

Chapter 1

Computer Networks and the Internet

Computer networking is one of the most exciting and important technological fields of our time. The Internet interconnects millions (and soon billions) of computers, providing a global communication, storage, and computation infrastructure. Moreover, the Internet is currently being integrated with mobile and wireless technology, ushering in an impressive array of new applications. Yes, computer networking has indeed come a long way since its infancy in the 1960s. But this is only the beginning—a new generation of creative engineers and computer scientists will drive the Internet to yet unforeseen terrains. This book will provide today’s students with the vehicles they need to journey to and explore the new lands in this exciting field.

This first chapter presents an overview of computer networking and the Internet. Our goal here is to paint a broad-brush picture of computer networking, to see the forest through the trees. We’ll cover a lot of ground in this introductory chapter and discuss a lot of “pieces” of a computer network, while not losing sight of the “big picture.” The chapter lays the groundwork for the rest of the book. It can also be used for a mini-course on computer networking.

In this chapter, after introducing some basic terminology and concepts, we will first examine the “edge” of a computer network. We’ll look at the end systems and network applications, and the transport services provided to these applications. We’ll then explore the “core” of a computer network, examining the links and the switches that transport data, as well as the access networks and physical media that

connect end systems to the network core. We'll learn that the Internet is a network of networks, and we'll learn about how these networks connect with each other.

After having completed this overview of the “edge” and “core” of a computer network, we'll take a broader view. We'll examine the causes of data-transfer delay and loss in a computer network, and provide simple quantitative models for end-to-end delay, models that take into account transmission, propagation, and queuing delays. We'll then introduce some of the key architectural principles in computer networking, namely, protocol layering and service models. Finally, we'll close this chapter with a brief history of computer networking.

1.1 ♦ What Is the Internet?

In this book we use the public Internet, a specific computer network, as our principal vehicle for discussing computer network protocols. But what is the Internet? We would like to be able to give you a one-sentence definition of the Internet, a definition that you can take home and share with your family and friends. Alas, the Internet is very complex, both in terms of its hardware and software components, as well as in the services it provides.

1.1.1 A Nuts-and-Bolts Description

Instead of giving a one-sentence definition, let's try a more descriptive approach. There are a couple of ways to do this. One way is to describe the nuts and bolts of the Internet, that is, the basic hardware and software components that make up the Internet. Another way is to describe the Internet in terms of a networking infrastructure that provides services to distributed applications. Let's begin with the nuts-and-bolts description, using Figure 1.1 to illustrate our discussion.

The public Internet is a worldwide computer network, that is, a network that interconnects millions of computing devices throughout the world. Most of these computing devices are traditional desktop PCs, UNIX-based workstations, and so-called servers that store and transmit information such as Web pages and e-mail messages. Increasingly, nontraditional Internet end systems such as PDAs (Personal Digital Assistants), TVs, mobile computers, automobiles, and toasters are being connected to the Internet. (Toasters [Toasty 2002] are not the only rather unusual devices to have been hooked up to the Internet; see “Internet Home Appliances” [Appliance 2002].) In the Internet jargon, all of these devices are called **hosts** or **end systems**. As of January 2002 there were 100–500 million end systems using the Internet; and this number continues to grow exponentially [ISC 2002].

End systems are connected together by **communication links**. We'll see in Section 1.4 that there are many types of communication links, which are made up of different types of physical media, including coaxial cable, copper wire, fiber optics,

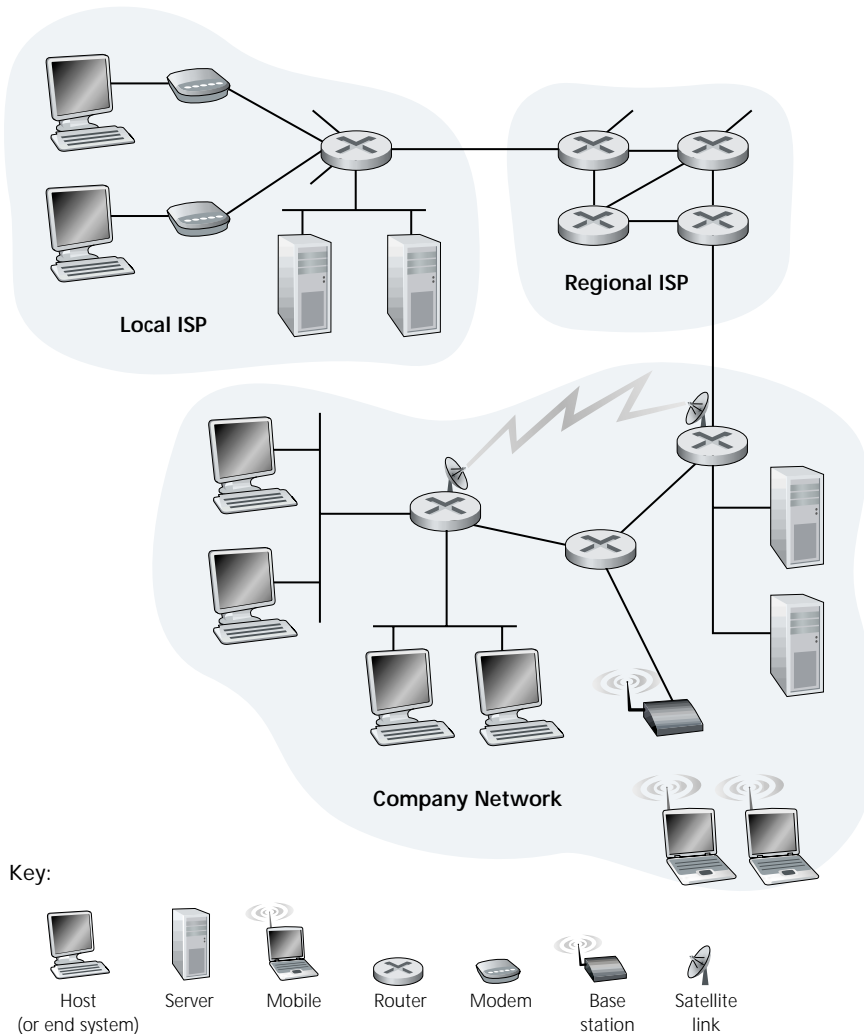


Figure 1.1 ♦ Some pieces of the Internet

and radio spectrum. Different links can transmit data at different rates. The link transmission rate is often called the **bandwidth** of the link, which is typically measured in bits/second.

End systems are not usually directly attached to each other via a single communication link. Instead, they are indirectly connected to each other through intermediate switching devices known as **routers**. A router takes a chunk of information arriving on one of its incoming communication links and forwards that chunk of

information on one of its outgoing communication links. In the jargon of computer networking, the chunk of information is called a **packet**. The path that the packet takes from the sending end system, through a series of communication links and routers, to the receiving end system is known as a **route** or **path** through the network. Rather than provide a *dedicated* path between communicating end systems, the Internet uses a technique known as **packet switching** that allows multiple communicating end systems to share a path, or parts of a path, at the same time. The first packet-switched networks, created in the 1970s, are the earliest ancestors of today's Internet.

End systems access the Internet through **Internet Service Providers (ISPs)**, including residential ISPs such as AOL or MSN, university ISPs such as Stanford University, and corporate ISPs such as Ford Motor Company. Each ISP is a network of routers and communication links. The different ISPs provide a variety of different types of network access to the end systems, including 56 Kbps dial-up modem access, residential broadband access such as cable modem or DSL, high-speed LAN access, and wireless access. ISPs also provide Internet access to content providers, connecting Web sites directly to the Internet. To allow communication among Internet users and to allow users to access worldwide Internet content, these lower-tier ISPs are interconnected through national and international upper-tier ISPs, such as the UUNet and Sprint. An upper-tier ISP consists of high-speed routers interconnected with high-speed fiber-optic links. Each ISP network, whether upper-tier or lower-tier, is managed independently, runs the IP protocol (see below), and conforms to certain naming and address conventions. We will examine ISPs and their interconnection more closely in Section 1.5.

End systems, routers, and other “pieces” of the Internet, run **protocols** that control the sending and receiving of information within the Internet. **TCP** (the Transmission Control Protocol) and **IP** (the Internet Protocol) are two of the most important protocols in the Internet. The IP protocol specifies the format of the packets that are sent and received among routers and end systems. The Internet's principal protocols are collectively known as **TCP/IP**. We begin looking into protocols in this introductory chapter. But that's just a start—much of this book is concerned with computer network protocols!

The public Internet (that is, the global network of networks discussed above) is the network that one typically refers to as *the* Internet. There are also many private networks, such as many corporate and government networks, whose hosts cannot exchange messages with hosts outside of the private network (unless the messages pass through so-called firewalls, which restrict the flow of messages to and from the network). These private networks are often referred to as **intranets**, as they use the same types of hosts, routers, links, and protocols as the public Internet.

At the technical and developmental level, the Internet is made possible through creation, testing, and implementation of **Internet standards**. These standards are developed by the Internet Engineering Task Force (IETF)[IETF 2002]. The IETF standards documents are called **RFCs** (request for comments). RFCs started out as general requests for comments (hence the name) to resolve architecture problems

that faced the precursor to the Internet. RFCs, though not formally standards, have evolved to the point where they are cited as such. RFCs tend to be quite technical and detailed. They define protocols such as TCP, IP, HTTP (for the Web), and SMTP (for open-standards e-mail). There are more than 3,000 different RFCs.

1.1.2 A Service Description

The preceding discussion has identified many of the pieces that make up the Internet. Let's now leave the nuts-and-bolts description and take a service-oriented view.

- ◆ The Internet allows **distributed applications** running on its end systems to exchange data with each other. These applications include remote login, electronic mail, Web surfing, instant messaging, audio and video streaming, Internet telephony, distributed games, peer-to-peer (P2P) file sharing, and much, much more. It is worth emphasizing that “the Web” is not a separate network but rather just one of many distributed applications that use the communication services provided by the Internet.
- ◆ The Internet provides two services to its distributed applications: a **connection-oriented reliable service** and a **connectionless unreliable service**. Loosely speaking, the connection-oriented reliable service guarantees that data transmitted from a sender to a receiver will eventually be delivered to the receiver in order and in its entirety. The connectionless unreliable service does not make any guarantees about eventual delivery. Typically, a distributed application makes use of one or the other (but not both) of these two services.
- ◆ Currently, the Internet does not provide a service that makes promises about *how long* it will take to deliver the data from sender to receiver. And except for increasing your access bandwidth to your Internet service provider, you currently cannot obtain better service (for example, bounded delays) by paying more—a state of affairs that some (particularly Americans!) find odd. We'll take a look at state-of-the-art Internet research that is aimed at changing this situation in Chapter 6.

This second description of the Internet—that is, in terms of the services it provides to distributed applications—is a nontraditional, but important, one. Increasingly, advances in the nuts-and-bolts components of the Internet are being driven by the needs of new applications. So it's important to keep in mind that the Internet is an *infrastructure* in which new applications are being constantly invented and deployed.

We have just given two descriptions of the Internet, one in terms of its hardware and software components, the other in terms of the services it provides to distributed applications. But perhaps you are still confused as to what the Internet is. What are packet switching, TCP/IP, and connection-oriented service? What are routers? What kinds of communication links are present in the Internet? What is a distributed application? If you feel a bit overwhelmed by all of this now, don't worry—the

purpose of this book is to introduce you to both the nuts and bolts of the Internet, as well as the principles that govern how and why it works. We will explain these important terms and questions in the subsequent sections and chapters.

1.1.3 What Is a Protocol?

Now that we've got a bit of a feel for what the Internet is, let's consider another important buzzword in computer networking: "protocol." What *is* a protocol? What does a protocol *do*? How would you recognize a protocol if you met one?

A Human Analogy

It is probably easiest to understand the notion of a computer network protocol by first considering some human analogies, since we humans execute protocols all of the time. Consider what you do when you want to ask someone for the time of day. A typical exchange is shown in Figure 1.2. Human protocol (or good manners, at least)

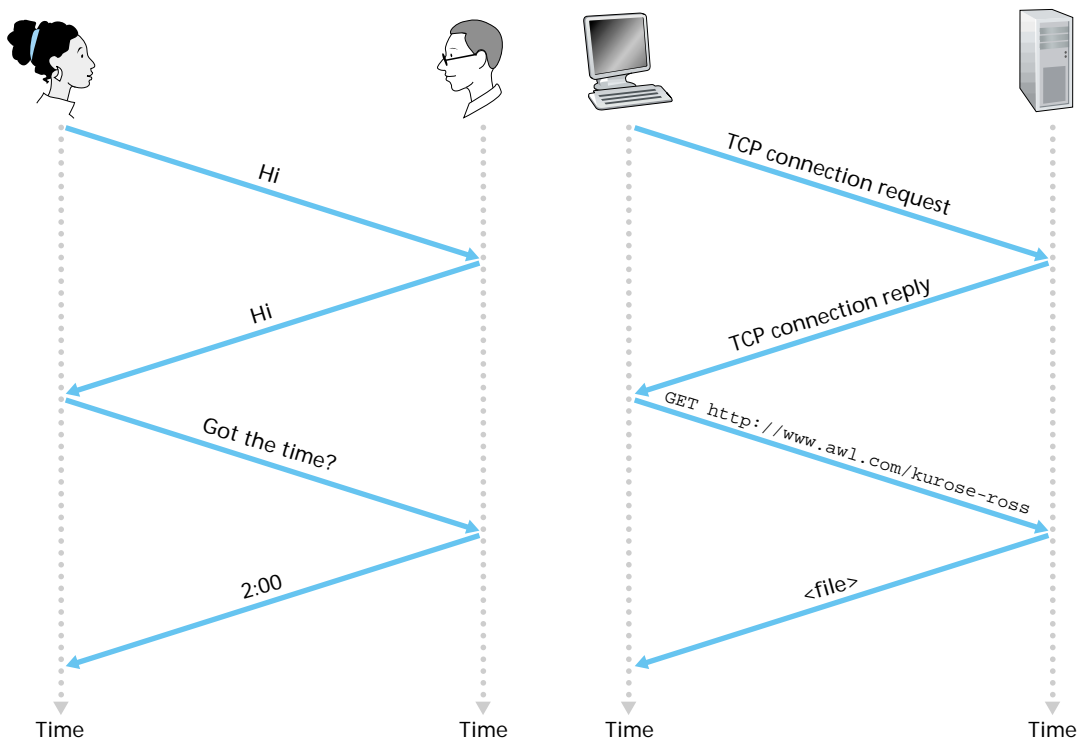


Figure 1.2 ♦ A human protocol and a computer network protocol

dictates that one first offer a greeting (the first “Hi” in Figure 1.2) to initiate communication with someone else. The typical response to a “Hi” is a returned “Hi” message. Implicitly, one then takes a cordial “Hi” response as an indication that one can proceed and ask for the time of day. A different response to the initial “Hi” (such as “Don’t bother me!” or “I don’t speak English,” or some unprintable reply) might indicate an unwillingness or inability to communicate. In this case, the human protocol would be not to ask for the time of day. Sometimes one gets no response at all to a question, in which case one typically gives up asking that person for the time. Note that in our human protocol, *there are specific messages we send, and specific actions we take in response to the received reply messages or other events (such as no reply within some given amount of time)*. Clearly, transmitted and received messages, and actions taken when these messages are sent or received or other events occur, play a central role in a human protocol. If people run different protocols (for example, if one person has manners but the other does not, or if one understands the concept of time and the other does not) the protocols do not interoperate and no useful work can be accomplished. The same is true in networking—it takes two (or more) communicating entities running the same protocol in order to accomplish a task.

Let’s consider a second human analogy. Suppose you’re in a college class (a computer networking class, for example!). The teacher is droning on about protocols and you’re confused. The teacher stops to ask, “Are there any questions?” (a message that is transmitted to, and received by, all students who are not sleeping). You raise your hand (transmitting an implicit message to the teacher). Your teacher acknowledges you with a smile, saying “Yes . . .” (a transmitted message encouraging you to ask your question—teachers *love* to be asked questions), and you then ask your question (that is, transmit your message to your teacher). Your teacher hears your question (receives your question message) and answers (transmits a reply to you). Once again, we see that the transmission and receipt of messages, and a set of conventional actions taken when these messages are sent and received, are at the heart of this question-and-answer protocol.

Network Protocols

A network protocol is similar to a human protocol, except that the entities exchanging messages and taking actions are hardware or software components of some device (for example, computer, router, or other network-capable device). All activity in the Internet that involves two or more communicating remote entities is governed by a protocol. For example, protocols in routers determine a packet’s path from source to destination; hardware-implemented protocols in the network interface cards of two physically connected computers control the flow of bits on the “wire” between the two network interface cards; congestion-control protocols in end systems control the rate at which packets are transmitted between sender and receiver. Protocols are running everywhere in the Internet, and consequently much of this book is about computer network protocols.

As an example of a computer network protocol with which you are probably familiar, consider what happens when you make a request to a Web server, that is, when you type in the URL of a Web page into your Web browser. The scenario is illustrated in the right half of Figure 1.2. First, your computer will send a “connection request” message to the Web server and wait for a reply. The Web server will eventually receive your connection request message and return a “connection reply” message. Knowing that it is now OK to request the Web document, your computer then sends the name of the Web page it wants to fetch from that Web server in a “GET” message. Finally, the Web server returns the Web page (file) to your computer.

Given the human and networking examples above, the exchange of messages and the actions taken when these messages are sent and received are the key defining elements of a protocol:

*A **protocol** defines the format and the order of messages exchanged between two or more communicating entities, as well as the actions taken on the transmission and/or receipt of a message or other event.*

The Internet, and computer networks in general, make extensive use of protocols. Different protocols are used to accomplish different communication tasks. As you read through this book, you will learn that some protocols are simple and straightforward, while others are complex and intellectually deep. Mastering the field of computer networking is equivalent to understanding the what, why, and how of networking protocols.

1.1.4 Some Good Hyperlinks

As every Internet researcher knows, some of the best and most accurate information about the Internet and its protocols is not in hard-copy books, journals, or magazines. Some of the best stuff about the Internet is in the Internet itself! Of course, there’s really too much material to sift through, and sometimes the gems are few and far between. Below, we list a few generally excellent Web sites for network- and Internet-related material. Throughout the book, we will also present links to relevant, high-quality URLs that provide background, original, or advanced material related to the particular topic under study. Here is a set of key links that you may want to consult as you proceed through this book:

- ◆ Internet Engineering Task Force (IETF), <http://www.ietf.org>: The IETF is an open international community concerned with the development and operation of the Internet and its architecture. The IETF was formally established by the Internet Architecture Board (IAB), <http://www.iab.org>, in 1986. The IETF meets three times a year; much of its ongoing work is conducted via mailing lists by working groups. The IETF is administered by the Internet Society, <http://www.isoc.org>, whose Web site contains lots of high-quality, Internet-related material.

- ◆ The World Wide Web Consortium (W3C), <http://www.w3.org>: The W3C was founded in 1994 to develop common protocols for the evolution of the World Wide Web. This is an outstanding site with fascinating information on emerging Web technologies, protocols, and standards.
- ◆ The Association for Computing Machinery (ACM), <http://www.acm.org>, and the Institute of Electrical and Electronics Engineers (IEEE), <http://www.ieee.org>: These are the two main international professional societies that have technical conferences, magazines, and journals in the networking area. The ACM Special Interest Group in Data Communications (SIGCOMM), <http://www.acm.org/sigcomm>, the IEEE Communications Society, <http://www.comsoc.org>, and the IEEE Computer Society, <http://www.computer.org>, are the groups within these bodies whose efforts are most closely related to networking.
- ◆ Computer Networking: A Top-Down Approach Featuring the Internet (that is, the Web site for this textbook!), <http://www.awl.com/kurose-ross>: You'll find a wealth of resources at the Web site, including hyperlinks to relevant Web pages, Java applets illustrating networking concepts, homework problems with answers, programming projects, streaming audio lectures coupled to slides, and much more.

1.2 ◆ The Network Edge

In the previous sections we presented a high-level overview of the Internet and networking protocols. We are now going to delve a bit more deeply into the components of a computer network (and the Internet, in particular). We begin in this section at the edge of a network and look at the components with which we are most familiar—namely, the computers that we use on a daily basis. In the next section we will move from the network edge to the network core and examine switching and routing in computer networks. Then in Section 1.4 we will discuss the actual physical links that carry the signals sent between computers and switches.

1.2.1 End Systems, Clients, and Servers

In computer networking jargon, the computers connected to the Internet are often referred to as **end systems**. They are referred to as end systems because they sit at the edge of the Internet, as shown in Figure 1.3. The Internet's end systems include many different types of computers. End users directly interface with some of these computers, including desktop computers (desktop PCs, Macs, and UNIX-based workstations) and mobile computers (portable computers and PDAs with wireless Internet connections). The Internet's end systems also include computers with which users do not directly interface, such as Web servers and e-mail servers. Furthermore, an increasing number of alternative devices, such as thin clients and household

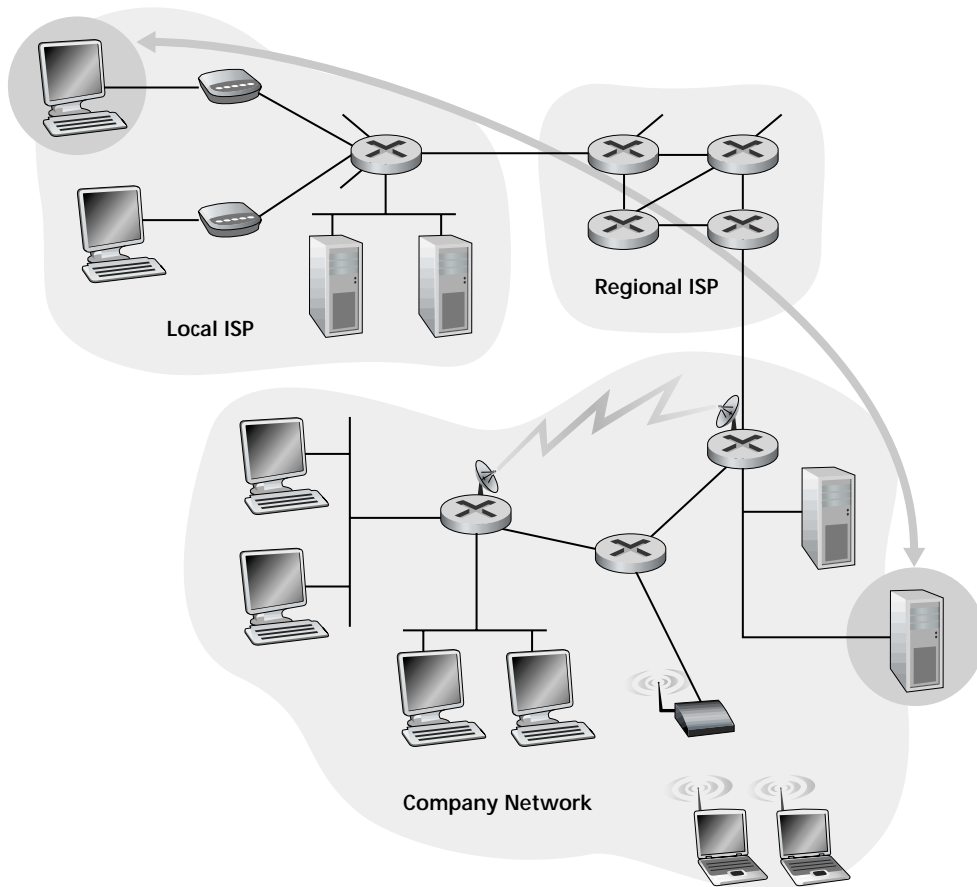


Figure 1.3 ♦ End-system interaction

appliances [Thinplanet 2002], Web TVs and set top boxes [Nesbitt 2002], and digital cameras are being attached to the Internet as end systems. For interesting discussions of the future of Internet appliances see [Manelli 2001; Appliance 2001; Dertouzos 2001; Odlyzko 1999].

End systems are also referred to as **hosts** because they host (that is, run) application programs such as a Web browser program, a Web server program, an e-mail reader program, or an e-mail server program. Throughout this book we will use the terms hosts and end systems interchangeably; that is, *host = end system*. Hosts are sometimes further divided into two categories: **clients** and **servers**. Informally, clients tend to be desktop and mobile PCs, PDAs, and so on, whereas servers tend to be more powerful machines hosting servers such as Web servers and mail servers.