO serve up pages, Web sites need a *host*—a computer—and server software that runs on the host. The host manages the communications protocols and houses the pages and the related software required to create a Web site on the Internet. The host machine often uses the Unix, Windows NT, Linux, or Macintosh operating systems, which have the TCP/IP protocols built in.

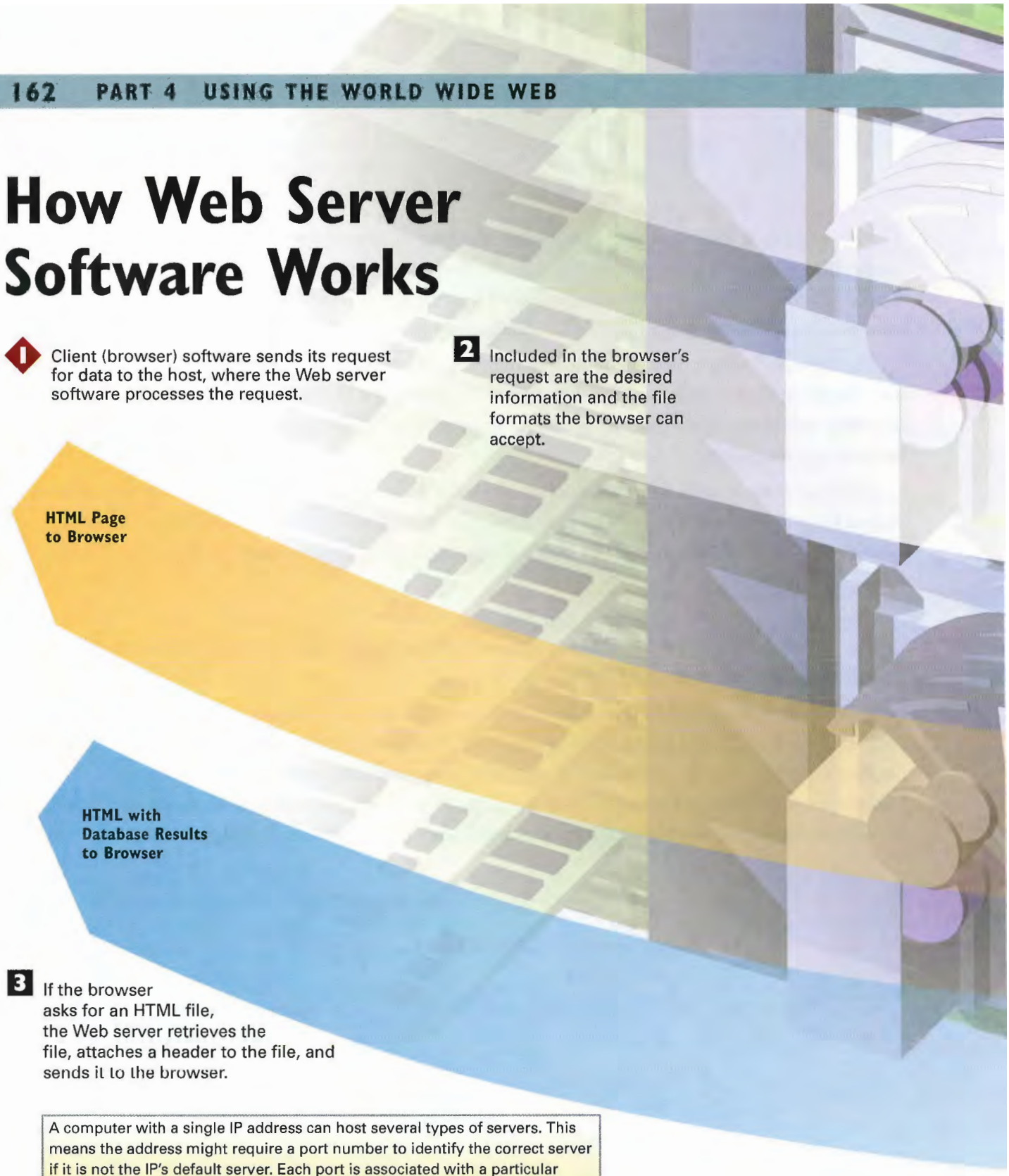
The server software resides on the host and serves up the pages and otherwise acts on the requests sent from the client browser software. The server is not responsible for TCP/IP communications—the host operating system does that—but instead the server handles the HTTP requests and communications with the host operating system.

Different types of server software (database servers or network servers, for example) exist that perform various types of services for various types of clients. Specifically, a *Web server* is an HTTP server, and its function is to send information to the client software (typically a browser) using the Hypertext Transfer Protocol (HTTP).

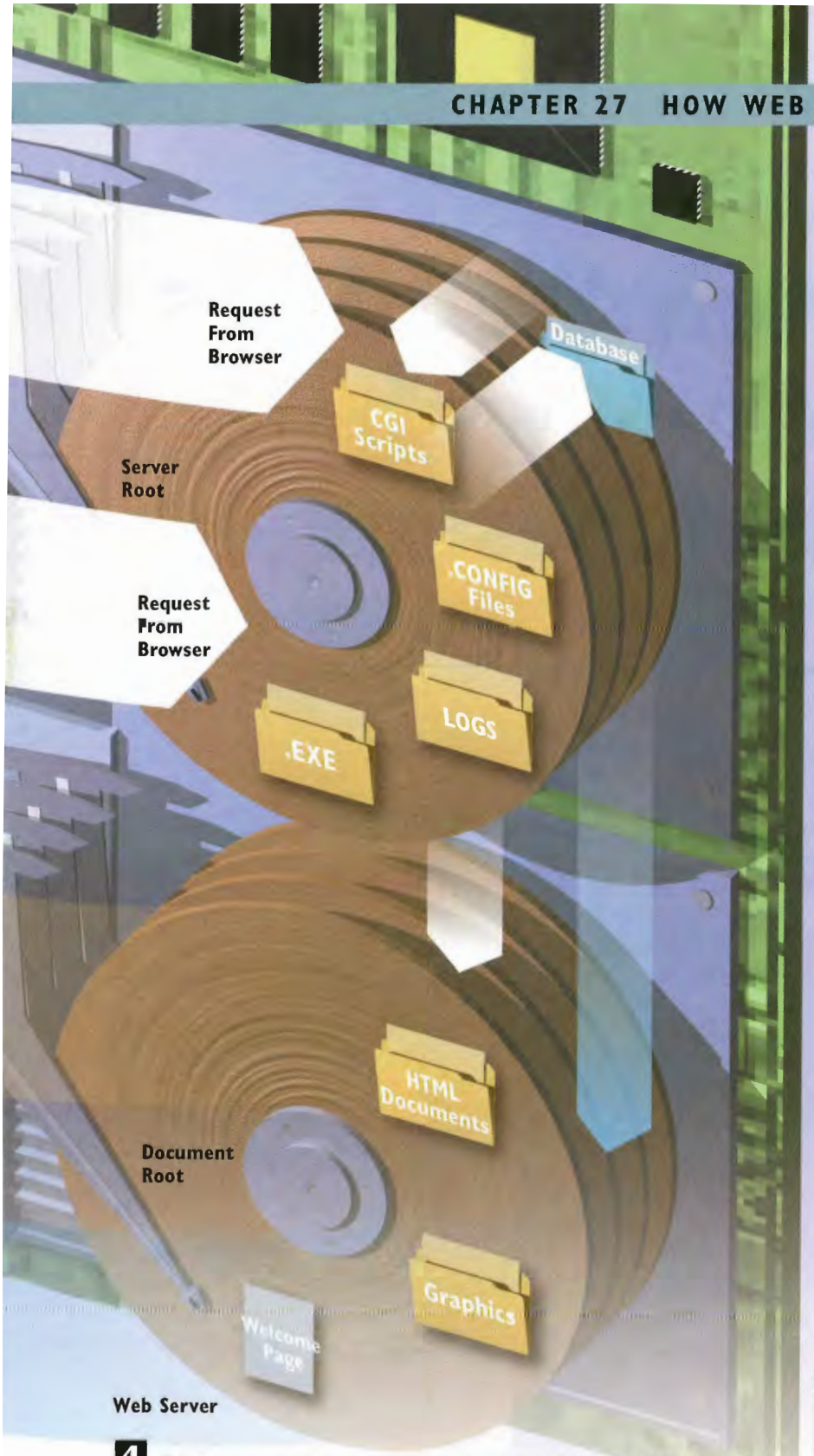
Usually, the client browser requests that the server return an HTML document. The server receives this request and sends back a response. The top portion of the response includes transmission information, and the rest of the response is the HTML file.

A Web server does more than send pages to the browser, however. It passes requests to run Common Gateway Interface (CGI) scripts to CGI applications. These scripts run external mini-programs, such as a database lookup or interactive forms processing. The server sends the script to the application via CGI and communicates the results of the script back to the browser, if appropriate. Moreover, the server software includes configuration files and utilities to secure and manage the Web site in a variety of ways.

How Web Server Software Works



A computer with a single IP address can host several types of servers. This means the address might require a port number to identify the correct server if it is not the IP's default server. Each port is associated with a particular server. Ports are identified by a number from 0 to 65,535, but common server types, such as FTP servers, are given the same number by convention.



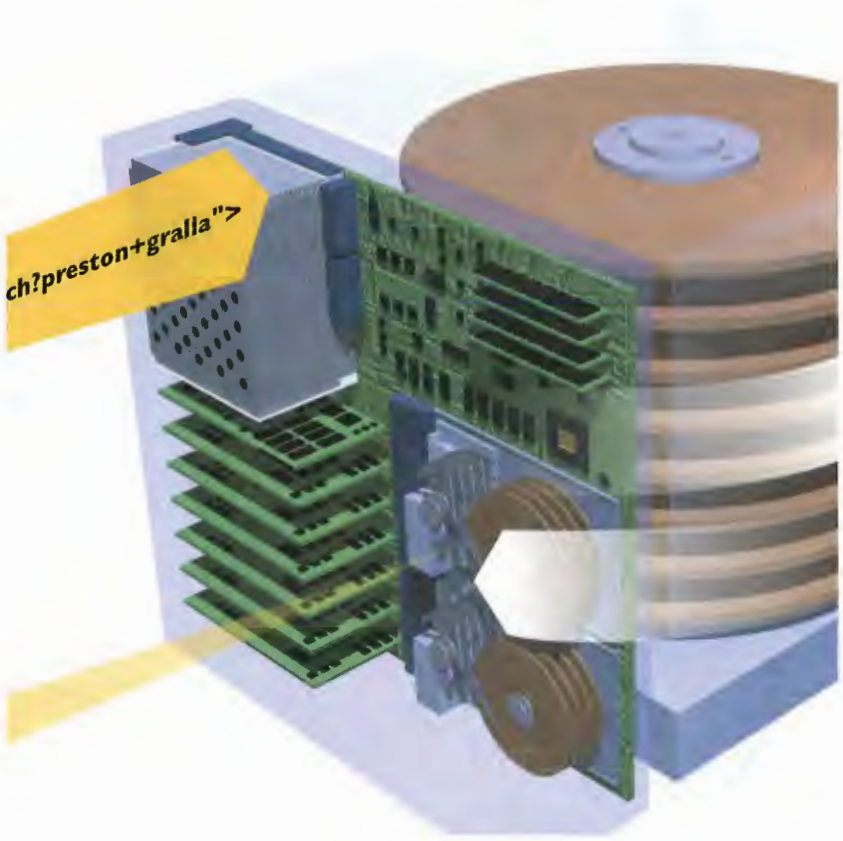
4 If the browser has asked for specific database information, the Web server passes a request through CGI to the application, which performs a database lookup, for example. The CGI script returns the results to the Web server, which in turn attaches a header to the data and sends it to the browser.

The host computer must have a unique IP address to send and receive information across the Internet. Because raw IP addresses are very intimidating, they are assigned a unique domain name, which is less daunting. The domain name is part of a hierarchical lookup system called the Domain Name System (DNS).

IP Address	=	Domain Name
83.382.1.838		books.que.com

CHAPTER

28 How Web Sites Work with Databases



ONE of the most useful applications of the Web is its capability to link a Web site with a database so Web surfers can search for information. In essence, the Web page becomes the front-end for database applications, enabling you to select search criteria and execute even complex searches of a database that resides on the host computer.

A well-known and widely used example of this type of linking between Web sites and databases is the popular Yahoo! Web site (`www.yahoo.com`). The Yahoo! site serves as a front-end to an extensive database of Web site descriptions, which can be searched according to keywords. The Welcome page includes a search dialog box in which you enter a keyword that represents the subject matter you are looking for. Clicking the Search button sends a request from the browser to the Web server to bring back a list of all Web sites that contain your keyword.

Furthermore, not only can the Web serve up data, it can also collect it. For example, many Web sites ask users to "register" their names, addresses, and other demographic information that is captured and stored in a database.

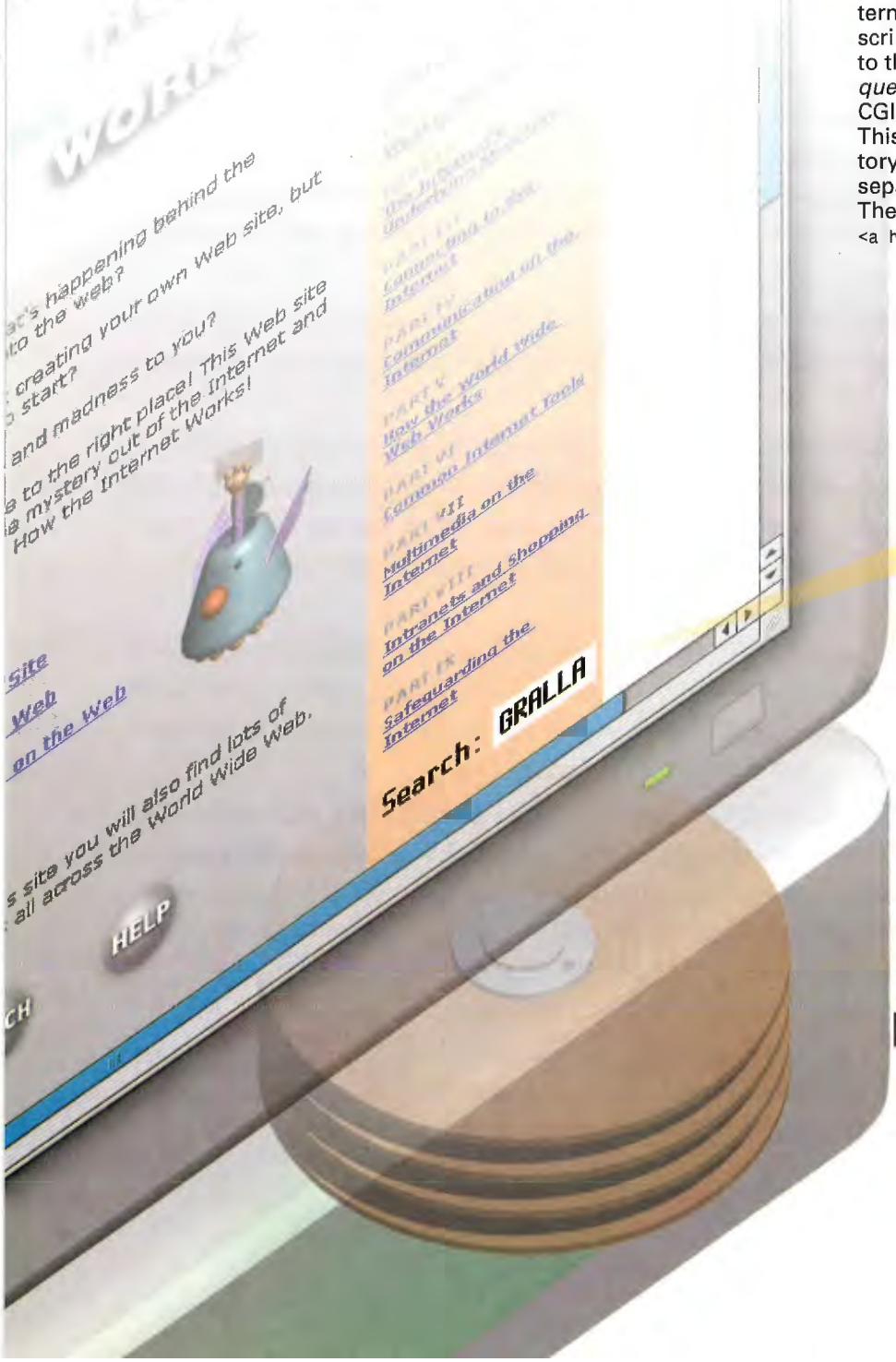
But how does this all work? You don't have to be a corporate giant—or for that matter even an able programmer—to link your Web site to a database. In fact, linking a Web site to a database can be relatively simple. The database can take just about any form and can be as simple as a FileMaker Pro database or as complex as an Oracle SQL database. The bridge that brings together Web sites and databases is the Common Gateway Interface (CGI).

On the client side of the database, you see a Web page that includes a form in which you enter your search terms. When you execute the search, the Web server passes your search information to a CGI script, which then searches the database. So a search on the Yahoo! site for public relations firms looks like this:

```
http://search.yahoo.com/bin/search?p=public+relations
```

When the Web server receives this URL, it identifies the URL as a trigger for a CGI script (called *script* in this example) and passes it along with the search criteria ("public relations," in this example) to the miniprogram using CGI. The CGI script then sends the search to the database, receives the results of the query, and passes it on to the Web server to be sent back to the client. That's a lot of handing off of requests and data, but typically even a search of a large database is very fast because the majority of Unix and Windows NT databases—the types most often used—can perform these tasks simultaneously. All of this happens behind the scenes, of course—you won't need to do any kind of database work or scripting yourself. Instead, the Web sites you visit have easy-to-use interfaces that take care of interacting with databases; you'll have to type only what you're looking for.

How the Web Works with Databases

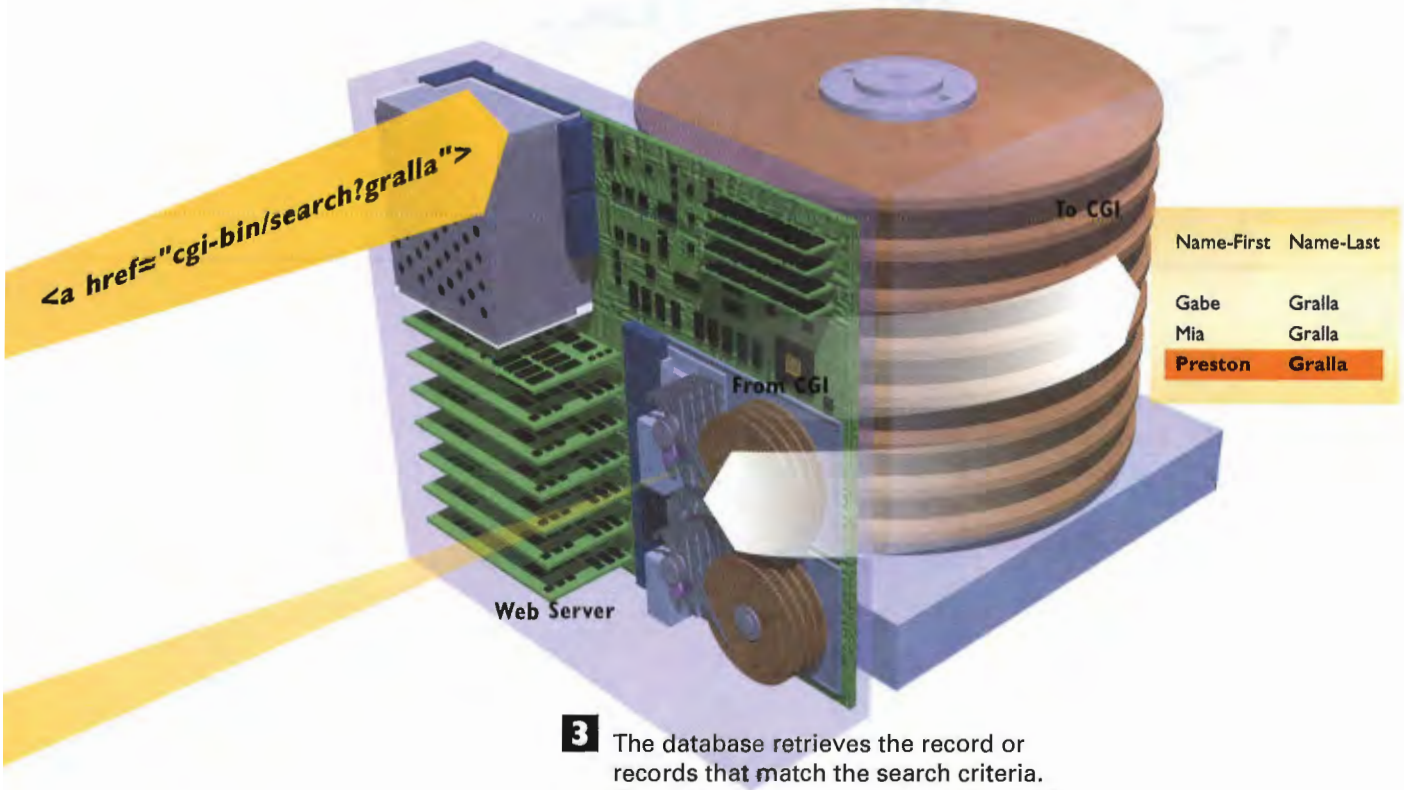


1 The search begins on a Web page that includes a form field to accept search terms and HTML codes to execute a CGI script. The browser might pass the data to the Web server in a query string. The *query string* contains the name of the CGI script in a directory called `cgi-bin`. This directory is followed by a subdirectory that includes the search terms, often separated by a question mark or slashes. The HTML code might look like this:

```
<a href="cgi-bin/search?preston+gralla/">.
```

4 The database returns the data to the Web server via CGI in the form of a new HTML page. The server then sends the page back to the client browser as a new HTML page.

2 When the Web server receives the URL with the embedded search terms, it sends the information through the CGI program to the database. Typically, the program is stored in a unique directory that contains all the CGI scripts serviced by the Web server.

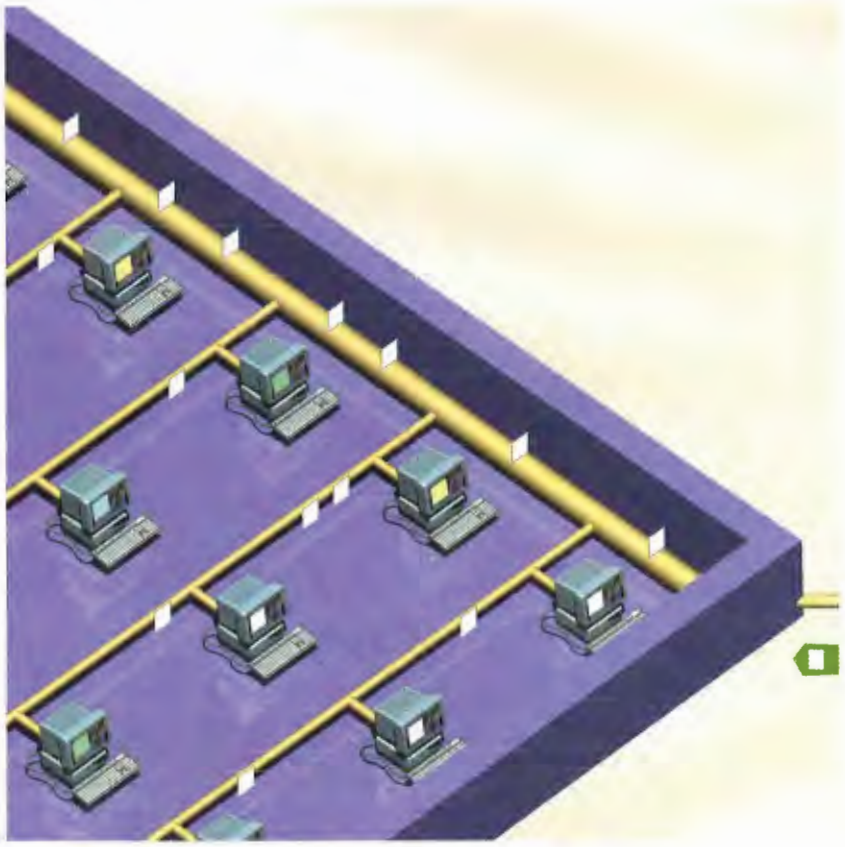


3 The database retrieves the record or records that match the search criteria. The database record might contain text and numeric data as well as references to graphics or other data types.

CHAPTER

29

How Internet-Based Software Works



SOME people believe that in the not-too-distant future, much of the software you run won't physically be found on your computer. Instead, it will be located, at least in part, on an Internet server somewhere. You'll turn on your computer, it will have instant, high-speed access to the Internet, you'll choose software to run from a server, and it will always be up-to-date because the company in charge will continually update it.

To a certain extent, this type of Internet-based computing has a *Back to the Future* kind of feel to it. In the days before the advent of the personal computer, all software was run on a mainframe or large minicomputer, and people accessed it via a dumb terminal that had no processing power itself. In some ways, Internet-based software goes back to that model. The difference is that instead of dumb terminals, people have powerful computers that can still do a lot on their own. And not many people expect that all software will be run this way.

No matter what the future holds, today some software is already being run that way. The software is being provided by application service providers (ASPs), which have a variety of ways of letting you run software over the Internet. The most common way is one in which you visit a Web site, and the software runs inside your browser. But there are variations as well, such as services that run Web servers, while you or your business runs client software that requires the use of those servers.

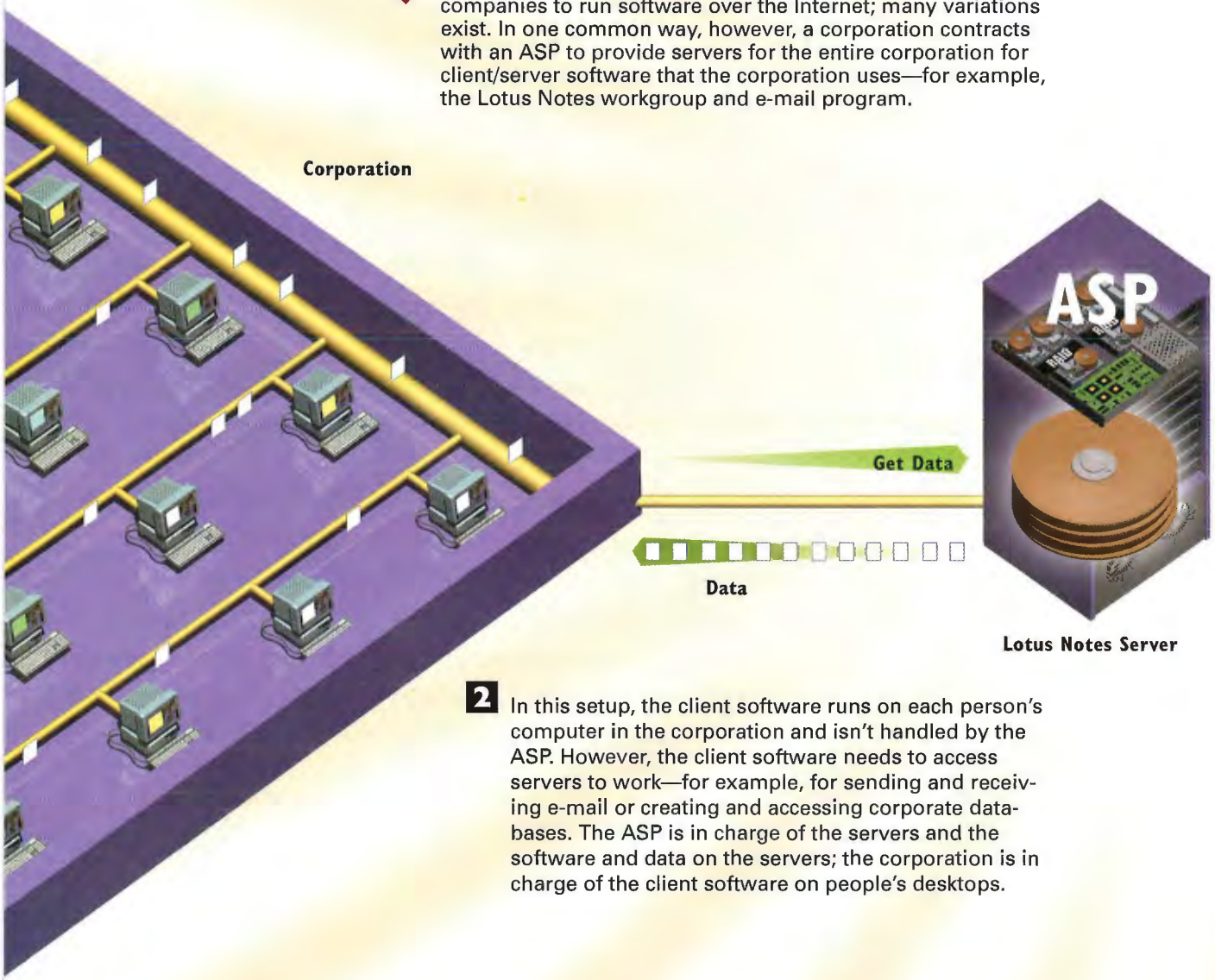
You might already be using an ASP without realizing it. For example, Web-based e-mail services, such as HotMail, in fact are forms of ASPs—in essence they allow you to run an e-mail program from within your browser.

Corporations turn to ASPs because they enable them to do business more cheaply. Rather than having to build a large infrastructure of internal servers or have a large support staff for installing and troubleshooting software and hardware, they contract with an ASP and let the ASP bear all those costs.

It's not only corporations, though, that eventually might turn more to ASPs. Microsoft has announced its vision for the future of software in its .Net plan, and that plan includes ASP-like features. At this point, the details of .Net are quite hazy. But what's come out so far indicates that a good part of it will include a far tighter relationship between the software on your computer and servers on the Internet, and it might even include software running similarly to ASPs.

How Application Service Providers Work

1 There is no single, standard way that ASPs enable people and companies to run software over the Internet; many variations exist. In one common way, however, a corporation contracts with an ASP to provide servers for the entire corporation for client/server software that the corporation uses—for example, the Lotus Notes workgroup and e-mail program.



2 In this setup, the client software runs on each person's computer in the corporation and isn't handled by the ASP. However, the client software needs to access servers to work—for example, for sending and receiving e-mail or creating and accessing corporate databases. The ASP is in charge of the servers and the software and data on the servers; the corporation is in charge of the client software on people's desktops.

3 Another type of ASP is completely Web based and can be used by individuals as well as corporations. It's a way for people to run software directly in their Web browsers, so they don't have to have any software on their own computers. In this instance, we'll look at how individuals use ASPs. To use the software, someone visits a Web site that allow her to run an application—for example, a piece of personal finance software or an e-mail program.

Personal Finance ASP Server

Get my finance data.

ActiveX

4 The Web site delivers the software to the person's Web browser. The software can be created using many kinds of tools, such as ActiveX or Java.

5 The software runs inside the person's browser. Depending on the software, it might allow the person to save data to her own PC, or she might be required to save it to the Web site. When the person leaves the Web site, the software stops running, and it leaves no traces behind on the person's computer.

6 Another type of ASP requires someone to download a small piece of software to her computer. It's a kind of "helper" software—it's not the application itself, but instead helps the application run.

Lotus Notes Server

Connect

Helper

7 After the person downloads the helper, she chooses the actual application she wants to run—a graphics program, for example, or a piece of personal finance software. The helper application goes out to a Web site and downloads a core part of the application to the person's hard disk.

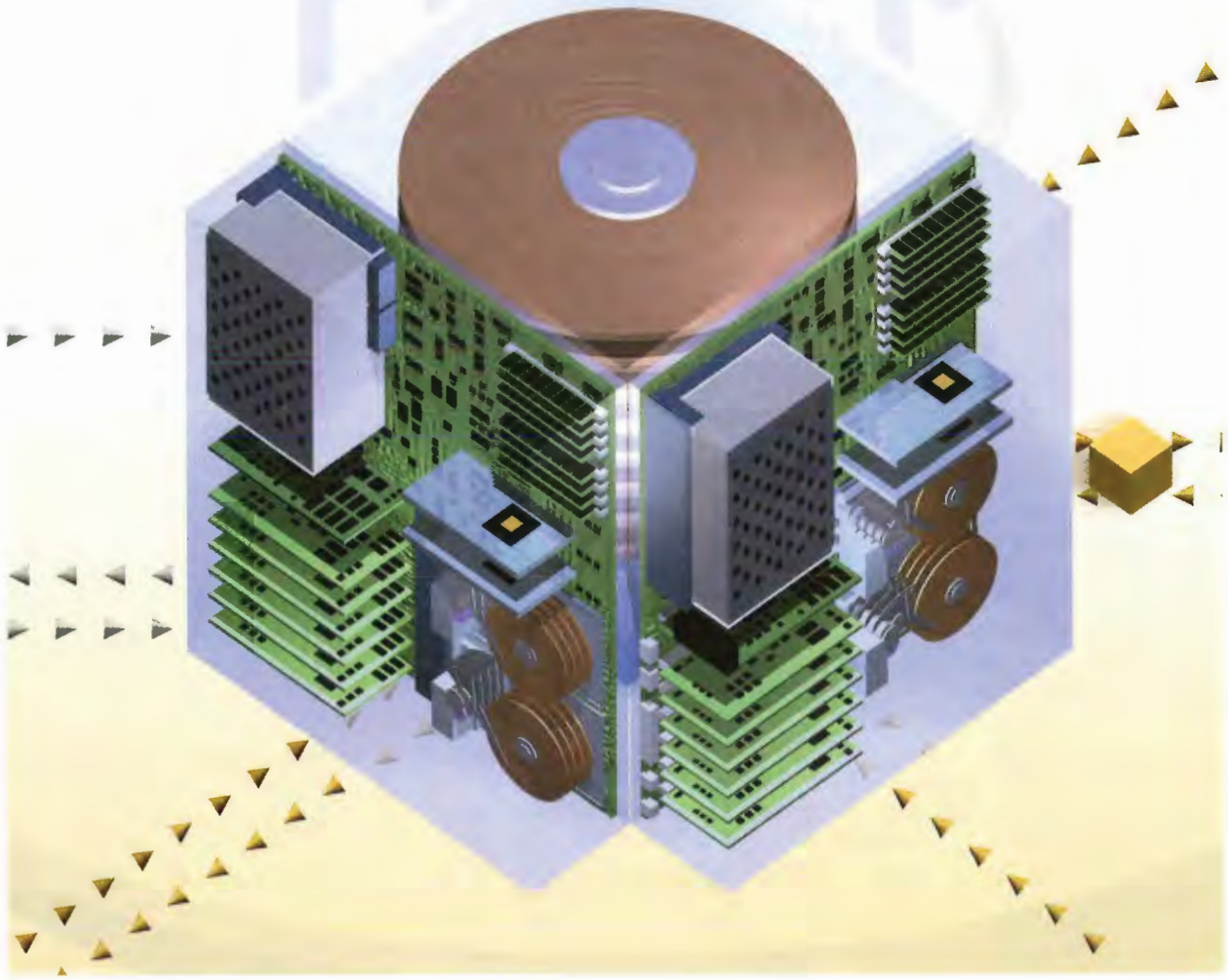
Get application.

Here it is.

Application

Application downloads and then launches.

8 The person now runs the application on her own computer. If she needs other elements of the application other than the core that was downloaded to her computer, the helper application goes out and gets it. When the person stops running the application, the helper program deletes it from the hard disk, leaving behind no trace.



PART 5

USING COMMON INTERNET TOOLS

AN enormous amount of information and entertainment is available on the Internet, but how do you access it? Although using the Internet gets easier every day, it's still not quite as simple as turning on your television or reading your daily newspaper.

The solution is to use a variety of Internet tools. These tools enable you to tap into the colossal resources of the Internet. Some of these resources, such as the World Wide Web, are quite well known. Others, such as FTP (file transfer protocol) are used quite often and sometimes people use them without even knowing it. Still others, such as Telnet, are not nearly as popular, although they are still useful. Many of these Internet tools predate the Web, but they are still useful today.

For many people, the term "Internet" really means the World Wide Web, but as this section of the book shows, a world exists well beyond the Web. (Turn to Part 4, "Using World Wide Web," for information about the Web—the fastest growing and most visible part of the Internet.) And this section of the book also shows you some of the advanced underlying technologies that make the Internet and the World Wide Web a richer, more interactive, more entertaining, and more productive medium. Many of these technologies have changed the very nature of the Internet and have turned it into a truly interactive medium—one that people can navigate efficiently. The technologies also enable Web publishers and Internet developers to more effectively present information to people.

This section looks at how the most common and useful Internet tools work.

Chapter 30, "How Telnet Works," covers one of the older Internet technologies, and one that is still in widespread use—Telnet. Telnet enables you to take over the resources of a distant computer while sitting at your own computer. What you type on your keyboard is sent across the Internet to the distant computer, the commands are carried out by the distant computer, and the results of your commands are sent to your own computer screen. It appears as if you're sitting at the distant computer's keyboard. Telnet is used in many ways, notably by libraries making their catalogs available over the Internet. When you log in to a distant computer using Telnet, you often use a menuing system.

Chapter 31, "How FTP Downloading Works," covers one of the most popular uses of the Internet—downloading files. Generally, files are downloaded from the Internet using FTP, the Internet protocol. Not only will you look at how FTP works, but you'll also look at how files are compressed and decompressed on the Internet. A compressed file takes less time to be sent over the Internet to your computer. You might not know it, but many times when you're on a Web site and download a file, you're actually using the FTP protocol.

Chapter 32, "How Internet Searching Works," examines Internet search engines. The Internet contains such a vast amount of information that it's often impossible to find exactly what you want. Search engines look through the entire Internet—not only Web pages, but other sites such as newsgroups—and find information you're looking for, based on keywords you type.

Chapter 33, "How Agents Work," looks at agents on the Internet. *Agents* are programs that do your bidding across the Internet automatically, without you doing anything. They can find the latest news and download it to your computer; they can find you the best deal on the

CD you want to buy; they can perform important Web maintenance tasks; and more. They are becoming so complex that systems are being developed to enable agents to interact with one another so they can perform jobs cooperatively.

Chapter 34, "How Java, ActiveX, and JavaScript Work," examines three other types of technologies that are transforming the Internet—Java, JavaScript, and ActiveX. These three technologies might do more to transform the Internet than almost any other technologies currently available. These technologies add multimedia and interactivity, but more importantly, they begin to treat the Internet as if it were an extension of your computer. In essence, they enable your computer and the Internet to interact as if they were one large computer system. This enables things such as news tickers, interactive games you can play with others, multimedia presentations combining animations, sounds, music, graphics, and much more.

Java, a computer language developed by Sun Microsystems, enables applications to be run from the Internet. The programs run inside your Web browser. One benefit of Java applications is that they can be run on any computer, such as a PC, a Macintosh, or a Unix workstation.

ActiveX, a competing technology from Microsoft, can also essentially turn the Internet into an extension of your computer. Similar to Java applets, ActiveX controls are downloaded to your computer and run there. They can do anything a normal application can do and can also interact with the Web, the Internet, and other computers connected to the Internet. To run them, a browser that supports ActiveX, such as Internet Explorer, is necessary.

JavaScript, which despite its name is not really related to Java, is simpler than Java and ActiveX and can be written by people who don't have substantial programming experience. JavaScript is commonly used to create interactive forms, site navigation, and similar features.

Finally, Chapter 35, "How CGI Scripting Works," examines CGI (Common Gateway Interface) scripting. This might appear as one of the more mundane Internet technologies, but without it, very little Web interactivity would take place. CGI is a standard way in which the Web interacts with outside resources—most commonly, databases. You've probably run CGI scripts many times without knowing it. If you've filled out a form on a Web page to register to use a site and then later received an e-mail notification with a password for you to use, you've probably run a CGI script. CGI enables programmers to write code that can access information servers (such as Web servers) on the Internet and then send the information to users.

CHAPTER 30

How Telnet Works

ONE of the more remarkable features of the Internet is the way it lets you use the resources of a distant computer somewhere else in the world. From your own home or office, you can log onto another computer, issue commands just as if you were at that computer's keyboard, and then gain access to all the computer's resources. You do this with an Internet resource called Telnet. Telnet follows a *client/server model*, which means that you run a piece of software on your own PC (the client) to use the resources of a distant server computer. This distant computer is called the *host*.

The host allows many clients to access its resources at the same time; it isn't devoted to a single user. To use Telnet and the host's resources, you must know the address of the Internet host whose resources you want to access.

When you use Telnet, before you can take over the resources of a host computer you typically have to log onto the host. Often, you can use the name "guest" to log on. Some systems require that you also give information about yourself, such as your name and address. And some might require that you choose a username and a password that you will use the next time you log in.

You can access many hosts on the Internet by using Telnet. They are all different computers, so many of them don't work or look alike. For example, some might be Unix-based systems, some might be NT-based computers, and some might be Macintoshes, as well as a variety of other computers, and they all work and look different from one another. As a way to make things easier, many hosts use a menuing system that gives you access to their resources.

Telnet gives you a way to use those menuing systems by using something called *terminal emulation*. It lets you use your computer to emulate the type of keyboard and computer that each of the different computer systems expect. Different computers often require different kinds of terminal emulation, but one common kind is called VT-100 emulation, so if you use Telnet software and tell it to use VT-100 emulation, that's a safe emulation to use.

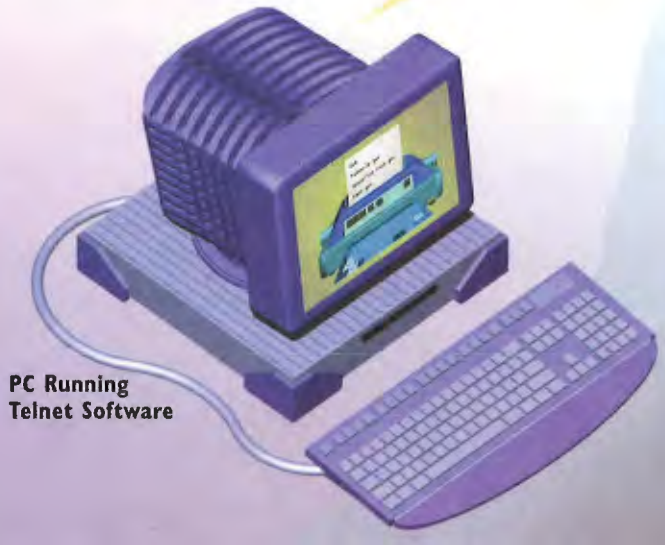
Telnet clients are available for all the major operating systems, including Linux, Unix, Macintosh, and all versions of Windows. If you use an Internet shell account instead of a SLIP/PPP connection, you'll typically use a Telnet client by simply typing the word **`Telnet`** followed by the Internet address of the computer you want to access. For example, if you wanted to gain access to a computer run by the federal government called Fed World that lets you access a great deal of government information, you'd type **`Telnet fedworld.gov`**.

A Windows- or Macintosh-based Telnet client is easier to use than a DOS- or Unix-based Telnet client because the former remembers hostnames for you. With clients, you can often keep an address book of hostnames so you can easily revisit them.

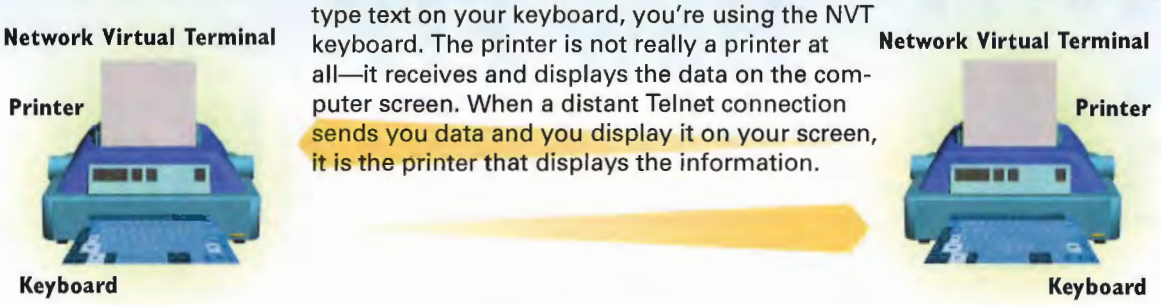
Understanding Telnet

1 To use Telnet, you need to know the Internet address of the host whose resources you want to use; your Telnet client contacts the host, using its Internet address.

2 When you contact the host, the distant computer and your computer negotiate how they will communicate with each other. They decide which terminal emulation will be used. Terminal emulation determines how your keyboard will transmit information to the distant computer and how information will be displayed on your screen. It determines, for example, things such as how certain keys like the backspace key will work. VT-100 is the most common type of terminal emulation.



3 When a client and a server communicate, they use the Telnet protocol. The Telnet protocol assumes that each end of the connection—the client and the server—is a network virtual terminal (NVT). Each NVT has a virtual "printer" and a virtual "keyboard" The keyboard sends data from one NVT to the other. When you type text on your keyboard, you're using the NVT keyboard. The printer is not really a printer at all—it receives and displays the data on the computer screen. When a distant Telnet connection sends you data and you display it on your screen, it is the printer that displays the information.



4 Typed text in a Telnet session accumulates in a buffer on your computer. When a complete line of data is ready for transmission, or when you give a command to transmit data (such as pressing the Enter key), the data is sent across the Internet from your NVT keyboard. Along with the data is the host's IP address, which ensures the packet is sent to the proper location.

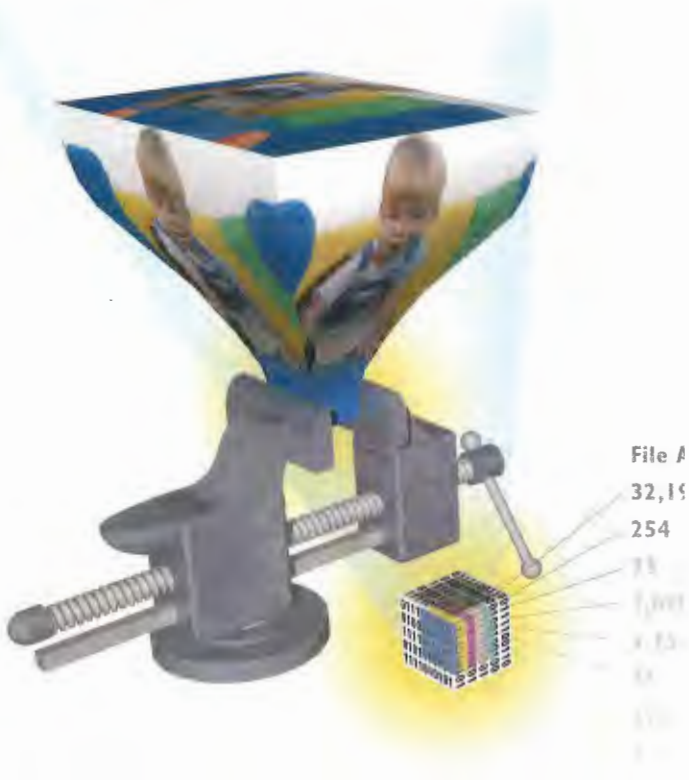
5 Your IP address is also sent, so that information can be routed back to you. Additionally, specific Telnet commands are sent that the other NVT uses to decide what to do with the data or how to respond to the data. For example, when data is sent from one NVT to another, and certain information must be sent back to the originating NVT for a process to proceed, the Telnet Go Ahead (GA) command is sent.

6 The Telnet host receives the data you've sent. It processes the data and returns to your screen (your NVT printer) the results of using the data or running the command on a distant computer. So, for example, if you type a series of keys with the letters `dir` and press Enter, the distant computer carries out the `dir` command. That computer also returns to your screen the `dir` command and sends the results of running that command on the distant computer.

7 Because packets must go through many Internet routers in each direction between your computer and the host, a delay might occur between the time you send a command and the time you see the results on your own computer screen.

CHAPTER

31 How FTP Downloading Works



ONE of the most popular uses of the Internet is to download files—that is, transfer files from a computer on the Internet to your computer. These files can be of many types: programs that you can run on your own computer; graphics you can view; sounds and music you can listen to; or text files that you can read. Many tens of thousands of files are downloaded every day over the Internet. Frequently, using the Internet's File Transfer Protocol, commonly referred to as FTP. You can also use FTP to upload files from your computer to another computer on the Internet.

FTP, like many Internet resources, works on a client/server model. You run FTP client software on your computer to connect to an FTP server on the Internet. On the FTP server, a program called an FTP daemon (pronounced "demon") allows you to download and upload files.

To log on to an FTP site and download files, you must type in an account number (or username) and a password before the daemon will allow you to enter. Some sites allow anyone to enter and download files, but an account number (or username) and password must still be entered. Often, to get in, you use `anonymous` as your username and your e-mail address as your password. Because of this, these sites are often referred to as *anonymous FTP sites*. Some FTP sites are private and allow only certain people with the proper account number and password to enter.

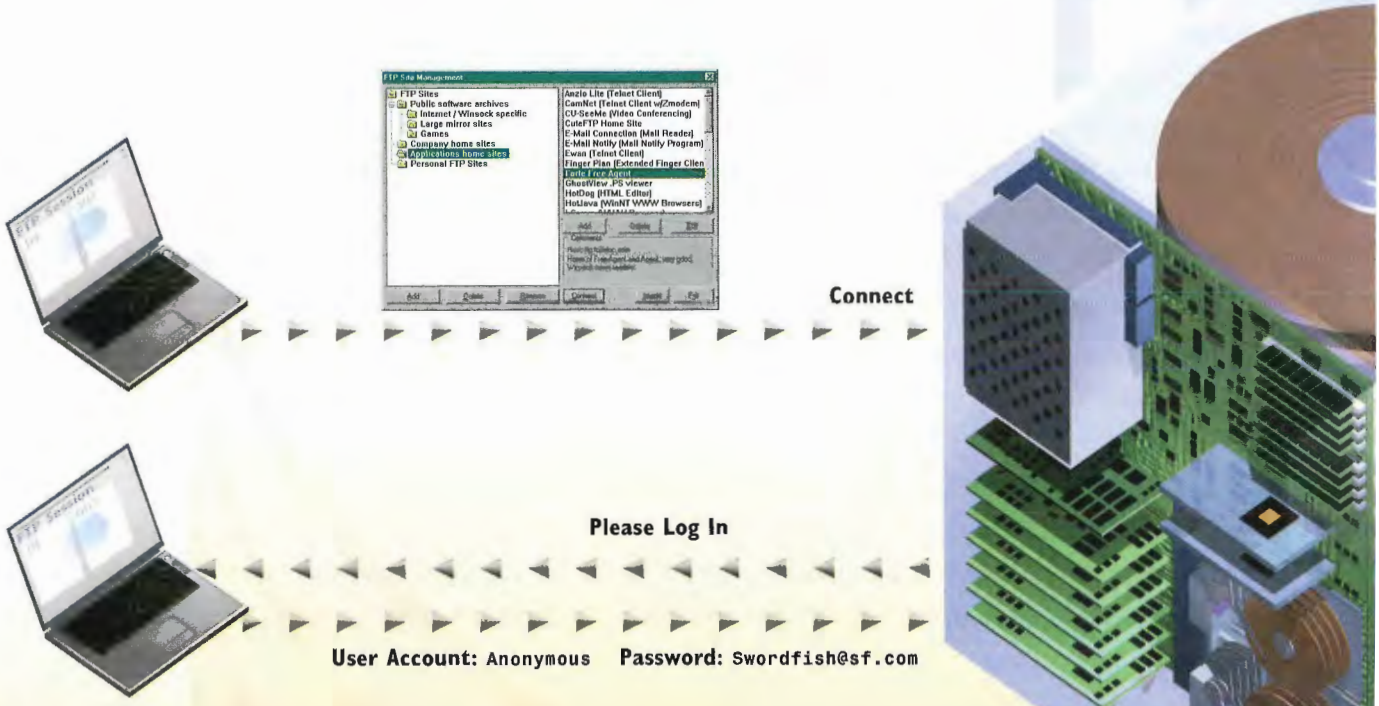
FTP is fairly simple to use. When you log on to an FTP site, you can browse through the available files by changing directories and seeing a listing of all the files available in each directory. When you see a file you want to download, use your client software to instruct the FTP server to send you the file.

As the World Wide Web gained popularity, downloading software became even easier. You can use your Web browser and click links to files. Behind the scenes, FTP is often still downloading the files. FTP remains the most popular way to download files from the Web and the Internet. The HTTP protocol of the Web can be used for downloading files from the Web, but it's not as efficient as FTP, so it isn't used as frequently.

One problem with downloading files over the Internet is that some files are so large that it can take a tremendous amount of time to download them; especially if the connection is made via modem. Even at 56Kbps, downloading files can be slow. As a way to speed up file transfers and save space on the FTP server, files are commonly compressed, or shrunk in size using special compression software. Many different methods are used to compress files. Depending on the file type, files are usually compressed from 10%–50%. After downloading the files, you'll need to run the compression software on your own computer to decompress the files so you can use them.

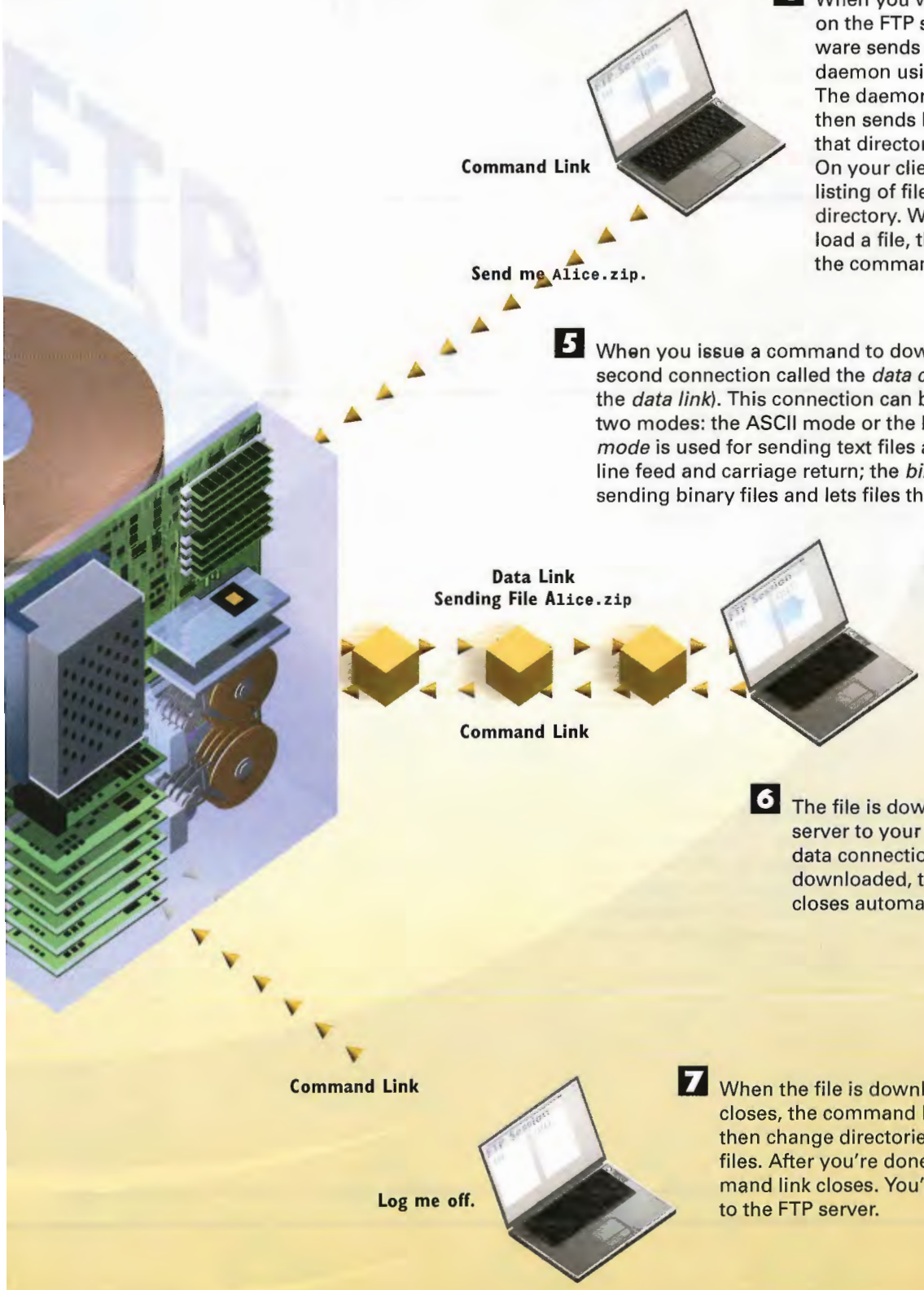
How an FTP Session Works

1 FTP, like many other Internet services, runs on a client/server model. To use it you'll need client software on your computer. To begin an FTP session, run the FTP client software and contact the FTP server from which you want to download files. You can get FTP client software in hundreds of places on the Internet, such as ZDNet Downloads at `www.zdnet.com/downloads`. A command-line FTP program is also included on Windows-based computers, but is much harder to use than these FTP clients.



2 The FTP daemon runs on the FTP server. This daemon handles all FTP transactions. When an FTP client contacts a server, the daemon will ask for an account number (or username) and password. Many FTP sites let anyone log on to them to download files and software. This is called anonymous FTP. With anonymous FTP, you often use `anonymous` for your account number and your e-mail address for your password. Note that some FTP clients will automatically log on to the FTP server for you when you connect so you won't be asked to log on.

3 When you log on to the FTP server, a connection called a *command link* is opened up between your computer and the server. Your computer uses this link for sending commands to the server, and the server uses this link for sending messages and information back to your computer.



4 When you want to change directories on the FTP server, your client software sends an instruction to the FTP daemon using the command link. The daemon changes directories and then sends back a listing of files on that directory via the command link. On your client software, you'll see a listing of files in the new server directory. When you want to download a file, the request is issued over the command link.

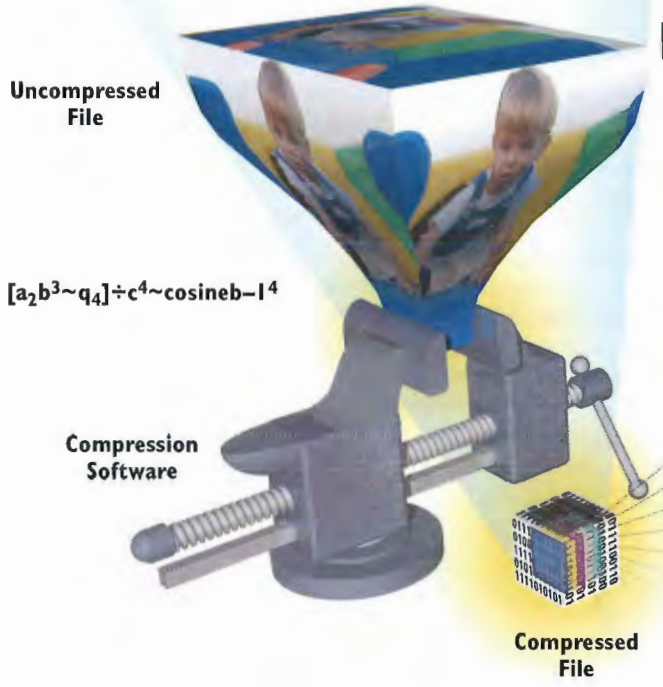
5 When you issue a command to download a file, you open a second connection called the *data connection* (also called the *data link*). This connection can be opened up in one of two modes: the ASCII mode or the binary mode. The *ASCII mode* is used for sending text files and alters things such as line feed and carriage return; the *binary mode* is used for sending binary files and lets files through untouched.

6 The file is downloaded from the server to your computer via the data connection. After the file is downloaded, the data link closes automatically.

7 When the file is downloaded and the data link closes, the command link stays open. You can then change directories or download more files. After you're done, log off and the command link closes. You're no longer connected to the FTP server.

How File Compression Works

1 Compression programs use *algorithms*—complex mathematical formulas—to shrink files. In the first step in the process, the algorithm examines the file to be compressed and looks for repeating patterns of data.



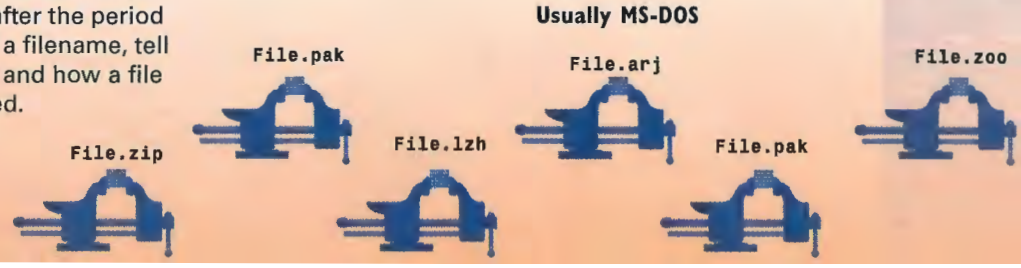
2 When the algorithm finds patterns of data that repeat, it replaces the patterns with smaller *tokens*. In a file that has many repeating patterns, many tokens are used to replace data so the compressed file is much smaller than the original file.

File Analysis

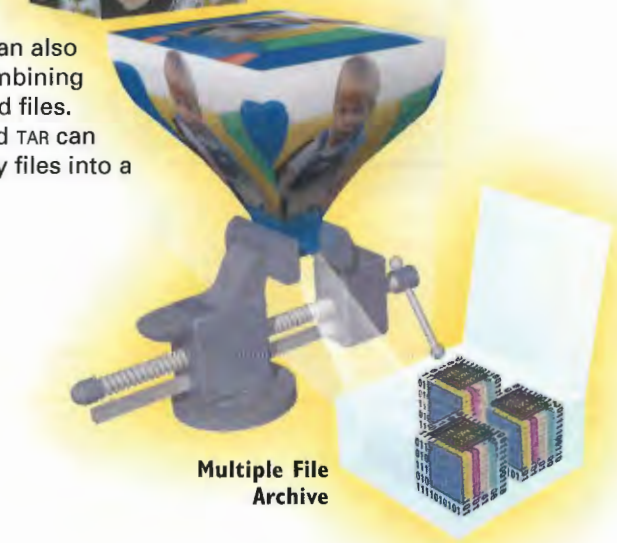
32,196	token A
254	token B
73	token C
1,098	token D

3 A *header* can also be added to the file as it is compressed. This header contains information about the file, such as the filename, the file size, and the compression method used. This information is used to help reconstruct the file when it is uncompressed.

7 File *extensions*, the letters that appear after the period at the end of a filename, tell you whether and how a file is compressed.



4 Some compression software, such as PKZIP for the PC, can also archive files by combining several compressed files. The Unix command TAR can also combine many files into a single archive.



Multiple File Archive

5 When you want to use a compressed file you find on the Internet, transfer it over the Internet to your computer.

Example .tif

6 To use the file, you'll need decompression software on your computer. The decompression software looks into the file's header and examines the tokens in the file. The decompression software uses a decompression algorithm to reconstruct the original file which you can then use on your computer.

CHAPTER

32

How Internet Searching Works

SO much information is available on the Internet, but there is so little organization to the Internet that it can seem impossible to find the information or documents you want. A number of solutions have sprung up to solve the problem. The two most popular ones are indexes and search engines.

Indexes present a highly structured way to find information. They enable you to browse through information by categories, such as arts, computers, entertainment, sports, and so on. In a Web browser, you click a category, and you are then presented with a series of subcategories. Under sports, for example, you'll find baseball, basketball, football, hockey, and soccer. Depending on the size of the index, several layers of subcategories might be available. When you get to the subcategory you're interested in, you are presented with a list of relevant documents. To get to those documents, you click the links to them. Yahoo! (`http://www.yahoo.com/`) is the largest and most popular index on the Internet. Yahoo! and other indexes also enable you to search by typing words that describe the information you're looking for. You then get a set of search results—links to documents that match your search. To get the information, you click a link.

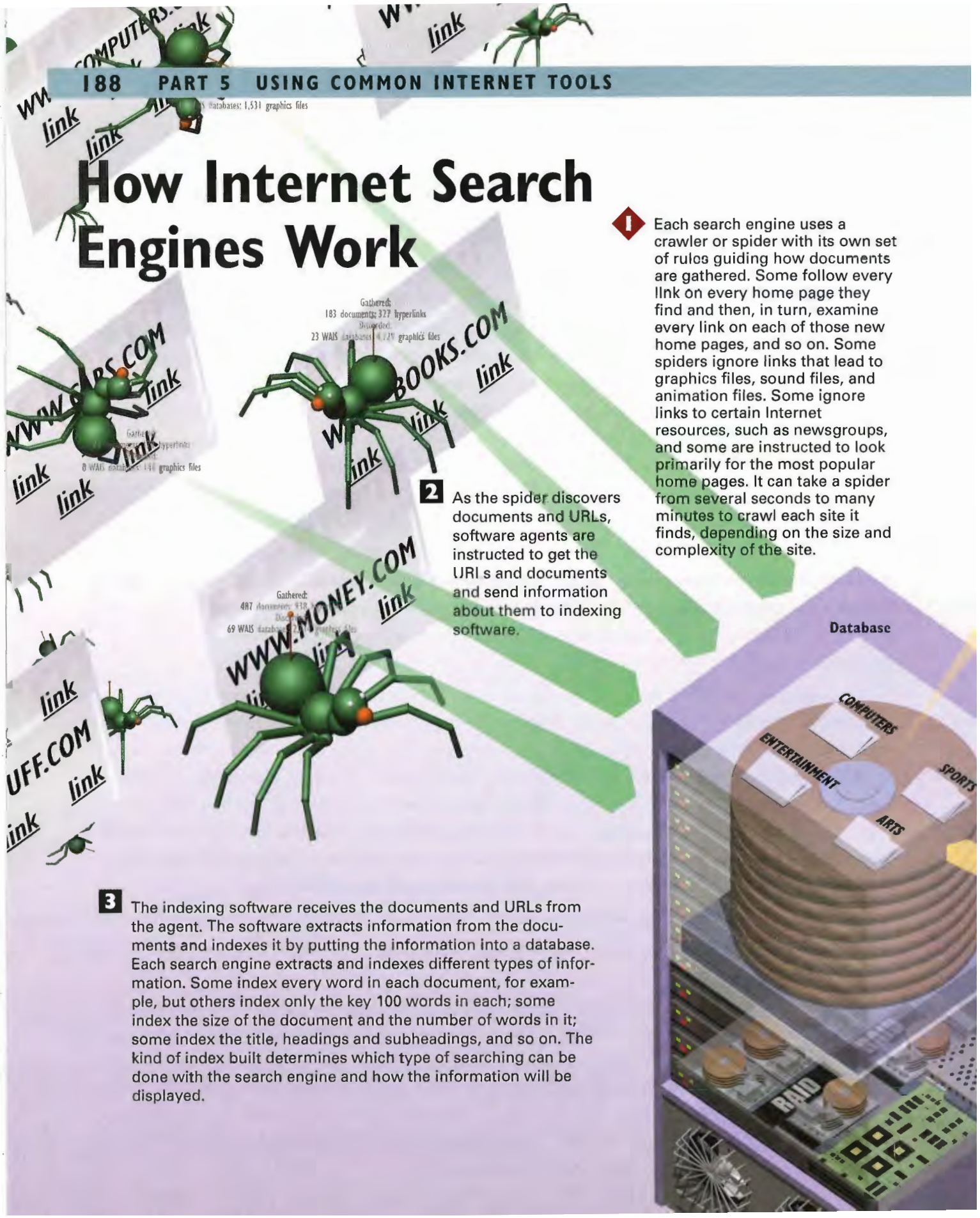
Another popular way of finding information is to use *search engines*, also called *search tools* and sometimes called *Web crawlers* or *spiders*. Search engines operate differently from indexes. They are essentially massive databases that cover wide swaths of the Internet. Search engines don't present information in a hierarchical fashion. Instead, you search through them as you would a database, by typing keywords that describe the information you want.

There are many popular Internet search engines, including Google, Lycos, Excite, and AltaVista. Although the specifics of how they operate differ somewhat, generally they are all composed of three parts: at least one spider, which crawls across the Internet gathering information; a database, which contains all the information the spiders gather; and a search tool, which people use to search through the database. Search engines are constantly updated to present the most up-to-date information, and they hold enormous amounts of information. Search engines extract and index information differently. Some index every word they find in a document, for example, and others index only the key 100 words in each document. Some index the size of the document; some index the title, headings, subheadings, and so on.

Additionally, each search engine returns results in a different way. Some weigh the results to show the relevance of the documents; some show the first several sentences of the document; and some show the title of the document as well as the URL.

Many search engines and indexes are on the Internet, each with its own strengths and weaknesses. To cast the widest possible net when looking for information, you should search as many of them as you can. The problem is that doing so is too time-consuming. So a type of software called *meta-search* software has been developed. With this software, such as Copernic, you type a search on your own computer. The software then submits the search to many Internet search engines and indexes simultaneously, compiles the results for you, and then delivers the results to your computer. To visit any resulting site, just click the link, the same as if you were on an index or a search engine site.

How Internet Search Engines Work

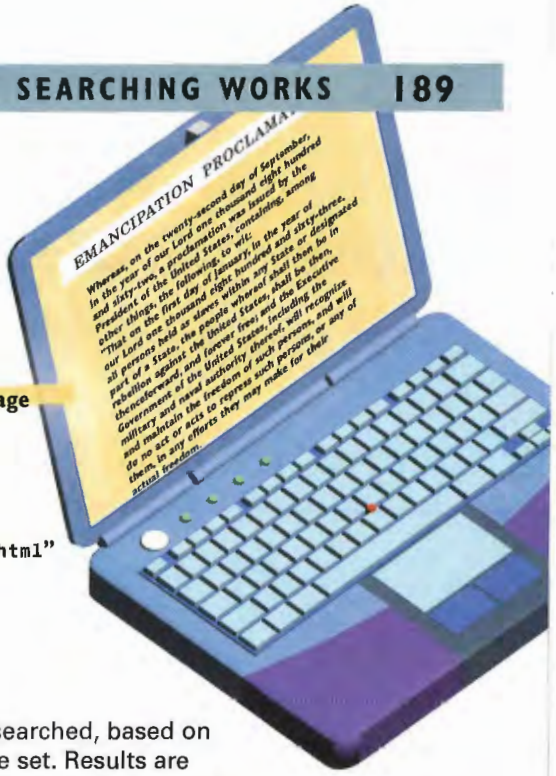


1 Each search engine uses a crawler or spider with its own set of rules guiding how documents are gathered. Some follow every link on every home page they find and then, in turn, examine every link on each of those new home pages, and so on. Some spiders ignore links that lead to graphics files, sound files, and animation files. Some ignore links to certain Internet resources, such as newsgroups, and some are instructed to look primarily for the most popular home pages. It can take a spider from several seconds to many minutes to crawl each site it finds, depending on the size and complexity of the site.

2 As the spider discovers documents and URLs, software agents are instructed to get the URLs and documents and send information about them to indexing software.

3 The indexing software receives the documents and URLs from the agent. The software extracts information from the documents and indexes it by putting the information into a database. Each search engine extracts and indexes different types of information. Some index every word in each document, for example, but others index only the key 100 words in each; some index the size of the document and the number of words in it; some index the title, headings and subheadings, and so on. The kind of index built determines which type of searching can be done with the search engine and how the information will be displayed.

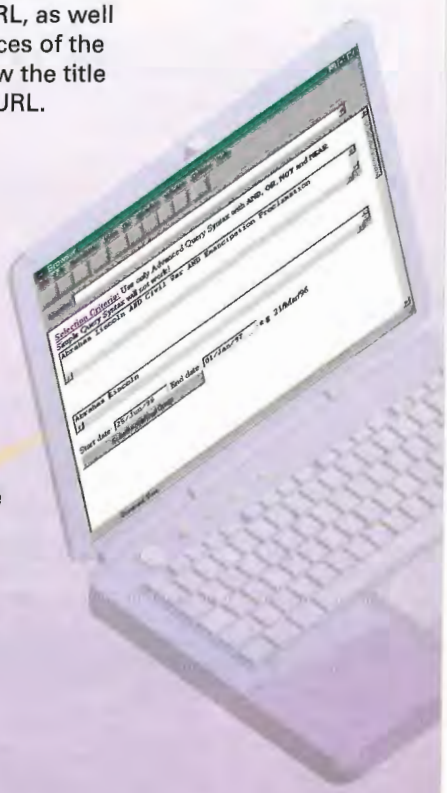
6 When you click a link to one of the documents that interest you, you're sent straight to that document. The document itself is not in the database or on the search engine site.



Web Page

"Request: `Lincoln.html`"

5 The database is searched, based on the criteria you've set. Results are returned in HTML pages. Each search engine returns results in a different way. Some weigh the results to show how relevant the document is to your search; some show the URL, as well as the first several sentences of the document; and some show the title of the document and the URL.

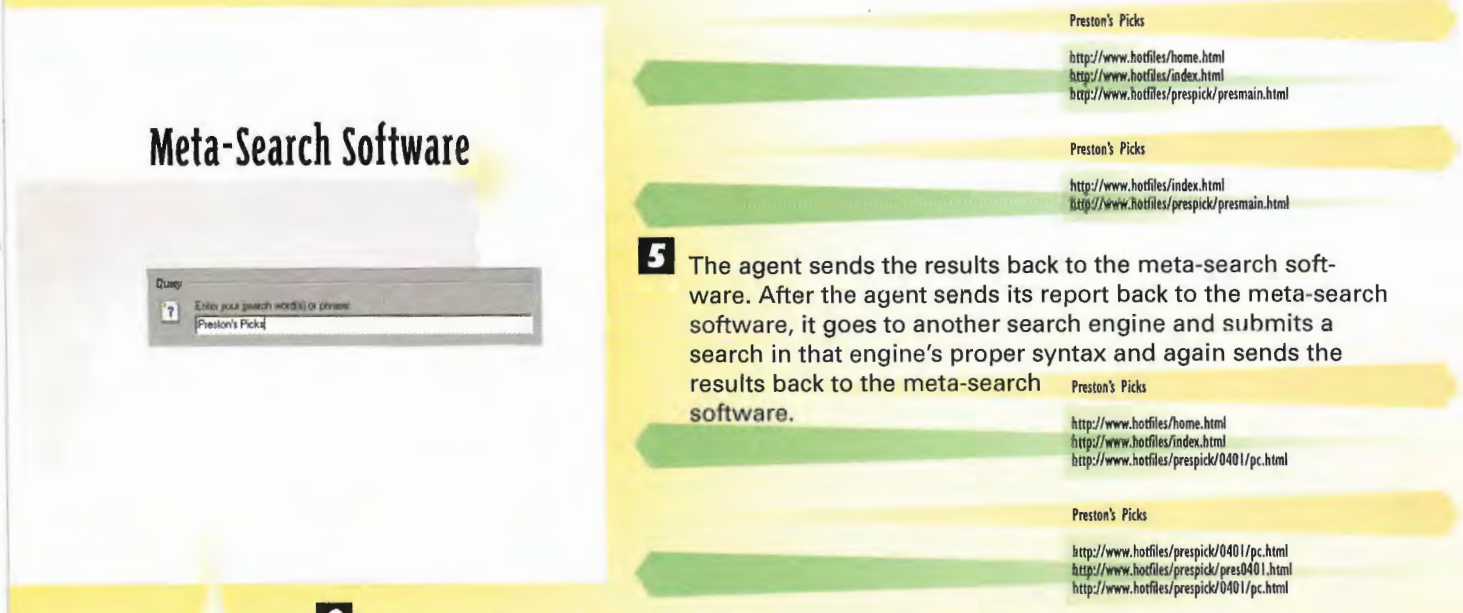


4 When you visit a search engine and want to search the Internet for information, you type words on a Web page that describe the information you want to find. Depending on the search engine, more than just keywords can be used. For example, you can search by date and other criteria with some search engines.

Server

How Meta-Search Software Works

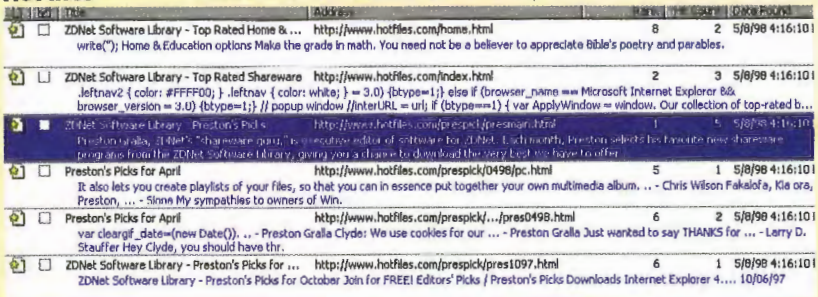
1 Meta-search software is software that sits on your computer and enables you to search through many Internet search engines simultaneously and to view and use the results. When you want to search for something on the Internet, you type descriptive words or a search term into the meta-search software.



5 The agent sends the results back to the meta-search software. After the agent sends its report back to the meta-search software, it goes to another search engine and submits a search in that engine's proper syntax and again sends the results back to the meta-search software.

6 The meta-search software takes all the results from all the search engines and examines them for duplicate results. If it finds duplicate results, it deletes them. It then displays the results of the search, ranking each *hit* by the likelihood that it contains the information you requested. It figures out the ranking by examining the title of the site found, the header information in the site, and the words on the site.

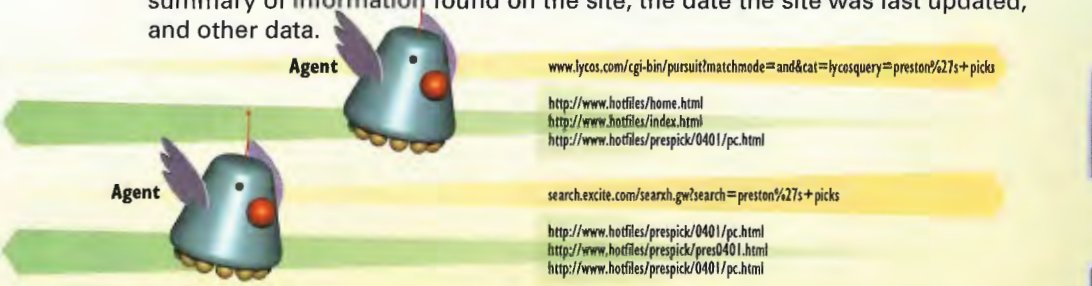
Results



2 The meta-search software sends many "agents" out onto the Internet simultaneously—depending on the speed of your connection, usually from 4 to 8, but it can be as many as 32 different agents. Each agent contacts one or more search engines or indexes, such as Yahoo!, Lycos, or Excite.

3 The agents are intelligent enough to know how each search engine functions—for example, whether a particular engine allows for Boolean searches (searching by using AND, OR, and other variables). The agents also know the exact syntax each engine requires. The agents put the search terms in the proper syntax required at each specific search engine and submit the search—they don't have to fill out forms, as users normally do at search engines.

4 The search engines report the results of the search to each agent. The results typically include the URL of each site that matches the search, and often a summary of information found on the site, the date the site was last updated, and other data.

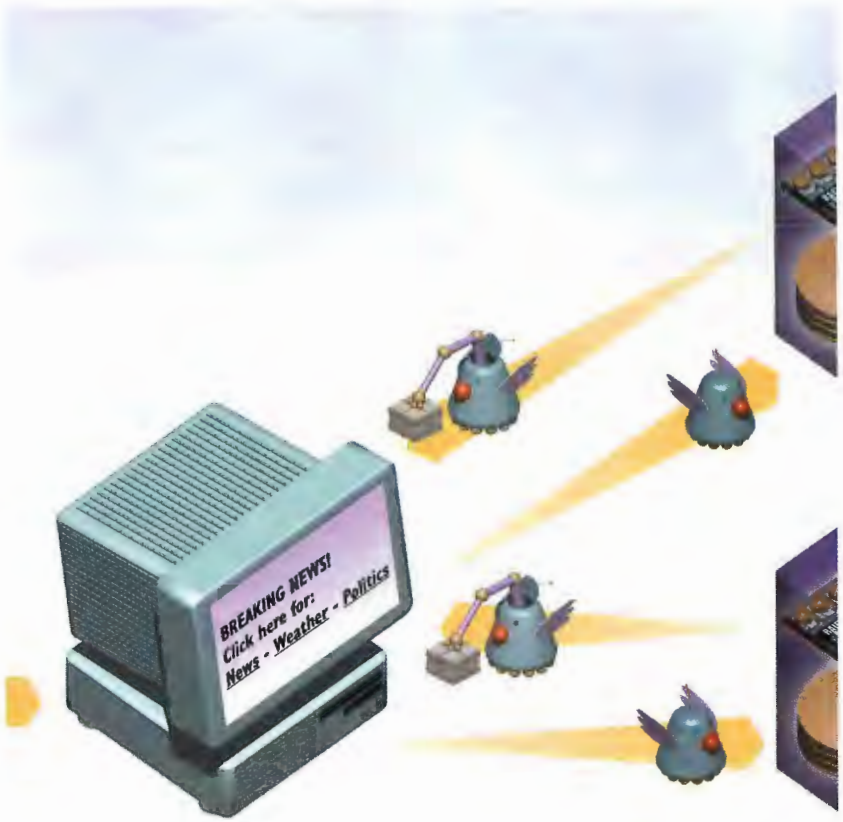


Web Page

7 You browse through the results in the meta-search software. When you see a page you're interested in, you double-click it. You then are sent to that site.

CHAPTER

33 How Agents Work



THE Internet has grown so quickly and its resources are so vast that we need help navigating around it. We can now use special software called agents to help us access the Net's resources.

Although there are a lot of technical definitions for agents, put simply, agents are programs that do your bidding automatically. Many of them run over the Internet or on individual computers every day. Agents can find the latest news for you and download it to your computer; they can automatically monitor Internet traffic and report on its total usage; they can find you the best deal on the CD you want to buy; they can perform important Web maintenance tasks; and they can do far more. They are becoming so complex that systems are being developed to allow agents to interact with one another so they can perform jobs cooperatively.

On the Internet, agents are commonly called *spiders*, *robots* (often shortened to "bots"), and *knowbots*, among other terms. Those used for searching automatically create indexes of almost every resource on the Web and then allow people to search through those indexes to find things more quickly. Common search tools such as Lycos, Infoseek, and AltaVista use spiders in this way. This specialized use of spiders is discussed in Chapter 32, "How Internet Searches Work."

All these agents are software programs that are invisible to the user. You just determine the task you want done, and behind the scenes the agent automatically goes off and performs that task. A variety of programming languages can be used to write agent programs.

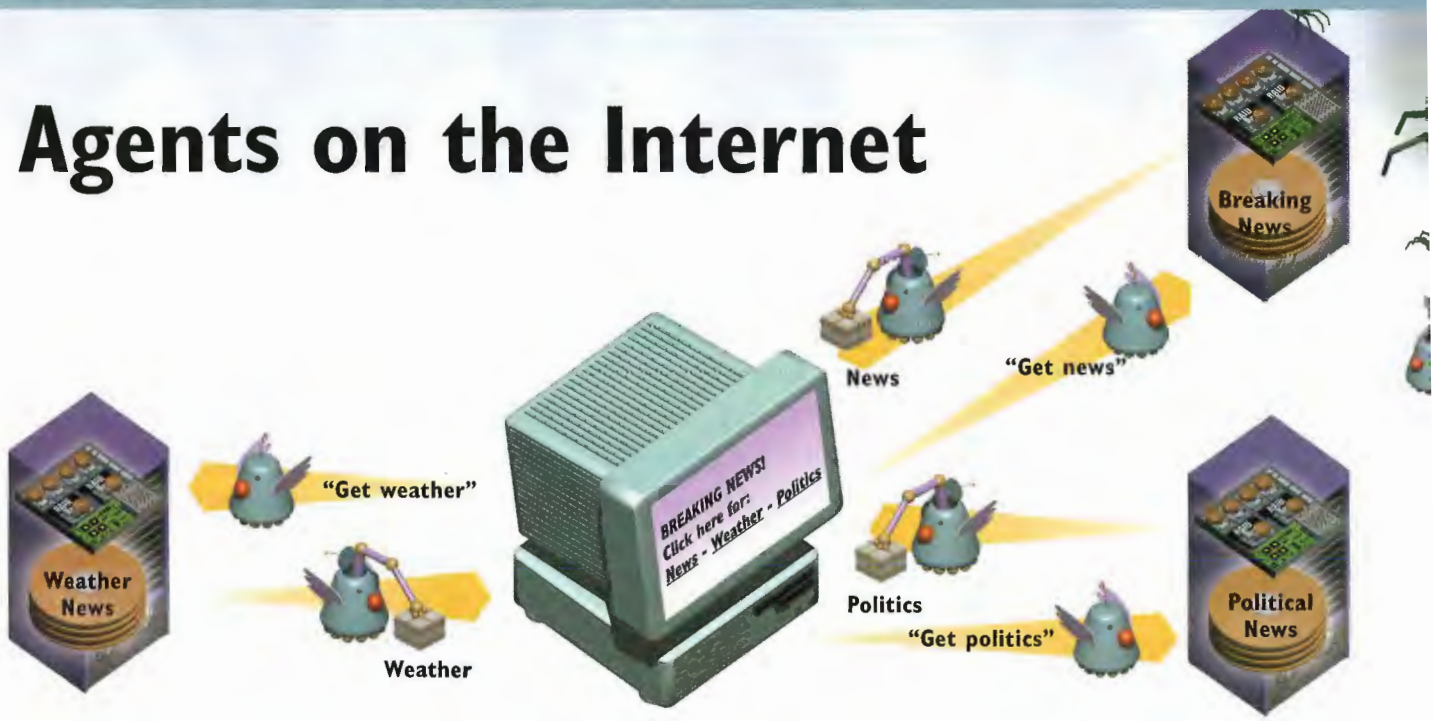
Agents might well alter the way we all use the Internet in the future. Not only do they respond to our requests, but they also "learn" from our requests the types of tasks and information that interest us. They then go off on their own and perform those tasks and get that information, even before we make these additional requests. As we use these types of agents more, they'll become even smarter and more efficient.

Robots and agents can cause problems for some Web sites. For example, they can overload Web servers by swamping them with too many requests in too short a time. That means users who try to get access to those Web pages will be denied access, or access will be exceedingly slow.

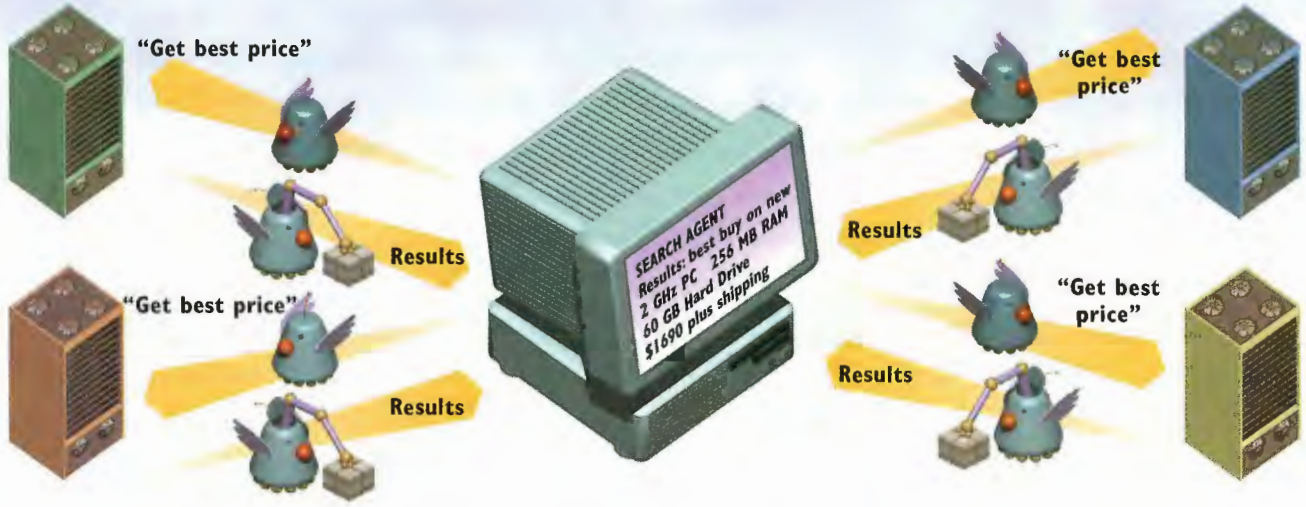
Another problem has to do with the way Web sites make money. Many Web sites sell ads to support themselves and charge advertisers based on the number of pages that have been viewed. If many of those pages "viewed" are in fact never seen by people and are instead accessed only by a computer via a robot, both advertisers and the Web site suffer.

Several ways exist to solve these problems and limit robot access. One way includes creating a file called `Robots.txt` that describes the areas that are off limits to robots, which the robots would automatically read, adhere to, and not visit. Another is to use a technology that automatically detects whether a robot or a human has visited a page and forgo charging advertisers whenever robots visit.

Agents on the Internet

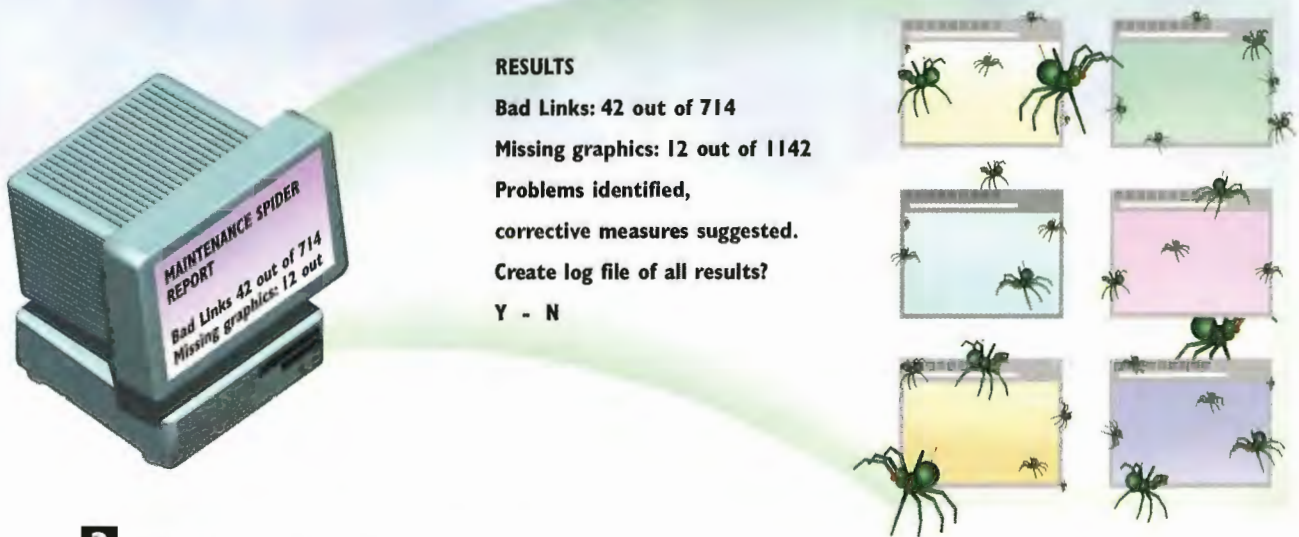


1 A simple Internet agent is one that gathers news from a variety of sources while you're not using your computer or while you are using your computer for another task. News agents can work in several ways. In the simplest example, you fill out a form saying which type of news you're interested in and on what schedule you want your news delivered. Based on that information, at preset intervals, the news agent connects to news sites around the Internet and downloads news stories to your computer, where you can read them as HTML pages.



2 Shopping agents let you search through all of the Internet for the best bargains. On the Web, you fill out a form detailing the product you want to buy. When you submit the form, the shopping agent launches programs that search through a variety of shopping sites and databases on the Internet. The agent looks into the databases of those sites and finds the best prices. It then sends back to you the links to the sites so you can visit the sites with the best prices and order from there.

4 When robots and spiders do their work on a remote Internet site from where they were launched, they can put an extra load on the site's system resources—for example, by swamping the server with too many requests in too short a time. Because of this, some system administrators are interested in ways of excluding robots in certain circumstances, such as not allowing robots into certain Web directories. A variety of ways have been devised to limit robot access, including creating a file called `Robots.txt` that describes the areas off limits to robots, which the robots would read, adhere to, and not visit. But there is nothing that guarantees the robots must adhere to this rule. It's up to the good faith of the person writing the robot to adhere to it.

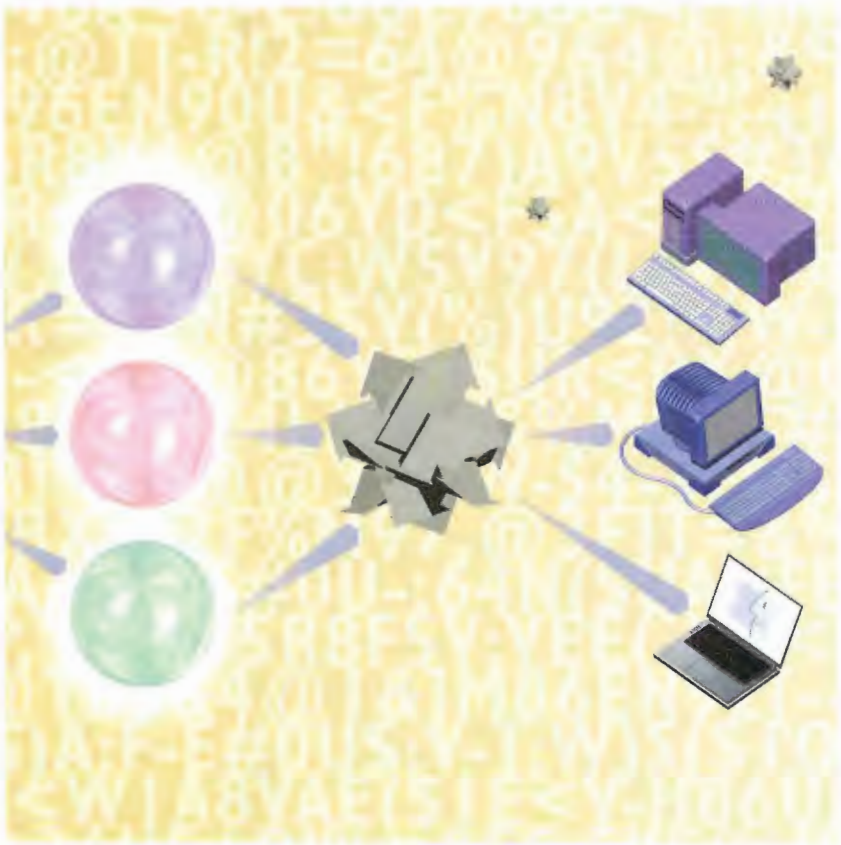


3 Web robots called *Web maintenance spiders* can perform important Web maintenance chores. On Web sites, particularly large ones, very often HTML pages can include links that become outdated. In other words, the object being linked to has been taken off the Internet. Whenever a user clicks the link, an error message is sent. A Web maintenance spider can look at every link on every HTML page on a Web site and trace each link to see whether the linked object still exists. It then generates a report of dead links. Based on that report, the system administrator can rewrite the HTML code, getting rid of the bad links.

CHAPTER

34

How Java, ActiveX, and JavaScript Work



THE Internet is no longer a place that you visit with your computer and merely look at documents or gather information—increasingly, it is an extension of your computer. You can now run programs that reside on the Internet rather than on your own computer, and tools have been developed that enable your computer and the Internet to interact as if they were one large computer system. This allows for all kinds of things never before possible: news tickers that flash breaking news; interactive games; multimedia presentations combining animation, sound, and music; and much more.

The three most important programming tools associated with Web technology are Java, ActiveX, and JavaScript. *Java*, a computer language developed by Sun Microsystems, enables applications to be run from the Internet—the same as word processing and spreadsheet programs that are run on your computer. It is similar to the C++ computer programming language and is object-oriented, which means programs can be created by using many preexisting components instead of by a programmer writing the entire program from scratch. Although most Java programs are run from the Internet, they don't have to be and can be run just like any other type of program.

Java programs run inside your Web browser if you have a Java-enabled browser. Most browsers are Java-enabled. When Java programs are run inside a browser, they are called *applets*. You don't need to do anything to run a Java applet. When you visit a Web site that has a Java applet on it, the applet is downloaded automatically from a Web server and then run automatically in your browser. Java applets can be run on any computer, such as a PC, a Macintosh, or a Unix workstation.

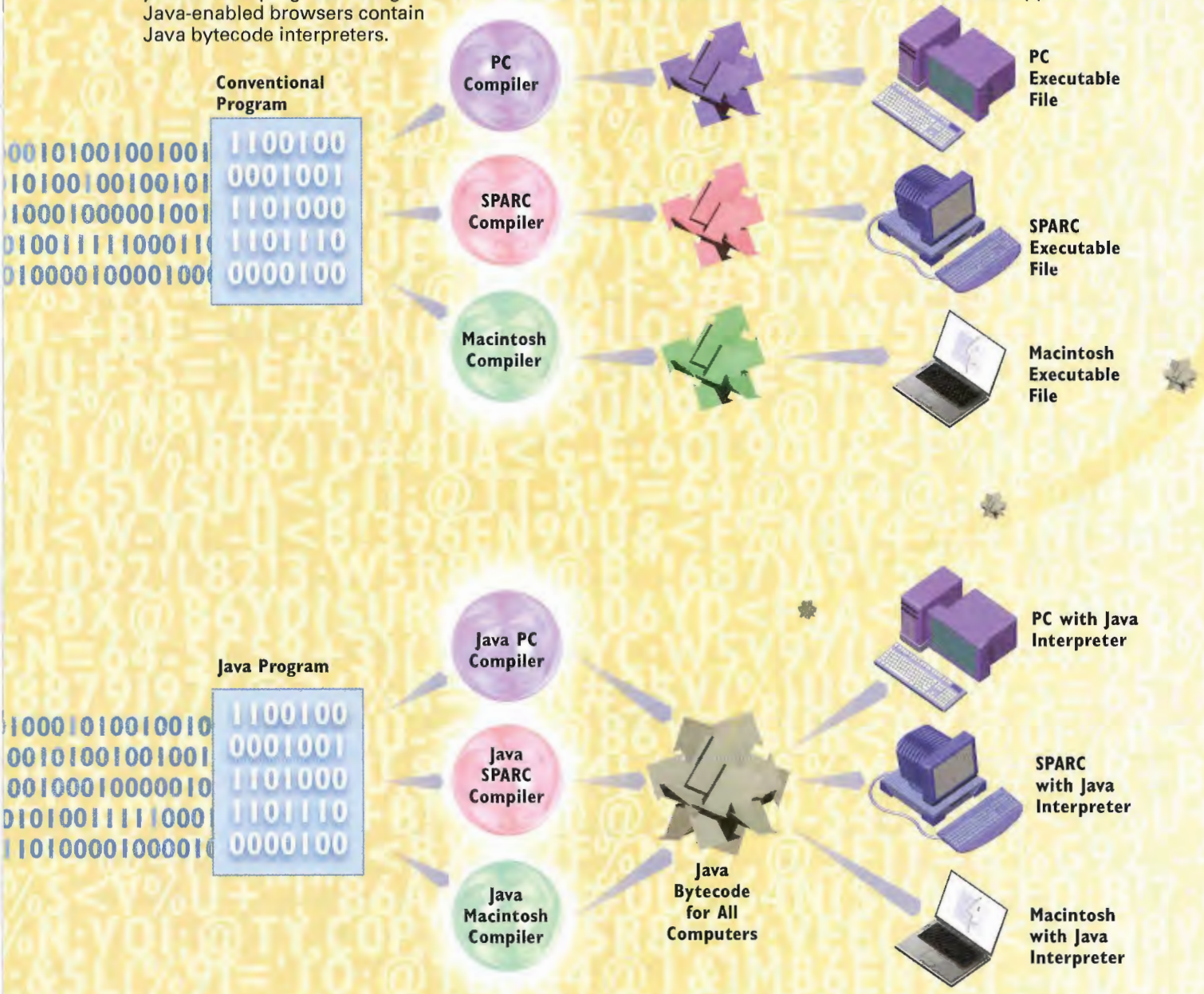
ActiveX, a technology from Microsoft, enables Internet programmers to create programs—commonly referred to as ActiveX *controls* or *components*—that can essentially turn the Internet into an extension of your computer. Similar to Java applets, these controls are downloaded to your computer and run there. They can do anything a normal application can do, in addition to interacting with the Web, the Internet, and other computers connected to the Internet. To run these controls, a browser that supports ActiveX, such as Internet Explorer, is necessary.

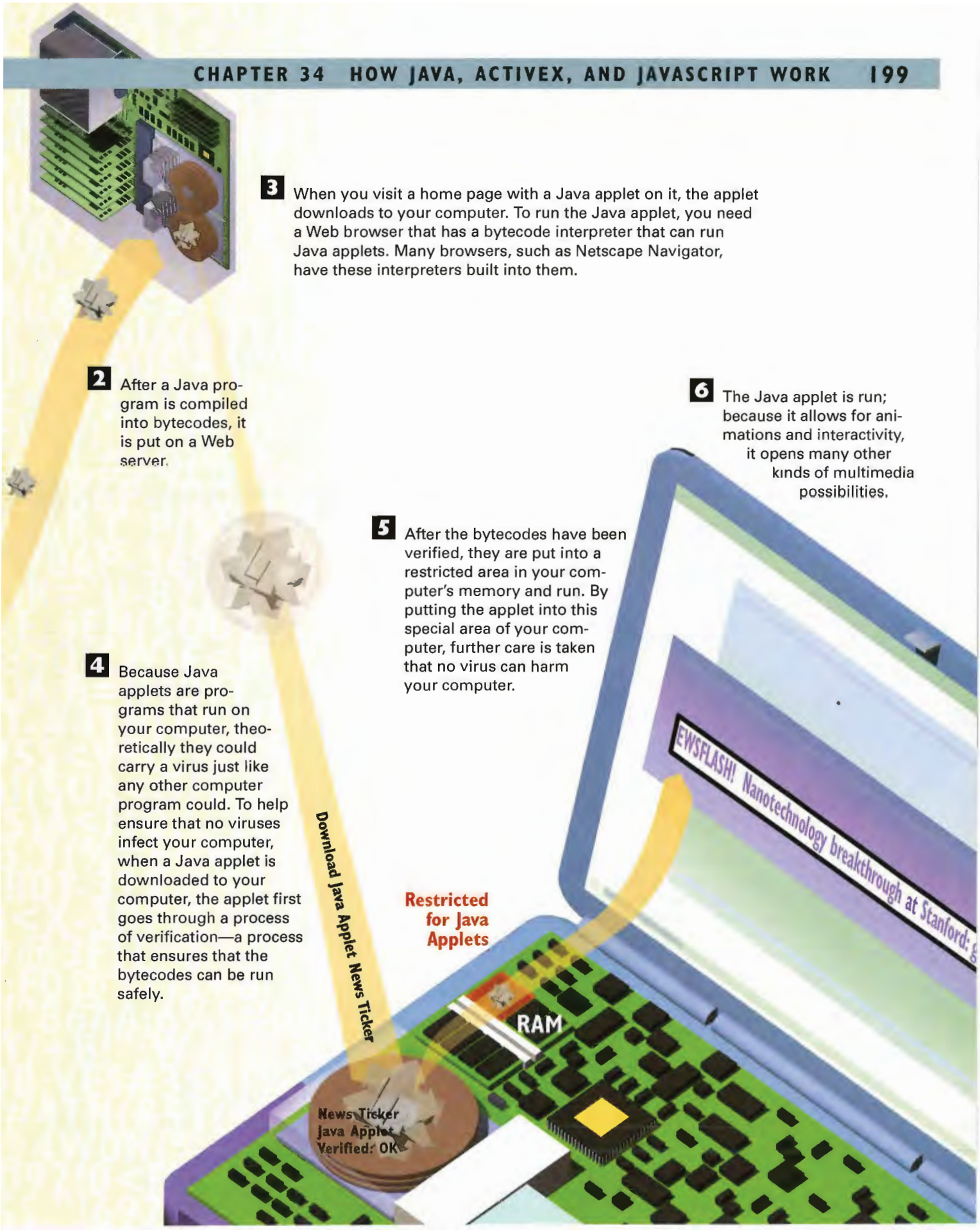
One benefit of ActiveX controls is that they are written as components, which means they can be put together, much like building blocks, to build larger and more complex applications. Another benefit is that, because you have already downloaded a component, you won't ever need to download it again. So, when you visit a page with a complex ActiveX application on it, you might need to download only a small portion of it because you might already have the other components on your computer.

JavaScript is a scripting language that is less complex and therefore much easier to learn than Java and ActiveX. People without substantial programming experience can write scripts with JavaScript. It's also an interpreted language, which means that its commands are executed by the browser in the order in which the browser reads them. It's commonly used for things such as creating drop-down boxes, navigational aids, and interactive forms, although it can be used for creating more complex applications, as well.

How Java Works

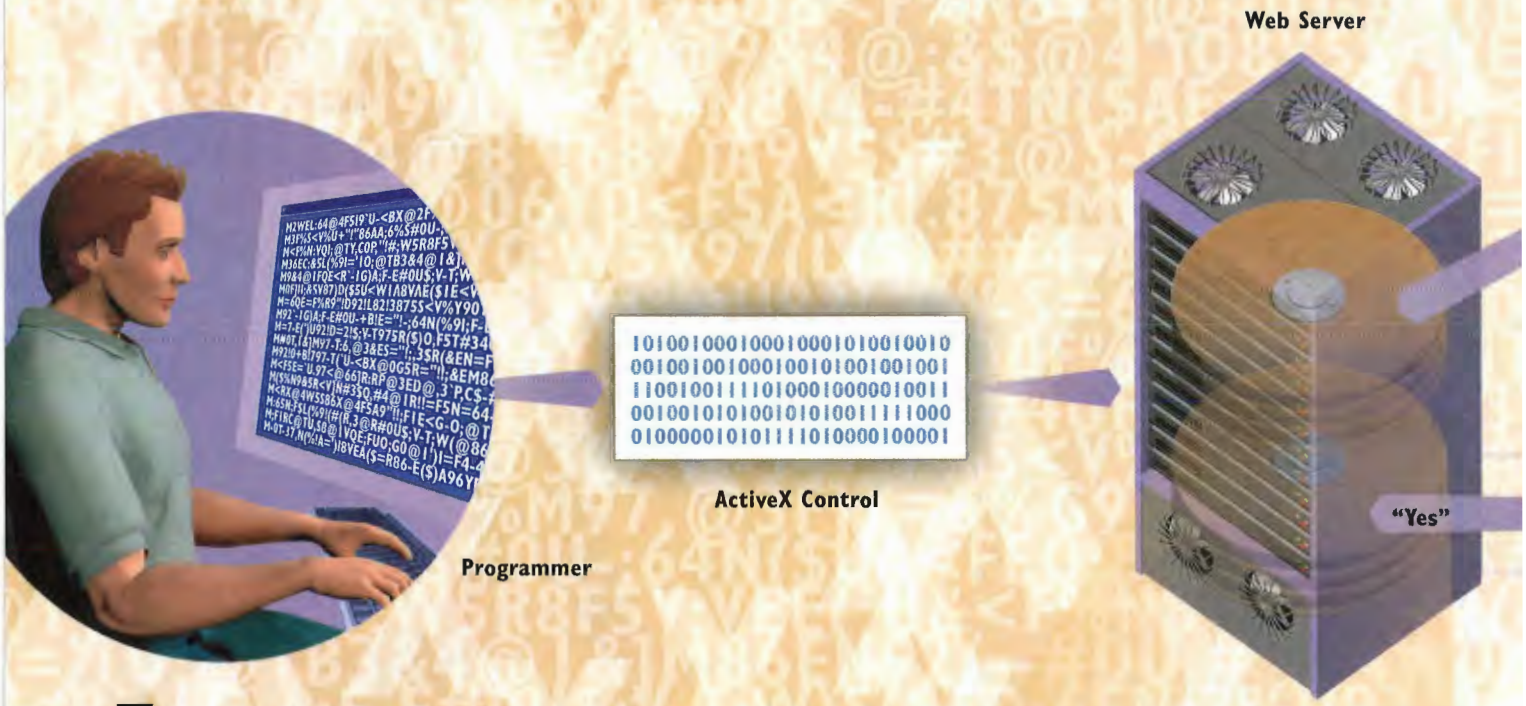
Java is a compiled language, which means that after a Java program is written, the program must be run through a compiler to turn the program into a language a computer can read. Java differs from other compiled languages, however. In other compiled languages, computer-specific compilers create distinct executable binary code for all the different computers on which the program can run. In Java, by contrast, a single compiled version of the program—called Java *bytecode*—is created by a compiler. Interpreters on different computers, such as a PC, Macintosh, or SPARC workstation, understand the Java bytecode and run the program. In this way, a Java program can be created once and then used on many types of computers. Java programs designed to run inside a Web browser on the World Wide Web are called applets. Java-enabled browsers contain Java bytecode interpreters.




3 When you visit a home page with a Java applet on it, the applet downloads to your computer. To run the Java applet, you need a Web browser that has a bytecode interpreter that can run Java applets. Many browsers, such as Netscape Navigator, have these interpreters built into them.
2 After a Java program is compiled into bytecodes, it is put on a Web server.
6 The Java applet is run; because it allows for animations and interactivity, it opens many other kinds of multimedia possibilities.
5 After the bytecodes have been verified, they are put into a restricted area in your computer's memory and run. By putting the applet into this special area of your computer, further care is taken that no virus can harm your computer.
4 Because Java applets are programs that run on your computer, theoretically they could carry a virus just like any other computer program could. To help ensure that no viruses infect your computer, when a Java applet is downloaded to your computer, the applet first goes through a process of verification—a process that ensures that the bytecodes can be run safely.
Download Java Applet News Ticker
Restricted for Java Applets
RAM
News Ticker Java Applet Verified: OK
EWSFLASH! Nanotechnology breakthrough at Stanford:

How ActiveX Works

1 First, a programmer creates an ActiveX control. A variety of programming tools can be used to create controls, such as Visual Basic or the C programming language. A control can be as complex as a program that checks your computer for viruses and then eradicates them, or as simple as showing a Web site in an outline view. After the control is created, it is posted on a Web server, and information about the control is coded into the Web page through use of the HTML `<object>` tag.



2 When you visit a Web page that contains an ActiveX control, your browser sees the HTML `<object>` tag. This tells your browser that an ActiveX control is present. The tag contains a variety of information required to run the control. It can tell the browser where the control is located on the server, for example, and which type of file it is. It can point to the control (which has an .ocx extension); to an installation file (which has an .inf extension); to a compressed file (which has a .cab extension); or to a variety of other kinds of files.

```
<OBJECT ID="ourMenu"
WIDTH=0, HEIGHT=0
CLASS ID="CLSID:9BC24D87-
E21A-10DF-B7D2-180089E9610A
CODEBASE="http://www.que.com/
ActiveX/QueControls.CAB#
Version=1, 0, 8, 0">
```

3 Some ActiveX controls need more than a single control to work. In that event, the HTML page contains multiple references to ActiveX controls needed to run the control on that page. The controls can be located on the same server, on a different server on the same Web site, or on another site and server on the Internet.

Get Active X Control

Web Server

4 Using the information in the `<object>` tag, the browser starts to download the ActiveX controls. If you have set your browser to a certain level of security, you get a message asking whether you want to download the control. To ensure even more security, ActiveX controls can be digitally *signed* by a digital certificate authority, such as VeriSign. This signing assures you that the control you want to run was written by the person to whom it is attributed. If a problem occurs with the control, you will be able to contact that person.

ActiveX Control

Web Browser

Do you want to download control XYZ?

ActiveX Control

5 If you have low security set on your browser, or if you give the okay to download the control, the control and its related ActiveX controls are downloaded to your computer. Some of the controls already might be on your system, so you won't need to download them. After the control is downloaded, the file is decompressed (if it was compressed), information about it is put into the Windows Registry, and it is installed on your computer. The control then runs. An ActiveX control can do anything any other program can do. It can interact with your computer and with any Internet resource, such as the Web, FTP, Telnet, or virtually any other Internet resource. It can also directly use the Internet's TCP/IP protocols so that it need not ride on top of another Internet resource.

How JavaScript Works

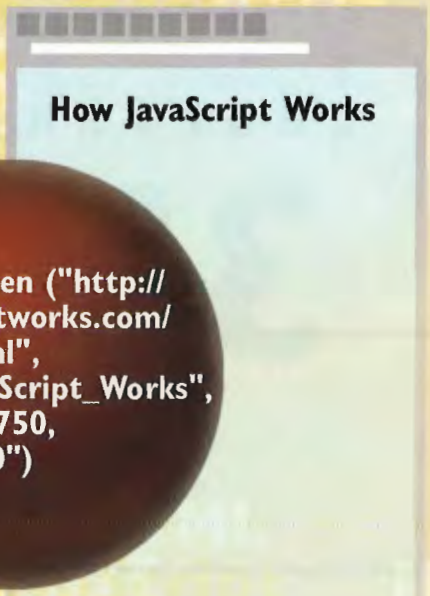
1 JavaScript is an *object-oriented* language, which means that it works by manipulating objects on a Web page, such as windows, buttons, images, and documents. It groups these objects into hierarchies, which enables programmers to manipulate them more easily. It's also an *interpreted* language, which means that its commands are executed by the browser in the order in which the browser reads them.

```
Window
    Document
        Elements
        Button
        Check Box
        File Upload
        Password
        Radio
```

2 JavaScript commands are put directly into the HTML file that creates a Web page. Depending on the script being run, the commands can be put into several places in the file. Often, the commands are put near the top of the file. Special codes set off the commands, alerting the browser that they're JavaScript commands. If the commands are put before the HTML `<body>` tag at the top of the file, the script can start executing while the HTML page is still loading.


```
window.open
then
document.write
then
window.status=
"these are the times
that try men's souls"
then
```

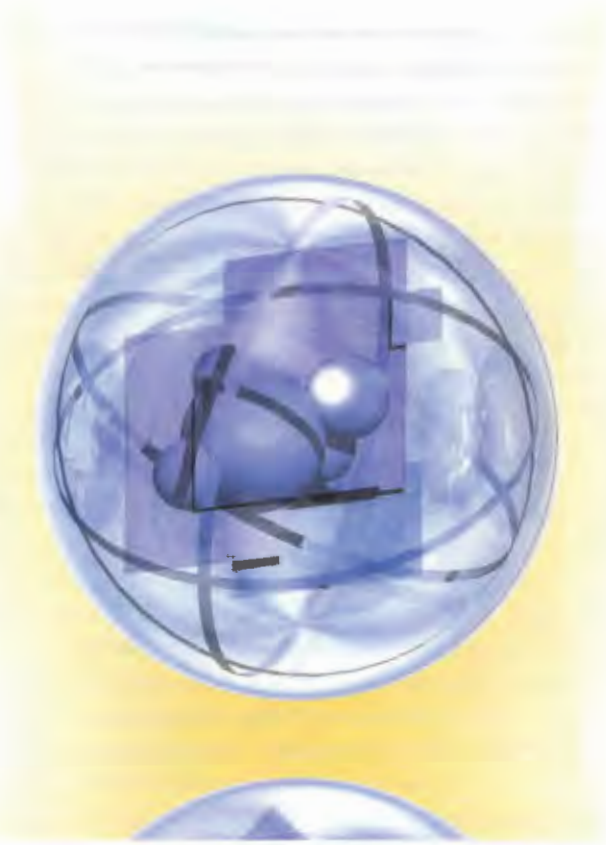
3 The heart of the way JavaScript works is to take actions on objects. These actions are called *methods*. Using this basic concept, JavaScript can be used for a wide variety of sophisticated, interactive features, but we'll look at a simple script that opens a new browser window to a specified size, puts a specific Web page in it, and names the window. In the basic syntax of JavaScript, first the object is named, and then a period appears, followed by the action taken on the object—the method. So, the command to open a new window in JavaScript is `window.open`. In this instance, `window` is the object, and `open` is the method. This command opens a new browser window.



4 You can add further instructions to the `window.open` command. You do this by adding parameters after the command. You put all the parameters inside one set of parentheses, put each individual parameter inside quotation marks, and then separate the parameters by commas. So, the command `window.open("http://www.howitworks.com/jscript.html","How_JavaScript_Works","height=1750,width=150")` opens a new browser window 1750 pixels high and 150 pixels wide with the `http://www.howitworks.com/jscript.html` in it.

CHAPTER 35

How CGI Scripting Works



IF you browse the Web for very long, you are sure to come across the term CGI, or *Common Gateway Interface*. CGI refers to the communications protocol by which a Web server can communicate with other applications. For example, a CGI application, sometimes called a *script*, is often used to enable Web users to access databases or to get information from forms people fill out. CGI can also be used to create agents that do things such as check a Web site to see whether it has any broken links.

Essentially, CGI is a standard way in which the Web interacts with outside resources. Often, that outside resource is a database. You've probably run CGI scripts many times without knowing it. For example, if you've filled out a form on a Web page to register to use a site and then later received an e-mail notification with a password for you to use, you've probably run a CGI script. In that case, the CGI script probably took the information you filled in on the form and performed several actions on it, including putting the information in a database, automatically creating a password, and then sending you mail.

CGI and CGI applications are often confused. *CGI applications* receive data from the server and return the data via the Common Gateway Interface. CGI applications usually are written in a programming language called Perl (Practical Extraction and Reporting Language), although they can be written in C, C++, Pascal, AppleScript, or others as well. CGI itself is a standardized means of communicating between a CGI application and the HTTP server. It's the "doorway" of sorts through which the Web server sends requests and the CGI application collects and returns data.

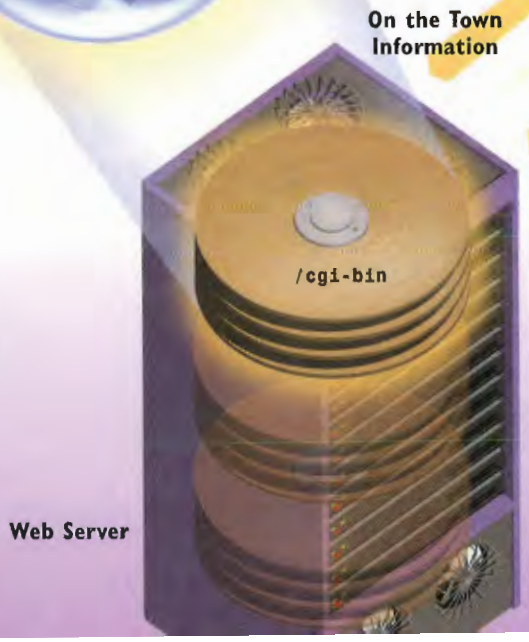
In the example of providing information on a Web page designed to accept user input, CGI performs many tasks. First, you submit unique information—such as a name or e-mail address—to the server for processing. Next, the server redirects the information to a CGI application that is called by the form "submit." CGI scripts are activated by the server in response to an HTTP request from the client. Lastly, a CGI application might send form data to another computer program, such as a database; save it to a file; or even generate a unique HTML document in response to the user's request. This is known as an *interactive form*.

In the illustration that accompanies this chapter, we'll look at a CGI program that enables someone to search a movie database for information.

Understanding CGI Scripting

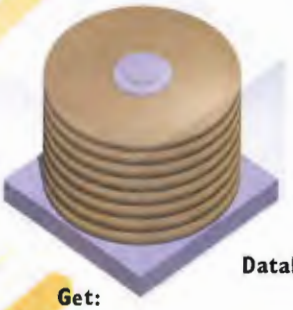
1 People who dial into the Web site don't need to know programming to access CGI programs. Instead, a programmer writes a CGI program. A number of languages can be used for CGI, such as C or C++, FORTRAN, Visual Basic, and AppleScript. An application written in a programming language such as C must be passed through a program called a *compiler* before it can be run. The compiler turns the application into a language CGI can understand. Other languages, called *scripting languages*, do not need to be compiled first. CGI scripts tend to be easier to debug, modify, and maintain than compiled programs so they are used more frequently. Perl is probably the most popular language used for writing CGI scripts.

2 After the program is written and compiled, or the script is written, the program is put into a special directory on the Web server, such as `/cgi-bin`, where all the CGI programs are stored and maintained. The person in charge of the Web server determines which directory should hold CGI programs. If someone writes a program and doesn't put it in the proper directory, it won't run. This is a security feature. If there were many different directories people could use to store and run CGI programs, keeping track of them all would be difficult, and someone from the outside could create and post a program that could be dangerous to the software that's already there.



6 The CGI program receives the data from the database and formats it in a way that will be understandable to the user. For example, the program might take the information and put it into HTML format so the user can read it using her Web browser. The CGI program sends the results in HTML format to the user, who displays it in a Web browser. The user can then access that HTML page. She can click links to visit other pages, print pages, and view graphics and multimedia files.

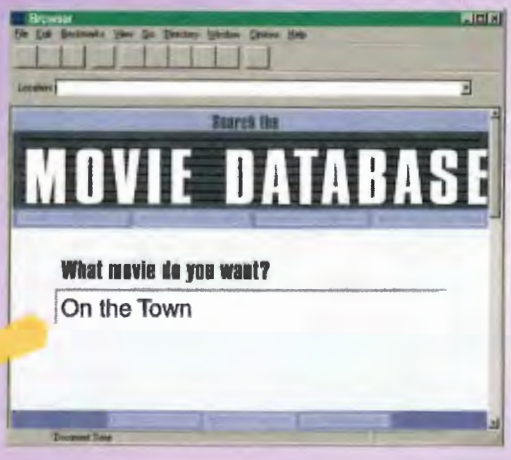
5 The CGI program contacts a database and requests the information the user is looking for. The database sends the information to the CGI program. The information can be in a variety of formats, such as text, graphics, sound and video files, and URLs.



Database

Get:
On the Town

4 When you visit the Web site and click the URL, the CGI program is launched. If the CGI program allows you to search a database, for example, it sends a form in HTML format. You then fill out the form detailing what you want to find. When you finish the form and click Send, the data from the form is sent to the CGI program.



3 After the CGI program is posted to a special directory, a link to it is embedded in a URL on a Web page.


real
RealPlayer Plus: WebSurfer
International
Search

PART

ENJOYING ENTERTAINMENT AND MULTIMEDIA ON THE INTERNET

POSSIBLY the most dramatic and remarkable part of the Internet is the multimedia content and entertainment you can find there. You can listen to music, sound clips, and live radio stations from your computer. You can share your favorite music files with others all over the world. You can watch astronauts live while they're on the space shuttle. You can watch video clips of the news and other events. And you can even have live videoconferencing with people from all over the world.

You can do all that with the Internet's audio and video capabilities. You won't need specialized hardware and software to do it—and in many cases, you won't even need a very high-speed Internet connection. An ordinary dial-in connection to the Internet will do, although the sound and video quality will be better at higher speeds. And you'll only need free or inexpensive software, and a sound card and speakers that ship with most computers, or that are available separately.

The Internet's multimedia capabilities go beyond mere playing of audio and video clips and listening to Internet radio stations. You can participate in virtual worlds and join in virtual chat sessions in which you build your own online persona, called an *avatar*, which communicates with other avatars. The Internet enables the creation of remarkable online multimedia content, combining animation, sound, and programming via technologies such as streaming audio and video, Shockwave, and Multicast IP.

This section of the book discusses how every aspect of multimedia and entertainment on the Internet works. Chapter 36, "How Music and Audio Work on the Internet," covers audio and music. You'll see how audio files are sent to your computer and played. You'll look at how streaming audio works in detail. Streaming audio enables you to play sounds and music on your computer while the audio file is being transferred to your computer, so you don't have to wait for the file to download.

Chapter 36 also looks at the hottest new type of music on the Internet—music that can be downloaded and then played on a computer using a format called MP3. MP3 files are near-CD quality sound files, yet are small enough so they easily can be downloaded. This chapter also shows how Internet radio broadcasting works. Increasingly, radio stations broadcast live over the Internet so you can listen in using special software or just your browser. Many of these stations are Internet-only stations that broadcast only online, although many real-life radio stations around the world also broadcast over the Internet.

Chapter 37, "How Napster and File Sharing Work," looks at one of the most controversial uses of the Internet—the way music files can be shared with others. It shows you the inner workings of Napster, the file-sharing software that enables anyone to download their favorite music from other music lovers.

Chapter 38, "How Video on the Internet Works," details how video works. You'll learn how streaming video works, which (like streaming audio) lets you watch a video while it is being downloaded to your computer. Today, you're able to watch news broadcasts, music videos, and even live launches of the space shuttle through streaming video technology.

Chapter 38 also examines videoconferencing. Videoconferencing enables people from various parts of the world to see each other and talk to each other—all through their computers. The voices and images are transferred over the Internet.

In addition, Chapter 38 covers a lighter topic: how NetCams work. NetCams are cameras on the Internet that broadcast a photograph or digital animation at regular intervals. NetCams are all over the world, from the top of Pike's Peak to the streets of Hong Kong.

Chapter 39, "How Multicast IP and the MBone Work," looks at Multicast IP and the MBone. Multicast IP is a technique that enables videos to be broadcast to many thousands of people simultaneously, without clogging up the Internet's backbone. And the MBone is a high-speed Internet backbone used for transmitting Multicast IP video across the Internet.

Chapter 40, "How Virtual Reality Works," examines virtual reality. Virtual reality enables the creation of virtual worlds—3D creations on the Web through which you can walk or fly, interacting with your surroundings. As the bandwidth of the Internet increases, these virtual worlds might become increasingly popular.

Finally, Chapter 41, "Animation on the Web," looks at some of the most popular types of animation technologies, from the very simple to the very sophisticated. You'll learn how client pull and server push technologies enable the easy creation of simple animations. You'll also look at Shockwave, an extremely sophisticated way in which animation, audio, and other types of interactive technologies can be used to create powerful multimedia presentations on the Web. You'll even learn about the newest and most powerful kind of Web animation—Flash—which goes several steps beyond the capabilities of Shockwave.

CHAPTER

36

How Music and Audio Work on the Internet

SOUNDS, voices, and music are now an everyday part of the Internet. Through the Internet, you can listen to radio stations, interviews, music, sound clips, and much more.

You can listen to all this music and sound by downloading *audio files*—files that have been digitized so that a computer can play them. You'll find many music files and sound clips in a variety of sound formats online. Each of these formats has a different extension associated with it, such as .WAV, .MP3, or .AU. To play these files, you'll first have to download them and then have audio-player software play them on your computer. Netscape and other browsers have some of these software players built in, as do many operating systems. For additional formats, you'll have to find and download the player and then configure your browser properly to play the files—or run special client software that can play them.

Most sound files tend to be quite large—even after being compressed. For some types of sound files, you won't be able to listen to them until you download the entire file, and this can take quite awhile. Downloading a sound file that has less than a minute of sound in it may take 15 minutes.

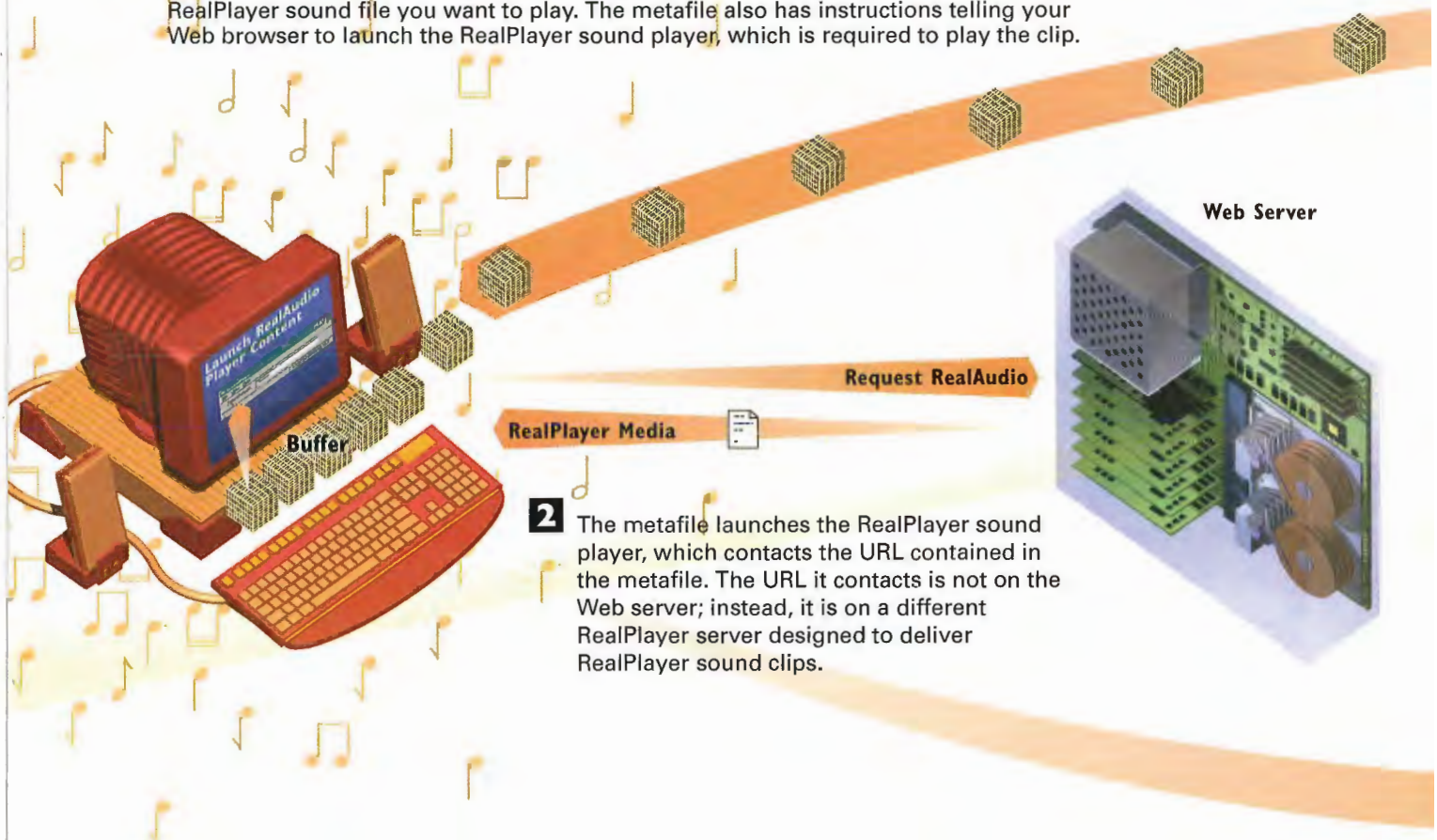
A far better and newer use of audio on the Internet is called *streaming audio*. It handles audio in a much more clever way. With streaming audio you don't have to wait until the entire audio file is downloaded to play it. Instead, you listen to the audio while it downloads to your computer. A variety of technologies allow for streaming audio. For all of them, you'll need to have the proper audio player for each specific kind of streaming audio. This chapter looks at the most popular audio streaming technology, called RealPlayer. Other kinds of streaming technologies exist, such as that used by the Windows Media Player. However, all streaming technologies work similarly.

One problem with streaming audio is that the sound quality generally isn't nearly as good as a music CD. However, MP3, a popular kind of audio file type, offers CD-quality audio. Furthermore, the MP3 files themselves aren't that large—usually less than 4MB or 5MB per song. With other kinds of computer music technology, these songs can take up 20MB and more. Technologies have been developed that enable MP3 files to be streamed so that you can listen to them as they download to your computer. This technology gives you the best of both worlds—high quality sound without having to wait for the whole file to download.

One of the more intriguing new audio uses of the Internet is the ability to listen to radio stations from all across the world. An increasing number of radio stations stream their live broadcasts over the Internet, and you can listen right from your browser or use software such as RealPlayer or the Windows Media Player. Entirely new radio stations have sprung up that broadcast only over the Internet.

How RealPlayer Streaming Audio Works

1 When you use your Web browser and click a link to a RealPlayer sound clip on a home page, the link doesn't lead directly to a sound file. Instead, your Web browser contacts the Web server, which then sends a file called a *RealPlayer metafile* back to your browser. This metafile is a small text file that has the true location—the URL—of the RealPlayer sound file you want to play. The metafile also has instructions telling your Web browser to launch the RealPlayer sound player, which is required to play the clip.



2 The metafile launches the RealPlayer sound player, which contacts the URL contained in the metafile. The URL it contacts is not on the Web server; instead, it is on a different RealPlayer server designed to deliver RealPlayer sound clips.

5 The packets are sent to a buffer on the receiving computer. When the packets exceed the capacity of the buffer, they are sent to the RealPlayer player, which then plays the sound file. RealPlayer allows you to jump ahead or back in a sound or music clip. When you move to a different place in the clip, the RealPlayer player contacts the server and tells it to start sending the file from that new place in the clip.

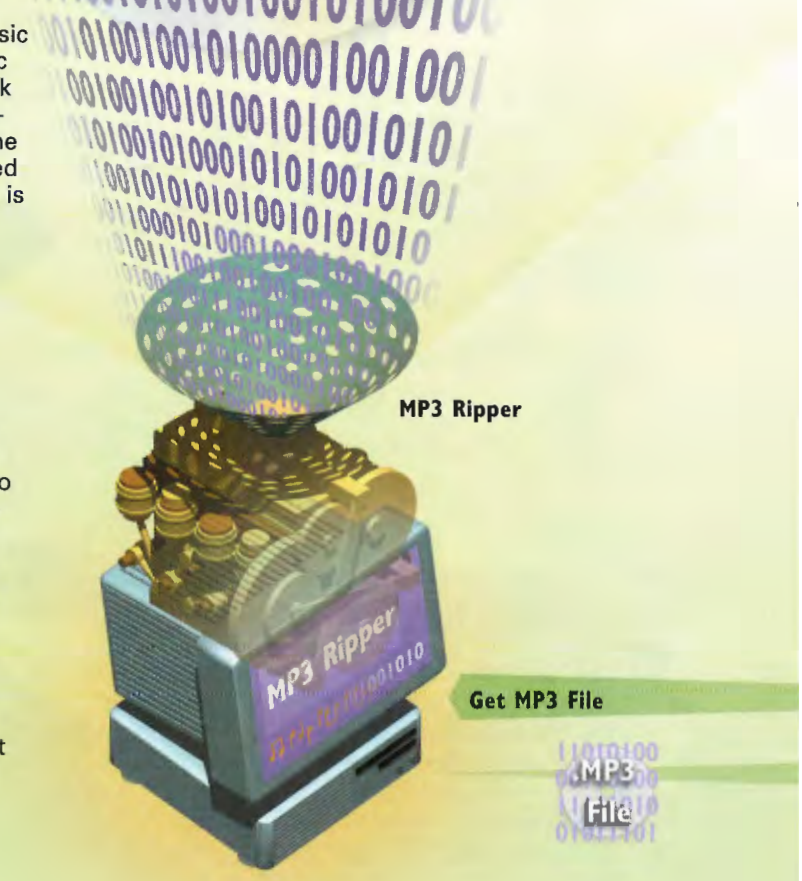
4 The RealPlayer clip is compressed and encoded. The sound file is too large and takes too long to send and play if it is not compressed. The clip is sent in IP packets using the UDP (User Datagram Protocol) instead of the Internet's normal TCP (Transmission Control Protocol). Unlike TCP, UDP doesn't keep resending packets if they are misplaced or other problems occur. If packets have to keep being re-sent, the sound player on the receiving end is constantly interrupted with packets and can not play the clip.

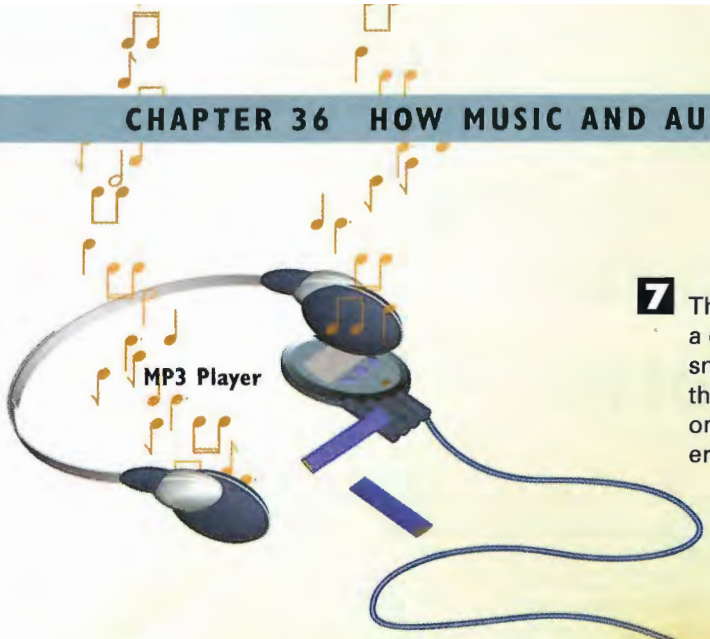
3 The RealPlayer server and the RealPlayer sound player "talk" to one another so that the server knows at what speed the user is connected to the Internet. If the connection is a low-speed connection, a smaller RealPlayer file is sent that contains less data. This file is of lesser quality than a file sent via a high-speed connection. If a high-speed connection is used, a larger, higher-quality sound file is sent. This provides for better sound quality.

How MP3 Music Files Work

1 One of the most popular ways to distribute music on the Internet is through the use of MP3 music files. The files use special algorithms that shrink the size of the files while retaining CD-like quality. Before MP3 files can be posted, however, the music has to be recorded. The music is recorded no differently than any other kind of music and is then put on a CD.

2 The music from the CD has to be converted to the MP3 format so that it can be posted on the Internet. A typical way to convert the music is to use a *ripper*—a piece of software that takes the music from the CD and converts it to MP3 format. The software uses special algorithms that shrink the size of the file dramatically so that a typical song is under 3MB and still maintains high quality. (In earlier kinds of PC music formats, those files would be 20MB and more.) This mix of small size and high quality is what sets the MP3 standard apart from other Internet music formats.

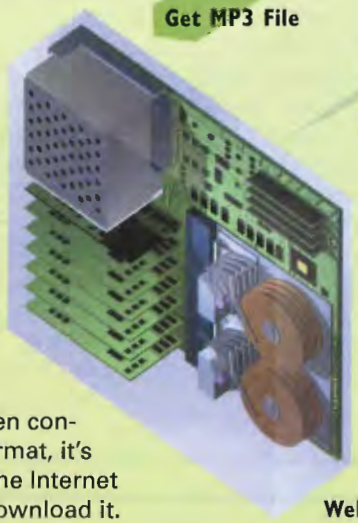




7 The MP3 files can also be transferred from a computer to a *portable MP3 player*—a small audio device that can play music in the MP3 format. The MP3 files are stored on a memory card in the device and can be erased or overwritten with new MP3 files.

6 One issue with MP3 files is that they can violate an artist's copyright—for example, if the file was created and posted without the artist's permission. In some instances, an MP3 player won't play an MP3 file if someone who didn't get the artist's permission ripped the file from a CD. In other instances, the file will play, but may contain copyright information about the MP3 file. However, in many instances, the file can be played and does not include copyright information.

4 When someone wants to download the MP3 file, he visits the Web site or FTP download site and downloads the file to his computer.

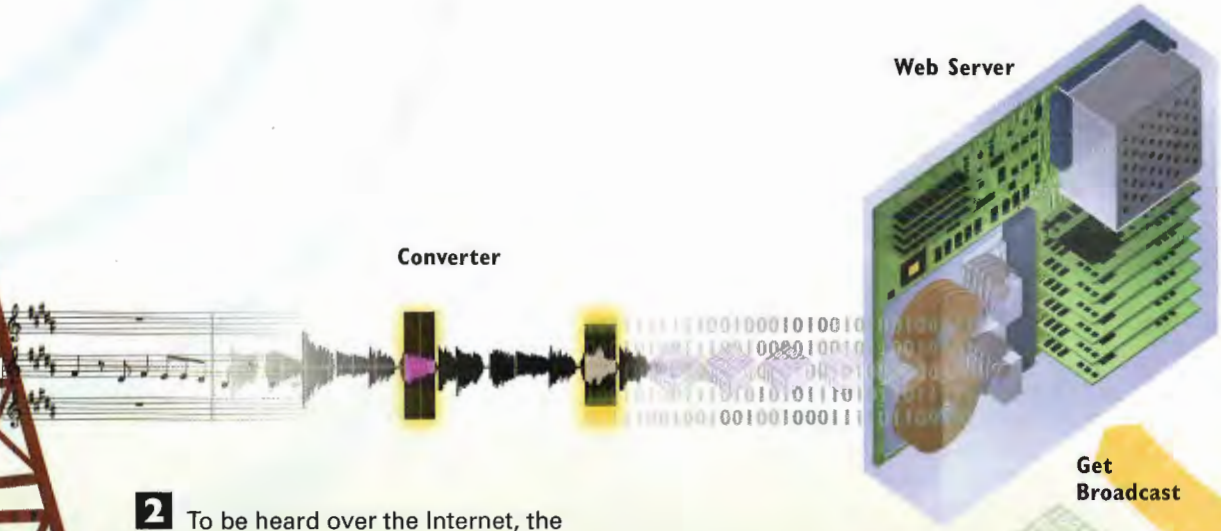


5 After the file is downloaded, it can be played with a special piece of software called an MP3 player. Some software and Internet servers can *stream* the MP3 file—play it while it's being downloaded. In most cases, however, the file is first downloaded and then played.

3 After the file has been converted to an MP3 format, it's posted to a site on the Internet where people can download it.

How Internet Radio Broadcasting Works

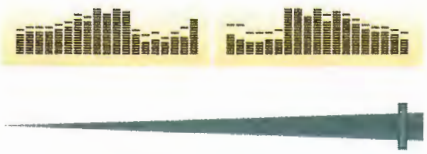
1 Two main types of radio stations broadcast over the Internet—traditional radio stations that also broadcast over the airwaves as well as Internet-only broadcasts. In both instances, the radio station plays music, news, or offers other kinds of broadcasts in much the same way that a normal radio station does.



2 To be heard over the Internet, the radio broadcast needs to be altered into a format that computers can read and play. A number of different formats exist—most notably RealPlayer and Windows Media Player. Software turns the normal broadcast into one or both of those formats. (To get the RealPlayer, go to `www.realnetworks.com`. For the Windows Media Player, go to `www.microsoft.com`.)

3 An Internet server hosts the live broadcast in the RealPlayer or Windows Media Player formats.

6 The client software now plays the broadcast live on the person's computer. The broadcast can be controlled like a radio broadcast—the sound can be lowered or raised and, depending on the software used, the quality of the sound can be altered. Generally, the higher the speed of the connection between the PC and the Web server, the higher the audio quality of the broadcast.



5 When the link is clicked, the client software contacts the server. The server sends the broadcast to the PC in a steady stream.

4 When someone wants to listen to a radio broadcast, he launches their client software radio player, such as RealPlayer. If he has Internet Explorer 5.0 or higher, he can also launch a radio toolbar that sits near the top of the browser window. From this toolbar, he can choose radio stations to listen to. To listen to a radio station, someone clicks a link in the player or in the browser. He can also visit a radio station site and click a link on the site.

CHAPTER

37

How Napster and File Sharing Work

EVERY once in a while a new feature or application takes the Internet by storm and not only changes the way many people use the Internet, but, at times, even changes the world beyond the Internet's borders.

Several years ago that's what Napster did. Not only did it change the way that many people used the Internet, but it also threatened the multi-billion dollar music industry. Napster forever changed the way people thought of and listen to music.

It did all this by putting into effect a very simple idea—letting people share their music with each other over the Internet. Despite all the hype, technology, and lawsuits, that's what it all comes down to.

People can make digital copies of their CDs by using a piece of *ripping* software that can turn CD tracks into digital files that can be played on a computer—most commonly files in the .MP3 music format. Napster isn't a ripper. Instead, it lets people find music in .MP3 format by searching through the music collections of thousands of other people. When someone finds a song he wants, he can download it to his computer from another person. Then he can listen to that song on his computer by transferring it to an MP3 player and listen to it there, or burn it onto a CD and listen to the song in a CD player.

The music industry cried "Foul!" and brought Napster to court on copyright violations. Napster was ordered to figure out a way to stop alleged copyright violations, while still allowing people to share music. The legal and technical fights continue, and it's unclear whether Napster will survive, and, if it does, whether it will remain free.

But no matter what happens, the genie is out of the bottle. Other software such as the Gnutella file-sharing network—software that lawyers and judges can't pursue—enables people to do the same thing.

This kind of technology, which enables people to share files directly with one another, is called *peer-to-peer*. Peer-to-peer has gone far beyond allowing people to share music—people can share any kind of files from spreadsheets to movies.

A whole new kind of application has sprung up in Napster's wake: business peer-to-peer software. The most notable example is called Groove—software that enables people in corporations to create their own private workspace where they can share files, messages, and software. Multi-billion dollar corporations have already signed on and are using the software. Ironically, a technology that started as guerilla music-sharing software might find fruition as a corporate mainstay.

How Napster Works

1 To use Napster, download and install it on your computer. After you've installed Napster, it compiles a database of all the music on your computer.

Here's my music (IP 192.68.1.10).

Search: Dylan

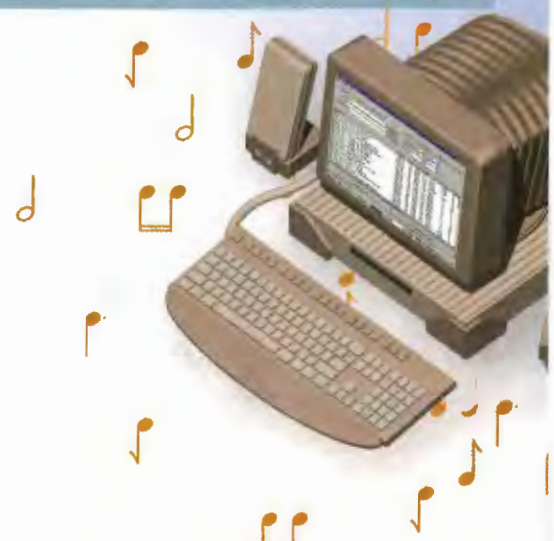
Here's how to find the song.

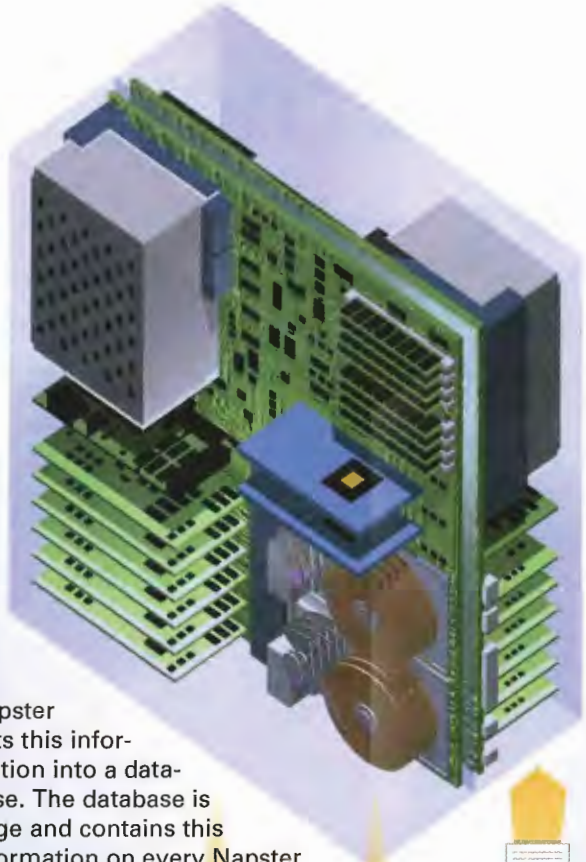
2 When you connect to the Internet and launch Napster, Napster contacts a Napster server. It tells the server all of the music on your computer, the location on your computer of that music, and your name and IP address.

Here it is.

6 You choose the song you want to download and from whose computer you want to download. You bypass Napster's servers and download the song directly from the person's computer.

Give me "Like a Rolling Stone."





4 When you want to search for music, type in the artist and song title for which you're looking in Napster.

3 Napster puts this information into a database. The database is huge and contains this information on every Napster user currently online and running Napster.

Here are the results.

Search: Dylan

Add songs from 192.168.1.120.

Artist/Song	Name	IP Address
Dylan/Like a Rolling Stone	GIGiggey	168.1.12.123
Dylan/Tangled up in Blue	GIGiggey	168.1.12.123
Davis/Kind of Blue	JazzHound	123.4.89.143
Coltrane/My Favorite Things	JazzHound	123.4.89.143
Coltrane/Giant Steps	JazzHound	123.4.89.143
Mozart/Eine Kleine Nachtmusik	ClassEguy	168.23.12.123
Cage/The Seasons	ClassEguy	168.23.12.123
Spann/Cryin'Time	BluSe	156.23.34.456
Ball/Red Beans	BluSe	156.23.34.456
Wolf/I'm a Man	BluSe	156.23.34.456
Earle/Transcendental Blues	CoLliYou	124.78.1.234

5 The Napster client contacts the Napster server. The server looks in its database and finds the artist and song title for which you're looking. The server then sends a list to you of every copy of every artist and song and shows you on whose computer you can find each.

CHAPTER

38

How Video on the Internet Works

THE Internet began as a way for people to share text-based information such as e-mail, discussion groups, and file transfers. Today, however, the technology has advanced far beyond text. Currently on the Internet, you can have video conferences in which you talk live with someone and see them live on your computer screen. You can use whiteboard applications that let you see and talk to other people at their computers, and you can also work on a file together live on your computer screens. You can watch live video footage of astronauts from outer space. Plus, you can watch taped videos whenever you want—not when a national broadcaster says you must watch them.

To understand how all this works, you must understand three types of technologies. The first is called the *MBone (Multicast Backbone)*; it's a special Internet high-speed backbone capable of sending vast amounts of information. Many video transmissions—especially live ones—are sent across the MBone because of its high bandwidth. Turn to Chapter 39, "How Multicast IP and the MBone Work," for more information on the MBone.

The second technology is called streaming video. *Streaming video* solves a long-standing problem of sending video signals across the Internet. Video files tend to be extremely large because they have so much information packed into them. Because of that, sending video was never very practical—it could take hours to send a single video file to someone's computer. The person on the other end would have to wait until the entire file was downloaded and then play it—and it might play for only a few minutes.

Streaming video solves the problem in two ways. First, it compresses the video file dramatically so it is much smaller as it is transmitted across the Internet. Secondly, streaming video lets the receiving computer start playing the video while the file is being transmitted. So if you receive a streaming video file, you watch the video as you receive it—no waiting for the entire file to download. Streaming video files are not usually live broadcasts. Instead, they are often files created ahead of time and then posted on the Internet. You can watch the video by clicking its hypertext link. You need a special player to watch the video. A number of ways exist to send streaming video across the Internet and watch it as it comes to your computer.

The third piece of technology is *videoconferencing*. It lets you use your computer to have live videoconferences across the Internet. Videoconferencing is done live, although the technology can also be used to broadcast taped videos as well. NASA, the National Aeronautics and Space Administration, sometimes uses the technology to broadcast live from the Space Shuttle and also to broadcast taped videos about space exploration.

Webcams, also called *digital cameras*, are another intriguing use of video technology. A *Webcam* is a camera attached to the Internet that automatically broadcasts photographs of moving images at certain intervals. Photos can be downloaded or browsed with your Web browser. There are hundreds of Webcams on the Internet, sending live pictures from all over the world.

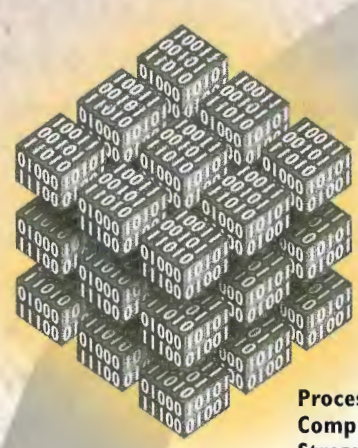
How Streaming Video Works

1 Streaming video refers to video you can play live on the Internet—you don't have to wait until the download is complete to see the video. Instead, you can play the video while it is being sent to your computer.

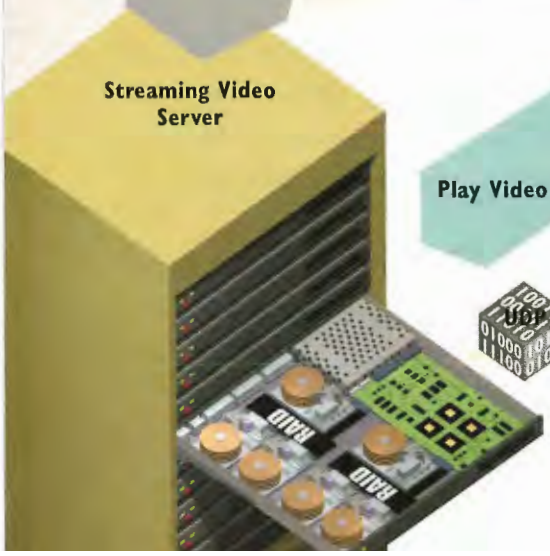
Analog Video Input

Digital Video Input

2 Before the video file is posted on the Internet, it is compressed and encoded in a special codec (coder/decoder)—an algorithm (mathematical formula) that compresses the video to a small size. This algorithm is required because without it, the video file would be so large that it would take an enormous amount of time to send it across the Internet.



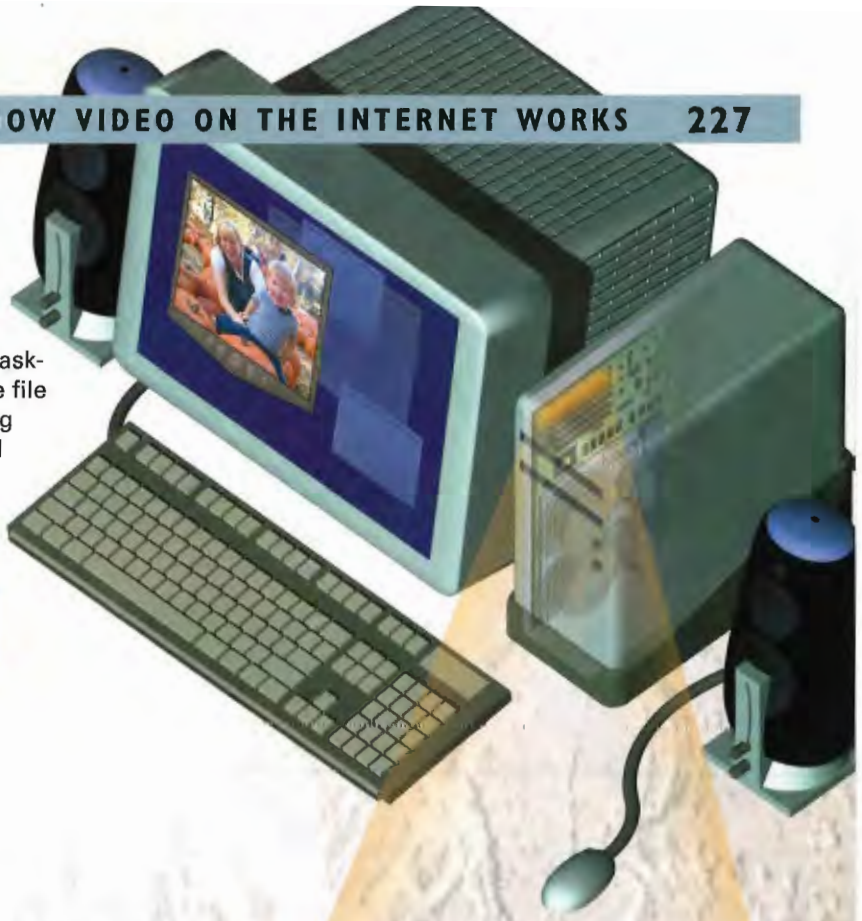
Processed and Compressed Streaming Video Files



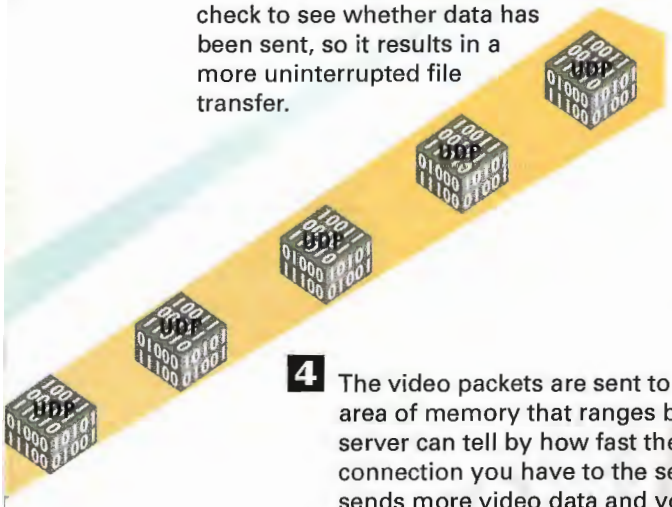
Streaming Video Server

Play Video File

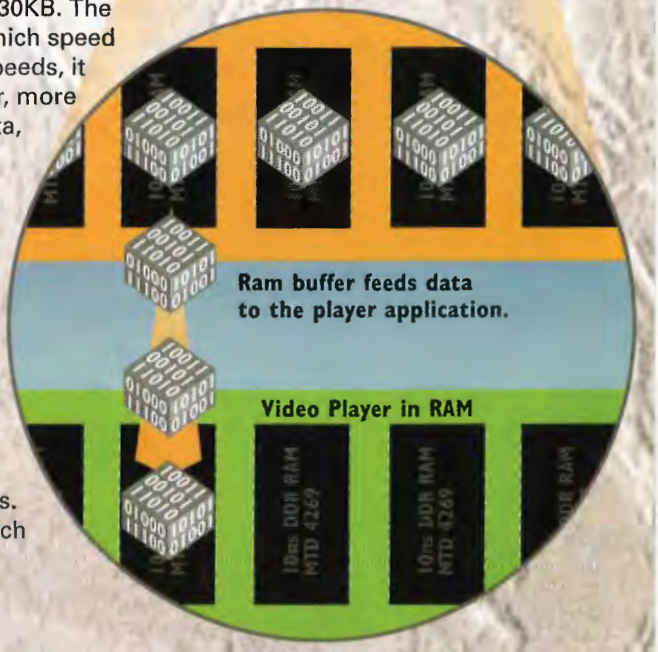
User Datagram Protocol



3 When you send a request to see the video by clicking an icon or a link on a World Wide Web page, you are sending a message from your computer to a server asking for the video file. The server sends the file to you in packets across the Internet, using the IP protocol. It does not use the normal Internet TCP, though. Instead, it uses the User Datagram Protocol (UDP). Unlike TCP, the UDP does not constantly check to see whether data has been sent, so it results in a more uninterrupted file transfer.



4 The video packets are sent to a buffer in your computer—an area of memory that ranges between 5KB and 30KB. The server can tell by how fast the buffer fills up which speed connection you have to the server. At higher speeds, it sends more video data and you get a smoother, more lifelike video. At lower speeds, it sends less data, which causes the video quality to suffer.



5 When the buffer fills up, which takes only a few seconds, a video player is launched on your computer. You can now watch the video on the player. As you watch the video, video packets are still being delivered to your buffer. Data from the buffer is continually sent to the player so you can watch an entire video. When all the video data has been sent, the video stops. The video file does not stay on your system; each section of the file is discarded after it is played.

How Videoconferencing Works

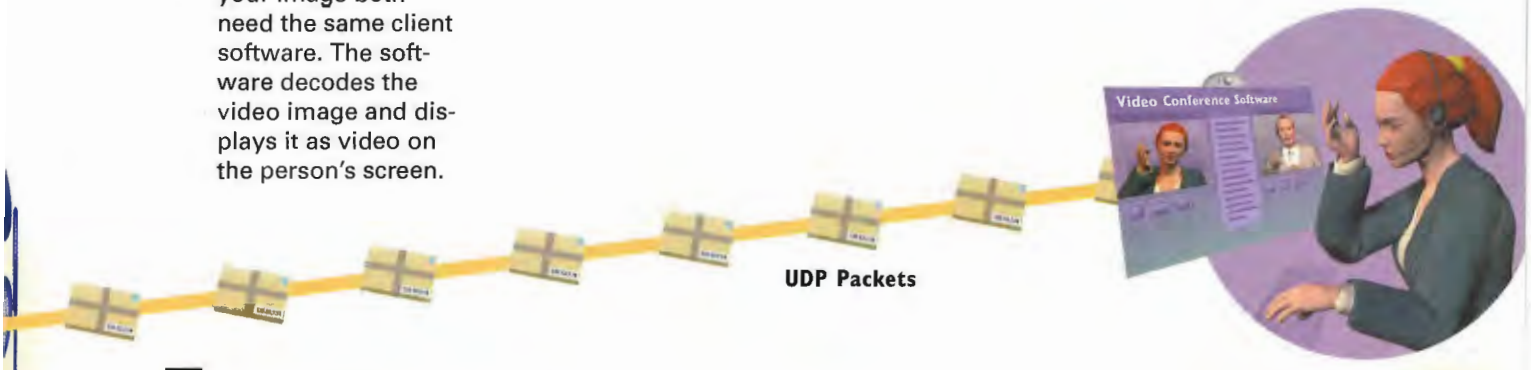
1 There are a variety of ways for people to videoconference across the Internet. Most require client software that enables people to send and receive video and audio signals. Also involved are special types of server hardware and software (sometimes called *reflectors*) that host the videoconference and send the signals to everyone connected to the server.

2 When one person wants to videoconference with someone else, he uses client software to log into a reflector. A *reflector* is an Internet computer that hosts many live videoconferences people can join. When you log into a reflector, you can join any conference that exists. When someone is logged into a reflector, a signal goes out regularly from the person's computer to the reflector, telling everyone connected to the reflector that the person is logged in and available for a videoconference.

3 To be a live participant in a videoconference, you need a video camera and a microphone on your computer. (Sometimes the microphone is built into the camera.) The camera converts your video image into digital data. The client software then compresses and encodes that data, enabling it to be sent across the Internet. If the data were not compressed and encoded, it would be too large to be sent.

Some types of videoconferencing software also provide a way to videoconference without having to go through a reflector. If you know the IP address of someone who wants to videoconference, you can connect directly to that person without having to go through a reflector. In this case, you can participate only in one-on-one conferences rather than group videoconferences.

6 You and the person who wants to see your image both need the same client software. The software decodes the video image and displays it as video on the person's screen.



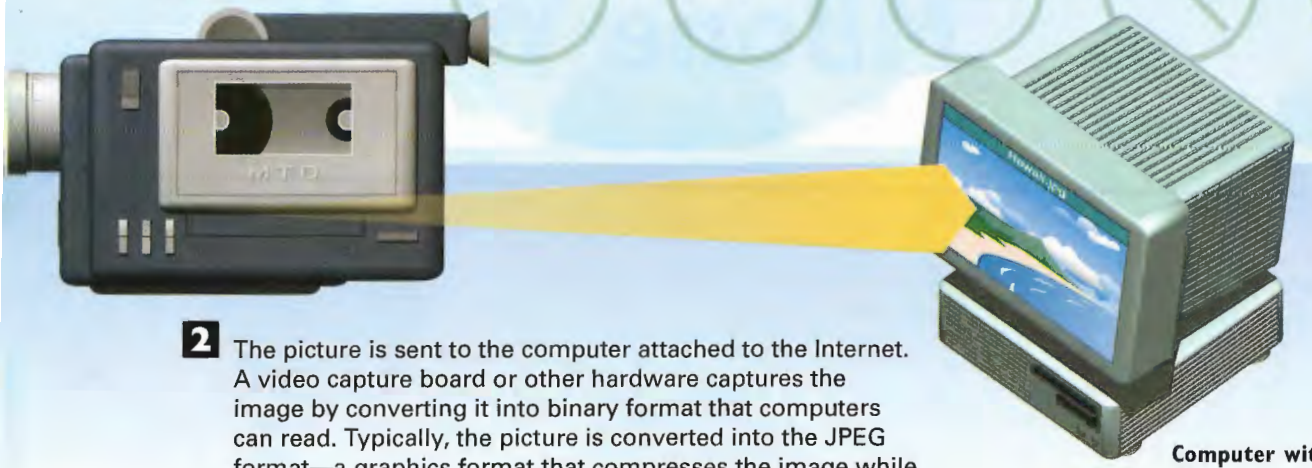
5 The video data is sent across the Internet using UDP, which is more efficient than TCP when you are sending video data.

4 As another way of cutting down on the amount of data that needs to be sent across the Internet, some kinds of client software send only part of the screen and not the entire video image. They send only the part of the image that has changed. So, for example, if someone moves her head, it transmits only the head moving and not the background, which hasn't changed.

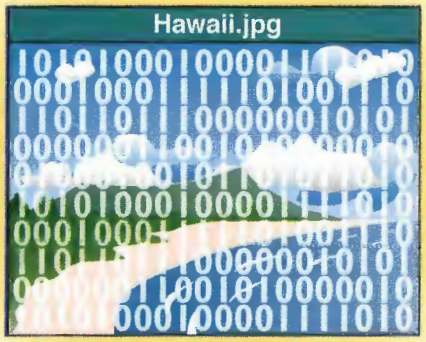
How WebCams Work

4 When someone clicks the link, the picture is sent to her Web browser and displayed. Some Webcams appear to send live video images, which means the Webcam image on your computer isn't a photograph, but instead appears to be a moving image. In fact, the "moving image" is a series of photographs sent every few seconds that give the illusion of movement. When you click the link to the image, the images automatically are sent to your Web browser as the video camera updates them.

1 A video camera is connected to a computer on the Internet. At regular intervals, one frame of the video is sent to a computer. It can be updated once every several seconds or once a day.



2 The picture is sent to the computer attached to the Internet. A video capture board or other hardware captures the image by converting it into binary format that computers can read. Typically, the picture is converted into the JPEG format—a graphics format that compresses the image while still retaining a high level of detail.



3 The JPEG image is linked to a specific URL on a Web home page. The link stays constant, even though the image itself changes regularly. That means whenever someone clicks the link, that person sees the most recent picture that was taken by the Webcam.

CHAPTER

39

How Multicast IP and the MBone Work

TO a great extent, the Internet of today is still in its infancy. Although multimedia elements can be found on the Internet, it's still largely made up of text and static pictures. These text and static pictures individually take up very little space—an entire Web page made up of text and pictures, for example, is often only 50 kilobytes (KB) or less.

The Internet of tomorrow, however, will be made up of a wide variety of multimedia elements: sound, video, animation, 3D objects, and more. Web pages will become interactive, and video-based shows might be broadcast over the Internet.

All this will cause serious congestion on the wires and networks that make up the Internet. (In fact, at times this congestion already occurs.) One more problem exists as well—there's no practical way for broadcasts to be sent out over the Internet because the files clog up the Internet. Suppose, for example, that someone wants to broadcast a telecast of a concert. The size of the file that contains that broadcast might be 50 megabytes (MB). Now imagine that 10,000 people want to watch the concert. That 50MB file needs to be sent individually to each of those 10,000 people. As you can imagine, that single broadcast could easily clog entire sections of the Internet, which would prevent the broadcast from being delivered.

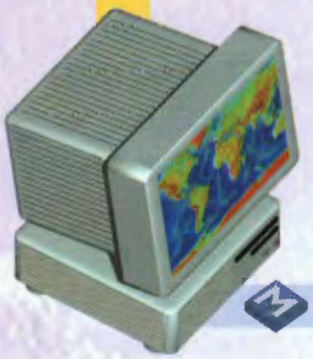
A potential answer is on the horizon, however, called the Multicast Backbone, or the MBone. The *MBone* is a high-capacity Internet backbone for transmitting broadcasts using the IP multicast protocol. The MBone enables broadcasts to start out as a single transmission instead of, for example, 10,000 transmissions. Inside that single transmission are the addresses of all the people who want to see the broadcast. As the file is sent across the Internet, it eventually makes copies of itself when necessary and delivers the broadcast to the networks and individuals who want to see it.

Suppose that 100 people want to see a broadcast of a 50MB file. Fifty people who want to see it are connected to the Internet via the WorldNet Internet service provider, 25 people are on a corporate network at `zd.com`, and another 25 use the Internet Access Company Internet service provider. When the broadcast goes out, it goes out as one single file, not 100 separate files. The file then splits into three parts: One part goes to WorldNet, one part goes to `zd.com`, and one part goes to the Internet Access Company. After the file is on each of those separate networks, it is delivered to the people inside the networks who want to see it. The key here, however, is that instead of 100 files of 50MB traveling across the Internet—5 gigabytes (GB) of data—only three 50MB files travel, or 150MB of data. As you can see, the MBone can cut down tremendously the amount of traffic traveling across the Internet.

How Multicast IP Travels Along the MBone

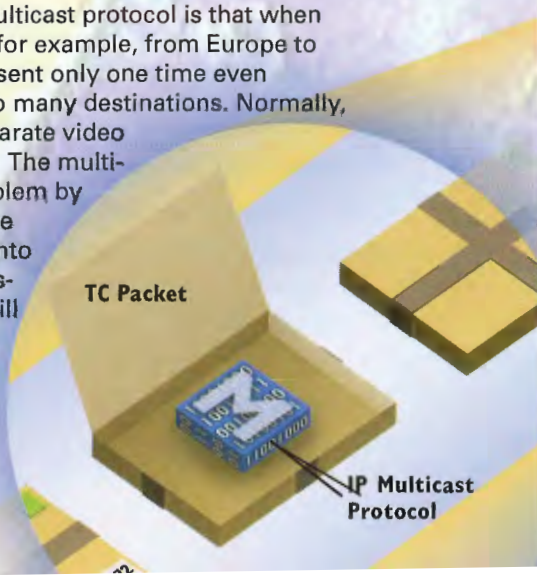
Video Camera

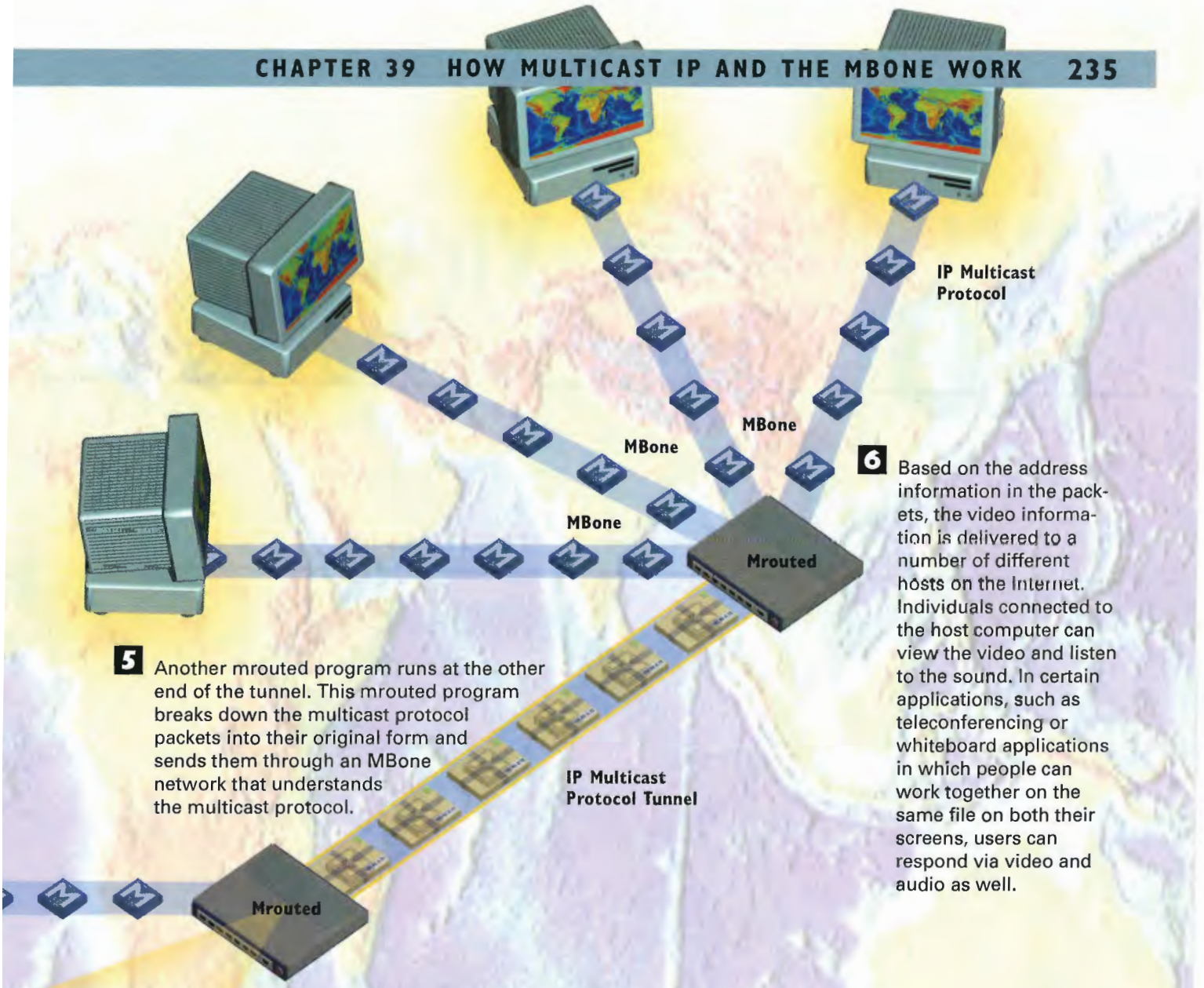
1 The *MBone* (Multicast Backbone) is a high-speed Internet backbone capable of sending live video and audio broadcasts. It's a network of host computers that communicate with one another using a technique called IP (Internet Protocol) Multicast. An MBone multicast begins when a video signal is digitized and compressed so that it can be sent over the Internet. Without compression, the signal would be too large and take too long to deliver.



2 The compressed, digitized signal is sent in packets using the IP multicast protocol instead of the Internet's normal TCP protocol. The multicast protocol enables the signal to be sent to a number of sites on the Internet simultaneously. Normally, the Internet is *unicast*, which means that each signal can be sent only to a single, specific location.

3 A major advantage of the multicast protocol is that when the video packets are sent—for example, from Europe to the United States—they are sent only one time even though they might be sent to many destinations. Normally, TCP would have to send separate video packets for each destination. The multicast protocol solves the problem by putting information about the many Internet destinations into one packet. Later in the transmission, the video signals will be delivered to each of the destinations.





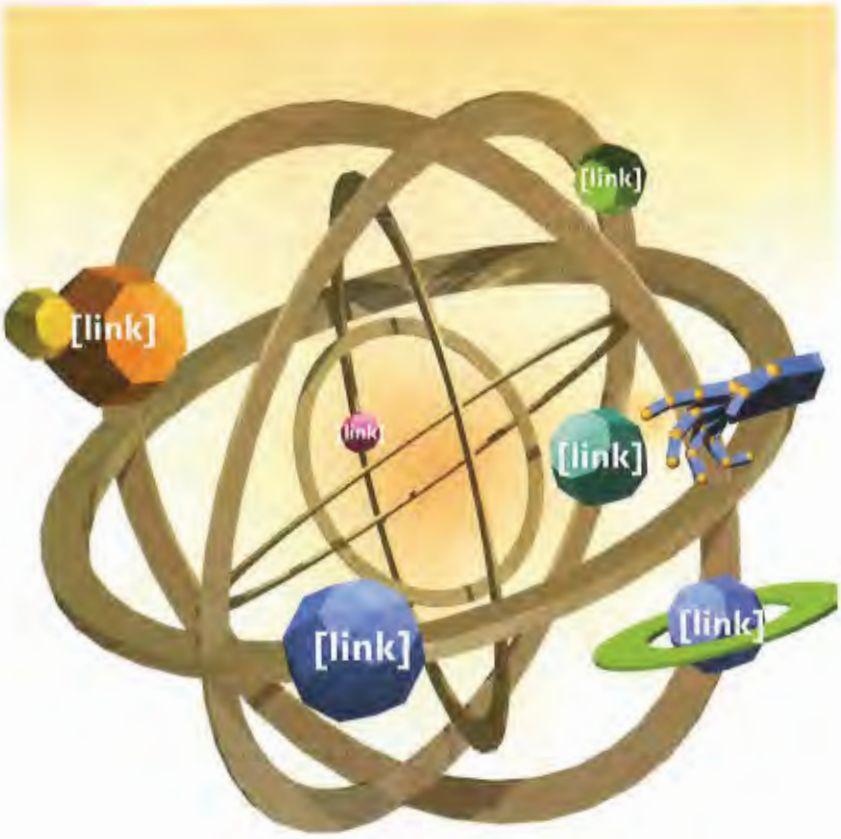
5 Another mrouted program runs at the other end of the tunnel. This mrouted program breaks down the multicast protocol packets into their original form and sends them through an MBone network that understands the multicast protocol.

6 Based on the address information in the packets, the video information is delivered to a number of different hosts on the Internet. Individuals connected to the host computer can view the video and listen to the sound. In certain applications, such as teleconferencing or whiteboard applications in which people can work together on the same file on both their screens, users can respond via video and audio as well.

4 The MBone understands the multicast protocol, but most networks and routers on the Internet do not. However, the MBone network often requires that data travel along normal Internet routes. To solve the problem, the MBone data travels in tunnels through existing Internet networks and routers. At one end of the tunnel is a Unix workstation that runs software called *mrouted* (multicast routing demon). This software encapsulates the multicast protocol packets inside normal TCP packets. To the Internet, the data now looks like normal TCP packets, and the data can be properly routed.

CHAPTER

40 How Virtual Reality Works



IMAGINE the Internet as a place where you could walk through three-dimensional worlds, pick up objects, examine them, and go to other Internet locations by flying or walking through doors. Picture home pages that were more than flat, two-dimensional surfaces that you could only read. What if you could be inside them, just like you can walk through a city or a building?

That's the promise of virtual reality (VR) on the Internet. In fact, it's more than just a promise—VR is already here. You'll find many virtual worlds you can explore on the Internet. You can walk through a giant computer, explore bizarre art galleries, visit outer space, go to the sites of what seem like ancient ruins, explore inside the human brain, and much more.

Virtual worlds are created using a computer language called Virtual Reality Modeling Language (VRML). This language instructs computers on how to build 3D geometric objects. Programmers and artists use the language to build complex worlds from these geometric objects. A VRML world is created by an ASCII text file containing VRML language commands—and for greater realism, graphics files can be added to this world as well. Because the virtual world is only an ASCII file, with perhaps a few graphics files, it can be downloaded quickly to your computer from the Internet, although some worlds with many graphics in them can be large.

When a virtual world is created, it is posted on an Internet server. When you want to visit that world, you either type in its URL or click a link to it, just as you do to visit any other location on the World Wide Web. To display the virtual world, you need a program capable of displaying the world—either a separate virtual reality browser, or more likely, a plug-in player that configures itself to your normal Web browser.

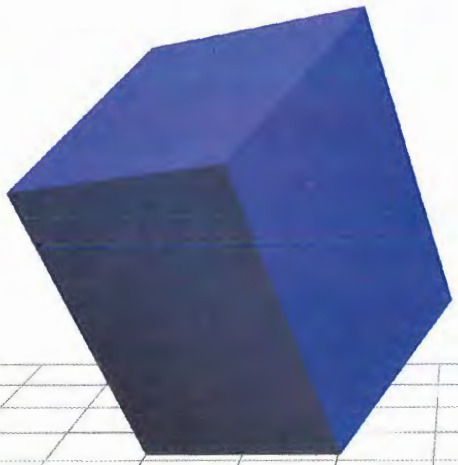
The VRML file describing the virtual world downloads to your computer. This can take a few minutes, or well over half an hour, depending on the size of the world and your connection speed. After the file is on your computer, your CPU computes the geometry of the world, based on the VRML commands in the file. Again, depending on the size of the world and the speed of your CPU, this can take only a minute or two, or up to ten minutes or more. After the world is computed, you can walk through it, fly through it, examine objects, and spin them. You can also visit other virtual worlds or places on the Internet by interacting with the world.

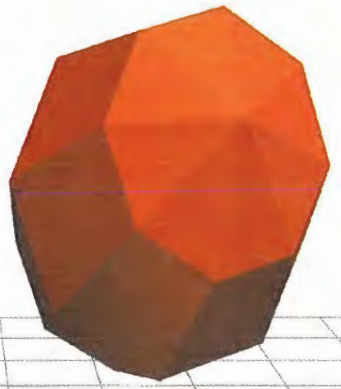
VR on the Internet is being used for far more than just creating virtual worlds people can walk through. For example, it has been used to create views of the brain and of molecules. It has been used by astronomers to show the rotation of molecular gas in a galaxy undergoing active star formation. Finally, as with everything else related to the Internet, VR will be eventually used for things that today none of us can imagine.

Despite all its appeal, one major problem with VR worlds is that they can be very large. Because of that, they can be very slow to load and interact with due to the current limited bandwidth of the Internet. As bandwidth increases, those problems might eventually go away.

How Virtual Reality Is Created by VRML

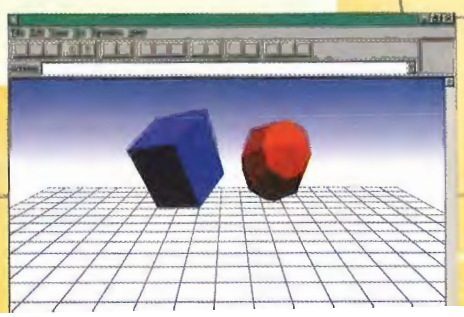
1 When someone wants to create a virtual world, she uses the Virtual Reality Modeling Language (VRML). VRML lets people create 3D worlds not by drawing them, but instead by using the VRML computer language to describe the geometry of a scene. VRML files are much smaller than graphics files. VRML files are simply text files that contain instructions for drawing the VRML world. VRML files end in a .WRL extension. After the world is created, it is posted on a Web server.



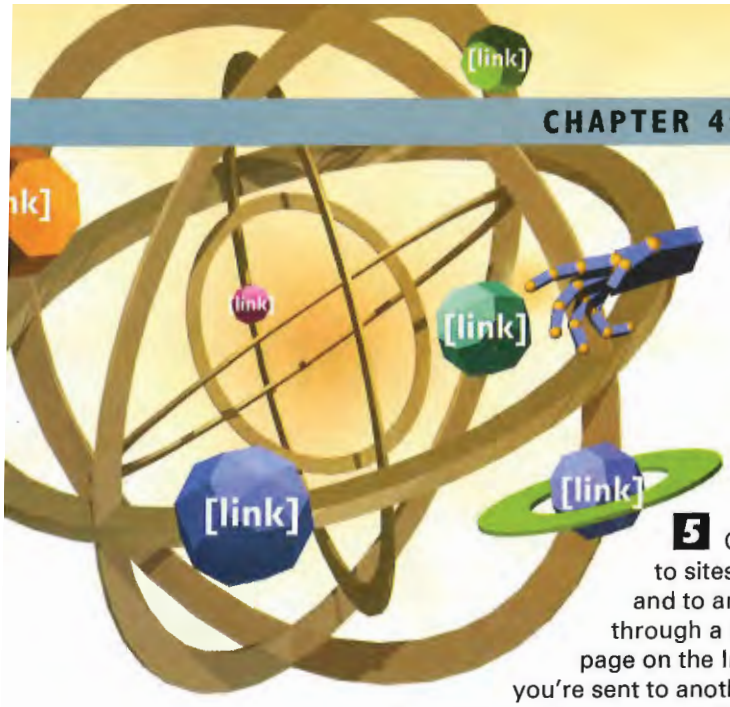


```
#VRML V1.Ø ascii
Separator {
   DirectionalLight {
      direction Ø Ø -1  # Light shining from viewer into scene
   }
   PerspectiveCamera {
      position    -8.6 2.1 5.6
      orientation -Ø.1352 -Ø.9831 -Ø.1233  1.1417
      focalDistance      1Ø.84
   }
   Separator {  # The red sphere
      Material {
         diffuseColor 1 Ø Ø  # Red
      }
      Translation { translation 3 Ø 1 }
      Sphere { radius 2.3 }
   }
   Separator {  # The blue cube
      Material {
         diffuseColor Ø Ø 1  # Blue
      }
      Transform {
         translation -2.4 .2 1
         rotation Ø 1 1  .9
      }
      Cube {}
   }
}
```

2 Here is an example of a VRML file describing a scene that has a red sphere and a blue cube in it, lit by a directional light.



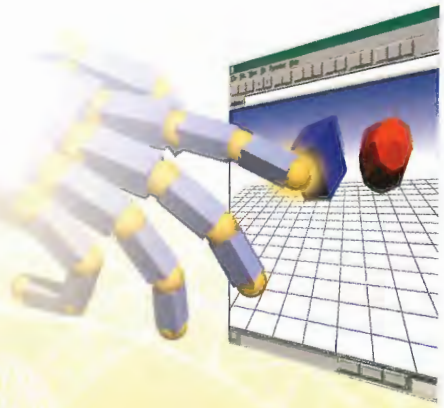
Web Server

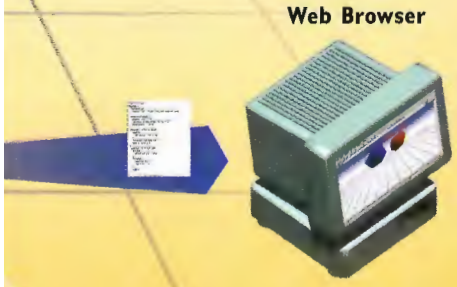


6 For greater realism and detail, graphics files can be "painted" on virtual reality objects—for example, to show paintings in an art gallery. When these graphics files are painted on objects, they must be downloaded along with the .wrl file as .gif or .jpeg files. When the browser displays the virtual world, it shows those graphics files on top of VR objects so they look like part of the scene.

5 Objects in the virtual world can be links to sites on the Web, to other virtual worlds, and to animations. So, for example, if you walk through a door, you might be sent to a home page on the Internet or to another virtual world. If you're sent to another virtual world, that virtual world must be downloaded from a Web server to your computer so your browser can compute the new world and you can interact with it.

4 As the file downloads, the VR plug-in is launched. It doesn't run separately from your Web browser. Instead, it takes over your Web browser while you're in the virtual world. After the file is downloaded, your VR plug-in creates the virtual world by taking the VRML commands in the file and having your computer compute the geometry of the scene. After the computation is done, the scene appears on your screen. The VRML file contains three-dimensional information that enables you to "walk" or "fly" through the scene using your browser. Depending on the complexity of the scene, your computer might have to do computations as you move through the scene.





3 When you have a VR plug-in installed on your Web browser, you can visit a virtual world by clicking its URL. First, the VRML file is sent from the Web server to your computer. Depending on the size of the virtual world and your connection speed, the file can take from a few minutes to a half-hour or more to download to your computer.

CHAPTER

41 Animation on the Web

ANIMATION on the Web works no differently than animation anywhere else. Just like in a flip book, animation is a series of still images displayed in succession to create an illusion of fluid motion. The faster the frames advance, the more fluid the animation becomes. Unfortunately, the Web can be a very slow place, and an animation that should run quickly often crawls across the screen unless special technology is used.

A number of different ways exist for creating Web animations, including client pull, server push, animated GIFs, and the Shockwave and Flash multimedia plug-ins. In *client pull*, an HTML page gives the browser instructions to request and load another document automatically, similar to a slideshow. Web pages are displayed one after the other with a specified time delay in between. This feature is useful for step-by-step instructions. But client pull is slowed by the need to load a whole page rather than a single cell of animation, which prevents the illusion of fluid animation.

Client pull requests are embedded within the HTTP response header of a Web page that the server sends back to the client. The `<META>` tag inserts meta-information into a response header. *Meta-information* helps parse a Web page, but the browser does not display it. A *response header* is the beginning of each HTTP response that a server sends back to a client with the requested Web page.

Server push is a complement to client pull, although server push is the more complex of the two. Server push requires a *Common Gateway Interface (CGI)* script that tells the server when to automatically serve a new document or image. It also requires the client browser to be capable of recognizing the MIME-type, `multipart/x-mixed-replace`. This MIME-type enables multiple documents to be sent within one message. To understand how server push works, imagine an e-mail message with text, hypertext, a digital movie, and sound. You can see how multiple "documents" (media types) can be sent within a single message. The multipart message is simply a series of images that displays one image right after the other. The server sends, or *pushes*, each image. In this way, a small animation can be embedded among the text and images of an otherwise static Web page.

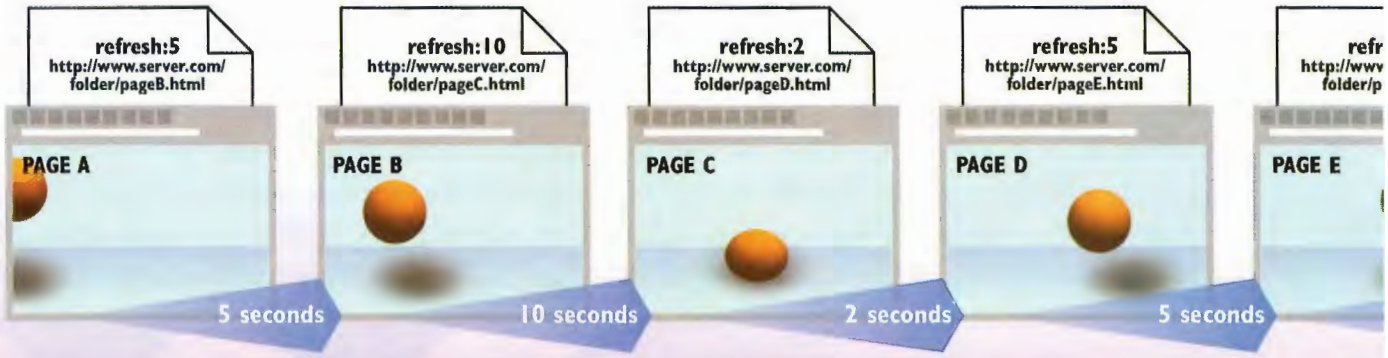
Animated GIFs are a series of graphical GIF images that roll up into a single image—much like the flip book seems to animate a series of drawings as you thumb through the pages. The animated GIFs load into a browser just like any GIF file, however, they load in a series to give the illusion of motion. Animated GIFs have the benefit of speed because images are cached on the client PC and loaded from memory rather than from the Internet. They represent an easy solution to adding motion to Web pages.

More complex multimedia animation has become possible using Macromedia's Shockwave and Flash plugs-in. Shockwave plays multimedia files that are created with Macromedia's popular Director and Authorware programs, whereas Flash uses its own special tools. You must first download and install the Shockwave and Flash plug-ins before you can view any Web pages that have Shockwave or Flash animations.

How Animation on the Web Works

Client Pull

Client pull is executed by the Refresh command. A refresh command is written into an HTML document using the <META> tag. The contents of the <META> tag are added to the header's meta-information that the server sends along with the HTTP response. During a client pull sequence, the browser reads this header information that instructs it to use your PC's internal clock to keep track of the time elapsed between pages retrieved. When the time has elapsed, the browser requests and displays the next page.



1. Each page in a client pull sequence can be located anywhere on the Web. The URL following the Refresh command might lead the browser to any active server. Page E is located on a different server than pages A–D, but is still requested automatically after five seconds.

2. If the next document to load also has a Refresh command in the header, the browser simply repeats the process. In this case, it retrieves and displays page C after 10 seconds.

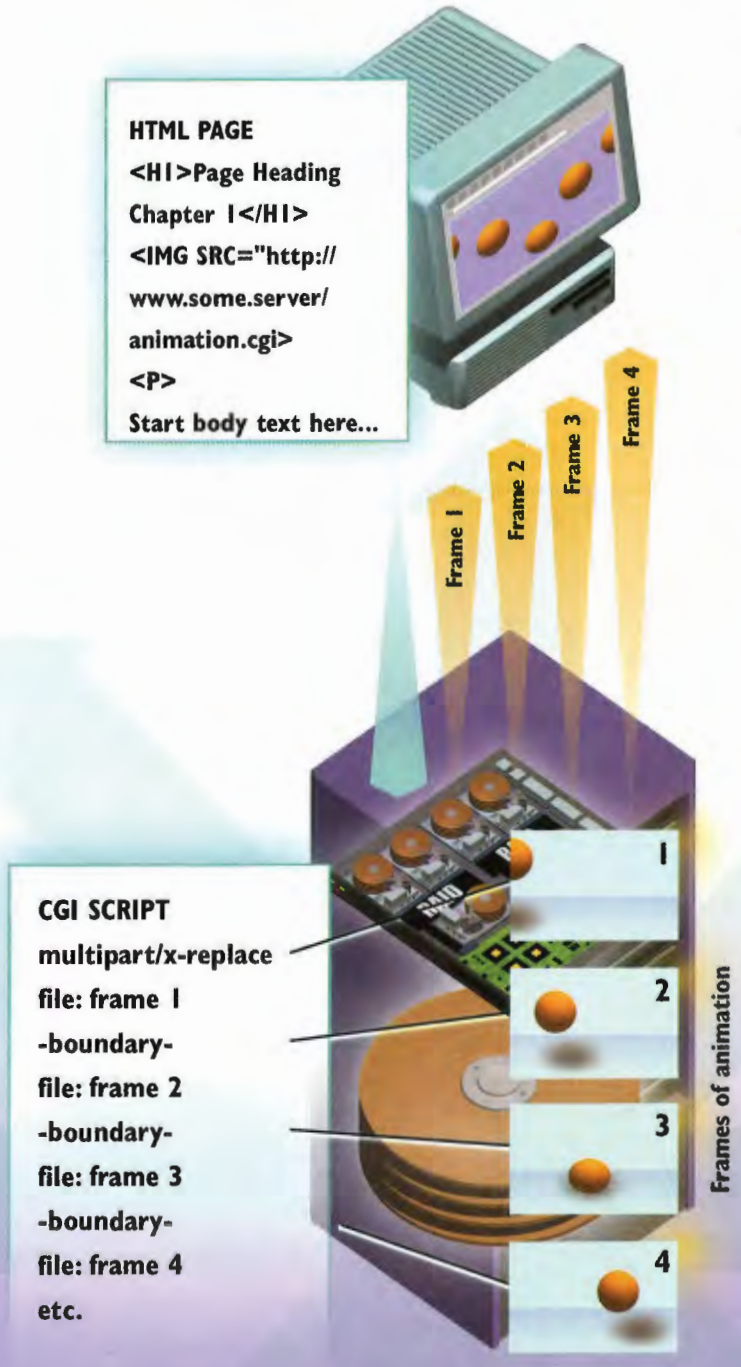
3. Whoever writes the HTML source code can specify how long it will be until the request for the next page is made. Page C refreshes after only 2 seconds, and page D follows.

4. The Refresh command does two things. First, it indicates the time before the next page request is sent or the same page reloads. For example, page A in the illustration refreshes after 5 seconds. Secondly, if the URL follows the number of seconds, a request for that page is sent automatically after 5 seconds. After the browser parses a document's meta-information and recognizes the Refresh command in the header, it knows to send a request for the page indicated by the URL following the command.

A client pull sequence might continue for as many or as few pages as the site designer wants. The last page will simply not have a Refresh command in the header. A user can stop the process manually by clicking the browser's Stop button.

Server Push

Server push is more complicated than client pull, but it enables inline animation that does not require an entire Web page to load each animation frame.



1. The HTML source code for a server push animation is deceptively simple. The `<IMG>` (image) tag references the animation just like a static picture or icon.

2. When the browser recognizes the `<IMG>` tag, it makes a single request to the server for a file. But rather than retrieving an image file, the HTML reference tag `<IMG SCR-URL>` gives directions to a CGI script that runs the animation.

3. When the request arrives at the server, the CGI script is opened and executed. (Recall that a programmer must write a CGI script—similar to authoring other kinds of software.)

4. The CGI script takes advantage of the `multipart/x-mixed-replace` MIME type. This enables the CGI script to send, or push, a series of still images from the server to the client as if it were transferring a single file. In this illustration, the animation has four frames, and each frame is a separate file. Each new frame that arrives at the client replaces the old one, which gives the illusion of fluid movement.

The server and client make one connection that is open for as long as the CGI script runs. You can manually end a server push animation by clicking the browser's Stop button.

How Shockwave Works

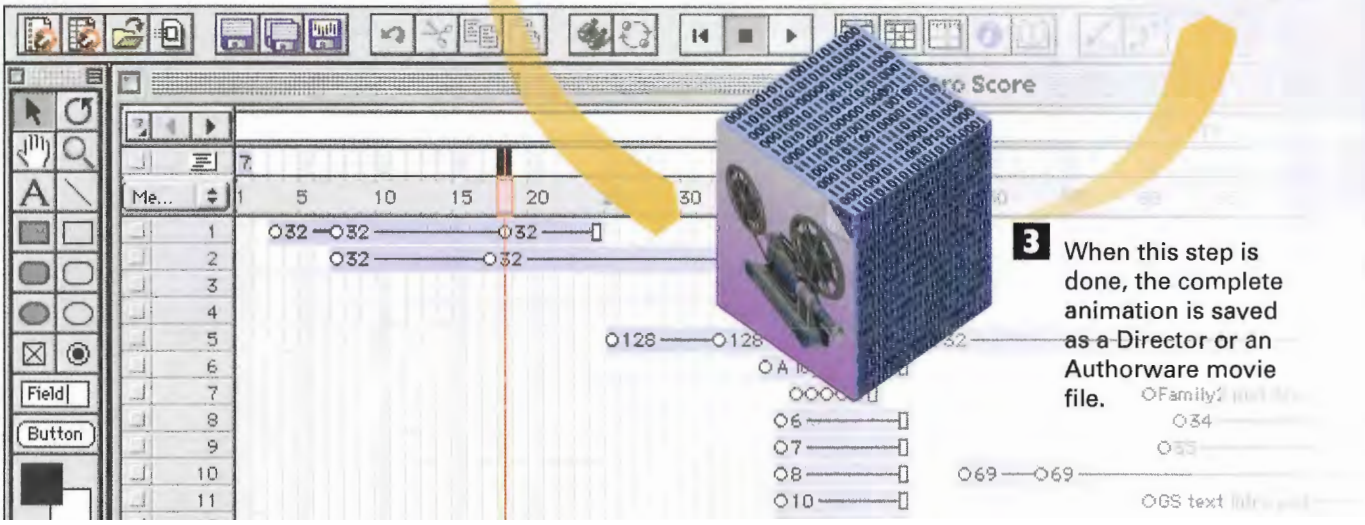
1 The first step in a Shockwave animation happens in a multimedia-authoring program such as Director or Authorware. An animation designer must gather the raw materials, such as still images, music, and sound effects, necessary for a short and compelling animation.

2 The authoring program then helps arrange the elements frame by frame along a time line. It also enables the designer to match a sound effect with a particular action in the animation.

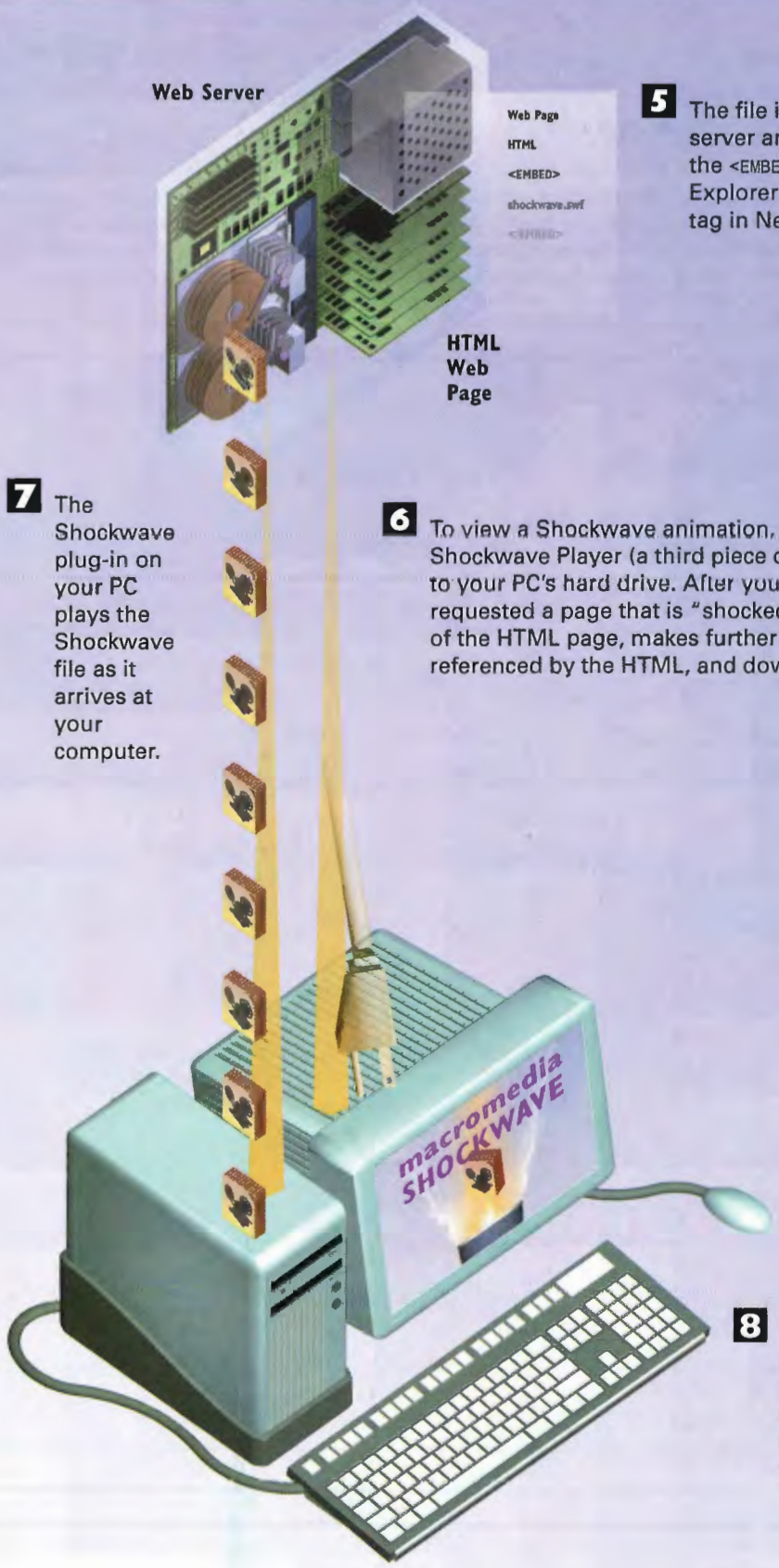
Director Software from Macromedia

4 Next, the movie file must be converted and compressed into a small file that can be quickly downloaded to a user's PC.

Converts Director File to Shockwave File



3 When this step is done, the complete animation is saved as a Director or an Authorware movie file.



5 The file is put onto a host server and referenced by the <EMBED> tag in Internet Explorer and the <OBJECT> tag in Netscape Navigator.

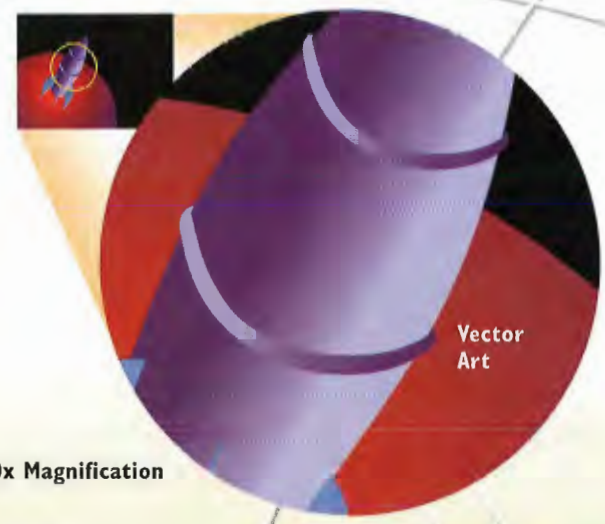
6 To view a Shockwave animation, you must have downloaded the Shockwave Player (a third piece of software from Macromedia) to your PC's hard drive. After you've downloaded the player and requested a page that is "shocked," the browser displays the text of the HTML page, makes further requests for other graphics referenced by the HTML, and downloads the Shockwave file.

7 The Shockwave plug-in on your PC plays the Shockwave file as it arrives at your computer.

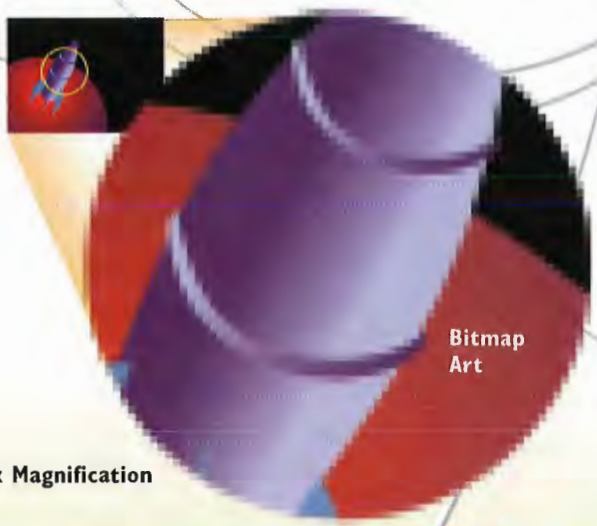
8 The Shockwave Player displays the final product as an inline animation. The player also reads any sound information and simultaneously sends it to your PC's audio hardware. Shockwave closes when it is no longer needed.

How Flash Works

1 Flash is an animation program that enables designers to add animation, sound, and interactivity to Web pages. Flash uses vector graphics instead of bit-mapped graphics. *Vector graphics* are mathematical descriptions of a shape, while a *bit-mapped* image is an actual pixel-by-pixel representation of the image. Not only are vector graphics smaller and more suitable to the Web because of reduced download time than bit-mapped images, but, unlike bit-mapped graphics, they can also be zoomed in on without any degradation in quality.



10x Magnification

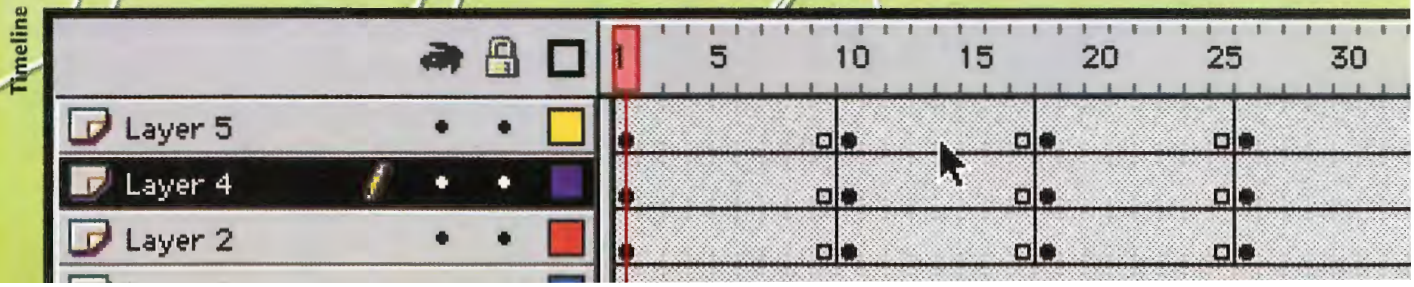


10x Magnification

2 To create a Flash movie, a designer uses traditional animation techniques. He draws a picture in a frame, draws another picture slightly different than the first, then a third slightly different from the second, and so on.

The art can change between frames with sound effects and more!

3 These pictures are all places on a Flash timeline and are displayed one after another—appearing to move. The animator has control over the speed at which the frames replace one another, or *frame rate*.



Web Browser

6 When someone visits a site with an HTML reference to the .swf file, the Flash movie begins to play. If the person doesn't have the Flash player, he is prompted to download it for free from the Macromedia site. The movie begins to play. A big benefit of Flash is that it *streams content*, which means that the movie can begin playing while the rest of the movie downloads in the background. This means that Flash movies can start playing very quickly even if they are very long ones.

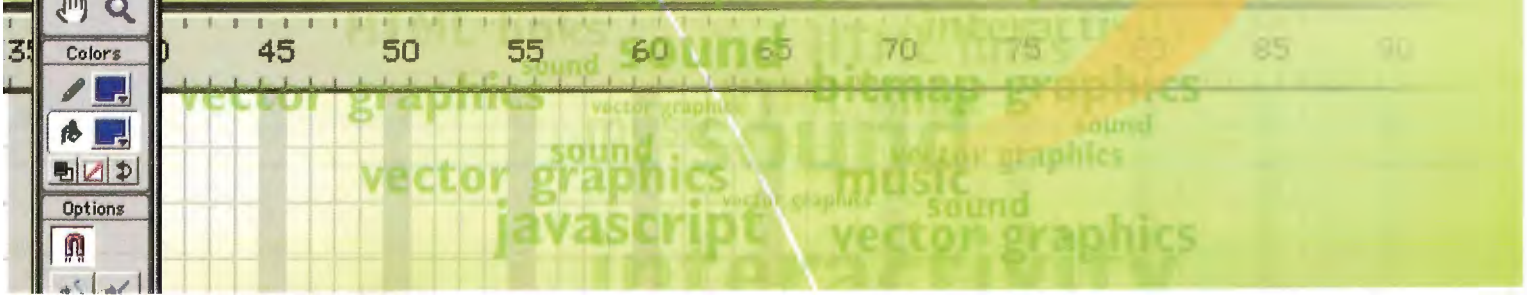
Web Server

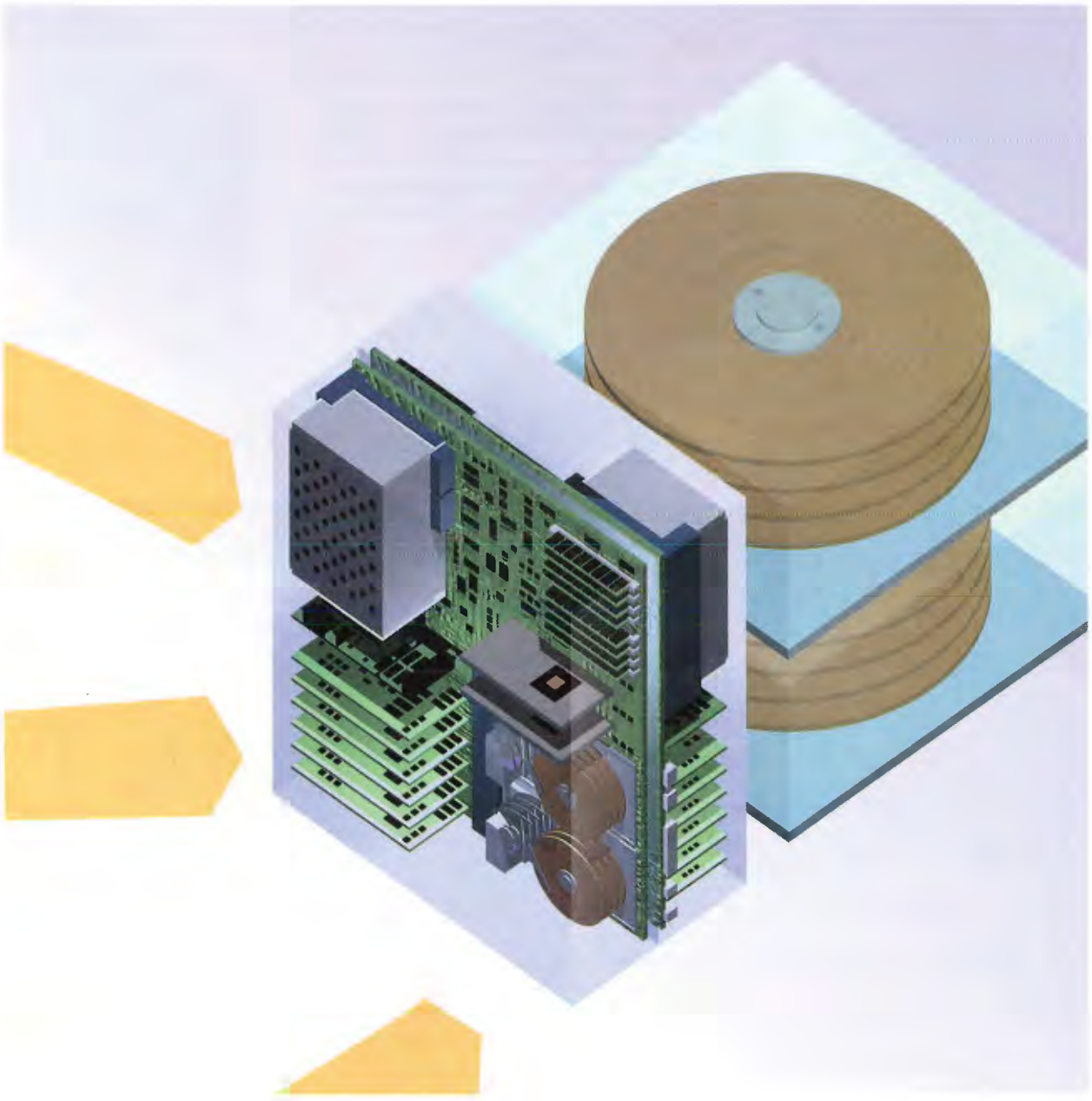
5 When a designer finishes making a movie, it is compiled into a file with an .swf format and posted on a Web site. Anyone who views it needs a Flash player, which is available for free and is bundled into many versions of browsers.

Flash .swf File

Tool Bar

4 The timeline is used not just for images, but for sounds as well. To play sounds at specific times, the animator drops the sound into the timeline. The Flash movie can also include links to Web pages, JavaScript commands, and other kinds of interactivity. Again, all these are dropped into a timeline.





PART 7

SHOPPING AND DOING BUSINESS ON THE INTERNET

THE Internet is no longer a self-enclosed club with no connection to the outside world. It has become intimately tied to the way we live and work, and it will become more so with each passing year. At work, for entertainment, to get information, and to shop—the Internet is increasingly becoming a part of our daily lives.

The Internet might have its roots in the military and in academia, but its dramatic growth has been fueled in large part by business and consumers. The Internet might become one of the primary places in which businesses operate, and where hundreds of billions of dollars in goods and services will be bought and sold every year. In fact, already billions of dollars are spent every year in online shopping, and the use of the Internet has become a way of life at just about every business in the country.

Thousands of businesses use the Internet to market and sell their products, and many people buy things from home and from their places of business through the Internet instead of at retail stores. You can use the Internet to browse through catalogs and make purchases online; to buy and sell stock, mortgages, and insurance; and even to participate in online auctions. Companies are figuring out ways not only to sell online, but also to hook those online transactions into their internal computer and billing systems.

This part of the book looks at the various ways the Internet is used for business and shopping online. It covers how the Internet is being used by businesses as their primary corporate networks and how business and commerce are being conducted on the Internet every day.

Chapter 42, "How Intranets Work," covers intranets. Intranets are private networks set up by companies for their employees, using Internet technology. They're used for many purposes, including e-mail, group brainstorming, group scheduling, and access to corporate databases and documents, among others.

Although intranets use TCP/IP networks and technologies, the network and its resources are used only by the businesses and are not available to people outside the company. Intranets are separated from the Internet by firewalls that don't allow unauthorized access to the intranet. People who work in the company can access the intranet and use its resources, but intruders are kept out by the firewalls.

This chapter also looks at one of the most important parts of an intranet—workgroup software. This type of software ties together everyone in a corporation and enables them to work together better. Among other things, it enables people to share files and information; to cooperate more easily on projects; and in general, to work together in ways never before possible. It allows people to go beyond simply communicating and enables them to work together on shared documents.

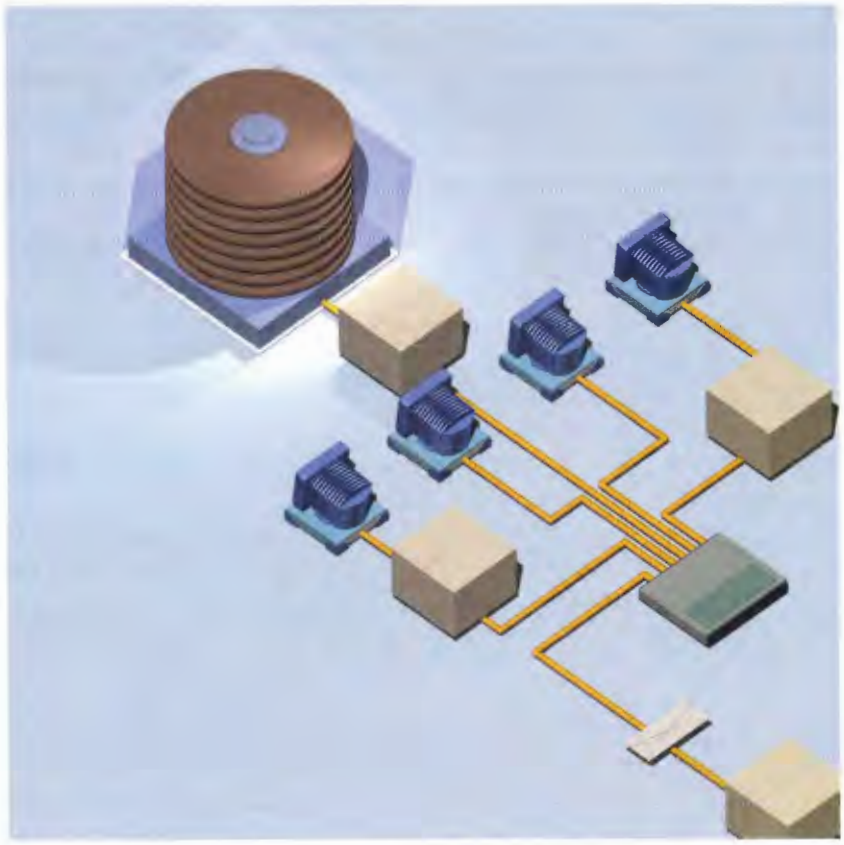
This chapter examines a variety of workgroup software. It covers messaging software that enables people to publicly participate in group discussions. It also looks at whiteboard software, which enables people to see what is on other people's computers and work together on documents. Several people could look at a spreadsheet together, for example, and one person could mark up the spreadsheet while everyone else sees what is being done.

Chapter 43, "Shopping on the Internet," covers what has become one of the most popular parts of the Internet—shopping online. Today, shopping on the Internet accounts for billions of dollars a year in revenue, and every year many more billions are spent. In fact, you can't turn on your television set or open a newspaper or magazine without being confronted by advertising for a variety of online shopping sites. Although many of the original shopping sites have gone out of business, they've been replaced by the very businesses they expected to supplant—existing retailers. So, today you'll find stores such as the Gap and Wal-Mart—so called bricks-and-mortar stores, or bricks-and-clicks—with big online shopping sites.

This chapter shows you what's going on behind the scenes when you shop online. It shows you how online shopping carts work—a technology that lets you gather together goods you're thinking of buying into a virtual shopping cart and then go through a checkout with them and pay by credit card. This chapter also details one of the newer and more intriguing shopping technologies—eWallets. When you use an eWallet, you don't need to put in personal and credit card information on the various sites where you shop; you just fill out the information once, in your eWallet, and that information is automatically sent to the shopping site when you want. Finally, this chapter covers one of the most popular ways to shop online—buying at online auctions. Every day, millions of people buy and sell millions of items through auction sites, such as the popular eBay, and through a variety of auctions powered by technology developed by a company called FairMarket. You'll see, in this chapter, how technology enables auctions to work.

CHAPTER

42 How Intranets Work



BUSINESSES increasingly use Internet technology to create private corporate networks called *intranets*. These intranets are used for a wide variety of purposes, such as e-mail, group brainstorming, group scheduling, access to corporate databases and documents, videoconferencing, and the buying and selling goods and services.

Intranets use TCP/IP networks and technologies as well as Internet resources such as the World Wide Web, e-mail, Telnet, and FTP. However, the network and its resources are used privately by businesses and are not available to people outside the company. An intranet is separated from the rest of the Internet by a *firewall*—a hardware and software combination that prohibits unauthorized access to the intranet. People who work in the company can access the Internet and use its resources, but firewalls keep out intruders. Turn to Chapter 44, "How Firewalls Work," to learn more about firewalls.

Intranets use a combination of off-the-shelf software, such as Web browsers, and customized software, such as database querying tools. Because intranets are based on Internet standard protocols, it will always be possible to quickly update them with the latest in network technologies.

In the long term, companies will make the most use of intranets in *workgroup applications*—software that enables people to work cooperatively with their computers. Many kinds of workgroup software exist. These programs enable people to participate in discussions and videoconferencing across the country and across the world, share databases, track documents, and much more.

The key to workgroup software is that it enables people to go beyond simply communicating and lets them work together on shared documents.

One of the most basic pieces of workgroup software is *messaging software*—programs that enable people to publicly participate in group discussions.

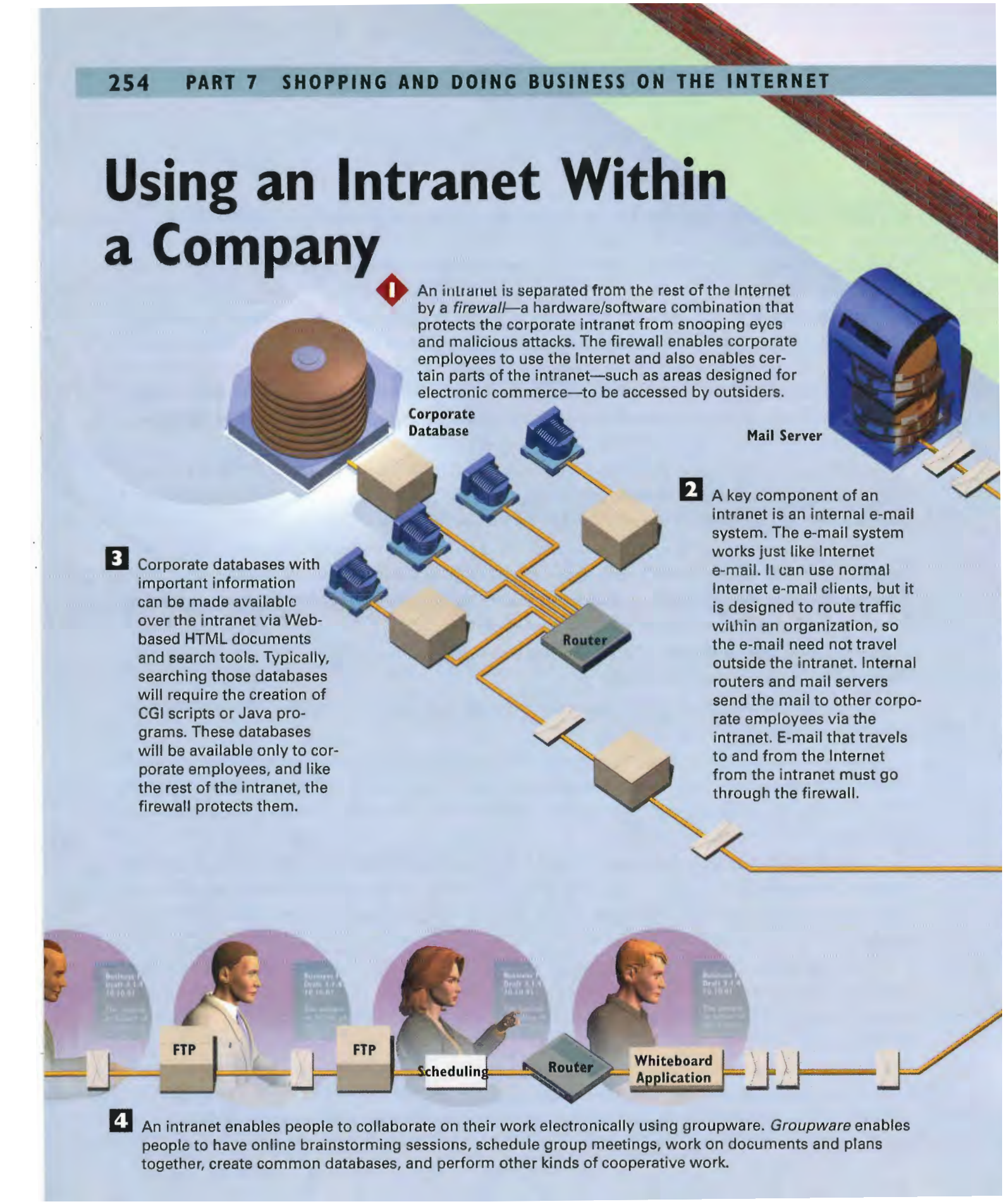
What makes intranet messaging software especially useful is the way it integrates with other Internet and intranet technologies. For example, some discussion software allows the use of Hypertext Markup Language (HTML) embedded inside messages. This means that from within a discussion, someone could embed a link to a Web page or other intranet resource.

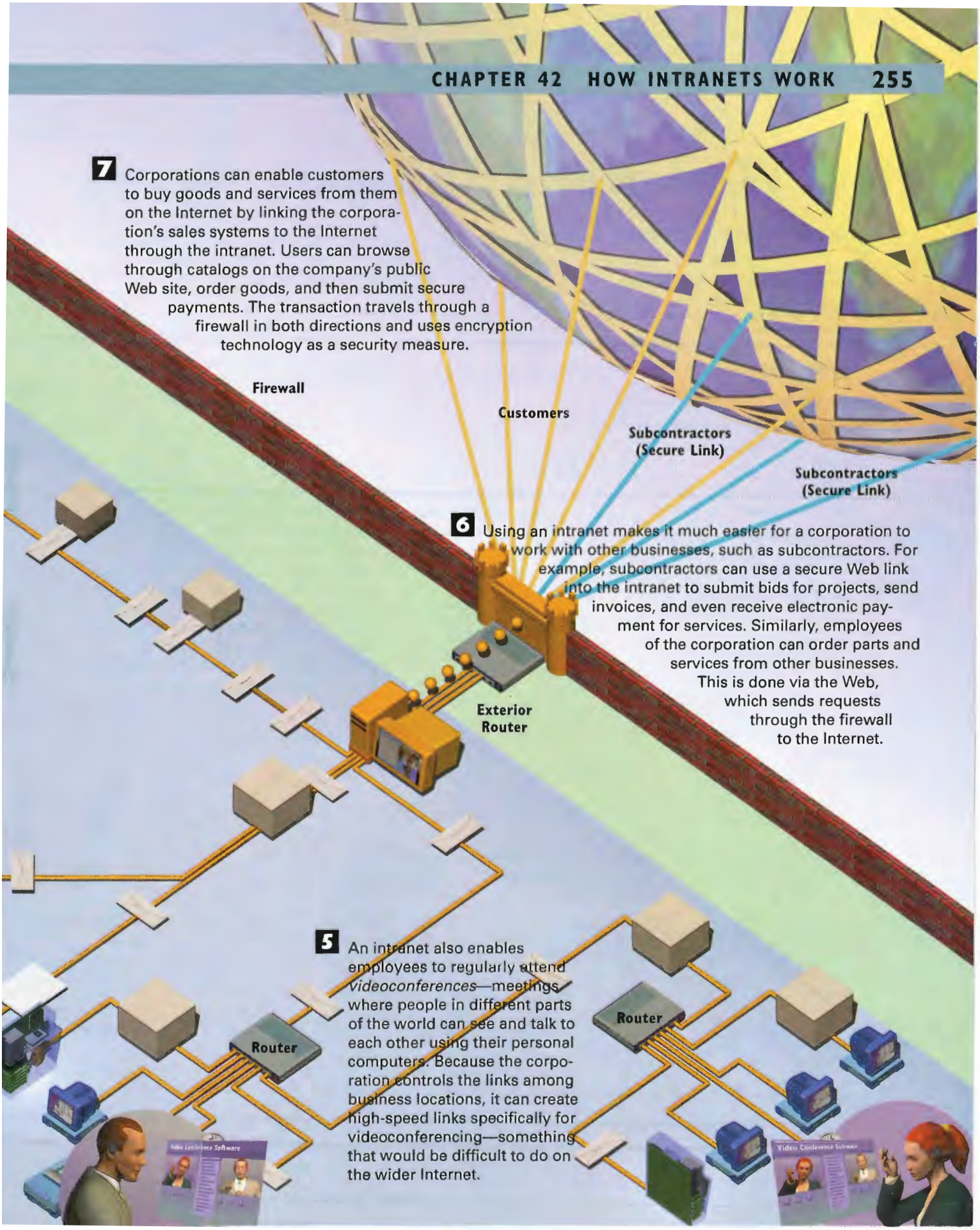
A more sophisticated workgroup application is *desktop videoconferencing*. This application requires that everyone involved have computer-linked video cameras as well as hardware and software that enables computers to send and receive voice and sound. While sitting at computers, people can see and speak to one another.

A related technology is called whiteboard software. *Whiteboard software* lets people see what is on someone else's computer on an intranet while sitting at their own computers. This means people on the same intranet—whether they're on opposite sides of the country from one another—can easily comment on one another's work.

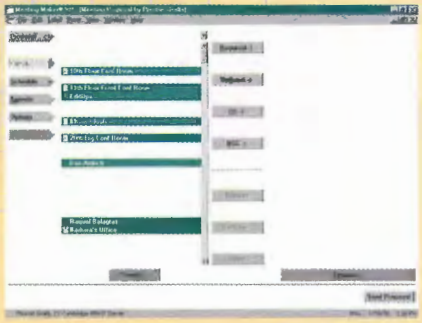
Document management software and workflow groupware are useful for intranets in companies that have complicated work procedures, or where many people must cooperatively put together a single document. These types of software streamline the way that work flows through a company and enable businesses to operate much more efficiently.

Using an Intranet Within a Company




7 Corporations can enable customers to buy goods and services from them on the Internet by linking the corporation's sales systems to the Internet through the intranet. Users can browse through catalogs on the company's public Web site, order goods, and then submit secure payments. The transaction travels through a firewall in both directions and uses encryption technology as a security measure.
Firewall
Customers
Subcontractors (Secure Link)
Subcontractors (Secure Link)
6 Using an intranet makes it much easier for a corporation to work with other businesses, such as subcontractors. For example, subcontractors can use a secure Web link into the intranet to submit bids for projects, send invoices, and even receive electronic payment for services. Similarly, employees of the corporation can order parts and services from other businesses. This is done via the Web, which sends requests through the firewall to the Internet.
Exterior Router
5 An intranet also enables employees to regularly attend *videoconferences*—meetings where people in different parts of the world can see and talk to each other using their personal computers. Because the corporation controls the links among business locations, it can create high-speed links specifically for videoconferencing—something that would be difficult to do on the wider Internet.
Router
Router
Video Conference Software
Video Conference Software

How Workgroup Software Works

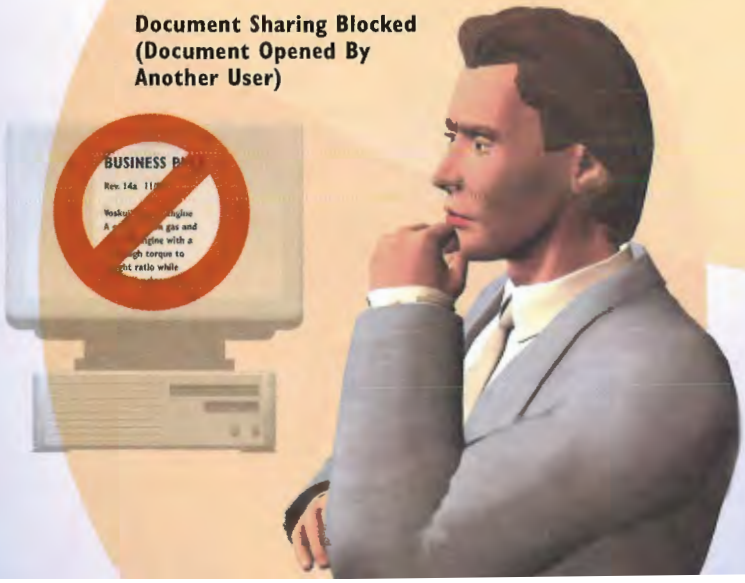


1 Discussion software enables people from within a corporation to exchange work and ideas. Links to other intranet resources are included in the software. From within a discussion, people can link out to a Web page on the Internet or intranet, or even link into intranet databases, servers, or shared applications. Additionally, software can replicate intranet discussions onto Internet newsgroups. From one discussion area, people on an intranet can hold discussions with people from within their company or people out on the Internet.

2 Document management software and workflow groupware enable intranet administrators to create systems that track and control access to documents through every aspect of their creation—for example, allowing only one person at a time to check a document out of a library. They can provide a *version history* of every document so anyone can see who has worked on a document along with the changes that person made. The administrators can also give certain people the right to lock the document so no further changes are allowed.

Document Sharing and Management

Document Sharing Blocked (Document Opened By Another User)



Whiteboard Software

5 Whiteboard applications will be a popular workgroup use of intranets. In a *whiteboard application*, two or more people can see and talk about what is on each other's computer screens across the intranet. Additionally, they can mark up what is displayed on each other's screens.

4 With intranet groupware, videoconferencing can finally be a corporate reality. Desktop-conferencing software enables two or more people to see and talk to each other on their computer screens as long as they have cameras connected to sound-equipped computers. Because intranets can be built using very high bandwidth connections, having a video-conference across an intranet is possible, although doing so across the Internet can be much more difficult because of the lower bandwidth of the Internet.

Video Conferencing

Desk-to-Desk Chat

3 Groupware can allow for desk-to-desk chats. People can sit at their computers and directly communicate with others sitting at their computers by typing on their keyboards. What one person types at the keyboard shows up on another person's computer screen and vice versa.

CHAPTER

43

Shopping on the Internet

BILLIONS of dollars are spent every year shopping on the Internet—and if think tanks and market research firms are to be believed, that's only the beginning. The Internet will eventually revolutionize shopping in the same way it's revolutionized working, getting information, and communicating.

Online shopping is possible through the use of *encryption techniques*—the ability to scramble information as it's sent through the Internet so that no one can read it except the intended recipient. Encryption is used to scramble credit card information—the primary way that people pay when buying online. (For more information about encryption, and to see how encryption works, turn to Chapter 48, "Cryptography, Privacy, and Digital Certificates." To see how encryption can keep e-mail private, turn to Chapter 16, "How E-mail Works.")

Most of what you see when you visit a shopping site on the Internet is contained in databases on Web servers. These databases have information about the products for sale at the site—and they're also used to automatically generate the HTML pages that make up the shopping site. So, for example, when a new product becomes available, information about that product is put into a database, and CGI scripts and a Web server then work with the database to create a new item on a Web page describing the product. You, in turn, can look at that product and decide whether you want to buy it.

Databases and cookies are also used when you use *virtual shopping carts*—portions of a Web site where you place items you're considering buying. Before buying, you can take items out of the cart or can put new items in. *Cookies* track everything you put into and take out of the cart, and then *databases* work with the cookies, CGI scripts, and Web servers to complete the transaction when you want to buy something.

Web databases are also used to complete the shopping transaction when you buy. When you decide you want to buy something from a site, fill out a form, and send in your credit card information, that information is sent to a Web database. The database, in turn, checks the validity of your credit card. If it's valid, the database sends a confirmation to you and then sends off an order to a warehouse or other distribution method that ships the product to you. Databases can't do all this by themselves—they work in concert with CGI scripts, Web servers, and cookies.

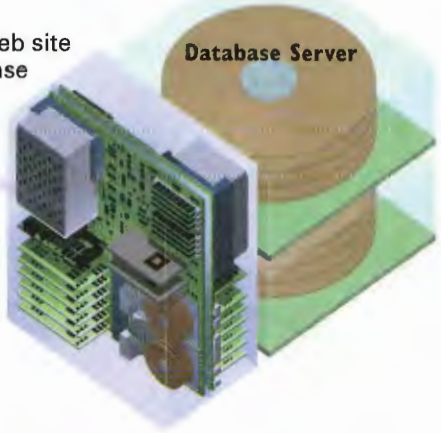
This chapter looks at how online buying and virtual shopping carts work. This chapter also examines one of the newer shopping technologies—electronic wallets, or *ewallets*. Finally, you'll see how one of the most popular kinds of buying sites work—*online auctions*.

How Online Buying Works

1 Most shopping sites are built on top of databases, so when customers visit a Web site and browse or search for a product, they're actually searching through a database that is searched from the Web.

Product Listing

VCRs
Digital Cameras
Video Cameras
CD Players
Printers
Monitors



Entering Secure Area

2 When customers see a product they want to buy, they'll usually pay by credit card. Before filling out a form with their credit card information, they're usually sent to a secure section of the Web site where encryption will be used to scramble the data.

6 The site confirms the order, and using CGI scripts, the Web page refreshes and displays a page that the customer can print out to confirm the order. Many sites also follow up by sending an e-mail message.

Thank you for your order.

3 After customers are in the secure area, they fill out a form that includes a request for their credit card information. As they fill the forms out, their information stays on their computers and isn't yet sent out over the Internet.

5 The transaction server receives the encrypted information and decrypts it. It then checks with the credit card company to ensure that the card is valid and can be used, in a similar way to how a retail stores checks that your card is valid, except that it's usually done over the Internet.

Transaction Server

4 When the form is filled out, the customer clicks a Submit button, or something similar, to send the information from the customer's computer to the site's secure transaction server. As the information is sent out over the Internet, it's encrypted so that it's nearly impossible to read, except by the site itself.

Ok to accept?

OK

Ok to ship the goods.

7 The transaction server sends an order to the warehouse or other designated area that fills the order, and the order is completed as any other order is, by shipping via the mail or express mail service.

How Online Shopping Carts Work

2 When the person completes the registration form, it's sent to the Web database. The database creates a record for that person and sends the person a *cookie*—a small piece of data that sits on the person's hard disk and can be used to identify that person.

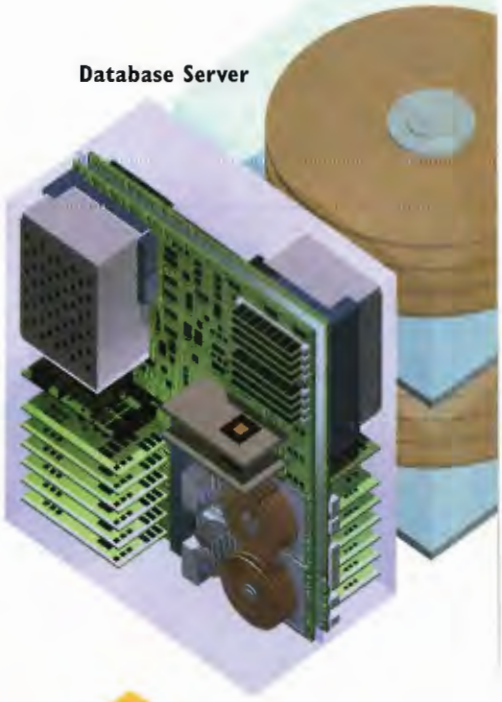
4 The Web server in turn writes a new piece of data to the cookie, which identifies the item that the person wants to put in his shopping cart. More than one item can be put in the shopping cart in this way.

5 When the person is ready to check out his items, he goes to a Web page containing his shopping cart. When he goes to the page, the cookie tells the Web server what items to display on the page.

1 Before a shopping cart can be used, a person must register with the site by filling out such information as name, address, and other personal information. Sometimes a credit card number is required as well.

Register Me

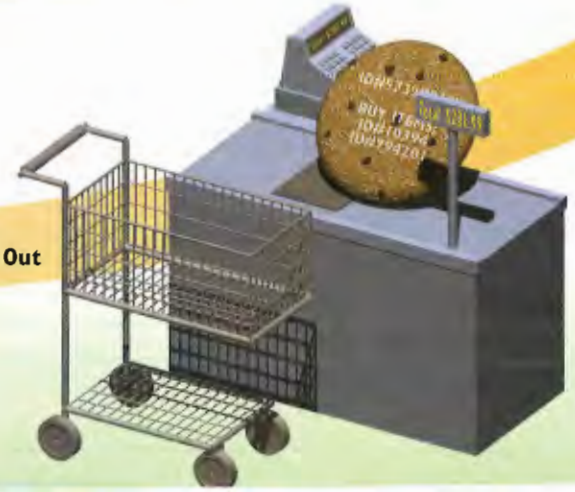
Database Server



3 When the shopper thinks he wants to buy something, he clicks it, which puts it in his shopping cart. When he clicks an item, that information is sent to the Web server.

Buy This

Buy Item/ Delete Item



Check Out

6 When someone decides to buy the items in the shopping cart, he sends in credit card information. When the items are bought, the server updates information about those items in the cookie. When the person visits his shopping cart, it is empty because the information about his purchases was deleted from the cookie, or else the cookie was designed so that it would expire after a certain amount of time.

How Electronic Wallets Work

1 Electronic wallets enable you to store information about your credit cards and similar information so that you don't need to fill out forms every time you want to buy something from a Web site. This illustration shows how one called eWallet works. *eWallet* is a piece of client software that runs on your computer. (Some electronic wallets, though not eWallet, run on Internet servers.) Typically, people either download eWallet from an Internet site, or it might be sent to them when they sign up for a credit card. The first step to using eWallet is to download the software and install it.

2 After installing the software, you enter a username and password that must be used to access the eWallet. That way, no one else can get at your credit card information. After entering a username and password, you enter information about the credit cards you'll be using to pay online.

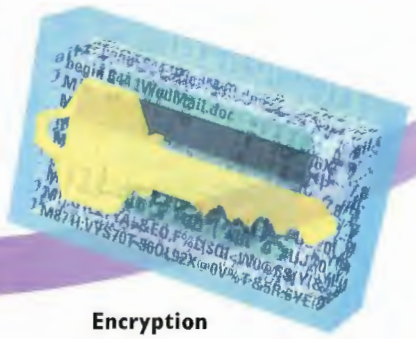
Web Server

Download Wallet

4 When customers are at a shopping site and want to use eWallet, they go to the page on the site where they must enter information about themselves and their credit card information. They open eWallet and drag the name of their credit card onto the Web page. eWallet recognizes where information needs to be filled in, such as name, address, and credit card information. It puts information into the form just as if the form were being filled in by hand. While the customer is filling out the information on the form, it is still local on the customer's computer; the information hasn't yet gone out over the Internet.

3 All the information in the eWallet is encrypted and stored locally on the computer. That way, no one can get at the credit card and personal information except someone who has the user name and password for the eWallet.

Personal Data and Credit Card Info



Encryption

Encryption

5 After the form is filled out, the customer clicks a button to place the order. As the credit card number and other information is sent over the Internet, it's encrypted by the shopping site so that no one can read it as it's sent. Only when the encrypted information is received by the shopping site can it be read.

How Online Auctions Work

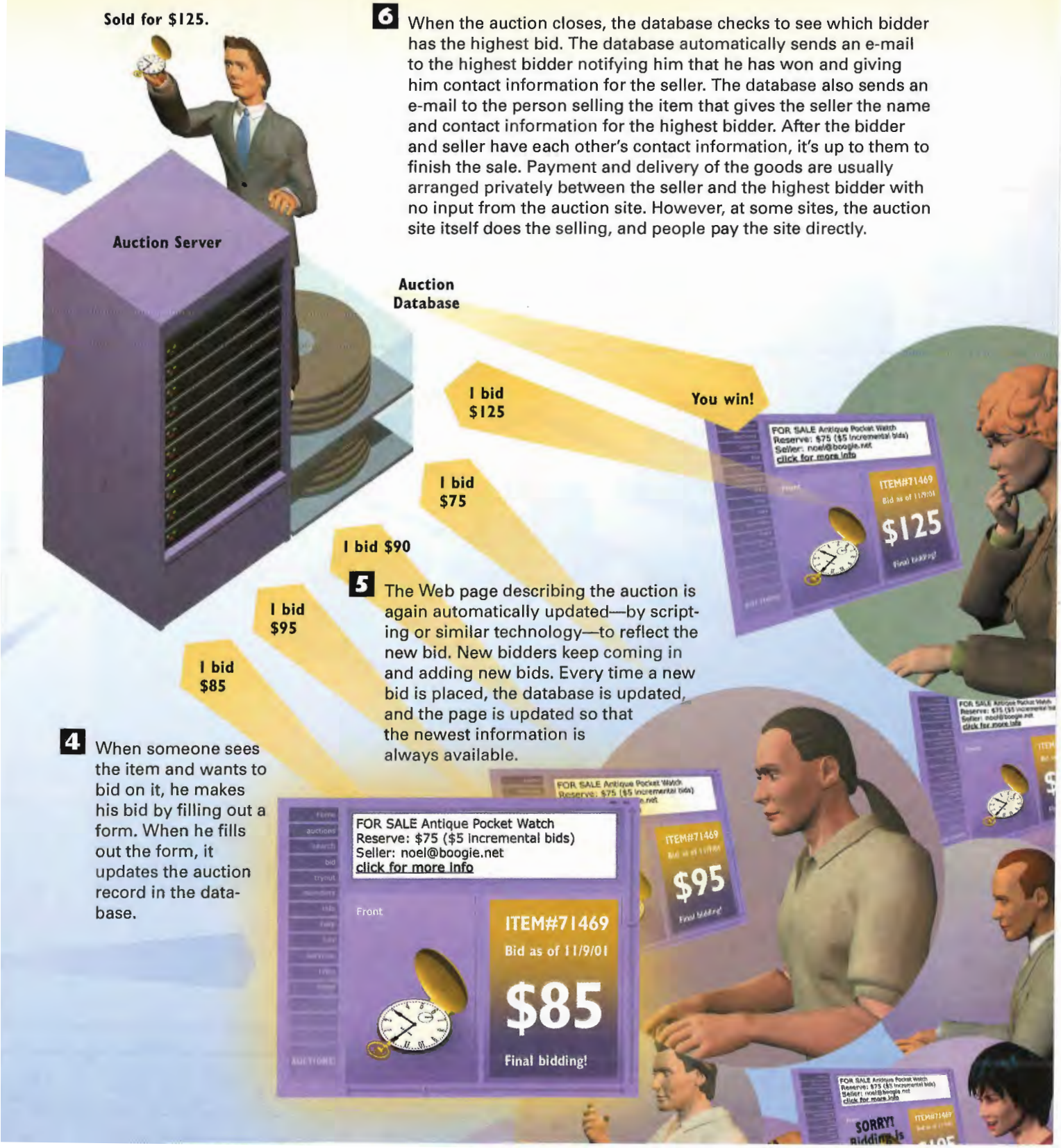
Register Me

1 Before someone can put something up for sale or bid on an item, he has to register by filling out a form on the auction site. When he fills out that form, the information from the form is sent to the auction site's database, which creates a unique record for that person.

Will sell antique watch.

2 When the person wants to put an item up for sale, the site checks whether he's registered. This can be done via a piece of information put on the person's computer called a *cookie*. It can also be done by asking the person to enter his username and password. When the site confirms the person is registered, the person puts an item up for sale by filling out a form detailing the item to be put on sale. When he does this, a new record is created in the auction site's database.

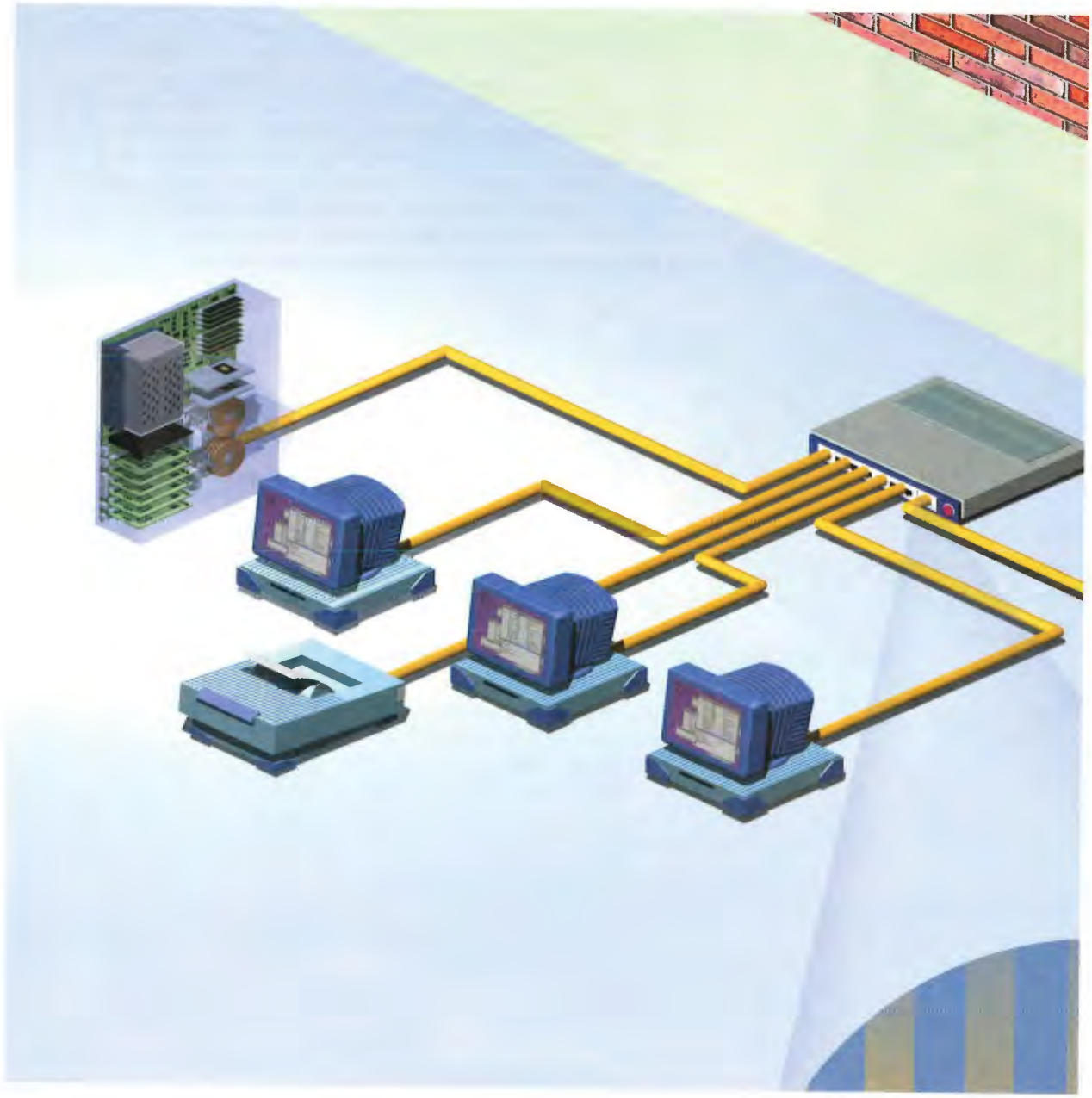
3 A program or script takes information from the database and automatically creates a Web page on the auction site, which describes the item up for sale and information about it. The Web page is now available to anyone on the auction site looking to bid on an auction.



6 When the auction closes, the database checks to see which bidder has the highest bid. The database automatically sends an e-mail to the highest bidder notifying him that he has won and giving him contact information for the seller. The database also sends an e-mail to the person selling the item that gives the seller the name and contact information for the highest bidder. After the bidder and seller have each other's contact information, it's up to them to finish the sale. Payment and delivery of the goods are usually arranged privately between the seller and the highest bidder with no input from the auction site. However, at some sites, the auction site itself does the selling, and people pay the site directly.

5 The Web page describing the auction is again automatically updated—by scripting or similar technology—to reflect the new bid. New bidders keep coming in and adding new bids. Every time a new bid is placed, the database is updated, and the page is updated so that the newest information is always available.

4 When someone sees the item and wants to bid on it, he makes his bid by filling out a form. When he fills out the form, it updates the auction record in the database.



PART 8

PROTECTING YOURSELF ON THE INTERNET

THE very nature of the Internet makes it vulnerable to attack. It was designed to allow for the freest possible exchange of information, data, and files—and it has succeeded admirably, far beyond its designers' wildest expectations. However, that freedom carries a price: Hackers and virus writers try to attack the Internet and computers connected to the Internet; those who want to invade others' privacy attempt to crack databases of sensitive information or snoop on information as it travels across Internet routes; and distasteful and pornographic sites have sprung up on the Web and on Usenet newsgroups.

This section of the book looks at a variety of security-related issues. You'll see how various tools have been developed to make transactions on the Net more secure and to help companies protect their sensitive data. You'll examine the thorny issue of pornography versus free speech and see how software can block children from visiting obscene sites or getting obscene materials. You'll also learn about some of the more controversial technologies on the Internet, such as cookies, which enable Web servers to track you as you move through their sites. This section also takes an inside look at an even more controversial technology, the FBI's Carnivore program, which enables the federal government to do things such as read the private mail of people it's investigating. You'll also look at how viruses work and how hackers attack Internet service providers (ISPs).

Chapter 44, "How Firewalls Work," looks at firewalls. Many companies whose networks are connected to the Internet have a great deal of sensitive information on their networks and want to ensure that their data and computers are safe from attack. The answer is to use firewalls—systems that allow people from inside a company to use the Internet but also stop people on the Internet from getting at the company's computers. This chapter also discusses personal firewalls—software people can use at home to ensure that hackers can't invade their own computers.

Chapter 45, "How Hackers Can Cripple the Internet and Attack Your PC," looks at attacks launched by hackers that can cripple Internet service providers and attack your computer as well. In a smurf attack, also called *smurfing*, a hacker targets an ISP and floods it with so much "garbage" traffic that none of the ISP's customers can use the service. Smurfing is one of the most common types of hacking attacks on the Internet. This chapter also examines the various ways hackers can attack your PC.

Chapter 46, "How Viruses Work," looks at viruses and how they are detected. Any program you download from the Internet has the potential for being infected with a virus, and it could, in turn, infect your computer. You'll see just how these nasty data-killers work and look at antivirus tools that can detect and kill them. This chapter also examines how a special type of virus called a Trojan Horse works. Trojan Horse viruses are becoming increasingly common on the Internet, so you'll look at one of the most recent Trojan Horse viruses—Melissa—and how it affected the Internet. As of this writing, Melissa is the most famous Internet Trojan Horse of all time.

Chapter 47, "How Internet Sites Can Invade Your Privacy," explores controversial technologies that enable Web sites to track what you do when you're online. It covers cookies, Web tracking, and Web Bugs as well as a technology that can help preserve people's privacy—Internet passports. Some people worry that cookies and Web tracking can invade their privacy. Others disagree, saying that cookies and Web tracking can help customize the

Web to users' interests. Cookies are bits of data put on a hard disk when someone visits certain Web sites. That data can be used for many purposes. One common use is to make it easier for people to use Web sites that require a username and password by storing that information and then automatically sending the information whenever it's requested. Passports enable people to decide what type of information about them can be tracked by Web sites. Web tracking enables those who run Web sites to see how people use their sites. Web bugs are another technique for tracking people's Internet use.

Chapter 48, "Cryptography, Privacy, and Digital Certificates," examines cryptosystems and digital certificates. An enormous amount of information is sent across the Internet every day—everything from personal e-mail to corporate data to credit card information and other highly sensitive material. All that information is vulnerable to hackers and snoopers. Because the information is sent in packets along public routers, the possibility exists that someone could intercept and decipher it. As a way to ensure that the sensitive material can't be looked at, sophisticated cryptosystems have been developed so that only the sender and receiver know what's in the packets.

The chapter also looks at digital certificates. On the Internet, no face-to-face communication takes place, so knowing whether people really are who they say they are can be difficult. Digital certificates are used to absolutely identify someone. If someone sends you an e-mail, for example, a digital certificate will let you know that the person is who he says he is.

Chapter 49, "How the FBI's 'Carnivore' Program Works," details an extremely controversial program that enables the federal government to read people's e-mail and follow their Internet activity without people knowing about it.

Finally, Chapter 50, "Parental Controls on the Internet," takes a detailed look at the issues of pornography and free speech on the Internet. Explicit sexual material is posted on the Internet, and some people would like to fine and jail people and organizations that allow such material to be posted. Passing those types of laws raises a host of constitutional issues about free speech. As a way to solve the problem, companies create and sell software for parents that enables them to block their children from seeing obscene and violent material on the Internet. In this chapter, you'll see how one of the most popular pieces of parental control software works.

CHAPTER

44 How Firewalls Work

EVERY time a computer is connected to the Internet, it faces potential danger. Corporate Local Area Networks (LANs) connected to the Internet, as well as PCs at home—especially those connected to a high-speed cable modem or DSL modem—are likely targets for hackers.

Due to the Internet's openness, every corporate network connected to it is vulnerable to attack. Crackers on the Internet could theoretically break into the corporate network and do harm in a number of ways: They could steal or damage important data; damage individual computers or the entire network; use the corporate computer's resources; or use the corporate network and resources as a way of posing as a corporate employee. The solution isn't to cut off the network from the Internet. Instead, the company can build firewalls to protect its network. These firewalls enable anyone on the corporate network to access the Internet, but they stop crackers, hackers, and others on the Internet from gaining access to the corporate network and causing damage.

Corporate firewalls are hardware and software combinations that are built using routers, servers, and a variety of software. They sit at the most vulnerable point between a corporate network and the Internet and can be as simple or complex as system administrators want to build them.

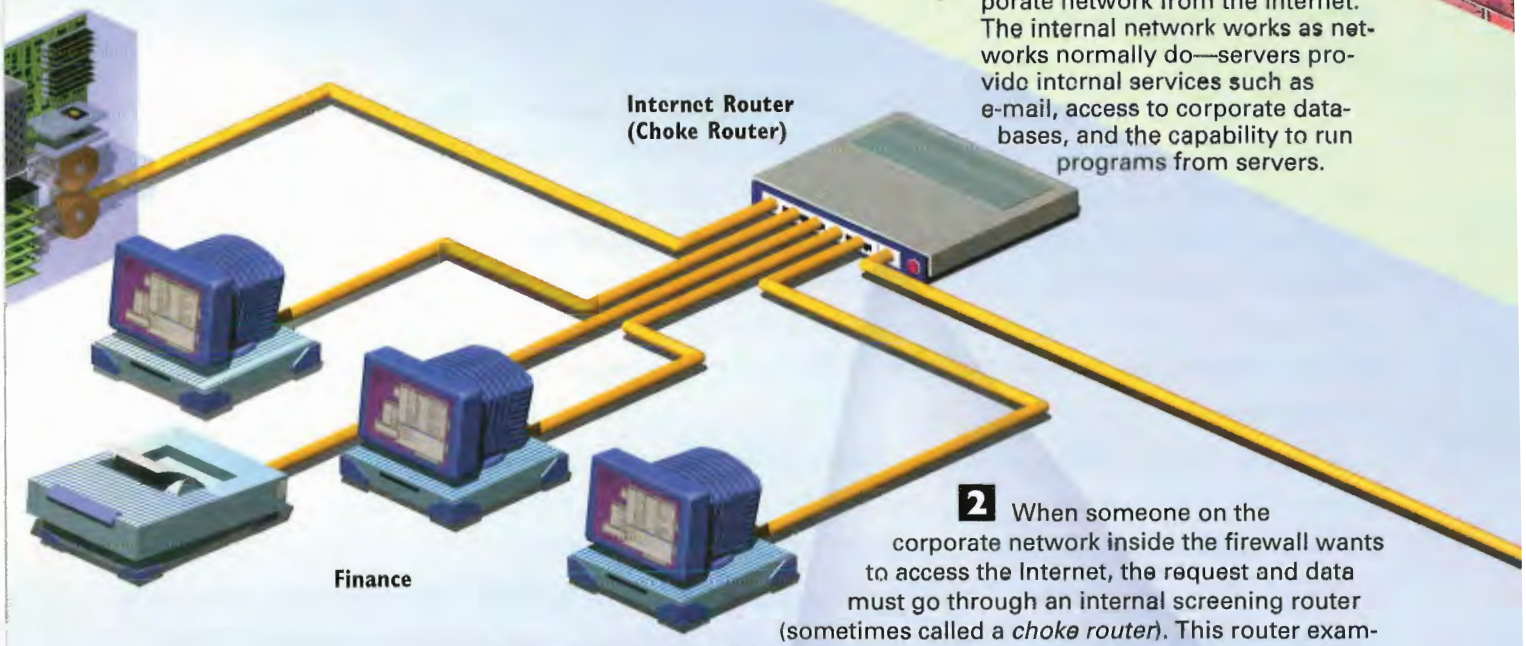
One of the simplest kinds of firewalls utilizes packet filtering. In *packet filtering*, a screening router examines the header of every packet of data traveling between the Internet and the corporate network. Packet headers have information in them such as the IP address of the sender and receiver, the protocol being used to send the packet, and other similar information. Based on that information, the router knows what kind of Internet service—such as FTP or rlogin—is being used to send the data, as well as the identity of the sender and receiver of the data. (The command, rlogin, is similar to Telnet, which enables someone to log into a computer. It can be dangerous because it enables users to bypass having to type in a password.) After this information is determined, the router can bar certain packets from being sent between the Internet and the corporate network. For example, the router could block any traffic except for e-mail. Additionally, it could block traffic to and from suspicious destinations or from certain users.

Proxy servers are also commonly used in firewalls. A *proxy server* is server software that runs on a host in a firewall, such as a bastion host. Because only the single proxy server (instead of the many individual computers on the network) interacts with the Internet, security can be maintained. That single server can be kept more secure than can hundreds of individual computers on a network.

Home PCs connected to the Internet via high-speed cable modems or DSL modems are targets as well because if hackers can break into them, they can use them as launching pads for their attacks, while covering their tracks. *Personal firewalls* have become popular—software and hardware that sits on a home computer and protects the home computer in similar ways to how corporate firewalls protect corporate LANs.

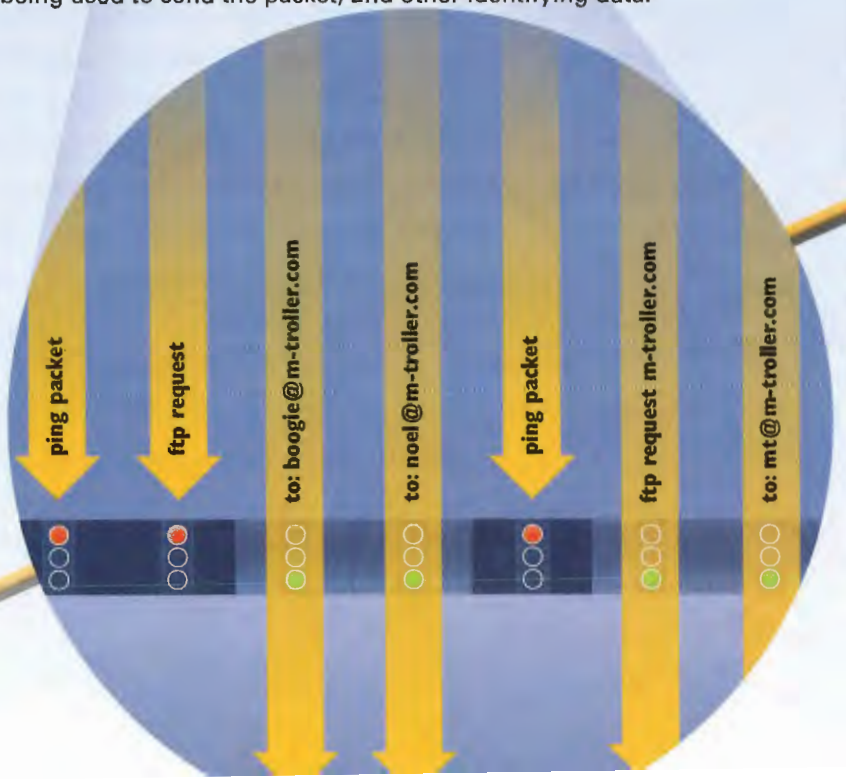
How Firewalls Work

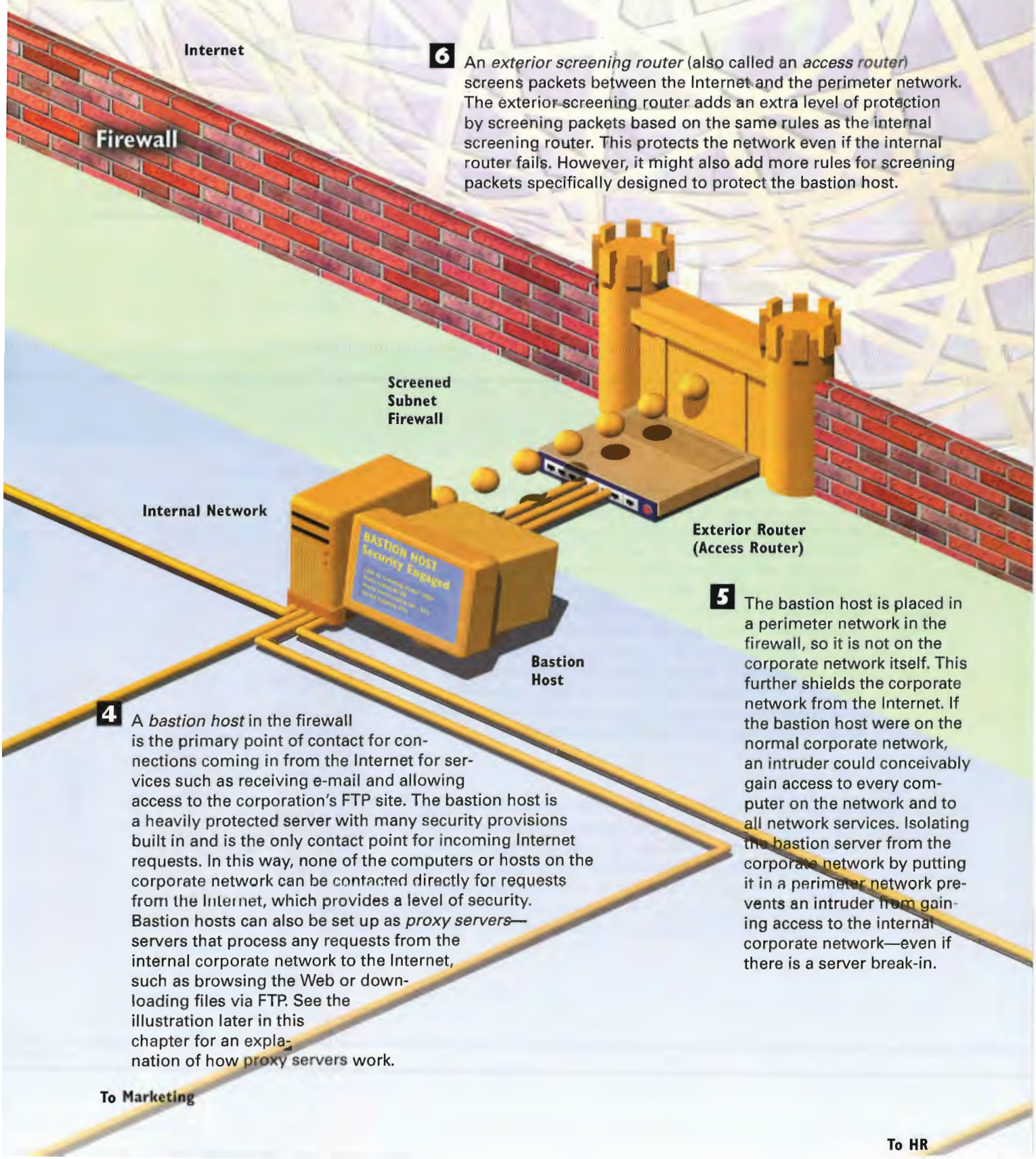
1 The *firewall* shields the internal corporate network from the Internet. The internal network works as networks normally do—servers provide internal services such as e-mail, access to corporate databases, and the capability to run programs from servers.



2 When someone on the corporate network inside the firewall wants to access the Internet, the request and data must go through an internal screening router (sometimes called a *choke router*). This router examines all the packets of data traveling in both directions between the corporate network and the Internet. Information in the packets' headers gives the router important information, such as the source and destination of the packet, the kind of protocol being used to send the packet, and other identifying data.

3 Based on the information in the headers, the screening router allows certain packets to be sent or received, but blocks other packets. For example, it might not allow some services, such as rlogin, to be run. The router also might not allow packets to be sent to and from specific Internet locations because those locations have been found to be suspicious. Conceivably, a router could be set up to block every packet traveling between the Internet and the internal network except for e-mail. System administrators set the rules for determining which packets to allow in and which ones to block.





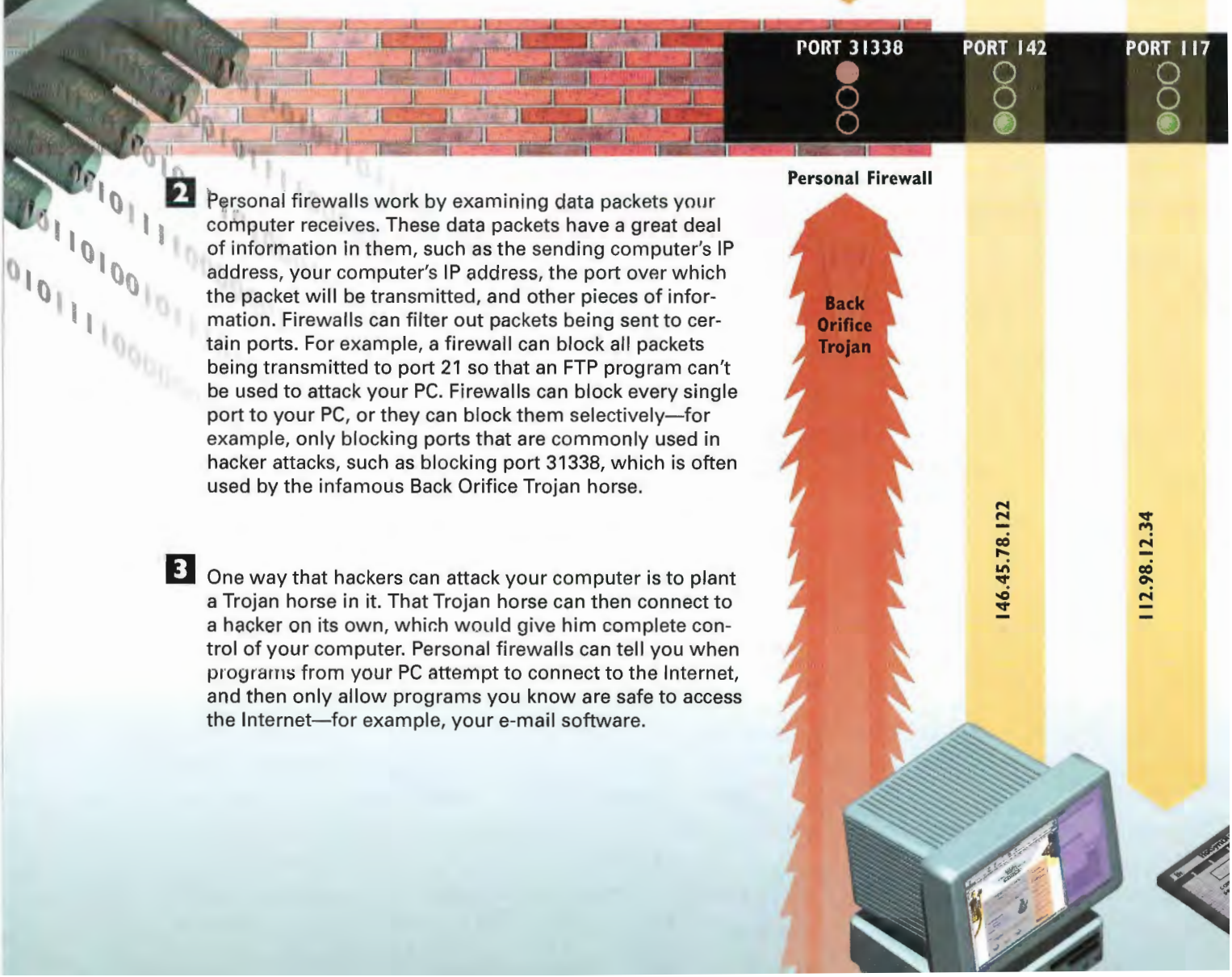
6 An *exterior screening router* (also called an *access router*) screens packets between the Internet and the perimeter network. The exterior screening router adds an extra level of protection by screening packets based on the same rules as the internal screening router. This protects the network even if the internal router fails. However, it might also add more rules for screening packets specifically designed to protect the bastion host.

5 The bastion host is placed in a perimeter network in the firewall, so it is not on the corporate network itself. This further shields the corporate network from the Internet. If the bastion host were on the normal corporate network, an intruder could conceivably gain access to every computer on the network and to all network services. Isolating the bastion server from the corporate network by putting it in a perimeter network prevents an intruder from gaining access to the internal corporate network—even if there is a server break-in.

4 A *bastion host* in the firewall is the primary point of contact for connections coming in from the Internet for services such as receiving e-mail and allowing access to the corporation's FTP site. The bastion host is a heavily protected server with many security provisions built in and is the only contact point for incoming Internet requests. In this way, none of the computers or hosts on the corporate network can be contacted directly for requests from the Internet, which provides a level of security. Bastion hosts can also be set up as *proxy servers*—servers that process any requests from the internal corporate network to the Internet, such as browsing the Web or downloading files via FTP. See the illustration later in this chapter for an explanation of how proxy servers work.

How Personal Firewalls Work

1 People who use high-speed connections such as cable modems at home might be prone to hackers' attacks because computers connected to the Internet in this way are more vulnerable and more enticing to the hackers. To protect home computers, many people have turned to *personal firewalls*—software that runs on the computer and protects the computer against Internet attacks. To understand how personal firewalls work, you first need to understand the concept of Internet ports. An *Internet port* isn't a physical device—rather it's a virtual entranceway between your computer and the Internet. When you make an Internet connection, many of these virtual connections are opened up, and each has its own number and purpose. For example, e-mail software usually uses port 110 on a mail server to get mail and uses port 25 on a mail server to send mail. FTP software usually connects to FTP servers using port 21.



2 Personal firewalls work by examining data packets your computer receives. These data packets have a great deal of information in them, such as the sending computer's IP address, your computer's IP address, the port over which the packet will be transmitted, and other pieces of information. Firewalls can filter out packets being sent to certain ports. For example, a firewall can block all packets being transmitted to port 21 so that an FTP program can't be used to attack your PC. Firewalls can block every single port to your PC, or they can block them selectively—for example, only blocking ports that are commonly used in hacker attacks, such as blocking port 31338, which is often used by the infamous Back Orifice Trojan horse.

3 One way that hackers can attack your computer is to plant a Trojan horse in it. That Trojan horse can then connect to a hacker on its own, which would give him complete control of your computer. Personal firewalls can tell you when programs from your PC attempt to connect to the Internet, and then only allow programs you know are safe to access the Internet—for example, your e-mail software.

4 Firewalls can also block specific IP addresses from contacting your computer. For example, if you know the IP of a hacker who has attacked you before, you can have your firewall block it from getting through to your computer.

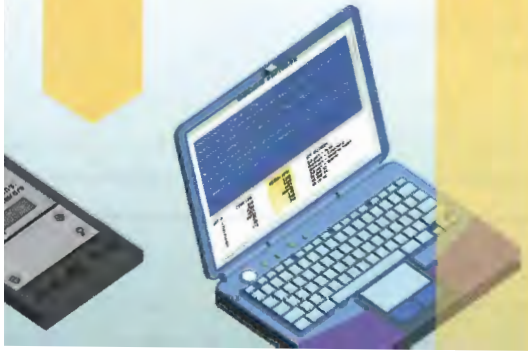
NAT (Network Address Translation)

6 Many personal firewalls keep a running log of every attempt made to attack or probe your PC. These logs can be sent to your ISP, which can use them to try to track down the hackers and shut them down.

102.147.12.32

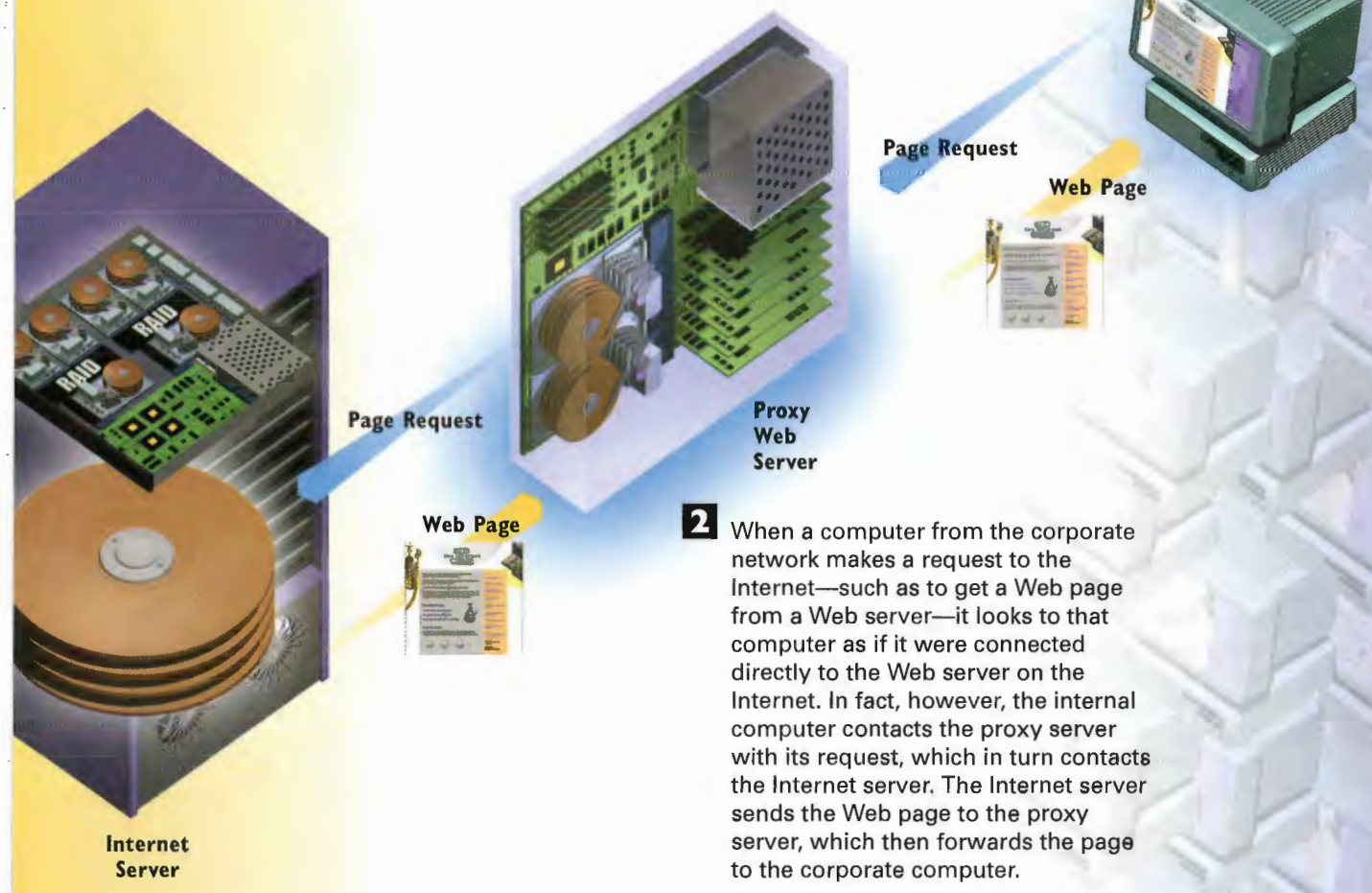
131.244.34.12

5 Many home network routers include a hardware-based personal firewall that protects you from the Internet using a technique called *Network Address Translation* (NAT). With NAT, your true IP address is shielded from the Internet—it can't be seen by anyone or any application outside your home network. In essence, it's invisible and can't be reached by hackers.



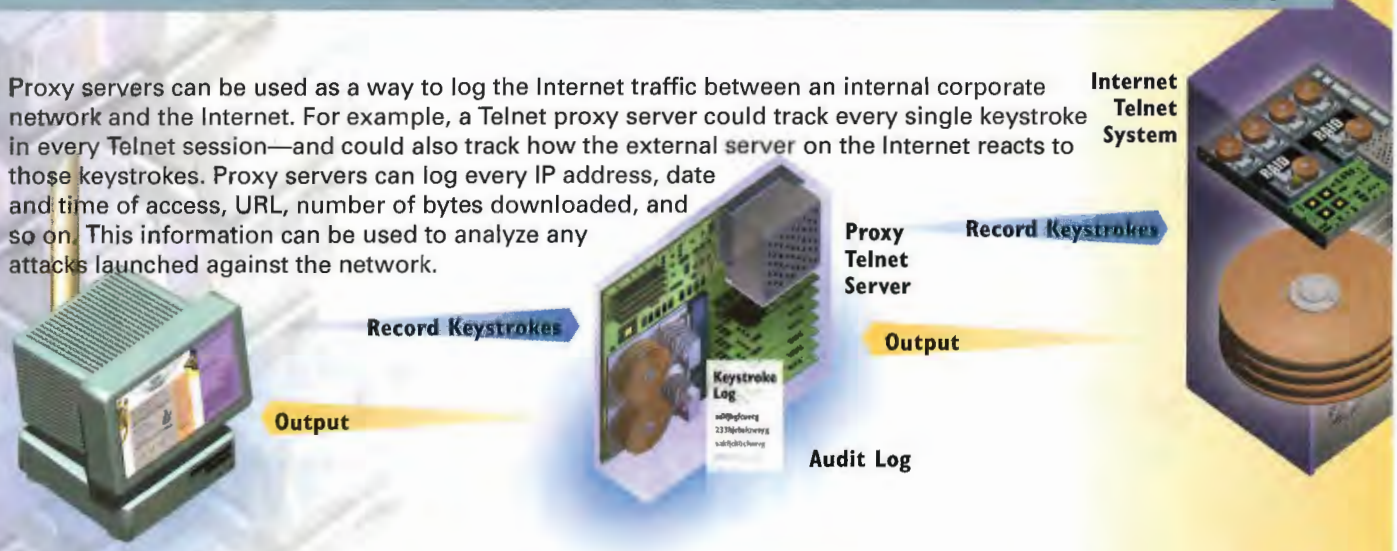
How Proxy Servers Work

1 System administrators can set up proxy servers to be used for many services, such as FTP, the Web, and Telnet. System administrators decide which Internet services must go through a proxy server. Specific proxy server software is required for each kind of Internet service.

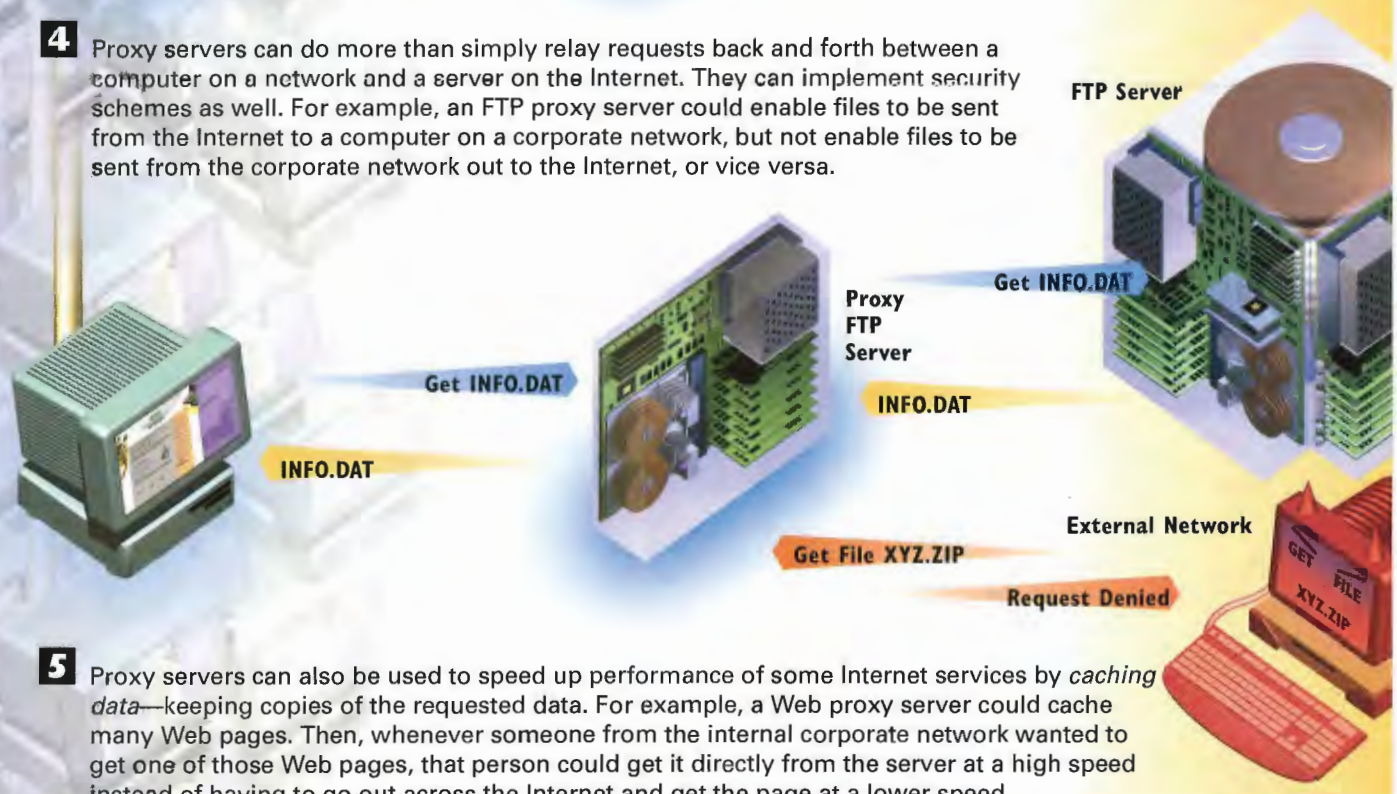


2 When a computer from the corporate network makes a request to the Internet—such as to get a Web page from a Web server—it looks to that computer as if it were connected directly to the Web server on the Internet. In fact, however, the internal computer contacts the proxy server with its request, which in turn contacts the Internet server. The Internet server sends the Web page to the proxy server, which then forwards the page to the corporate computer.

3 Proxy servers can be used as a way to log the Internet traffic between an internal corporate network and the Internet. For example, a Telnet proxy server could track every single keystroke in every Telnet session—and could also track how the external server on the Internet reacts to those keystrokes. Proxy servers can log every IP address, date and time of access, URL, number of bytes downloaded, and so on. This information can be used to analyze any attacks launched against the network.



4 Proxy servers can do more than simply relay requests back and forth between a computer on a network and a server on the Internet. They can implement security schemes as well. For example, an FTP proxy server could enable files to be sent from the Internet to a computer on a corporate network, but not enable files to be sent from the corporate network out to the Internet, or vice versa.

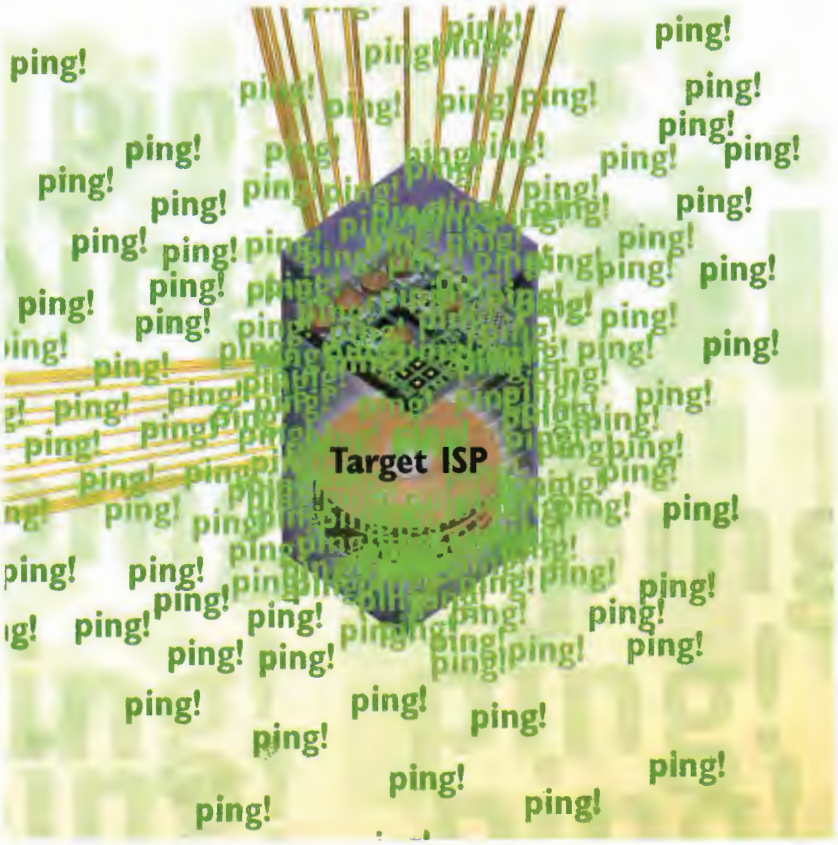


5 Proxy servers can also be used to speed up performance of some Internet services by *caching data*—keeping copies of the requested data. For example, a Web proxy server could cache many Web pages. Then, whenever someone from the internal corporate network wanted to get one of those Web pages, that person could get it directly from the server at a high speed instead of having to go out across the Internet and get the page at a lower speed.

CHAPTER

45

How Hackers Can Cripple the Internet and Attack Your PC



HACKERS are equal opportunity attackers—they attack individuals as well as Web sites by using a variety of software and malicious programs. Among hackers' many targets are *Internet service providers* (ISPs)—companies that sell access to the Internet. A hacker might target an ISP for several reasons: He might be angry at the ISP or at someone using the ISP, or he might attack the ISP for the mere thrill of it.

One of the most common attacks against an ISP is called a *smurf attack*, or *smurfing*. (Smurfing is a kind of Denial of Service, or DOS attack. There are several ways that a hacker can launch a DOS attack; smurfing is one of the most popular ones.) In a smurf attack, a hacker floods the ISP with so many garbage packets that all the ISP's available bandwidth is used up. The ISP's customers can't send or receive data and can't use e-mail, browse the Web, or use any other Internet service.

In a smurf attack, hackers exploit a commonly used Internet service—ping (Packet Internet Groper). People normally use *Ping* to see whether a particular computer or server is currently attached to the Internet and working. When a computer or server is sent a ping packet, it sends a return packet to the person who sent the ping, which in essence says, "Yes, I'm alive and attached to the Internet." In a smurf attack, hackers forge the return addresses on ping requests so that, instead of going back to them, the return packets go to the hackers' target ISP. The hackers are able to use networks attached to the Internet as a way of relaying their ping requests and magnifying each ping request many times. In this way, a hacker can use networks attached to the Internet to flood the ISP with so many return ping packets that the ISP's customers can't use the ISP's services. Hackers can use multiple networks attached to the Internet in a single smurf attack.

ISPs have difficulty fighting smurf attacks because the ping answering packets come from legitimate networks and not from the hacker. The ISP has to track down where the ping answering packets are coming from and then contact each of those networks to ask it to turn off the ping answering packets. Making this more difficult is that when an ISP goes down, often its customers will send ping requests to it to see whether it is alive and connected to the Internet. The ISP has a difficult time separating the legitimate ping packets from the smurf attack packets.

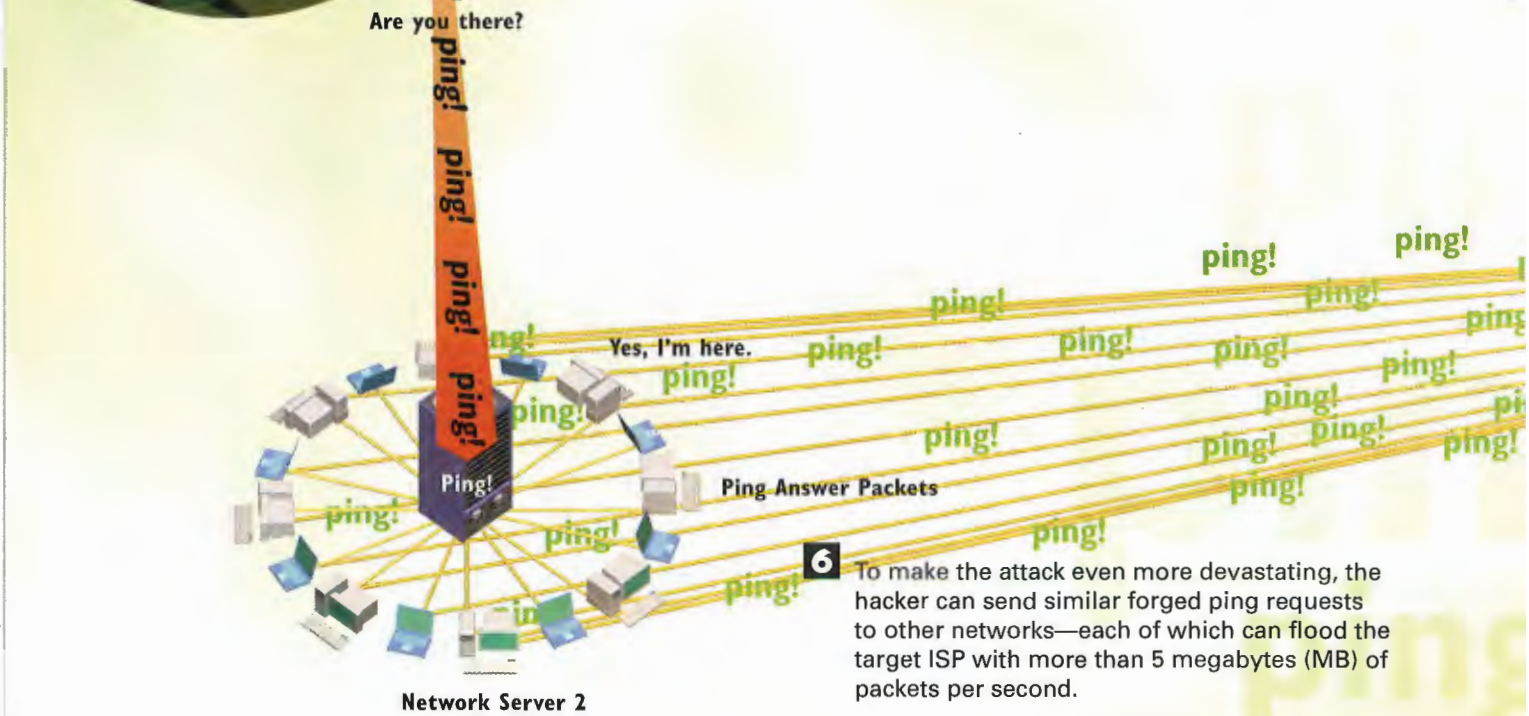
Smurf protection features and software have become available for ISPs and to put on Internet routers. But few companies are using those features and software because they have yet to gain widespread acceptance and not everyone recognizes how big a problem smurf attacks have become.

Hackers don't just target ISPs, of course. They attack individuals as well. As you'll see in the illustration later in this chapter, hackers can take over people's computers to delete and steal files, steal personal information and passwords, and even use the person's computer as a launch pad for attacks on ISPs and Web sites.

How Smurf Attacks and DOS Attacks Work

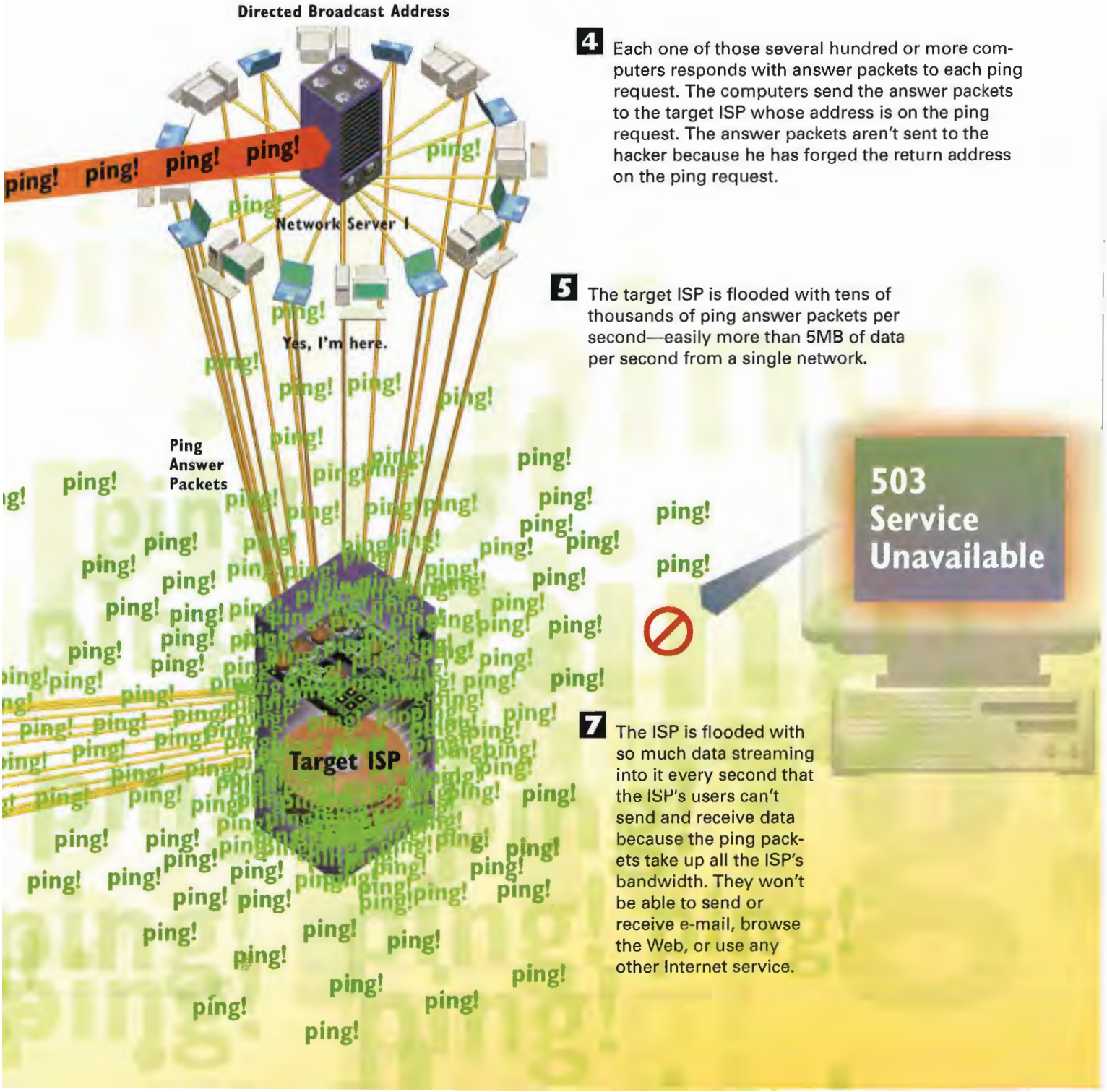
1 In a *smurf attack*, or "smurfing," and a *denial of service*, or DOS attack, a hacker targets an Internet service provider (ISP) and floods it with so much garbage traffic that none of the ISP's customers are able to use the service. Smurf attacks have become one of the most popular kinds of hacker attacks on the Internet. The attack starts when a hacker sends a series of *ping* (Packet Internet Groper) packets to a network attached to the Internet. Ping uses the *Internet Control Message Protocol*—a widely used protocol for, among other things, determining whether a particular computer is attached to the Internet and working properly. The network being pinged is not the target of the attack. Instead, it will be used as a way to attack the ISP.

2 The hacker forges the return address on the ping packets. Instead of having his address, it has the address of the ISP that the hacker will be attacking. This serves two purposes: It attacks the ISP, and it also shields the hacker from being caught because his address is not on the ping requests.



6 To make the attack even more devastating, the hacker can send similar forged ping requests to other networks—each of which can flood the target ISP with more than 5 megabytes (MB) of packets per second.

3 The ping requests are sent in a constant stream to the network's *directed broadcast* address. This address, in turn, sends the ping requests to every computer attached to the network—which can be several hundred or more computers.



4 Each one of those several hundred or more computers responds with answer packets to each ping request. The computers send the answer packets to the target ISP whose address is on the ping request. The answer packets aren't sent to the hacker because he has forged the return address on the ping request.

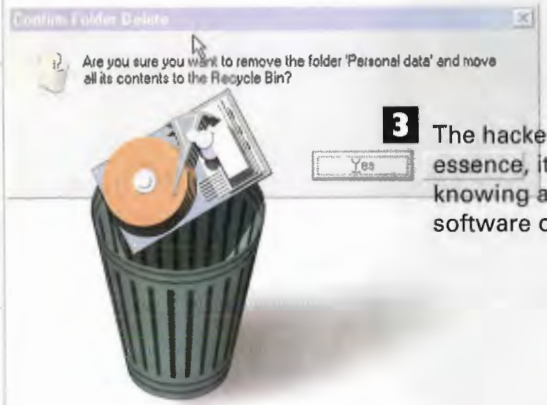
5 The target ISP is flooded with tens of thousands of ping answer packets per second—easily more than 5MB of data per second from a single network.

7 The ISP is flooded with so much data streaming into it every second that the ISP's users can't send and receive data because the ping packets take up all the ISP's bandwidth. They won't be able to send or receive e-mail, browse the Web, or use any other Internet service.

How Hackers Can Attack Your Computer

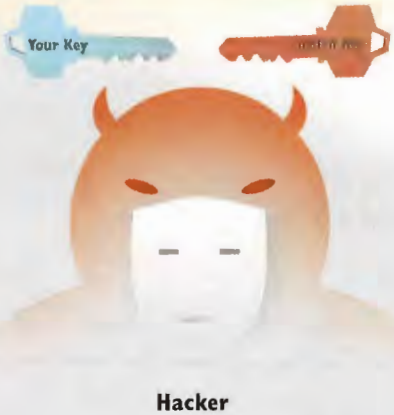
1 Hackers not only attack big Web sites and corporations, but also individual computers in homes or businesses. Hackers can do damage and use your computer in many ways. As a start to many of hacker's nefarious deeds, they need access to your computer. One common way they gain it is through the use of a program called Back Orifice. Before the hackers can use the program, you have to get it on your computer. You can unwittingly get a copy of Back Orifice on your computer in many different ways—for example, you can open a file in an e-mail message and it can be installed to your computer without you realizing it, or you can be sent it when you use Internet's IRC chat protocol.

2 Hackers have automated tools that scan thousands of different computers to see which ones have Back Orifice running on them. These tools send out *port probes*—packets that look at a specific virtual ports that all computers have when connected to the Internet. Back Orifice uses port 31337, and if it's running on a computer, it will open that port. A port probe will alert the hacker that port 31337 is open so that he knows he can take control of your PC.



3 The hacker can do many things when he takes control of your computer—in essence, it's as if he's sitting at your keyboard and monitor without you knowing about it. He can, for example, copy or delete all the files, data, and software on your computer.

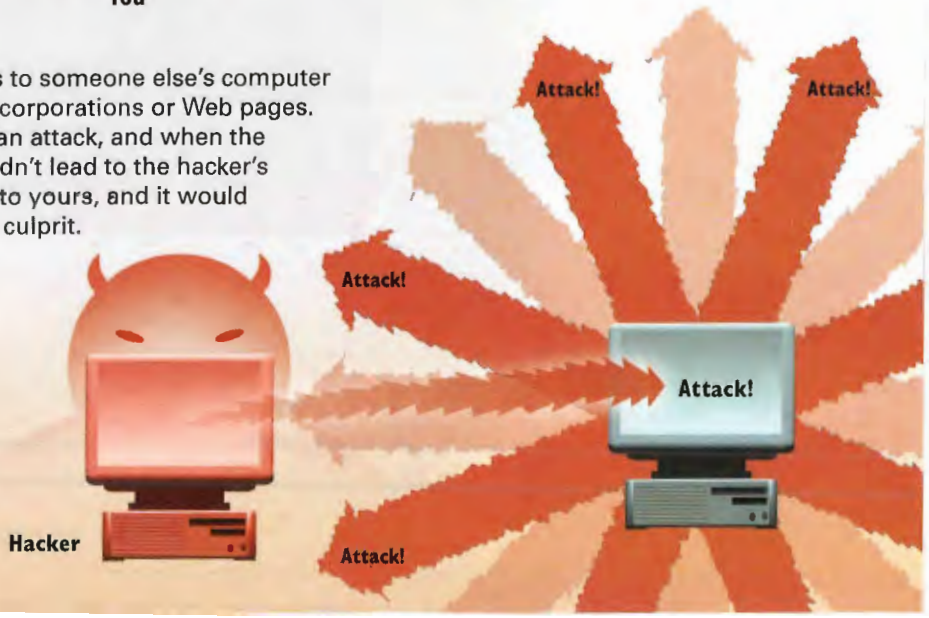
4 He can find out personal information about you by looking through your files. For example, he might be able to gain access to your credit card number, bank account, and social security number, and then use that information illegally.



5 He can gain access to all your passwords, which would enable him to pose as you on Web sites and break into data on your computer that you've tried to protect with passwords.

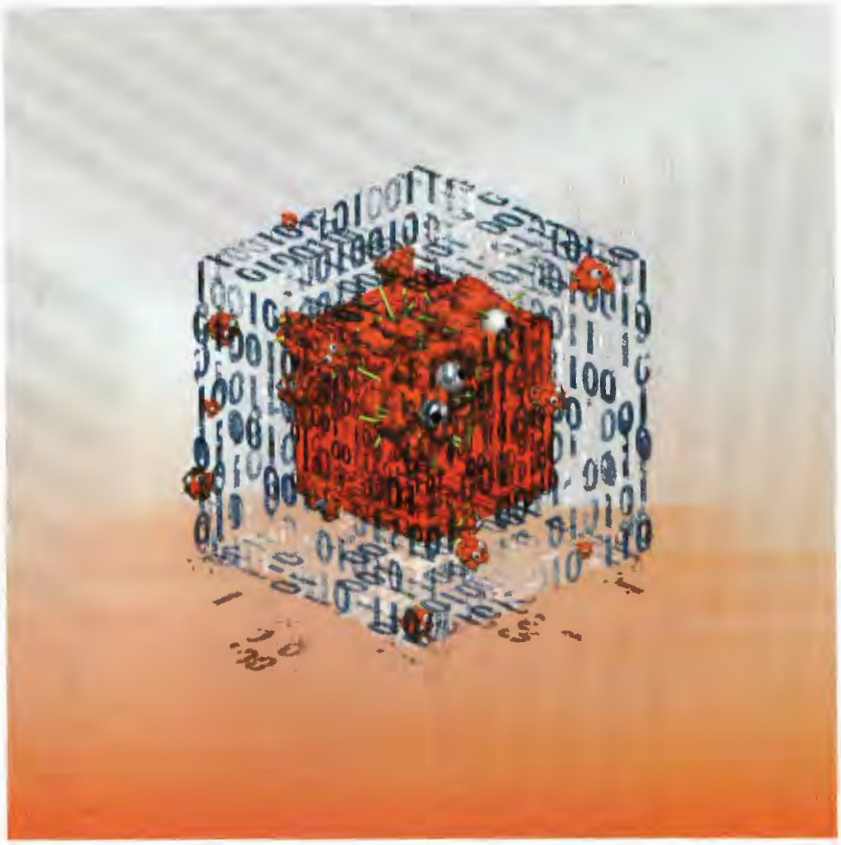
6 He can upload any files to, or download any files from, your computer. For example, he could use your computer to store copies of illegally copied software and could even enable other hackers to then download those illegal copies.

7 Often, hackers use access to someone else's computer to launch attacks against corporations or Web pages. The hacker could launch an attack, and when the attack was traced, it wouldn't lead to the hacker's computer—it would lead to yours, and it would appear that you were the culprit.



CHAPTER

46 How Viruses Work



THE Internet, just like the rest of the world, is not a perfectly safe place to visit. If you download files from the Internet, there is a chance—a very small chance, but nonetheless a chance—that your computer could become infected with a virus.

Viruses are malicious programs that invade your computer. They can cause many different kinds of damage, such as deleting data files, erasing programs, or destroying everything they find on your hard disk. Not every virus causes damage; some simply flash annoying messages.

Although you can get a virus from the Internet by downloading files to your computer, the Internet is not the only place where viruses can be picked up. If you've sent files via e-mail or on your company's internal network, you can get viruses that way as well. Instances have occurred when commercially bought, shrink-wrapped software has contained viruses.

The term virus is a somewhat generic term applied to a wide variety of programs. Viruses are written for specific kinds of computers, such as PCs or Macintoshes, because the files they infect run only on one kind of computer.

Traditional viruses attach themselves to programs or data files, infect your computer, replicate themselves on your hard disk, and then damage your data, hard disk, or files. Viruses usually attack four parts of your computer: its executable program files; its file-directory system that tracks the location of all your computer's files (and without which, your computer won't work); its boot and system areas that are needed to start your computer; and its data files. At one time it was believed that data files could not be infected by viruses, but recently, viruses have been written that infect data files too. For example, some viruses attach themselves to Microsoft Word macros inside a Word data file and are launched whenever a particular macro is run.

Trojan horses are files or programs that disguise themselves as normal, helpful programs or files, but in fact are viruses. For example, if a program purported to be a financial calculator, but really deleted every file on your hard disk, that program would be called a Trojan horse. The most famous Trojan horse of all, Melissa, was disguised as a Word document sent via e-mail—and it wreaked enough havoc that it crashed many Internet and corporate mail servers by making use of the Outlook and Outlook Express e-mail programs.

Worms are programs designed to infect networks such as the Internet. They travel from networked computer to networked computer and replicate themselves along the way. The most infamous worm of all was released on November 2, 1988. The worm copied itself to many Internet host computers and eventually bringing the Internet to its knees.

The best way to protect your computer against viruses is to use antiviral software. Several kinds of antiviral software exist. A scanner checks to see if your computer has any files that have been infected; whereas an eradication program will wipe the virus from your hard disk. Sometimes eradication programs can kill the virus without having to delete the infected program or data file, while other times those infected files must be deleted. Still other programs, sometimes called *inoculators*, do not allow a program to be run if it contains a virus and stop your computer from being infected. Malicious e-mail programs can sometimes be stopped by disabling a built-in capability to run scripts in e-mail software.

How Viruses Infect Computers

1 A virus hides inside a legitimate program where it remains dormant until you run the infected program. The virus springs into action when you actually run the infected program. Sometimes the first thing the virus will do is infect other programs on your hard disk by copying itself into them.

Infected Application

Newly Infected File

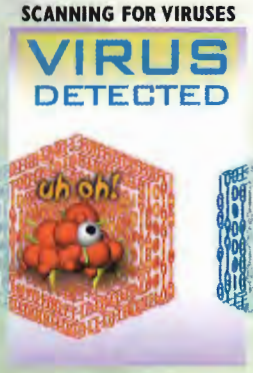
2 Some viruses place messages called *v-markers* or *virus markers* inside programs that they infect, and these messages help manage the viruses' activities. Each virus has a specific v-marker associated with it. If a virus encounters one of these markers in another program, it knows that the program is already infected so it doesn't replicate itself there. When a virus cannot find more unmarked files on a computer, that signals to the virus that there are no more files to be infected. At this point, the virus might begin to damage the computer and its data.

Virus Marker

Damaged File or Application

3 Viruses can corrupt program or data files so that they work oddly, not at all, or cause damage when they do run. They can destroy all the files on your computer, change the system files that your computer needs when it is turned on, and cause other types of damage.

4 Software programs called *scanners* check for viruses and alert you to the viruses' presence. They work in many different ways. One method of detection is to check your program files for telltale virus markers that indicate the presence of a virus. Other methods include checking to see whether a program's file size has changed. Some types of antiviral programs run continuously on your computer and check any program for the presence of a virus before the program is run or downloaded.



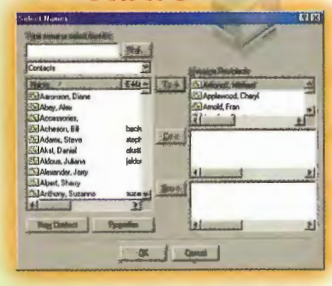
5 *Eradication programs* disinfect, or remove, viruses from software. Sometimes they can eradicate the virus without damaging the program that the virus has infected. In other instances, they have to destroy the program as well as the virus.

How Trojan Horses Work

1 Trojan horses are programs that disguise themselves as normal, helpful programs or files, but in fact are viruses. The most well known Trojan horse of all time, called Melissa, used e-mail to spread itself, and damaged many Internet and corporate mail servers. Here's how it did its work.

2 Melissa arrived in people's e-mail inboxes—disguised as a normal e-mail message—with a Microsoft Word file as an attachment. The subject line of the e-mail read, "important message from," followed by a person's name—and that name might have been the name of a friend, acquaintance, or co-worker of the person receiving the message. The body of the e-mail read, "Here is that document you asked for...don't show anyone else."

3 When people opened the attached Word file, Melissa sprang to work. If the file wasn't opened, then Melissa could do no damage. The attached file appeared to be a normal Word file that contained a list of pornographic sites. However, when the file was opened, a *macro* ran without the user knowing it. A macro is a set of automated commands—much like a program.



4 The macro checked to see whether the person had Outlook on his computer. Outlook is a Microsoft e-mail program. If Outlook wasn't present, the macro wasn't able to do any damage. If Outlook was present, Melissa looked at the first 50 names in Outlook's address book, then made a copy of itself and mailed itself to all those 50 names without the person knowing this was happening. The e-mail sent to each of those people looked exactly like the e-mail the person received: The subject line of the email read, "important message from," followed by the name of the person who had been infected by Melissa. It appeared that the infected person was sending out a personal message.

125,000 Possible Infections

5 Each of these 50 people, in turn, received the infected e-mail and attached Word document. When he opened the attached file, Melissa did the same thing to him—automatically sent itself to 50 more people.

6 The volume of e-mail being sent quickly became so great that Internet and corporate e-mail servers were unable to keep up with the demand for sending and receiving messages—and many of them crashed. Many Internet and corporate mail servers were overwhelmed by the huge demand for sending and receiving e-mail, and so normal mail—not just Melissa-related mail—couldn't be sent or received. The problem was finally resolved when anti-virus software was updated to include features that could detect and kill Melissa.

CHAPTER

47

How Internet Sites Can Invade Your Privacy

PRIVACY issues are a big concern on the Net. Much information can be gathered about people when they use the Net, and it's not always clear who will use that information or how it will be used. In particular, three technologies that concern people are cookies, Web tracking, and Web bugs. Cookies and Web tracking both serve useful purposes, but many people worry that there is a "Big Brother" aspect to them. Web bugs, on the other hand, can be used for nefarious purposes. One technology, Internet passports, might enable people to ensure that their privacy isn't invaded while still enabling Web sites to gather information that can be used to deliver specialized services to Web surfers.

Cookies are bits of data put on a hard disk when someone visits certain Web sites. The most common use of this data is to make it easier for people to use Web sites that require a username and password. The cookie on the hard disk has the username and password in it, so people don't have to log in to every page that requires that information. Instead, the cookie sends the information to the server, and the person can visit the page freely.

Cookies can contain virtually any kind of information, such as the last time a person visited the site, the person's favorite sites, and similar, customizable information. They can be used to track people as they go through a Web site and to help gain statistics about what types of pages people like to visit. Although some people view them as invaders of privacy, they can also make the Web a much better place to visit by doing things such as making it easier to conduct electronic commerce.

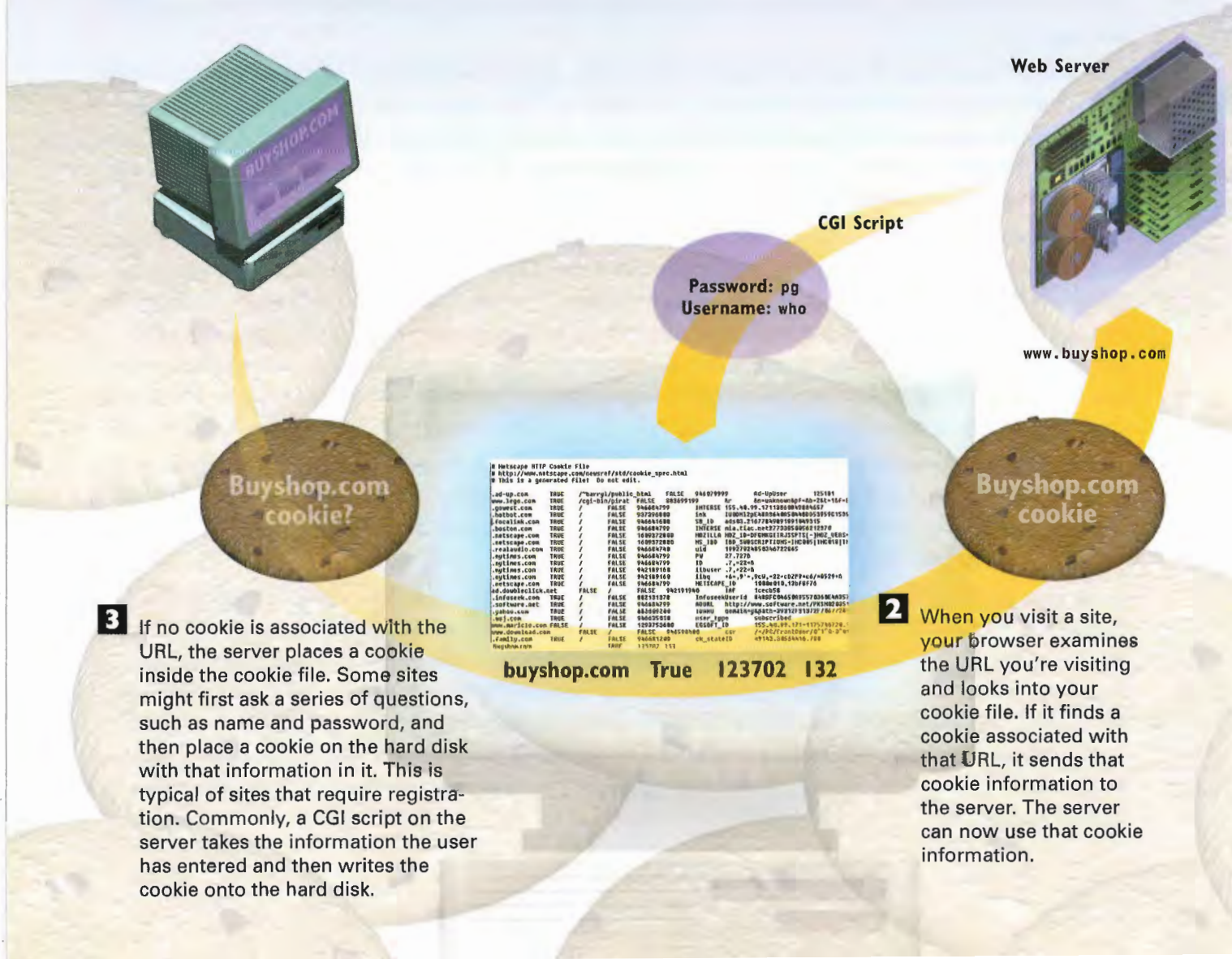
Although cookies can be used to track how people use a Web site, many other methods can be used, as well. In one method, Web server logs are examined in detail. This would make it possible, for example, to identify the most popular pages on the site, the sites people have just visited, how many pages people read in a typical visit, and similar information. Other methods include using software *sniffers* that examine every packet coming into or going out of a Web site. Webmasters can use this tracking information to help create better sites—but they can also use it to assemble demographic information to sell to advertisers. The second illustration in this chapter shows the functionality of Web tracking software from a company called Accrue.

Web bugs can also trace people's paths through a Web site. Web bugs get their name not in reference to an error in a program, but instead from the term *to bug* as in "to wiretap." More dangerously, Web bugs can be included in e-mail, and they can actually enable people to view some of your e-mail, as you'll see in the illustration later in this chapter.

To allay people's privacy concerns, a variety of technologies and standards are being developed. They include the Platform for Privacy Preferences (P3P), the Internet Content and Exchange standard (ICE), and the Open Profiling Standard (OPS). Generically, we'll call these Internet *passports*. These Internet passports let people control which information about themselves they'll allow to be released to Web sites and how that information can be used. And they let people control what type of information can be gathered about their surfing and how that information can be used, as well. In general, the more information that people allow to be gathered about themselves, the more specialized services they'll gain on the Web, such as customized news feeds.

How Cookies Work

1 *Cookies* are pieces of data placed on a computer's hard drive by a Web server; they can be used for a variety of purposes. They can store usernames and passwords, for example, so that people don't have to continually log on to a site that requires registration; or they can enable people to fill electronic shopping carts with goods they want to buy. Cookies also store the name of the site that placed the cookie. Only that site can read the cookie information, so information from one site can't be shared with information from another site. Cookie information is put into a special file on a hard disk. The location and files vary according to the type of computer and the browser. On PCs using Netscape, for example, the information is put into a file called COOKIES.TXT. That single text file holds all the cookies, and each cookie is one line of data in the file.



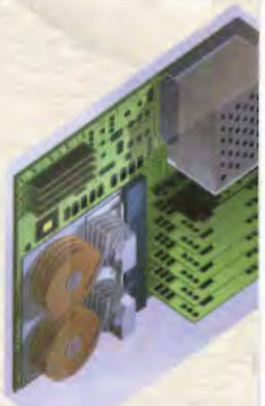
3 If no cookie is associated with the URL, the server places a cookie inside the cookie file. Some sites might first ask a series of questions, such as name and password, and then place a cookie on the hard disk with that information in it. This is typical of sites that require registration. Commonly, a CGI script on the server takes the information the user has entered and then writes the cookie onto the hard disk.

2 When you visit a site, your browser examines the URL you're visiting and looks into your cookie file. If it finds a cookie associated with that URL, it sends that cookie information to the server. The server can now use that cookie information.

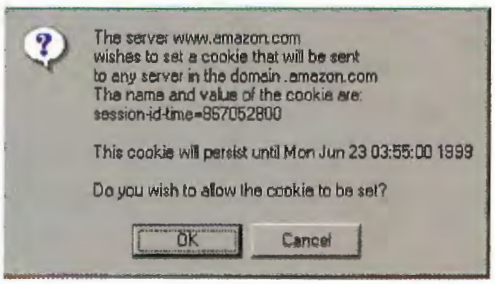
4 As you travel through a Web site, more information might need to be put into your cookie. On a site where you can purchase goods online, for example, you might put goods into an electronic shopping cart. Every time you did this, new cookie information would be added, detailing the goods you wanted to buy. When new cookie information is put in, a CGI script deletes the old cookie information and puts in a new cookie. When you leave a site, your cookie information remains on your hard disk so the site can recognize you the next time you decide to visit—unless the cookie has specifically been written to expire when you leave the site.

5 The server takes actions based on your cookies—for example, displaying your electronic shopping cart. If the site enables you to buy online, it might have asked for your credit card number. For security reasons, that number is not stored in your cookie. Instead, it is stored on a secure server. When you decide to buy something, you enter a secure area with your browser. Your cookie then sends an ID to the server that identifies you, and the server then displays your credit card information, enabling you to buy online.

6 After you order something from your electronic shopping cart—or after you decide to delete something out of the shopping cart—a new cookie is put on your hard disk; this one does not include the goods you bought or decided to take out of your shopping cart.



Web Server



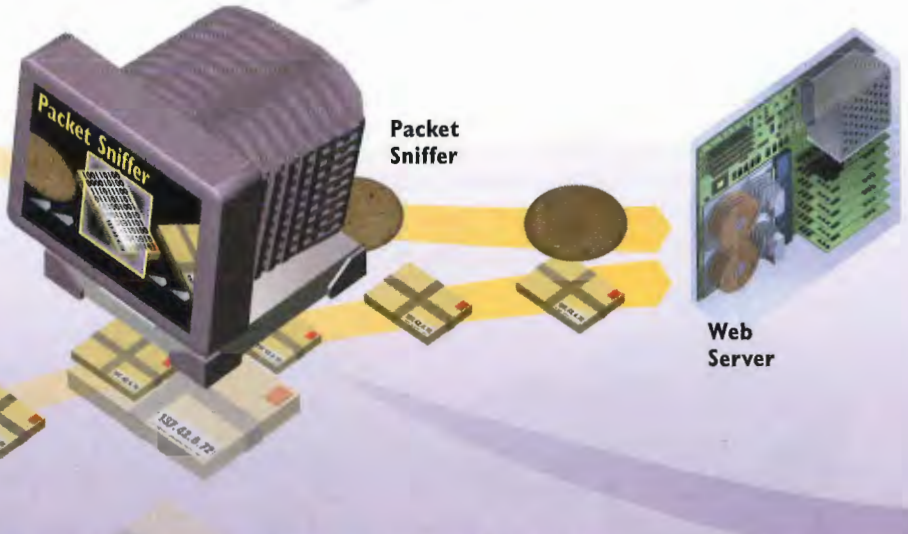
7 Because some people don't like cookies to be placed on their hard disks, browsers give people control over whether to accept cookies, to not accept cookies, or to ask each time a cookie is being placed on the hard disk. Pictured is the message you get if you've asked to be told each time a cookie is placed on your hard disk.

How Web Tracking Works

User

Cookie

1 Accrue's software doesn't rely on analyzing Web logs to figure out how a Web site is being used. Instead, a *sniffer* sits on the Internet and analyzes traffic to the site. This sniffer is a computer that runs software that examines all the TCP/IP packets coming in and out of the Web site.



Packet Sniffer

Web Server

IP Packet

2 To track traffic through a Web site, the sniffer must first identify who is coming to the site. Accrue can do this in a number of ways. If the site uses cookies, the software uses the cookie as a way to identify someone. Accrue can also use the Open Profiling Standard (OPS) information stored on a person's Web browser. OPS enables people to determine the type of information about them that can be made public. If no cookies or OPS information are present, the software uses the person's IP address.

User

Start Download

End Download

Begin FTP

3 The sniffer examines packets as they come into and go out of the site. It notes any time an action is taken, such as when someone requests a Web page, and whenever that action is completed, such as when the final packets from the page are delivered. It tracks who is making the requests, where they are coming from, where they are going, and similar information. This information is contained in the TCP/IP packets. The sniffer discards all the intermediate packets transmitted during each action—only the beginning and ending packets are necessary. It discards all the intermediate packets because they provide no useful information.

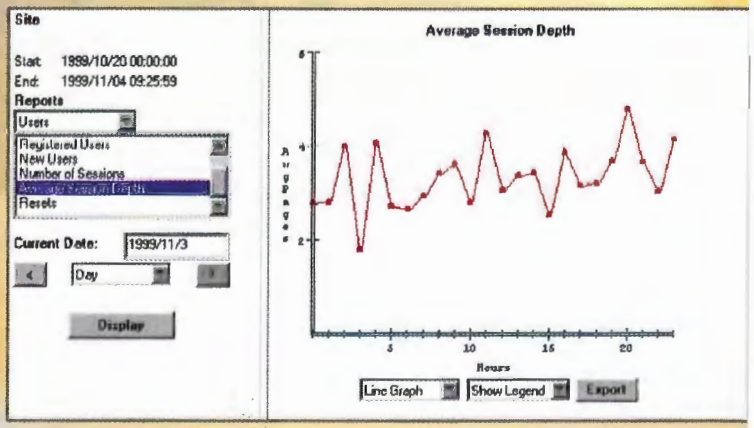
Cookie Data

4 Information is sent from the sniffer to a database, where all the information is stored.

Database

IP Data

5 Many types of reports can be created out of the database, such as the average amount of time people spend on a site, the average number of pages they read per visit, the most popular pages on a site, sites people have just visited, sites they're going to visit, and other information.



Server Traffic Analysis

How Web Bugs Can Invade Your Privacy

1 A *Web bug* is a piece of HTML code placed on Web pages or in e-mail messages that can be used to silently gather information about people, track their Internet travels, and even allow the creator of the bug to secretly read a person's e-mail. In this illustration we'll look at Web bugs used in e-mail. E-mail Web bugs can be placed only in HTML e-mail, so the person creating the bug must create an HTML-based e-mail message.

2 In the HTML code for the message, the person puts in a small piece of JavaScript code that has the capability to read the entire contents of an e-mail message.

3 The person also puts a Web bug into the e-mail message. The Web bug is an HTML reference to a tiny graphic—the smallest possible on a computer screen is one pixel by one pixel—that is transparent so that it can't be seen. This tiny graphic is also called a *clear gif* because *gif* refers to a common Web graphics format. When someone reads her HTML message, her computer gets the graphic from a server—and that server then can get information about the person's computer.

4 The person sends the e-mail message. The recipient opens the message in an HTML-enabled e mail reader, such as Outlook. (Note: If the recipient doesn't have an e-mail reader, the Web bug won't work.)

5 The JavaScript runs and reads the entire e-mail message. The person's e-mail software contacts the remote server to get the clear gif. It does more than get the gif, though—it also send identifying information about the computer, such as its IP address and the time the message is being read. It also sends the contents of the e-mail message, as taken by the JavaScript. At this point, that doesn't really matter because the sender of the message knows the contents of the message because she created it. But the sender now knows identifying information about the recipient of the mail.

Get Graphic

IP 47.32.21.2 read 5:10 09/12/02

Get Graphic

IP 132.21.23.4 read 6:20 10/13/02

Cookie

clear.gif

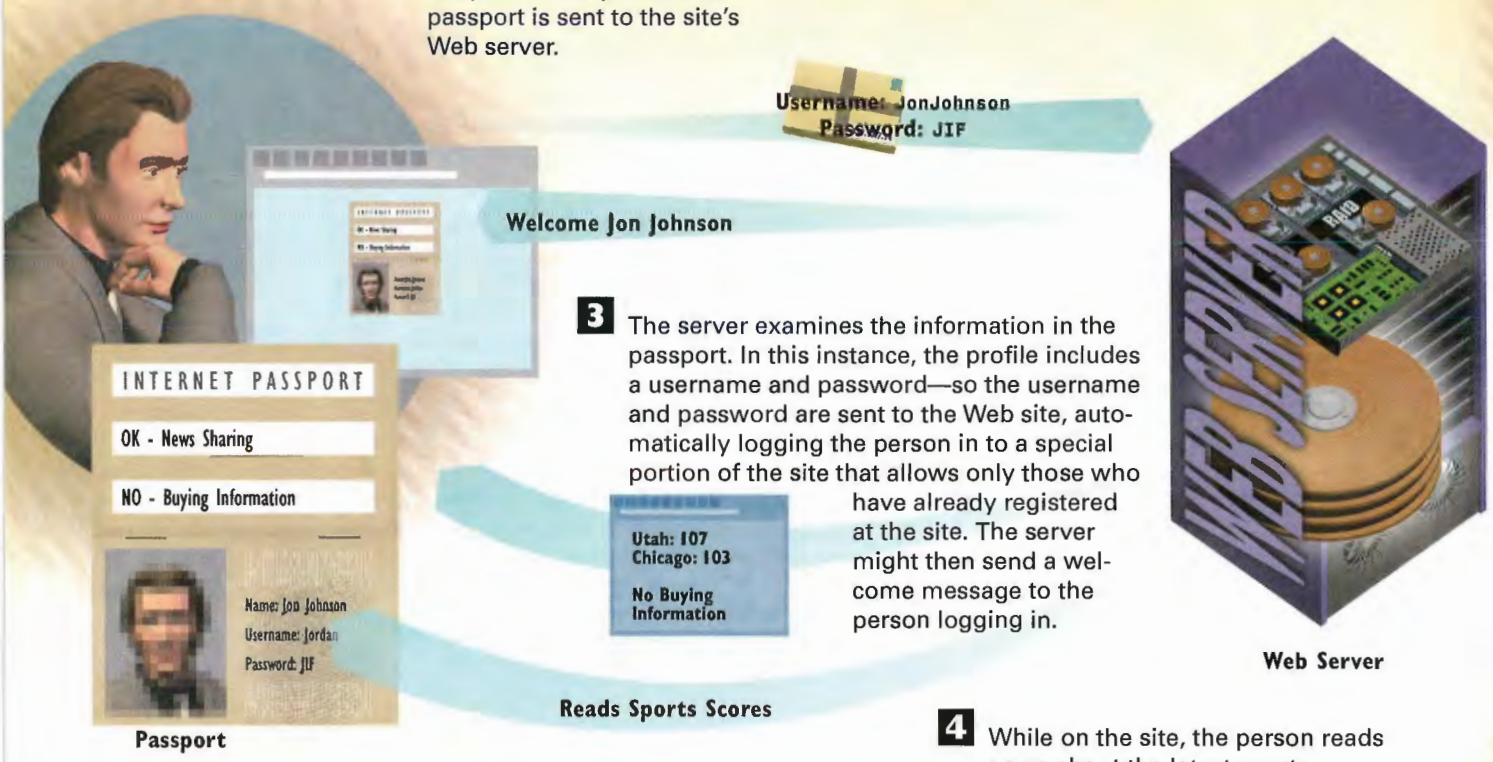
6 The server sends the gif, but can also send a cookie along with it. It can match this cookie with the identifying information send via the Web bug, and with those pieces of information track a person's use of the Internet. For example, if the piece of mail that set all this in motion was a piece of junk mail, the sender would be able to know who responded to the offer, and track what he did in response—visiting a particular Web page or buying specific products, for example. That information could then be kept in a database.

7 If the recipient of the message sends the message along to someone else, and sends a message along with it, the whole process starts all over. Now, however, when the Web bug sends the contents of the e-mail message, it contains the person's comments—so the mail has effectively been wiretapped. This can keep continuing so that every time a new person gets the message, the wiretap continues.

Forwarded E-mail

How Internet Passports Work

2 When the user visits a Web site, the information in the profile the person has put into his passport is sent to the site's Web server.



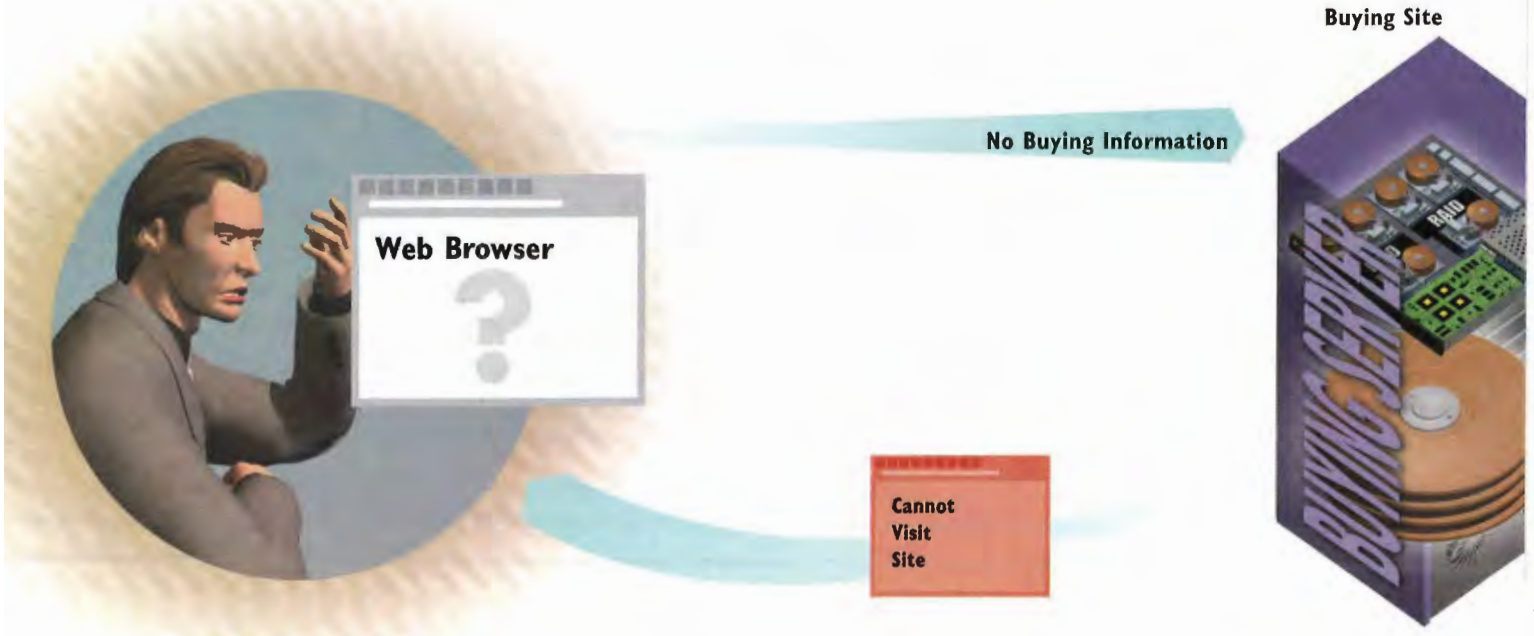
3 The server examines the information in the passport. In this instance, the profile includes a username and password—so the username and password are sent to the Web site, automatically logging the person in to a special portion of the site that allows only those who have already registered at the site. The server might then send a welcome message to the person logging in.

1 Internet *passports* are designed to let people decide what personal information they will allow to be released to Web sites. A variety of technologies are involved with Internet passports, including the Platform for Privacy Preferences (P3P), the Internet Content and Exchange standard (ICE), and the Open Profiling Standard (OPS). The passport lives inside a Web browser. A user fills out a profile in the browser, determining what information can be made available to Web sites, such as name, address, occupation, username and password, and age. The user also decides which type of information about his surfing habits can be shared among Web sites—and which can't. In this instance, the person has decided that information about what news stories he reads can be shared, but not information about what products he buys.

4 While on the site, the person reads news about the latest sports scores, then buys an electric razor and a book about sailing. The Web site puts into the person's profile that he has read sports scores, but not that he has bought a razor and a book—because the person's profile said it would allow information about what news stories he reads to be shared but not information about what he buys.

5 The person visits another Web site. The information in the profile the person has put into his passport is sent to the site's Web server. The server sees that the person has recently read a story about sports scores, so it sends to him a daily digest of the latest sports news. Because the profile doesn't have any information about what the person has bought, it doesn't send any information about special sales on the site.

6 The person now surfs to a different Web site. This site allows in only people who have agreed, in their profiles, to allow their online buying habits to be shared among sites. Because the person has said he doesn't want that information to be shared, the person is not allowed onto the site.



CHAPTER

48

Cryptography, Privacy, and Digital Certificates

THE Internet is a notoriously insecure network. Anything you send across it can be easily snooped upon. This is of particular concern when highly confidential information, such as corporate data and credit card numbers, is transmitted across the Internet. Unless there is some way to protect that type of information, the Internet will never be a secure place to do business or send private, personal correspondence.

Another related concern is that knowing that the person sending the information across the Internet, such as credit card information, is really who he says he is can be impossible. There are ways for people to forge identities and steal credit card numbers, and financial institutions and other businesses require ways to know that the person sending information really is who he says he is.

Several ways have been developed to solve these problems. At the heart of them is *encryption*—a way of altering information so that to anyone other than the intended recipient it will look like meaningless garble. When the recipient gets the information, it needs to be *decrypted*—that is, turned back into the original message by the recipient, and *only* by the recipient. Many complex cryptosystems have been created to enable this type of encryption and decryption.

Cryptosystems use what are called *keys*—secret values computers use in concert with complex mathematical formulas called *algorithms* to encrypt and decrypt messages. If someone encrypts a message with a key, only someone else with a matching key will be able to decrypt the message.

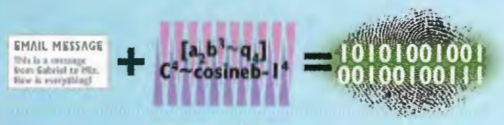
There are two kinds of common encryption systems: secret-key cryptography and public-key cryptography, also called asymmetric cryptography. Public key cryptography is what is commonly used on the Internet.

In *public-key cryptography*, two keys are involved: a public key and a private key. Every person has both a public key and a private key. The public key is made freely available, whereas the private key is kept secret on the person's computer. The public key can encrypt messages, but only the private key can decrypt messages the public key has encrypted. If someone wants to send a message to you, for example, she would encrypt it with your public key. But only you, with your private key, would be able to decrypt the message and read it. Your public key could not decrypt it.

Digital certificates use encryption to verify that the person sending information—such as a credit card number, a message, or anything else over the Internet—really is who she says she is. The certificates place information on a person's hard disk and use encryption technology to create a unique digital certificate for each person. When someone with a digital certificate goes to a site or sends e-mail, that certificate is presented to the site or attached to the e-mail, and it verifies that the user is who she claims to be.

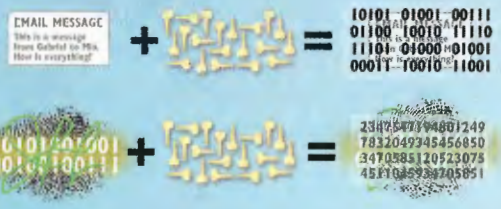
Digital certificates are issued by certificate authorities. These certificate authorities are private companies who charge either users or companies for the issuance of the certificates. You might be familiar with one such certificate authority, called VeriSign. Digital certificates contain information such as your name, the name of the certificate authority, the certificate's serial number, and similar information. The information has been encrypted in a way that makes it unique to you.

How Cryptosystems Work



1 Gabriel wants to send a confidential message over the Internet to Mia. Mia will need some way to decrypt the message as well as a way to guarantee that Gabriel—and not an imposter—has actually sent the message. First, Gabriel runs his message through an algorithm called a *hash function*. This produces a number known as the *message digest*. The message digest acts as a sort of "digital fingerprint" that Mia will use to ensure that no one has altered the message.

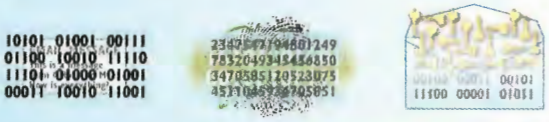
2 Gabriel now uses his private key to encrypt the message digest. This produces a unique digital signature that only he, with his private key, could have created.



3 Gabriel generates a new random key. He uses this key to encrypt his original message and his digital signature. Mia will need a copy of this random key to decrypt Gabriel's message. This random key is the only key in the world that can decrypt the message—and at this point, only Gabriel has the key.

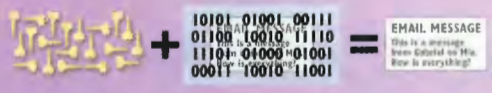
4 Gabriel encrypts this new random key with Mia's public key. This encrypted random key is referred to as the *digital envelope*. Only Mia will be able to decrypt the random key because it was encrypted with her public key, so only her private key can decrypt it.

5 Gabriel sends a message to Mia over the Internet that is composed of several parts: the encrypted confidential message, the encrypted digital signature, and the encrypted digital envelope.



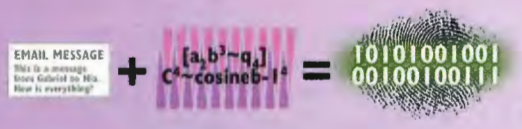
Original Message | Hash function | Message digest | Gabriel's public key | Gabriel's private key | Mia's public key

6 Mia gets the message. She decrypts the digital envelope with her private key and out of it gets the random key Gabriel used to encrypt the message.

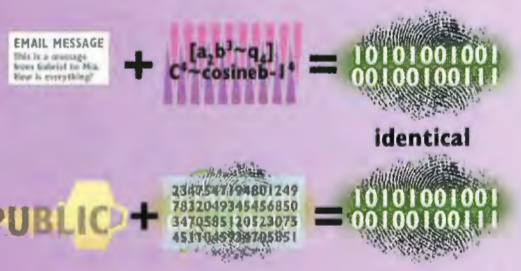


7 Mia uses the random key to decrypt Gabriel's message. She can now read the confidential message he sent to her. However, she can't yet be sure that the message hasn't been altered en route to her or that Gabriel was definitely the sender.

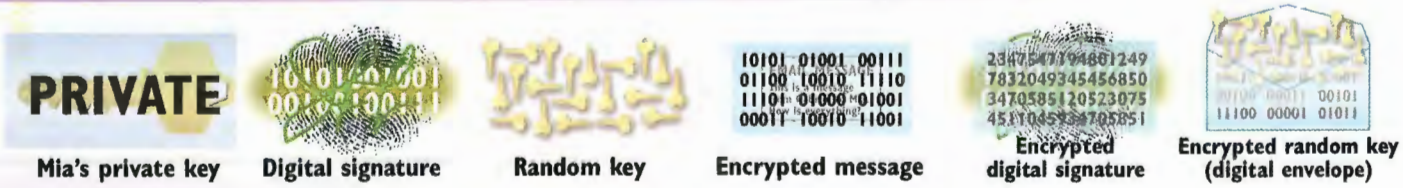
8 Mia now uses the random key and Gabriel's public key to decrypt his encrypted digital signature. When she does this, she gets his message digest, the message's "digital fingerprint."



9 Mia uses this message digest to see whether Gabriel indeed sent the message and that it was not altered in any way. She takes the message she decrypted and runs it through the same algorithm—the *hash function*—that Gabriel ran the message through. This produces a new message digest.



10 Mia compares the message digest she calculated to the one she got out of Gabriel's digital signature. If the two match precisely, she can be sure that Gabriel signed the message and that it was not altered after he composed it. If they don't match, she knows that either he didn't compose the message or that someone altered the message after he wrote it.



How Digital Certificates Ensure Internet Security

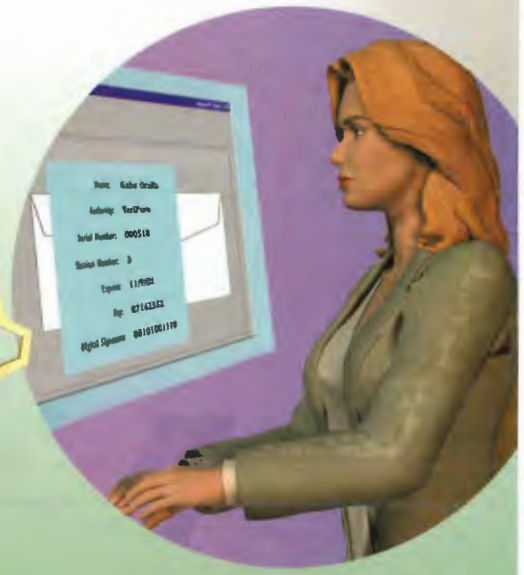
1 A digital certificate is used to guarantee that the person who sends information or e-mail over the Internet or who makes a financial transaction really is who he says he is. Digital certificates are issued by certificate authorities (CAs). To get a digital certificate, you typically visit a CA site and request a certificate. You then provide information about yourself, such as your name and other identifying information.

2 You are issued a digital certificate, which has been digitally signed to guarantee its authenticity. The certificate is data unique to you and is put on your hard disk, along with a private key.

3 The digital certificate is composed of information such as your name, the name of the CA, the unique serial number of the certificate, the version number of the certificate, the expiration date of the certificate, your public key, and the digital signature of the CA. The exact format of the certificate is defined by a standard known as X.509.

Gabe Gralla Requests Certificate

4 When you want to send e-mail to someone and have her know for certain that it is you and no one else who has sent the mail, you attach the digital certificate to your e-mail message. One of the things the certificate does is sign the message with a private key that you were given as part of the digital certificate.



5 The person to whom you're sending e-mail gets your digital certificate along with your e-mail. The key is used to read the private key's signature. That signature matches information found in the digital certificate, so the receiver is assured that the message really came from you.

CHAPTER

49

How the FBI's "Carnivore" Program Works

THE most controversial law enforcement tool related to the Internet is the FBI's so-called Carnivore system. Carnivore is the Internet equivalent of a wiretap—it allows FBI agents to examine everything that a suspect does on the Internet, from sending and reading e-mail, to browsing the Web, sending and receiving files via FTP, and, in fact, anything else someone does. It literally allows agents to examine and keep copies of every bit of information sent to and from an individual.

Just as a special warrant is required for a law enforcement agency to obtain a wiretap, so is a special warrant required for the FBI to obtain a Carnivore tap.

With Carnivore, the FBI places a Pentium-class computer at the Internet service provider (ISP) of the target of the investigation. It then taps into the line of the ISP and examines all the ISP's Internet traffic, using filters and software to get copies of the target's traffic and discarding the rest. Depending on the warrant, the FBI might target only portions of the target's Internet usage—for example, e-mail but not FTP.

Although no public records about specific the use of Carnivore exist, indications are that law enforcement officials are increasingly asking for warrants to examine the activities of Internet users. Between 1997 and 1999, for example, an increase of more than 800% occurred in warrants served to America Online by state and local investigators investigating the online activities of America Online subscribers.

Civil libertarians and those concerned with privacy issues have criticized Carnivore. They worry that it can easily be abused, and they argue that the simple fact of tapping into someone's Internet usage is an invasion of privacy.

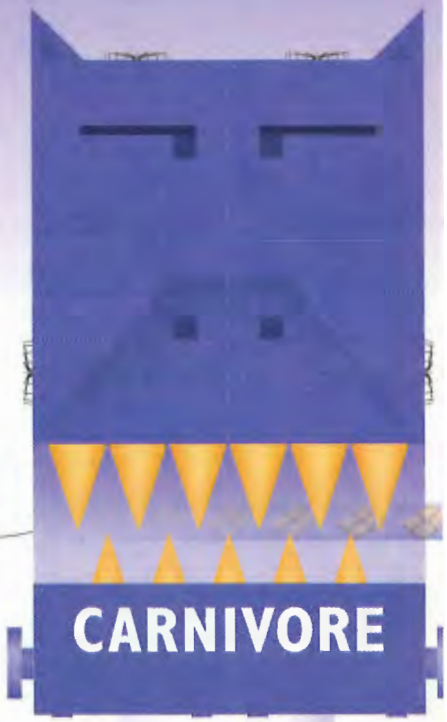
Chief among the concerns is that there is no way for people to ensure that Carnivore isn't being abused. The system can in fact examine every Internet packet of every subscriber in a given ISP. That means the federal government could have access to everything you see and do online and could keep a record of it. The FBI counters that it will not abuse its system and that it will adhere to the warrants it asks for. It claims that it asks for a Carnivore tap only in extreme circumstances, and that even in those circumstances, it limits the tap as much as possible.

But critics worry still. They point to the numerous problems the FBI has had in recent years—among them, not revealing thousands of documents it was required to by law in the Oklahoma bombing case. And they point to the FBI's past history of abuse of power in the 1960s and 1970s. The FBI says, however, that those incidents are deep in its past and that without the power to use Carnivore, it won't be able to protect the country against criminals and terrorists. In the end, it will be our country's usual democratic system of checks and balances that will determine how, when, and whether Carnivore will be used.

How Carnivore Can Invade Your Privacy

FBI

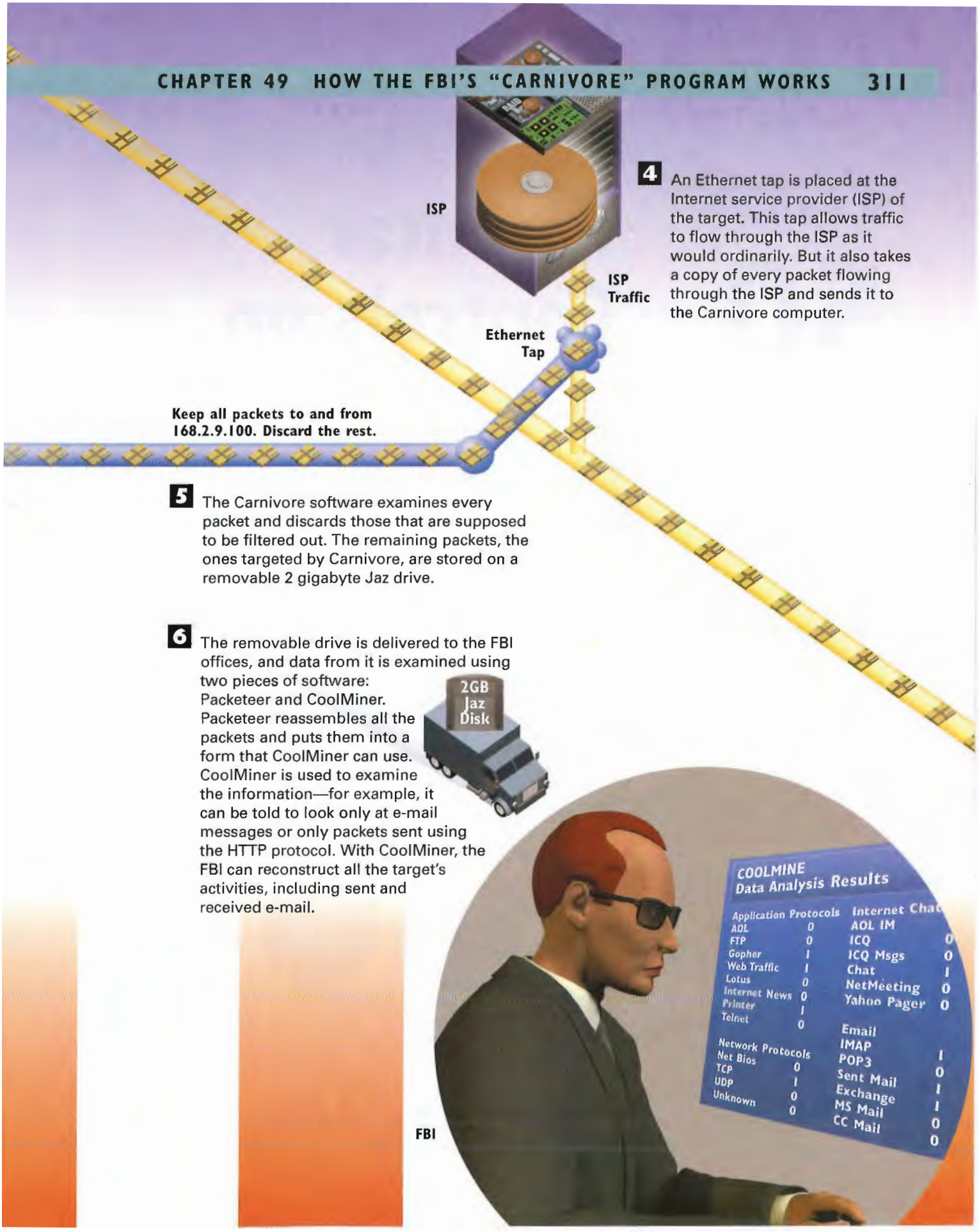
1 The FBI's system for tracking people's Internet use, including reading their e-mails and seeing what Web sites they visit, is called Carnivore. The data-gathering portion of Carnivore is a Pentium-based system, on which the Carnivore packet sniffing software runs. No keyboard or monitor is attached to the computer, so no one at the ISP can make use of it.



Dedicated Phone Line

2 The computer is attached via a dedicated phone line and a 56K modem to the FBI offices. The FBI runs an off-the-shelf program called PcAnywhere to enable it to control the Carnivore software and computer remotely. The dedicated connection is not connected to the Internet, and all data is encrypted, using both PcAnywhere's encryption and other encryption programs.

3 FBI agents have the right to monitor only someone against whom they've obtained a wiretap warrant. The warrant also might require that they gather only certain types of information about that person—for example, only his e-mail messages. They use the Carnivore software to set filters, which filter out all the data they don't want and focus only on the data they do want. So, for example, they set a filter that says only to track packets to and from a particular person, or only to examine his e-mail, or e-mail and Web usage.



4 An Ethernet tap is placed at the Internet service provider (ISP) of the target. This tap allows traffic to flow through the ISP as it would ordinarily. But it also takes a copy of every packet flowing through the ISP and sends it to the Carnivore computer.

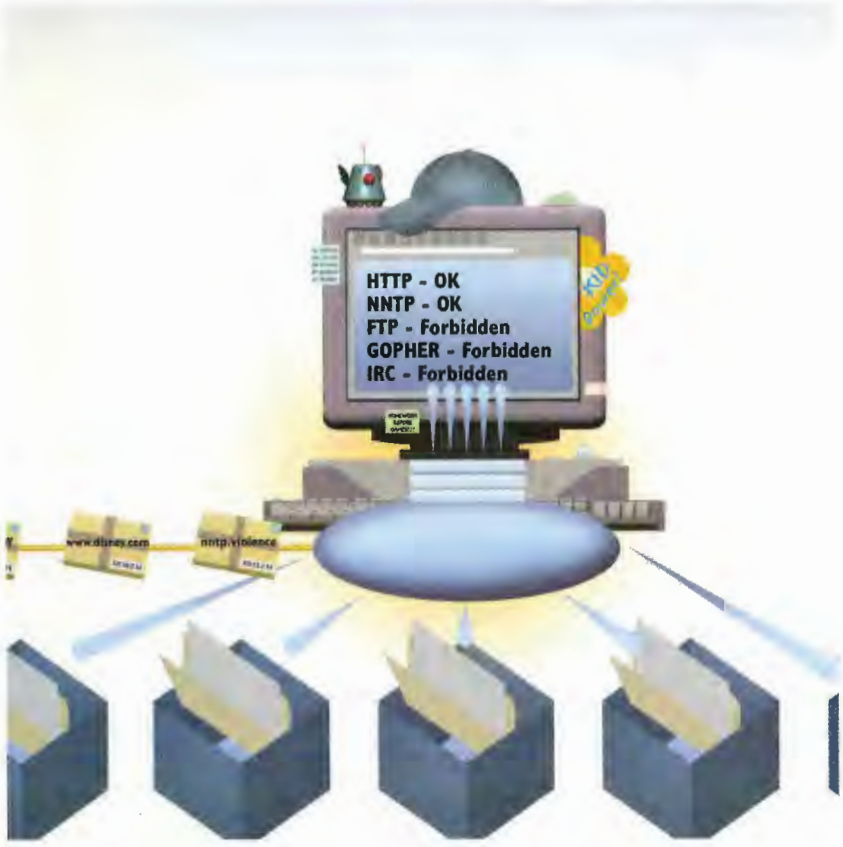
5 The Carnivore software examines every packet and discards those that are supposed to be filtered out. The remaining packets, the ones targeted by Carnivore, are stored on a removable 2 gigabyte Jaz drive.

6 The removable drive is delivered to the FBI offices, and data from it is examined using two pieces of software: Packeteer and CoolMiner. Packeteer reassembles all the packets and puts them into a form that CoolMiner can use. CoolMiner is used to examine the information—for example, it can be told to look only at e-mail messages or only packets sent using the HTTP protocol. With CoolMiner, the FBI can reconstruct all the target's activities, including sent and received e-mail.

CHAPTER

50

Parental Controls on the Internet



THE very nature of the Internet—the way it allows the free, unfettered flow of information among people—has gotten it a lot of bad publicity. Much has been made of the fact that erotic and pornographic information is available on the Internet, everything from pictures to discussions about subjects many people find objectionable. The truth is, that kind of content makes up a very small part of what's available on the Internet. Furthermore, the objectionable content is not exactly in public view—you have to do a bit of digging to find it.

However, just the fact that this type of information is available to anyone who wants to see it, including children, has made people uncomfortable. In fact, Congress and other legislative bodies have tried to take steps to ban certain types of content from being available on the Internet. As a result of these efforts, a controversial law was passed against online pornography. The law is called the Communications Decency Act, which the Supreme Court ruled unconstitutional.

The real answer to the problem, though, doesn't lie with legislation. The answer lies with technology—software that enables parents to ensure that their children are not seeing objectionable material. A number of companies make and sell software that will do this, such as SurfWatch, CyberNanny, and CyberPatrol. They each check sites for content and then bar children from getting to those sites containing content that is unsuitable for them. Some routers used for home networks also include filtering capabilities built into them.

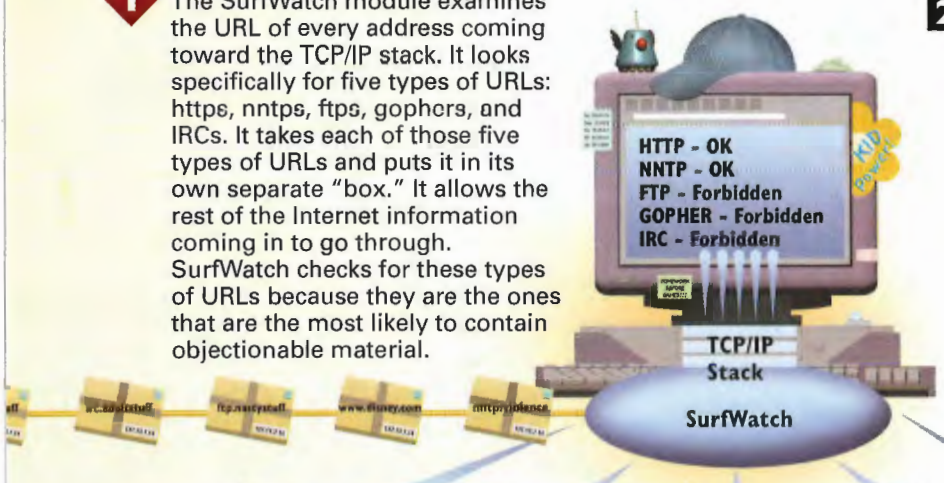
Online services such as CompuServe and America Online have a variety of ways to block access to objectionable material on the Internet. Some allow parents to block children from using services such as the World Wide Web, chat, or newsgroups completely. Others, such as America Online, license technology from software makers, such as those that manufacture SurfWatch, to enable anyone on their service to block Internet sites they don't want their children to visit.

One group working on the issue is Platform for Internet Content Selection (PICS), which is trying to give parents control over the type of material to which their children have access. The group is trying to develop industry standards for technology that would allow the content of all sites and documents on the Internet to be rated according to its suitability for children. Additionally, the group would create standards to enable software to be developed for blocking sites based on those suitability ratings.

Businesses are also concerned with the type of Internet material their workers are accessing over corporate networks. There is a feeling that getting at and displaying sexual material could be interpreted as sexual harassment. Furthermore, some companies simply don't want their workers accessing that material on company time. Some companies now lease the same software parents are buying. Instead of installing the software on individual computers, though, the software is installed on a server, and it checks all incoming Internet traffic to every computer in the company.

How Parental Controls Work

1 The SurfWatch module examines the URL of every address coming toward the TCP/IP stack. It looks specifically for five types of URLs: https, nntps, ftps, gophers, and IRCs. It takes each of those five types of URLs and puts it in its own separate "box." It allows the rest of the Internet information coming in to go through. SurfWatch checks for these types of URLs because they are the ones that are the most likely to contain objectionable material.



2 SurfWatch software is installed on a computer a parent wants to monitor to ensure that children can't get to objectionable material on the Internet. When a child launches software to get onto the Internet, SurfWatch latches onto Winsock or MacTCP, depending on whether a PC or Macintosh is being used. A SurfWatch software module sits "in front" of Winsock or MacTCP (which we'll generically call a TCP/IP stack) and monitors the TCP/IP data stream coming to the TCP/IP stack from the Internet.

Objectionable URLs: `meanstuff.com`, `naughtystuff.com`, `sexystuff.com`, `violentstuff.com`

3 Every URL in each of the boxes is checked against a database of the URLs of objectionable sites. If SurfWatch finds that any of the URLs are from objectionable sites, it doesn't allow that information to be passed on to the TCP/IP stack, blocking the site and preventing information from being viewed. It alerts the child that the site has been blocked. SurfWatch checks thousands of sites and lists in its database the ones that are found to be objectionable.

Objectionable words: `sex`, `violence`, `pornography`, `guns`

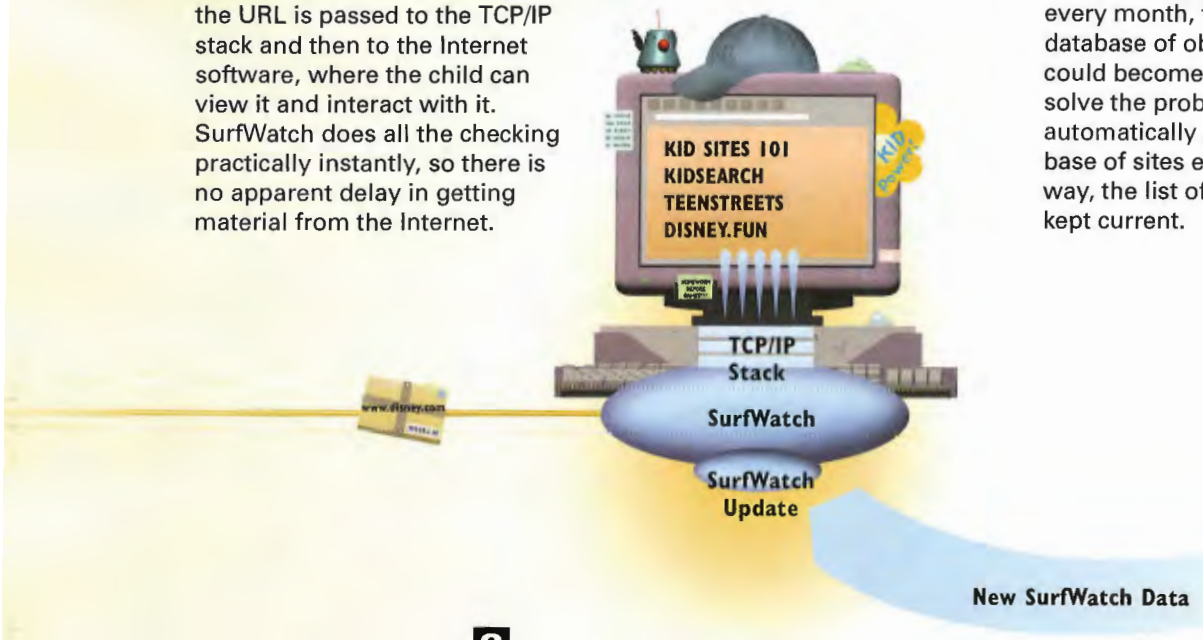
4 If the URL is not in the database, SurfWatch does another check of the URL. This is called *pattern matching*. It looks at the words in the URL and checks them against a database of words to see whether any of them indicates a request for objectionable material. Often, people creating objectionable material put representative words in the URL to draw attention to the site. If SurfWatch finds a matching pattern, it doesn't allow that information to be passed on to the TCP/IP stack, blocking the site and preventing information from being viewed. It also alerts the child that the site has been blocked.

PICS

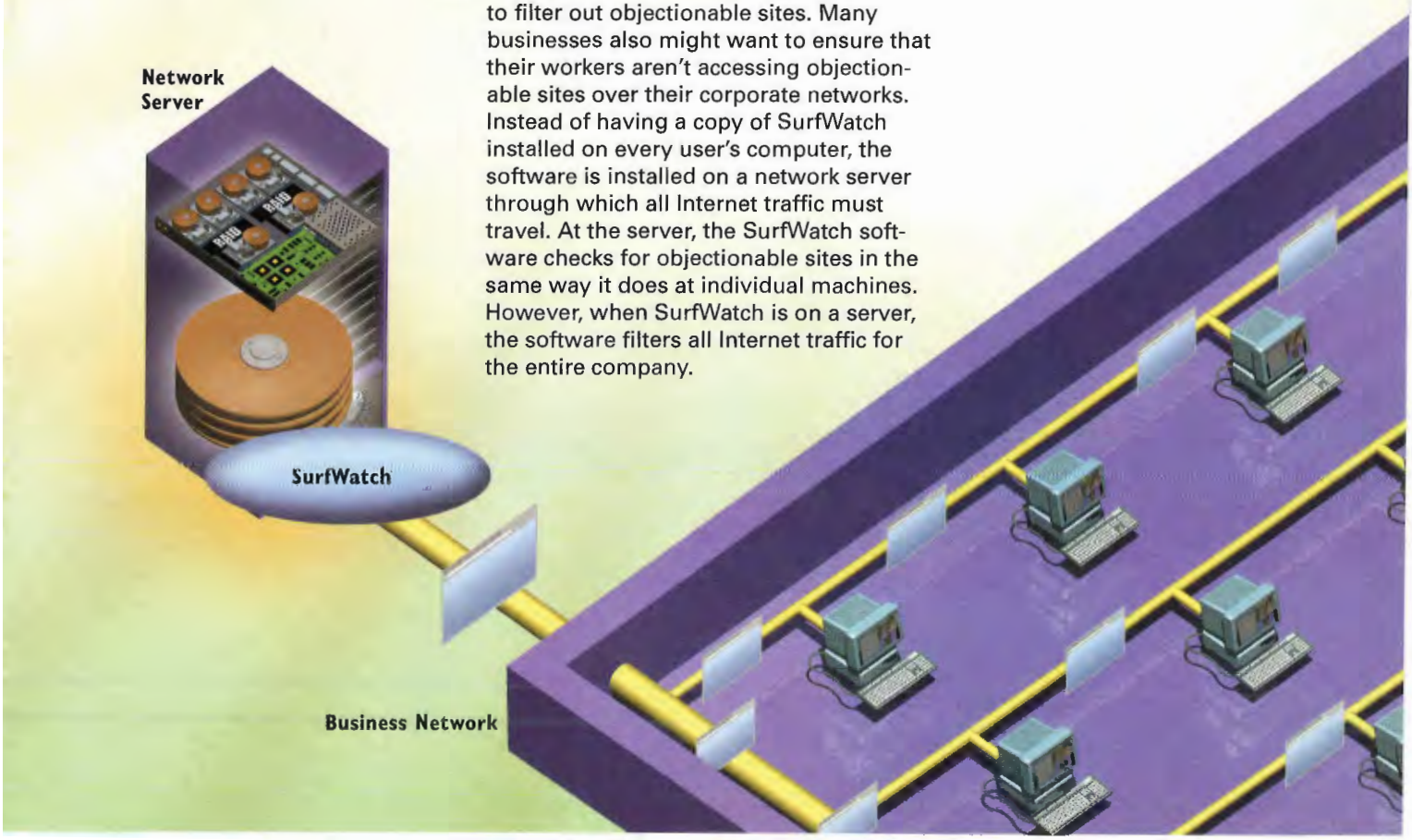
5 There is another way that SurfWatch eventually might check for objectionable sites. A rating system called PICS (Platform for Internet Content Selection) is being developed that will embed information about the content in its documents—saying, for example, whether objectionable material can be found there. If SurfWatch uses this system and finds that the URL is of a site containing objectionable material, it won't allow that information to be passed on to the TCP/IP stack, blocking the site and information from being viewed. It also will alert the child that the site has been blocked.

6 If the URL is not found to be of an objectionable site after the checks have been completed, the URL is passed to the TCP/IP stack and then to the Internet software, where the child can view it and interact with it. SurfWatch does all the checking practically instantly, so there is no apparent delay in getting material from the Internet.

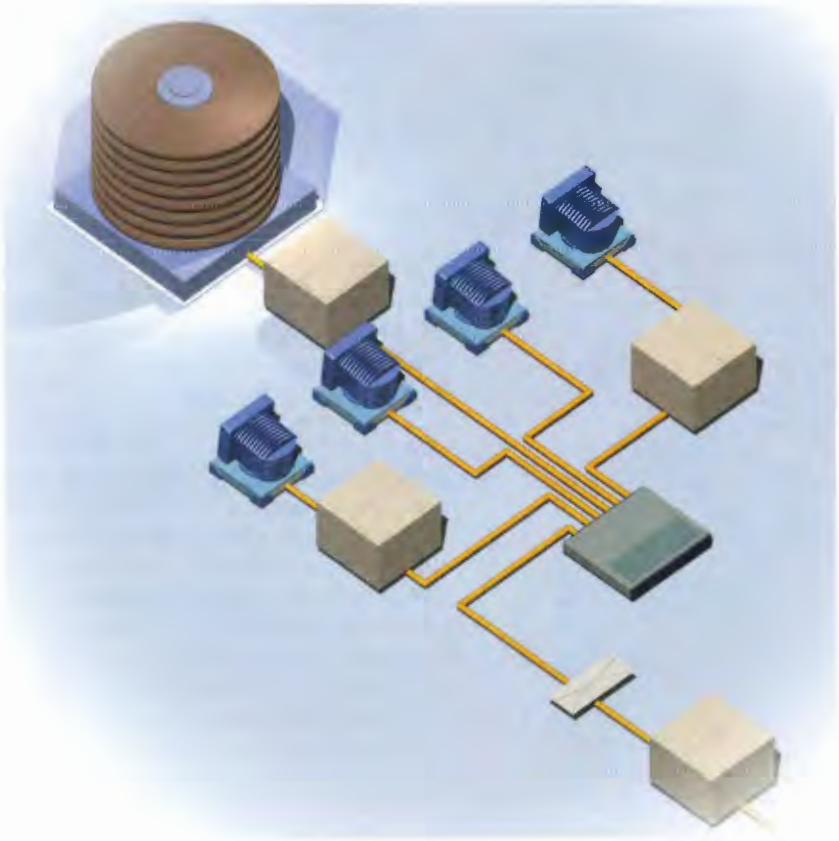
7 Because the Internet is growing so quickly, and so many new sites are being created every month, the SurfWatch database of objectionable sites could become outdated. To solve the problem, SurfWatch automatically updates the database of sites every month. That way, the list of sites is always kept current.



8 Parents aren't the only people who want to filter out objectionable sites. Many businesses also might want to ensure that their workers aren't accessing objectionable sites over their corporate networks. Instead of having a copy of SurfWatch installed on every user's computer, the software is installed on a network server through which all Internet traffic must travel. At the server, the SurfWatch software checks for objectionable sites in the same way it does at individual machines. However, when SurfWatch is on a server, the software filters all Internet traffic for the entire company.



Glossary



ActiveX A technology for Microsoft Internet Explorer that enables programs to be downloaded and run in your browser.

Address An Internet location, such as a URL, an IP address, or an e-mail address.

Agent A piece of software that goes out across the Internet and does a job for you—for example, finding the best prices on a product you want to buy.

Application service provider (ASP) A service that enables you to run software over the Internet without installing it on your computer. Many ASPs enable the software to run inside your Web browser when you visit a Web site.

ASCII characters Plain-text characters that you get by pressing keys on your keyboard.

Attachment See *File attachment.*

Audio file A file you can download or play from the Internet that has music or sounds in it.

Avatar A picture that represents you in picture-oriented chat rooms.

Bandwidth A measure of the amount of data that can be sent across an Internet connection over a unit of time.

Bluetooth A wireless networking standard that allows devices of many different kinds to communicate in a peer-to-peer fashion, that is, without having to use a server or other hardware to connect them.

Bridge A device that connects local area networks with each other.

Broadband connection A very fast Internet connection, such as via a cable modem, or DSL.

Browser See *Web browser.*

Buddy list In instant messaging software, a list of friends you create so that you are alerted whenever one of your "buddies" comes online.

Cable modem A device used to connect a computer to the Internet at very high speeds over cable TV lines. The device isn't a true modem, though, and uses a network card inside the computer to connect to the cable line.

Cache A place on a computer or server that temporarily stores items such as Web pages and graphics so they can be more quickly retrieved.

Carnivore An FBI hardware and software system that can be used to read people's e-mail and track everything they do when they are on the Internet.

Chat A way that two or more people can communicate in real-time by typing messages on their keyboards.

Chat room A location in cyberspace where people go to chat.

Client A piece of software running on a local computer or device that communicates with a central server.

Client pull animation A Web animation technique in which the Web browser requests a series of images that, when displayed one after another, appear to be animated.

Client/server architecture A model of computing in which clients on local computers cooperate with distant servers to compete tasks. The Internet is largely based on client/server architecture.

Coaxial cable The type of cable used for cable TV connections. It can also be used to provide high-speed access to the Internet, via cable modems.

Common Gateway Interface (CGI) A communications protocol that enables Web servers to communicate with applications, such as databases.

Cookies A bit of data put on your computer by a Web server that can be used to track what you do when you are on the Web.

Cryptosystems Systems used to encrypt data and then decrypt data so that only the intended recipient can read it.

Decryption A method of unscrambling encrypted data so that it can be understood.

Digital certificate A key used to encrypt and decrypt information; it can be used to guarantee that you're the sender of a message or to verify the authenticity of a person sending you a message.

Digital signature An encrypted electronic "signature" that identifies you as the sender of a message—and that can't be forged.

Digital subscriber line (DSL) A way of giving a computer high-speed access to the Internet using existing phone lines. A DSL modem is required.

Domain An area of the Internet owned by a company or person, such as `zdnet.com`.

Domain name server A server that translates Internet addresses, such as `www.zdnet.com`, into their IP addresses, such as `128.42.23.68`, and vice versa.

Domain Name System (DNS) The system that translates Internet addresses, such as `www.zdnet.com`, into their IP addresses, such as `128.42.23.68`, and vice versa.

Download To transfer information or files from the Internet to your computer.

Dynamic Host Configuration Protocol (DHCP) A protocol used to renew an IP address or provide a new IP address to a computer when it connects to a server. When you connect to the Internet using an Internet service provider, you usually are given a different IP address every time you connect.

Dynamic HTML (DHTML) A group of HTML-related technologies that allows for greater interactivity and animation on Web pages.

Dynamic IP address An IP address delivered via DHCP; with a dynamic IP address, the IP address of a computer will be different each time it goes onto the Internet.

E-mail filter A way of automatically sorting incoming e-mail so that some are automatically routed to certain folders or deleted, based on the sender and the content of the message. E-mail filters can be used to cut down on spam sent to you.

E-mail reader A piece of software used to send and receive e-mail.

Encryption A method of scrambling data so that it can be read only by its intended recipient.

Ethernet The most common local area networking standard.

Ewallet An electronic wallet that contains your credit card information or electronic money so that you can use it to easily shop at many online shopping sites.

eXtensible Markup Language (XML) An extension of HTML that separates the content of a Web page from its display. It can be used to allow designers to easily create Web pages to be displayed on many different devices, such as computers, cell phones, and PDAs.

File attachment A file attached to an e-mail message or a newsgroup posting. Any type of file can be attached to e-mail or newsgroup postings.

File compression Shrinking the size of a file down so that it can be transferred more quickly over the Internet.

File extension The letters on the end of a filename that are used to identify the type of file it is. For example, files with .doc extensions are Microsoft Word files.

Firewall A hardware or hardware/software combination that protects computers on a network from being attacked by hackers or snoopers.

Flash movie An animated movie played over the Web, created using Macromedia's Flash software.

FTP (File Transfer Protocol) A way of downloading files on the Internet. See also *Download*.

Gateway A device that connects local area networks with each other and can translate data from one network to another.

GIF (Graphics Interchange Format) A common graphics format used on Web pages. Files in this format end in .gif.

Helper application See *Plug-in*.

Hops The number of times a packet of information needs to be sent to different routers before reaching its destination.

Host See *Server*.

HTML (Hypertext Markup Language) The computer language that forms the basis of the World Wide Web. Web browsers interpret HTML commands and display Web pages based on the HTML commands.

HTTP (Hypertext Transfer Protocol) An Internet protocol that defines the way Web browsers and Web servers communicate with each other.

Hub A device that connects several computers to one another on a network.

Hub/Router A combination of a hub and router that connects computers, routes data among them, and provides access to the Internet or other networks. Home networks commonly use a hub/router.

Hyperlink A link on a Web page that sends you to another Web page or resource.

Hypertext Text that, when clicked, sends you to another piece of text or location.

Image map A static image that has been turned into a clickable image with different clickable parts.

Instant message A chat-like message sent to another individual in a private, one-on-one conversation.

Instant messaging software Software that enables people to know when their friends are online and lets them send person-to-person messages.

Internet-enhanced TV The use of Internet technologies to add interactivity and Web links to television broadcasts.

Internet service provider (ISP) A company that provides dial-in or some other type of access to the Internet for a monthly fee.

Internet telephony The use of the Internet to make telephone calls.

Intranet A private network inside a corporation that uses Internet technology.

IP address An Internet address that is a series of four numbers separate by dots, such as `155.40.112.23`. Every time you go onto the Internet, you use an IP address; without it you can't do things such as surf the Web.

IRC (Internet Relay Chat) A standard that enables people to chat with each other over the Internet. You need special IRC software to chat via IRC.

ISDN (Integrated Services Digital Network) A method of establishing a high-speed connection to the Internet using telephone lines. Special lines and modems are needed for ISDN.

Java A programming language used to create programs that can be run inside Web browsers or on a variety of computers. The strength of Java-written programs is that they can be written once and can then run as is in a variety of computers.

JavaScript A technology that enables Web designers to use a variety of interactive features on Web pages.

JPEG (Joint Photographic Experts Group) A common graphics format used for Web pages. Files in this format end in .jpg.

Key A piece of data used to encrypt or decrypt information.

Lightweight Directory Access Protocol (LDAP) A protocol that enables the creation of Internet white pages, which let people look up other people's e-mail addresses.

listserv A type of software that manages sending and receiving e-mail broadcasts and discussions. The term often is used generically to describe an e-mail broadcast.

Local area network (LAN) A network that connects computers to each other so they can easily communicate.

MacTCP Software for Macintosh computers that interprets TCP/IP commands.

Mail header The part of an e-mail message that contains the subject line, the sender, the receiver, and similar information.

Mail server A server that delivers or receives e-mail.

Mailing list See *listserv*.

MBone (Multicast backbone) A high-capacity Internet backbone used for transmitting broadcasts using the Multicast IP protocol.

Message board A public area online where people can read and send messages.

Metasearch software Software that can search through many search engines simultaneously and report back the results.

Microbrowser A browser that a cell phone or similar device uses to browse the World Wide Web.

Microsoft Outlook A popular e-mail program.

Moderated newsgroup A newsgroup in which all postings first have to go through a moderator before being posted.

MP3 file A special music format that's of almost CD quality, but that produces files that aren't very large and so don't take a long time to download.

Multicast IP A protocol that enables video and audio broadcasts to take place, while using a minimum of bandwidth.

Name server A server that translates Internet addresses, such as `www.zdnet.com`, into their IP addresses, such as `145.45.23.45`.

Napster A popular program used for sharing music files in the MP3 format with others.

NetCam A video camera that attaches to a computer and often is used for Internet videoconferencing or videochat.

Network address translation (NAT) A technique in a local area network that provides an internal IP address to computers inside the network, while masking the IP address to the outside world. It also enables several computers on the local area network to share an external IP address.

Network card An add-in card put into a computer so that it can get onto a network.

Newsgroup A discussion area on the Internet.

Newsgroup reader A piece of software used to read newsgroups.

Node A portion of a network through which many computers are connected.

Online auction Just like a real-life auction, except that it's done online.

Opt out A policy that lets you say you don't want to receive junk mail or similar information.

Packet A piece of data that has been broken down into pieces for transmitting over the Internet or another network.

Packet switched network A network in which there is no unbroken connection between sender and receiver; instead data is broken into packets, sent, and then reassembled when received. The Internet is a packet switched network.

Palm Query Application (PQA) A small piece of software on a wireless Palm device that enables it to get information from the Internet using Web clipping.

Palmtop computer A small computer, such as the Palm, that fits in the palm of your hand and is often used for keeping track of schedules, to-do lists, and a calendar; it also can be used for wireless communications.

Parental controls/content filtering A feature of America Online, some routers, and some add-in software that lets parents decide where kids can go on America Online and the Internet and how they can use America Online and the Internet.

Passport A technique that enables people to determine what information to give to Web sites and what information to keep private.

Password A set of private letters and numbers or words you type in to give you access to a service or site.

Peer-to-peer network A network that enables computers or other devices to connect directly with one another without having to use a server or other hardware to connect them.

Personal Digital Assistant (PDA) A small handheld computer, such as a Palm device or Windows CE device.

Personal firewall A hardware or hardware/software combination that protects an individual computer from being attacked by hackers or snoopers.

Piconet A network formed by the connection of two or more Bluetooth devices with one another.

Plug-in A piece of software that installs in a browser or works in concert with a browser, such as for displaying different types of video.

Point-to-Point Protocol (PPP) A protocol for computers to connect to the Internet using dial-up modems.

POP 3 (Post Office 3) A communications protocol used by e-mail servers to deliver e-mail.

Pretty Good Privacy (PGP) A program used to encrypt and decrypt information. It's especially useful for sending out private e-mail that only the sender and recipient can understand.

Private key Someone's key in an encryption scheme that only one person can use. It's used in concert with that individual's public key to encrypt and decrypt information. See also *Key* and *Public key*.

Proxy server A server located between a client, such as a Web browser, and the server the client is trying to contact, and which tries to fulfill the request before sending it to the server. For example, a proxy server could be used to speed up the delivery of Web pages.

Public key Someone's key in an encryption scheme that anyone can use. It's used in concert with that individual's private key to encrypt and decrypt information. See also *Key* and *Private key*.

RealPlayer A popular piece of software that plays video and audio files.

Registrars Private companies that accept payment from companies and individuals who buy Internet domains.

Router A piece of hardware that sends packets to their proper destinations on the Internet.

Routing table A database in a router that details the various paths packets can take en route to their destinations.

Search engine A Web site that lets you perform searches throughout the Internet.

Secure site A site that encrypts your credit card information as it's sent across the Internet so the credit card number can't be stolen.

Serial Line Internet Protocol (SLIP) A protocol for computers to connect to the Internet using dial-up modems. It's not as effective as a similar protocol, called PPP.

Server A computer that performs some task for other computers, such as sending or receiving e-mail or delivering Web pages.

Server push animation A Web animation technique in which a server sends a series of images to a browser that, when displayed one after another, appear to be animated.

SET (Secure Electronic Transactions) The electronic encryption and payment standard that a group of companies, including Microsoft, Netscape, VISA, and MasterCard, is pushing to become the standard for doing electronic commerce on the Internet.

Shockwave An animated movie played over the Web, created using Macromedia's Director software.

Shopping cart A list of items someone wants to buy when at a online store.

SMTP (Simple Mail Transfer Protocol) A communications protocol used to send e-mail.

Smurf attack A method hackers use to attack Internet service providers or Web sites.

Socket Software that understands and interprets TCP/IP commands.

Spam Junk e-mail sent to people who haven't requested it. Most spam are commercial offers and can also be fraudulent.

Spam filter Software that can filter out spam before it is received.

Spamoflauge The act of hiding of a spammer's true e-mail address so that the true sender of the spam can't be traced.

Spider Software that gathers information from the Web and puts it into a large database that can be searched by search engines.

SSL (Secure Sockets Layer) A technology that scrambles information as it's sent across the Internet so hackers can't read it.

Static IP address A fixed IP address that never changes. Unlike a dynamic IP address, it is permanent, so the IP address of the computer never changes whenever it goes onto the Internet.

Streaming audio See *Streaming media.*

Streaming media A technique that enables you to view and listen to audio and video files from the Internet while they're still downloading to your computer. With streaming media, you can view and listen to audio and video files only a few seconds after you click them.

Streaming video See *Streaming media.*

T1 line A high-speed line that can carry data at a rate of 1.544Mbps.

T3 line A high-speed line that can carry data at a rate of 44.746Mbps.

TCP/IP (Transmission Control Protocol/Internet Protocol) The communications protocols that underlie the Internet.

TCP/IP stack See *Socket.*

Telnet A way of controlling a host computer from your own computer over the Internet.

Trojan horse A malicious program that appears to be benign, but in fact is doing damage to your computer. Some Trojan horses can give hackers complete access to the computers of people who run them.

TRUSTe A company that sets voluntary standards for privacy on the Internet and that gives out "seals" that companies can post on their Web sites if the companies adhere to those privacy rules.

Uniform resource locator (URL) An address on the Internet, such as `www.zdnet.com`, that enables computers and other devices to visit it.

Universal Serial Bus (USB) A technology that enables many devices to connect to a computer, such as NetCams, scanners, and digital cameras. The devices can be attached to one another in daisy-chain fashion, allowing many to be connected at once.

Unmoderated newsgroup A newsgroup in which postings don't have to go through a moderator before being posted.

Upload To transfer a file from your computer to another computer or to a server.

Usenet An Internet service for newsgroups.

Videoconference A conference among several people in which they can talk to one another and see each other using video cameras over the Internet.

Virtual Private Network (VPN) A encryption technique that enables people to connect to their corporation's network over the Internet, while protecting the data from being seen by anyone else.

Virtual reality (VR) The simulation of reality on a computer screen or over the Internet.

Virtual Reality Modeling Language (VRML) The language used to create Virtual reality Web sites.

Virus A malicious program that attacks a computer.

WAP Transaction Protocol (WTP) A communications protocol, part of the Wireless Access Protocol (WAP), that is the equivalent of the Internet's TCP/IP protocols and enables cell phones and similar devices to access the Internet.

Web browser A piece of software that enables people to browse the World Wide Web.

Web bug A technique that enables Web sites or people to track people's activities when they visit the Web or use e-mail.

WebCam A video camera that sends live still or video images to a Web site.

Web clipping A technique that enables Palm devices to get information from the Internet.

Web page template A preformatted design for a Web page that includes colors, fonts, layout, and other elements. Templates make creating Web pages easy—you only have to put in your own words, pictures, and content.

Web tracking A technique used by Web sites that tracks what people do when they visit a Web site.

WebTV A product that lets you get access to the Web on your television set.

Web white pages Web sites that contains information that can be searched through, for identifying information such as e-mail addresses, phone numbers, and addresses.

Whiteboard In videoconferencing, an application that enables several people to work on the same screen simultaneously.

Winsock Software for Windows that interprets TCP/IP commands.

Wireless access point A device that connects wireless devices, such as a computer equipped with a wireless network card, to a network.

Wireless access protocol (WAP) An Internet protocol that defines the way cell phones and similar devices can access the Internet.

Wireless Markup Language (WML) A markup language related to HTML that is used to create Web sites that cell phones and similar devices can visit.

WMLScript A scripting language that enables interaction between cell phones and the Internet.

Workgroup software Software that enables groups in a corporation to work more closely and effectively with each other and does things such as route documents among people and allow people to run whiteboard applications.

World Wide Web The most popular portion of the Internet, it allows you to view pages that include text, pictures, video, sound, and various forms of interactivity.

World Wide Web Consortium (W3C) The group that develops standards for the evolution of the World Wide Web.

Index

Symbols

A

B

D

E

N

Q-R

S

U

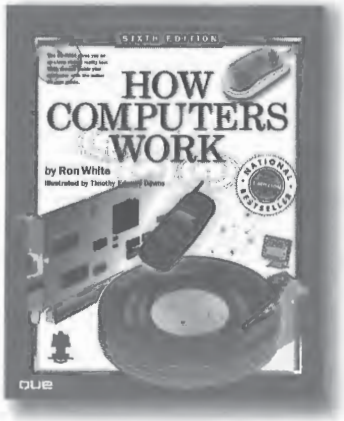
V

W

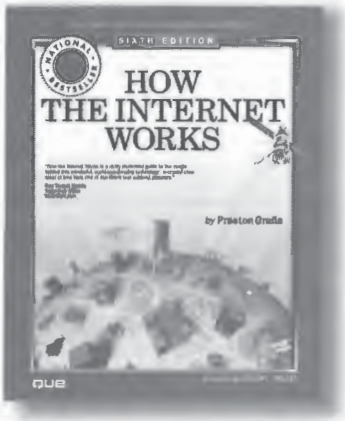
X-Z

HOW IT WORKS

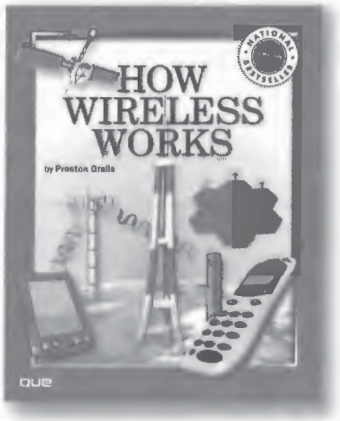
The *How It Works* series offers a unique, visual, four-color approach designed to educate curious readers. From machine code to hard-drive design to wireless communication, the *How It Works* series offers a clear and concise approach to understanding technology—a perfect source for those who prefer to learn visually. Check out other books in this best-selling series by Que:



How Computers Work, Sixth Edition offers a unique and detailed look at the inner workings of your computer. From keyboards to virtual reality helmets, this book covers it all.
ISBN: 0-7897-2549-5
US $34.99
CAN $52.95 UK £25.50



How the Internet Works, Sixth Edition clearly explains how the Internet works and gives you behind-the-scenes information. Find out what actually happens when you send an e-mail or purchase goods over the Web.
ISBN: 0-7897-2582-7
US $29.99
CAN $44.95 UK £21.99



How Wireless Works is your key to understanding current and emerging wireless technologies. This book shows you how wireless home networks communicate, how cell phones access the Internet, and much more.
ISBN: 0-7897-2487-1
US $29.99
CAN $44.95 UK £21.99

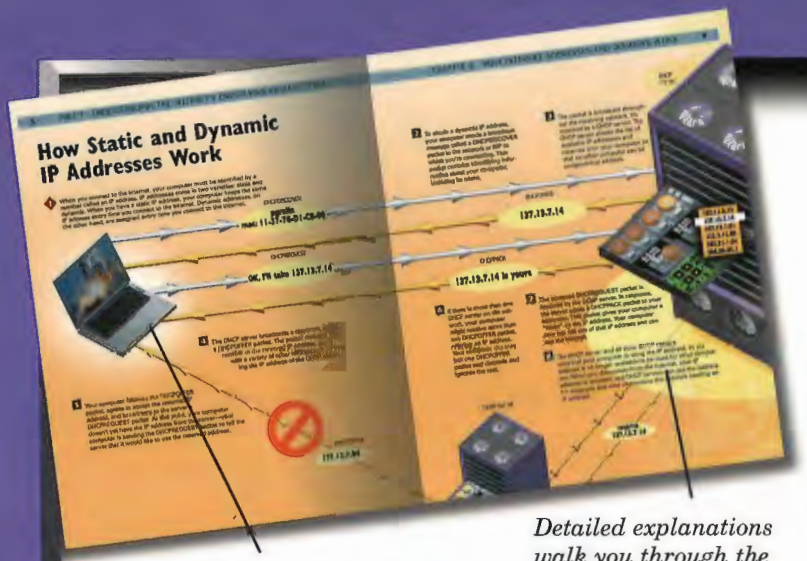
OTHER GREAT BOOKS IN THIS SERIES

How the Mac Works, Millennium Edition
ISBN: 0-7897-2428-6
US $29.99
CAN $44.95 UK £21.99

How Networks Work, Millennium Edition
ISBN: 0-7897-2445-6
US $29.99
CAN $44.95 UK £21.99

How to Expand and Upgrade PCs, Second Edition
ISBN: 0-7897-2500-2
US $29.99
CAN $44.95 UK £21.99

que www.quepublishing.com



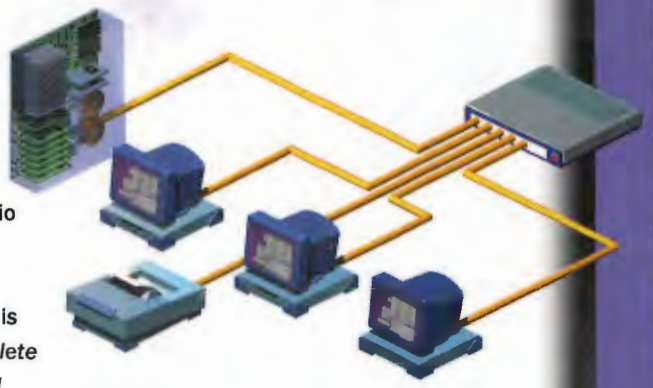
See how information really moves on the Internet.

Detailed explanations walk you through the technology.

The Internet has completely changed the way people work, communicate, socialize, and have fun. All around the world, millions of people log on the cyberspace network that knows no political, racial, ethnic, or religious boundaries to meet, conduct research, send e-mail, and play games. It's a revolutionary medium that is rewriting the script of human interaction.

FINALLY UNRAVEL THE MYSTERY OF HOW THE INTERNET REALLY WORKS.

- **Learn about MP3 music files and see how music-sharing software like Napster works.**
- **Understand the latest technologies, such as wireless Internet access, cable and DSL modems, and home networks.**
- **Discover how hackers can attack the Internet and your computer, and get a look inside the FBI's controversial Carnivore Web-tapping system.**
- **Closely examine new technologies, such as Java™, ActiveX™ Agents, and Flash™.**
- **Learn how to use e-mail, surf the Internet using a Web browser, and more.**
- **Explore the growing world of Web commerce and learn how to purchase products securely through the Internet.**



Preston Gralla is a long-time journalist with Ziff-Davis Publications and is Executive Editor for CNet and ZDNet and a technology columnist for the *Dallas Morning News*. As a well-known Internet guru, he appears regularly on many TV and radio shows, including CNN, MSNBC, the *CBS Early Show*, and CNBC. While at *PC/Computing* magazine, he was instrumental in the development of the *How It Works* features on which the series is based. He is also the author of *How Wireless Works*, *The Complete Idiot's Guide® to Protecting Yourself Online*, and *How to Expand and Upgrade PCs*.

Category	The Internet
Covers	Online/Communications
User Level	Beginning—Intermediate

$29.99 USA / $44.95 CAN / £21.99 Net UK

ISBN 0-7897-2582-7

0 29236 72582 2

9 780789 725820 52999

que

www.quepublishing.com