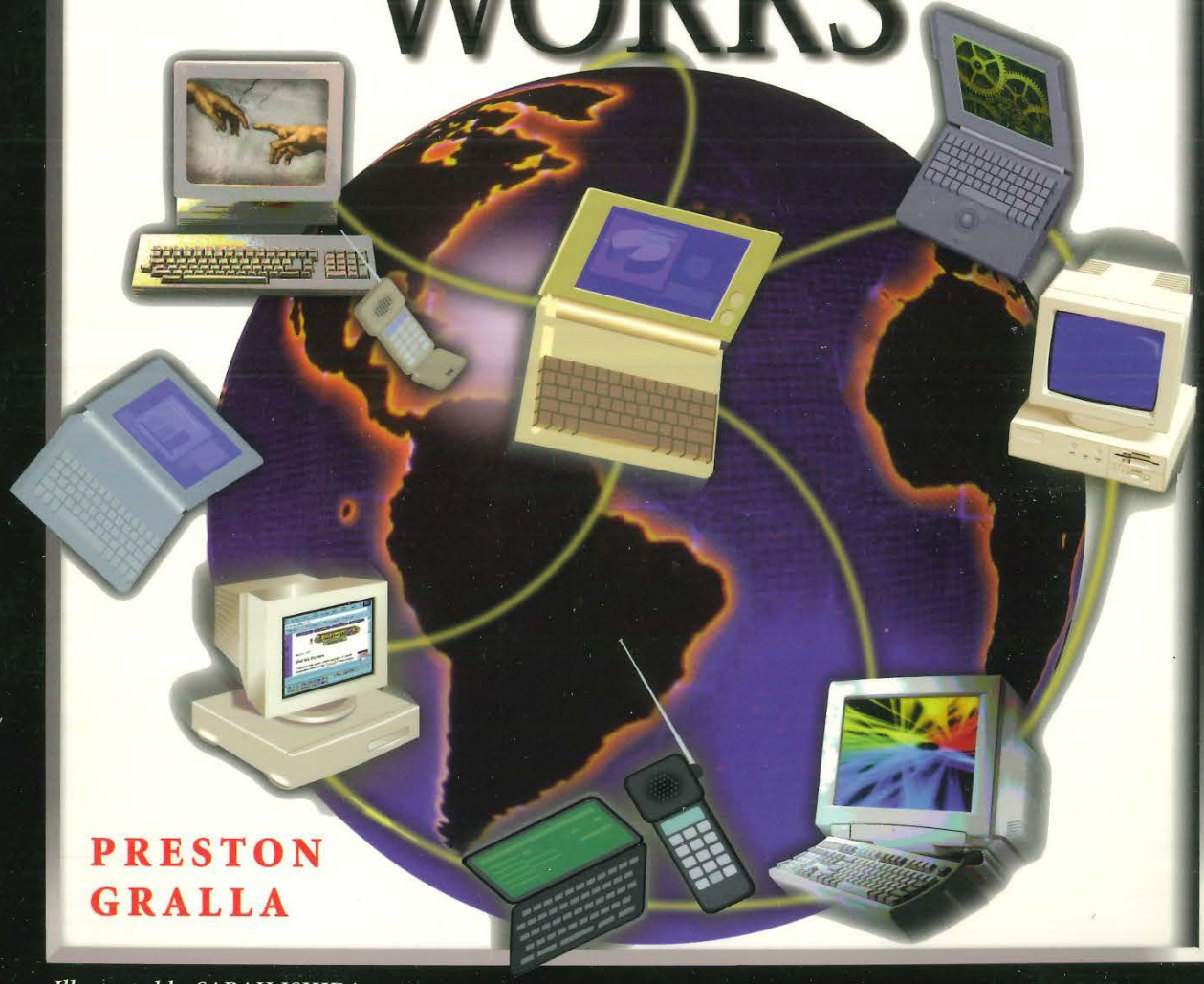
NATIONAL BESTSELLER
OVER 2 MILLION HOW IT WORKS BOOKS SOLD

HOW THE INTERNET WORKS

FOURTH EDITION



PRESTON GRALLA

Illustrated by SARAH ISHIDA, MINA REIMER, & STEPHEN ADAMS

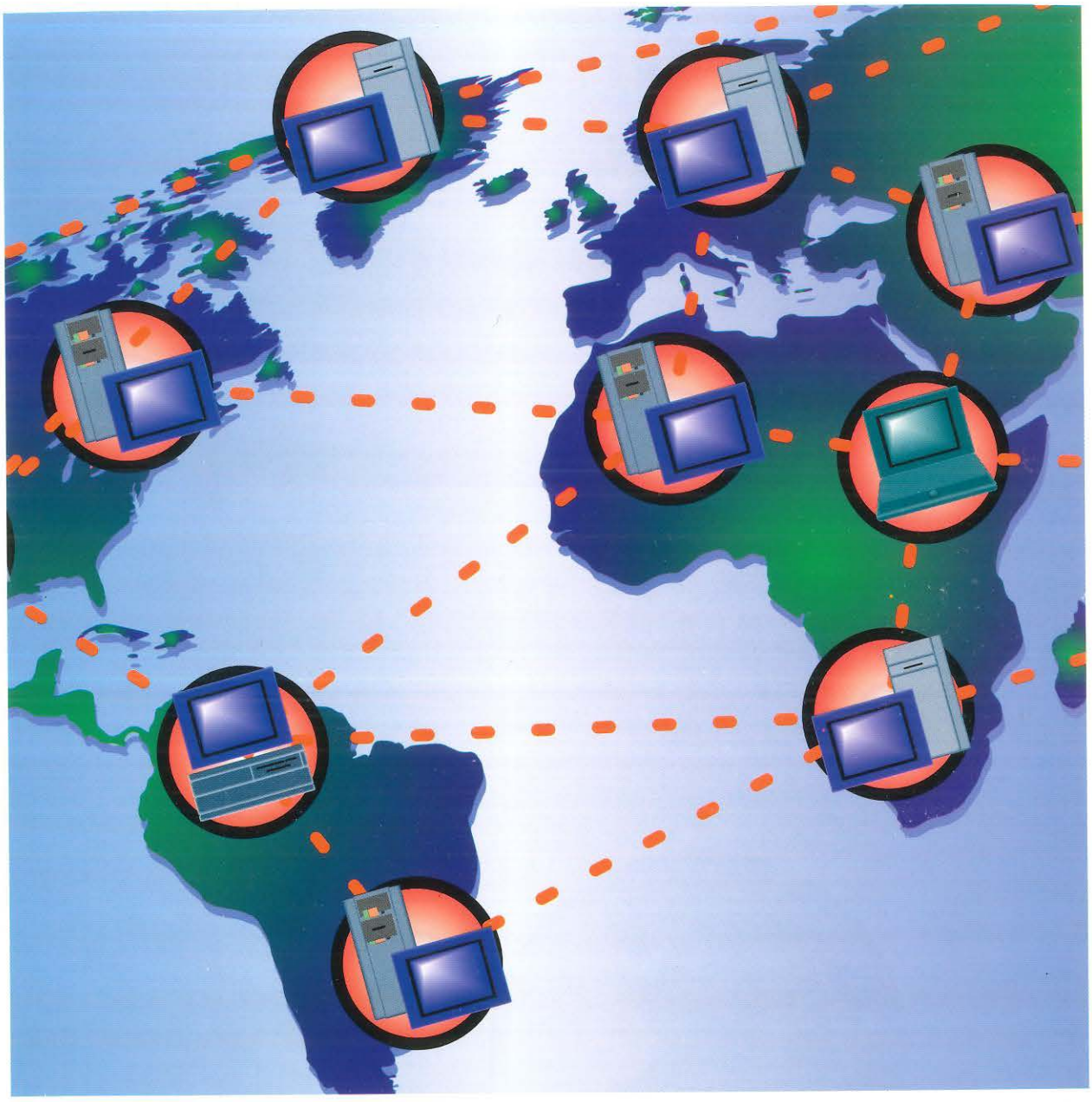
que®

VIANT-1047, Part 1 of 2

How the Internet Works

Fourth Edition





How the Internet Works

Fourth Edition

Preston Gralla

Illustrated by Sarah Ishida, Mina Reimer, and Stephen Adams



A division of Macmillan Computer Publishing, USA
201 W. 103rd Street
Indianapolis, IN 46290

Copyright© 1998 by Que® Corporation
First printing: September 1998

Executive Editor	Karen Reinisch
Acquisitions Editor	Renee Wilmeth
Development Editor	Noelle Gasco
Technical Editor	Doug Klippert
Managing Editor	Thomas F. Hayes
Project Editor	Karen A. Walsh
Copy Editors	Barbara Hacha and Julie McNamee
Illustrators	Sarah Ishida, Mina Reimer, and Stephen Adams
Book Designers	Carrie English and Bruce Lundquist
Production	Betsy Deeter, Lisa England, John Etchison, and Trina Wurst
Indexers	Sandra Henselmeier and Cheryl Jackson

All rights reserved. Printed in the United States of America. No part of this book may be used or reproduced in any form or by any means, or stored in a database or retrieval system, without prior written permission of the publisher, except in the case of brief quotations embodied in critical articles and reviews. Making copies of any part of this book for any purpose other than your own personal use is a violation of United States copyright laws. For information, address Que Corporation, 201 W. 103rd Street, Indianapolis, IN, 46290. You may reach Que's direct sales line by calling 1-800-428-5331.

Library of Congress Catalog No.: 98-85292

ISBN: 0-7897-1726-3

This book is sold as is, without warranty of any kind, either express or implied, respecting the contents of this book, including but not limited to implied warranties for the book's quality, performance, merchantability, or fitness for any particular purpose. Neither Que Corporation nor its dealers or distributors shall be liable to the purchaser or any other person or entity with respect to any liability, loss, or damage caused or alleged to have been caused directly or indirectly by this book.

00 99 98 6 5 4 3 2

Interpretation of the printing code: The rightmost double-digit number is the year of the book's printing; the rightmost single-digit number, the number of the book's printing. For example, a printing code of 98-1 shows that the first printing of the book occurred in 1998.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Que cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Introduction

Part 1: What Is the Internet? 1

Chapter 1
The Wired World of the Internet 4

Chapter 2
How Information Travels Across the Internet 8

Chapter 3
How TCP/IP Works 12

Chapter 4
Understanding Internet Addresses and Domains 16

Chapter 5
Anatomy of a Web Connection 20

Chapter 6
Internet File Types 24

Part 2: The Internet's Underlying Architecture 28

Chapter 7
How the Domain Name System Works 32

Chapter 8
How the Routers Work 36

Chapter 9
The Internet's Client/Server Architecture 40

Part 3: Connecting to the Internet 44

Chapter 10
Connecting Your Computer 48

Chapter 11
Connecting to the Internet from Online Services 52

Chapter 12
How ISDN Works 56

Chapter 13
The Internet/Television Connection 60

Chapter 14
How Network Computers (NCs) Work 66

Chapter 15
How the Digital Subscriber Line (DSL) Works 70

Part 4: Communicating on the Internet 74

Chapter 16
How Email Works 78

Chapter 17
EMail Spam and Blocking 90

Chapter 18
How Usenet Newsgroups Work 96

Chapter 19
How Internet Chat and Instant Messaging Work 102

Chapter 20
Making Phone Calls on the Internet 108

Part 5: Common Internet Tools 114

Chapter 21
How Gophers Work 118

Chapter 22
How Telnet Works 122

Chapter 23
FTP and Downloading Files 126

Chapter 24
Searching the Internet 132

Part 6: How the World Wide Web Works 138

Chapter 25
How Web Pages Work 142

Chapter 26
How Web Browsers Work 150

Chapter 27
How Markup Languages Work 158

Chapter 28
How Hypertext Works 164

Chapter 29
How URLs Work 168

Chapter 30
How Imagemaps and Interactive Forms Work 172

Chapter 31
How Web Host Servers Work 176

Chapter 32
How Web Sites Work with Databases 180

Part 7: Advanced Internet Tools 184

Chapter 33
How Push Technology Works 188

Chapter 34
How Java, ActiveX, and JavaScript Work 196

Chapter 35
How Agents Work 204

Chapter 36
How CGI Scripting Works 208

Part 8: Multimedia on the Internet 212

Chapter 37
Audio on the Internet 216

Chapter 38
Video on the Internet 220

Chapter 39
How Multicast IP and the MBone Work 228

Chapter 40
Virtual Reality on the Internet 232

Chapter 41
Animation on the Web 236

Part 9: The Internet, Intranets, and the Outside World 242

Chapter 42
How Intranets Work 246

Chapter 43
How Workgroup Software Works 250

Chapter 44
Shopping on the Internet 254

Part 10: Safeguarding the Internet 262

Chapter 45
How Firewalls Work 266

Chapter 46
How "Smurf Attacks" Can Cripple ISPs 272

Chapter 47
How Viruses Work 276

Chapter 48
How Cookies, Passports, and Web Tracking Work 280

Chapter 49
Cryptography and Privacy 288

Chapter 50
How Digital Certificates Work 292

Chapter 51
Parental Controls on the Internet 296

Index 300

THIS book, like the Internet, is a collaborative work. My name may be on the cover, but I am far from the only person involved in its creation.

Development Editor Noelle Gasco managed to pull together all the elements of a complex project, all the while keeping a cool head, improving the illustrations, and offering editorial insight and help. Renee Wilmeth, once again, helped shape this book and was instrumental in getting it off the ground.

Without Mina Reimer, Sarah Ishida, and Stephen Adams, the illustrators, there would be no book because this book is such a visual experience. And many thanks to the entire team at Que who produced the book including Karen Walsh, project editor; and Barbara Hacha, copy editor.

Thanks also have to go to the many, many people I interviewed for this book. People from Quarterdeck Corporation, Chaco Communications, Progressive Networks, White Pine Software, Microsoft, Netscape, Headspace, SurfWatch Software, WebTV, Accrue, VDONet Corporation, America Online, Yahoo!, Hilgraeve, and Nuborn Technologies are only a few of the folks who gave their time to help me understand the nitty-gritty of how various Internet technologies work. Tim Smith and Jonathan Spewak from ZDNet offered me vital help as well.

I gleaned much information from the many FAQs and similar documents widely available on the Internet. I'd like to thank the anonymous authors of those documents, whoever they are.

Many thanks go to Doug Klippert, the technical editor for the book, who did a superb job of ensuring that I always got the information right.

Finally, big thanks have to go to my wife, Lydia. She put up with those occasional glassy-eyed looks that were replies to simple questions like, "Did you leave your keys in the refrigerator again?" She also endured my extreme absent-mindedness while I was figuring out ways to explain how firewalls, ISDN, or Web robots work when I should have been concentrating on more immediate matters.

IN the course of cruising the World Wide Web and clicking on a link, have you ever wondered, "How does that work?" Or perhaps this question popped into your mind while you were transferring a file to your computer via FTP, or reading a newsgroup message, or when you first heard about technologies such as Spam, cookies, and firewalls. Maybe you've wondered how a message sent from your computer travels through the vastness of cyberspace and ends up in the right email box halfway across the world. Have you ever wanted to know how search tools find the exact piece of information you want out of the millions of pieces of information on the whole Internet? How can you listen to music and view animations while surfing the Web?

This book is designed for everyone interested in the Internet. Its guiding principle is this: No matter how much of a cyberpro you are—or how much of a novice—there's a lot you don't understand about the Internet. Here's just one small example. I have a friend who has made his living with companies involved with the Internet for many years. He's a complete cyberpro who lives and breathes the Internet. One day, he almost whispered to me, "I don't like to admit this, but I don't know what a proxy server is. How *does* it work, anyway?"

He's not alone. The Internet changes so quickly, and the technology advances so rapidly, that it can seem almost impossible to keep up with all of it. If you're like just about everyone else involved in the Internet, your questions are similar to those of my friend. You'll find your answers here.

In Part 1, I explain the underlying basics of the Internet: who runs it, how TCP/IP works, how to understand Internet addresses and domains, and similar topics.

Part 2 covers the Internet's underlying architecture. Here's where you'll find out about things such as routers and how the client/server architecture underpins virtually every aspect of the Internet.

Part 3 depicts the various ways that you can connect your computer to the Internet. Here's where to turn if you're interested in any of the following topics: how a cable modem, Digital Subscriber Line (DSL), or ISDN works; how online services connect to the Internet; how Network Computers (NCs) work; and a host of similar subjects.

Part 4 covers every aspect of Internet communications. It shows how email and newsgroups work, how IRC chat works, what email "Spam" is and what you can do to prevent it, how instant messaging works, and how you can use the Internet to make telephone calls anywhere in the world.

Part 5 shows how the common Internet tools and services work. Here's where you'll find out how gophers and Telnet work, how search engines and search tools work, and what happens when you use FTP to download a file to your computer.

Part 6 covers what has become by far the most popular part of the Internet—the World Wide Web. You'll learn virtually every aspect of how the Web works. It delves into how browsers work, how Web server software works, and how Hypertext Markup Language (HTML) works. This section also covers the ways in which the Web is becoming integrated directly into your computer, how Web pages are published and organized on a site, and every other aspect of the Web that is likely to be of interest to you.

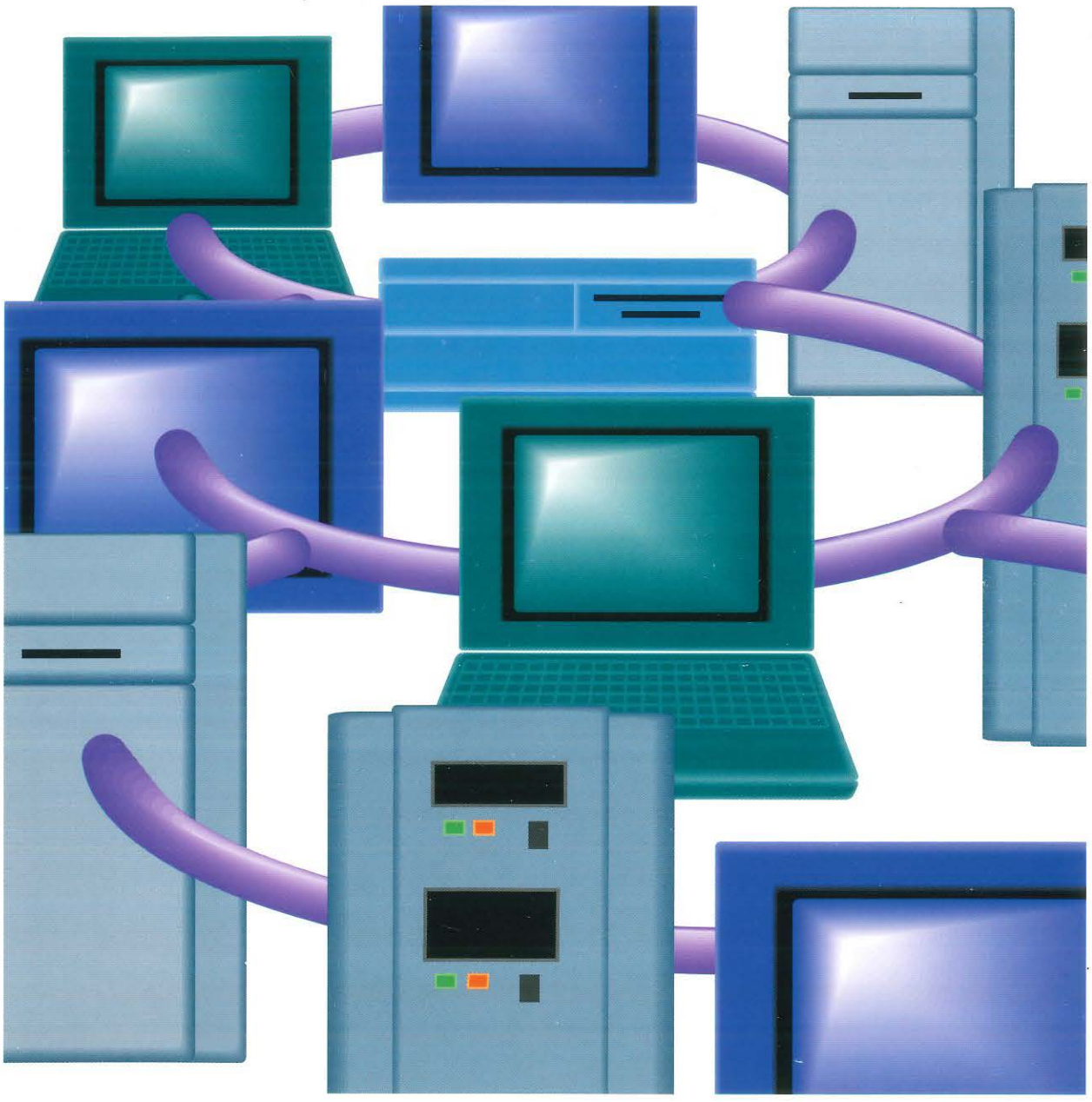
Part 7 takes a close look at advanced Internet tools. Here's the place to learn about cutting-edge Internet technology. You'll find out about how push technology works, how the Java and ActiveX programming languages work, how JavaScript works, how agents can silently do your bidding for you, and how the Common Gateway Interface (CGI) works, which is a little-known but vital part of the Internet.

Part 8 shows you how some of the most exciting parts of the Internet works—the various multimedia technologies. Whether you want to know how virtual reality or animations work, how streaming video works, how videoconferencing works, or how similar technologies work, you'll find it all here.

Part 9 covers intranets and how the Internet works with the outside world. We'll see how companies use Internet technologies to build their own private networks called *intranets.* We'll take a close look at the underlying technologies that let you shop on the Web, which account for billions of dollars a year in sales.

Finally, Part 10 covers security concerns. It explains the controversial *cookie* technology that lets Web servers put bits of information on your hard disk and use that information to track you. It shows how firewalls work, how viruses can attack your computer, and how cryptosystems allow confidential information to be sent across the Internet. It delves into how hackers can attack Internet service providers (ISPs) using so-called "smurf attacks." And it covers the issue of pornography on the Internet and shows how parental-control software can prevent children from seeing objectionable material.

So, come along and see how the vast Internet works. Even if you're a cyberpro (and especially if you're not), you'll find out a lot you never knew.



P A R T

I

WHAT IS THE INTERNET?

Chapter 1: The Wired World of the Internet
4

Chapter 2: How Information Travels Across the Internet
8

Chapter 3: How TCP/IP Works
12

Chapter 4: Understanding Internet Addresses and Domains
16

Chapter 5: Anatomy of a Web Connection
20

Chapter 6: Internet File Types
24

FOR the first time ever, the world is truly at your fingertips. From your computer you can find information about anything you can name or even imagine. You can communicate with people on the other side of the world. You can set up a teleconference, tap into the resources of powerful computers anywhere on the globe, search through the world's best libraries, and visit the world's most amazing museums. You can watch videos, listen to music, and read special multimedia magazines. You can shop for almost anything you can name. You can do all this by tapping into the largest computer network in the world—the Internet.

The Internet isn't a single network; it is a vast, globe-spanning network of networks. No single person, group, or organization runs the Internet. Instead, it's the purest form of electronic democracy. The networks communicate with each other based on certain protocols, such as the Transmission Control Protocol (TCP) and the Internet Protocol (IP). More and more networks and computers are being hooked up to the Internet every day. Tens of thousands of these networks exist, ranging from university networks to corporate local area networks to large online services such as America Online and CompuServe. Every time you tap into the Internet, your own computer becomes an extension of that network.



We'll spend the first section of this book defining the Internet. We'll also examine the architectures, protocols, and general concepts that make it all possible.

In Chapter 1, "The Wired World of the Internet," we will examine how the Internet runs. We'll look at who pays for the high-speed data backbones that carry much of the Internet's traffic, and at the organizations that make sure that standards are set for networks to follow so that the Internet can run smoothly. We'll also look at the various kinds of networks that are connected to the Internet.

Chapter 2, "How Information Travels Across the Internet," explains how information travels across the Internet and describes how hardware such as routers, repeaters, and bridges sends information among networks. It also shows how smaller networks are grouped into larger regional networks—and how those large regional networks communicate among themselves.

In Chapter 3, "How TCP/IP Works," we'll look at the Internet's basic protocols for communications and learn a little about basic Internet jargon, such as TCP/IP (short for Transmission Control Protocol and Internet Protocol). The chapter will explain how those protocols work and how special software such as Winsock enables personal computers to get onto a network originally designed for larger computers.

Chapter 4, "Understanding Internet Addresses and Domains," takes the mystery out of the Internet's often confusing addressing scheme. You'll learn about Internet domains and addresses and will even be able to make sense of them.

Chapter 5, "Anatomy of a Web Connection," covers the basics of what has become the most popular portion of the Internet—the World Wide Web. We'll look at the anatomy of a Web connection and see how hardware, such as Web servers, and software, such as Web browser client software, talk to one another and let you surf through the entire known world.

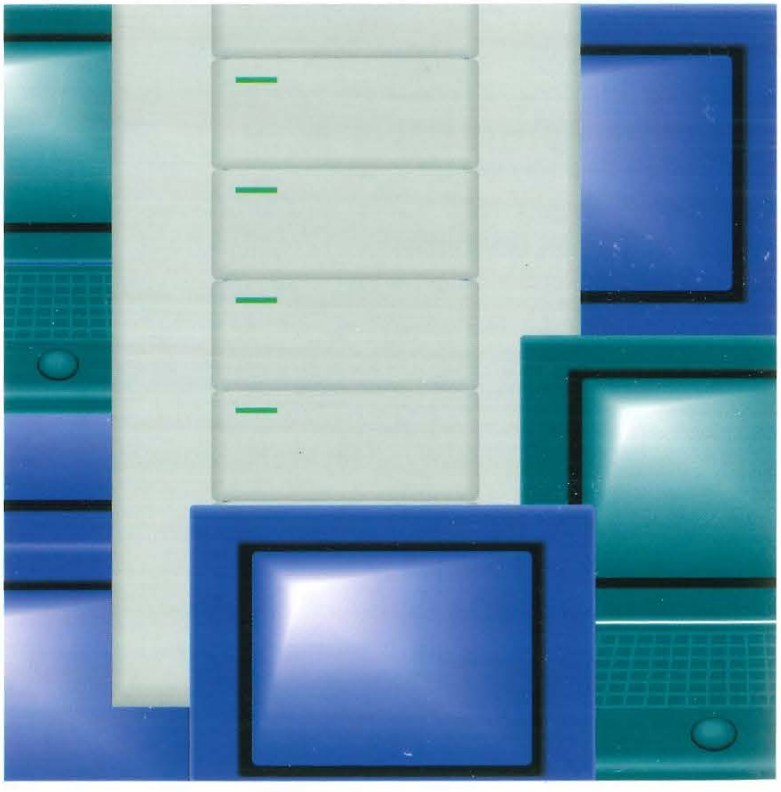
Chapter 6, "Internet File Types," will give you an understanding of the most common types of files you'll come across as you browse the Net. Compressed files, video files, graphics files—you'll learn about almost every kind of file you might encounter.

Whether you're a newbie or cyberpro, this section will teach you the basics of the Internet.



CHAPTER 1

The Wired World of the Internet



ONE of the most frequently asked questions about the Internet is: Who runs it? The truth is that no centralized management of the Internet exists. Instead, it is a collection of thousands of individual networks and organizations, each of which is run and paid for on its own. Each network cooperates with other networks to direct Internet traffic so that information can pass among them. Together, these networks and organizations make up the wired world of the Internet. For networks and computers to cooperate in this way, however, a general agreement must take place about things such as Internet procedures and standards for protocols. These procedures and standards are laid out in RFCs (requests for comment) agreed upon by Internet users and organizations.

A variety of groups guide the Internet's growth by helping to establish standards and by educating people on the proper way to use the Internet. Perhaps the most important is the Internet Society, a private, nonprofit group. The Internet Society supports the work of the Internet Activities Board (IAB), which handles much of the Internet's behind-the-scenes and architectural issues. The IAB's Internet Engineering Task Force is responsible for overseeing how the Internet's TCP/IP protocols evolve. (See Chapter 3, "How TCP/IP Works," for details on protocols.)

The World Wide Web Consortium (W3C) develops standards for the evolution of the fastest-growing part of the Internet, the World Wide Web. The W3C is an industry consortium run by the Laboratory for Computer Science at the Massachusetts Institute of Technology (MIT).

Private companies, such as `www.zdnet.com`, oversee the registering of Internet domains. These companies are overseen by boards made up of people from business, the government, and individual Internet users.

Although these organizations are important as a kind of glue for holding together the Internet, at the heart of the Internet are individual local networks. These networks can be found in private companies, universities, government agencies, and online services. They are funded separately from each other and in a variety of manners, such as fees from users, corporate support, taxes, and grants. Many Internet service providers (ISPs), which provide Internet access for individuals, have networks as well. Individuals who want to access the Internet pay ISPs a monthly connection rate, so in that sense, everyone who uses the Internet helps pay for it.

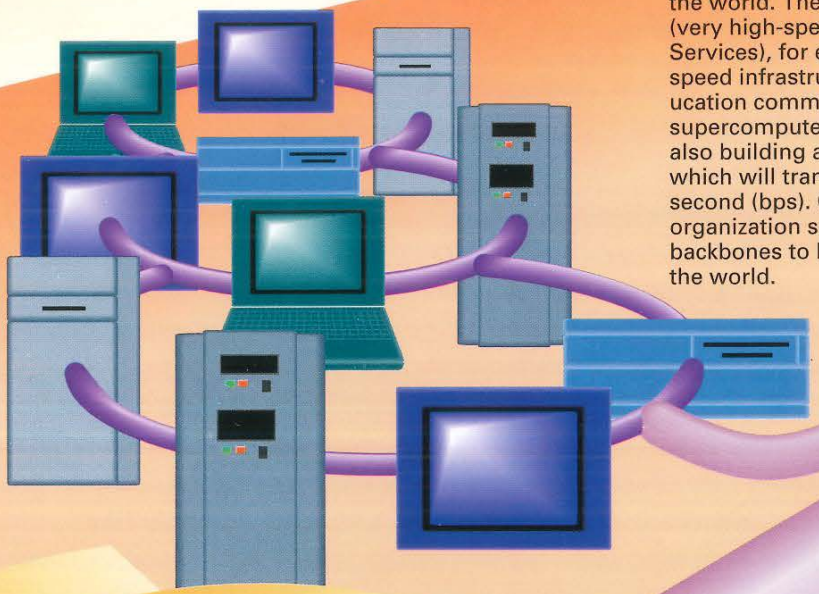
The networks are connected in a variety of ways. For efficiency's sake, local networks join in consortiums known as regional networks. A variety of leased lines connect regional and local networks. The leased lines that connect networks can be as simple as a single telephone line or as complex as a fiber-optic cable with microwave links and satellite transmissions.

Private companies who make money by selling access to their lines build backbones, which are very high-capacity lines that carry enormous amounts of Internet traffic. Government agencies, such as NASA, and large private corporations pay for some of these backbones. The National Science Foundation also pays for some backbones.

How the Internet Runs

1 Because the Internet is a loose organization of networks, no single group runs it and pays for it. Instead, many private organizations, universities, and government agencies pay for and run parts of it. They all work together in a democratic, loosely organized alliance. Private organizations range from small, homegrown networks to commercial online services, such as America Online and CompuServe, and private Internet service providers (ISPs) that sell access to the Internet.

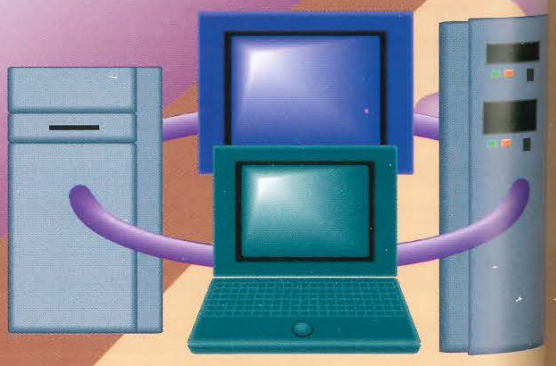
2 Through agencies such as the National Science Foundation, the federal government pays for some high-speed backbones that carry Internet traffic across the country and the world. The extremely high-speed vBNS (very high-speed Backbone Network Services), for example, provides a high-speed infrastructure for the research and education community by linking together supercomputer centers. The government is also building an even faster backbone, which will transmit data at 9.6 billion bits per second (bps). Often, a large corporation or organization such as NASA will provide backbones to link sites across the country or the world.



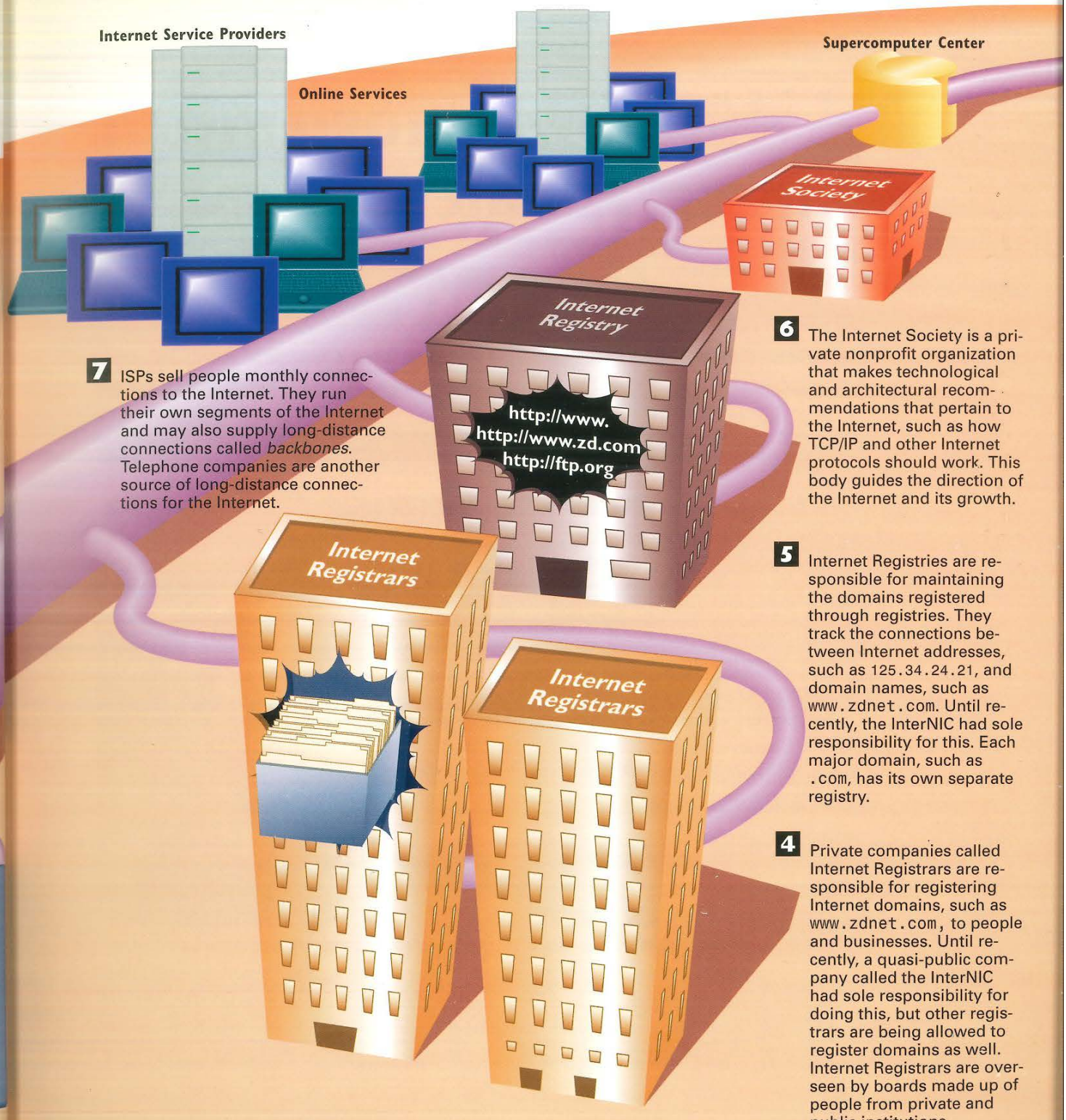
Regional Network

vBNS Backbone

3 Regional networks provide and maintain Internet access within a geographic area. Regional nets may consist of smaller networks and organizations within the area that have banded together to provide better service.



Supercomputer Center



7 ISPs sell people monthly connections to the Internet. They run their own segments of the Internet and may also supply long-distance connections called *backbones*. Telephone companies are another source of long-distance connections for the Internet.

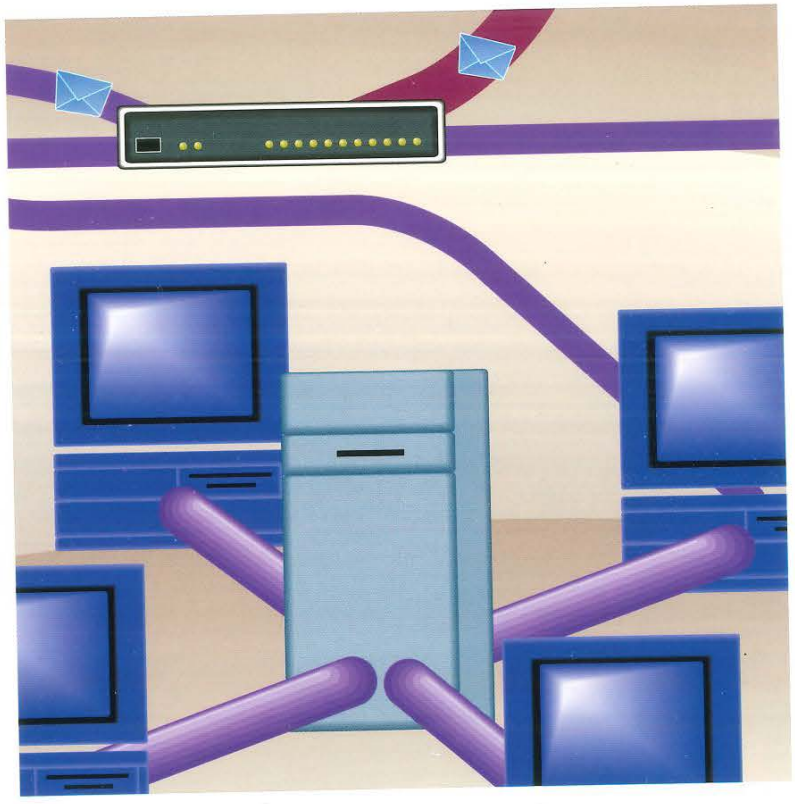
6 The Internet Society is a private nonprofit organization that makes technological and architectural recommendations that pertain to the Internet, such as how TCP/IP and other Internet protocols should work. This body guides the direction of the Internet and its growth.

5 Internet Registries are responsible for maintaining the domains registered through registries. They track the connections between Internet addresses, such as `125.34.24.21`, and domain names, such as `www.zdnet.com`. Until recently, the InterNIC had sole responsibility for this. Each major domain, such as `.com`, has its own separate registry.

4 Private companies called Internet Registrars are responsible for registering Internet domains, such as `www.zdnet.com`, to people and businesses. Until recently, a quasi-public company called the InterNIC had sole responsibility for doing this, but other registrars are being allowed to register domains as well. Internet Registrars are overseen by boards made up of people from private and public institutions.

CHAPTER 2

How Information Travels Across the Internet



YOU may take for granted that when you send a piece of information across the Internet, it will always reach its intended destination. However, the process of sending that information is remarkably complex.

When you send information across the Internet, the Transmission Control Protocol (TCP) first breaks it up into packets. Your computer sends those packets to your local network, Internet service provider (ISP), or online service. From there, the packets travel through many levels of networks, computers, and communications lines before they reach their final destination, which may be across town or around the world. A variety of hardware processes those packets and routes them to their proper destinations. This hardware is designed to transmit data between networks and makes up much of the glue that holds the Internet together. Five of the most important pieces of hardware are hubs, bridges, gateways, repeaters, and routers.

Hubs are important because they link groups of computers to one another and let computers communicate with each other. *Bridges* link local area networks (LANs) with one another. They enable data destined for another LAN to be sent there, while keeping local data inside its own network. *Gateways* are similar to bridges, but they also translate data from one kind of network to another.

When data travels across the Internet, it often crosses great distances, which can be a problem because the signal sending the data can weaken over the distance. To solve the problem, *repeaters* amplify the data at intervals so the signal doesn't weaken.

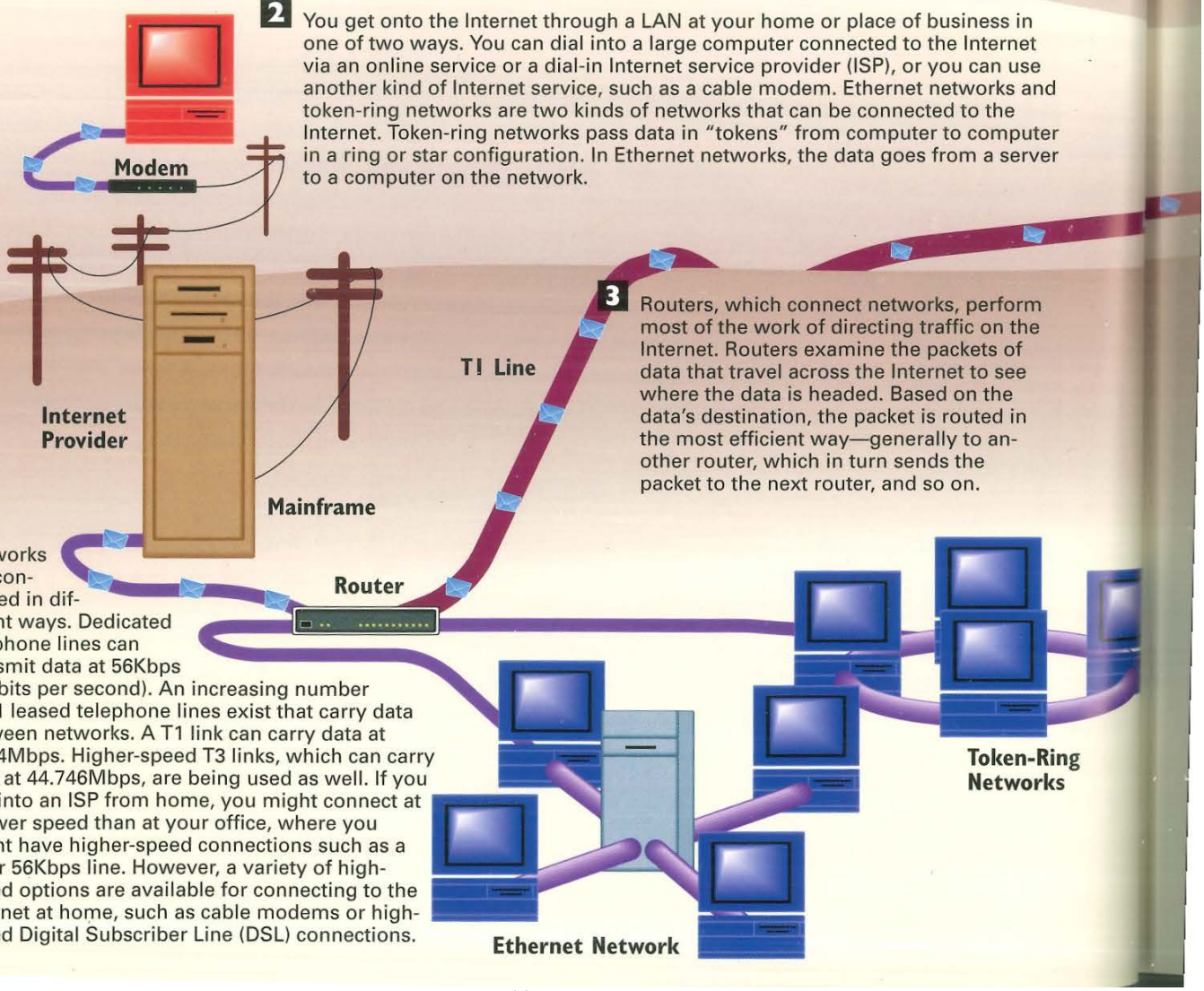
Routers play a key role in managing Internet traffic. Their job is to make sure the packets always arrive at the proper destination. If data is being transferred among computers that are on the same LAN, routers often aren't needed because the network itself can handle its internal traffic. Routers come into play when the data is sent between two different networks. Routers examine packets to determine their destination. They take into account the volume of activity on the Internet, and they send the packets to another router—one that is closer to the packet's final destination.

Midlevel networks hook LANs together using high-speed telephone lines, ethernet, and microwave links. A regional network is a midlevel network in a geographic area. A wide area network (WAN) is another kind of midlevel network. A WAN consists of an organization with many networked sites linked together.

When a packet travels from a computer on a LAN in a midlevel network to a computer somewhere else on the midlevel network, a router (or a series of routers) sends the packets to their proper destination. However, if the destination lies outside the midlevel network, the packets are sent to a NAP (Network Access Point), where they are sent across the country or the world on a backbone. High-speed backbones such as the vBNS (very high-speed Backbone Network Services) can transmit enormous amounts of data—155 megabits (millions of bits) per second (Mbps). Even faster backbones are being built that will transmit data at an astonishing 9.6 billion bits per second.

Linking Networks to the Internet

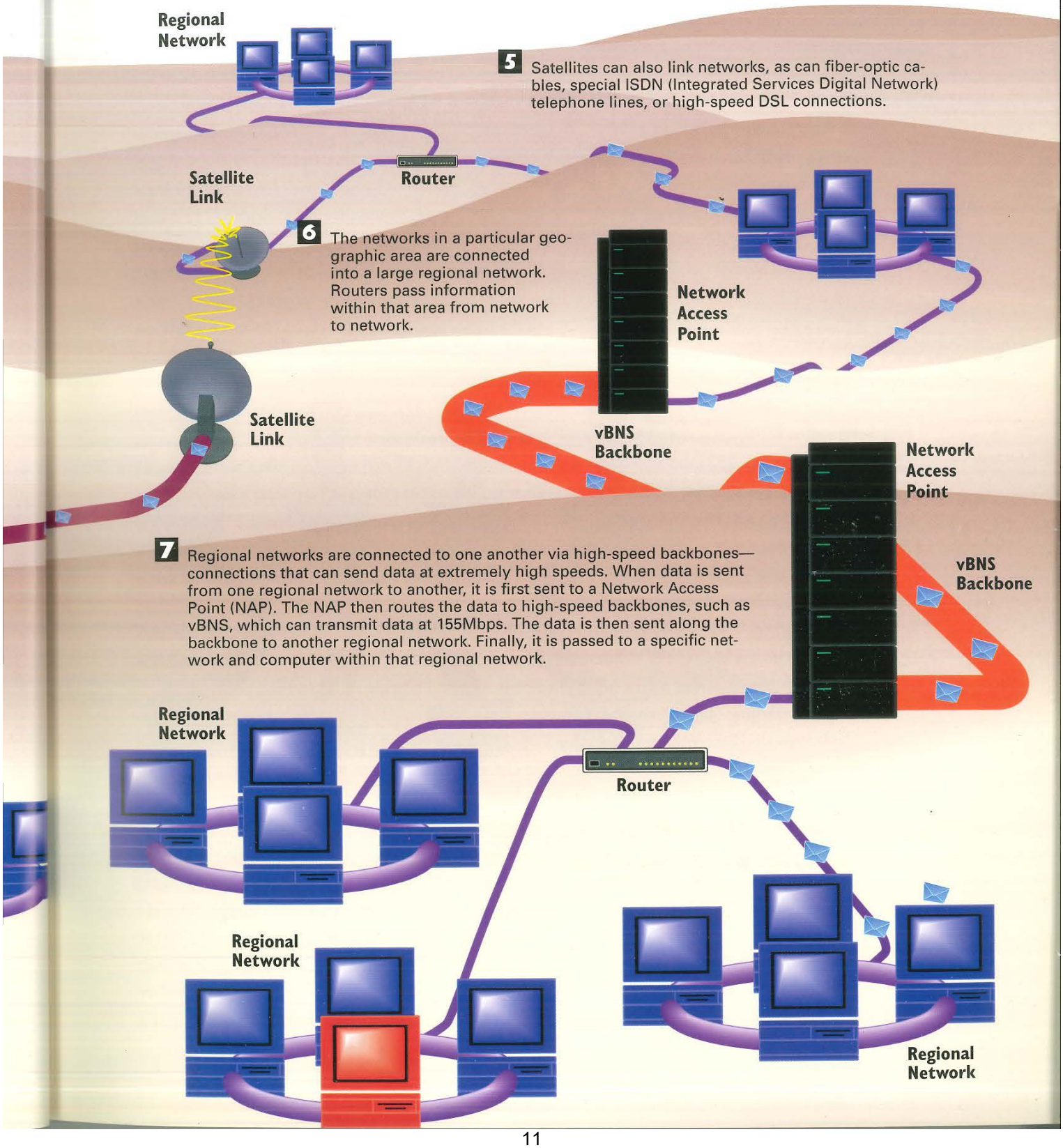
1 The Internet comprises networks that are attached to one another via pathways that facilitate the exchange of information, data, and files. Being connected to the Internet means having access to these pathways. Your computer can send packets of data over these pathways to any other computer connected to the Internet.



2 You get onto the Internet through a LAN at your home or place of business in one of two ways. You can dial into a large computer connected to the Internet via an online service or a dial-in Internet service provider (ISP), or you can use another kind of Internet service, such as a cable modem. Ethernet networks and token-ring networks are two kinds of networks that can be connected to the Internet. Token-ring networks pass data in "tokens" from computer to computer in a ring or star configuration. In Ethernet networks, the data goes from a server to a computer on the network.

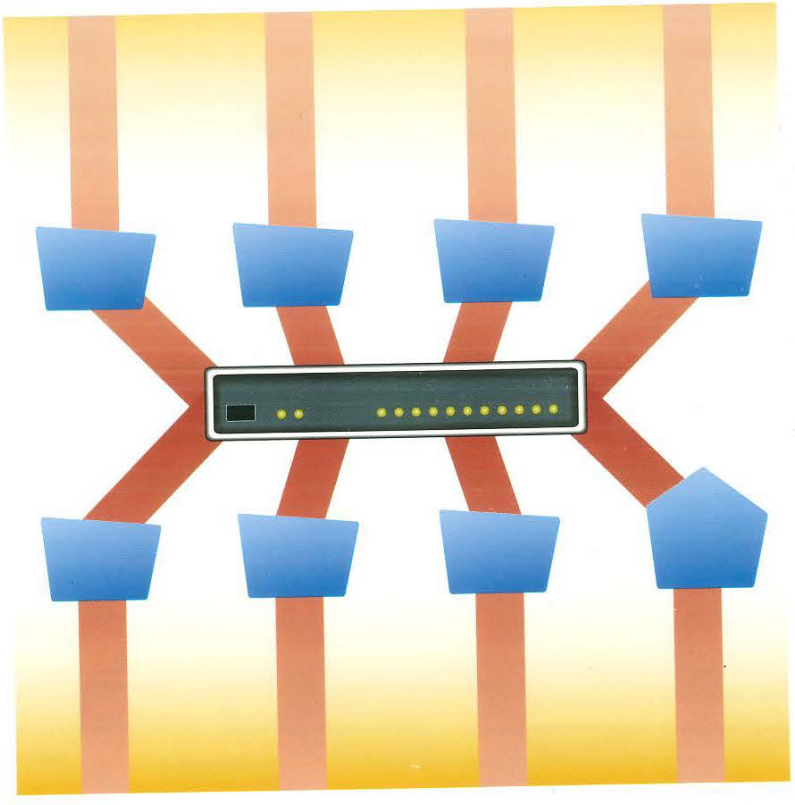
3 Routers, which connect networks, perform most of the work of directing traffic on the Internet. Routers examine the packets of data that travel across the Internet to see where the data is headed. Based on the data's destination, the packet is routed in the most efficient way—generally to another router, which in turn sends the packet to the next router, and so on.

4 Networks are connected in different ways. Dedicated telephone lines can transmit data at 56Kbps (kilobits per second). An increasing number of T1 leased telephone lines exist that carry data between networks. A T1 link can carry data at 1.544Mbps. Higher-speed T3 links, which can carry data at 44.746Mbps, are being used as well. If you dial into an ISP from home, you might connect at a lower speed than at your office, where you might have higher-speed connections such as a T1 or 56Kbps line. However, a variety of high-speed options are available for connecting to the Internet at home, such as cable modems or high-speed Digital Subscriber Line (DSL) connections.


Regional Network
5 Satellites can also link networks, as can fiber-optic cables, special ISDN (Integrated Services Digital Network) telephone lines, or high-speed DSL connections.
Satellite Link
Router
6 The networks in a particular geographic area are connected into a large regional network. Routers pass information within that area from network to network.
Network Access Point
Satellite Link
vBNS Backbone
Network Access Point
7 Regional networks are connected to one another via high-speed backbones—connections that can send data at extremely high speeds. When data is sent from one regional network to another, it is first sent to a Network Access Point (NAP). The NAP then routes the data to high-speed backbones, such as vBNS, which can transmit data at 155Mbps. The data is then sent along the backbone to another regional network. Finally, it is passed to a specific network and computer within that regional network.
vBNS Backbone
Regional Network
Router
Regional Network
Regional Network

CHAPTER 3

How TCP/IP Works



A seemingly simple set of ideas makes it possible for computers and networks all over the world to share information and messages on the Internet: Break up every piece of information and message into pieces called *packets*, deliver those packets to the proper destinations, and then reassemble the packets into their original form after they've been delivered so the receiving computer can view and use them. That's the job of the two most important communications protocols on the Internet—the Transmission Control Protocol (TCP) and the Internet Protocol (IP). They are frequently referred to as TCP/IP. TCP breaks down and reassembles the packets, whereas IP is responsible for making sure the packets are sent to the right destination.

TCP/IP is used because the Internet is what is known as a packet-switched network. In a *packet-switched network*, there is no single, unbroken connection between sender and receiver. Instead, when information is sent, it is broken into small packets, sent over many different routes at the same time, and then reassembled at the receiving end. By contrast, the telephone system is a circuit-switched network. In a *circuit-switched network*, once a connection is made (as with a telephone call, for example), that part of the network is dedicated only to that single connection.

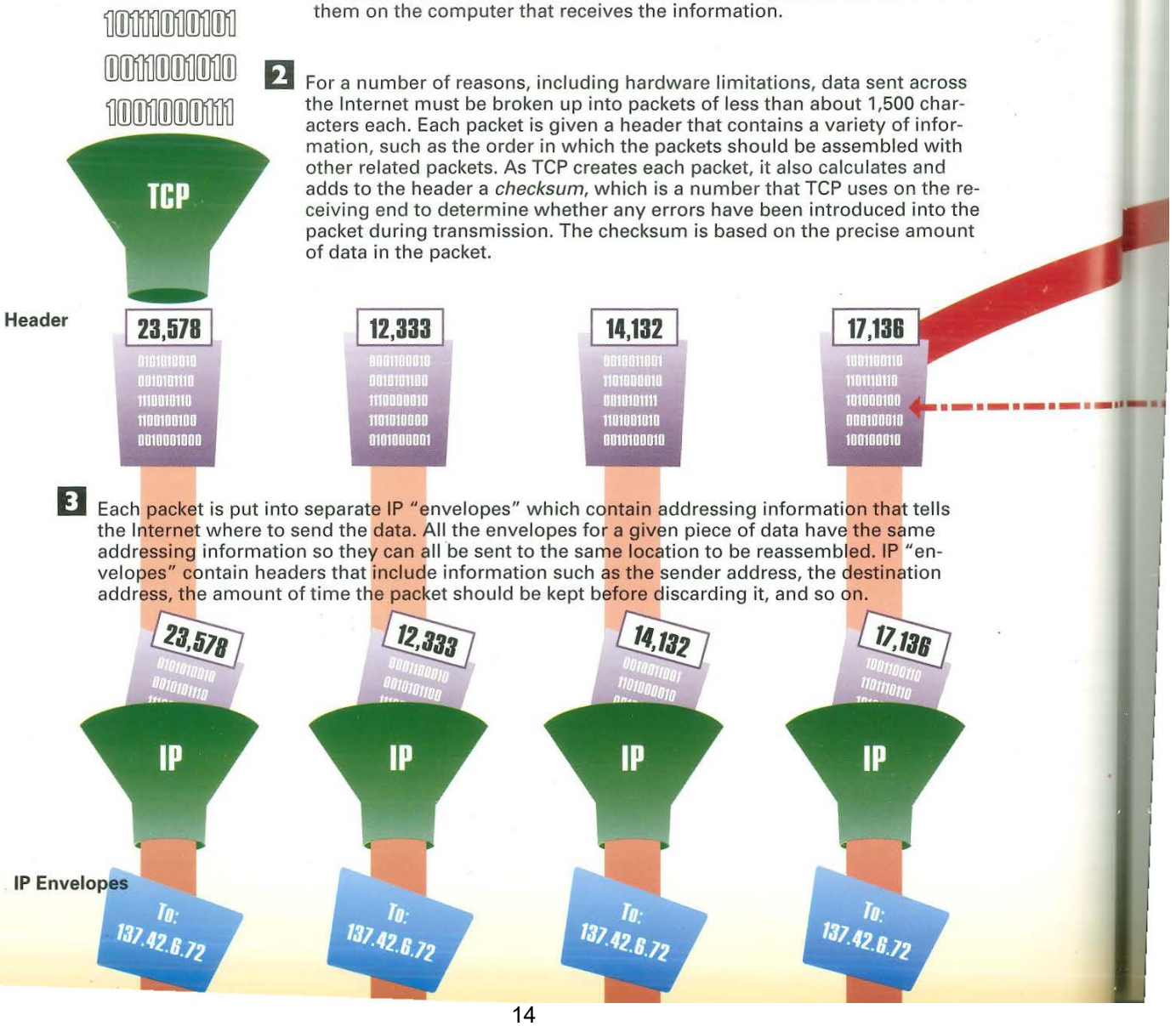
In order for personal computers to take full advantage of the Internet, they need to use special software that understands and interprets the Internet's TCP/IP protocols. This software is referred to as a *socket* or a *TCP/IP stack*. For PCs, the required software is called Winsock. There are many different versions of Winsock available for PCs. For Macintoshes, the software is called MacTCP. In both cases, this software serves as an intermediary between the Internet and the personal computer. Personal computers can take advantage of the simplest, most rudimentary parts of the Internet without using Winsock or MacTCP. For full access to the Internet, TCP/IP stacks are necessary.

A computer can be connected to a LAN (local area network) with a network card. In order to communicate with the network, the network card requires a hardware driver—software that mediates between the network and the network card. If a computer is not physically connected to a LAN with a network card, it can instead connect to the Internet by dialing in and using a modem. The computer will still need a TCP/IP stack in order to use the TCP/IP protocols. However, it won't require a network card or a hardware driver. Instead, the computer needs to use one of two software protocols: either SLIP (Serial Line Internet Protocol) or PPP (Point-to-Point Protocol).

SLIP and PPP are designed for computers connected to the Internet over a serial connection via a modem. In general, the newer PPP provides a more error-free connection than does the older SLIP. Computers can also dial into the Internet without using TCP/IP stacks, SLIP, or PPP. However, they won't be able to tap into the full power of the Internet, most specifically, the World Wide Web. To make full use of the Web, you need a computer that is part of the Internet, which requires a TCP/IP stack. The SLIP or PPP must then be used when dialing into an ISP for Web access.

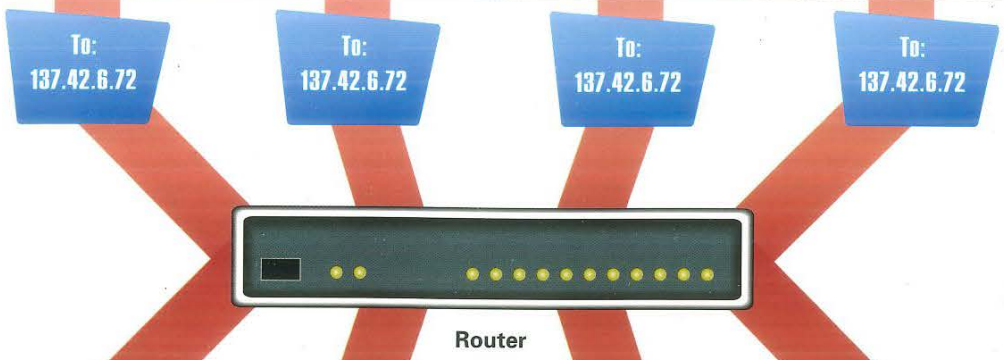
How TCP/IP Works

1 The Internet is a packet-switched network, which means that when you send information across the Internet from your computer to another computer, the data is broken into small packets. A series of switches called routers send each packet across the Net individually. Once all the packets arrive at the receiving computer, they are recombined into their original, unified form. Two protocols do the work of breaking the data into packets, routing the packets across the Internet, and then recombining them on the other end: The Internet Protocol (IP), which routes the data, and the Transmission Control Protocol (TCP), which breaks the data into packets and recombines them on the computer that receives the information.

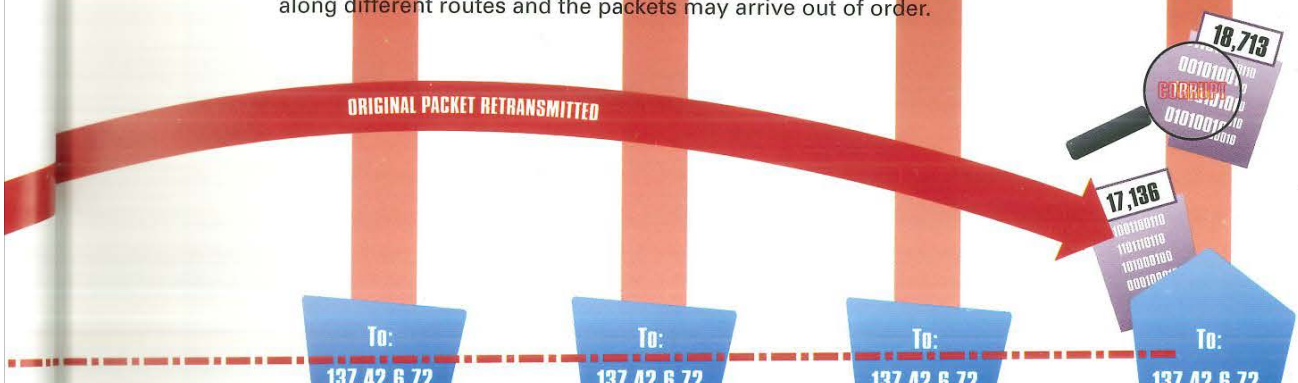


2 For a number of reasons, including hardware limitations, data sent across the Internet must be broken up into packets of less than about 1,500 characters each. Each packet is given a header that contains a variety of information, such as the order in which the packets should be assembled with other related packets. As TCP creates each packet, it also calculates and adds to the header a *checksum*, which is a number that TCP uses on the receiving end to determine whether any errors have been introduced into the packet during transmission. The checksum is based on the precise amount of data in the packet.

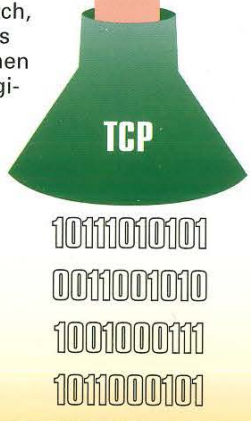
3 Each packet is put into separate IP "envelopes" which contain addressing information that tells the Internet where to send the data. All the envelopes for a given piece of data have the same addressing information so they can all be sent to the same location to be reassembled. IP "envelopes" contain headers that include information such as the sender address, the destination address, the amount of time the packet should be kept before discarding it, and so on.



4 As the packets are sent across the Internet, routers along the way examine the IP envelopes and look at their addresses. These routers determine the most efficient path for sending each packet to the next router closest to its final destination. After traveling through a series of routers, the packets arrive. Because the traffic load on the Internet changes constantly, the packets may be sent along different routes and the packets may arrive out of order.



5 As the packets arrive at their destination, TCP calculates a checksum for each packet. It then compares this checksum with the checksum that has been sent in the packet. If the checksums don't match, TCP knows that the data in the packet has been corrupted during transmission. It then discards the packet and asks that the original packet be retransmitted.



6 When all the non-corrupt packets are received by the computer to which the information is being sent, TCP assembles them into their original, unified form.

CHAPTER

4 Understanding Internet Addresses and Domains



TO do just about anything on the Internet, and especially to send email, you'll need to understand Internet addresses. The Internet Protocol (IP) uses Internet address information to deliver mail and other data from computer to computer. Every IP address on the Internet is actually a series of four numbers separated by periods (called dots), such as `163.52.128.72`. It would be difficult, if not impossible, for you to remember numeric addresses when you wanted to get in touch with someone. Also, numeric IP addresses sometimes change. This means not only would it be impossible for people to keep track of the original numeric addresses, but it would also be impossible to know every time those numeric addresses were altered.

Sun Microsystems developed the Domain Name System (DNS) in the early 1980s as an easier way to keep track of addresses. It has been the addressing system on the Internet ever since. The DNS establishes a hierarchy of *domains*, which are groups of computers on the Internet. The DNS gives each computer on the net an Internet address, or domain name, using easily recognizable letters and words instead of numbers. The domains at the top level of the hierarchy maintain lists and addresses of the domains just beneath them. Those subordinate domains have similar responsibilities for the domains just beneath them, and so on. In this way, every computer on the net gets an Internet address. The DNS also helps Internet computers send email to the proper destination by converting the textual Internet address into its IP numeric equivalent.

An Internet address is made up of two major parts separated by an @ (at) sign. One example is

```
pgralla@ziffdavis.com
```

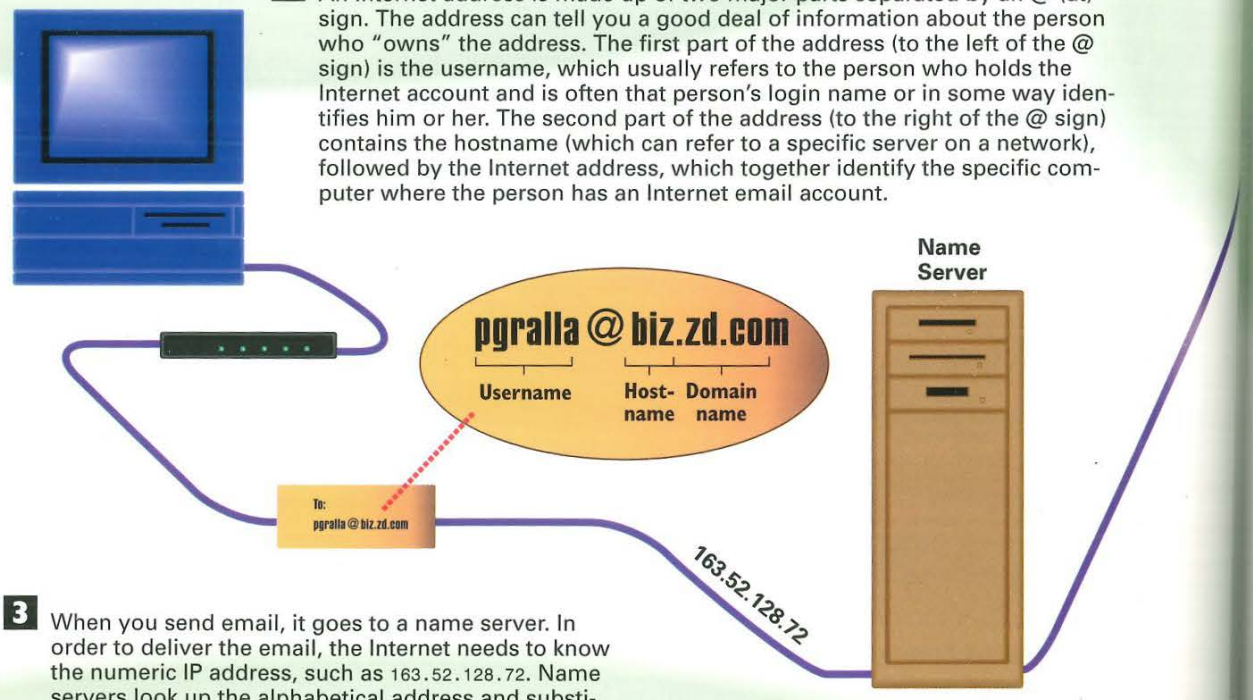
The first part of the address—to the left of the @ sign—is the user name, which usually refers to the person who holds the Internet account and is often that person's login name. The second part of the address, to the right of the @ sign, is the host name or domain name, which identifies the specific computer where the person has an Internet email account. If the email account is located at the person's work, this host or domain name may reflect the name of the company, such as `pgralla@ziffdavis.com`. Otherwise, it will probably reflect where the person has an Internet account—either with an Internet service provider (ISP) or online service, such as `jjcale@tic.net` or `jonejoan@aol.com`.

The rightmost portion of the domain section of the address identifies the largest domain and kind of organization where the person has his or her address. Common domains in the United States are `.com` for commercial; `.edu` for education; `.gov` for government; `.mil` for military; `.net` for network (most commonly used by ISPs, but also for companies and groups concerned with the organization of the Internet); and `.org` for organization. A variety of plans have been proposed for seven or more domains to be added, such as `.web` for Web. Outside the United States, only two letters are used to identify the domains, such as `.au` for Australia; `.ca` for Canada; `.uk` for United Kingdom; and `.fr` for France.

How Internet Addresses and Domains Work

1 The Internet Protocol (IP) delivers mail based on the specific email address. This address is expressed as four numbers, separated by periods (called dots), such as 163.52.128.72. However, because it would be difficult to remember such complex addresses, you can instead use Internet addresses made up of words and letters. Computers called *domain name servers* translate the alphabetical address into a numerical address, so email can be sent to the proper location.

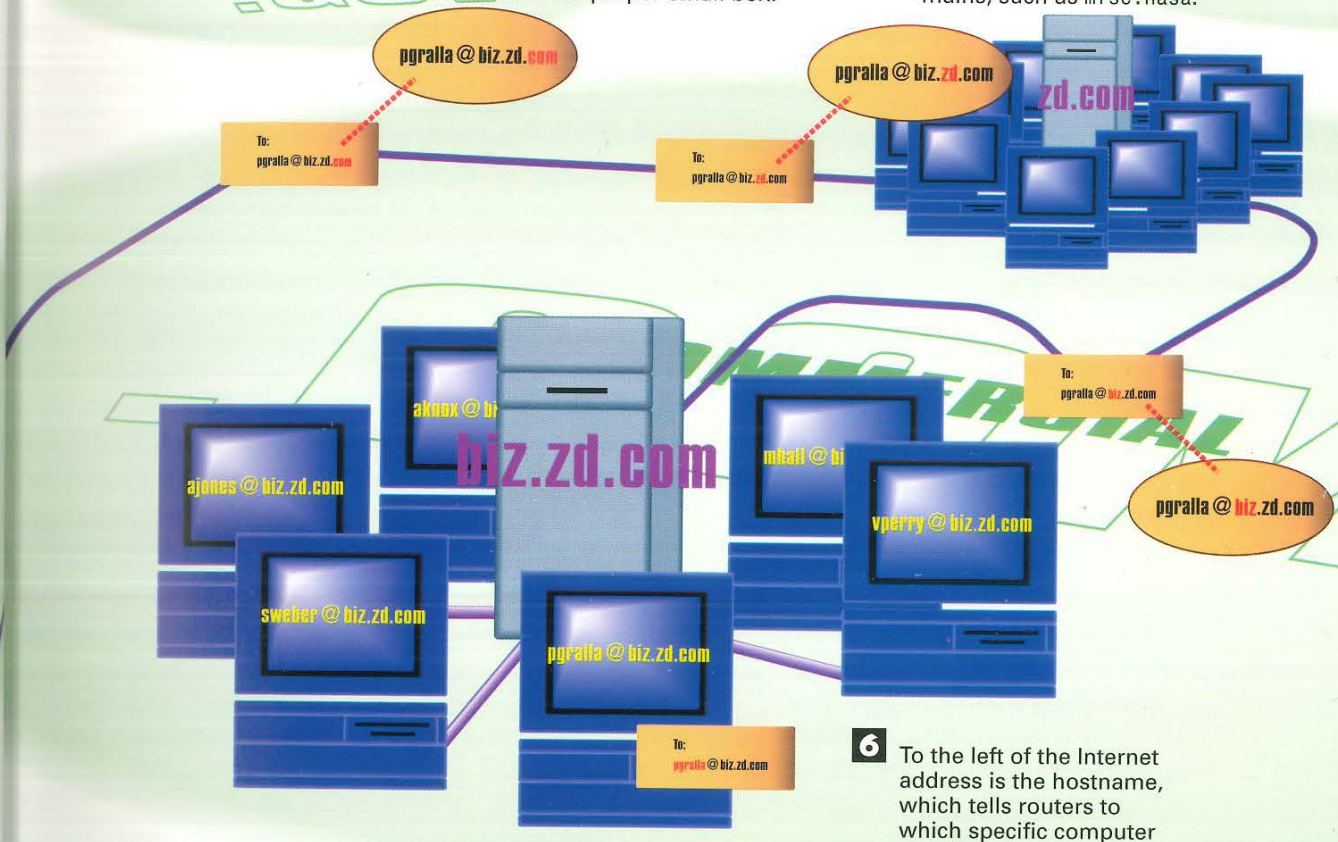
2 An Internet address is made up of two major parts separated by an @ (at) sign. The address can tell you a good deal of information about the person who "owns" the address. The first part of the address (to the left of the @ sign) is the username, which usually refers to the person who holds the Internet account and is often that person's login name or in some way identifies him or her. The second part of the address (to the right of the @ sign) contains the hostname (which can refer to a specific server on a network), followed by the Internet address, which together identify the specific computer where the person has an Internet email account.



3 When you send email, it goes to a name server. In order to deliver the email, the Internet needs to know the numeric IP address, such as 163.52.128.72. Name servers look up the alphabetical address and substitute the numeric IP address for it so the email can be delivered properly.

4 The Domain Name System (DNS) divides the Internet into understandable groups, or domains. Note the portion of the domain section at the far right of the address. It identifies the largest domain names and kind of organization where the person's address resides. The domain names and host-names identify the host computer where the Internet should deliver the email. The receiving host computer looks at the username and delivers the mail to the proper email box.

5 To the left of the largest domain is specific information about the organization, which tells routers to which network the email should be sent. It can be a single Internet address, such as `zd` (for Ziff-Davis), or it can be a group of domains and subdo-mains, such as `mfsc.nasa`.

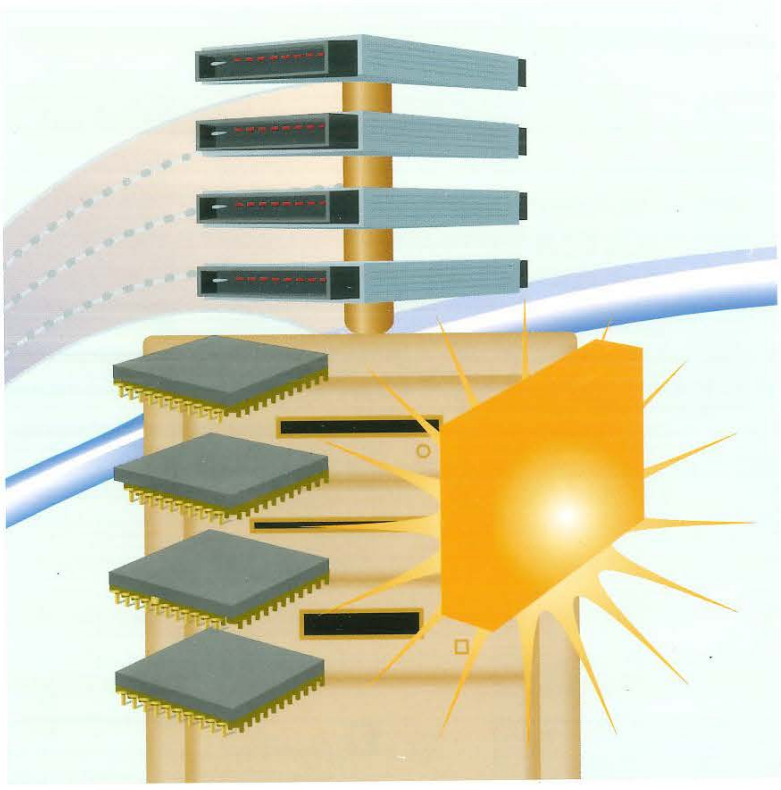


6 To the left of the Internet address is the hostname, which tells routers to which specific computer within the domain the email should be delivered.

7 The domain and hostnames tell the Internet to which computer the email should be delivered. The receiving host computer looks at the username and delivers the mail to the proper email box.

CHAPTER

5 Anatomy of a Web Connection



A typical Web connection consists of four basic parts: the desktop, client computer, or similar device; the Internet access provider; the host or server computer at the other end of the connection; and the communications networks that link together the first three components. The desktop is the command and control station from which you direct your navigation of the Web. If you're getting onto the Internet via a personal computer and a modem, at the starting point of the connection, you need a few basic hardware components—the personal computer, a fast modem and a telephone line, Web "browser" software, and an integrated communications program. The browser sends instructions to the communications program to contact a specific computer on the Web and retrieve a specific Web document or page. The communications program dials the modem, which enables the browser to send the request using TCP/IP. To reach the Internet, most desktop computers call an intermediary computer at a company that serves as an Internet access provider, also known as the Internet service provider, or ISP. The ISP acts as a conduit for individuals to dial in before getting patched through to the Internet backbone.

If you're connecting from work, your connection might look somewhat different. A personal computer on a network can share a modem with other computers on the network, for example, or the business may be directly connected to the Internet via a LAN (local area network) or other connection. In the latter case, an ISP wouldn't be needed to get onto the Web. You can get onto the Internet at home in a variety of ways, as well, without using a traditional modem—you can use a cable modem or not even use a computer at all; instead, you can go onto the Internet via your television set with a service such as WebTV. And even personal digital assistants (PDAs)—little handheld digital devices—can get you onto the Web these days.

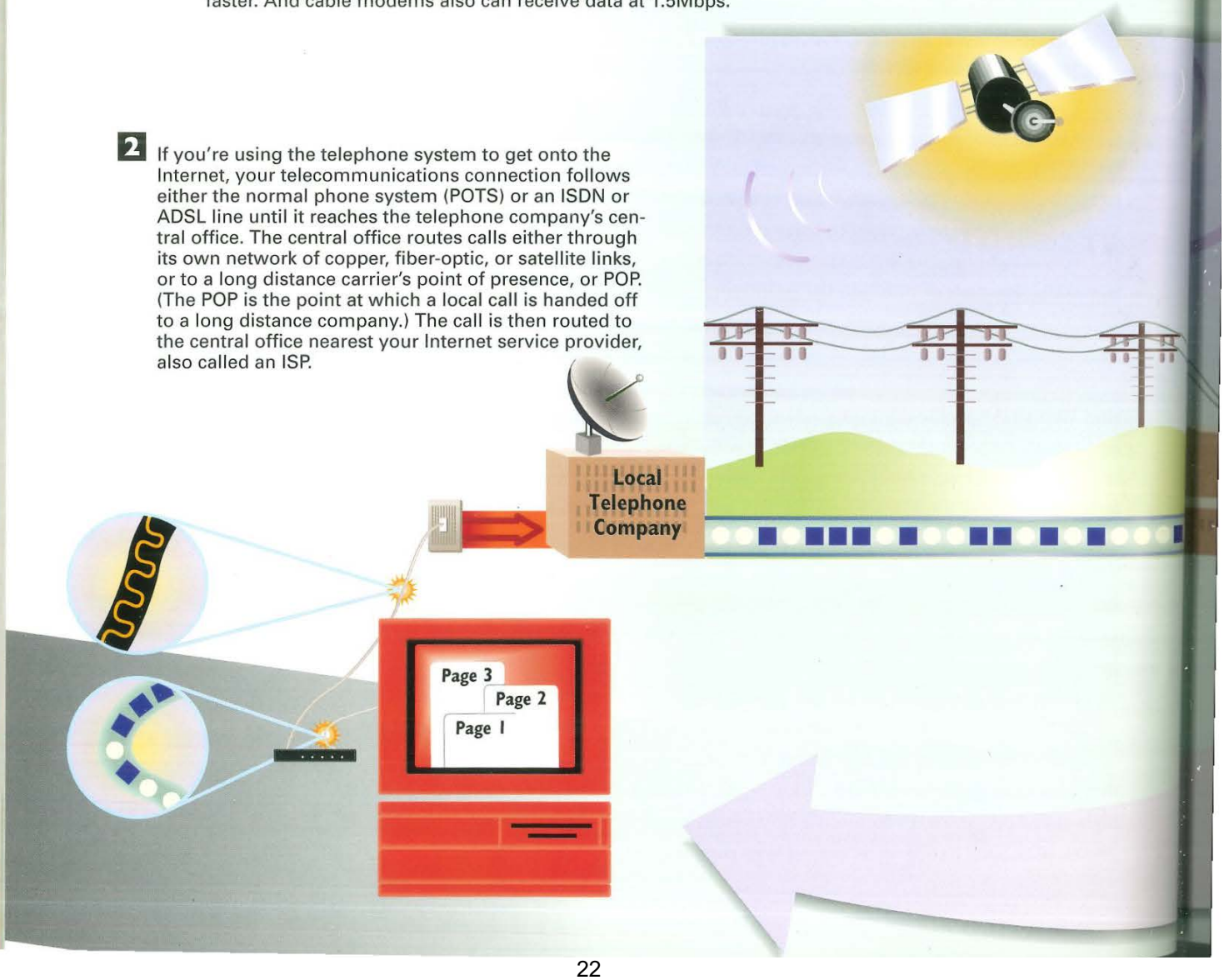
After you're connected to the Internet and you want to get to a Web site, you'll need to know that Web site's specific domain, such as `zdnet.com`. After you type in that domain, preceded by `www.`, you'll be sent to the site. You won't see it, but when you type in those letters, some nearly instantaneous electronic wizardry takes place. The computers that run the Internet can't understand letters and names such as `zdnet.com`, and so the plain English name is changed to an Internet IP address (such as `205.181.112.65`) before you're sent to the site. Names are "resolved" to IP addresses using the Domain Name System, which is covered in Chapter 4, "Understanding Internet Addresses and Domains." By the way, anyone can own a domain on the Internet. All an individual needs to do is pay a small fee to an Internet registrar, and if the desired name isn't already taken, he or she can own it. To set up a site with that domain name, the new owner will need to get a company such as an Internet service provider (ISP) to host the site.

Host computers spread throughout the Internet house Web pages. Hosts differ from desktop computers in that they can handle multiple telecommunications connections at one time. Usually, they also have gigabytes of hard-disk storage, considerable random access memory (RAM), and a high-speed processor. In fact, some host systems may actually be several computers linked together, with each handling incoming Web page requests.

How Web Connections Work

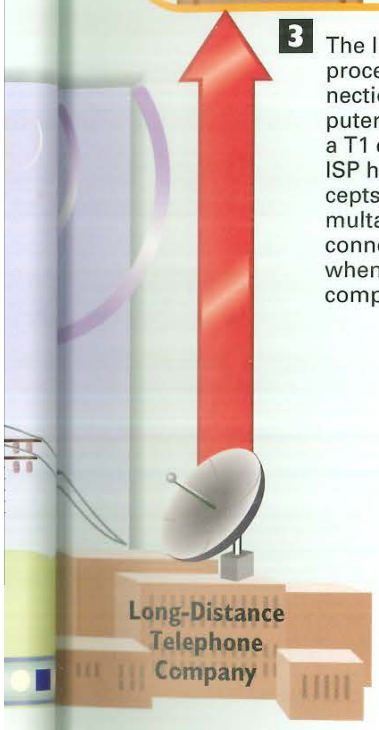
1 A fast connection is essential for Web cruising. You may use a 28.8Kbps (kilobits per second) or faster modem that converts digital information from the computer into the analog signals transmitted by ordinary phone lines (sometimes called POTS, or plain old telephone service). ISDN lines that transfer digital data at speeds up to 128Kbps are available in some residential areas. Faster still are ADSL (asynchronous digital subscriber line) lines that transmit and receive digital data at 1.5Mbps or faster. And cable modems also can receive data at 1.5Mbps.

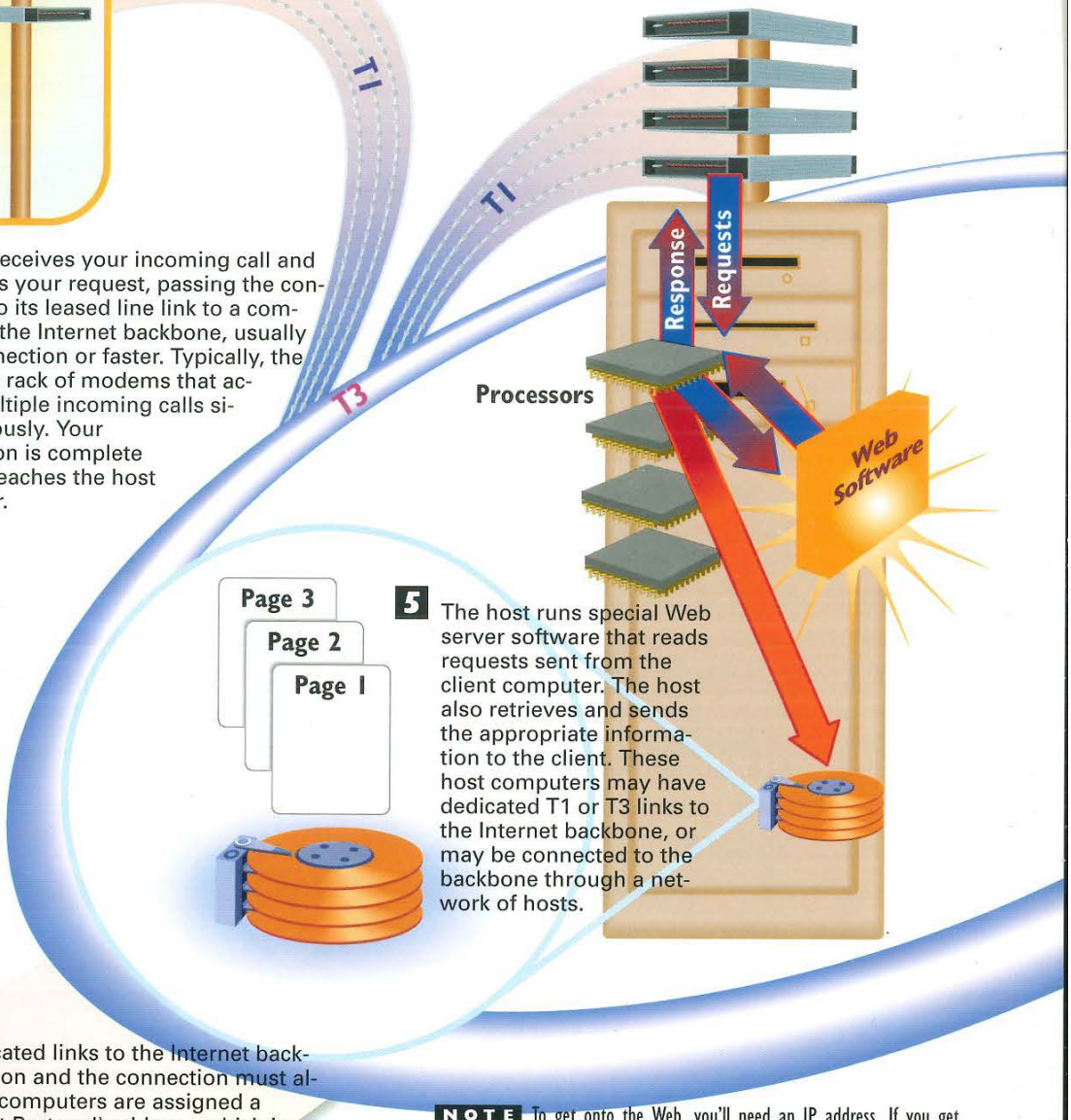
2 If you're using the telephone system to get onto the Internet, your telecommunications connection follows either the normal phone system (POTS) or an ISDN or ADSL line until it reaches the telephone company's central office. The central office routes calls either through its own network of copper, fiber-optic, or satellite links, or to a long distance carrier's point of presence, or POP. (The POP is the point at which a local call is handed off to a long distance company.) The call is then routed to the central office nearest your Internet service provider, also called an ISP.





3 The ISP receives your incoming call and processes your request, passing the connection to its leased line link to a computer on the Internet backbone, usually a T1 connection or faster. Typically, the ISP has a rack of modems that accepts multiple incoming calls simultaneously. Your connection is complete when it reaches the host computer.





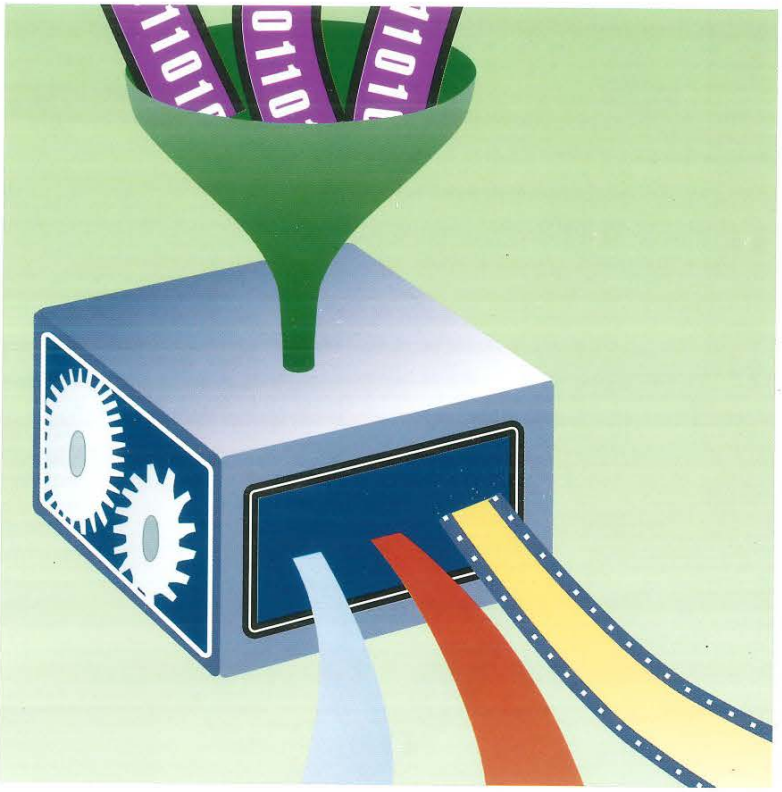
5 The host runs special Web server software that reads requests sent from the client computer. The host also retrieves and sends the appropriate information to the client. These host computers may have dedicated T1 or T3 links to the Internet backbone, or may be connected to the backbone through a network of hosts.

4 Computers with dedicated links to the Internet backbone must always be on and the connection must always be open. These computers are assigned a permanent IP (Internet Protocol) address, which is a multidigit number. This address enables other machines to find it. When you connect to the Internet via an ISP, your connection to the Internet backbone is opened, and your computer is assigned a temporary IP address—although some ISPs give you a permanent IP address, which is always yours whenever you connect to the Internet.

NOTE To get onto the Web, you'll need an IP address. If you get onto the Web at work via a LAN, the odds are you have a single IP address that is used every time you go onto the Internet. However, when you dial in from home, your IP address need not be the same every time you dial in. Often, the ISP you use will assign you a temporary IP address every time you dial in. By doing this, ISPs keep down the number of IP addresses they need to maintain.

CHAPTER

6 Internet File Types



THERE are millions of files on the Internet that show you pictures, let you hear music and sounds, allow you to watch videos, and let you read articles and run software on your computer. Some of these files can be used on any computer, regardless of whether it is a Macintosh or an IBM-compatible PC. Graphics files, sound files, and video files, for example, can be played on many different kinds of computers, as long as those computers have the special software (often called *players* or *plug-ins*) required to view, run, or listen to them. Some files require special hardware in order to work. Other files, such as software programs that you download, can only run on a PC or a Mac, or they may require extra memory.

In general, on the Internet you'll find three kinds of files: ASCII (American Standard Code for Information Interchange); EBCDIC (Extended Binary Coded Decimal Interchange Code); and binary. ASCII and EBCDIC are ways of organizing data into something we can understand. ASCII codes are computer codes that represent the characters on your screen, such as the uppercase letter A or a dollar sign ($). ASCII text files contain nothing but simple character data. They lack the sophisticated formatting commands that word processing or desktop publishing programs can apply to a document. EBCDIC is like ASCII but you won't normally find EBCDIC files because they're primarily used by mainframe computers. Binary files contain special coded data and can only be run or read by specific computers and software.

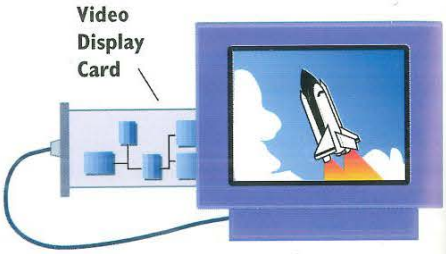
In contrast to plain ASCII text files, there are files that contain sophisticated formatting and graphical information in them. Files that adhere to the PostScript page description language are of this type. These files are in fact ASCII files, but they contain information about how to format and print the file. There are other kinds of files that contain formatting and graphical information, such as those in the Adobe Acrobat PDF format, but many of them are binary files instead of ASCII files. In order to view or print either of these types of files, you'll need special software readers. In the case of PostScript files, you'll generally need a special PostScript printer that can print files in that language. Files of this type can be viewed in two ways: either when you're online and connected to the Internet, or when you're offline and not connected. To view the files when online, you'll use special plug-in modules or helper applications for your Web browser. To view the files when offline, you'll need software readers.

Sound, visual image, animation, and video files are also common on the Internet. These files are all binary files, which are made up of bits that are represented by 0s and 1s. Files of this type are often large, requiring special software readers, players, and sometimes hardware in order to play and read them. Some of these files can only be played or viewed on a specific type of computer, whereas others can be played or viewed on many types of computers if you have the right software. Some of the files, such as streaming audio files and streaming video files, can be viewed while you're online, connected to the Internet. Other types of files need to be viewed with special readers or players when you're off line.

File Types on the Internet

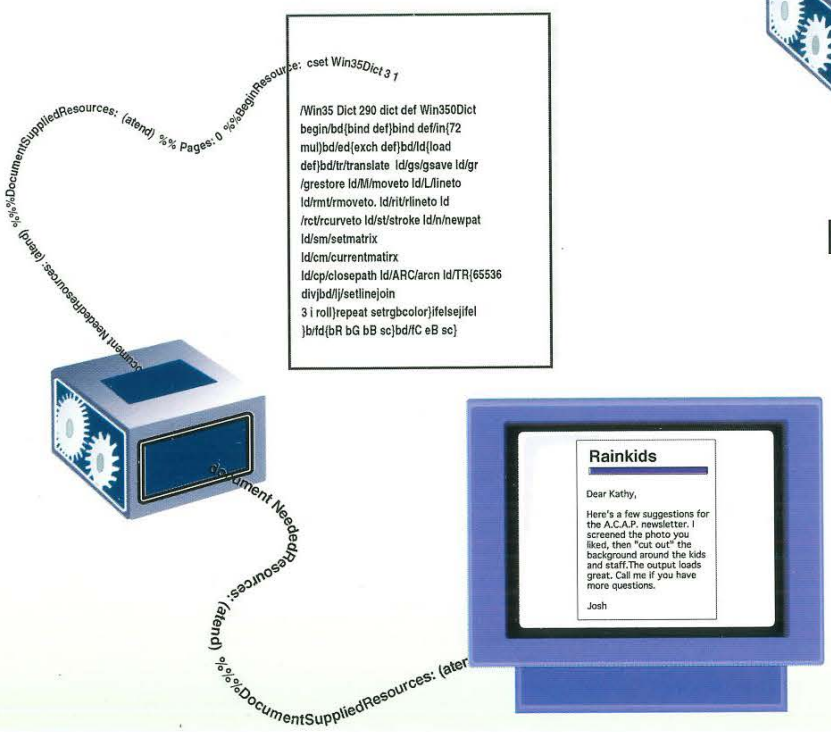
1 There are many different types of files on the Internet. Each file has information contained in its header. The header is simply the first line of a file or a specific number of bytes at the beginning of a file. Software that can read or view the files looks into the header of a file to distinguish what kind of file it is, and then processes the file accordingly.

2 You'll find graphics files in many formats, including GIF, JPEG, PNG, PCX, and TIF. GIF and JPEG are two of the most common graphics formats. They allow the exchange of image files between different kinds of computers and can be downloaded to your computer quickly. A GIF reader or JPEG reader can read binary GIF and JPEG files and display them on your computer. Graphics programs will typically read both types of files. A GIF file contains the binary data that will display an image when viewed with the proper reader. Your computer's video card takes information from the reader and displays it on your computer's monitor. Both GIF and JPEG are the commonly used graphics formats in Web pages. Your Web browser has the ability to read and display these files on Web pages without needing any added software.

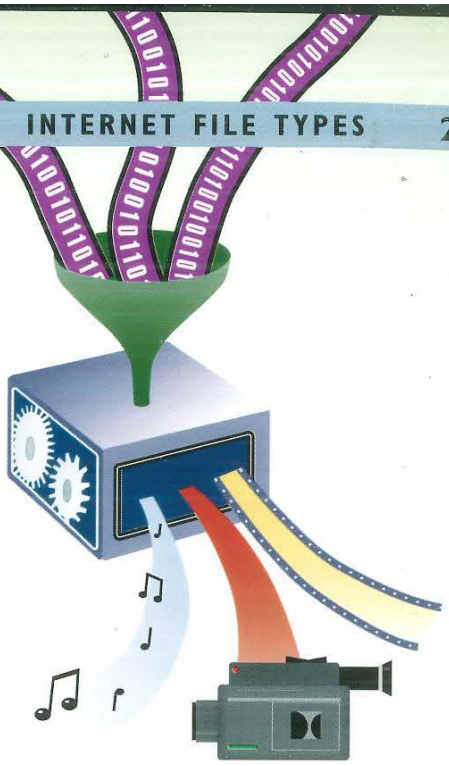




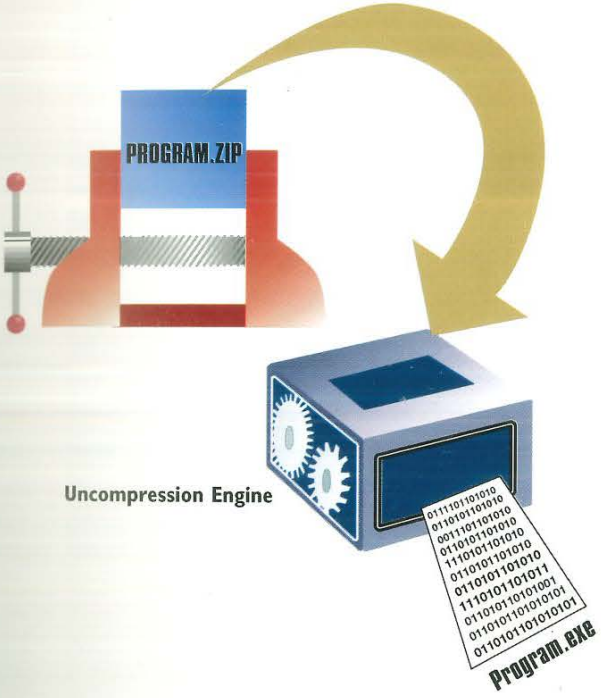
Software Engine



3 Several different file types, such as PostScript (which ends in a .PS extension) and Adobe Acrobat (.PDF) files, contain complex information about documents. This includes placement of pictures, size and type of fonts, and the complex shapes and formatting information that is needed in order to view the page. You'll need specific readers for PostScript and Acrobat files. These readers allow you to see the fully formatted pages on your screen over the Internet and print them if you wish.

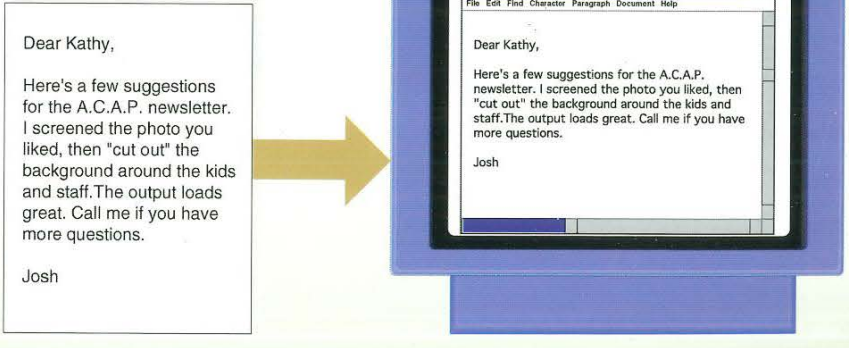


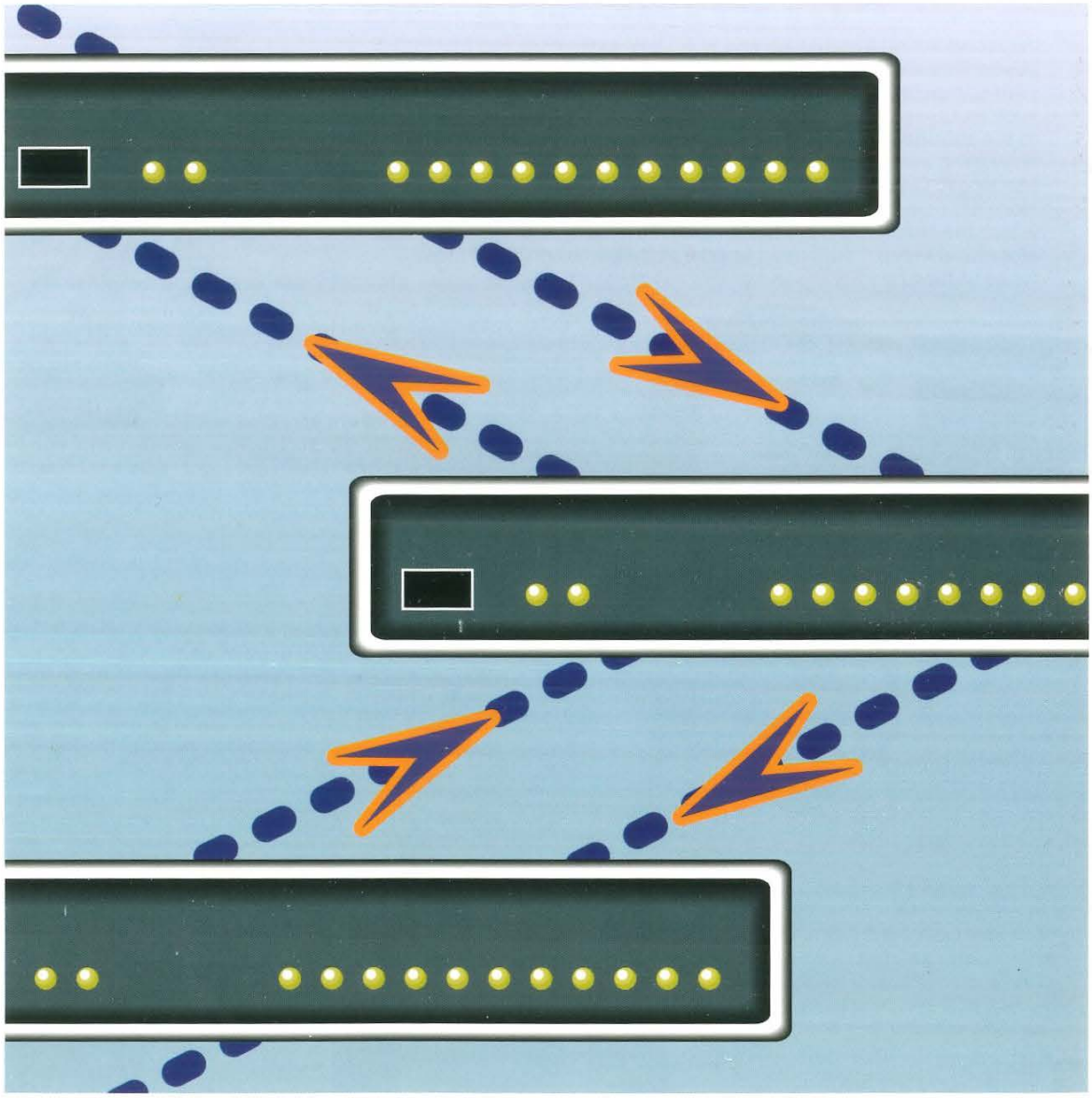
4 You'll find many different kinds of binary multimedia files on the Internet that let you listen to or play sounds, music, and videos. Windows sound files (which end in a .WAV extension) and Macintosh sound files are common types of files that you can download to your computer and play with a sound player. Other sound files, such as those created by Real Audio software, allow you to listen to the sounds while the file is downloading. This process is called *streaming audio*. Windows .AVI files are common animation files. Macintosh QuickTime and MPEG files (which end in an .MPG extension) are common video formats. Other animation files allow you to watch the video while the file is downloading. This process is called *streaming video*. You'll need special software to use all of these types of files.



5 There are many executable programs you'll find on the Internet that you can download to your computer and use just like any other kind of software. Because these software files can be large and take a long time to download, it is often compressed to make it download faster. Once it's on your computer, you uncompress it with special decompression software and run it like any other program. Frequently, PC software has been compressed with PKZIP and has a .ZIP extension, whereas Macintosh software can be compressed with a variety of compression software. In general, PC files can only be used with PCs, while Macintosh files can only be used with Macintoshes.

6 You'll find many ASCII text files on the Internet that you can read with a text editor or word processor. These can be articles, FAQs (frequently asked questions and their answers), or any other kind of informational files.





PART 2

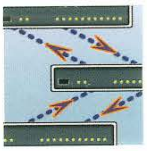
THE INTERNET'S UNDERLYING ARCHITECTURE

Chapter 7: How the Domain Name System Works
32

Chapter 8: How the Routers Work
36

Chapter 9: The Internet's Client/Server Architecture
40

AS you browse the World Wide Web or do anything else on the Internet, you take for granted that all information will be handled in a timely way. When you send data, you expect it to get to its proper location quickly. It all appears so seamless, as simple as the U.S. mail being delivered to the proper address. In fact, though, every time you do the simplest thing, such as contacting a Web site, you rely upon a remarkably complex architecture made up of complicated hardware and software spread across the planet—architecture that has been developed over the years and continues to be refined. This architecture is made up of your own computer and the software it runs; hardware, such as powerful servers that take your requests and then relay information to you; routers that examine every packet being sent across the Internet and ensure that it is sent to its proper destination; name servers that translate between alphanumeric Internet addresses, such as `pgralla@ziffdavis.com`, and IP addresses, such as `123.22.22.112`; and sets of rules that govern how computers on the Internet communicate with each other. When you look at it this way, it appears a remarkable feat that a simple piece of email you send to a friend manages to circumnavigate the globe and is received in a single piece.



Consider what has to happen for that piece of email to be sent to your friend. After you compose it on your computer and send it, it has to be broken down into many small pieces, and every one of those pieces has to be sent to the proper place and reassembled in the proper order. The email address, `myfriend@hishome.com`, for example, must be translated into its IP address, such as `123.22.22.112`, so that the Internet's hardware and software can figure out how to get it to its proper destination. That translation has to occur in every single packet in the message. After the IP address is known, the packets have to travel through many routers across the Internet, each of which sends the packets on their way, closer to their final destination. Your email software and your friend's email software have to communicate with mail servers to send and receive the mail. When you look at it this way, it's not quite so simple a task.

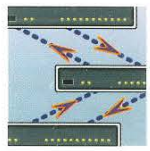
The chapters in this section cover the underlying architecture of the Internet—the hardware, the software, and the general organization that you rarely see or think of, but without which the Internet would not exist.

Chapter 7, "How the Domain Name System Works," covers the *Domain Name System (DNS)* and *domain name servers*. The DNS and its associated servers perform the most basic of tasks for the Internet—translating between the common alphanumeric Internet addresses with which we're all familiar, such as `prepg@tiac.com`, and the IP addresses that Internet computers use, such as `112.45.45.113`. Computers and people "think" differently; we think in letters as well as numbers, but computers can handle only numbers. So the DNS translates between human-speak

and computer-speak. It ensures that when you send email or try to contact a Web site, the site you think you're contacting matches the one on the Internet.

Chapter 8, "How the Routers Work," details how the most basic piece of hardware on the Internet—a router—works. *Routers* are combinations of hardware and software that perform the job of making sure that all data is sent to the proper destination. Think of routers as the traffic cops of the Internet. They use the IP addresses that the name servers have translated to route the data. Routers look at the address and then send the data to the next-closest router to the destination and so on, until the data is finally delivered. They use *routing tables* to determine how to route the traffic, and they can also adjust the routes as the traffic on the Internet changes, ensuring that the data is routed in the most efficient way possible.

Finally, Chapter 9, "The Internet's Client/Server Architecture," looks at the Internet's client/server architecture. *Servers*—also called *hosts*—are powerful computers that perform functions such as delivering information or Web pages, hosting databases, or handling email. A *client* is your own computer and the software that sits on it, such as a Web browser or a piece of email software. Clients request information from servers, which do the heavy-duty processing and then send the information back to the client, which displays the information.



So the next time you send a piece of email or view a Web page, realize that a lot is going on behind the scenes.

CHAPTER 7

How the Domain Name System Works



THE heart of how the Internet works is the Domain Name System (DNS), the way in which computers can contact each other and do things such as exchange email or display Web pages. When someone on the Internet wants to contact a location—for example, to visit a Web site—he or she will type in an address, such as `www.metahouse.com`. The Internet Protocol (IP) uses Internet address information and the DNS to deliver mail and other information from computer to computer.

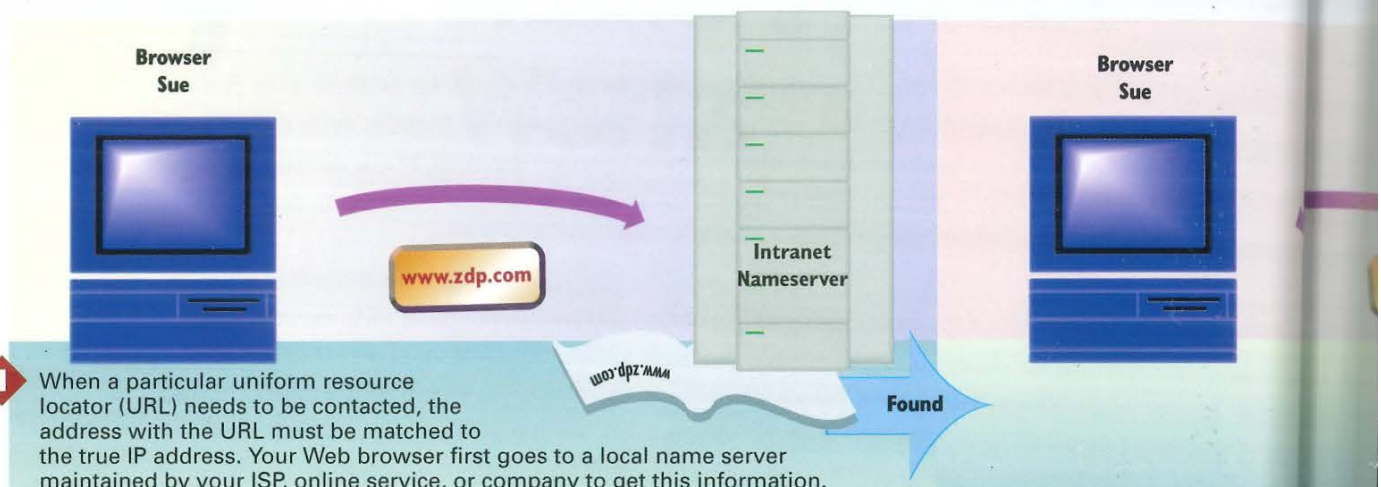
The DNS translates the plain English address, `www.metahouse.com`, for example, into numbers that Internet computers can understand, such as `123.23.43.121`. In order to do this efficiently, the Internet has been organized into a number of major domains. Major domains refer to the letters at the end of a plain English address, such as `.com`. A number of common domains are used in the United States: `.com` (commercial); `.edu` (education); `.gov` (government); `.mil` (military); `.net` (Internet service providers and networks—companies and groups concerned with the organization of the Internet); and `.org` (organization). Because the number of Internet sites has been growing exponentially, the domain name system is being expanded and may also include at least seven additional domains, such as `.web` for Web. Only two letters are used outside the United States to identify the domains; for example, `.au` for Australia; `.ca` for Canada; `.uk` for United Kingdom; and `.fr` for France.

Domains are organized in a hierarchical manner, so that beneath major domains are many minor domains. As an example of how the DNS and domains work, look at NASA's SPACElink Internet address: `spacelink.msfc.nasa.gov`. The top domain is `.gov`, which stands for government. The domain just below that is `.nasa`, which is the NASA domain. Then below that, `.msfc` (Marshall Space Flight Center) is one of NASA's many computer networks. Spacelink identifies the NASA computer that runs the SPACElink program. SPACElink's numeric IP address has changed through the years, but its Internet address has stayed the same.

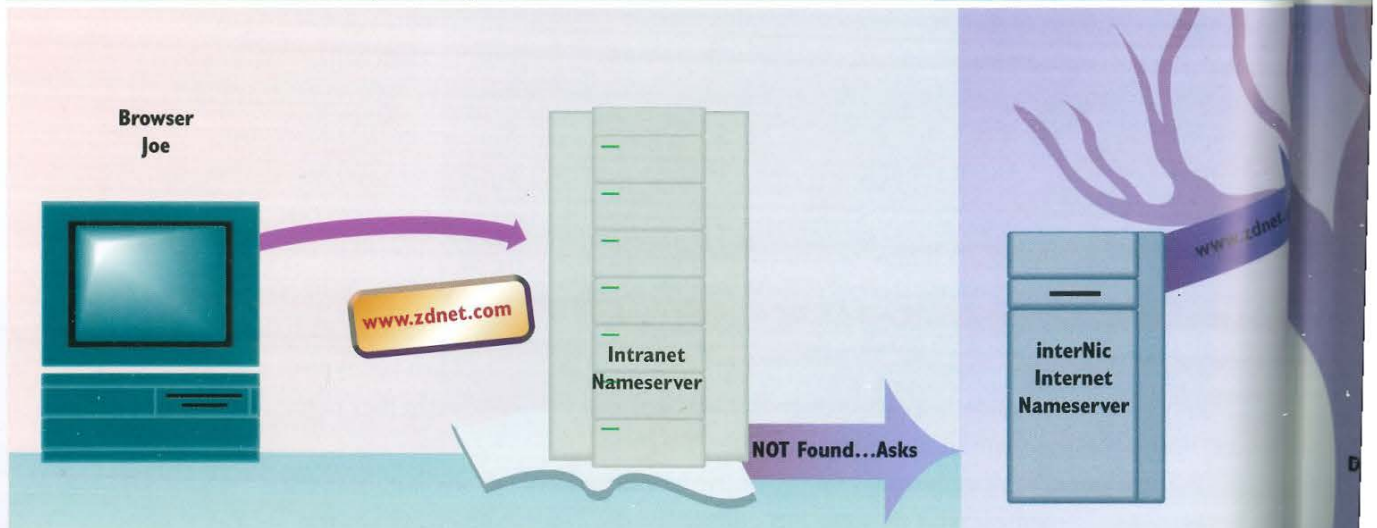
The DNS system keeps track of changes such as the one just mentioned so that even when an IP address changes, if the Internet address is used, email will always be delivered to the proper place. Computers called *name servers* are responsible for keeping track of such changes and translating them between IP addresses and domain addresses. Name servers also work with the DNS to ensure that mail is delivered to the right person. They make sure that when you type in a Web Uniform Resource Locator (URL) you're sent to the proper location. Lastly, they're responsible for properly routing all messages and traffic on the Internet.

The Internet can't understand alphanumeric Internet addresses such as `pgralla@ziff-davis.com`, so name servers translate that address into its proper numeric IP address, such as `163.52.128.72`. Name servers contain tables that match alphanumeric Internet addresses to numeric IP addresses.

How Domain Name System (DNS) Servers Work



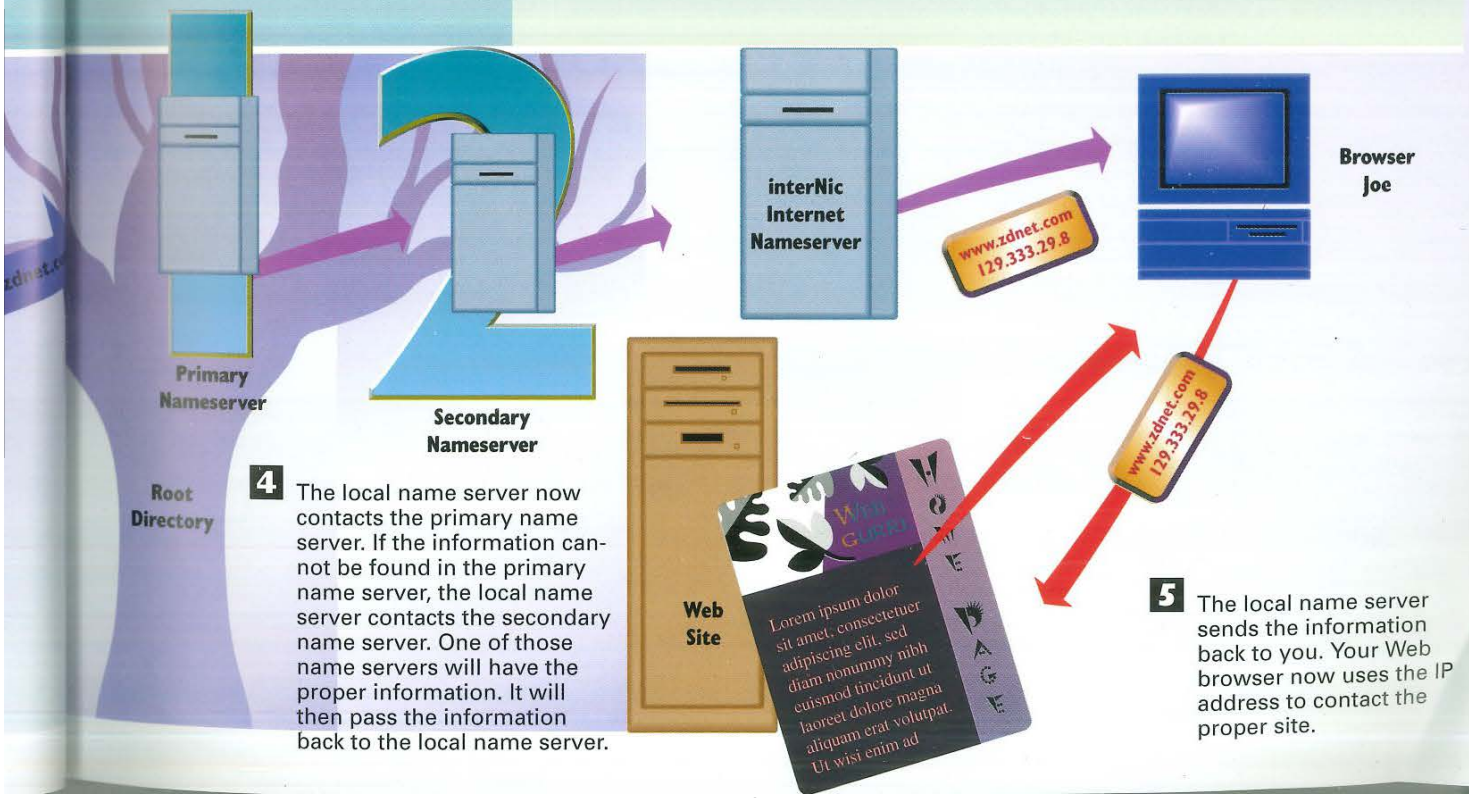
1 When a particular uniform resource locator (URL) needs to be contacted, the address with the URL must be matched to the true IP address. Your Web browser first goes to a local name server maintained by your ISP, online service, or company to get this information. If the IP address is a local one—on the same network as the one you are on—the name server will be able to resolve the URL with the IP address. It will send the true IP address to your computer.



3 If the information you have requested isn't on a local network, the local name server may not have the address you're looking for. In that instance, the local name server must get the information from a name server on the Internet. The local name server contacts the root domain server. The root domain server tells the local server which primary name server and secondary name server have the information about the requested URL.

2 Your Web browser now has the true IP address of the place you're trying to locate. It uses that IP address and contacts the site. The site sends you the information you've requested.



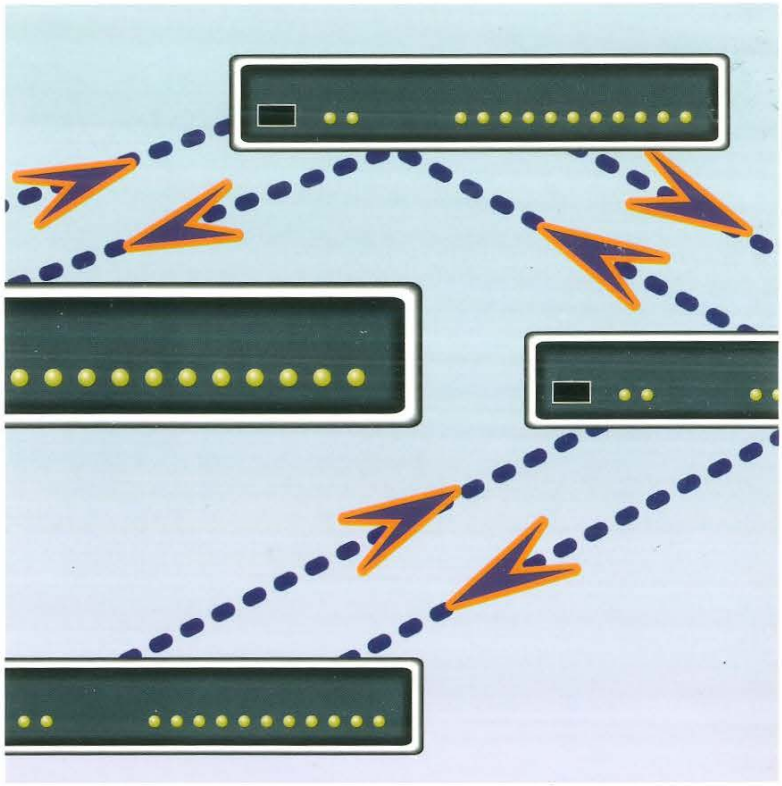


4 The local name server now contacts the primary name server. If the information cannot be found in the primary name server, the local name server contacts the secondary name server. One of those name servers will have the proper information. It will then pass the information back to the local name server.

5 The local name server sends the information back to you. Your Web browser now uses the IP address to contact the proper site.

CHAPTER

8 How the Routers Work



ROUTERS are the traffic cops of the Internet. They make sure all data gets sent to where it's supposed to go via the most efficient route. When you sit down at your computer on the Internet and send or receive data, generally that information must first go through at least one router and often more than one router before it reaches its final destination.

Routers open the IP packets of data to read the destination address, calculate the best route, and then send the packet toward its final destination. If the destination is on the same network as the sending computer, such as within a corporation, the router will send the packet directly to the destination computer. If the packet is destined for a destination outside the local network, the router will instead send the packet to another router closer to the destination. That router in turn will send the packet to a yet-closer router, and so on, until the packet reaches its final destination.

When routers determine the next router to receive packets, factors such as traffic congestion and the number of *hops* (routers or gateways on any given path) come into play. The IP packet carries a segment that holds the maximum hop counts it can travel and a router will not use a path that would exceed that predefined number of hops.

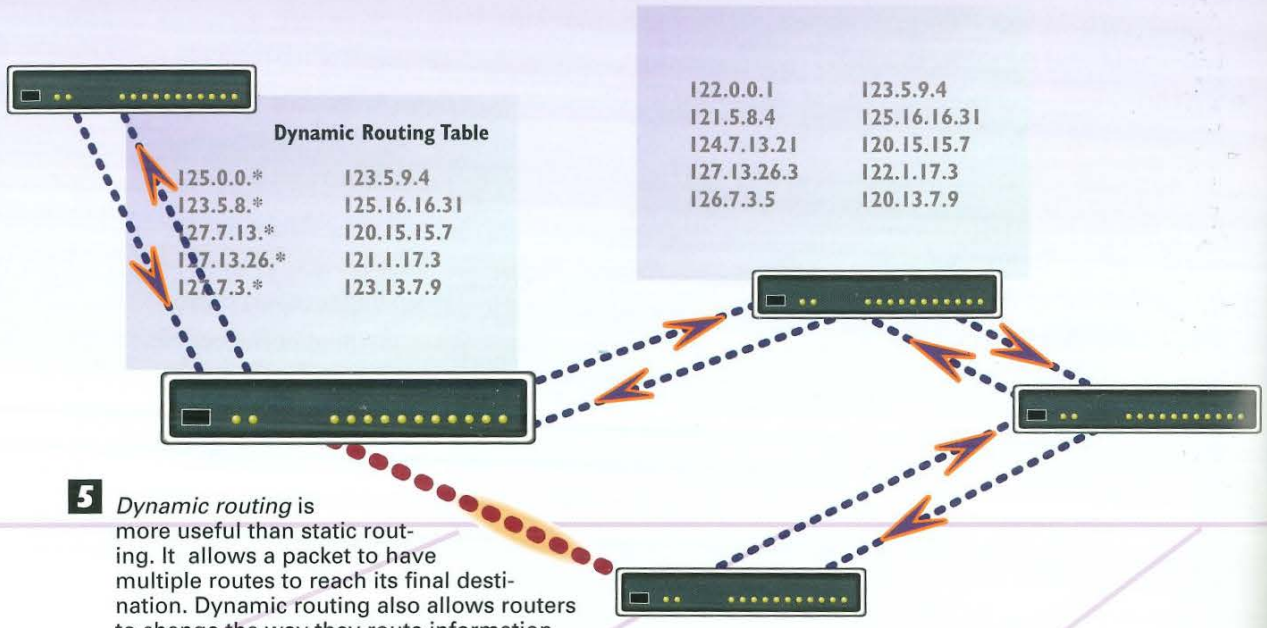
Routers have two or more physical ports: receiving (input) ports and sending (output) ports. In actuality, every port is bidirectional and can receive or send data. When an input port receives a packet, a software routine called a *routing process* is run. This process looks inside the header information in the IP packet and finds the address where the data is being sent. It then compares this address against an internal database called a *routing table.* The routing table has detailed information about the ports to which packets with various IP addresses should be sent. Based on what it finds in the routing table, the router sends the packet to a specific output port. This output port then sends the data to the next router or to the destination itself.

At times, packets are sent to a router's input port faster than the port can process them. When this happens, the packets are sent to a special holding area called an *input queue*, which is an area of RAM on the router. That specific input queue is associated with a specific input port. A router can have more than one input queue. Each input port processes packets from the queue in the order in which they were received so the first packets sent in are the first to get processed and sent out.

If the number of packets received exceeds the capacity of the queue (called the *length* of the queue), packets may be lost. When this happens, the TCP protocol on the sending and receiving computers will have the packets re-sent.

How Routers Work

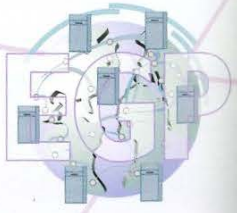
1 A router has *input ports* for receiving IP packets and *output ports* for sending those packets toward their destination. When a packet comes to an input port, the router examines the packet header and checks the destination in it against a *routing table*–a database that tells the router how to send packets to various destinations.



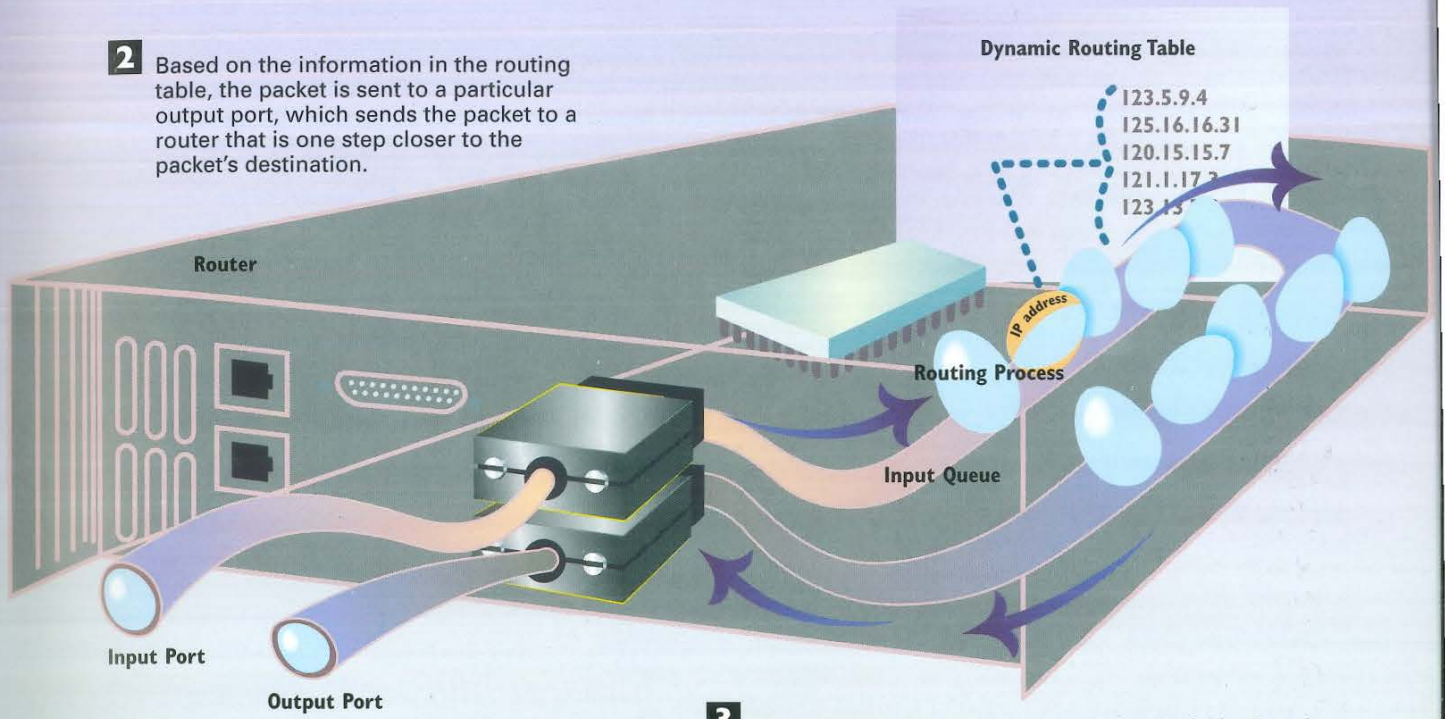
5 *Dynamic routing* is more useful than static routing. It allows a packet to have multiple routes to reach its final destination. Dynamic routing also allows routers to change the way they route information based on the amount of network traffic on some paths and routers. In dynamic routing, the routing table is called a *dynamic routing table* and changes as network conditions change. The tables are built dynamically by routing protocols, constantly changing according to network traffic and conditions.

6 There are two broad types of routing protocols: interior and exterior. *Interior routing protocols* are typically used only on routers in a company's *intranet*, or internal network. These interior routing protocols route traffic bound only for inside the intranet. A common interior routing protocol is the Routing Information Protocol (RIP). Exterior protocols are typically used for routers located on the Internet. A common exterior routing protocol is the Exterior Gateway Protocol (EGP).



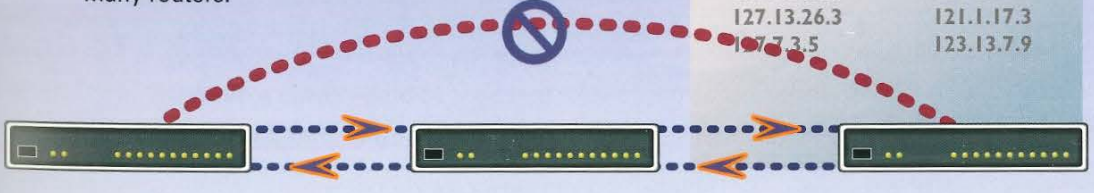


2 Based on the information in the routing table, the packet is sent to a particular output port, which sends the packet to a router that is one step closer to the packet's destination.



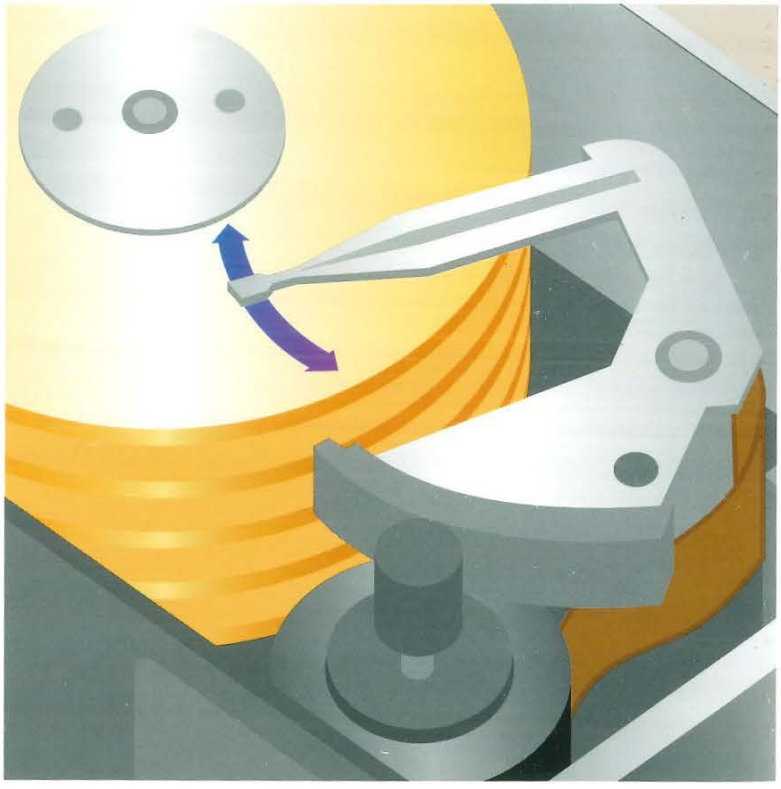
3 If packets come to the input port more quickly than the router can process them, they are sent to a holding area called an *input queue*. The router then processes packets from the queue in the order they were received. If the number of packets received exceeds the length of the queue, packets may be lost. When this happens, the TCP protocol on the sending and receiving computers will have the packets re-sent.

4 A simple kind of routing table is called a *static routing table*. In static routing, the routing table has specific ways of routing data to other networks. Only those pathways can be used. New routes can be added to the routing table. Static routing can't adjust routes as network traffic changes so it isn't an optimal alternative for many routers.



CHAPTER 9

The Internet's Client/Server Architecture



THE Internet works on the client/server model of information delivery. In this model, a client computer connects to a server computer on which information resides; the client depends on the server to deliver information. In effect, the client requests the services of the larger computer. These services may involve searching for information and sending it back to the client, such as when a database on the Web is queried. Other examples of these services are delivering Web pages, and handling incoming and outgoing email. Whenever you use the Internet, you're connected to a server computer and requesting the use of that server's resources.

Typically, the client is a local personal computer and the server (also known as the *host*) is usually a more powerful computer that houses the data. These computers can be of many different kinds: powerful Windows-based PCs, Macintoshes, and a wide variety of hardware that runs the UNIX operating system, such as computers made by the Digital Equipment Corporation.

The connection to the server is made via a LAN (local area network), a phone line, or a TCP/IP-based WAN (wide area network) on the Internet. A primary reason to set up a client/server network is to allow many clients to access the same applications and files that are stored on a server.

In the case of the Internet's World Wide Web, the client is actually the browser on your PC and the server is a host computer located somewhere on the Internet. Typically, the browser sends the server a request for a specific Web page. The server processes that request and sends an answer back to the browser (again, most often in the form of a Web page).

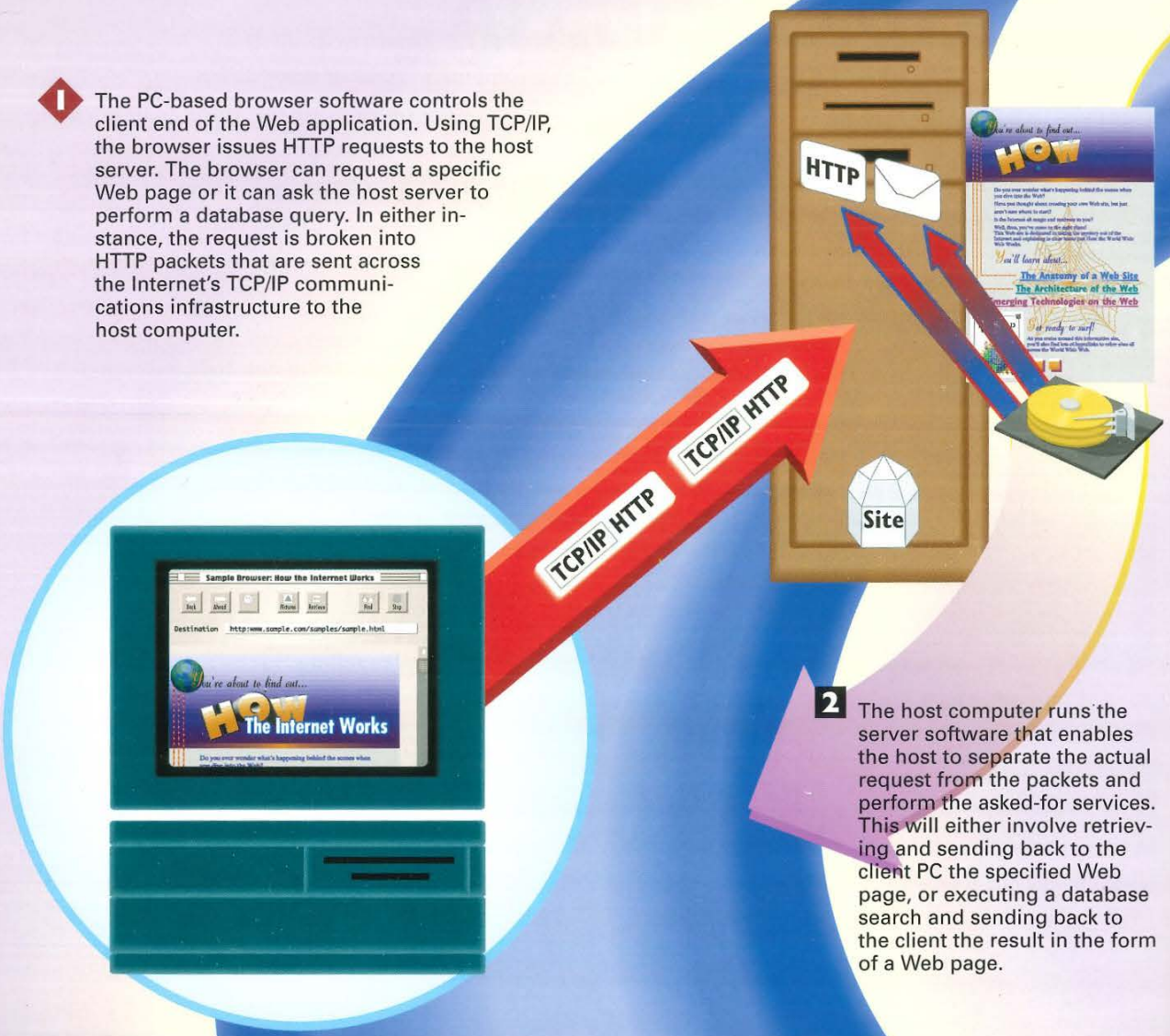
The connection between the client and server is maintained only during the actual exchange of information. Thus, after a Web page is transferred from the host (or server) computer, the HTTP connection between that computer and the client is broken. (HTTP stands for Hypertext Transfer Protocol; it's the protocol used by the World Wide Web.) Even though the HTTP connection is closed, the ISP maintains the TCP/IP connection to the Internet.

The client/server model enables the desktop PC to run the browser software to search the Web, yet still access host servers around the Internet to execute search and retrieval functions. In essence, this architecture enables the Web to be conceived of as a limitless file storage medium and database, distributed among thousands of host computers, all accessible by any individual PC.

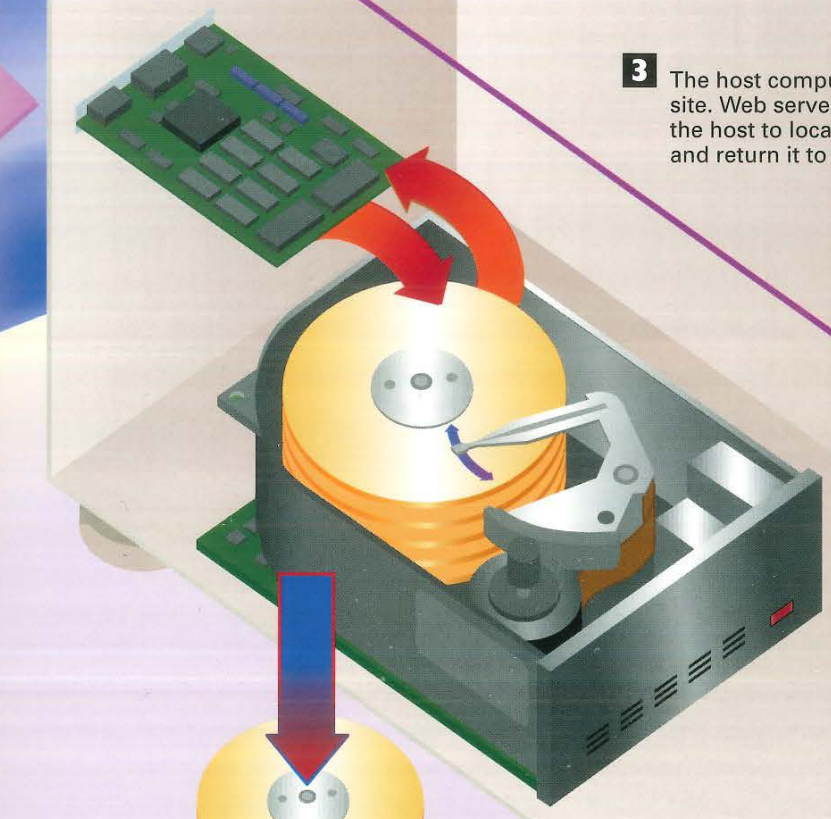
The following illustration shows how the Web runs on a client/server architecture. Keep in mind that all other resources on the Internet run on the client/server model as well. For example, in email transactions the client would be the email software on your computer while the server would be the email server into which you connect.

How Client/Server Architecture Works

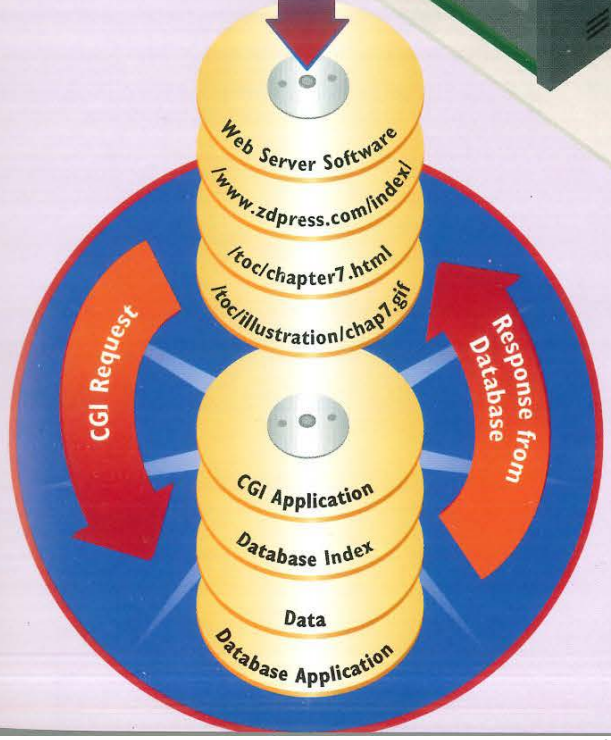
1 The PC-based browser software controls the client end of the Web application. Using TCP/IP, the browser issues HTTP requests to the host server. The browser can request a specific Web page or it can ask the host server to perform a database query. In either instance, the request is broken into HTTP packets that are sent across the Internet's TCP/IP communications infrastructure to the host computer.



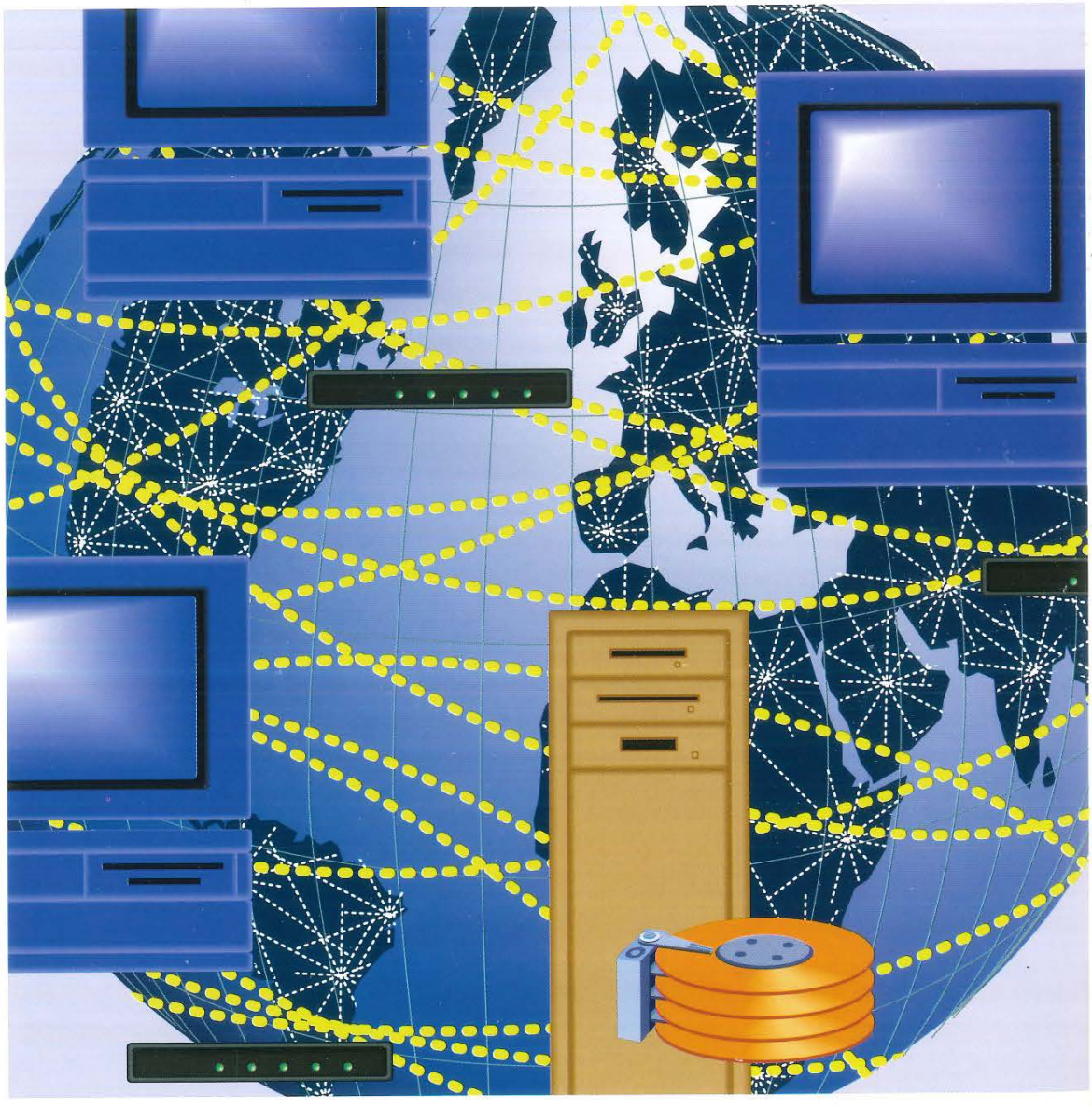
2 The host computer runs the server software that enables the host to separate the actual request from the packets and perform the asked-for services. This will either involve retrieving and sending back to the client PC the specified Web page, or executing a database search and sending back to the client the result in the form of a Web page.



3 The host computer houses the Web site. Web server software enables the host to locate a requested page and return it to the client.



4 Database and other similar applications on the host computer are accessed and controlled through a common gateway interface (CGI). When the server receives the request for the database query, it forwards the request to a separate, dedicated database server or application for processing via the CGI.



PART 3

CONNECTING TO THE INTERNET

Chapter 10: Connecting Your Computer
48

Chapter 11: Connecting to the Internet from Online Services
52

Chapter 12: How ISDN Works
56

Chapter 13: The Internet/Television Connection
60

Chapter 14: How Network Computers (NCs) Work
66

Chapter 15: How the Digital Subscriber Line (DSL) Works
70

NOW you know the basics of what the Internet comprises, and you know about its underlying architecture. So how do you actually connect to it?

You can connect to the Internet in many ways—and many more pop up practically every day. They range from simple telephone dial-in connections to special ISDN high-speed dial-ins, to television connections, Digital Subscriber Lines, and more. You can have a direct connection at your place of work or your university. You can browse the World Wide Web through a telephone line connection at home. Your computer can be merely a "dumb terminal" that can gain access to only a few Internet resources, such as a library database, or your computer can be a full Internet computer with access to everything the Internet has to offer. In this part of the book, we'll look at the myriad ways that people and computers gain access to the Internet.

One general rule is true about Internet connections: the faster the better. People want the fastest connection possible because many pictures, sounds, and videos are available on the Internet. Today, the two most common ways you can connect to the Internet are through a corporate or university LAN (local area network) or over telephone lines. Direct connections over LANs are generally faster than telephone line connections—although a new telephone technology called Digital Subscriber Lines (DSL) is changing that. Tomorrow, television and cable connections may be just as common.



In this section we'll look at all the ways that computers can connect to the Internet. Chapter 10, "Connecting Your Computer," gives an overview of the types of Internet connections that are possible. Not only will we examine different kinds of network connections and phone-line connections, but we'll also look at newer, high-speed connections that are becoming available; cable modems and DSL are two such examples.

Online services, such as CompuServe, America Online, and the Microsoft Network, at one time competed directly with the Internet for users. Now, instead of competing, those online services have joined the Internet club and provide Internet access to users. An online service can be one of the simplest ways to get an Internet connection. When you connect to an online service, you also get a full connection to the Internet. Chapter 11, "Connecting to the Internet from Online Services," examines how online services make that Internet connection.

Someday, we all expect to have extremely high-speed connections to the Internet, possibly via fiber-optic cabling in our homes. In fact, some communities are already installing such fiber optics. However, for the whole country to be wired will cost billions of dollars, and that capability appears to be at least several years away, if not longer. A quicker solution to reasonably priced, high-speed connections is ISDN (Integrated Services Digital Network). ISDN enables us to connect at high speeds to the Internet via normal, existing telephone lines. Not every area in the

country has ISDN yet. In some areas it can be an expensive option, but in others it is reasonably priced. It requires a special modem, as well. However, it is an increasingly popular way to get to the Internet at very high speeds—much faster than even the highest-speed modems. Chapter 12, "How ISDN Works," explains that new technology.

Chapter 13, "The Internet/Television Connection," looks at what may become one of the primary ways that many of us connect to the Internet—through some kind of television connection. The chapter looks at the most common types of television connections, using cable television as a way to get onto the Internet, and how to use an ordinary television to browse the Internet. A cable connection allows you to use your normal cable TV wires to hook up your PC to the Internet—and does it at exceedingly high speeds, far beyond even what ISDN has to offer. Meanwhile, services such as WebTV go a different route—they enable you to browse the Web using a normal TV, although they often make the Internet connection using telephone wires instead of cable wires.

Chapter 14, "How Network Computers (NCs) Work," details one of the more intriguing ways to hook yourself into the Internet, via Network Computers, or NCs (sometimes also called information appliances). This calls for very low-cost computers—at perhaps $500 or so—that have no local storage or software; instead, all software and storage are held on the Internet.

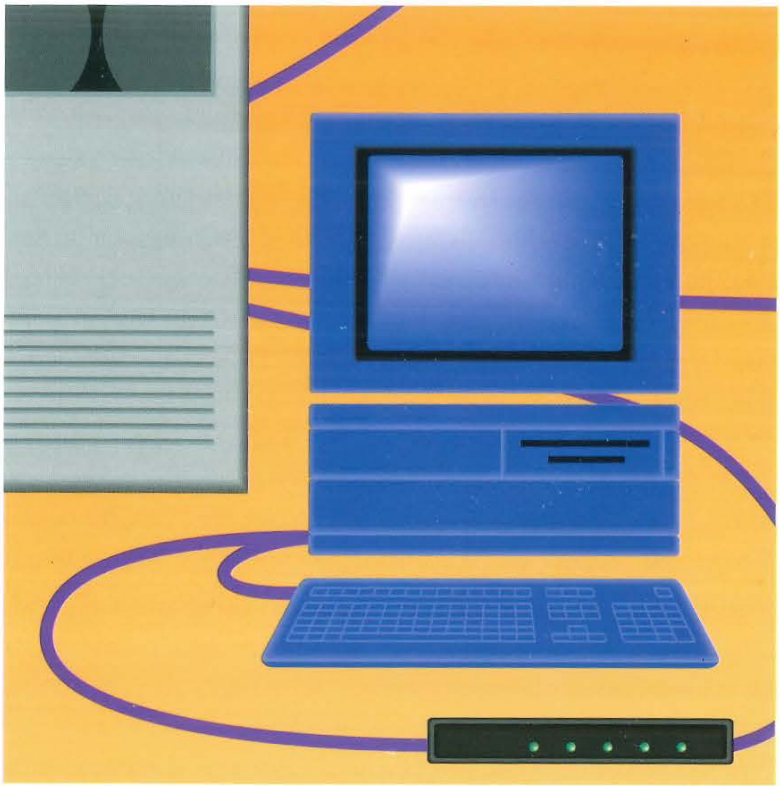


Chapter 15, "How the Digital Subscriber Line (DSL) Works," looks at one of the newer high-speed ways to connect to the Internet—DSLs. These connections enable you to use existing telephone lines to hook up to the Internet at extremely high speeds. To use it, you'll need a special DSL modem, and your phone company needs the special modems as well.

As you'll see in this part of the book, it will become increasingly easy to connect to the Internet—and at increasingly higher speeds. Not only will you be able to do things faster on the Internet, but you will also be able to access entirely new services that will contain video, animation, and other high-bandwidth content.

CHAPTER

10 Connecting Your Computer



THERE are many different ways your computer can connect to the Internet, ranging from dial-in connections to LANs (local area networks) to connections over cable TV wires. If you are connected to a LAN or campus network at your business or school, you may already be connected to the Internet. If the LAN or campus network that you're on is connected to the Internet via a router or bridge, that means your computer is also connected to the Internet. Often, this offers higher-speed access than when you dial into the Internet. If you're not connected to the Internet via a network, there are a variety of options for accessing the Internet and many more appear every day.

One option is to find a network or ISP (Internet service provider) that lets you dial into it with your home computer. You can do this with terminal emulation software that makes your computer function like a "dumb terminal." In this case, the Internet software doesn't actually run on your computer. Instead, you run it on the computer you've connected to and your screen merely shows what's happening on the computer you've dialed into. If you connect in this way, the Internet services you can access are severely limited. For example, you generally won't be able to see graphics while you browse the most popular part of the Internet, the World Wide Web.

If you want to tap the full power of the Internet over telephone lines, you'll need full Internet connections, such as those that are available with SLIP (Serial Line Internet Protocol) and PPP (Point-to-Point Protocol). When you dial into the Internet with these protocols, your computer becomes a part of the Internet and you can use all the processing power you've got. PPP is newer and more stable than SLIP and can retransmit packets if they get garbled, which is a fairly common situation you'll encounter when sending information over telephone lines.

ISDN (Integrated Services Digital Network) lines provide high-speed connections to the Internet. Typically, these speeds range from 64Kbps to 128Kbps. You'll need a special ISDN modem and an ISP that offers ISDN access. Asymmetric Digital Subscriber Line (ADSL), a newer technology, allows for very high-speed connections over existing telephone wires, but your ISP will have to support the technology.

Another option is a special cable modem that allows you to use your coaxial television cable to access the Internet. Your local cable company may be able to provide you with this service. Speeds can be up to 100 times faster than regular modem speeds.

There are many different kinds of connections you can use to hook up to the Internet. Start at your local college or library. Contact an ISP and find out about their fees. There are many national and local service providers, so shop around. Lastly, the major online services now include Internet access, so you can easily use them to connect to the Net.

Connecting Your Computer to the Internet



1 Dumb terminal A terminal that is connected to a mainframe, minicomputer, or other kind of large computer. This type of connection can often be found in libraries or universities although those institutions are moving toward offering fuller Internet access than dumb terminals provide.



2 Terminal emulation A personal computer can connect via modem to a large computer and run a terminal emulation program. The most common terminal emulation is the VT-100. The computer works like a dumb terminal, except it is connected via a phone line instead of a direct connection. Often, you won't be able to use the graphical part of the Internet, such as the World Wide Web, with this kind of access, although you'll be able to browse the text-only portion of the Web. This kind of Internet account is sometimes called a *shell account*. Terminal emulation has become increasingly rare.

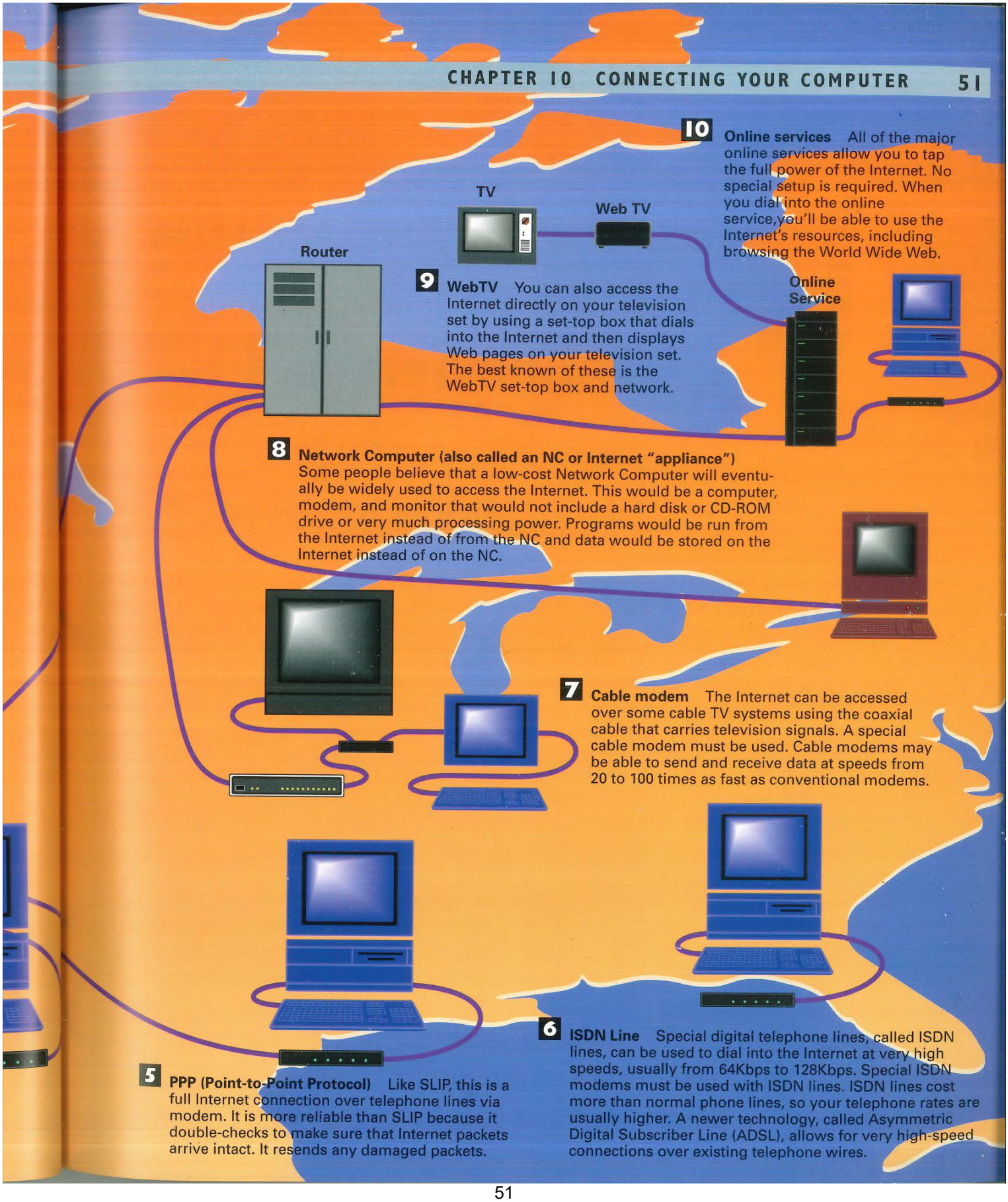
3 Direct connection LANs or large computers such as mainframes can be directly connected to the Internet. When a LAN is connected to the Internet, all the computers on the network can have full access to the Net. This kind of access is common inside corporations.



Mainframe



4 SLIP (Serial Line Internet Protocol) A full Internet connection over telephone lines that sends Internet packets via 9,600bps modems or better.



10 **Online services** All of the major online services allow you to tap the full power of the Internet. No special setup is required. When you dial into the online service, you'll be able to use the Internet's resources, including browsing the World Wide Web.

9 **WebTV** You can also access the Internet directly on your television set by using a set-top box that dials into the Internet and then displays Web pages on your television set. The best known of these is the WebTV set-top box and network.

8 **Network Computer (also called an NC or Internet "appliance")** Some people believe that a low-cost Network Computer will eventually be widely used to access the Internet. This would be a computer, modem, and monitor that would not include a hard disk or CD-ROM drive or very much processing power. Programs would be run from the Internet instead of from the NC and data would be stored on the Internet instead of on the NC.

7 **Cable modem** The Internet can be accessed over some cable TV systems using the coaxial cable that carries television signals. A special cable modem must be used. Cable modems may be able to send and receive data at speeds from 20 to 100 times as fast as conventional modems.

6 **ISDN Line** Special digital telephone lines, called ISDN lines, can be used to dial into the Internet at very high speeds, usually from 64Kbps to 128Kbps. Special ISDN modems must be used with ISDN lines. ISDN lines cost more than normal phone lines, so your telephone rates are usually higher. A newer technology, called Asymmetric Digital Subscriber Line (ADSL), allows for very high-speed connections over existing telephone wires.

5 **PPP (Point-to-Point Protocol)** Like SLIP, this is a full Internet connection over telephone lines via modem. It is more reliable than SLIP because it double-checks to make sure that Internet packets arrive intact. It resends any damaged packets.

CHAPTER 11

Connecting to the Internet from Online Services



THERE are many ways to get access to the Internet; one is by using an online service such as CompuServe or America Online. These online services have long provided their own unique content, special areas, and services available only to their subscribers. They use proprietary software and interfaces to give subscribers access to their resources. Unlike most of the Internet, the content, areas, and services the online companies provide are not always free. In order to get them, you may have to pay a monthly subscription fee to the online service. These online services are different from Internet service providers (ISPs) that offer only access to the Internet and don't have their own private areas and services.

The services let you access the Internet in a number of ways. To make it easier to use the Internet's resources, the services often use their own proprietary software or "user interface." The user interface requires you to issue a command that is then sent out over a gateway to the Internet. The information you've asked for is retrieved, sent back over the gateway to your online service, and is displayed for you using the online service's proprietary software.

Most online services also allow you to use your own client software to get to the Internet. To do this, you essentially bypass the online service's proprietary interface. You use the online service as you would a dial-in ISP. First, you dial into the online service to establish a TCP/IP connection. Then you launch the client software on your computer. It is the TCP/IP connection that gives you access to the Internet's resources.

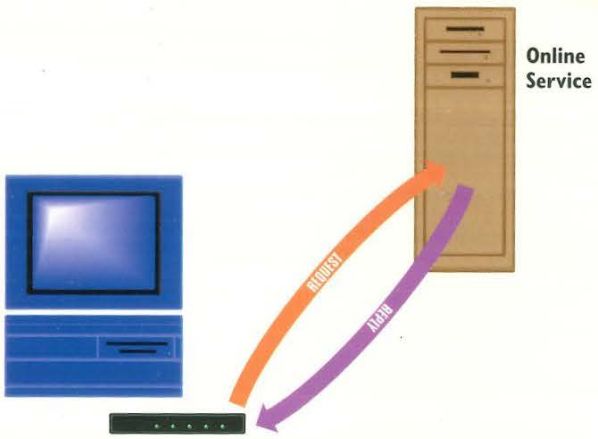
Online services also let you browse the Web. Some let you use only their own proprietary Web browsers. Others have their own proprietary Web browsers, but also allow you to browse the Web using any Web browser of your choice. To use your own Web browser, you first dial into the service to establish a TCP/IP connection. Then you launch your own Web browser and you can browse the Web just as you would with any other dial-in ISP.

In addition to letting you browse the Web, online services also let you use other Internet resources such as Telnet (which lets you gain access to a distant computer on the Internet), Gopher (a menuing system that gives you access to a great deal of information and resources), IRC (chat on the Internet), and Usenet newsgroups (public Internet discussion areas). In some instances, you'll gain access to those resources using the online service's proprietary software while in other instances you'll use normal Internet client software.

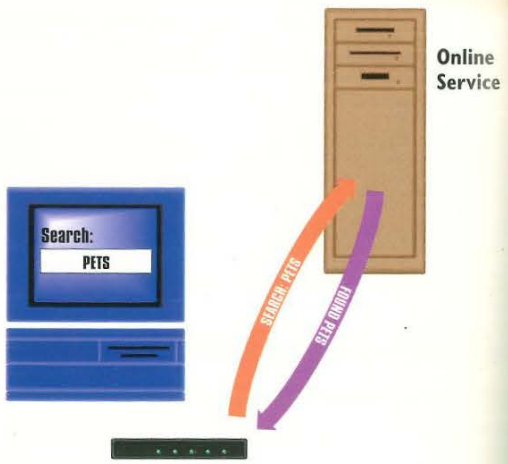
Increasingly, the online services are moving toward eliminating their proprietary architecture altogether. The trend is to build the service on the same framework as the Internet, such as the TCP/IP protocol and related technologies. The Microsoft Network, for example, is now based entirely on TCP/IP and Prodigy has launched an Internet-specific version of its service. America Online and CompuServe are both incorporating elements of Internet technology inside their services and may one day go entirely to an Internet-based model.

How Online Services Connect to the Internet

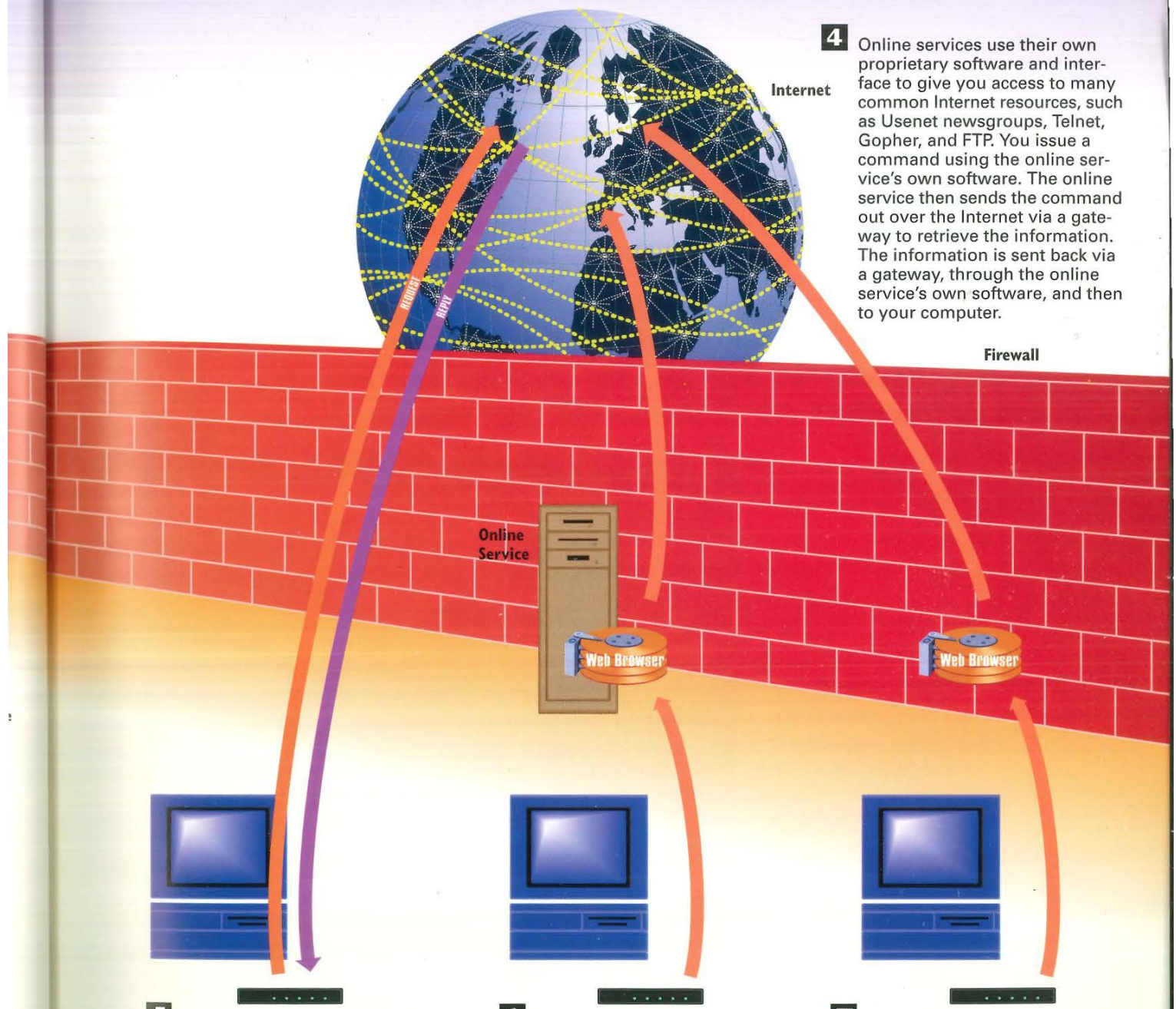
1 Online services such as CompuServe or America Online provide a very convenient way to use Internet resources. These services have their own software that makes it easy to access Internet resources such as Telnet, Gopher, and FTP. Some also allow you to use your own Telnet, FTP, Gopher, or other software instead of theirs. Using these resources, you may browse the World Wide Web, either with their own proprietary Web browser or by using any other Web browser. Finally, some services let you use special software required to access Internet resources such as IRC (Internet Relay Chat).



2 Each online service has a great deal of resources, content, and special areas that are not available to anyone except people who subscribe to that particular service. These areas use the online service's own proprietary software and interface; they do not work like the Internet. You dial directly into the online service to get at these resources.



3 When you use the resources of an online service, you don't go outside the service to the Internet—instead, you stay behind a firewall. A *firewall* is a security system of accepting or blocking packets as they are transmitted across a network. Individual online services establish firewalls that only let subscribers into the service.



4 Online services use their own proprietary software and interface to give you access to many common Internet resources, such as Usenet newsgroups, Telnet, Gopher, and FTP. You issue a command using the online service's own software. The online service then sends the command out over the Internet via a gateway to retrieve the information. The information is sent back via a gateway, through the online service's own software, and then to your computer.

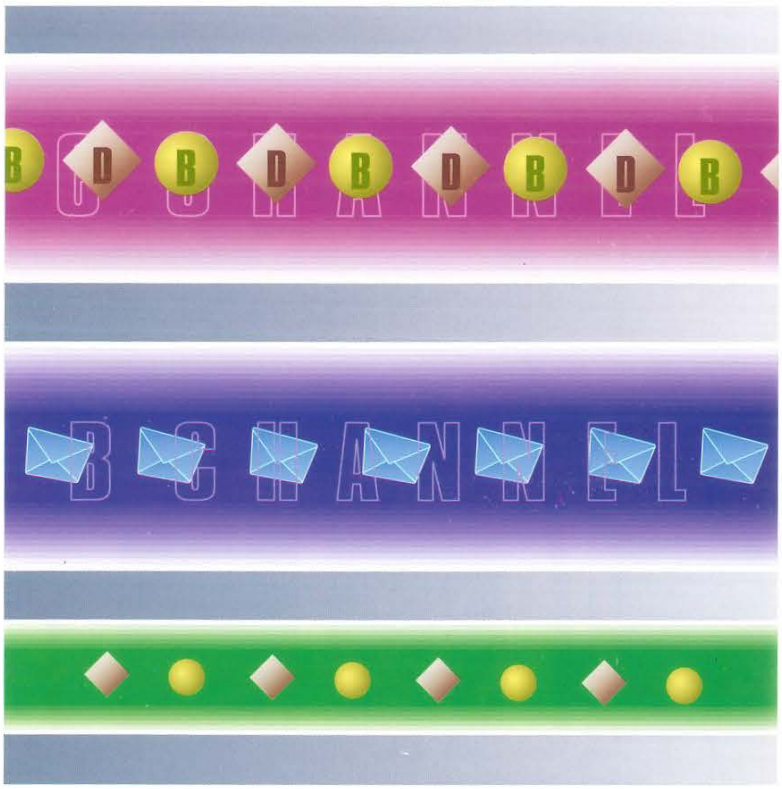
5 Most services also let you use your own client software to access Internet resources such as Telnet, Gopher, FTP, and IRC that the services do not have proprietary software for. In order to do this, you dial into the online service and establish a TCP/IP connection. You can then get directly onto the Internet using your own client software.

6 Online services also let you browse the World Wide Web. One way to browse the Web is to use the online service's own proprietary Web browser. Typically, you can launch a browser by clicking on an icon or going into a special area of the online service.

7 Some online services let you use any Web browser you want to browse the World Wide Web. In that instance, you can click on an icon on the online service, go into a special area of the online service, or simply launch your Web browser on your own computer, and then begin browsing the Web.

CHAPTER 12

How ISDN Works



ONE common complaint about the Internet is that ordinary telephone connections are too slow. Even "high-speed" modems that can connect at 28,800Kbps can be too slow to take advantage of the rich graphics now available. The Internet is full of graphics, videos, sounds, and other multimedia files that are very large and take a very long time to be sent across the Internet to your computer.

One affordable way to get a faster Internet connection is to use ISDN (Integrated Services Digital Network). It's a technology that's been around for years, but the advent of the Internet and the need for higher-speed connections has made it more popular.

ISDN allows for high-speed connections to the Internet using existing copper cable telephone wires. It isn't available to everyone in the country yet because it requires telephone companies to install special ISDN digital-switching equipment. However, every day more areas of the country offer ISDN service. If ISDN is available in your area, you'll pay extra for the service compared to a normal telephone line. In order to use ISDN and the Internet, the number you're dialing will have to be equipped for ISDN access. Many private Internet dial-in providers and online services allow for ISDN access.

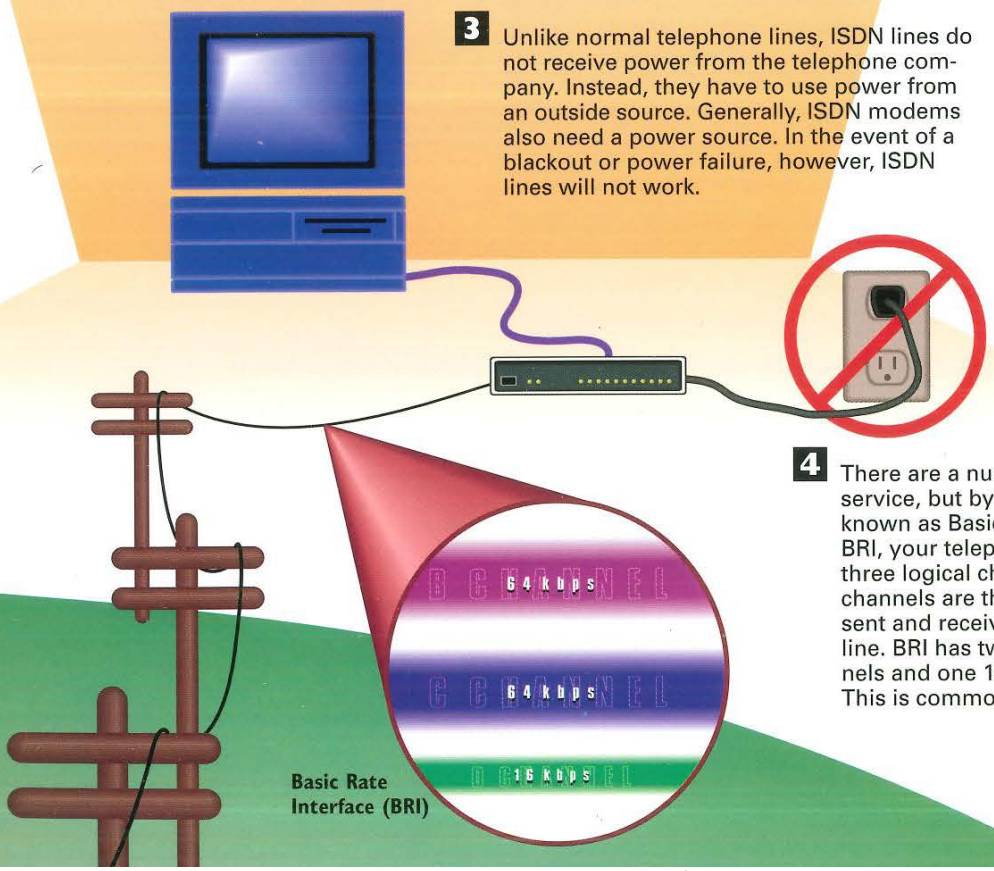
ISDN also requires that you have an ISDN modem. Although it looks like a modem, an ISDN modem isn't really a modem at all. Instead, it's a terminal adapter, a piece of hardware that lets you send and receive digital signals over ISDN phone lines. Normal modems transform digital signals from your computer into analog signals that can be sent over normal analog telephone lines. Because ISDN is a digital technology, only digital data is sent by the ISDN adapter. Some ISDN devices include the capability to function as a regular modem. This capability is needed because not every place you can dial into with your computer lets you use ISDN; some services presently only allow regular modem access.

There are different kinds of ISDN access, but the most common one is known as Basic Rate Interface (BRI). BRI divides your telephone line into three logical channels. The logical channels are not separate wires, but instead are ways in which data is sent and received over your telephone lines. BRI has two 64Kbps B (bearer) channels and one 16Kbps D (data) channel—commonly referred to as 2B+D. The D channel sends routing information while the two B channels send data. You can talk on one B channel while cruising the Internet on the other channel. Or, if your hardware and ISP or online service allows it, you can combine the two B channels into a single high-speed 128Kbps channel.

How ISDN Works

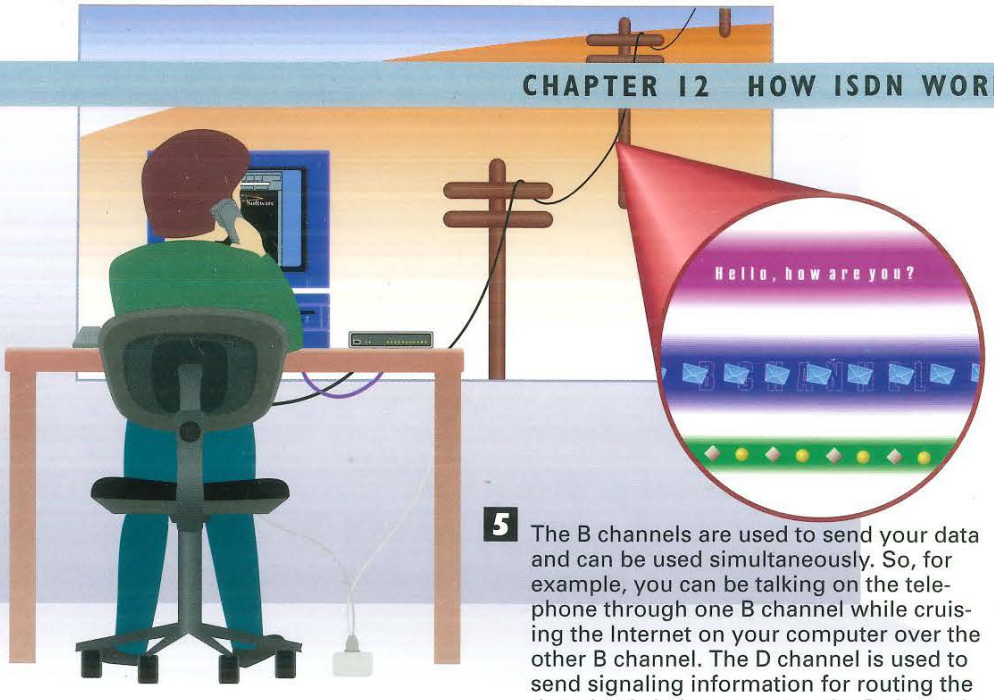
1 ISDN (Integrated Services Digital Network) is a way of establishing very high-speed connections to the Internet using existing copper telephone wires. No new telephone wires are needed in order for you to use ISDN. To allow you to use ISDN, your local telephone company will need to have installed special ISDN digital switches. In ISDN, all the information sent between your computer and the Internet is digital. Not all areas of the country have ISDN access yet. Call your telephone company to find out if it is available to you.

2 In order to use ISDN service, a computer requires an ISDN modem. In fact, this "modem" isn't really a modem at all and should be called a terminal adapter instead. A normal modem (short for modulator/demodulator) converts the information inside your computer from digital signals into analog signals that can be sent over telephone lines. ISDN is a digital technology so there is no need to convert information from digital to analog. Instead, the ISDN "modem" sends digital information from your computer over ISDN telephone lines and receives digital information from telephone lines.

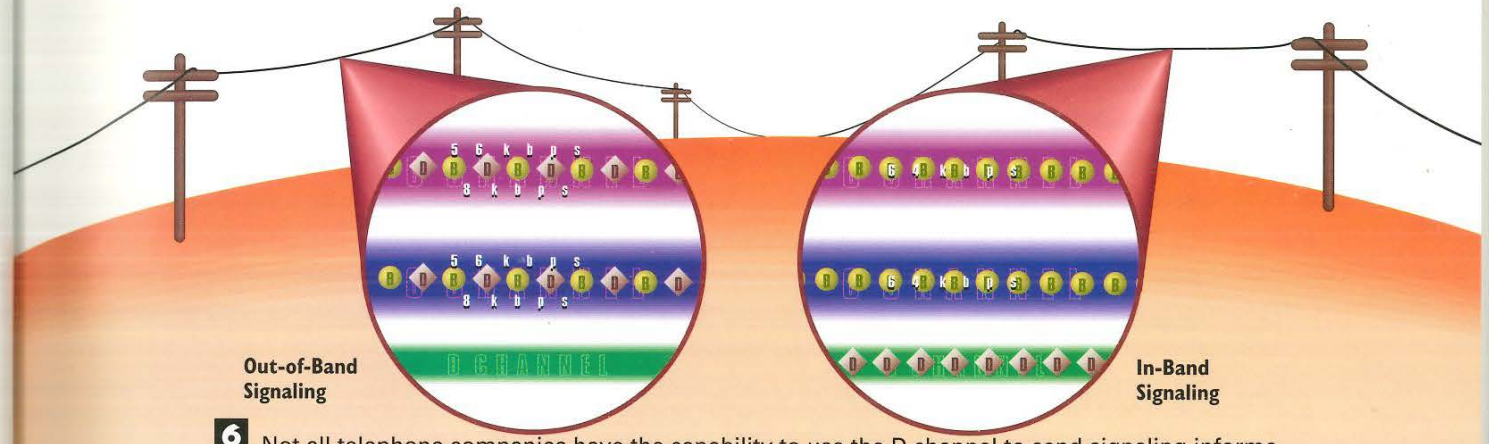


3 Unlike normal telephone lines, ISDN lines do not receive power from the telephone company. Instead, they have to use power from an outside source. Generally, ISDN modems also need a power source. In the event of a blackout or power failure, however, ISDN lines will not work.

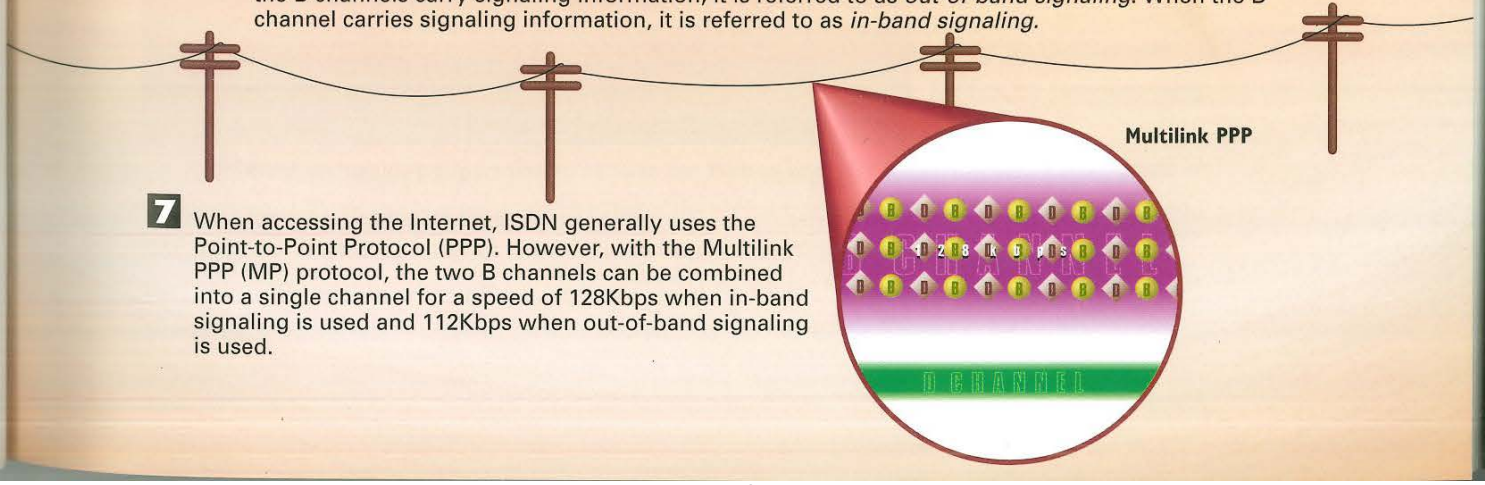
4 There are a number of variations of ISDN service, but by far the most common is known as Basic Rate Interface (BRI). With BRI, your telephone line is divided into three logical channels. These logical channels are the means by which data is sent and received over your telephone line. BRI has two 64Kbps B (bearer) channels and one 16Kbps D (data) channel. This is commonly referred to as 2B+D.



5 The B channels are used to send your data and can be used simultaneously. So, for example, you can be talking on the telephone through one B channel while cruising the Internet on your computer over the other B channel. The D channel is used to send signaling information for routing the data that is being sent over the B channels.



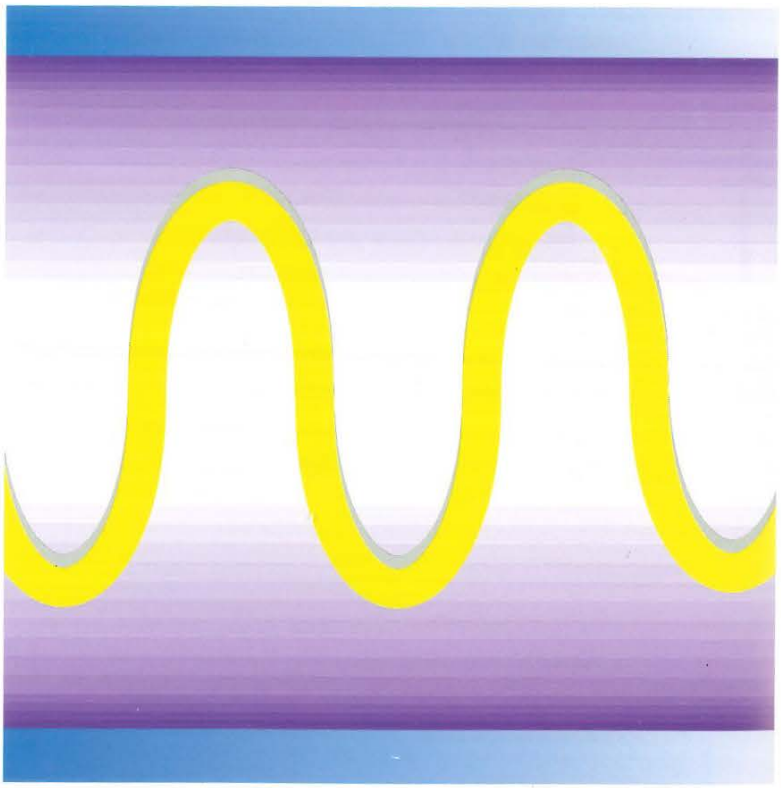
6 Not all telephone companies have the capability to use the D channel to send signaling information. In cases where the D channel can't be used for signaling information, that information is instead delivered over the B channels. In this instance, each B channel can only send and receive data at 56Kbps because the other 8Kb in each channel is used for signaling information. When the B channels carry signaling information, it is referred to as *out-of-band signaling*. When the D channel carries signaling information, it is referred to as *in-band signaling*.



7 When accessing the Internet, ISDN generally uses the Point-to-Point Protocol (PPP). However, with the Multilink PPP (MP) protocol, the two B channels can be combined into a single channel for a speed of 128Kbps when in-band signaling is used and 112Kbps when out-of-band signaling is used.

CHAPTER

13 The Internet/Television Connection



WHEN people first began talking about the "information superhighway" about five years ago, it wasn't the Internet they were talking about. Instead, it was television—and specifically cable television—that they believed would change the way we lived and worked. There were visions of 500 TV channels, "interactive television," shopping from home, and customized news available whenever you wanted it. This information superhighway was going to hook us all together electronically so we could more easily communicate and get information, services, goods, and entertainment. It didn't pan out that way. Instead, the Internet has become an information superhighway that can do almost everything people imagined could be done by using cable TV.

How much longer the Internet will be the sole driving force behind the information superhighway remains to be seen. Every day, TV and the Internet are drawing closer to one another. The Internet is gaining more TV-like qualities, such as the capability to play videos and music, and even the capability to broadcast live video feeds. Television technology is also developing in ways that will make it more interactive, like the Internet.

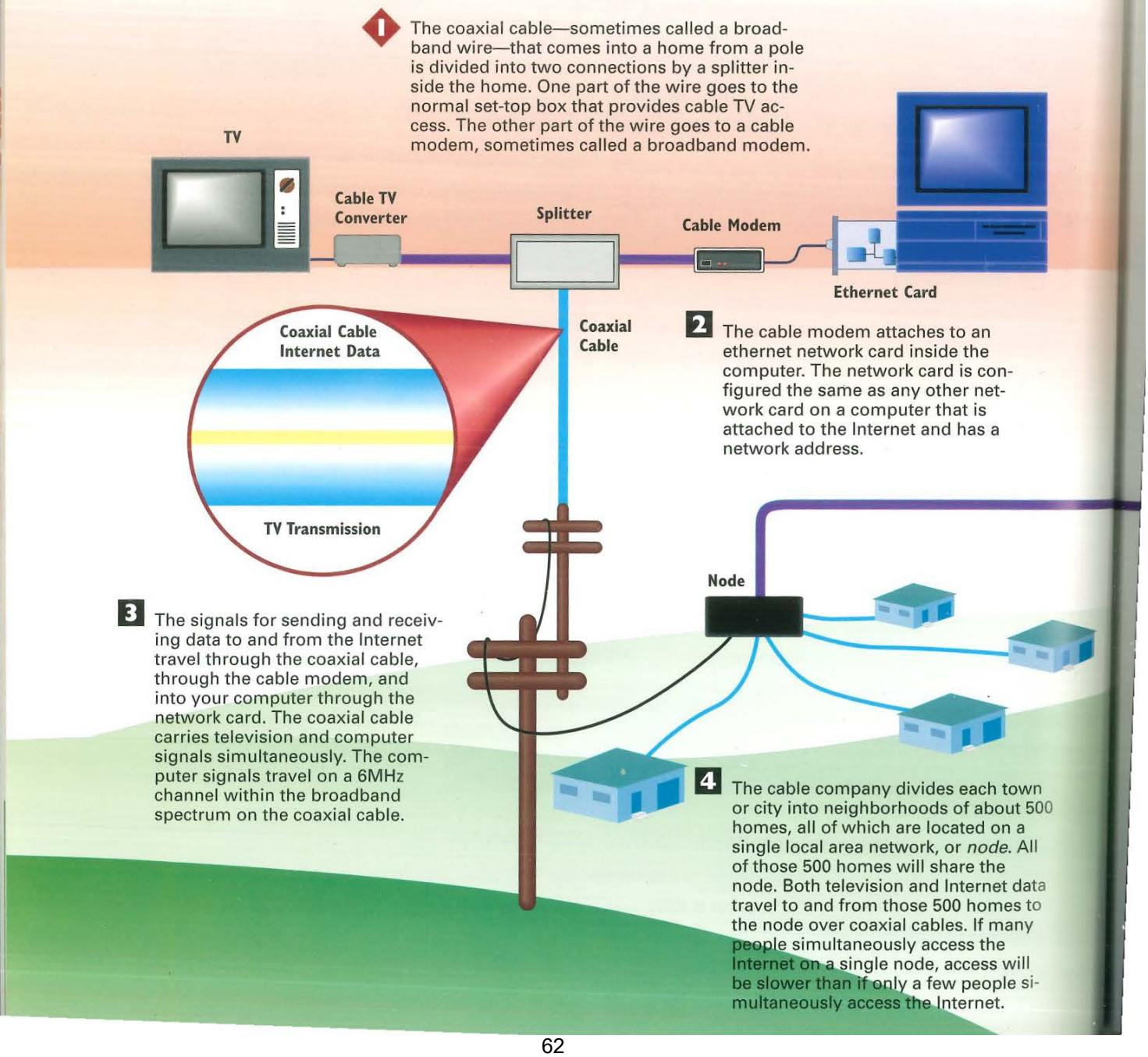
Eventually, the Net and TV will most likely merge in some way. Here's one example of how this might work. You may be able to watch a sporting event and simultaneously chat with others while you're watching—all on the same screen. Furthermore, when a batter came up to bat during a baseball game, you would be able to use the Net to get detailed statistics about the batter, and even past videos and highlights of his career.

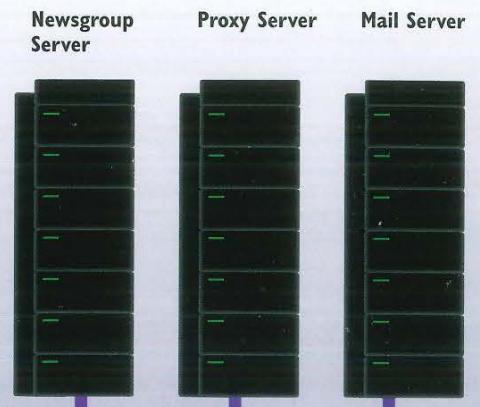
Television and the Internet have already begun to merge in a very real way, primarily through two different technologies: cable modems and TVs that connect to the Net, allowing you to browse the Net through your TV set, such as WebTV.

Cable modems offer extremely high-speed access to the Internet. They allow you to access the Net using the existing coaxial TV cable that comes into your house. Cable modems can receive data at up to 1.5 million bits per second and can send data at up to 300,000bps—far faster than normal modems and ISDN lines. They offer T1-like speeds, but at a fraction of the cost of T1 lines. They can deliver these high speeds because they are sent through high-capacity cable lines. Because Internet data and the normal cable signal coexist on the same lines, you can access the Net and watch TV simultaneously. The illustration in this chapter describes how MediaOne provides cable modem access to its customers.

A different technology allows you to browse the Web using your television set. A *set-top box* connects your TV to the Net via a modem, takes the signal from the modems, and sends it to the TV. A remote control-like device allows you to browse the Web while watching TV at the same time. As of this writing, the connection is still made via normal modem speeds, so access is far slower than with a cable modem. That, however, may change in the future.

How Cable Modems Work

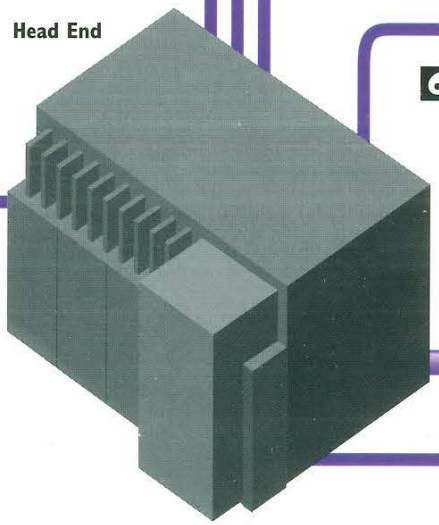




7 The head end also has high-speed Internet servers. A news server offers access to Internet Usenet newsgroups at a high speed because people access the server via the high-speed fiber-optic wires and coaxial cables instead of having to go out over the slower Internet. Additionally, *proxy servers* cache in their memory the most current versions of the more frequently accessed sites on the Internet. In that way, cable modem customers can get high-speed access to the sites because they get them via high-speed fiber-optic wires and coaxial cables instead of having to go out over the slower Internet.



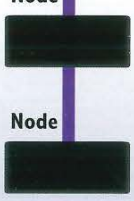
Head End



6 The head end receives television transmissions from satellites and has Internet access via high-speed links to the Internet. These feeds provide the cable programming and Internet access to cable customers.

High-Speed Internet Access

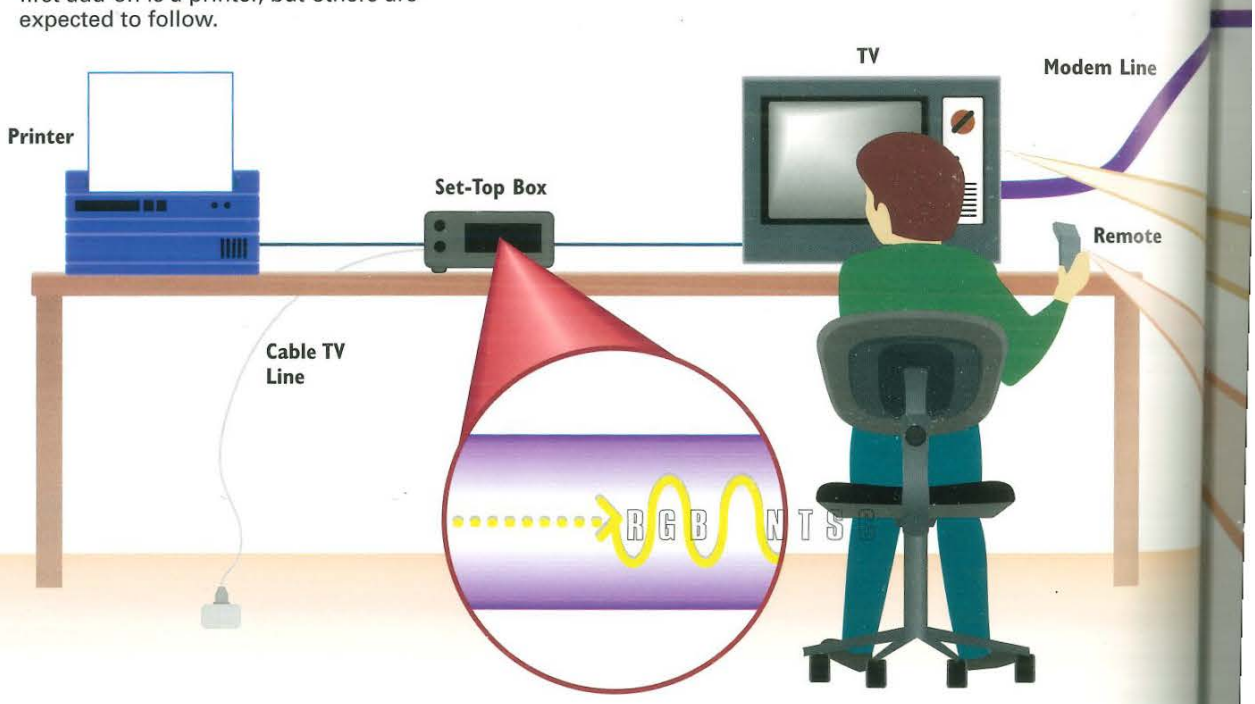
5 The nodes are connected via high-speed fiber-optic lines to a "head end" cable facility. A single head end typically handles all the nodes for four to ten towns. The head end is responsible for delivering television programming and Internet access to the cable customers.



How WebTV Works

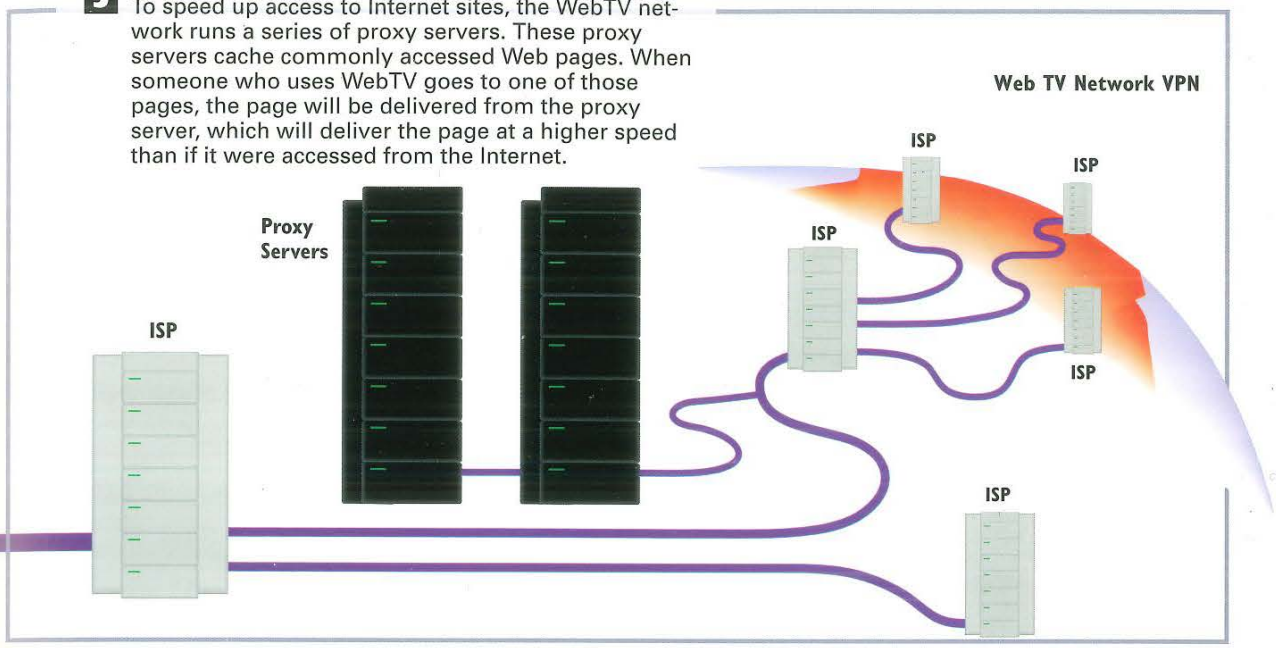
1 WebTV allows people to browse the Internet through a TV set and watch television at the same time. A special WebTV set-top box sits next to a television set. The box has normal connectors to the television and/or cable service, and special connectors to a modem and phone line. It also has the memory required to run a Web browser along with other hardware and software.

2 The WebTV set-top box has the capability to accept add-ons in a similar way to how computers can accept add-ons. The first add-on is a printer, but others are expected to follow.



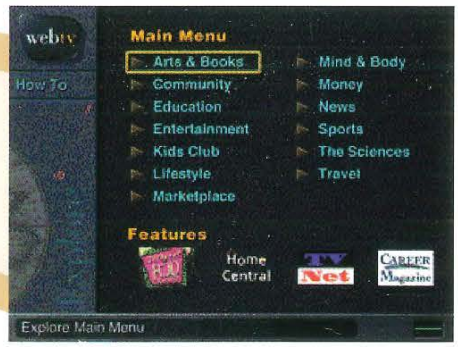
3 Television screens and computer monitors use different technologies to display pictures and information. Computer monitors use Red-Green-Blue (RGB) technology in which the three colors are combined to form all the colors you see on your monitor. Televisions use NTSC (National Television Standards Committee) technology. The WebTV set-top box converts the signal from the RGB of the Web to the NTSC used by televisions so that the signal will be as clear as possible.

5 To speed up access to Internet sites, the WebTV network runs a series of proxy servers. These proxy servers cache commonly accessed Web pages. When someone who uses WebTV goes to one of those pages, the page will be delivered from the proxy server, which will deliver the page at a higher speed than if it were accessed from the Internet.



4 When you want to access the Internet, the WebTV box connects to the WebTV network to get access to the Internet. The WebTV network is not a single network, but instead a Virtual Private Network (VPN). WebTV makes deals with ISPs (Internet service providers) across the country to provide access via local POPs (points of presence); encryption technology knits it all together into what appears to be a single network. This VPN has many Web servers on it to provide capabilities such as email and other services.

6 WebTV uses its own proprietary Web browser, which adheres to HTML standards and can run common plug-ins. The browser is designed for optimal display on a television set. Some sites, such as *TV Guide,* have developed Web sites that make the best use of the browser.





7 The ability to browse the Web is controlled by a simple remote control unit that allows people to scroll, move around the screen, and "click" on objects. The control also allows people to either browse the Web, watch television, or to view the Web and television simultaneously.

CHAPTER 14

How Network Computers (NCs) Work



THERE are those who believe that because of the widespread use of the Internet, the days of traditional computers are numbered. To some, it stands to reason that the Network Computer (NC) could take over. The NC will cost less than the traditional systems and will be able to do anything you need it to do because of its intimate tie-in to the Net. Some NCs are available now and their use is expected to grow in the future.

The NC can cost $500 or less because it doesn't require nearly as much hardware as a traditional system. It does not include a hard disk or CD-ROM drive, which cuts the cost of the computer significantly.

NCs don't require local storage or CD-ROM drives because the files and software aren't stored locally on the NC. Instead, the files and software are stored on the Internet, on special, secure NC servers. The software is Java-based, so it's portable across a variety of hardware. Rather than being written for a specific computer, such as the Macintosh or PC, Java programs will run on a variety of platforms.

NCs require that sophisticated servers be built to manage storing and serving up software and data. These servers would be accessible via the Internet. For security reasons, NCs may also be used with special "smart cards," credit-sized plastic cards with chips built into them.

To use an NC, you would put your card into a special reader, type in a personal identification number (PIN), and you would then be logged onto the network. A personalized desktop would appear on your NC that took into account your preferences about the kind of software you wanted to run, and the kind of information you wanted to receive. When you ran software, it would be delivered to you over the Internet by the NC server. When you saved files, you would save them to your own secure, private portion of the Internet.

There is no one standard for an NC. Oracle Corporation makes databases for the Internet and has been an early backer of NCs. Oracle would benefit by the NC's widespread adoption because the company would sell the server software to make NCs work.

NCs have a certain lure for corporations. They could cut the cost of buying and maintaining hardware and software dramatically by thousands of dollars per computer. Furthermore, because all software is delivered straight to the NC from a server, corporations could also ensure that everyone in the company used the same standardized software. It may well save significant time and money when systems needed to be configured and upgraded. Instead of upgrading 150 separate machines, a system administrator could merely install new software on the server and all the machines would automatically run that new software.

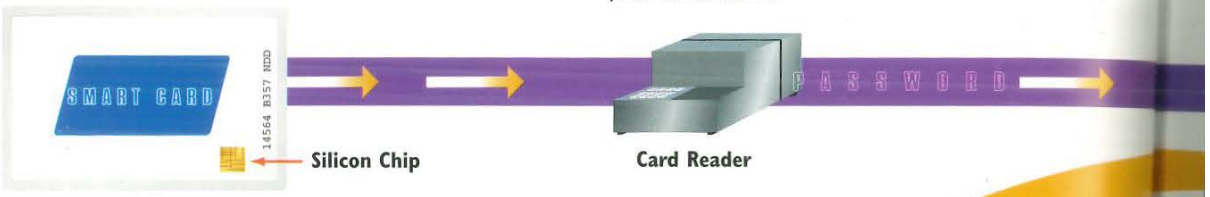
NCs could also be used in various consumer devices, as a way to bring Internet information into people's living rooms.

How Network Computers (NCs) Work

Network Computers (NCs) are low-cost computers with little or no local storage, and no CD-ROM drives or similar peripherals. The Internet handles all the software they run and the storage of data.

1 To use an NC, you'll need a smart card—a card with a computer chip embedded in it that identifies you to the network, much like a bank Automated Teller Machine (ATM) card identifies you to your bank's network.

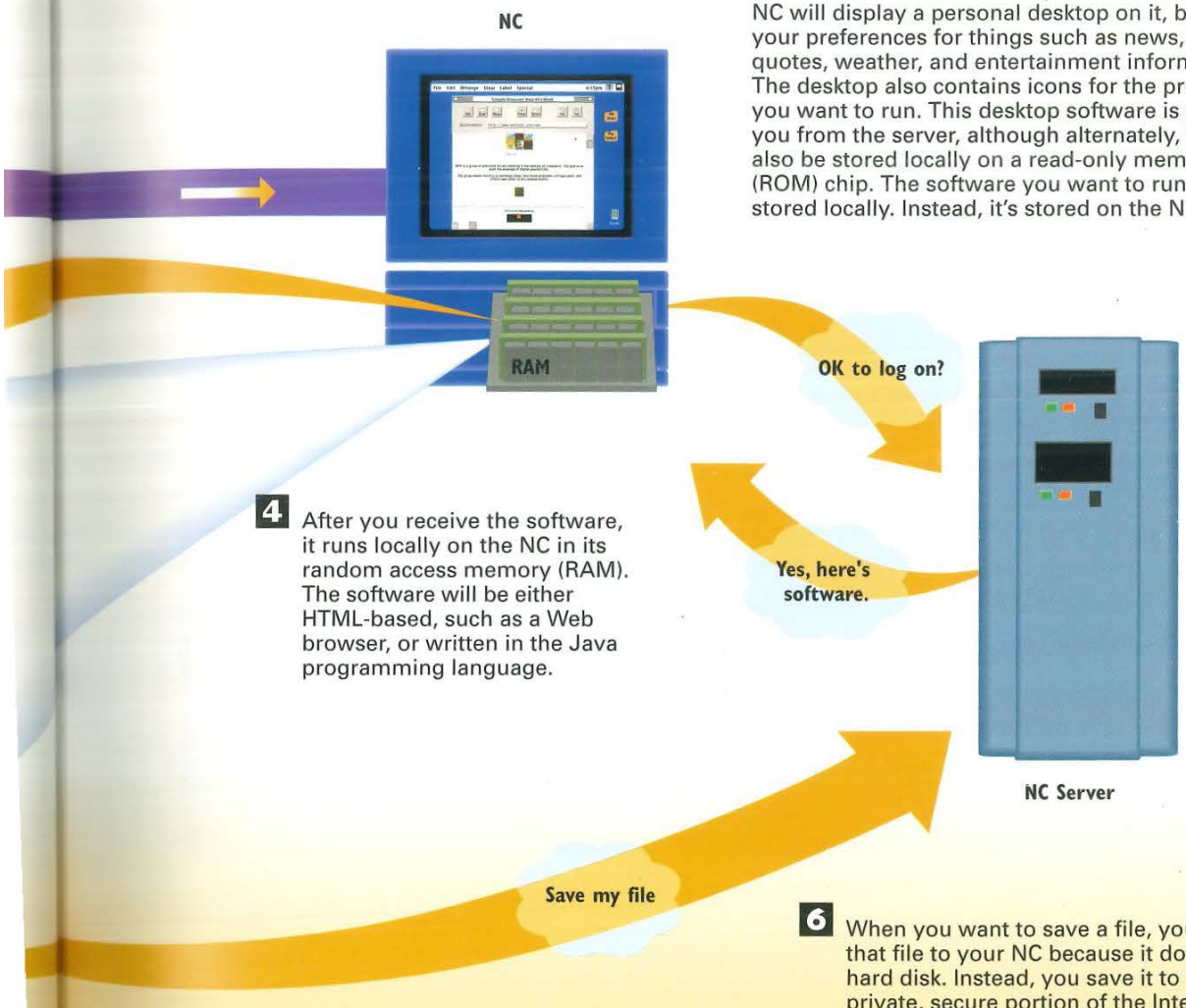
2 You put the NC card into a special card reader and will then be asked for a personal identification number (PIN). A server will check that your PIN matches your smart card.



Smart Card



5 You can use the software to browse the Web, use the Internet, send and receive email, play games, or use a word processor.



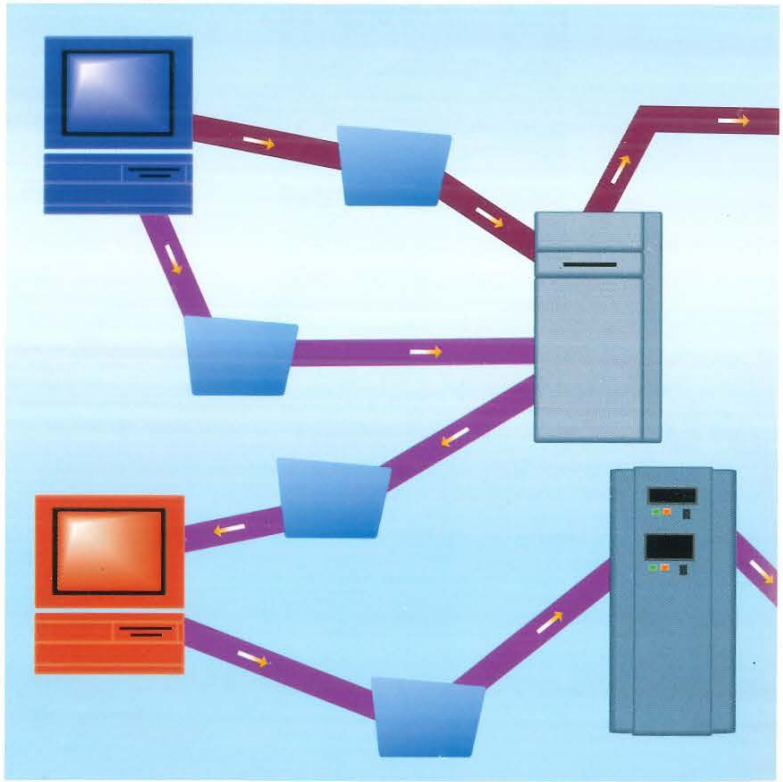
3 When the server authorizes you to use the NC, the NC will display a personal desktop on it, based on your preferences for things such as news, stock quotes, weather, and entertainment information. The desktop also contains icons for the programs you want to run. This desktop software is sent to you from the server, although alternately, it could also be stored locally on a read-only memory (ROM) chip. The software you want to run isn't stored locally. Instead, it's stored on the NC server.

4 After you receive the software, it runs locally on the NC in its random access memory (RAM). The software will be either HTML-based, such as a Web browser, or written in the Java programming language.

6 When you want to save a file, you don't save that file to your NC because it doesn't have a hard disk. Instead, you save it to your own private, secure portion of the Internet that is managed by the NC server. When you want to use the file again, you'll call it up from your own private area.

CHAPTER

15 How the Digital Subscriber Line (DSL) Works



IN recent years, a plethora of options could be found for gaining high-speed access to the Internet. Among the newest, which may eventually prove to be one of the most popular, is the Digital Subscriber Line, or DSL. Several kinds of DSL technologies are available, but they all work on the same principles. They enable you to use your existing telephone lines to access the Internet at very high speeds—potentially as high as 55Mbps—using VDSL (Very high data rate DSL), although that kind of extremely high-speed connection is used primarily by large corporations because of its cost. Home users will probably get speeds of 1.5Mbps, which is still very fast—about 50 times the speed of a 28,800bps modem. Additionally, DSL technologies will enable you to use the Internet at high speeds and simultaneously talk on your telephone, using the same single phone line.

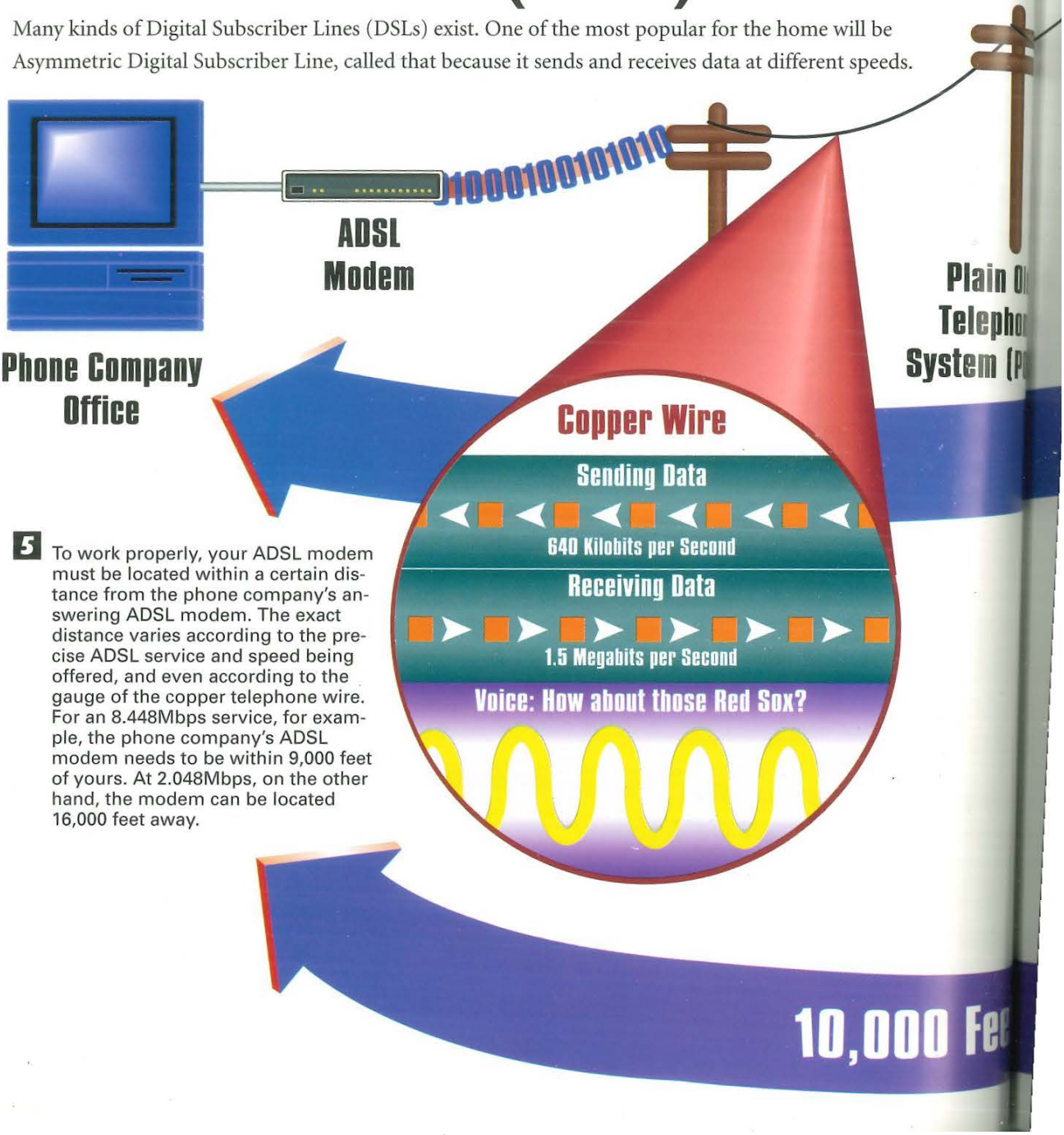
DSL technologies require that DSL modems be used on each end of the phone line. In fact, the term DSL doesn't really refer to a phone line, because an ordinary, existing copper phone line can be used for DSL. Instead, it refers to the DSL modems themselves. More confusing still, DSL "modems" aren't really modems at all. Traditional modems modulate the digital signals from your computer into analog signals so that they can be sent over telephone lines, and then they demodulate them back into digital signals so that your computer can understand them—hence the name modem, from the terms "MOdulate-DEModulate." DSL modems don't convert signals from digital to analog. Instead, they transmit and receive all data as digital signals. Because the signals don't need to be converted and can be sent as digital data instead of analog data, data can be sent at much higher speeds than data sent via traditional modems.

DSL technology has one drawback; it requires that your house (and DSL modem) be located within a certain distance from the telephone company office and its DSL modem. In cities, this should not be a major problem, but it could be a problem in rural areas. The exact distance required will depend on the kind and speed of DSL service you use. Higher speeds require that you be closer to the phone company office.

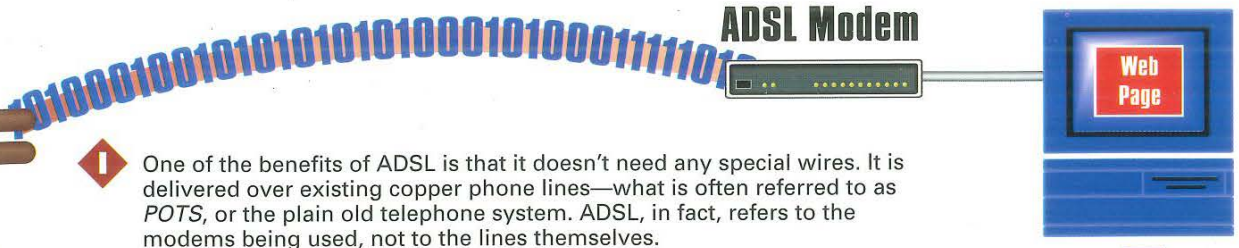
Although many kinds of DSL service are available, probably the most popular for people at home will be Asymmetric Digital Subscriber Line (ADSL) and DSL Lite. With them, you can receive data at a faster rate than you can send data—for example, you can receive data at 1.5Mbps and send data at 640Kbps.

How Asymmetric Digital Subscriber Line (ADSL) Works

Many kinds of Digital Subscriber Lines (DSLs) exist. One of the most popular for the home will be Asymmetric Digital Subscriber Line, called that because it sends and receives data at different speeds.

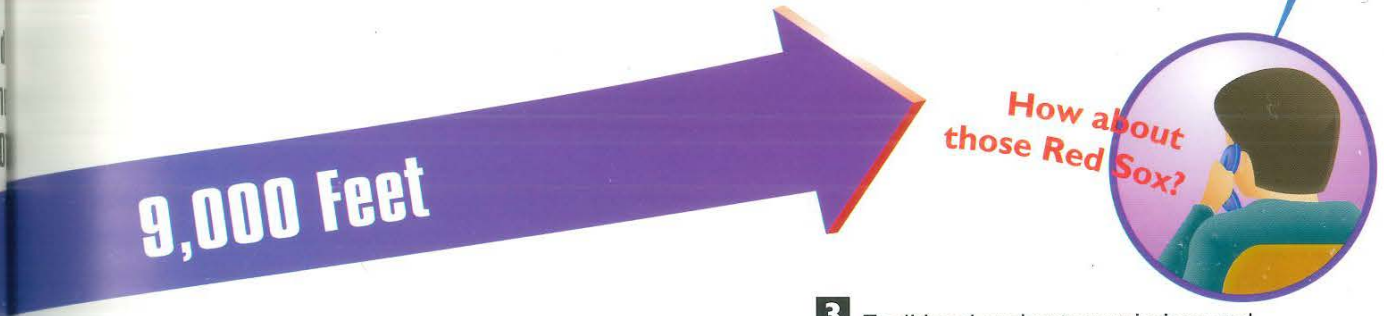


5 To work properly, your ADSL modem must be located within a certain distance from the phone company's answering ADSL modem. The exact distance varies according to the precise ADSL service and speed being offered, and even according to the gauge of the copper telephone wire. For an 8.448Mbps service, for example, the phone company's ADSL modem needs to be within 9,000 feet of yours. At 2.048Mbps, on the other hand, the modem can be located 16,000 feet away.



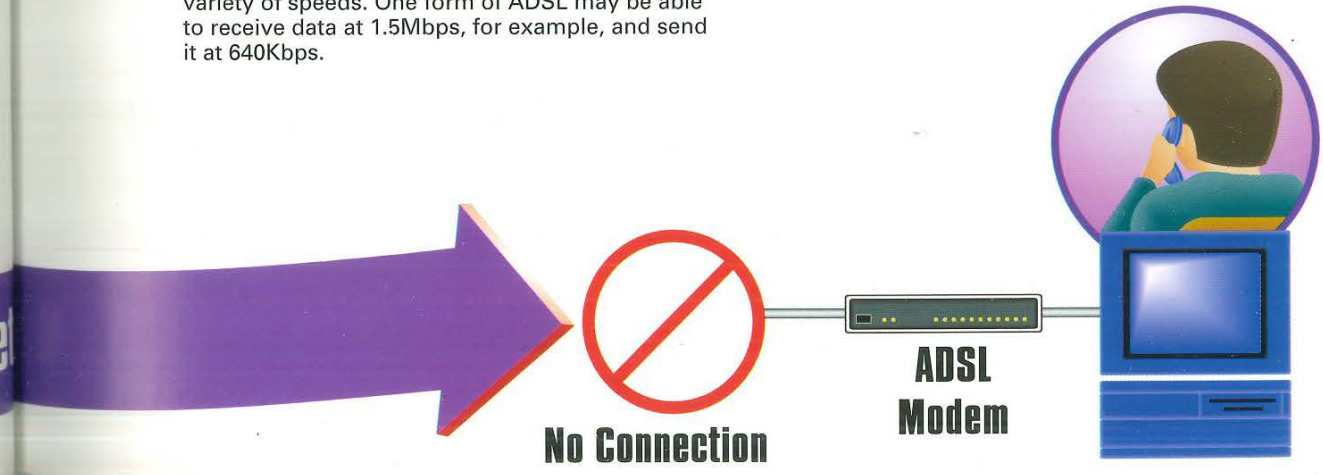
1 One of the benefits of ADSL is that it doesn't need any special wires. It is delivered over existing copper phone lines—what is often referred to as *POTS*, or the plain old telephone system. ADSL, in fact, refers to the modems being used, not to the lines themselves.

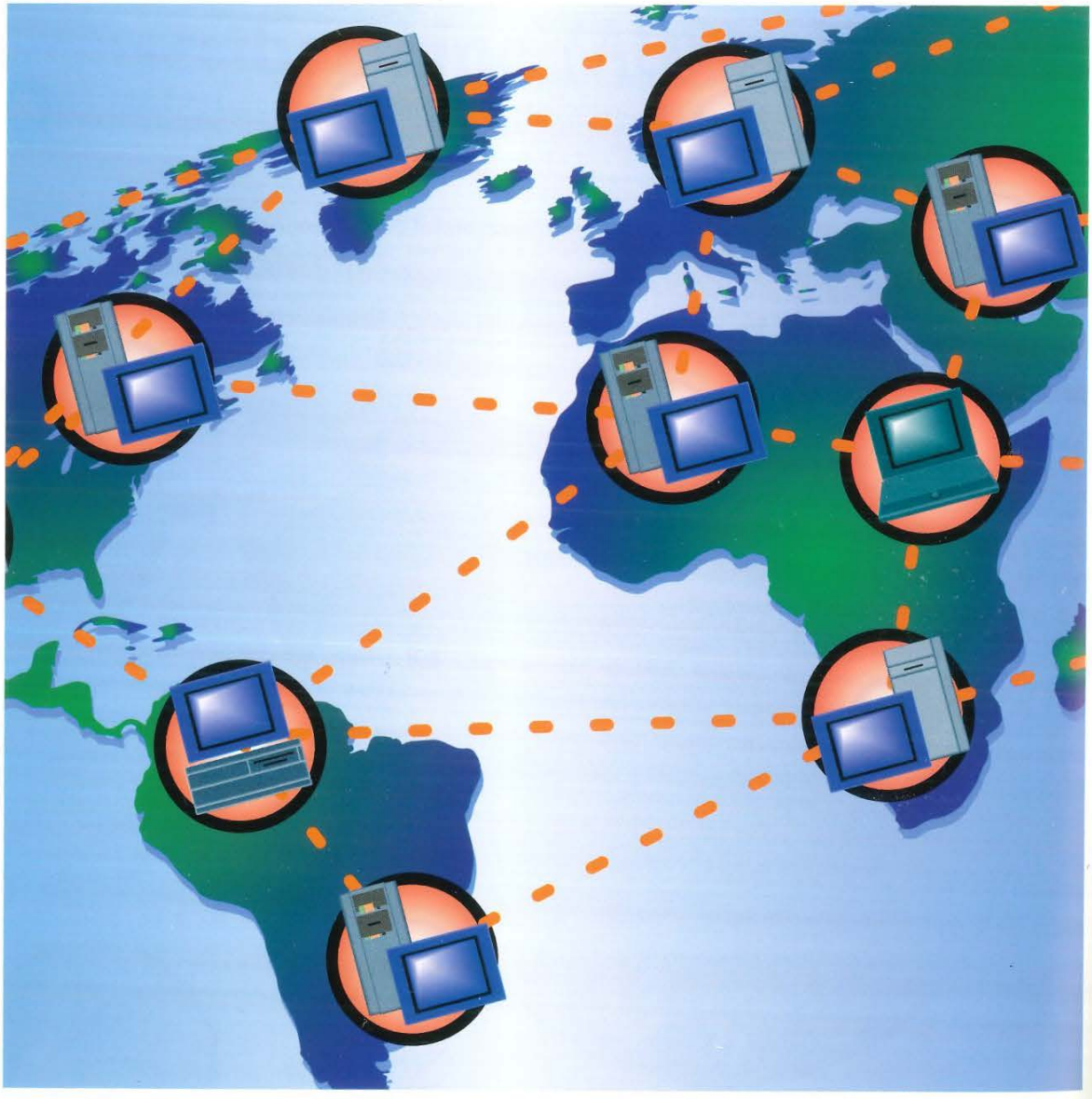
2 ADSL service requires that ADSL modems be on each end of the connection—in your home and at the telephone company office. ADSL modems are different than traditional modems. Traditional modems take the digital signals from your computer, translate them into analog signals, and send them over the phone line as analog signals. The receiving traditional modem then takes those signals and translates them back into digital information. ADSL modems instead send and receive all data as digital data—no translation to analog signals ever takes place.



3 Traditional analog transmissions and voice calls sent over copper telephone wires use only a small portion of the potential bandwidth that can be transmitted over copper telephone wires. ADSL enables people to talk on the telephone and use the Internet at high speeds simultaneously—all over a single phone line.

4 ADSL divides the phone line into three channels: one for receiving data, one for sending data, and one for talking over the telephone. That means that a single phone line can be used for browsing the Internet and talking over the telephone simultaneously. The phone line is not always physically separated into these channels. Rather, modulation techniques can be used to separate the three types of signals: voice, sending, and receiving. The sending and receiving channels can be divided into a variety of speeds. One form of ADSL may be able to receive data at 1.5Mbps, for example, and send it at 640Kbps.





PART 4

COMMUNICATING ON THE INTERNET

Chapter 16: How Email Works
78

Chapter 17: Email Spam and Blocking
90

Chapter 18: How Usenet Newsgroups Work
96

Chapter 19: How Internet Chat and Instant Messaging Work
102

Chapter 20: Making Phone Calls on the Internet
108

FROM its very earliest days, the Internet has been concerned primarily with one task—making it easier for people to communicate with one another using computers. The Internet was created to enable university researchers to share their thoughts, work, and resources, and for military people to communicate with each other in case of war and even, theoretically, a nuclear attack.

Today, more than two decades after the inception of the first networks that grew into the Internet, it is still primarily a communications medium. Millions of people from all over the world share their thoughts, hopes, work, gossip, and comments on the wires and computers that make up the Internet. Many of the means of communication, such as electronic mail, have changed very little in the past 20 years. Yet other, entirely new ways of communicating have been devised, such as enabling you to use the Internet as your telephone, completely cutting out long-distance charges, even if you're calling to the other side of the world. Technologies enable people to communicate privately, one-on-one; others allow for vast discussion groups that span the globe, and still others allow for both private communication with one person and public communication with large groups.



In this section of the book, we'll look at the main ways that people communicate on the Internet.

In Chapter 16, "How Email Works," we'll take a long look at what continues to be the most popular way for people to communicate on the Internet—electronic mail, or email. Email remains possibly the greatest use of the Internet and is used for business and personal communication. We'll see how email gets routed from your computer through the maze of wires that makes up the Internet, and then ends up in the proper recipient's in-box. We'll look at all the elements of a mail message and learn how you can send binary files, such as pictures and sounds, through email. We'll also explore mailing lists, where you can subscribe to any one of thousands of public discussions via email, or receive what are essentially electronic newsletters delivered to your email in-box. And we'll look at how you can look up anyone's email address using white page directories that use a technology called the Lightweight Access Directory Protocol (LADP).

Chapter 17, "Email Spam and Blocking," covers one of the most controversial modes of communication on the Internet—the use of *Spam*, the equivalent of junk mail via email, that is sent to millions of people each day. Although Spam is a problem on newsgroups as well as email, email is the greatest area of controversy. Spam annoys people and wastes their time while they clean out their mailboxes; it clogs the Internet so that other messages may be delivered late—or not at all; and it can be delivered by piggybacking onto other people's email servers, costing them money. The chapter looks at how Spam is sent, as well as ways that it can be blocked.

Chapter 18, "How Usenet Newsgroups Work," explores *Usenet newsgroups*—public discussion groups in which anyone can participate. Many thousands of these groups focus on every subject conceivable. You'll see how newsgroups work and how you can decipher their often arcane names.

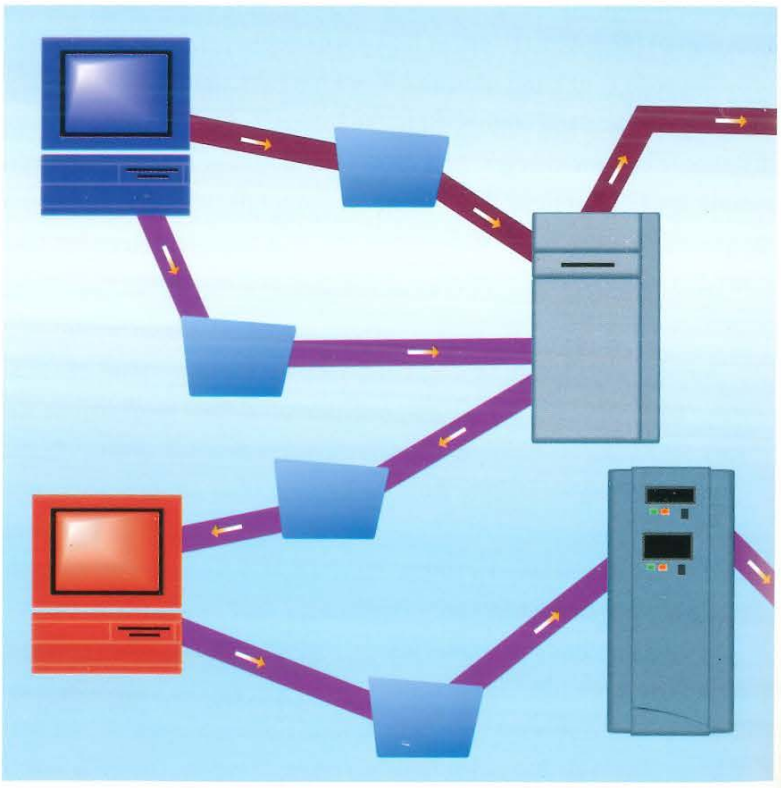
Chapter 19, "How Internet Chat and Instant Messaging Work," covers the different ways people can "chat" on the Internet. The most popular one is *IRC*, or *Internet Relay Chat*. IRC is a way for people to publicly chat with one another on the Internet. They don't actually speak, but type comments on their keyboards instead, and then people all over the world can read and respond to them. The chapter also covers *instant messaging*—ways in which people chat one-on-one with others when they log onto the Internet.

Finally, Chapter 20, "Making Phone Calls on the Internet," details one of the more intriguing new uses of the Internet—using it as your telephone. Today you can dial into your local Internet provider with your computer, and if you have the right hardware and software, you can talk with anyone similarly connected to the Internet anywhere in the world, all without paying long-distance telephone charges. And it covers a new kind of low-cost telephone service that has sprung up: IP Telephony. In *IP Telephony*, you talk over your telephone as you would normally—the phone call appears no different to you. But the Internet, rather than the normal phone system, is used to make the connection to the person you're calling.



CHAPTER

16 How Email Works



ELECTRONIC mail, or email, may be the most heavily used feature of the Internet. You can use it to send messages to anyone who is connected to the Internet or connected to a computer network that has a connection to the Internet, such as an online service. Millions of people send and receive email every day. Email is a great way to keep up with far-flung relatives, friends, co-workers in different branches of your company, and colleagues in your field.

Email messages are sent in the same way as most Internet data. The TCP protocol breaks your messages into packets, the IP protocol delivers the packets to the proper location, and then the TCP reassembles the message on the receiving end so that it can be read.

You can also attach binary files, such as pictures, videos, sounds, and executable files to your email messages. Because the Internet isn't able to directly handle binary files in email, the file first must be encoded in one of a variety of encoding schemes. Popular schemes are MIME and uuencode. The person who receives the attached binary file (called an attachment) must decode the file with the same scheme that was used to encode the file. Many email software packages do this automatically.

When you send email to someone on the Internet, that message often has to travel through a series of networks before it reaches the recipient—networks that might use different email formats. Gateways perform the job of translating email formats from one network to another so that the messages can make their way through all the networks of the Internet.

A mailing list is one of the most intriguing uses of email. It connects a group of people who are interested in the same topic, such as Japanese cartoons or home schooling. When one person sends email to the mailing list, that message is automatically sent to everyone on the list. You can meet others and talk to them on a regular basis about your shared interests, hobbies, or professions. To get onto a mailing list, you send an email note to the mailing list administrator and include your email address.

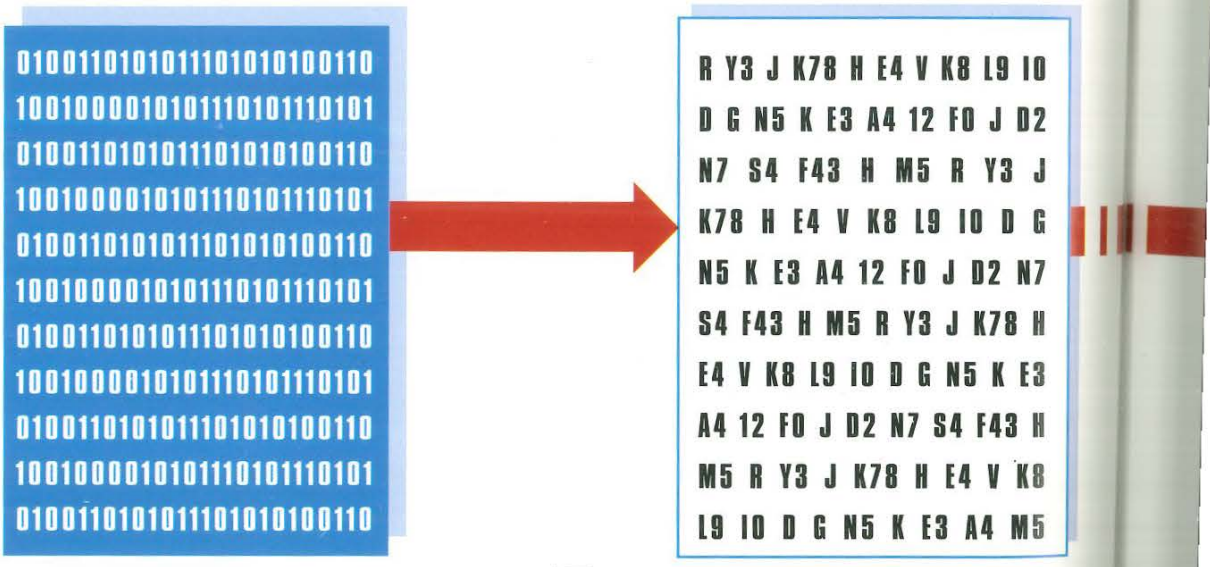
Mailing lists can be moderated or unmoderated. A *moderated mailing list* is screened by the list administrator, who may kill duplicate messages or messages that are not related to the list's theme. An *unmoderated mailing list* is wide open; all mail sent to it is automatically sent to everyone on the list.

Often, when you want to subscribe to a mailing list, you send a message to a computer instead of a person. That computer, known as a list server (also called a *listserv*), reads your email and automatically subscribes you to the list. You can unsubscribe to the list in the same way.

In the past, it was very difficult to find the email address of someone if you knew only his or her name. These days, it's not so hard. A variety of "white page" directories have sprung up on the Internet that enable you to look up people's email addresses easily. These sites mainly use a standard called the Lightweight Directory Access Protocol (LDAP), which enables you to find people's email addresses without even having to visit a Web site. Using the protocol, you can search for email addresses on the Internet from directly within your email program.

Anatomy of a Mail Message

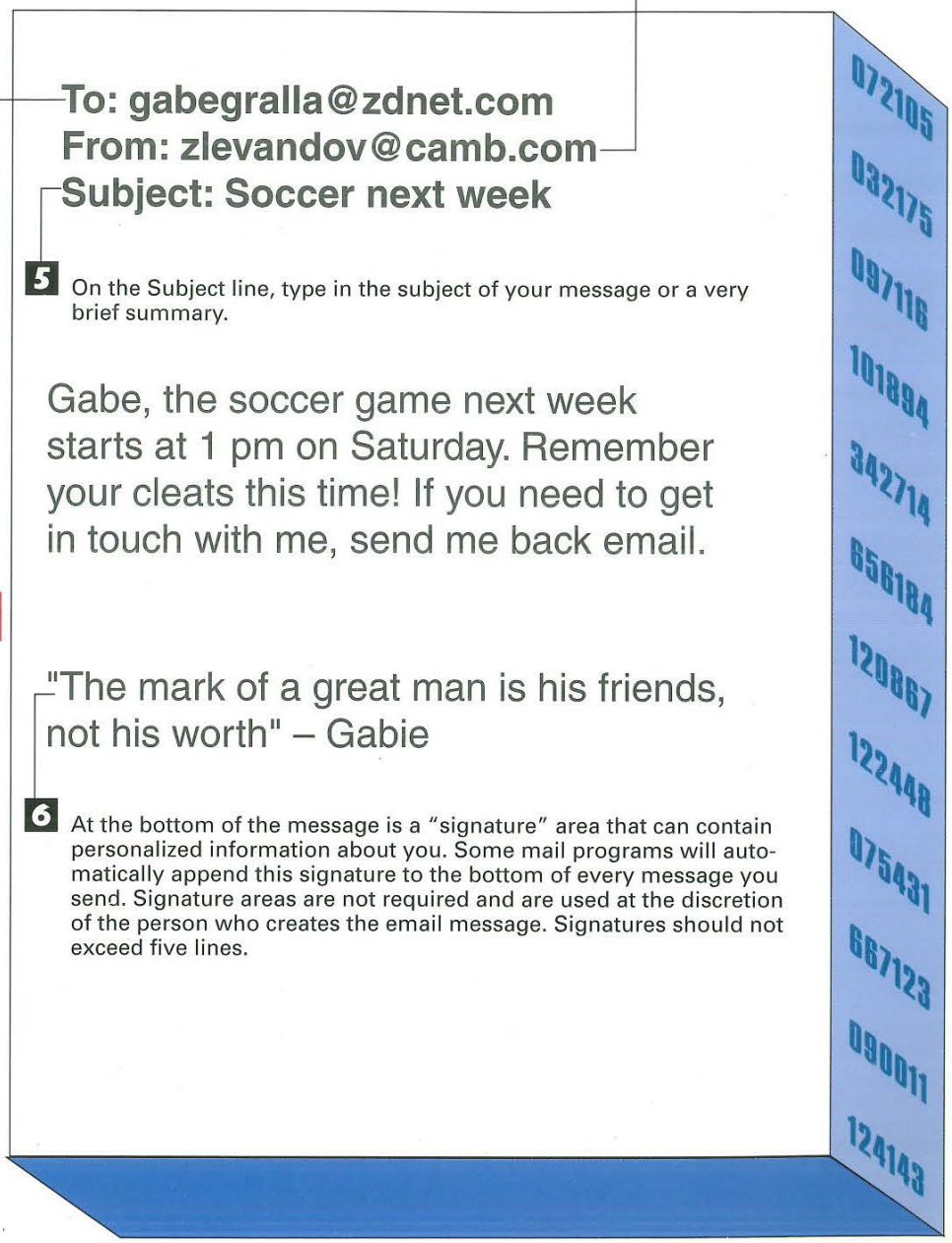
1 An email message is made up of binary data, usually in the ASCII text format. ASCII is a standard that enables any computer, regardless of its operating system or hardware, to read the text. ASCII code describes the characters you see on your computer screen.



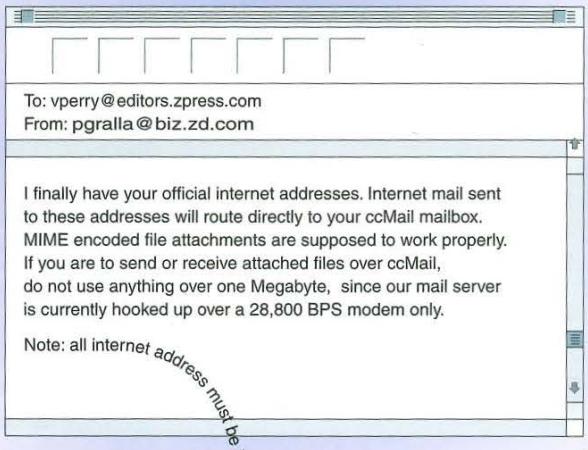
2 You can also attach pictures, executable programs, sounds, videos, and other binary files to your email message. To do this, you'll have to encode the file in a way that will enable it to be sent across the Internet. The receiver will also have to be able to decode the file after it is received. A variety of encoding schemes can be used. Some email software will automatically do the encoding for you and will also do the decoding on the receiving end.

3 In the To line, type in the email address of the person to whom you're sending a message. The address must be typed in following very strict rules. If you get a single letter or the syntax wrong, your message won't get to the intended recipient.

4 Your email address will appear on the From line. Using this address, the recipient of your message will be able to respond to you.



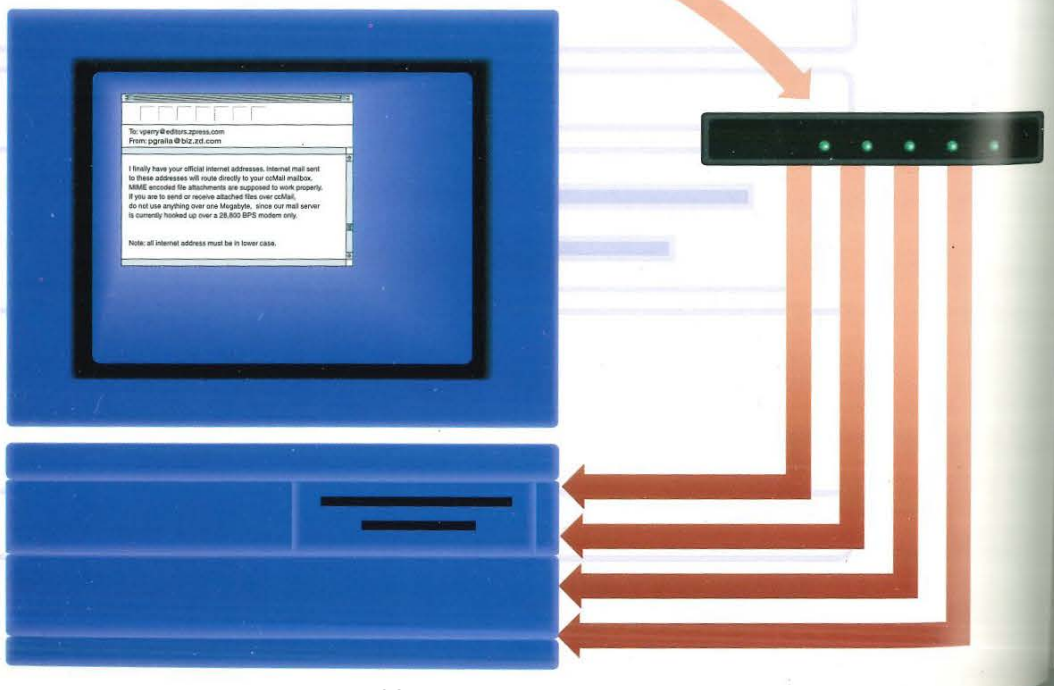
How Email Is Delivered over the Internet





1 After you create and send an email message, it is sent as a stream of packets using the Internet's TCP/IP protocol. Each packet bears the address of the destination.

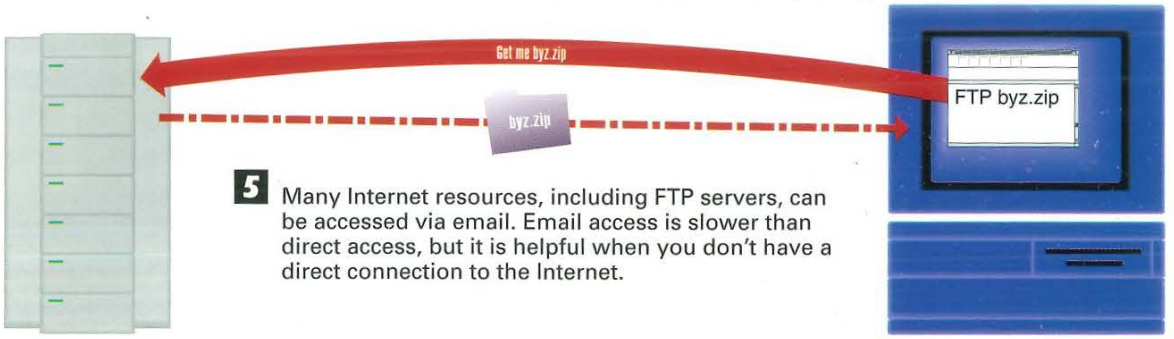
2 Routers on the Internet look at the addresses in each packet and send the packets on the best path to get there. Many factors go into how the packets are routed, including the traffic volume on different backbones. Each packet may take a different route, so the mail packets may arrive at the destination out of order.



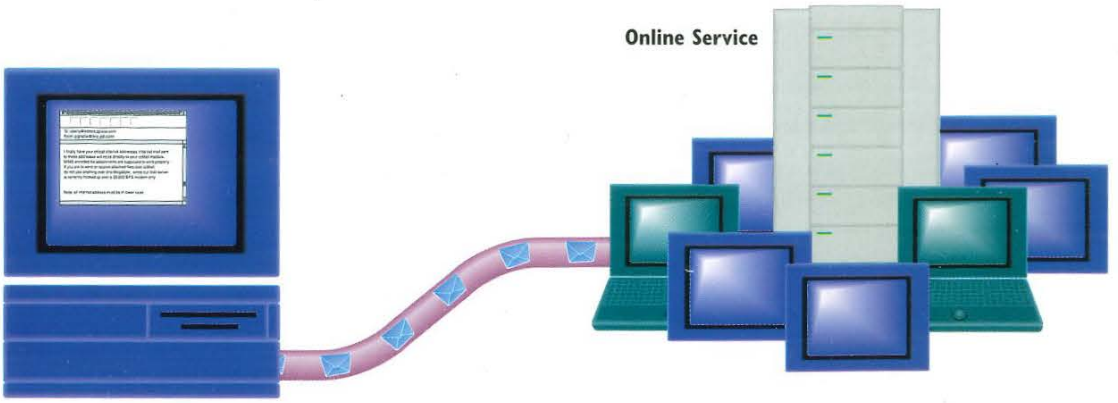
3 When all the packets have been received at the receiving address, they are recombined into an email message that the recipient can read.



4 Using a mailing list, you can send a single message to a group of people. A *mail reflector* is the program that runs on an Internet computer and routes mail to members of a mailing list. In a different kind of mailing list, known as a *listserv*, you subscribe to a mailing list by sending it your email address. You will get every message that everyone sends to the list. In another kind of email mailing list you can subscribe to, you only receive mail that a single person sends; only that person can send to the list. Often, electronic newsletters are distributed in this way.

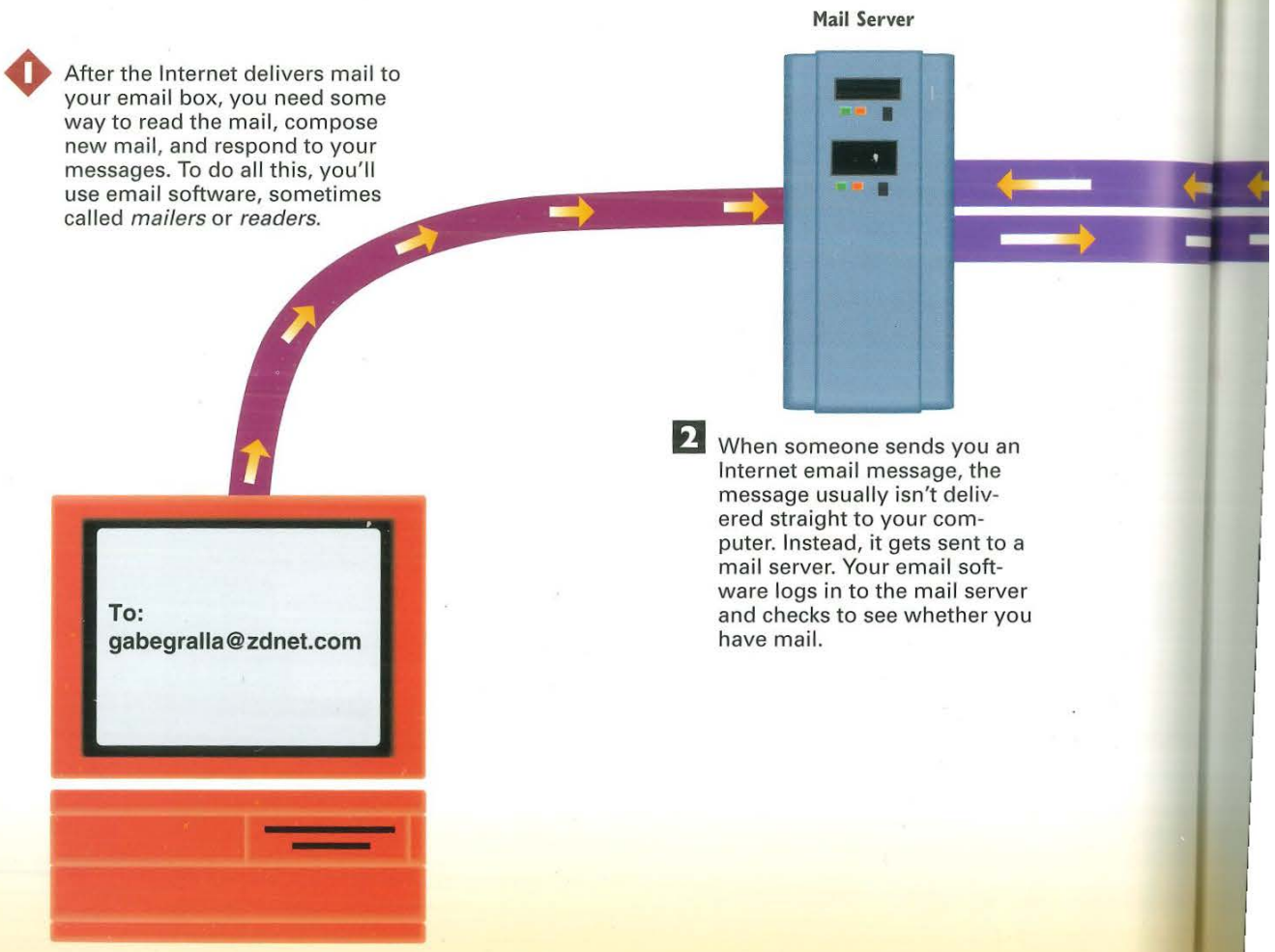


5 Many Internet resources, including FTP servers, can be accessed via email. Email access is slower than direct access, but it is helpful when you don't have a direct connection to the Internet.



6 Using the Internet, email can be exchanged among all the major online services, computer bulletin boards, and other networks. From the Internet, you can send email to any of those networks—and from any of those networks, mail can be sent to the Internet. When mail is sent from one of those networks to another, it often must pass through the Internet as a way of routing the mail.

How Email Software Works

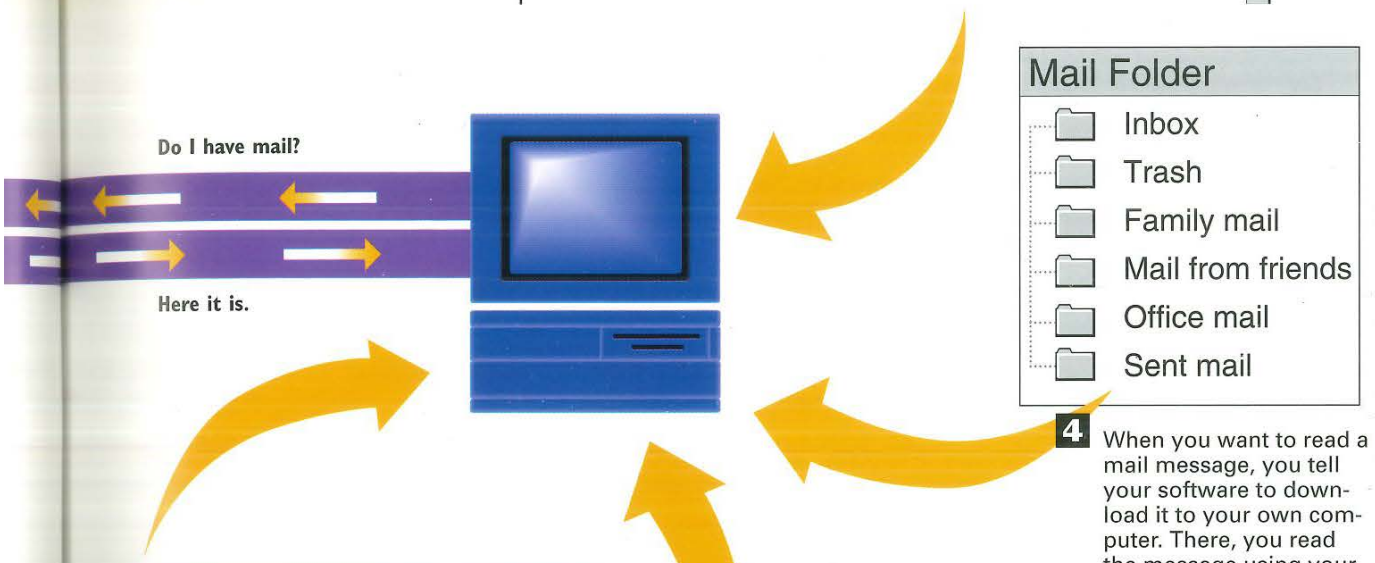


1 After the Internet delivers mail to your email box, you need some way to read the mail, compose new mail, and respond to your messages. To do all this, you'll use email software, sometimes called *mailers* or *readers*.

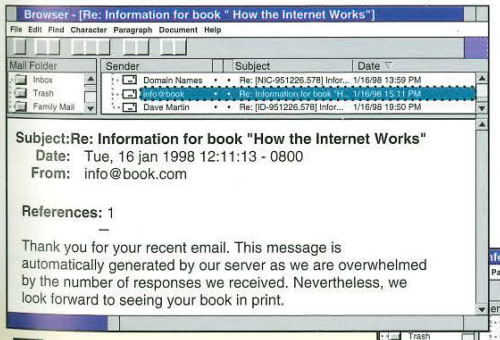
2 When someone sends you an Internet email message, the message usually isn't delivered straight to your computer. Instead, it gets sent to a mail server. Your email software logs in to the mail server and checks to see whether you have mail.

3 If you have new mail, you'll see a list of your new mail messages when you log in to the server. You'll often see the name of the sender, the subject of the message, and the date and time that the message was sent.

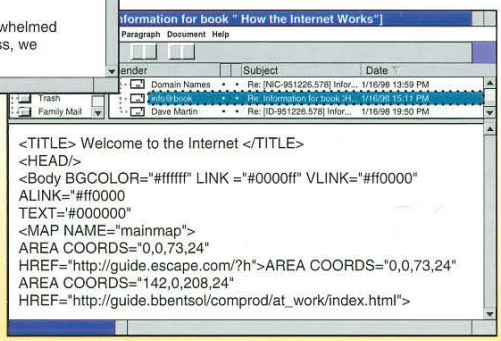
Sender	Subject	Date
News Flash	Re: Access to newsgroups f...	12/27/97 16:28 PM
Bill Jones	Re: WebWatch.	1/3/98 19:49 PM
Mary Peterson	tt#8522p Mail Server Upgra...	1/6/98 13:15 PM
Mary Peterson	FAQ: Club Member Discoun...	1/9/98 20:32 PM
Domain Names	Re: [ID-951226.578] Infor...	1/16/98 13:59 PM
info@book.com	Re: Information for book "H...	1/16/98 15:11 PM
Dave Martin	Re: [ID-951226.578] Infor...	1/17/98 19:50 PM



4 When you want to read a mail message, you tell your software to download it to your own computer. There, you read the message using your mail reader, and then you can file it, delete it, or respond to it.



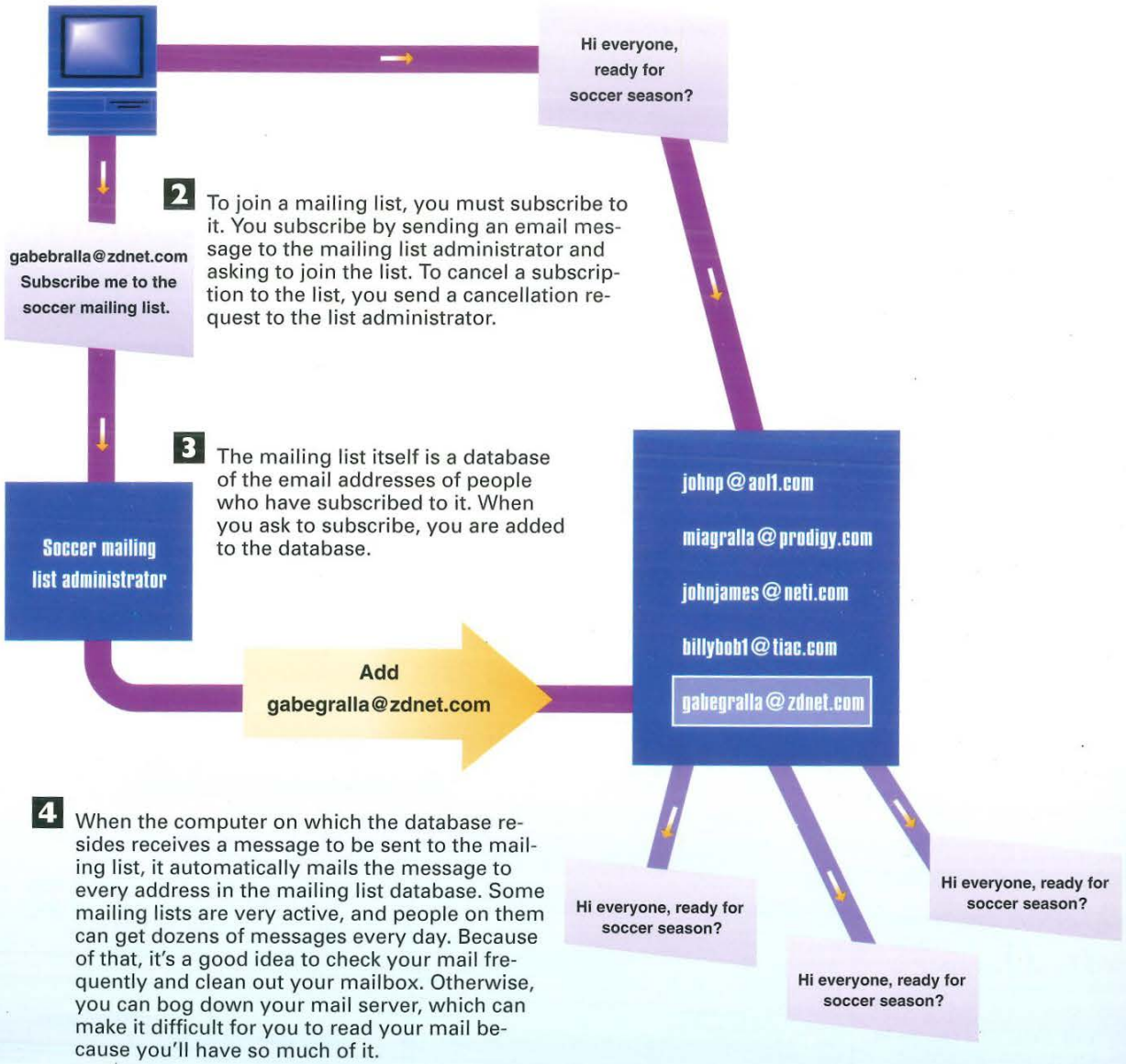
5 Email software enables you to do things such as create folders for storing mail, search through your messages, keep an address book of people to whom you send mail, create group mailing lists, create and add a signature file, and more.



6 The newest generation of email software (found in browsers such as Netscape Navigator) will read HTML-based pages that are sent to you so that you can receive, in your mailbox, fully formatted Web pages. When you click the links in them, your browser will launch and visit the page it is linked to.

How a Mailing List Works

1 Mailing lists are a way for groups of people to have public discussions via email. After you join a mailing list, every message you write to the list can be read by everyone on the list.

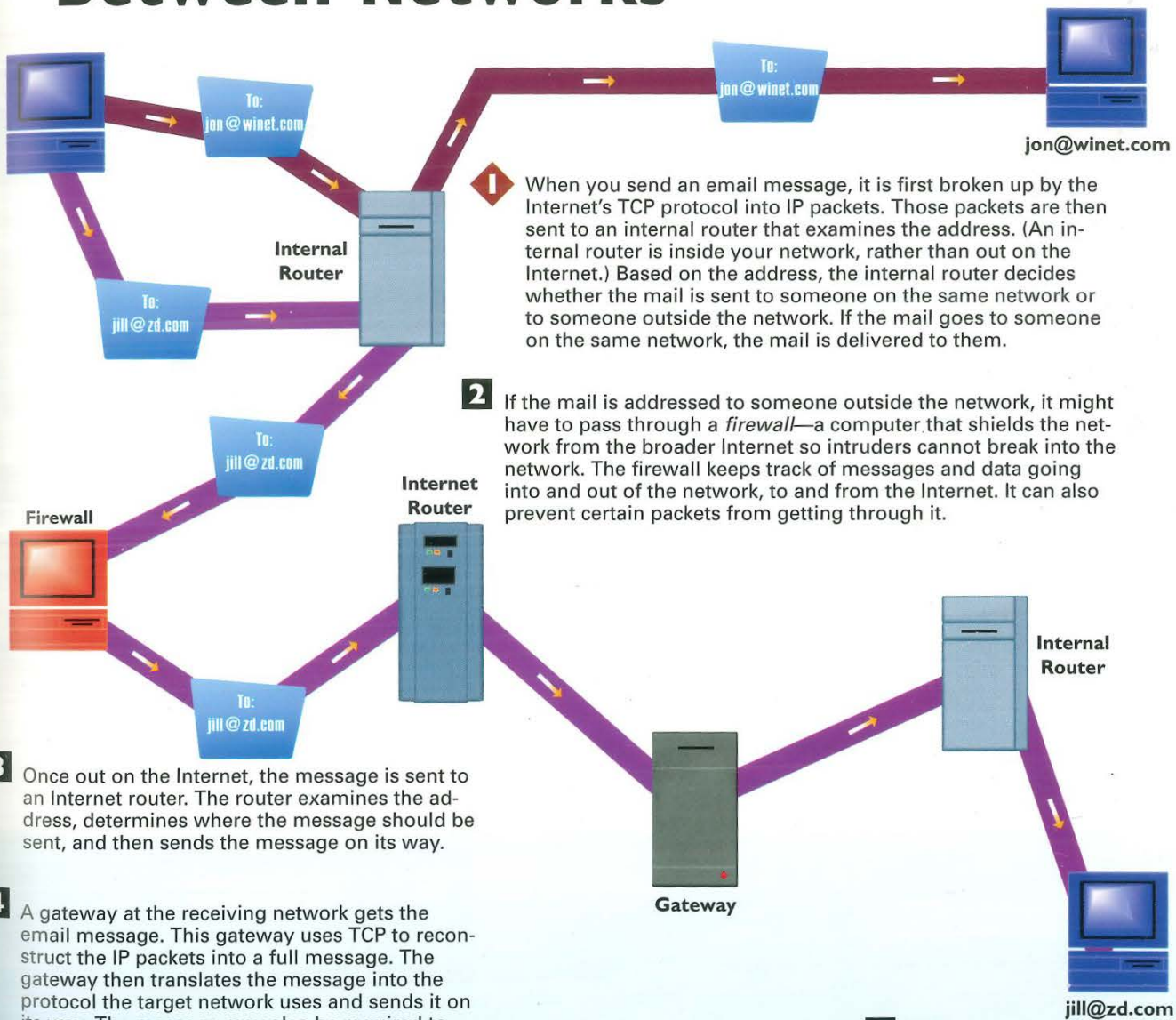


2 To join a mailing list, you must subscribe to it. You subscribe by sending an email message to the mailing list administrator and asking to join the list. To cancel a subscription to the list, you send a cancellation request to the list administrator.

3 The mailing list itself is a database of the email addresses of people who have subscribed to it. When you ask to subscribe, you are added to the database.

4 When the computer on which the database resides receives a message to be sent to the mailing list, it automatically mails the message to every address in the mailing list database. Some mailing lists are very active, and people on them can get dozens of messages every day. Because of that, it's a good idea to check your mail frequently and clean out your mailbox. Otherwise, you can bog down your mail server, which can make it difficult for you to read your mail because you'll have so much of it.

How Email Is Sent Between Networks



1 When you send an email message, it is first broken up by the Internet's TCP protocol into IP packets. Those packets are then sent to an internal router that examines the address. (An internal router is inside your network, rather than out on the Internet.) Based on the address, the internal router decides whether the mail is sent to someone on the same network or to someone outside the network. If the mail goes to someone on the same network, the mail is delivered to them.

2 If the mail is addressed to someone outside the network, it might have to pass through a *firewall*—a computer that shields the network from the broader Internet so intruders cannot break into the network. The firewall keeps track of messages and data going into and out of the network, to and from the Internet. It can also prevent certain packets from getting through it.

3 Once out on the Internet, the message is sent to an Internet router. The router examines the address, determines where the message should be sent, and then sends the message on its way.

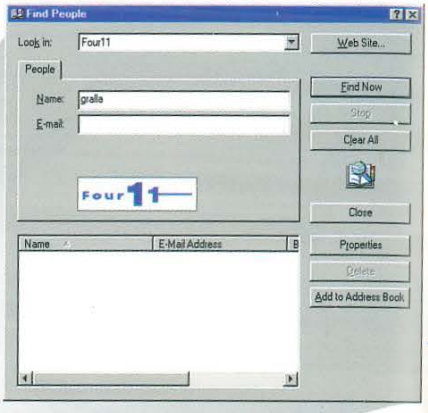
4 A gateway at the receiving network gets the email message. This gateway uses TCP to reconstruct the IP packets into a full message. The gateway then translates the message into the protocol the target network uses and sends it on its way. The message may also be required to pass through a firewall on the receiving network.

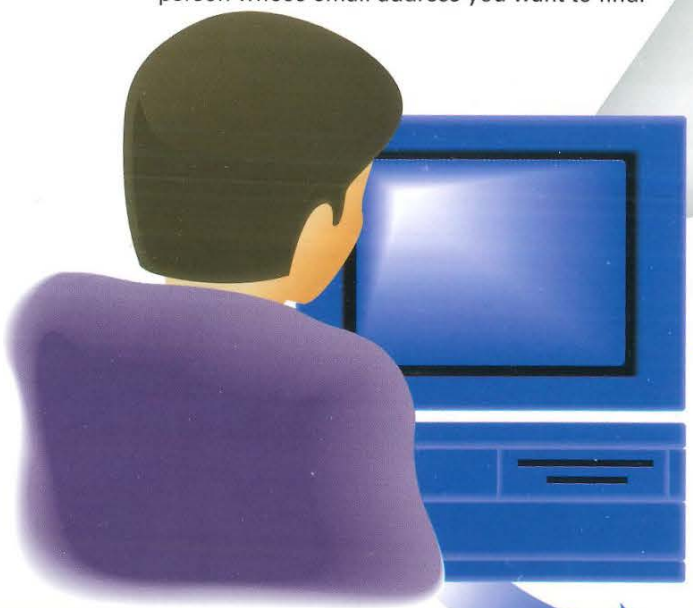
5 The receiving network examines the email address and sends the message to a specific mailbox.

How White Page Directories (LDAP) Work

Millions of people have email addresses on the Internet, but it's always been difficult to find an email address if you know only a name. The Lightweight Directory Access Protocol (LDAP) makes creating White Page-style directories easier, so it's now possible to look up people's addresses if you know only their name.

2 Like much other Internet technology, LDAP works on a client/server model. The client runs on a local computer. Many kinds of clients are available for many kinds of computers. To find someone's email address, you run the client on your computer and type in the name of the person whose email address you want to find.





HERE ARE THE RESULTS



7 The client receives the results and displays them on the local computer.

1 An LDAP directory is a database that follows LDAP rules and protocols for organizing its information. It is located on an Internet server. A company or business maintains the server and is responsible for keeping the information on it up-to-date by adding new names, deleting old names, and updating existing names.

Lightweight Directory Access Protocol (LDAP)

Name	Address
Allan Scott	a_scott@lantern.com
Bill Jones	BOJ@pol.net
Mary Peterson	Brandywine@rocketmail.com
Crash Corrigan	Crash_98@mailexcite.com
Stephen Adams	s_adams_98@yahoo.com
Jenna MacCimbri	CGF@Zamora.net
Dave Martin	Martini@olive.com
Flash Flannagan	FF_100@access.com
Carolyn Jones	morticia@noctem.net
Carl Magnus	aachen@imperator.com
Darlena Morraine	DCM@blonde.net
Clutch Cargo	CC_cartoon@tvland.com
Patricia Douglas	p_ddouglas@yahoo.com

3 Before sending the request itself, the client sends an LDAP command, called the BindRequest, to the server. Basically, this is a command telling the server that the client wants to make a request.

BINDREQUEST

4 The server answers with a command called the BindResponse, telling the client that it can go ahead and make a request.

BINDRESPONSE

FIND "GRALLA"

5 The client then sends the request to the server, sending a person's name and asking for his or her email address.

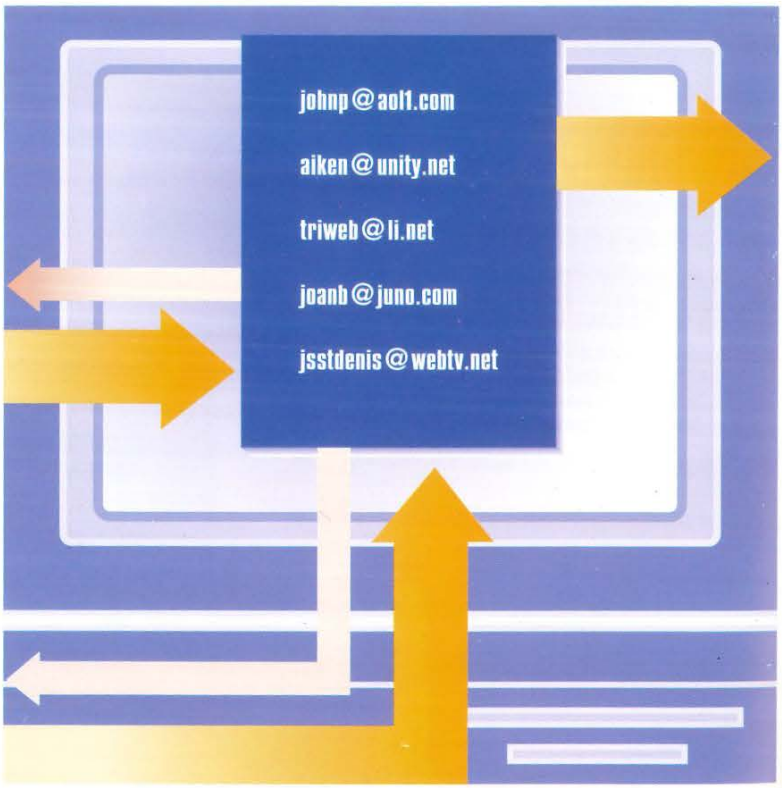
6 The LDAP directory performs a search and finds the email address or addresses in its database. It sends the results to the client.



8 Depending on the software you're using, you can now add the name and email address to your local address book, or you can immediately send an email to the person.

CHAPTER

17 Email Spam and Blocking



ONE of the most contentious issues to surface on the Internet in the last few years has to do with what Internet users call "Spam." Spam is unsolicited junk email that commercial companies send out, asking you to buy their goods and services. At times it may contain inducements to visit the seller's site. The email usually contains a phone number to call, an address to send money to, or a Web site to visit to buy the goods and services.

The term "Spam" comes from a Monty Python skit in which every item on a menu contained Spam luncheon meat. It was originally used to refer to unsolicited postings for commercial products or services on Usenet, especially when they were cross-posted to several newsgroups.

Spam may seem like a minor annoyance, but the truth is, it can cause major problems. Spam floods the Internet with unwanted mail, which can lead to delayed or lost mail. It clogs the Internet pipeline, making other information slower to send. It wastes time for those who have to go through their email boxes deleting unsolicited mail, especially when they pay for their email service by the hour. Additionally, it's fairly common for Spammers to hide their real email addresses by forging other people's names onto the From or Sender header of an email message. So, those people whose names were forged may be the target of angry mail. This makes it difficult for Webmasters and mail administrators to filter Spam messages by From address or domain name. Sometimes Spammers even use other people's servers to deliver their bulk email; in essence, forcing someone else to pay the costs of the Spammer's mail delivery.

In some ways, Spam is not very different from traditional junk mail. Spammers buy or compile massive lists of email addresses, in the same way that junk mailers buy or compile U.S. Postal addresses. The Spammer then uses special software to send a solicitation to every person on the list—not uncommonly, tens of thousands of pieces of email in a single Spam mailing. To hide their true identity, Spammers forge names onto the headers of email messages, and even "relay" their Spam to another mail server on the Internet, so that it is impossible to find out where the mail comes from. Often, a user will request to be taken off the list by replying to email addresses that the Spammers provide. However, this verifies the user's address and he or she will get even more Spam.

A variety of ways have been devised to block Spam, including having email filters on email software ignore any mail from known Spammers. This doesn't always work well, however, because Spammers often change or forge their email addresses. There are also calls for the courts or Congress to take action. Congress has considered several laws, including one that would ban Spam entirely, in the same way that junk faxes were banned several years ago. Until then, try doing an Internet search for Web sites that will help you download and install software to help filter your email and stop Spam.

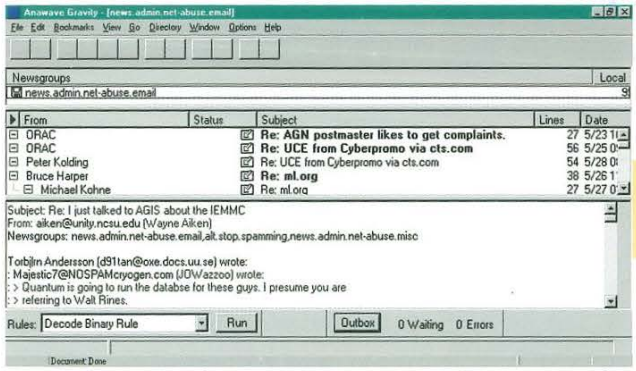
How Email Spam Is Sent

1 *Spam* is a term used to describe unsolicited email sent to you, often by commercial firms that attempt to sell you goods and services. Spam is sent out as bulk email, often to lists of 10,000 or more people at once. It's inexpensive to send, so its use has exploded on the Internet to the point where it is common for someone to get several dozen of these messages in a single day. Spam has become enough of an annoyance to warrant calls to ban it outright.



Bulk Email List

2 To send out unsolicited bulk email, a Spammer first needs to get a list of email addresses. Often, Spammers buy the lists from companies that compile them. These companies use automated software robots to get the email addresses. The robots get the lists from a number of sources. One way is to go into Usenet newsgroups and harvest email addresses by looking inside every message, which usually has in it the email address of the person who posted it.



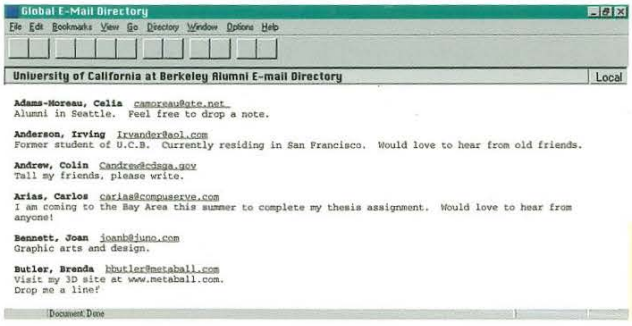
Usenet Newsgroups

Get Email Addresses

ADD: aiken@unity.net
triweb@li.net

johnp@aol1.com
aiken@unity.net
triweb@li.net
joanb@juno.com
jsstdenis@webtv.net

3 Email addresses can also be harvested from email directories on the Web sites that allow people to look up others' email addresses. Software robots can go into the directory and grab every address in the directory. Robots also can go into chat areas, such as those on America Online, and gather email addresses.



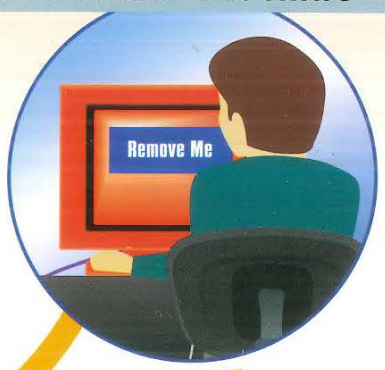
Web Email Directory

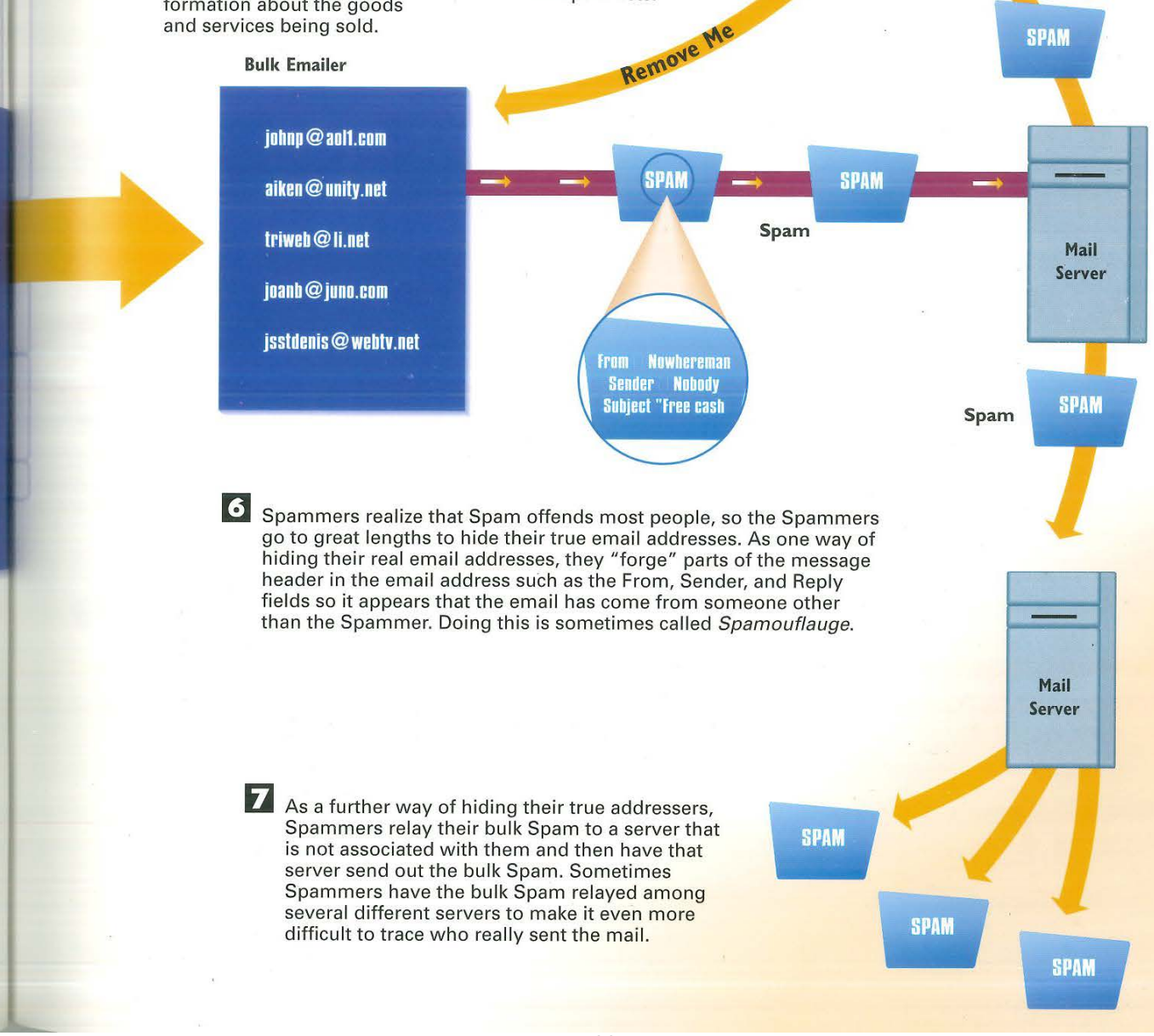
Get Email Addresses

ADD: aiken@unity.net
triweb@li.net

4 The Spammer either buys the resulting email list or compiles one of his or her own. The Spammer uses the list, along with bulk mailing software, and sends a Spam message to every person on the list. In the message may be a return address, Web site, or phone number where the reciever can get more information about the goods and services being sold.

5 Some Spammers will include in the email a return address where someone who no longer wants to receive Spam can send a message and be taken off the Spam list. When the remove message is received, a robot automatically takes the person off the list. However, Spammers rarely do this because most people would opt not to be on the Spam lists.



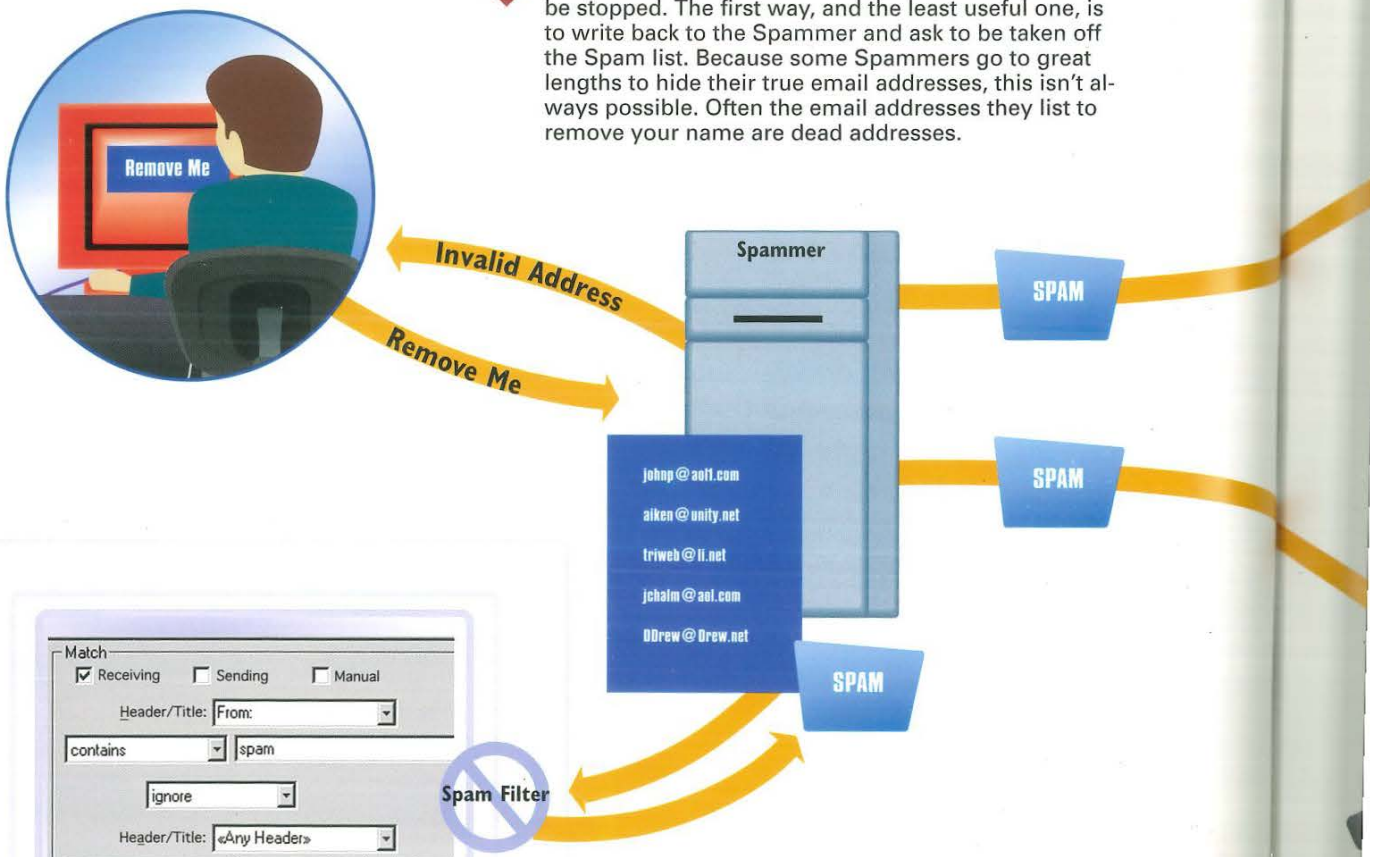


6 Spammers realize that Spam offends most people, so the Spammers go to great lengths to hide their true email addresses. As one way of hiding their real email addresses, they "forge" parts of the message header in the email address such as the From, Sender, and Reply fields so it appears that the email has come from someone other than the Spammer. Doing this is sometimes called *Spamouflauge*.

7 As a further way of hiding their true addressers, Spammers relay their bulk Spam to a server that is not associated with them and then have that server send out the bulk Spam. Sometimes Spammers have the bulk Spam relayed among several different servers to make it even more difficult to trace who really sent the mail.

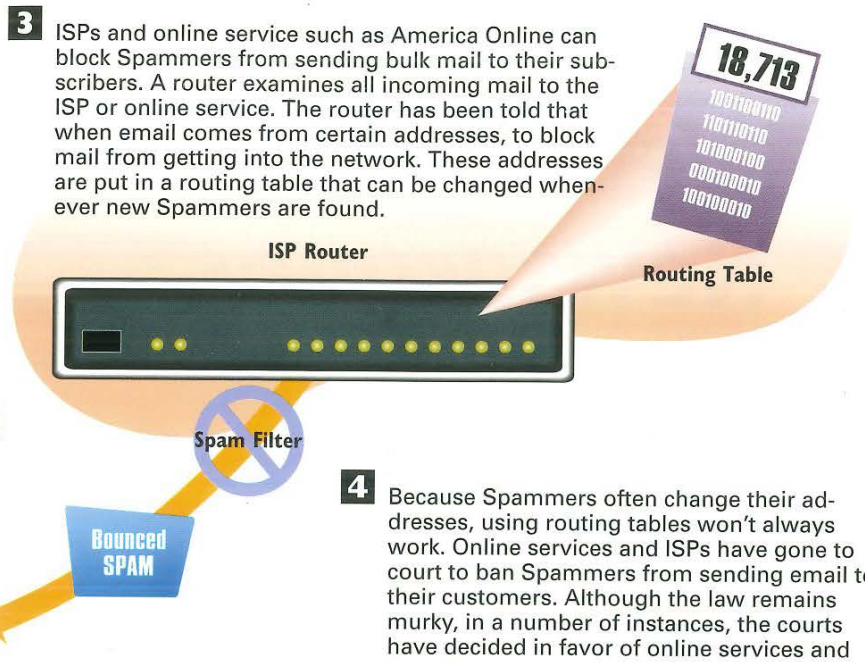
How Email Spam Is Blocked

1 There are a number of different ways that Spam can be stopped. The first way, and the least useful one, is to write back to the Spammer and ask to be taken off the Spam list. Because some Spammers go to great lengths to hide their true email addresses, this isn't always possible. Often the email addresses they list to remove your name are dead addresses.



2 Some email software allows you to filter out messages from certain addresses; when email comes from that address, your email software won't allow it through. This feature is referred to as a *Spam filter* or *Spam blocker*. Whenever you receive Spam, you can put that address into your Spam filter and you'll never get email from that address again. Spam filters don't always work, however, because Spammers often change their addresses and forge the From, Sender, and Reply fields in an email header. Online services such as America Online will let you block email from any address you specify—in essence, it's a Spam filter for those who use America Online. Additionally, some Internet service providers (ISPs) will block Spam.

3 ISPs and online service such as America Online can block Spammers from sending bulk mail to their subscribers. A router examines all incoming mail to the ISP or online service. The router has been told that when email comes from certain addresses, to block mail from getting into the network. These addresses are put in a routing table that can be changed whenever new Spammers are found.



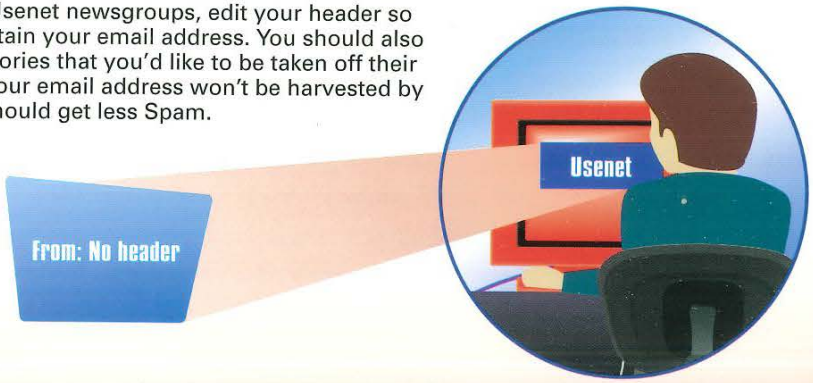
4 Because Spammers often change their addresses, using routing tables won't always work. Online services and ISPs have gone to court to ban Spammers from sending email to their customers. Although the law remains murky, in a number of instances, the courts have decided in favor of online services and the ISPs, and have banned Spammers from sending mail through them.



5 A number of laws and schemes have been proposed to regulate or outlaw Spam. In one plan, every piece of Spam would have to contain a specific piece of information in the message header, identifying it as unsolicited email. In this way, people could set their Spam filters to block every single piece of Spam, filtering out that piece of information. Some laws have been proposed that would outlaw Spam entirely, in the same way that junk faxes were banned.

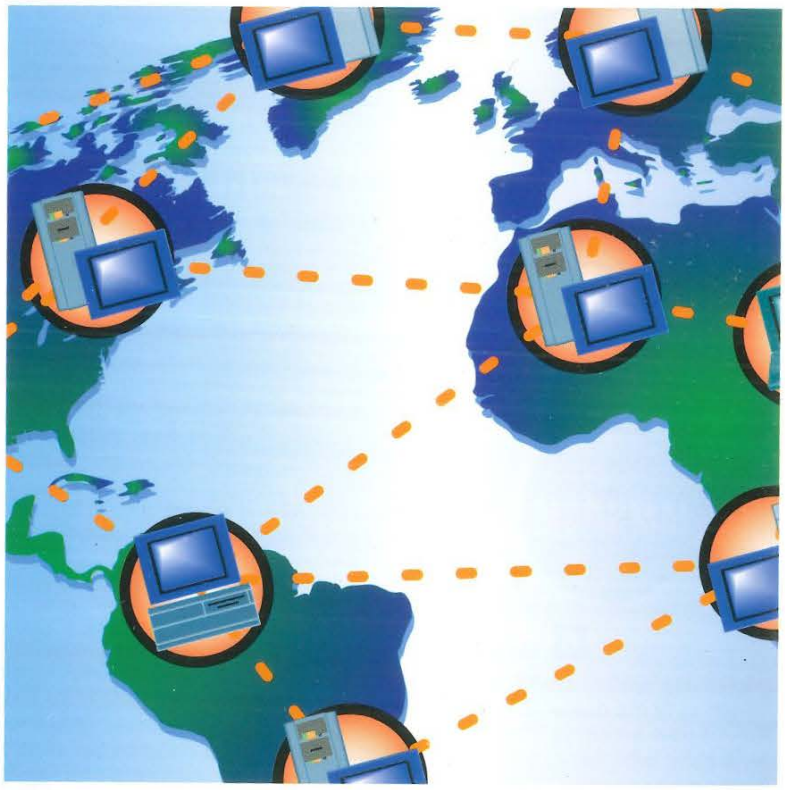


6 One way to prevent being Spammed is to make sure that your email address isn't added to Spam lists. To do this, when posting to Usenet newsgroups, edit your header so that it doesn't contain your email address. You should also notify email directories that you'd like to be taken off their lists. In this way your email address won't be harvested by robots, and you should get less Spam.



CHAPTER 18

How Usenet Newsgroups Work



USENET, the world's biggest electronic discussion forum, provides a way for messages to be sent among computers across the entire Internet. People from all over the world participate in discussions on thousands of topics in specific areas of interest called *newsgroups.*

There are at least 20 different major hierarchies of newsgroups, such as recreation (identified by the letters "rec") and computers (identified by the letters "comp"). Within these major hierarchies are subcategories (such as rec.arts) and further subcategories (such as rec.arts.books). Individual newsgroups can cover anything from movies to parenting, ecology, sports teams, clip art, and news about Usenet itself. Not all Internet sites carry all newsgroups. An administrator at each site decides which newsgroups to carry.

To participate in newsgroups, you'll need special software to read and respond to them. There are readers for PC, Macintosh, and UNIX computers. Online services such as CompuServe and America Online have their own proprietary software that lets you participate in newsgroups.

A good newsgroup reader will let you view the ongoing discussions as threads. *Threads* are ongoing conversations that are grouped by topic. So, for example, in the rec.arts.books newsgroup there may be many different threads going on at one time, each discussing a different book.

Many newsgroups have a list of Frequently Asked Questions, or FAQs (pronounced "facks"), associated with them. These FAQs answer common questions about the newsgroup. It's a good idea to read the FAQ before submitting questions to the newsgroup as a whole.

You participate in newsgroups by reading the messages and responding to them. There are moderated and unmoderated newsgroups. In a *moderated newsgroup*, each message goes to a human moderator. The moderator looks at the messages, making sure they're appropriate for the group. If they are appropriate, the messages are posted. All messages sent to an *unmoderated newsgroup* are automatically posted.

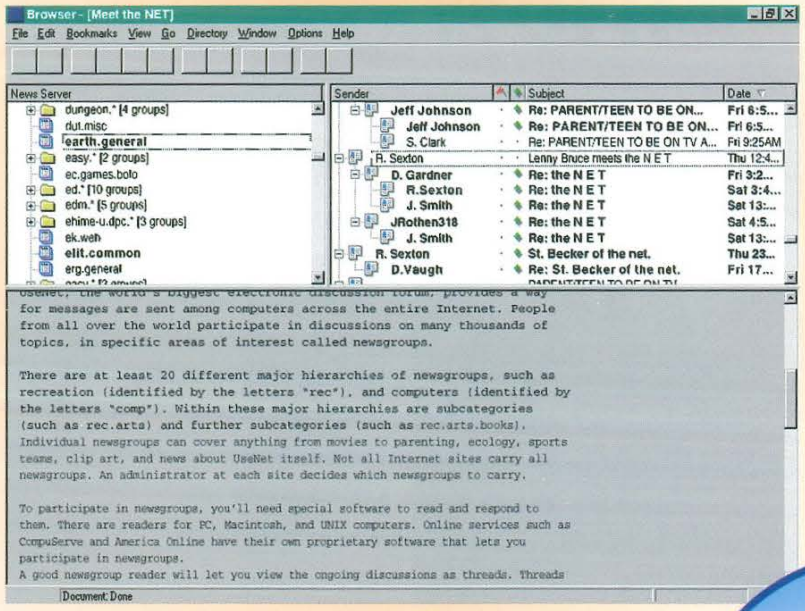
When messages are posted, Usenet servers distribute them to other sites that carry the newsgroup. A site usually carries only the most current messages; otherwise, they would soon run out of storage space. Some sites *archive*, or store, old discussions.

A convenient way to check newsgroups is to subscribe to those that interest you. That way, whenever you check the Usenet server, new messages in your subscribed newsgroup will be delivered to you. You can also cancel your subscription to a newsgroup if you are no longer interested in it. You can read newsgroups without subscribing to them; in that case, you'll have to manually ask to read specific newsgroups instead of having it done automatically for you.

Binary files such as pictures and multimedia can be posted in newsgroups. These files must be specially encoded in order for them to be posted. To view them, you'll have to transfer them to your computer and then unencode them with special software. A common encoding scheme used on newsgroups is called uuencode.

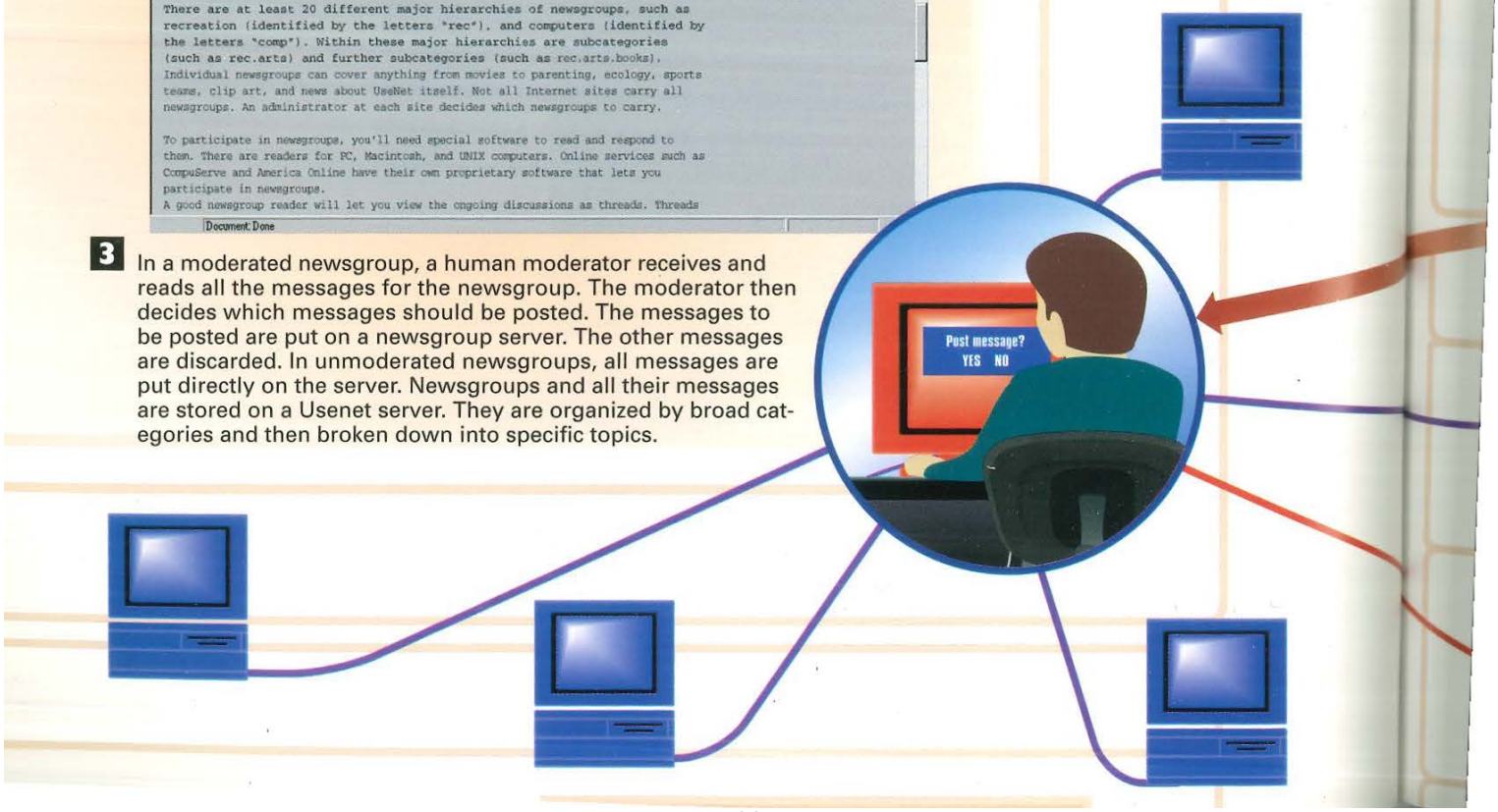
How Usenet Works

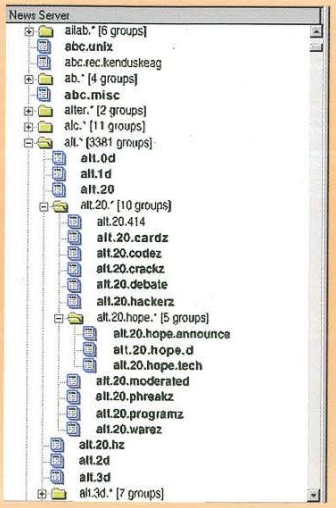
1 Usenet is a global bulletin board and discussion area. It collects messages about many thousands of different topics into newsgroups, which are freewheeling discussion areas in which anyone can participate. Newsgroups can be found on many host computers across the Internet. There are thousands of newsgroups that cover just about every topic you've ever imagined and many you probably haven't thought of.



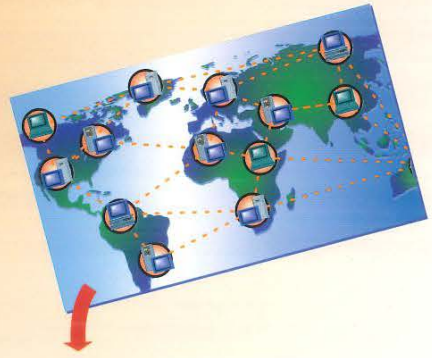
2 To participate, people send and read messages that are posted to the newsgroup. There are two kinds of newsgroups: moderated newsgroups and unmoderated newsgroups.

3 In a moderated newsgroup, a human moderator receives and reads all the messages for the newsgroup. The moderator then decides which messages should be posted. The messages to be posted are put on a newsgroup server. The other messages are discarded. In unmoderated newsgroups, all messages are put directly on the server. Newsgroups and all their messages are stored on a Usenet server. They are organized by broad categories and then broken down into specific topics.

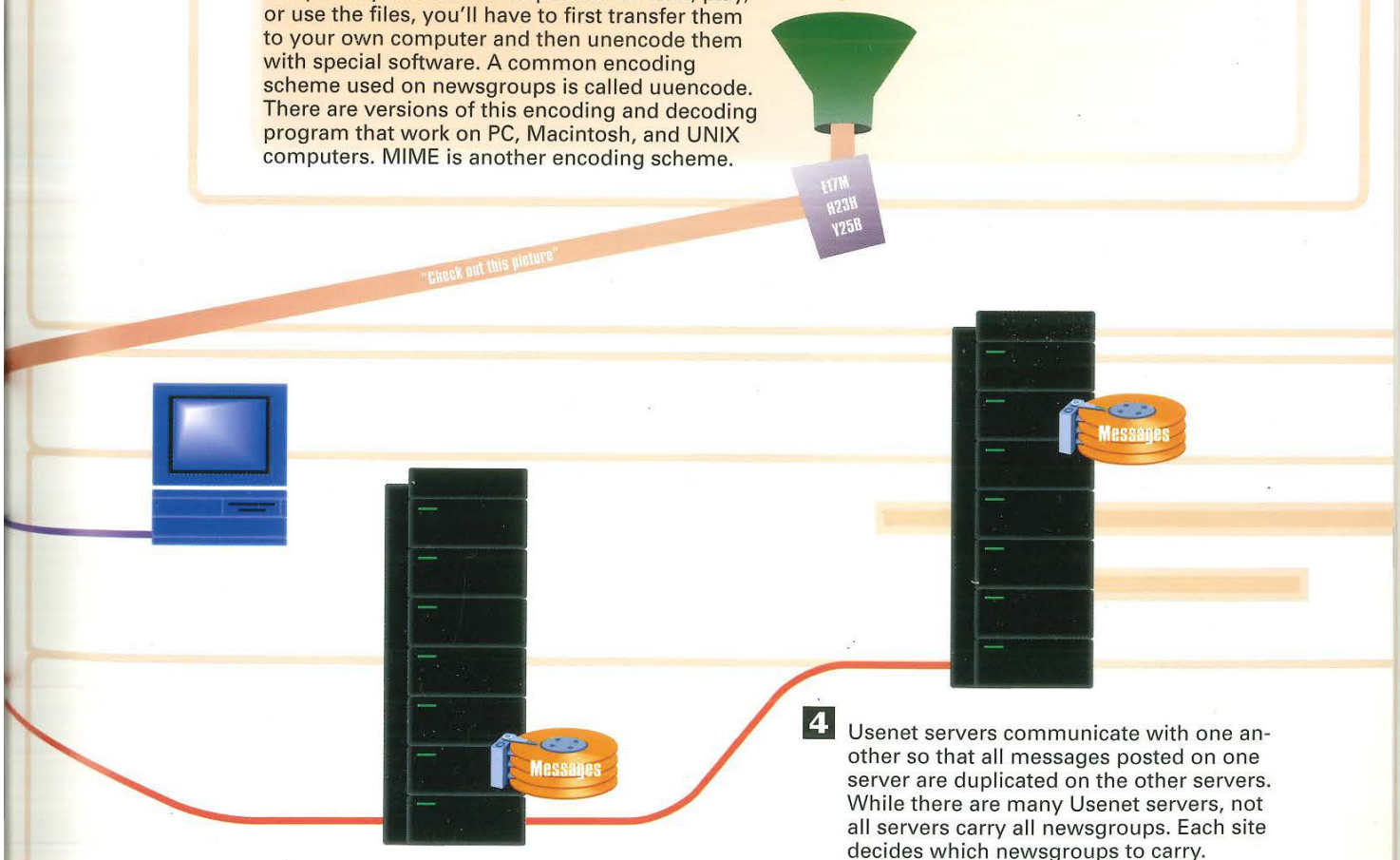




6 Newsgroup reader software lets you read messages and respond to newsgroups. The software gives you ways to manage your newsgroups by also allowing you to subscribe to newsgroups, which means that new messages will automatically be delivered to you when you check the server. You'll also be able to cancel your subscription.



5 Pictures, multimedia files, and even executable programs can be posted in newsgroups for other people to see and use. However, because of the technology used in newsgroups, those files must be specially encoded to be posted. To view, play, or use the files, you'll have to first transfer them to your own computer and then unencode them with special software. A common encoding scheme used on newsgroups is called uuencode. There are versions of this encoding and decoding program that work on PC, Macintosh, and UNIX computers. MIME is another encoding scheme.

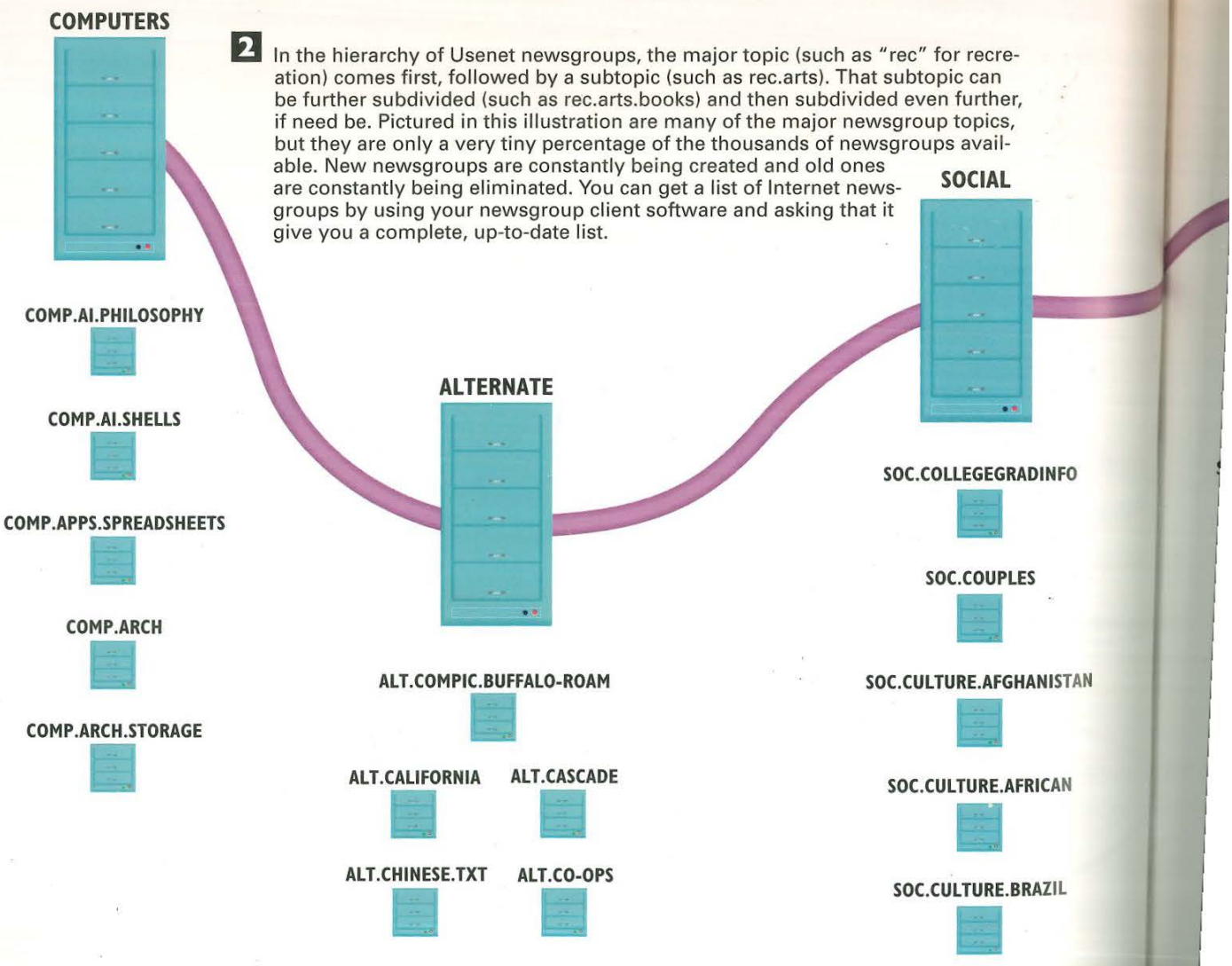


4 Usenet servers communicate with one another so that all messages posted on one server are duplicated on the other servers. While there are many Usenet servers, not all servers carry all newsgroups. Each site decides which newsgroups to carry.

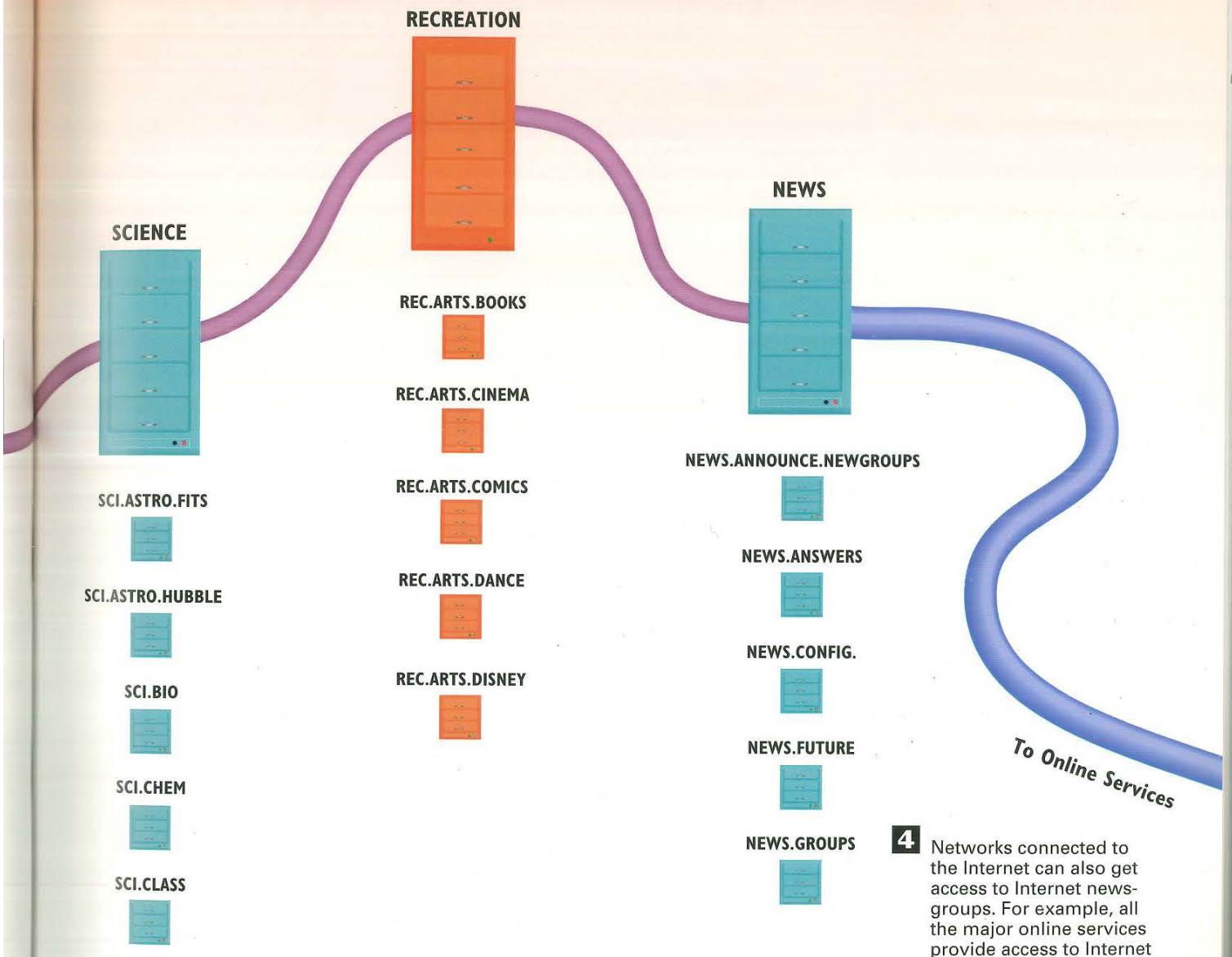
Understanding the Hierarchy of Usenet Newsgroups

1 There are thousands of Usenet newsgroups. They are divided into hierarchies of topics to make it easier for you to find the particular newsgroups you want to participate in. Not all sites carry all newsgroups. System administrators decide which newsgroups to carry at their site. Online services, as well as Internet providers that give direct connections to the Internet, allow access to Usenet newsgroups.

2 In the hierarchy of Usenet newsgroups, the major topic (such as "rec" for recreation) comes first, followed by a subtopic (such as rec.arts). That subtopic can be further subdivided (such as rec.arts.books) and then subdivided even further, if need be. Pictured in this illustration are many of the major newsgroup topics, but they are only a very tiny percentage of the thousands of newsgroups available. New newsgroups are constantly being created and old ones are constantly being eliminated. You can get a list of Internet newsgroups by using your newsgroup client software and asking that it give you a complete, up-to-date list.



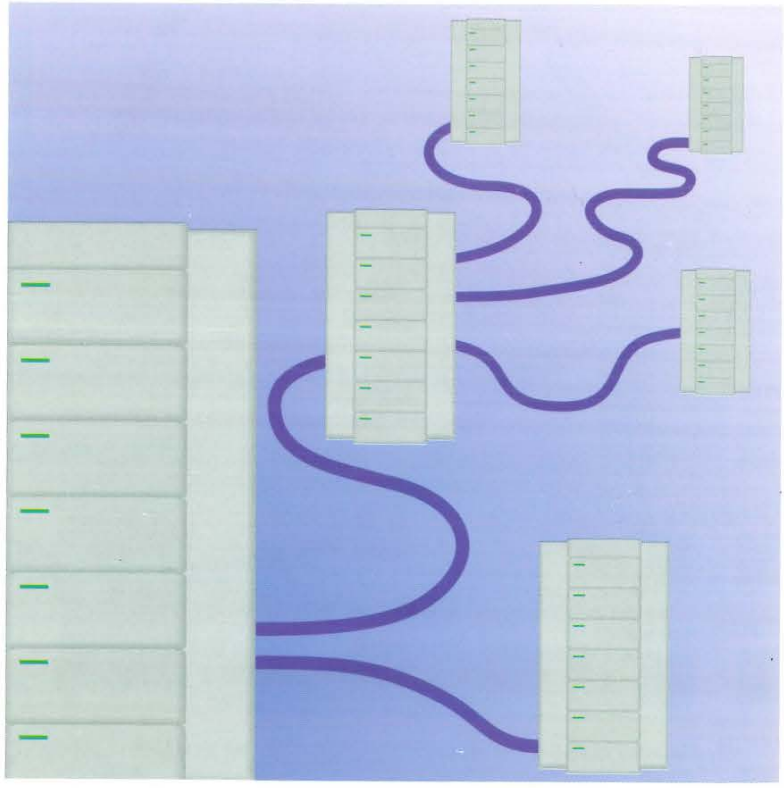
3 You can often identify the newsgroup by its name. Major Usenet topics include "comp" for discussions of computer-related topics; "soc" for discussions of societal topics; "sci" for discussions of scientific topics; "news" for discussions about newsgroups; "rec" for discussions of recreation topics; and "alt" for discussions of alternate topics, among others. So, for example, the rec.arts.books newsgroup carries discussions about books.



4 Networks connected to the Internet can also get access to Internet newsgroups. For example, all the major online services provide access to Internet newsgroups.

CHAPTER

19 How Internet Chat and Instant Messaging Work



ONE of the most immediate ways to communicate with others via the Internet is to participate in live "chat." Chat doesn't refer to people actually talking to each other and hearing each other's voices. Instead, it means that you hold live keyboard "conversations" with other people on the Internet—that is, you type words on your computer and other people on the Internet can see those words on their computers immediately, and vice versa. You can hold chats with many people simultaneously all over the world.

A number of ways are available to chat on the Internet, but one of the most popular ones is called *IRC*, or *Internet Relay Chat*. Every day, thousands of people all over the world hold conversations on many topics via IRC. Each different topic is called a "channel." When you join a channel, you can see what other people on the channel type on their keyboards. In turn, everyone in the channel can see whatever you type on your keyboard. You can also hold individual side conversations with someone. Channels live on different servers around the world. Some servers have only a few channels, and others have many of them.

IRC has facilitated communications during natural disasters, wars, and other crises. In 1993, for example, during the attempted Communist coup in Russia when Russian legislators barricaded themselves inside the Parliament building, an IRC "news channel" was set up for relaying real-time, first-person accounts of the events taking place.

IRC follows a client/server model, which means that both client and server software are required in order to use it. Many IRC clients are available for many kinds of computers, so whether you have a PC, a Macintosh, or a UNIX workstation, you'll be able to use IRC.

Your IRC client communicates with an IRC server on the Internet. You log on to a server using the client and pick a channel on which you want to chat. When you type words on your keyboard, they are sent to your server. Your server is part of a global IRC server network. Your server sends your message to other servers, which, in turn, send your messages to people who are part of your channel. They can then see and respond to your message.

You can chat on the Internet in other ways, as well. Many Web sites, for example, use proprietary chat software that doesn't use the IRC protocol, but enables you to chat when you're on the site.

Another kind of chat is called *instant messaging*. In instant messaging, you communicate privately, one-on-one, with another person. You can create special lists so that you're informed when your "buddies" come online ready to chat, and they're informed when you come online.

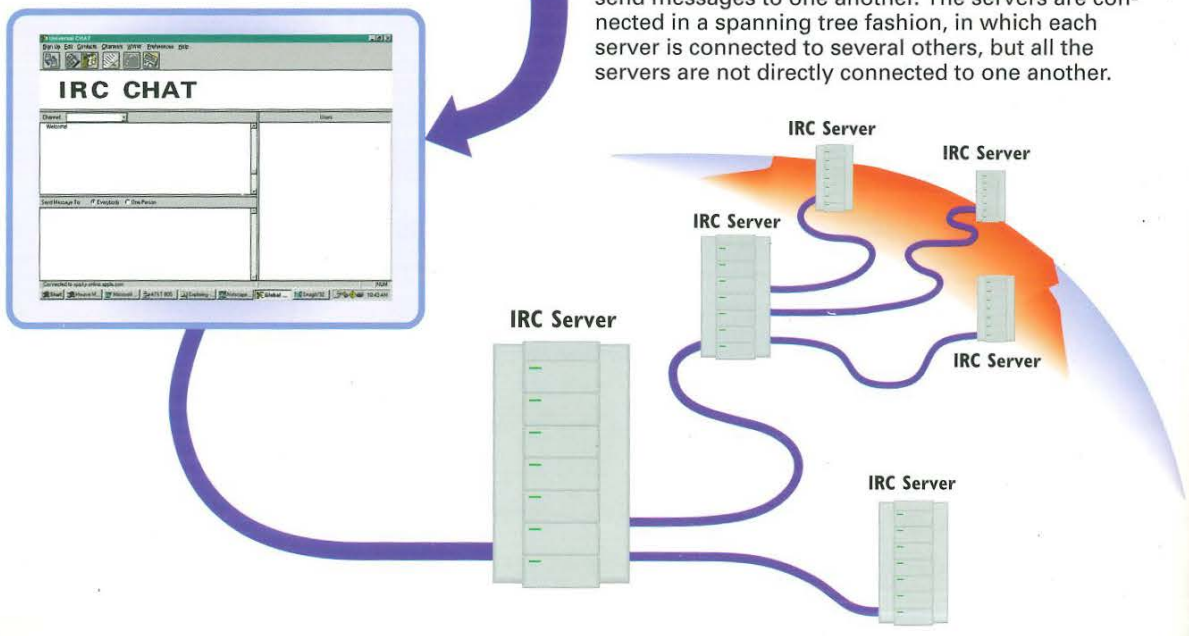
How IRC Works

1 IRC (Internet Relay Chat) is a way for people all over the world to "chat" with one another by using their keyboards. The typed words are instantly relayed to computers all over the world, where recipients can read them. This process happens in real-time, which means everyone sees the words as people type them.



2 IRC runs on a client/server model, which means that to use it, you need client software on your computer. Many IRC clients are available for PCs, Macintoshes, UNIX workstations, and other kinds of computers.

3 When you want to chat, you make a connection to the Internet and then start your client software. Next, you need to log on to an IRC server located on the Internet. Many IRC servers are located all over the world. They are connected together in a network so that they can send messages to one another. The servers are connected in a spanning tree fashion, in which each server is connected to several others, but all the servers are not directly connected to one another.



How Instant Messaging Works

People can send instant messages—one-to-one, private communications—to one another on the Internet in many ways. The various instant messaging systems don't work with one another. Pictured here is one of the most popular instant messaging systems—America Online's Instant Messenger on the Internet. Yahoo! Pager works similarly, and is popular as well.

1 The Internet version of America Online's Instant Messenger (AIM) runs as a piece of client software on your computer. To use it, you must be connected to the Internet. When you run the software, it opens up a TCP connection to an Instant Messenger login server. The software sends your screen name and password over the connection to log you into the server.

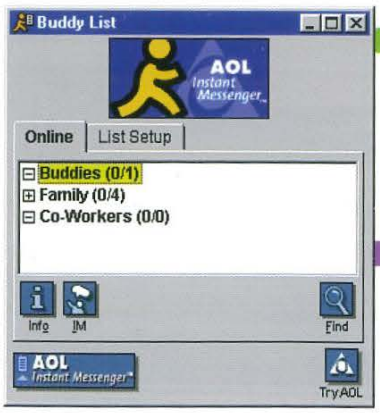


LOG IN

CLOSE CONNECTION

2 The server checks the screen name and password. If they're correct, the login server instructs the Instant Messenger software to close the connection to the login server and to open up a new connection to a different AIM server—the one that will handle your instant message session. This connection uses a special communications protocol that allows for AIM functionality.

3 Instant Message software includes "buddy list" capabilities. That means you keep a list of people you'd like to send instant messages to, and when they come online, you are notified so that you can send instant messages to and receive instant messages from them. You create a buddy list in your AIM software by adding your buddies' screen names to it.

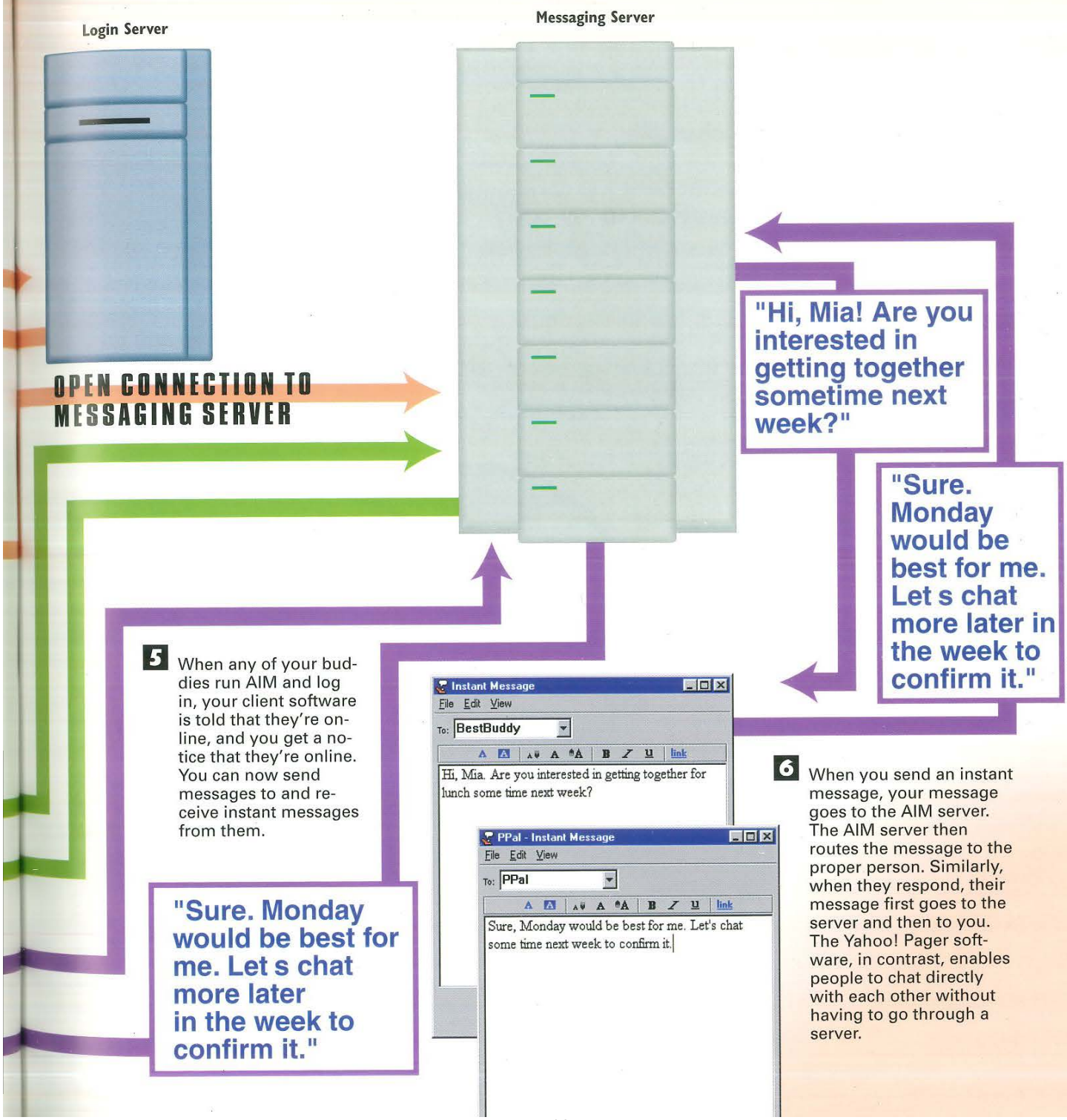


4 When you establish a connection with the AIM server, your client software sends a list of your buddies to the server. The server checks to see if any of the buddies are online—and it continues to do that for as long as you run the software on your computer. If you change the list of buddies during your session, that information is sent to the server as well so that it can keep track of new buddies or ignore buddies that you've deleted from your list.

"Are my buddies here?"

COOL LIZARD IS ONLINE

"Hi, Mia! Are you interested in getting together sometime next week?"

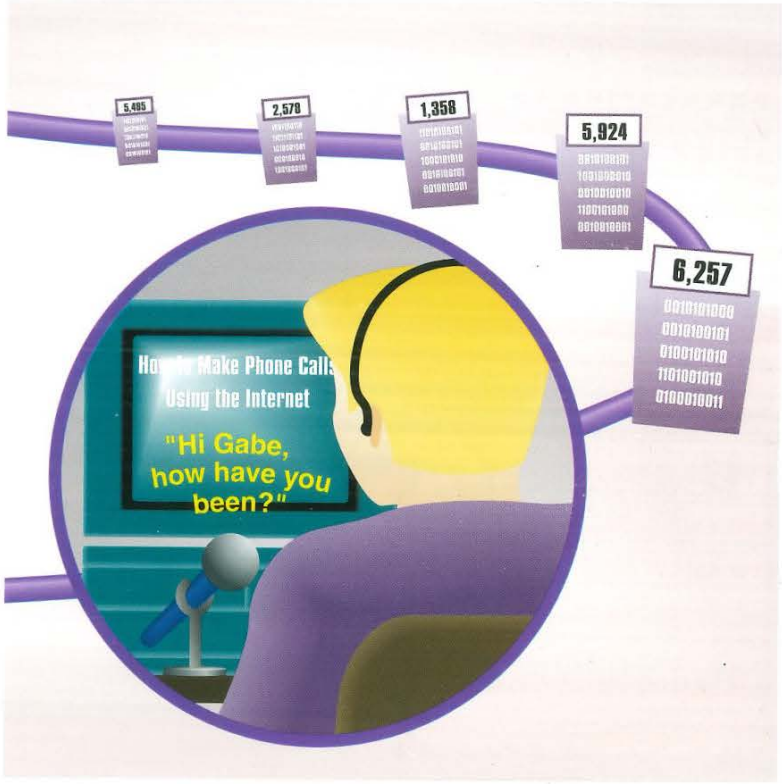


5 When any of your buddies run AIM and log in, your client software is told that they're online, and you get a notice that they're online. You can now send messages to and receive instant messages from them.

6 When you send an instant message, your message goes to the AIM server. The AIM server then routes the message to the proper person. Similarly, when they respond, their message first goes to the server and then to you. The Yahoo! Pager software, in contrast, enables people to chat directly with each other without having to go through a server.

CHAPTER

20 Making Phone Calls on the Internet



THE Internet has pioneered many new ways to communicate, such as email, live chat, and newsgroups. But it can be used to enable some old-fashioned communications as well: You can make telephone calls using the Internet. When you do so, the sound of your voice and that of the person you're talking to are broken down into packets. Those packets are delivered using the Internet's TCP/IP protocols.

You can make telephone calls over the Internet in two ways. In the first way, you use your computer and special hardware and software to make calls, so that you communicate through your PC. In the second way, often referred to as Internet telephony, you make a phone call as you normally do on your telephone, except your call is routed over the Internet rather than through the normal phone service. This usually cuts the cost to you of making long-distance phone calls.

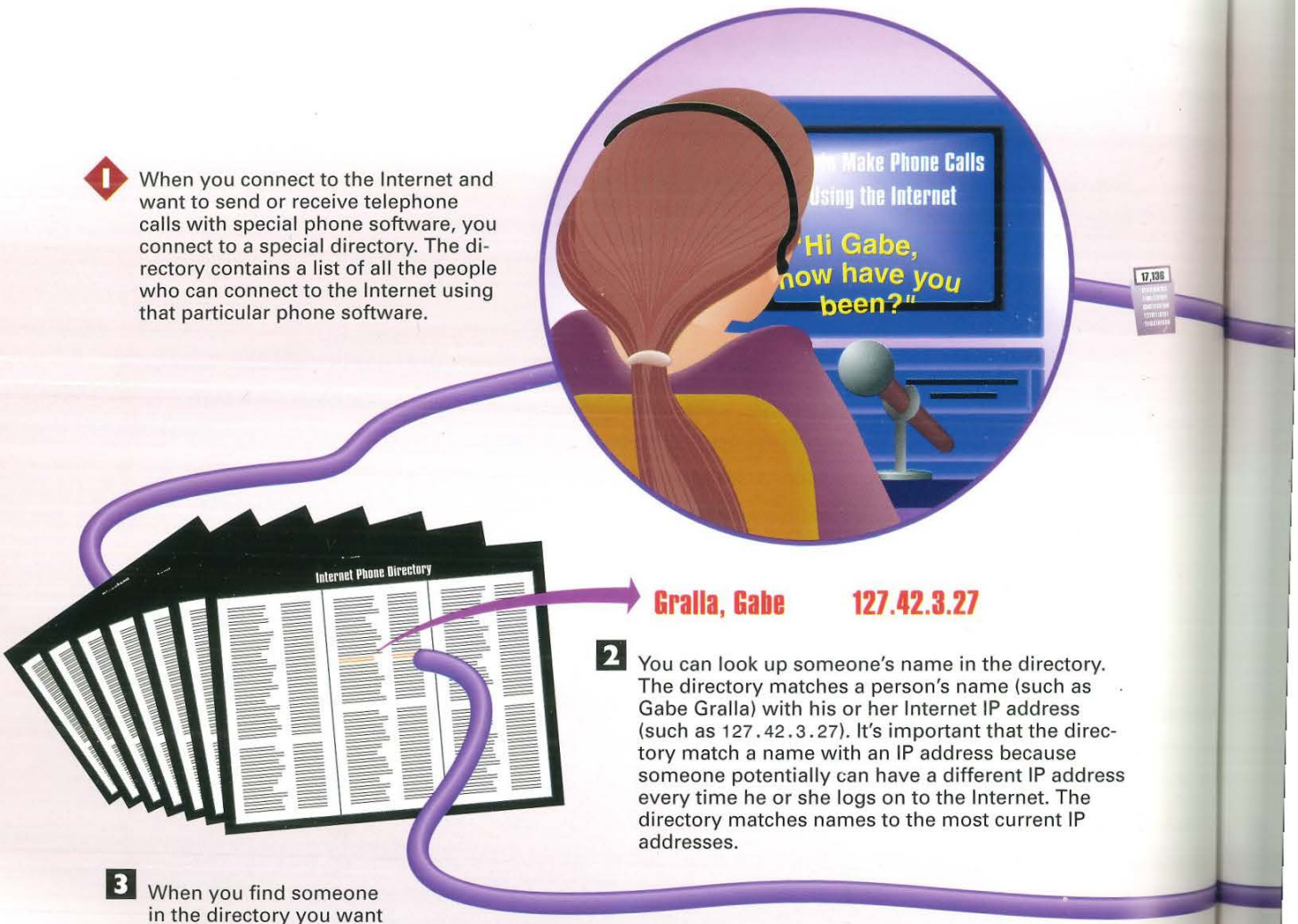
The revolutionary part of using the Internet to make phone calls from your PC is the price: it's free. You pay only for your Internet connection, as if you were browsing the Web or sending email. You don't actually have to pay for the phone call itself. In addition, you can make calls anywhere in the world. However, you will only be able to make telephone calls to and receive calls from someone who has an Internet address, so you won't be able to replace your telephone by using the Internet. A number of different competing products enable you to talk over the Internet, but as yet, they don't communicate with one another. No standard way of making Internet telephone calls exists, so you'll be able to talk only to people who use the exact same software you do for making and receiving phone calls.

Many different schemes and software packages will enable you to make phone calls from your PC over the Internet. You don't actually use your telephone with any of them. Instead, you speak into a microphone attached to your computer and listen through speakers and a sound card.

Although each company's software works somewhat differently, in general, they use similar ways of enabling people to make phone calls over the Internet. To make a phone call, first you'll have to know someone's Internet address. (You can consult directories that function like telephone books to get this information.) After you find the person's Internet address, you can connect to them by running special software and then by double-clicking their name. (The person on the other end must be running the same software you are using.) A message goes out to them over the Internet to see if they're available to talk.

When you make phone calls over the Internet using your normal phone, things work differently. You make the call as you would any other phone call. After you make the call, however, your voice will be digitized and converted into IP packets, and then delivered using the same TCP/IP technology as the rest of the Internet uses. On the other end of the call, the digitized voice data is converted and can be listened to over the phone.

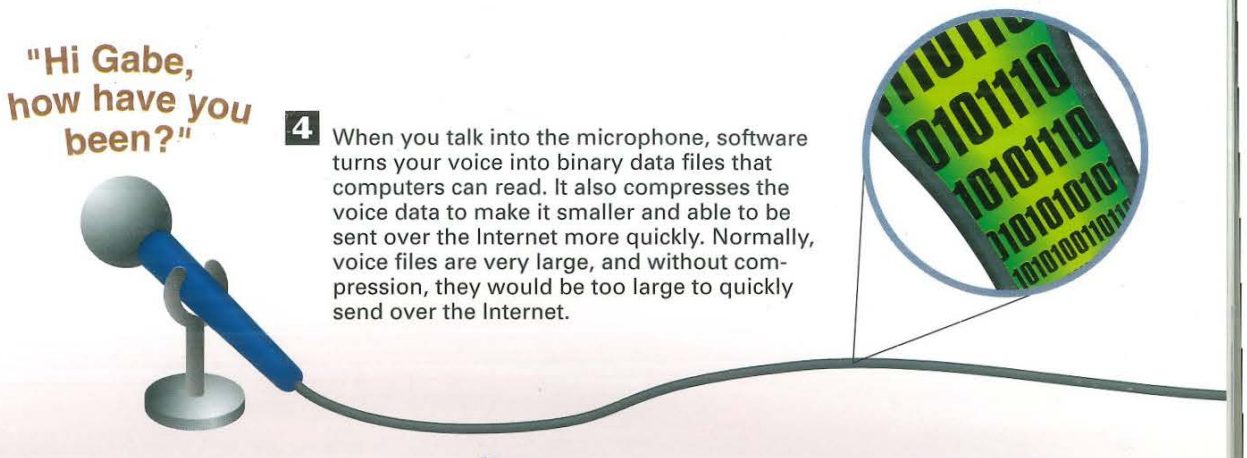
How to Make Phone Calls Using Your PC and the Internet



1 When you connect to the Internet and want to send or receive telephone calls with special phone software, you connect to a special directory. The directory contains a list of all the people who can connect to the Internet using that particular phone software.

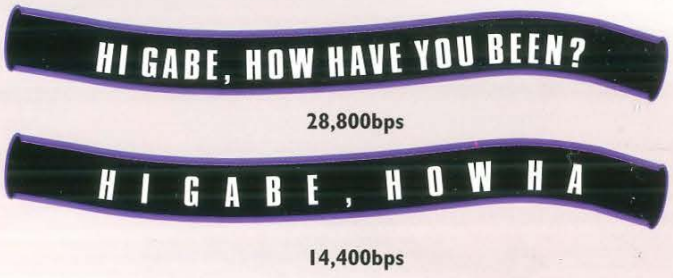
2 You can look up someone's name in the directory. The directory matches a person's name (such as Gabe Gralla) with his or her Internet IP address (such as 127.42.3.27). It's important that the directory match a name with an IP address because someone potentially can have a different IP address every time he or she logs on to the Internet. The directory matches names to the most current IP addresses.

3 When you find someone in the directory you want to call, the directory looks up the IP address and then routes your call directly to that computer. The phone "rings" at the person's computer. He or she picks up the receiver. The two of you can now talk, using the sound cards and microphones attached to your computers.



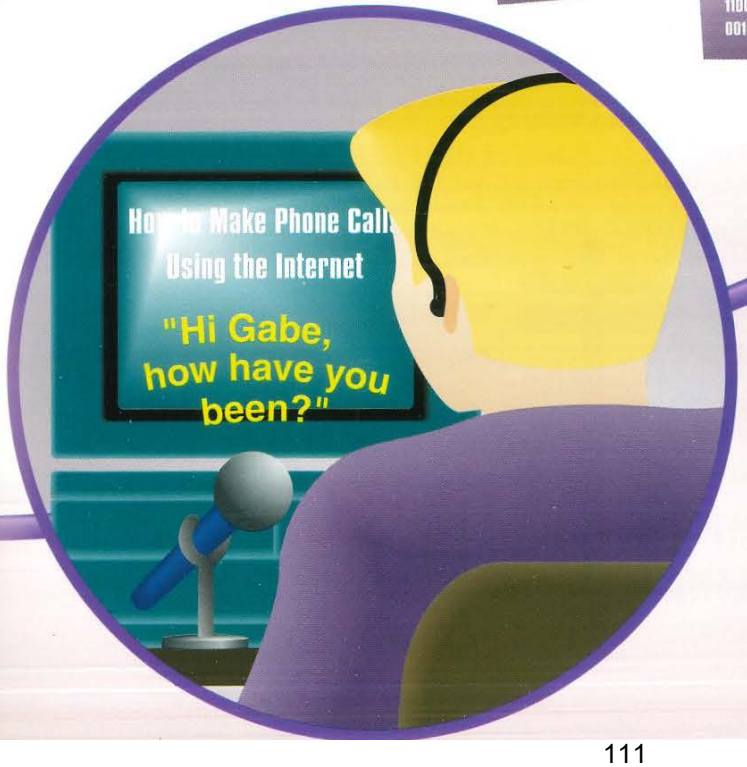
4 When you talk into the microphone, software turns your voice into binary data files that computers can read. It also compresses the voice data to make it smaller and able to be sent over the Internet more quickly. Normally, voice files are very large, and without compression, they would be too large to quickly send over the Internet.

5 The software also determines the speed of the Internet connection. If the connection is a high-speed connection, it creates voice files of high quality. If it's a lower-speed connection, it creates files of a lesser sound quality because at lower speeds it would take too long to send voice files, even after they had been compressed.





6 The software breaks up the voice files into a series of packets to be sent over the Internet using the Internet's TCP/IP protocols.



7 The packets are sent to the receiver. On the receiving computer, software decompresses the packets so they can be played and heard. Sometimes packets will arrive out of order on the receiving end, so some of them will be missing. When that happens, the software, based on the received packets, is able to estimate what the sounds are in the missing packet that has yet to be received. When the missing packet arrives, the software knows that it has already recreated what that packet would have sounded like. The software will then discard the packet.

How Internet Telephony Works



Hi, Mia. How about lunch next Tuesday?

2 The first part of the call goes over the normal Public Switched Telephone Network (PSTN), as does any other call; at this point, it is no different in any way.

1 When you make a phone call using IP telephony, you use a normal telephone, not a computer. Depending on the IP telephony service you use, you may have to dial in a special code first, as a way of routing the call to an IP telephony service.

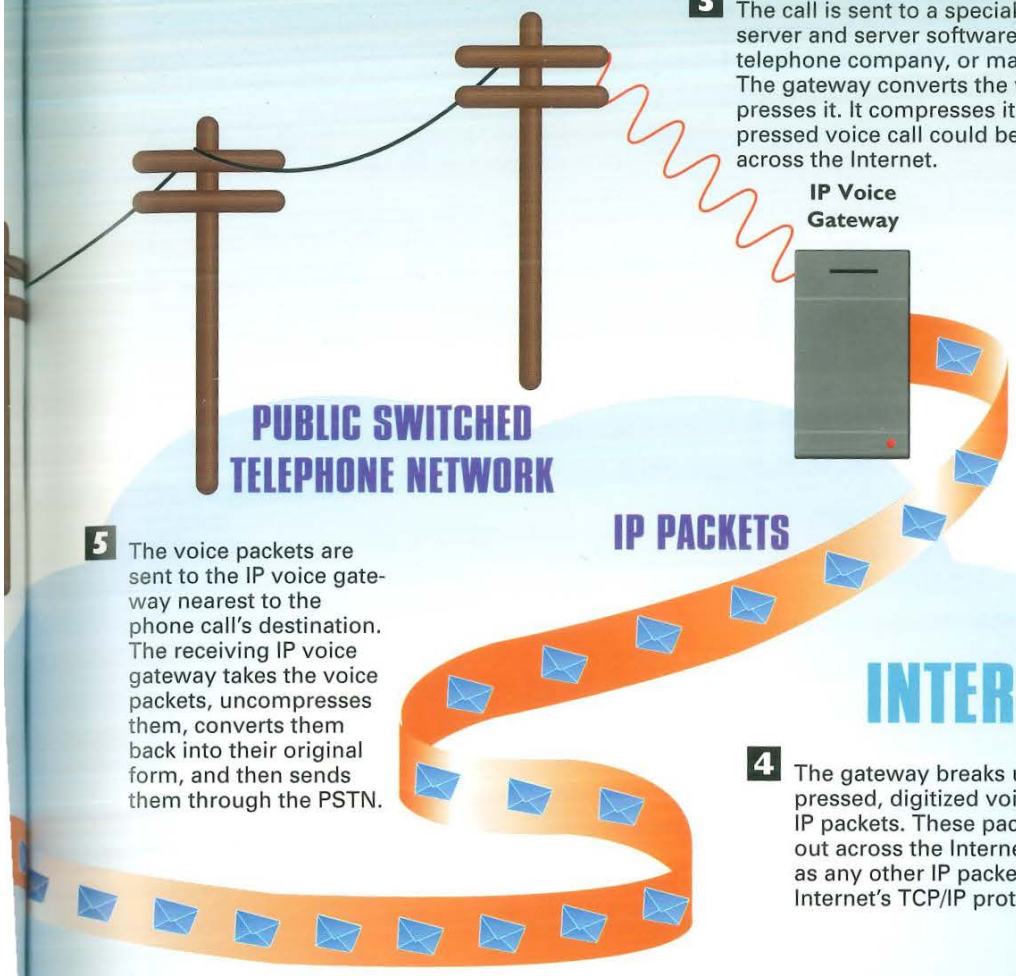
IP Voice Gateway

Hi, Mia. How about lunch next Tuesday?

IP (Internet Protocol) Telephony refers to telephone calls that are made over normal telephones but are delivered, in part, through the Internet instead of solely through a voice network, commonly called the Public Switched Telephone Network, or PSTN. Because the calls are routed through the Internet, they can be delivered at less cost than normal phone calls. But the voice quality may not be as good as calls delivered over the normal phone network.

PUBLIC SWITCHED TELEPHONE NETWORK

3 The call is sent to a special IP voice gateway. This gateway—a server and server software—may be located at the branch of the telephone company, or may instead be in a different location. The gateway converts the voice signal to digital data and compresses it. It compresses it because the data files of an uncompressed voice call could be too large to deliver in a timely fashion across the Internet.

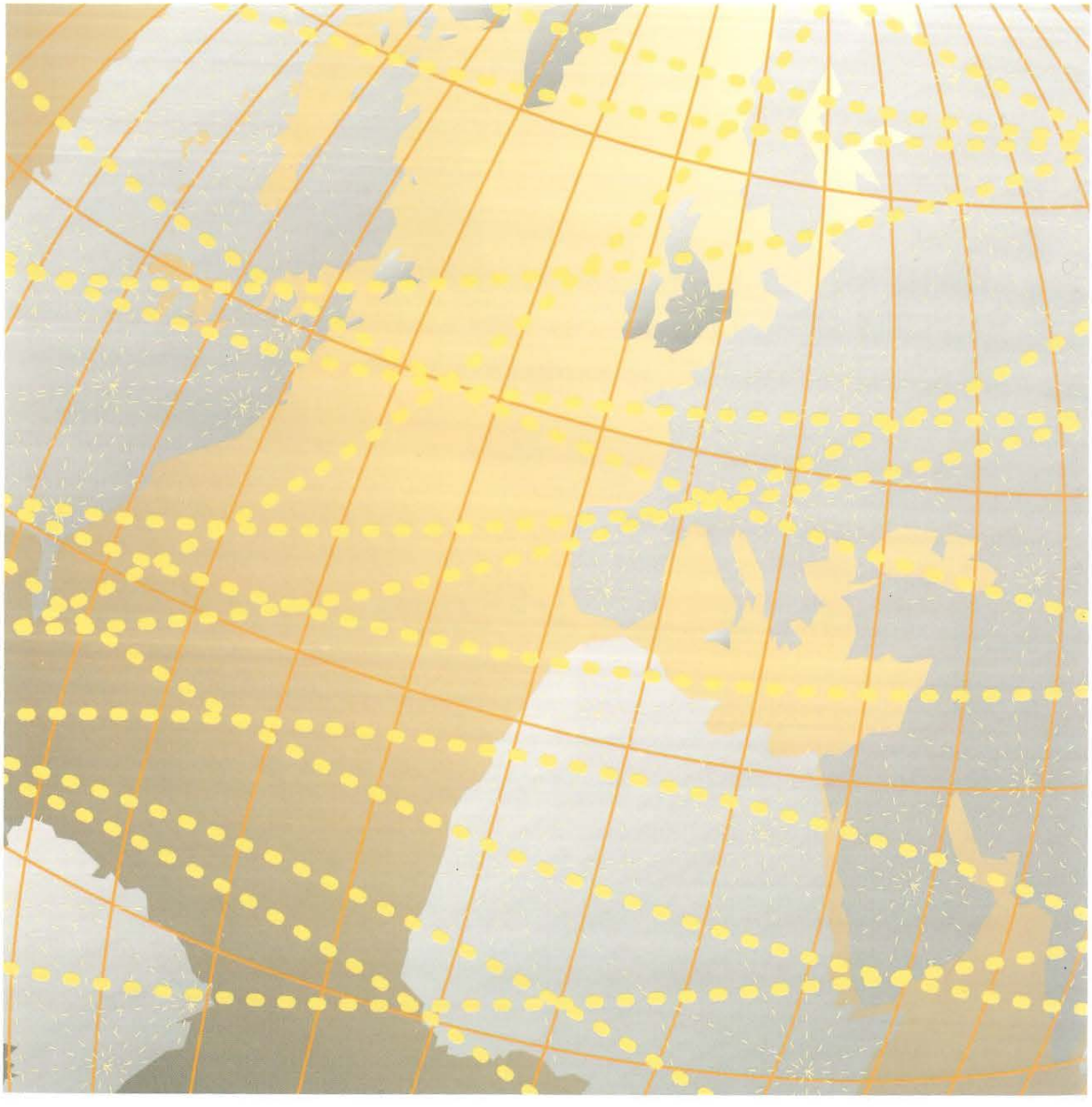


5 The voice packets are sent to the IP voice gateway nearest to the phone call's destination. The receiving IP voice gateway takes the voice packets, uncompresses them, converts them back into their original form, and then sends them through the PSTN.

4 The gateway breaks up the compressed, digitized voice signal into IP packets. These packets are sent out across the Internet the same as any other IP packets, using the Internet's TCP/IP protocols.

6 The call travels on the PSTN the same as any other telephone call. The person receiving the call picks up the phone and talks, the same as with any other telephone conversation. When the receiving person speaks into the phone, the voice is sent back through the PSTN to the IP voice gateway and across the Internet in the same way that the call was initially made.





PART 5

COMMON INTERNET TOOLS

Chapter 21: How Gophers Work
118

Chapter 22: How Telnet Works
122

Chapter 23: FTP and Downloading Files
126

Chapter 24: Searching the Internet
132

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 will show, a world exists well beyond the Web. Turn to Part 6, "How the World Wide Web Works," for information about the Web—the fastest growing and most visible part of the Internet.



The Web, of course, is only one part of the Internet. A lot more information is available on the Internet than that found via the Web. However, just as you need a browser to use the Web, you'll need special client software—what we can generically call Internet tools—to get at these other parts of the Internet. In this section, we'll look at how the most common and useful Internet tools work. And we'll also look at how Internet search engines work. Chapter 21, "How Gophers Work," describes how to use gophers. *Gophers* were one of the first attempts to corral the enormous amount of information available on the Internet. They help solve a major problem with the Internet—so much information is available that it can be difficult to find what you want. Gophers work on a simple menuing system and enable you to access information no matter where it's found, without having to remember to launch other Internet resources such as FTP or Telnet. Some of the ideas first introduced in gophers were later taken up by the World Wide Web.

Chapter 22, "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 into a distant computer using Telnet, you often use a menuing system.

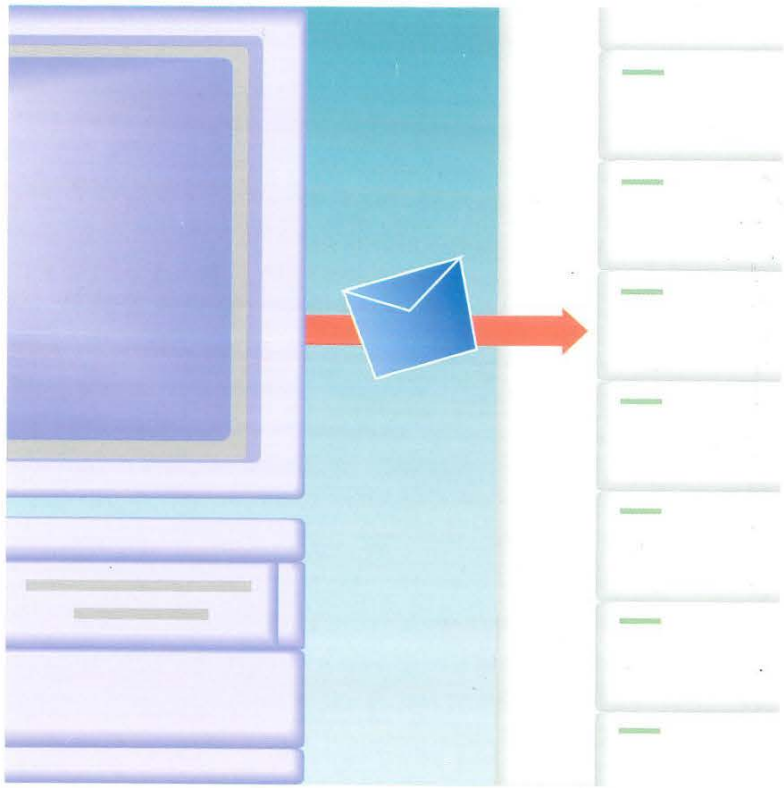
In Chapter 23, "FTP and Downloading Files," we'll cover one of the most popular uses of the Internet—downloading files. Generally, files are downloaded from the Internet using FTP, the Internet protocol. Not only will we look at how FTP works, but also at how files are compressed and decompressed on the Internet. A compressed file will take less time to be sent over the Internet to your computer. You may not know it, but many times when you're on a Web site and download a file, you're actually using the FTP protocol.

Finally, in Chapter 24, "Searching the Internet," we'll look at 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 that you type.



CHAPTER 21

How Gophers Work



TO a certain extent, the Internet has been a victim of its success. There is such an enormous amount of information on it that it's often difficult to find what you want.

This problem led to the development of a software program called an Internet Gopher. A Gopher organizes information in a logical fashion to lead you to files, Internet resources, data, and anything else you might search for on the Internet. Gophers were the first client software programs to allow access to many types of protocols and servers from within one client.

Because Gophers make it easy to access many types of information, they helped fuel the Internet's growth. Because they are easy to build, they set in motion a wave of publishing on the Internet. Gophers are not used as frequently now as they were in the early days, however, largely due to the explosion of the World Wide Web.

There are many Gophers located on the Internet. Organizations and universities generally create Gophers, often for internal use by employees or students. Most of the Gophers you'll come across are created and maintained by universities, and have been set up as a way to make it easier for students to tap into the resources of the university and the Internet. Although the Gophers were created by the university or an organization, they are often made available to anyone on the Internet.

Gophers work in a client/server model. You run a Gopher client on your own computer. The client sends requests to a Gopher server located on the Internet. The server sends back the information to the client, which displays the information on your computer screen. There are Gopher clients available for PCs, Macintoshes, and UNIX computers. Online services also have Gopher client software built into them.

Gophers are organized in a tree-like menu fashion. When you first engage a Gopher, you'll be greeted with the Gopher's main menu, called the *root gopher*. From this menu you choose items that interest you. Frequently, those items will be other menus, called submenus. Keep burrowing down through menus until you reach the information you want.

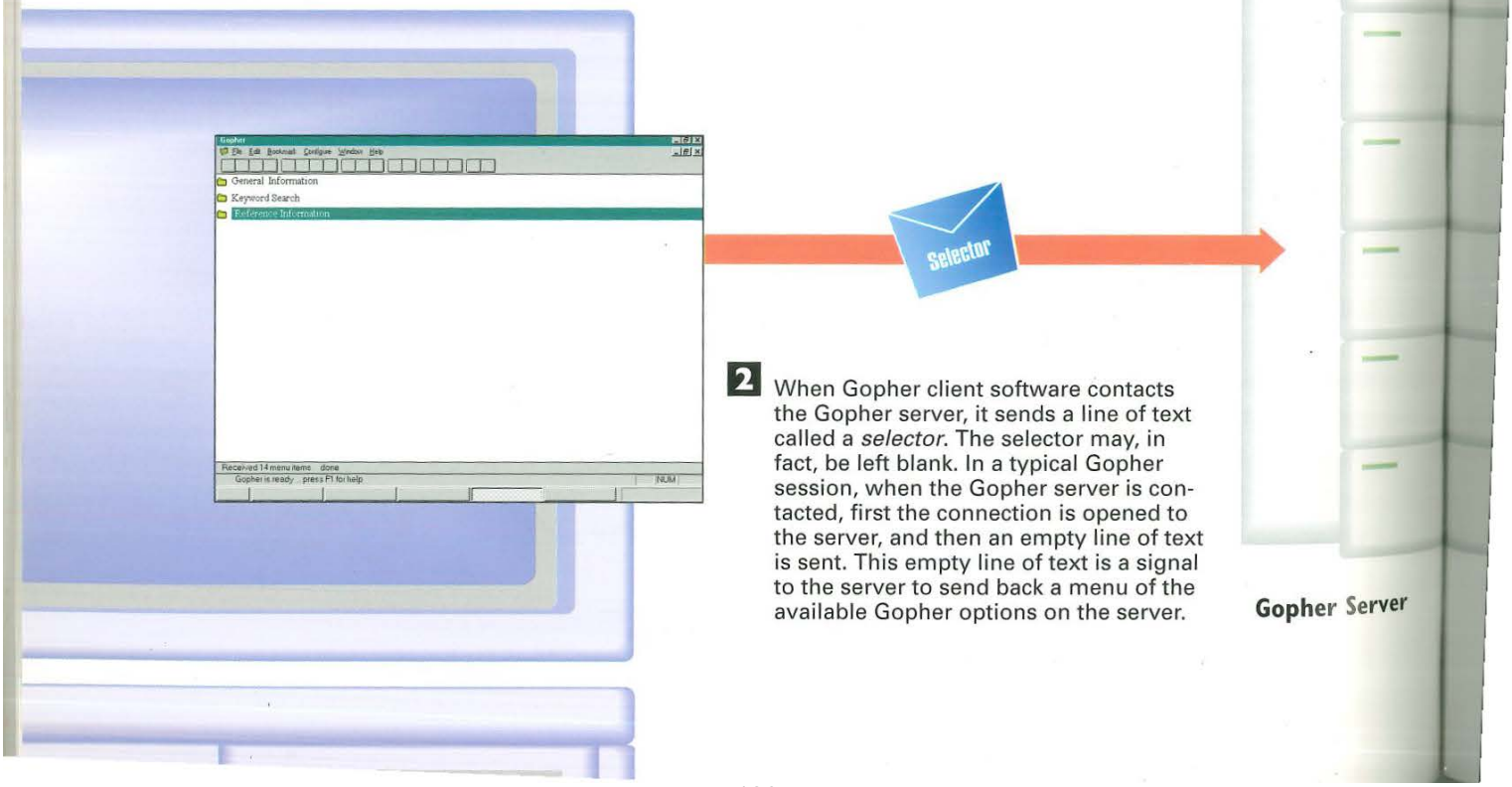
Gophers can link you to many types of Internet resources. The information you request often doesn't reside on the Gopher. Instead, the Gopher points you to other locations on the Internet. Gophers can lead you to text files, binary files, logins to remote computers, graphics files, multimedia files, WAIS database searches, other Gophers, and more.

What's helpful about Gophers is that you won't need to know which kind of Internet resource is required to get the information—a Gopher takes care of it for you. For example, if you request a file through a Gopher, the Gopher will automatically send it through FTP to your computer; you don't need to launch FTP software. (See Chapter 23, "FTP and Downloading Files," for details on FTP.)

Lastly, if the Gopher leads you to log on to a remote computer, the Gopher will do it for you; you won't need to launch a separate Telnet session. (Telnet is the topic of the next chapter.)

How Gophers Work

1 A Gopher follows the client/server model. The Gopher client sends requests to the Gopher server, which then sends back the requested information to the client. Next, the client displays the information on your computer. You can use special Gopher client software or a Web browser, which has the capability to function as a Gopher client. Online services have their own built-in Gopher clients. If you don't have a Web browser or Gopher client software, you can use Telnet to go to a public Gopher on the Internet. No matter how you access Gophers, however, they will generally work the same.



2 When Gopher client software contacts the Gopher server, it sends a line of text called a *selector*. The selector may, in fact, be left blank. In a typical Gopher session, when the Gopher server is contacted, first the connection is opened to the server, and then an empty line of text is sent. This empty line of text is a signal to the server to send back a menu of the available Gopher options on the server.

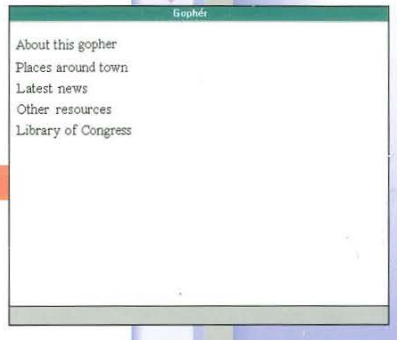
3 The server sends a series of lines back to the client, each of which ends with a carriage return and line feed. Each line is made up of a series of components separated by tab characters—a number; text called a *user display string* that the Gopher user will see in a menu; a selector string that will be sent to the Gopher server to retrieve a document or a directory; the domain name of the host that has the document or directory; and a port number, which tells the Gopher client and server how to connect over the TCP/IP connection.

User Display String

Selector String

```
0└About this gopher└dog.mbu.edu└70
1└Places around town└town. mbu.edu└70
1└Latest news└ news.mbu.edu└70
1└Other resources└ other.mbu.edu└70
8└ Library of Congress └ locis.loc.gov└70
```

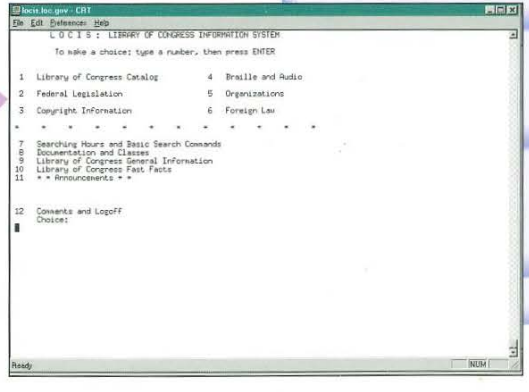
4 The Gopher client displays only the user display string of each line. All the user display lines taken together make up the Gopher menu that you'll see.



8└Library of Congress└locis.loc.gov└70

5 When you choose an item from the Gopher menu, the entire string—not just the string you see—is sent from the client back to the Gopher server. The number at the beginning of the string tells the server what kind of resource is being requested. A 0 means the item is a file; a 1 means the item is a directory; a 5 or a 9 means the item is a binary file; a 7 means a search will be launched; and an 8 points to a Telnet session.

Use Telnet on Library of Congress Info System



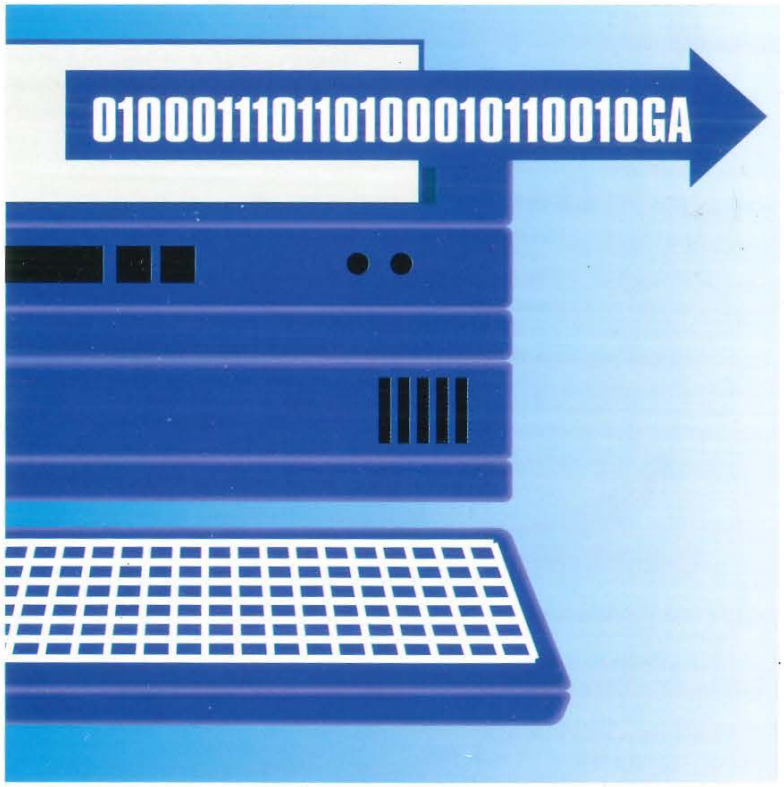
6 Using the selector string, the domain name, and the port number, in addition to the number at the beginning of the string, the server retrieves the information requested by the client. If the information requested was a directory, the Gopher server will send a series of lines, each of which displays a menu option, just as was done when the server was first requested.

7 If the information requested was a search, the client software will be sent data allowing for a search.

8 If the information requested is a file or binary document, that file or document will be sent. Often, the information requested will be on a different server than the Gopher server. In that event, the Gopher server will contact the other server and have the requested material sent to the Gopher client. The Gopher client can now read or otherwise use the material.

CHAPTER 22

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 different 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'll need to know the address of the Internet host whose resources you want to access.

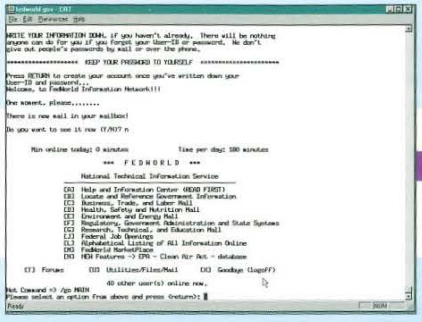
When you use Telnet, before you can take over the resources of a host computer you'll typically have to log onto the host. Often, you can use the name "guest" to log on. Some systems require that you also give information about yourself, such as your name and address. And some may require that you choose 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, 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. Also, when you connect using Telnet, you'll have to use terminal emulation; in essence you make sure your keyboard and monitor function as the host expects. The most common terminal emulation is the VT-100 emulation, so if you use Telnet software, that's a safe emulation to use.

Telnet clients are available for all the major operating systems, including 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 will remember hostnames for you. With clients, you can often keep an address book of hostnames so you can easily revisit them.

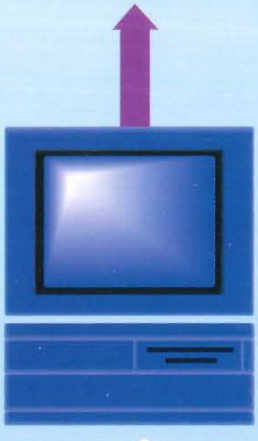
How Telnet Works

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.

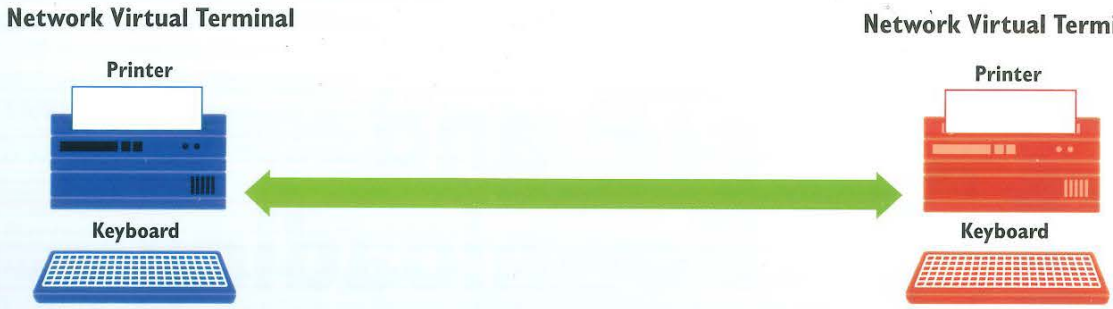








2 When you contact the host, the distant computer and your computer negotiate how they will communicate with each other. They decide which terminal emulation will be used. Terminal emulation determines how your keyboard will transmit information to the distant computer and how information will be displayed 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.



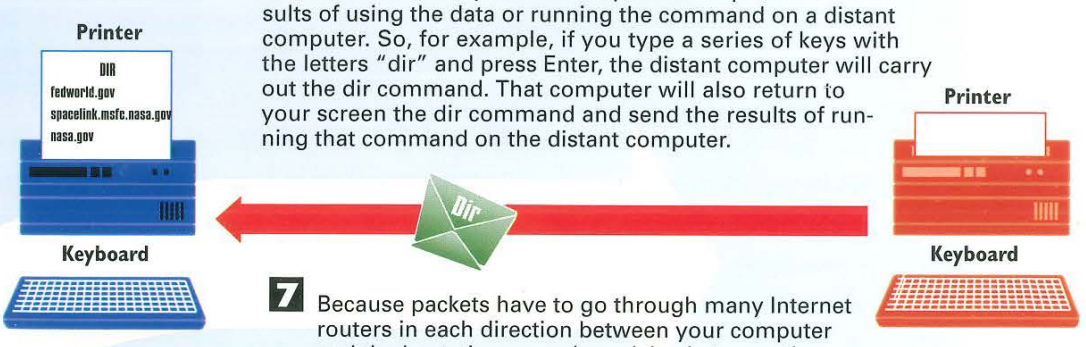
3 When a client and a server communicate, they use the Telnet protocol. The Telnet protocol assumes that each end of the connection—the client and the server—is a Network Virtual Terminal (NVT). Each NVT has a virtual "printer" and a virtual "keyboard." The keyboard sends data from one NVT to the other. 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 makes sure 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 will use 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.

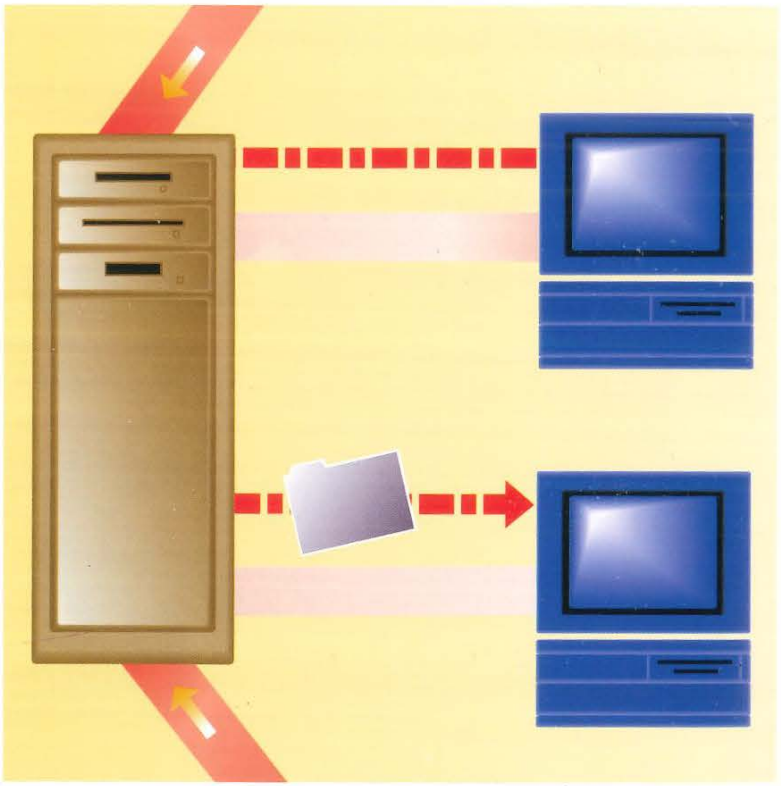
6 The Telnet host receives the data you've sent. It processes the data and returns to your screen (your NVT "printer") the results of using the data or running the command on a distant computer. So, for example, if you type a series of keys with the letters "dir" and press Enter, the distant computer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



7 Because packets have to go through many Internet routers in each direction between your computer and the host, there may be a delay between the time you send a command and the time you see the results on your own computer screen.

CHAPTER

23 FTP and Downloading Files



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 on the Internet. Most of those files are downloaded using the Internet's File Transfer Protocol, commonly referred to as FTP. FTP can also be used 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, an account number (or username) and a password must be typed in 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 email address as your password. Because of this, these sites are often referred to as anonymous FTP sites. Some FTP sites are private, only allowing certain people with the proper account number and password to enter.

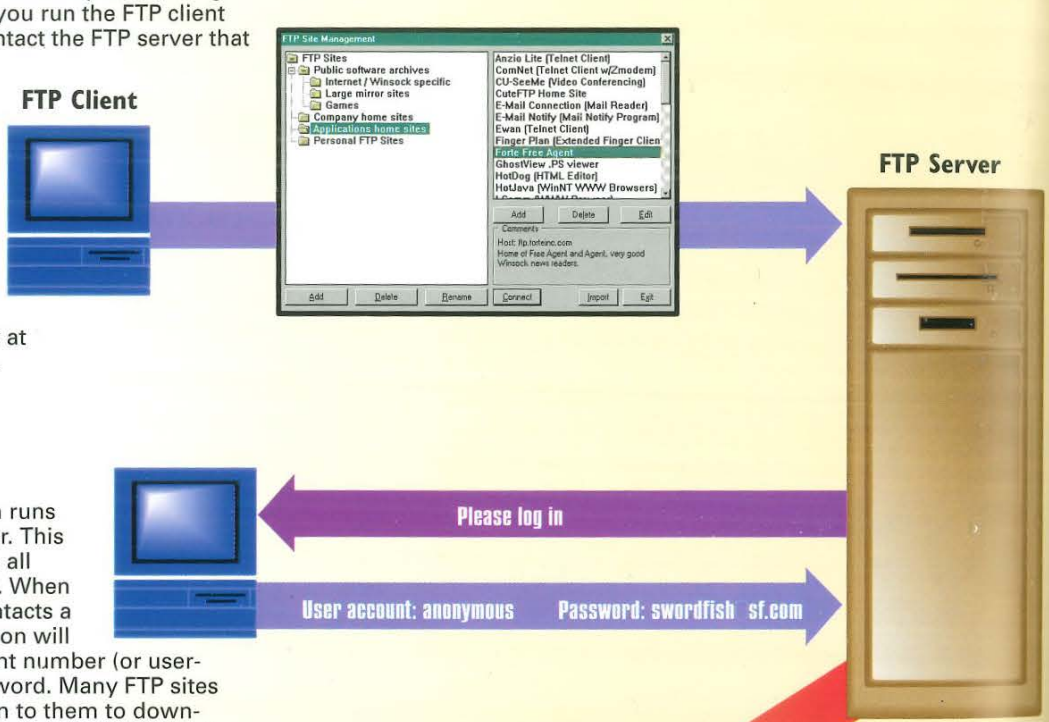
FTP is fairly simple to use. When you log on to an FTP site, you can browse through the available files by changing directories and you can see a listing of all the files available in each directory. When you see a file you want to download, you use your client software to instruct the FTP server to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

As the World Wide Web gains popularity, downloading software is becoming even easier. You can use your Web browser and click on 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.

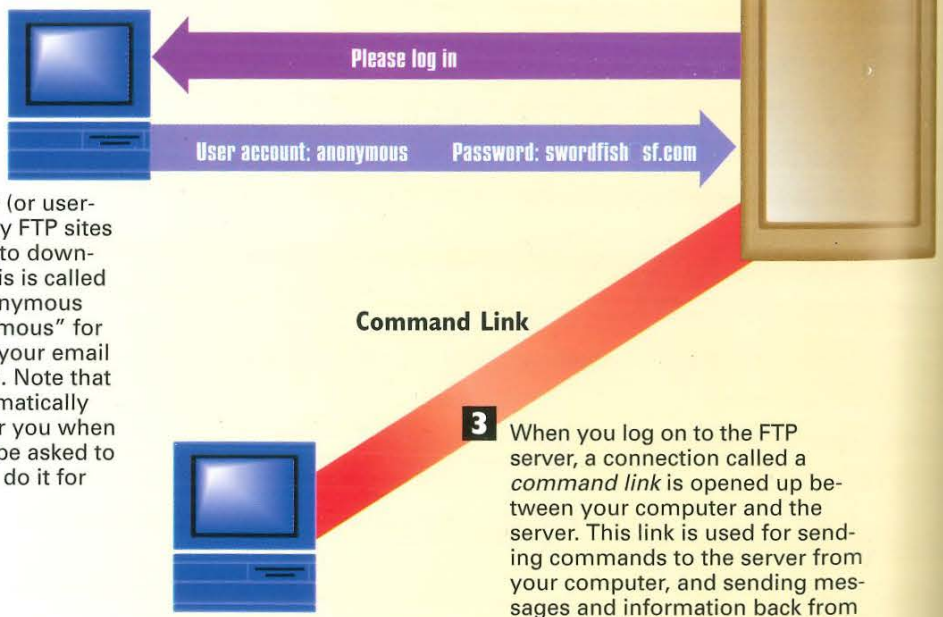
One problem with downloading files over the Internet is that some files are so large it can take a tremendous amount of time to download them, especially if the connection is made via modem. Even at 28,800bps, downloading files can be slow. As a way to speed up file transfers and save space on the FTP server, files are commonly compressed—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 to 50%. After the files have been downloaded, 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 resources, runs on a client/server model. So, in order to use it you'll need client software to run on your computer. To begin an FTP session, you run the FTP client software and contact the FTP server that you want to download files from. You can get FTP client software in hundreds of places on the Internet, including the ZDNet Software Library at `www.hotfiles.com`.



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 email 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—your software will do it for you automatically.



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

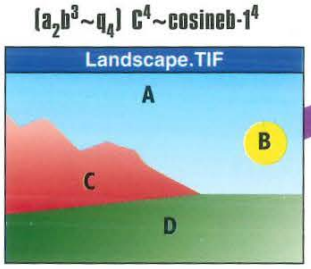
5 When you issue a command to download a file, a second connection is opened up, called the *data connection* (also called the *data link*). This connection can be opened up in one of two modes: the ASCII mode or the binary mode. The ASCII mode is used for sending text files and alters things such as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

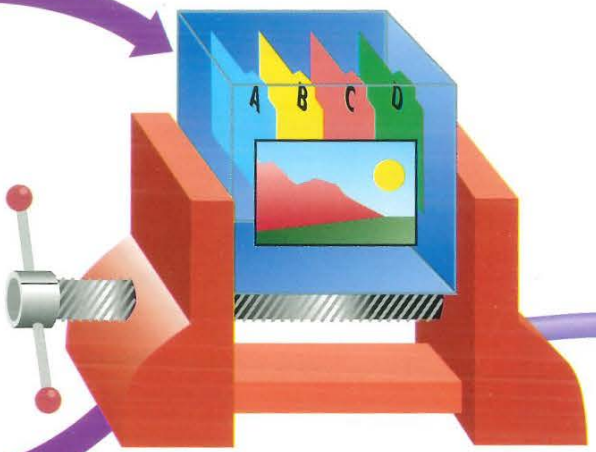
7 When the file is downloaded and the data link closes, the command link stays open. You can then change directories or download more files. After you're done, you log off and the command link closes. You're no longer connected to the FTP server.

How File Compression Works

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.



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, looking for repeating patterns of data.



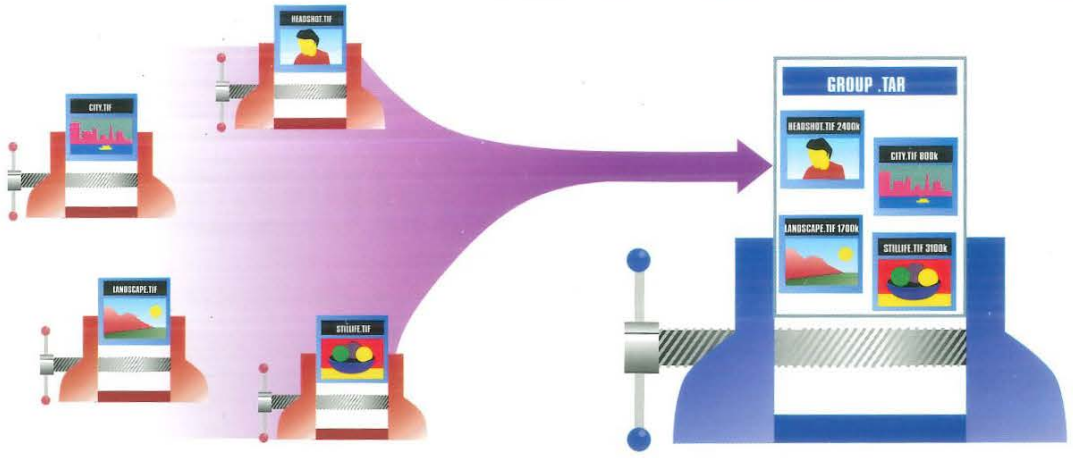


3 A "header" can also be added to the file as it is compressed. This header contains information about the file, such as the filename, the file size, and the compression method used. This information is used to help reconstruct the file when it is uncompressed.

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



Types of File-Compression Schemes You'll Find on the Internet

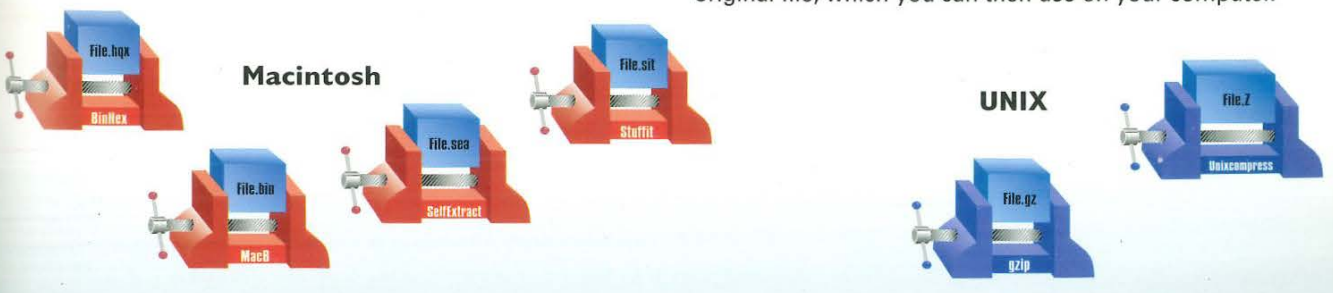


4 Some compression software, such as PKZIP for the PC, can also archive files, combining several compressed files. The UNIX command TAR can also combine many files into a single archive.

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

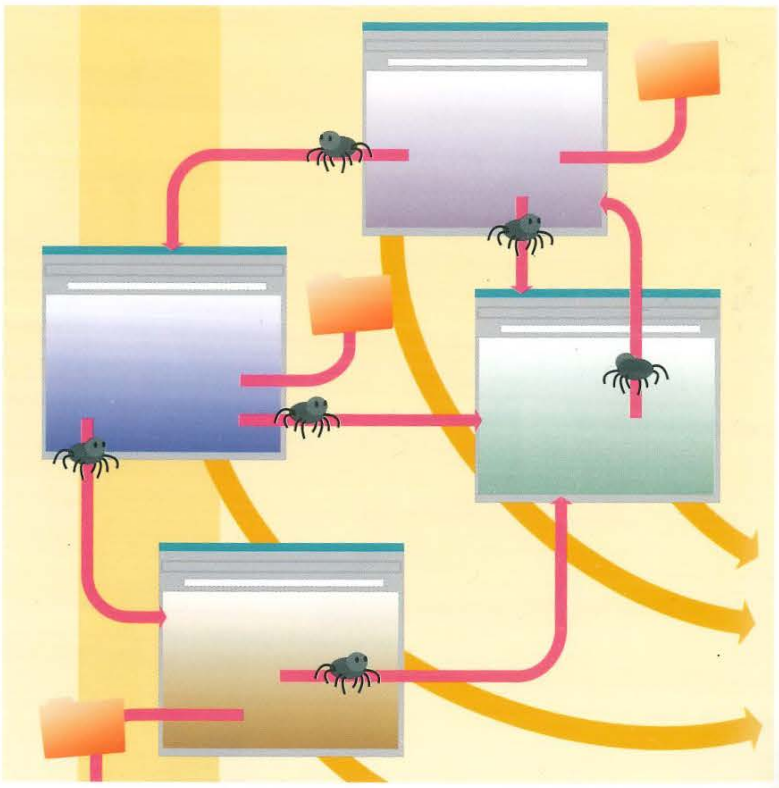


6 To use the file, you'll need decompression software on your computer. The decompression software looks into the file's header and examines the tokens in the file. Using a decompression algorithm, it reconstructs the original file, which you can then use on your computer.



CHAPTER

24 Searching the Internet



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 may be available. When you get to the subcategory you're interested in, you'll be 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.

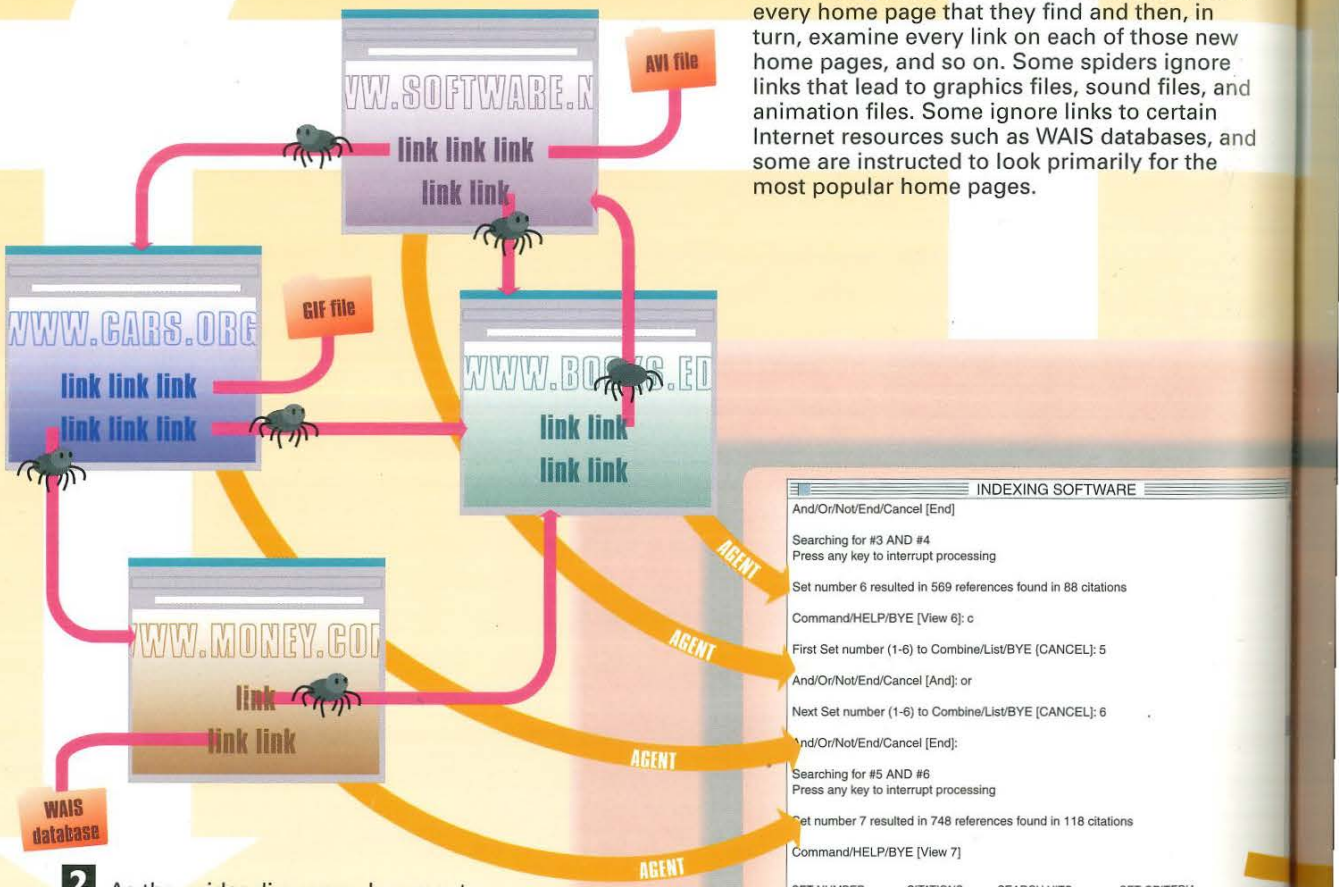
Many popular Internet search engines exist, including Lycos, Excite, and AltaVista. Although the specifics of how they operate differ somewhat, generally they are all composed of three parts: at least one spider, which crawls across the Internet gathering information; a database, which contains all the 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'd like to 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, 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 search engine site.

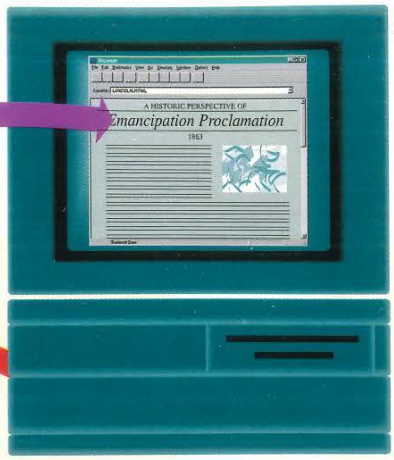
How Internet Search Engines Work

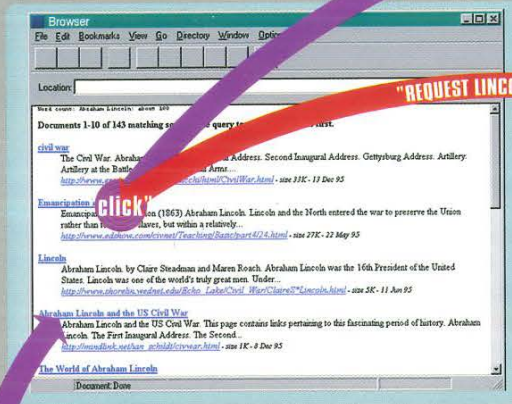
1 Each search engine uses a crawler or spider with its own set of rules guiding how documents are gathered. Some follow every link on every home page that they find and then, in turn, examine every link on each of those new home pages, and so on. Some spiders ignore links that lead to graphics files, sound files, and animation files. Some ignore links to certain Internet resources such as WAIS databases, and some are instructed to look primarily for the most popular home pages.



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.

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.

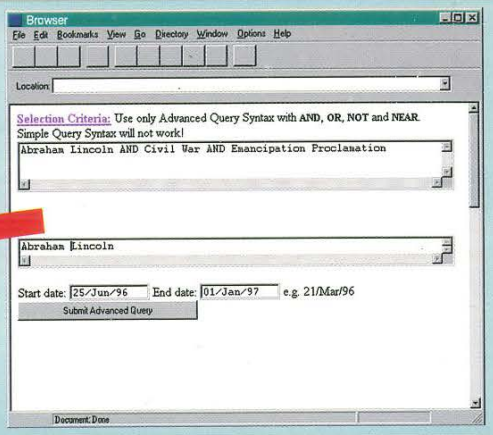




"REQUEST LINCOLN.HTML"

click!

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'll 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.



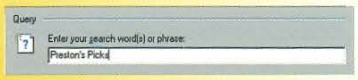
Database

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 kinds 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 will determine what kind of searching can be done with the search engine and how the information will be displayed.

How Meta-search Software Works

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. You can browse through the results on your computer and click those pages that you want to visit. This illustration shows how one such piece of meta-search software, called Copernic, works.

1 When you want to search for something on the Internet, you type descriptive words or a search term into the meta-search software.



Meta-Search Software

2 The meta-search software sends many "agents" 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.

Preston's Picks
http://www.hotfiles.com/home.html
http://www.hotfiles.com/index.html
http://www.hotfiles.com/prespick/presmain.html

Preston's Picks
http://www.hotfiles.com/index.html
http://www.hotfiles.com/prespick/presmain.html

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 then again sends the results back to the meta-search software.

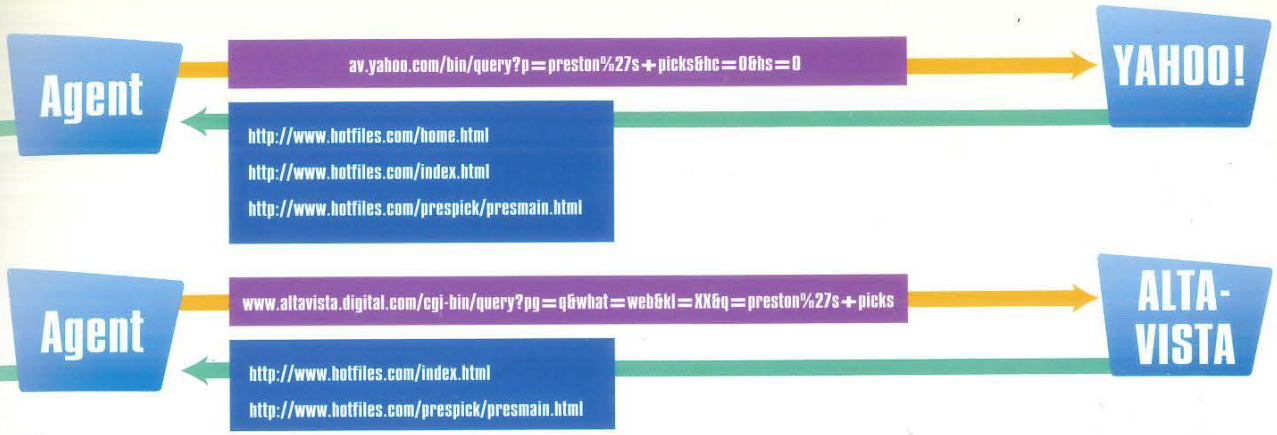
Preston's Picks
http://www.hotfiles.com/home.html
http://www.hotfiles.com/index.html
http://www.hotfiles.com/prespick/0498/pc.html

Preston's Picks
http://www.hotfiles.com/prespick/0498/pc.html
http://www.hotfiles.com/prespick/.../pres0498.html
http://www.hotfiles.com/prespick/prest097/pc.html

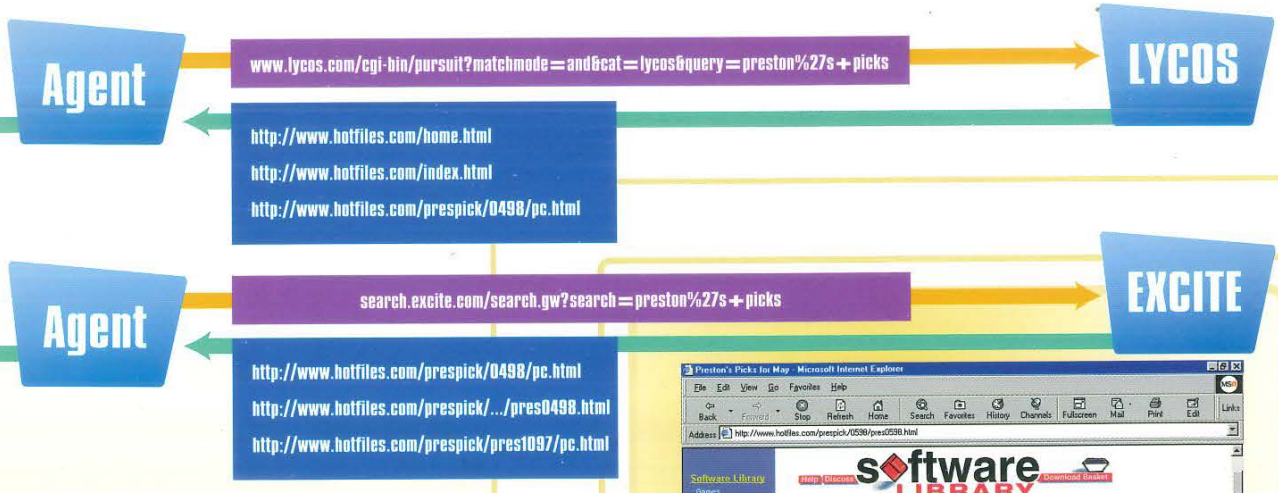
6 The meta-search software takes all the results from all the search engines and examines them for duplicate results. If it finds duplicate results, it deletes them. It then displays the results of the search, ranking each "hit" by the likelihood that it contains the information you requested. It figures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



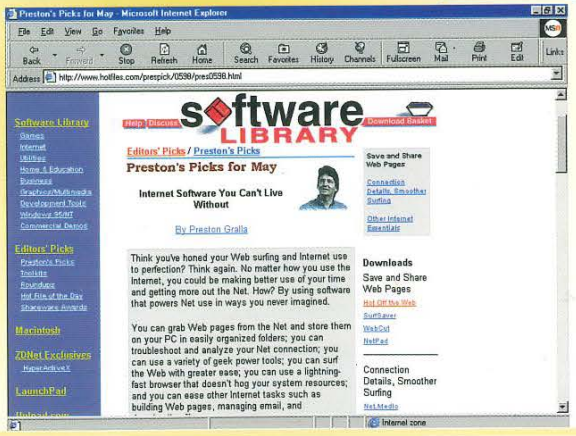
3 The agents are intelligent enough to know how each search engine functions—for example, whether a particular engine allows for Boolean searches (searching by using AND, OR and other variables). The agents also know the exact syntax that each engine requires. The agents put the search terms in the proper 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.

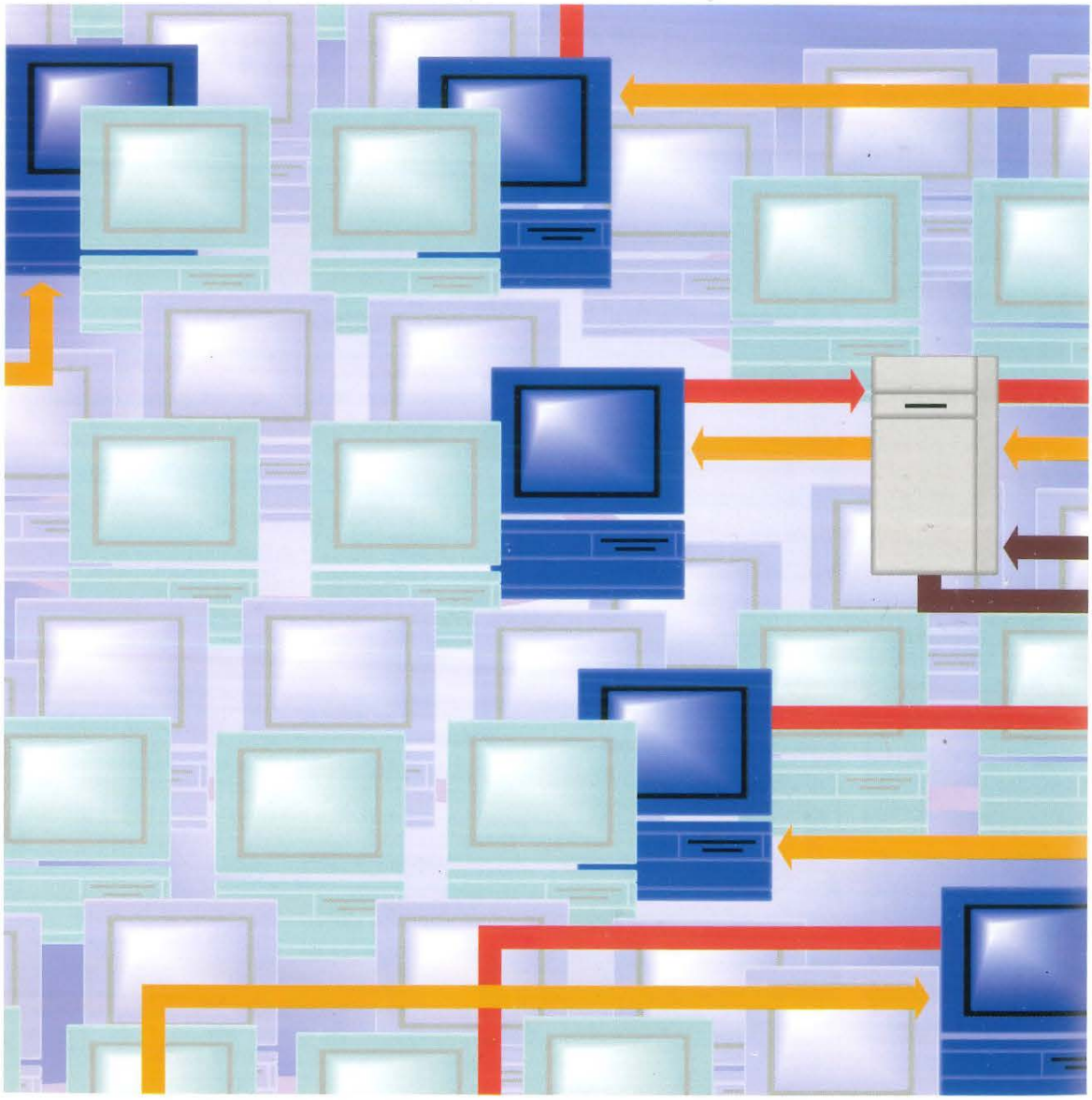


4 The search engines report the results of the search to each agent. The results typically include the URL of each site that matches the search, and often a summary of information found on the site, the date the site was last updated, and other data.



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'll then be sent to that site.





PART 6

HOW THE WORLD WIDE WEB WORKS

Chapter 25: How Web Pages Work
142

Chapter 26: How Web Browsers Work
150

Chapter 27: How Markup Languages Work
158

Chapter 28: How Hypertext Works
164

Chapter 29: How URLs Work
168

Chapter 30: How Imagemaps and Interactive Forms Work
172

Chapter 31: How Web Host Servers Work
176

Chapter 32: How Web Sites Work with Databases
180

WHEN many people use the word "Internet," they are really talking about the World Wide Web. The Web is the most interesting, the most innovative, the most visible, and the fastest growing part of the Internet. To a great degree, the explosive growth of the Web has been what's fueled the enormous amount of interest in the Internet in the past several years. When people refer to "surfing the Net," they're usually talking about using the World Wide Web.

This section of the book looks in great detail at how the World Wide Web works, from basic technologies such as how Web pages work, to more advanced features such as imagemaps, and the way that the Web is becoming increasingly integrated directly into your computer. We'll learn what the Web comprises, how a Web browser works, and thoroughly investigate URLs (uniform resource locators), as well as many other aspects of the Web and Web browsers.

Chapter 25, "How Web Pages Work," examines the most basic part of the Web. It starts by covering the general technology of the Web and describes how the Web works. Web pages are, in essence, multimedia publications that can include music, audio, video and animation, as well as graphics and text. Web pages are connected via hypertext that enables you to jump from any page to any other page, and to graphics, binary files, multimedia files, and other Internet resources. To jump from one page to another, or to another resource, you merely click a hypertext link.



The chapter details how all that works and looks at the client/server model of the Web. It shows what happens behind the scenes when you type a URL into your browser—how that information gets routed properly so that you visit the Web site that interests you. The chapter also delves into how Web pages are organized on a site and how an entire Web site works as a unit to deliver its information to you.

Chapter 26, "How Web Browsers Work," examines browsers. *Web browsers* are pieces of software that interpret the language of the Web, Hypertext Markup Language (HTML), and then display those results on your computer. The chapter begins by delving into how browsers do that job.

Browsers, however, now go far beyond merely allowing you to visit the Web. Both Netscape Communicator and Internet Explorer have grown to become full-blown application suites. Communicator, for example, includes the Navigator browser along with many other programs for communicating with others on the Internet. It includes a news reader, an email package, conferencing and whiteboard software, and even a powerful editing tool that enables you to create and publish your own Web pages.

Browsers also increasingly blur the distinction between the Internet and your own computer, as we'll see in Chapter 26. Internet Explorer allows your Windows desktop to be HTML-enabled,

which means you can have hyperlinks directly on your desktop to bring you to the Web. Even more significantly, it enables components created on the desktop to take information from the Internet automatically and put it on your PC; a real-time stock ticker is an example of this.

As mentioned earlier, HTML is the language of the Web. Chapter 27, "How Markup Languages Work," explains how HTML forms the building blocks for creating Web pages. The language is an essential set of directions that tells your browser how to display and manage a Web document. The chapter shows in detail how all that works.

Chapter 28, "How Hypertext Works," looks at hyperlinks, which set the Web apart from most other portions of the Internet. This chapter describes the various ways that documents can link to each other on the Web and explains concepts such as relative links and absolute links.

Chapter 29, "How URLs Work," looks in great detail at *URLs*, the addresses you type into your browser to visit a Web site. We'll look at the underlying structure of a URL and gain a better understanding of how a URL is put together and what it can tell you about the site you're visiting. We'll also see precisely how URLs help retrieve documents from the Web.



Chapter 30, "How Imagemaps and Interactive Forms Work," looks at two technologies you use every day on the Web without realizing it: imagemaps and interactive forms. *Imagemaps* aren't maps in a traditional sense. Instead, they're graphics with URLs embedded inside them. When you click one part of the graphic, you'll be sent to one site, and when you click another, you'll be sent to a different site. An imagemap could be a picture of a house, for example, and when you clicked the living room, you'd be sent to an entertainment site; when you clicked on a home office, you'd be sent to a business site.

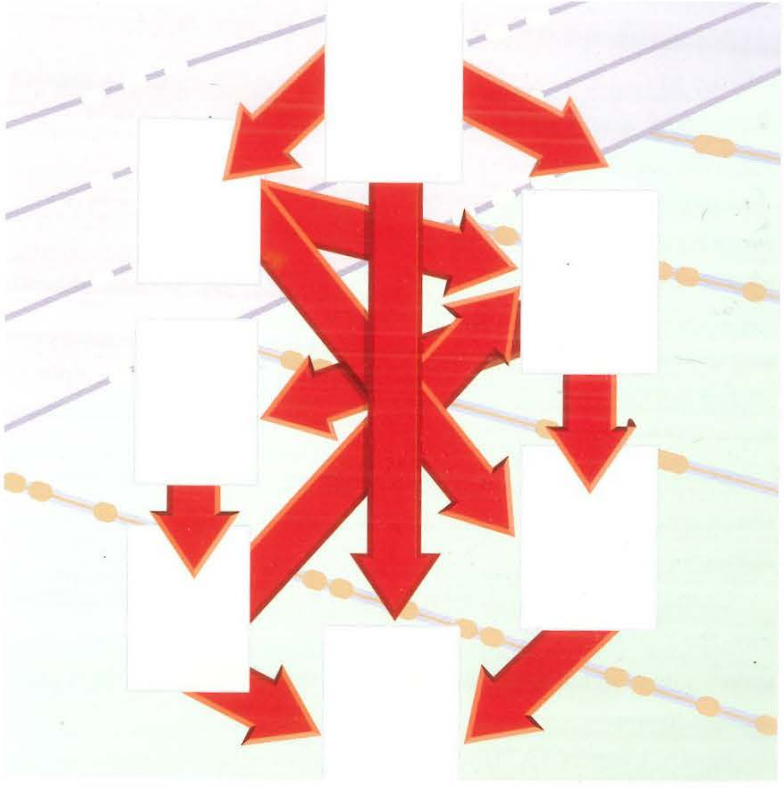
Interactive forms are the forms you fill out on the Web for doing things such as registering at a site or sending information about yourself before you're allowed to download a particular piece of software free of charge.

Chapter 31, "How Web Host Servers Work," looks at Web server software. As I mentioned before, the Web works on a client/server model. Your Web browser (the client) contacts the Web server. This chapter explains how Web server software interacts with your browser to deliver Web content to you.

Finally, Chapter 32, "How Web Sites Work with Databases," looks at databases. Databases are used on the Web for many things. Web indexes and search sites such as Yahoo! are, in essence, databases that interact with the Web.

CHAPTER

25 How Web Pages Work



THE World Wide Web is the fastest growing, and in many ways, the most exciting and intriguing part of the Internet. When people refer to "surfing the Net," more often than not they're talking about using the World Wide Web.

As its name implies, the World Wide Web is a globally connected network. The Web contains many things, but what makes it so fascinating to so many are the Web "pages" that incorporate text, graphics, sound, animation, and other multimedia elements. In essence, each page is an interactive multimedia publication that can include videos and music as well as graphics and text.

Pages are connected to each other using *hypertext* that allows you to move from any page to any other page, and to graphics, binary files, multimedia files, as well as any Internet resource. To jump from one page to another, click on a hypertext link.

The Web operates on a client/server model. You run Web client browser software such as Netscape Navigator or Microsoft's Internet Explorer on your computer. That client contacts a Web server and requests information or resources. The Web server locates and then sends the information to the Web browser, which displays the results.

Pages on the Web are built using a markup language called HTML (Hypertext Markup Language). The language contains commands that tell your browser how to display text, graphics, and multimedia files. It also contains commands for linking the page to other pages, and to other Internet resources.

The term "home page" is often used to refer to the first, or top, page in a collection of pages that make up a Web site. This is to distinguish it from the many pages collected together as a single "package" of sorts that often make up Web sites. The home page is like a magazine cover or the front page of a newspaper. Usually, it acts as an introduction to the site, explaining its purpose and describing the information found on other pages throughout the site. In this way, the home page often acts as the table of contents for the rest of the site.

In general, Web sites use three kinds of organizational structures to organize their pages. In a *tree structure*, a pyramid or outline format makes it easy for users to navigate through the site and find the information they want. In a *linear structure*, one page leads to the next, which then leads to the next, and so on, in a straight line. Finally, in a *random structure*, pages are connected to each other, seemingly at random.

The last illustration in this chapter shows how someone might build his or her own pages using an HTML editor. Once you create your pages, you use FTP software to post them onto a small portion of a Web server. You can either rent the server space from a local ISP or set up your own Web server.

How the World Wide Web Works

1 The World Wide Web is the fastest growing and most innovative part of the Internet. When you browse the Web, you view multimedia pages composed of text, graphics, sound, and video. The Web uses hypertext links that allow you to jump from one place to another on the Web. The language that allows you to use hypertext links and to view Web pages is called Hypertext Markup Language, more commonly known as HTML.

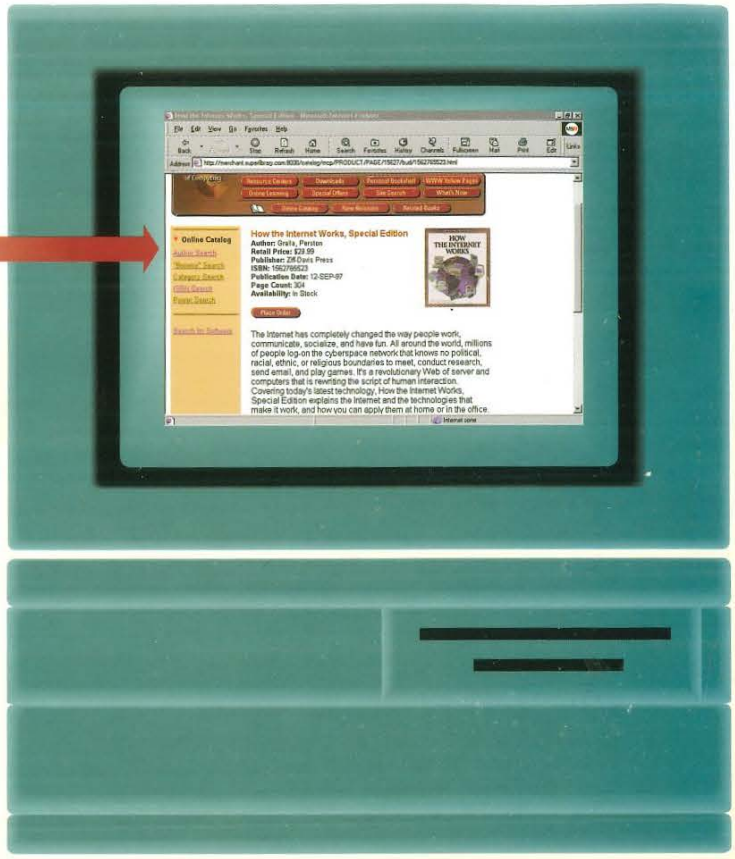


2 The Web works on a client/server model in which client software—known as a Web browser—runs on a local computer. The server software runs on a Web host. To use the Web, you first make an Internet connection, and then launch your Web browser.



3 In a Web browser, you type the URL for a location you want to visit or click on a link that will send you to the desired location. The names for Web locations are URLs (uniform resource locators). Your Web browser sends the URL request using HTTP (Hypertext Transfer Protocol), which defines the way the Web browser and the Web server communicate with one another.

7 When the server finds the requested home page, document, or object, it sends that home page, document, or object back to the Web browser client. The information is then displayed on the computer screen in the Web browser. When the page is sent from the server, the HTTP connection is closed and can be reopened.



6 The Web server receives the request using the HTTP protocol. It is told which specific document is being requested.

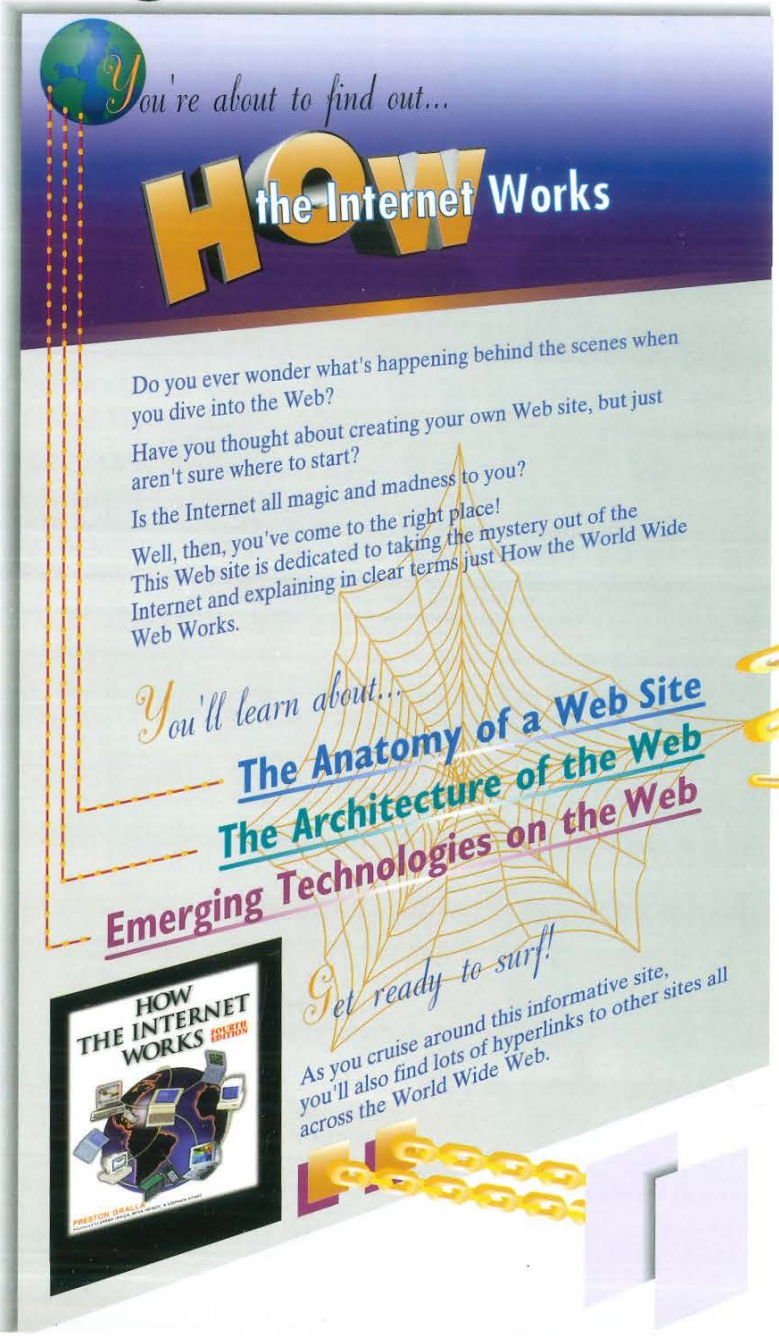


5 The request is sent to the Internet. Internet routers examine the request to determine which server to send the request to. The information just to the right of the http:// in the URL tells the Internet on which Web server the requested information can be found. Routers send the request to that Web server.



4 URLs contain several parts. The first part—the http://—details which Internet protocol to use. The second part—the part that usually has a www in it—sometimes tells what kind of Internet resource is being contacted. The third part—such as zdnet.com—can vary in length and identifies the Web server to be contacted. The final part identifies a specific directory on the server and a home page, document, or other Internet object.

How Web Pages Are Organized on a Web Site



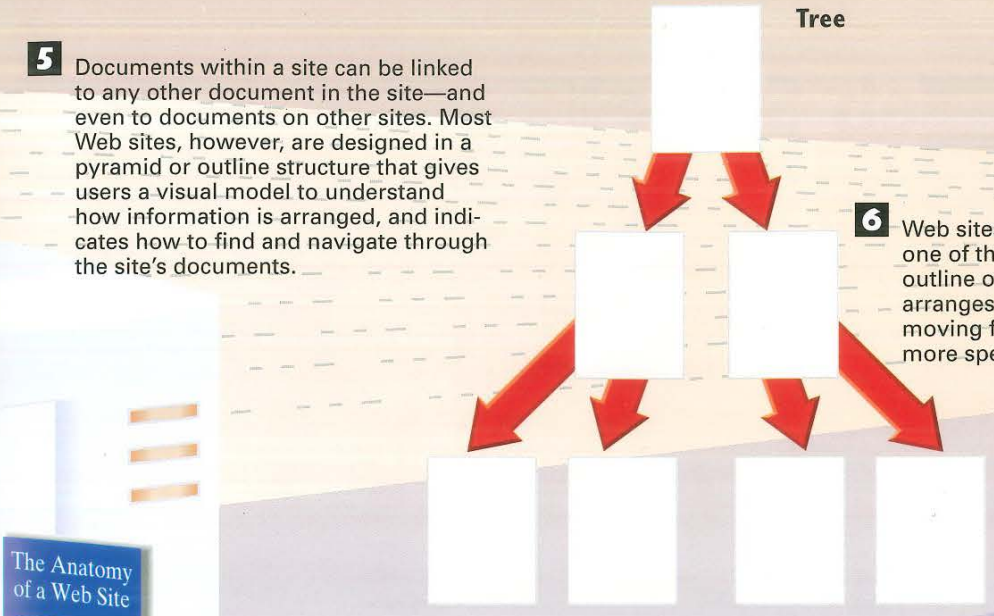
1. The home page is the first or top page of any Web site. A site can be just one page, or can comprise dozens or even hundreds of pages. In the latter case, the home page acts as a table of contents to organize the site and help users find information available on the site.

2. Underlined or highlighted hyperlink text is often embedded in the home page. The hyperlinks serve to connect the top page with other pages throughout the site.

3. Related documents residing together on a Web host computer make up a Web site. However, a single server can host multiple Web sites, each contained in a separate area or directory, much like a hard drive can accommodate multiple directories. Some Web sites are so large and heavily trafficked that they cannot fit on a single server and require multiple servers.

4. Good Web design principles suggest that pages throughout a site link back to the home page. This approach allows users to always find their way back to the top of a site to navigate in other directions.

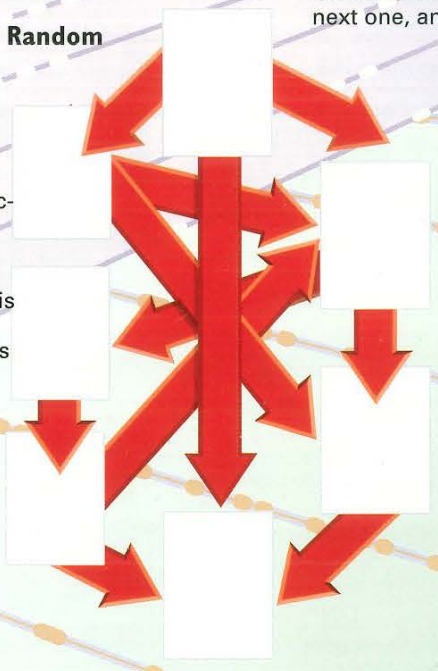
5 Documents within a site can be linked to any other document in the site—and even to documents on other sites. Most Web sites, however, are designed in a pyramid or outline structure that gives users a visual model to understand how information is arranged, and indicates how to find and navigate through the site's documents.



6 Web sites are typically organized in one of three ways. The first is an outline or tree structure, which arranges information hierarchically, moving from general information to more specific data.



7 The second organizational method is linear, in which one page leads to the next, which leads to the next one, and so on.



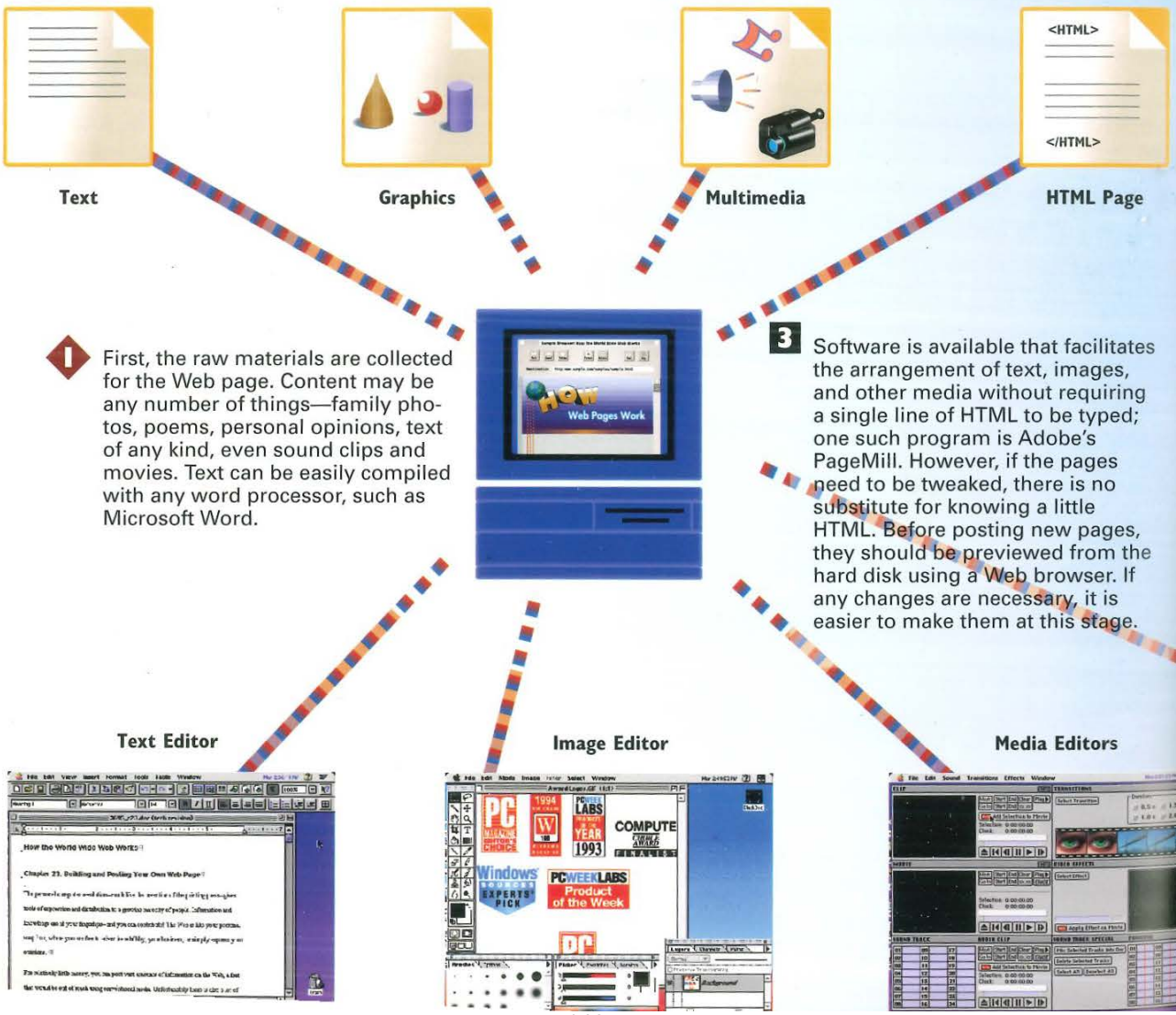
8 The third organizational structure is really a lack of structure, in which pages are connected to one another seemingly at random. (It's this last structure, though, that makes it clear why the Web is called the Web.)

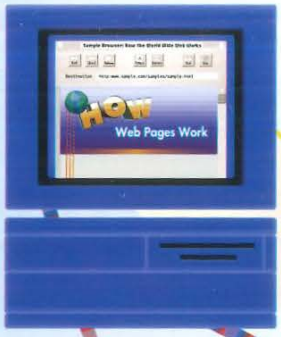






How Web Sites Work

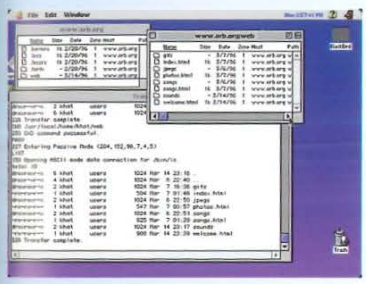




5 The Web page can typically be posted by means of a dial-up connection to the Internet (via a local ISP). Many ISPs provide the space for doing this free of charge as part of their monthly fee.



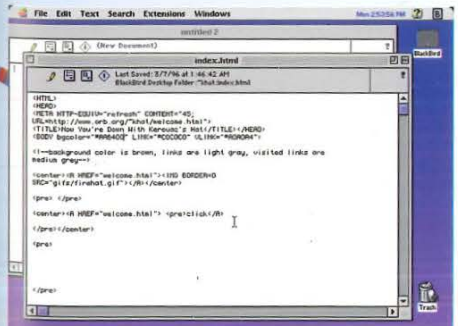
6 An FTP program is used to log on to the ISP's Web server; many good ones are available as shareware on the Internet. The ISP issues a username and password at sign-up time, which are used to log on to the directory where the HTML page, graphics, and other media will be placed. Once the various elements are put in the right folder, the "live" pages can be checked on the Web using the URL provided by the ISP.





2 A page can come to life with a few well-chosen images. In designing a family home page, for example, a scanned photo might work well on the Web page. Icons or colored bullets spruce up a text list, and navigational icons such as arrows or pointers let a user move logically around the site. A whole range of shareware image editors is available on the Internet. The most full-featured commercial image editor is Photoshop by Adobe. Photoshop is a professional-grade program that allows you to modify, create, or resize virtually any digital image.

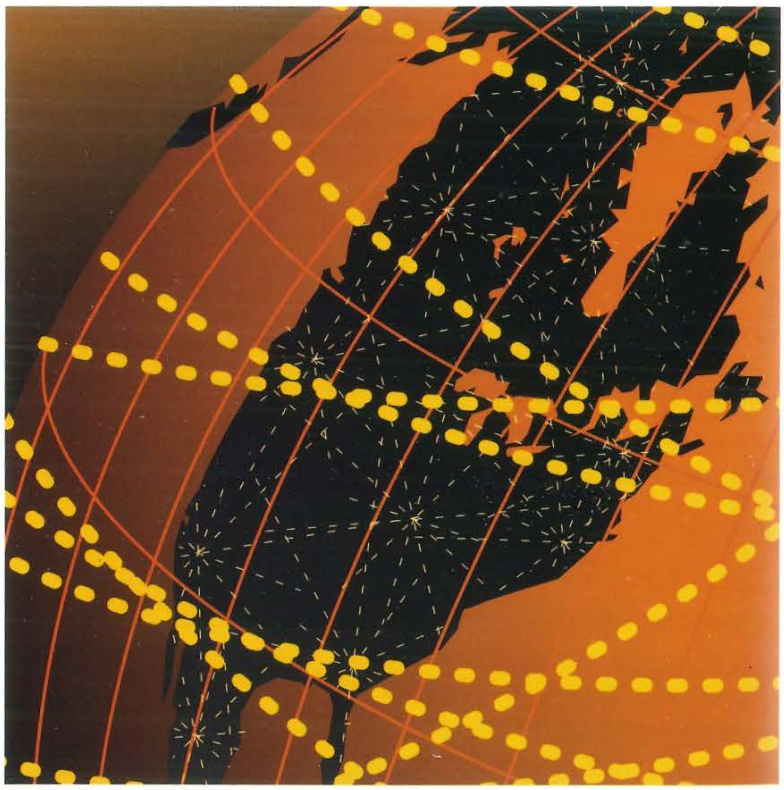
HTML Editor



4 If you want to add other media to your pages—such as digitized sound, music, or video—you will need access to the digital files, or you can digitize them yourself with additional computer hardware. When you are dealing with media with large file sizes—like a 4MB movie—it will take users a very long time to download the file. During editing, try to trim all the fat out of video and audio clips. Also try to scale back the length and resolution of audio and video to decrease file size.

CHAPTER 26

How Web Browsers Work



LIKE much of the Internet, the World Wide Web operates on a client/server model. You run a Web client on your computer—called a Web browser—such as Netscape Communicator, or Microsoft's Internet Explorer. That client contacts a Web server and requests information or resources. The Web server locates and then sends the information to the Web browser, which displays the results.

When Web browsers contact servers, they're asking to view pages built with Hypertext Markup Language (HTML). They interpret those pages and display them on your computer. They also can display applications, programs, animations, and similar material created with programming languages such as Java and ActiveX, and scripting languages such as JavaScript.

Sometimes, home pages contain links to files that the Web browser can't play or display, such as sound or animation files. In that case, you'll need a plug-in or a helper application. You configure your Web browser to use the helper application or plug-in whenever it encounters a sound or animation file that the browser can't run or play.

Over the years, Web browsers have become increasingly sophisticated. Browsers have now become full-blown software suites that can do everything from videoconferencing to letting you create and publish HTML pages. They have also begun to blur the line between your local computer and the Internet—in essence, they can make your computer and the Internet function as a single computer system.

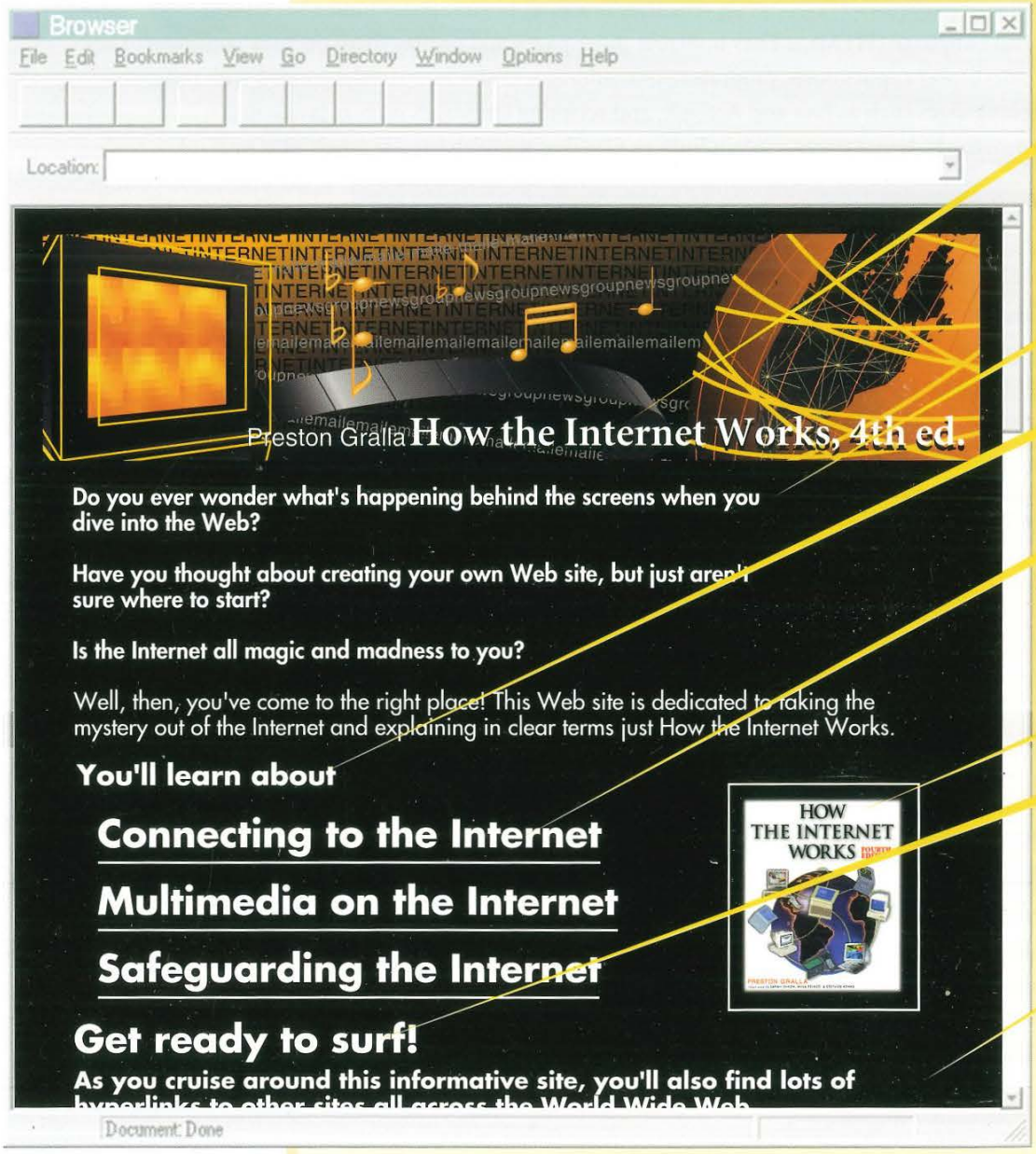
Toward this end of bringing the Internet more directly into computers, Microsoft has integrated Web browsing and the Internet directly into the operating system. For example, with Internet Explorer 4.0 and above, and with Windows 98, the Windows desktop can be HTML-based. This means Web links can be directly embedded into the desktop. So, for example, you can have links to your favorite Web pages right on the desktop. And even applications such as word processors now have Web capabilities built into them, such as being able to browse the Web, or build home pages. Even more significantly, using technology that Microsoft calls Active Desktop, Internet-based *desktop components* can live on the desktop. These components can be things such as stock tickers, which deliver live Web content directly to the desktop. You don't need to go fire up your Web browser to get the information; it's delivered straight to your Windows desktop without your having to do anything.

Both Microsoft and Netscape have also built entire suites of software around their browsers. Netscape, for example, calls its suite Netscape Communicator. Communicator includes modules for reading newsgroups; for reading, sending and managing Internet mail; for audio conferencing; for collaborative work on *whiteboard applications* in which people can view and mark up the same documents simultaneously; and more. These enhancements will help usher in an era of collaborative computing. Not only will the Internet be used to transmit and receive information, but it will also alter the way we do business and help us communicate more effectively.

How a Web Browser Works

1 Web browsers consist of client software that runs on your computer and displays home pages on the Web. There are clients for PC, Macintosh, and UNIX computers.

2 A Web browser displays information on your computer by interpreting the Hypertext Markup Language (HTML) that is used to build home pages on the Web. Home pages usually display graphics, sound, and multimedia files, as well as links to other pages, files that can be downloaded, and other Internet resources.



```
<HTML>
<HEAD>
<TITLE>
</HEAD>
```

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 that 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.

```
<BODY>
<P> Do you ever wonder what's happening behind
the screens when you dive into the Web?

<H3> You'll learn about:

<A HREF="URL"> Go TO URL</A>
```

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 will perform different actions. For example, if the HTML code specifies the link as another home page, the browser will retrieve the URL specified in the HTML file when the user clicks on the underlined link on the page. If the HTML code specifies a file to be downloaded, the browser will download the file to your computer.

```
<P> <IMG SRC="BOOK.GIF"

<H2> Get ready to surf!

<P> As you cruise around this informative site,
you'll also find lots of hyperlinks to other sites all
across the World Wide Web.
```

```
</BODY>
<HTML>
```

NOTE There are many kinds of files on the Internet that Web browsers cannot 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 kinds of files on Web pages. To view or play these files, you'll need what are called "helper applications" and plug-ins. You must configure your Web browser to launch these helper applications and plug-ins whenever you click on an object that needs them in order to be viewed. Helper applications and plug-ins can also be used for displaying virtual reality pages, for chatting on the Internet, and for doing other Internet tasks.

NOTE 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 will frequently be surrounded by tags like this: <P> Paragraph of text.</P>. Also, tags are not case sensitive. <P> equals <p>.

How Netscape Communicator Works

1 Netscape Communicator is a complete suite of Internet applications, not merely a Web browser. Any individual component can be run alone, or it can be run in concert with other components.



2 The centerpiece of Communicator is Netscape Navigator, the Web browser component.



7 Communicator also includes an HTML editor that allows you to edit, create, and post HTML to the Web.



Web Server



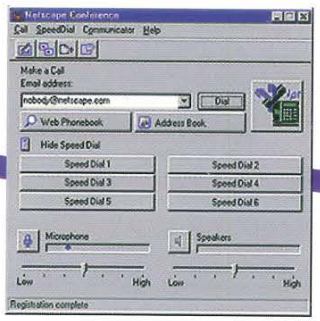


6 The Netcaster component of Communicator allows for push technology to send information to you over the Internet, and for "subscriptions" to Web sites. This saves time and allows information to be read offline while you're not connected to the Internet.



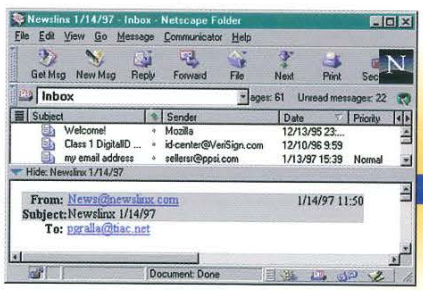
3 Netscape Conference is the workgroup and multimedia communications component of Communicator, used primarily on intranets. It allows people to hold group conference calls from their computers, as well as view and use shared files and documents together. This is called a *whiteboard* application.

Conference





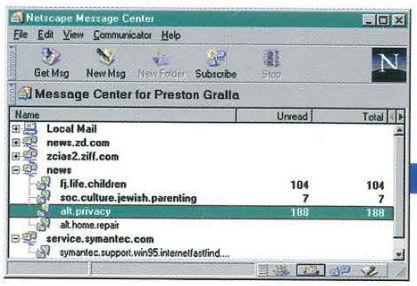
Messenger

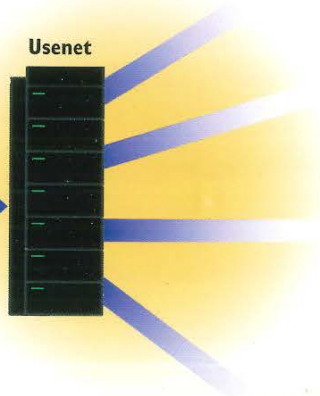


4 Communicator includes an email component called Netscape Messenger. Unlike some other email packages, Messenger can read HTML documents so you can receive entire Web pages as well as messages that include graphics and multimedia content.



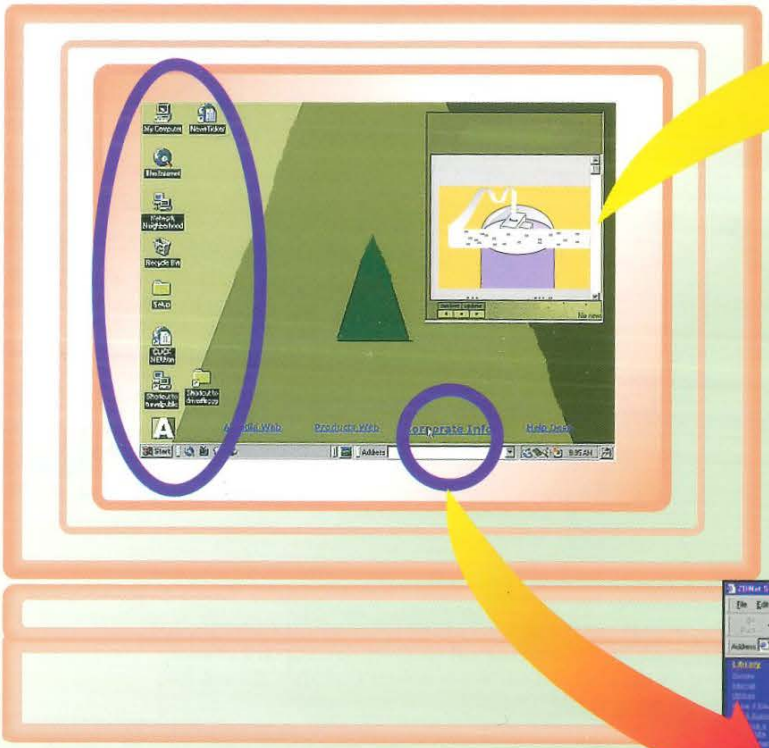
5 Two other Communicator components allow for reading newsgroups and online discussions on the Internet. A Usenet newsgroup reader gives access to Usenet, whereas software called Collabra allows for intranet-based discussions.



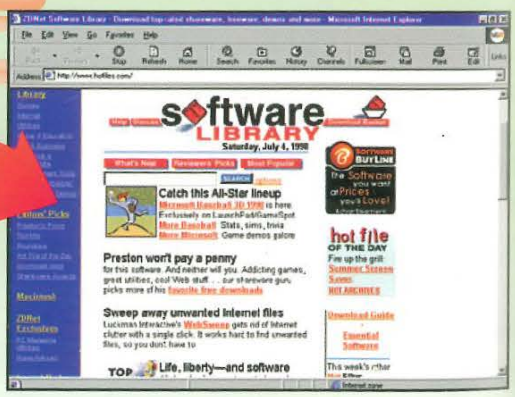


How Internet Explorer Integrates into Windows

1 Internet Explorer 4.0 and above—and Windows 98—blur the distinction between your computer and the Internet. This change has a great impact on Web browsing. The Active Desktop is primarily responsible for this enhancement.

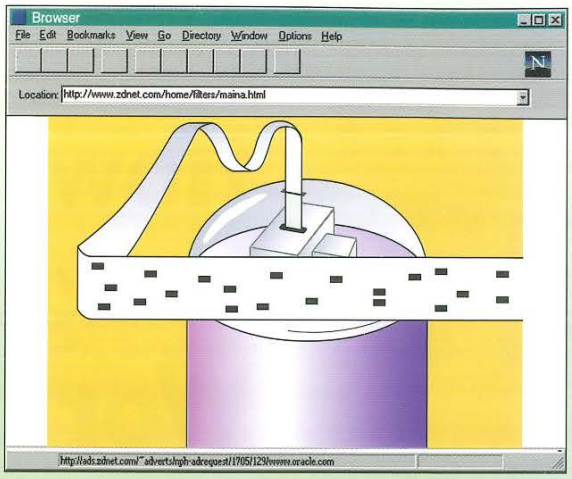


2 The Active Desktop is built with several layers. One layer, the background, allows your normal Windows 95 background to function as an HTML page instead of wallpaper or a color. So, for example, Web links could be placed on the background of your desktop. When you click on the link, it launches Internet Explorer and sends you to the Web site.



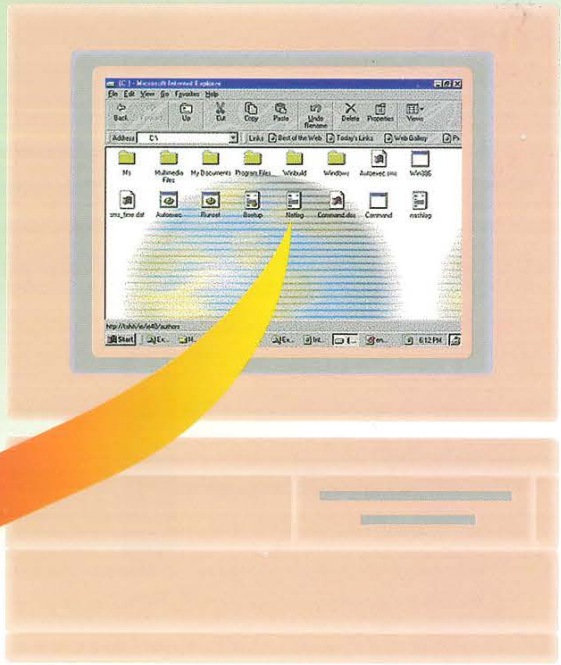
5 On top of the background HTML layer is the icon layer. The icons are the same shortcuts that exist in Windows. When icons are clicked upon, they launch a program or a file, open a folder, or do anything else that icons normally do in Windows.

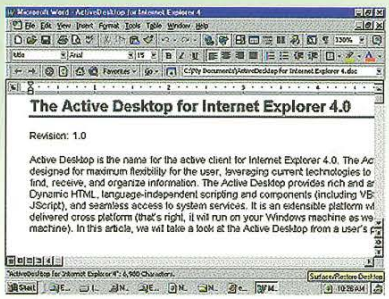
3 *Desktop components* can be placed on this background HTML layer. Desktop components are HTML frames that contain HTML-based content, sites, and applications. For example, a desktop component may be a stock ticker that displays constantly changing stock prices.



4 A desktop component, such as a stock ticker, can actually be a Web page or component that exists on the Internet. The Active Desktop contacts the Web server hosting the page, downloads it to your local PC, and then displays it within the desktop component frame, resizing the component or page to fit into the frame.

6 Another way Internet Explorer 4.0 and above and Windows 98 integrate your local PC with the Internet is by allowing you to browse the contents of your computer while you're using your browser on the Internet. When you type the drive letter of your computer (such as C:), you will be able to see the folder structure of your computer from within Internet Explorer.

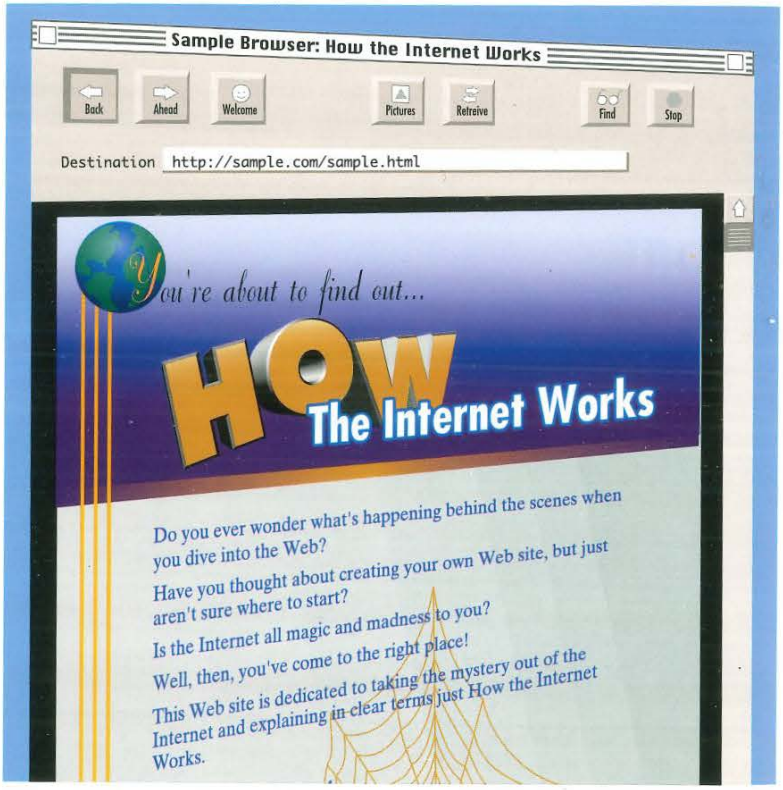




7 You can then launch any document by double-clicking it, just as you would normally do in Windows. You'll read the document in your normal application, not within Internet Explorer.

CHAPTER 27

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 or she 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 that 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.

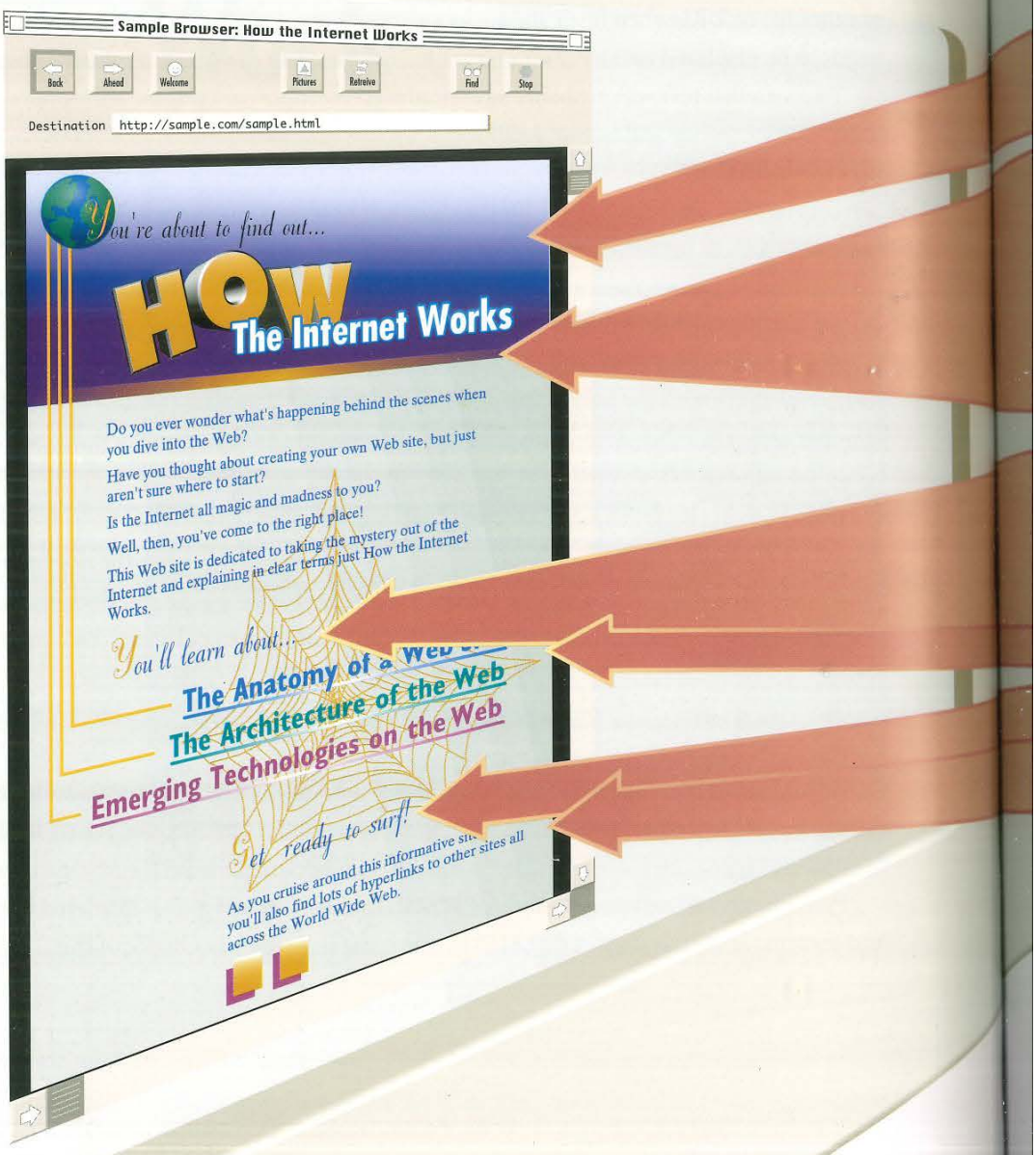
Hypertext Markup Language (HTML) is the markup language of the Web. It defines the format of a Web document and enables hypertext links to be embedded in the document. You can use any text editor or word processor to add HTML tags to an ASCII text document, although a number of shareware and commercially 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, which together are termed *Dynamic HTML* or 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. It may take awhile for this technology to catch on, but when it does, it should change the way everyone uses the Web.

Perhaps the best way to get a feel for HTML is to look at text as it is displayed on your screen and text that is "marked up" using HTML. Turn the page for an example of this.

How HTML Works

1 To display Web pages in any browser, you need to 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.

Display text "How the Internet Works"

Display text "Do you ever wonder..."

Display text "You'll learn about. . ."

Create a link to this site

Display text "Get ready to surf!"

Display text "As you cruise..."

```
<HTML>
<HEAD>
<TITLE>Sample Browser: How The Internet Works</TITLE>
</HEAD>

<BODY background= "Spider.GIF">
<IMG SRC = "top.gif">
<P>
<BLOCKQUOTE>
Do you ever wonder what's happening behind the scenes when you dive into the Web?
<P>
Have you thought about creating your own Web site, but just aren't sure where to
start?
<P>
Is the Internet all magic and madness to you?
<P>
Well, then, you've come to the right place!
<P>
This web site is dedicated to taking the mystery out of the Internet and explaining in
clear terms just How the Internet Works.

<H1> You'll learn about...</H1>
<BLOCKQUOTE>
<A HREF = "anatomy.html">The Anatomy of a Web Site</A>
<P>
<A HREF = "architecture.html">The Architecture of the Web</A>
<P>
<A HREF = "emerging.html">Emerging Technologies on the Web</A>
</BLOCKQUOTE>
</BLOCKQUOTE>

<IMG SRC = "footer.gif" ALIGN=LEFT>

<BLOCKQUOTE>
<H2> Get ready to surf!</H2>

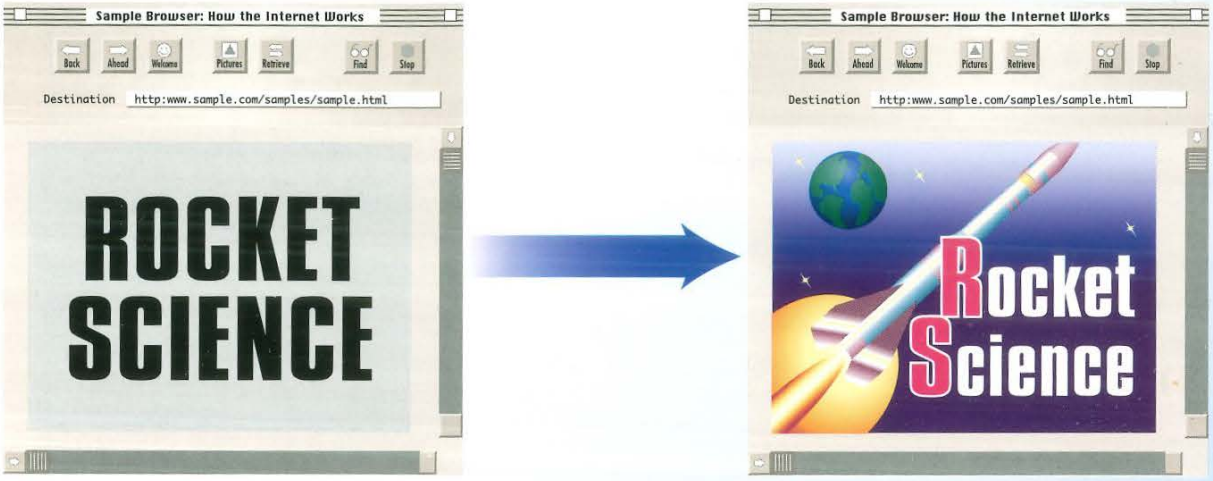
As you cruise around this informative site, you'll also find lots of hyperlinks to other
sites all across the World Wide Web.
</BLOCKQUOTE>

</BODY>
</HTML>
```

4 Most Web browsers will 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 may 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, once 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 you doing anything.



2 DHTML does its work without having to contact the server after the page downloads, which means that 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.

