PART I

Introduction To Networking And Internet Applications

An overview of networking and the interface that application programs use to communicate across the Internet

Chapters

Chapter Contents

1

Introduction And Overview

1.1 Growth Of Computer Networking

Computer networking has grown explosively. Since the 1970s, computer communication has changed from an esoteric research topic to an essential part of the infrastructure. Networking is used in every aspect of business, including advertising, production, shipping, planning, billing, and accounting. Consequently, most corporations have multiple networks. Schools, at all grade levels from elementary through post-graduate, are using computer networks to provide students and teachers with instantaneous access to online information. Federal, state, and local government offices use networks, as do military organizations. In short, computer networks are everywhere.

The growth and uses of the global Internet† are among the most interesting and exciting phenomena in networking. In 1980, the Internet was a research project that involved a few dozen sites. Today, the Internet has grown into a production communication system that reaches all populated countries of the world. Many users have high-speed Internet access through cable modems, DSL, or wireless technologies.

The advent and utility of networking has created dramatic economic shifts. Data networking has made telecommuting available to individuals and has changed business communication. In addition, an entire industry emerged that develops networking technologies, products, and services. The importance of computer networking has produced a demand in all industries for people with more networking expertise. Companies need workers to plan, acquire, install, operate, and manage the hardware and software systems that constitute computer networks and internets. In addition, computer programming is no longer restricted to individual computers — network programming is re-

†Throughout this text, we follow the convention of writing *Internet* with an uppercase "I" to denote the global Internet.

quired because all programmers are expected to design and implement application software that can communicate with applications on other computers.

1.2 Why Networking Seems Complex

Because computer networking is an active, exciting field, the subject seems complex. Many technologies exist, and each technology has features that distinguish it from the others. Companies continue to create commercial networking products and services, often by using technologies in new unconventional ways. Finally, networking seems complex because technologies can be combined and interconnected in many ways.

Computer networking can be especially confusing to a beginner because no single underlying theory exists that explains the relationship among all parts. Multiple organizations have created networking standards, but some standards are incompatible with others. Various organizations and research groups have attempted to define conceptual models that capture the essence and explain the nuances among network hardware and software systems, but because the set of technologies is diverse and changes rapidly, models are either so simplistic that they do not distinguish among details or so complex that they do not help simplify the subject.

The lack of consistency in the field has produced another challenge for beginners: instead of a uniform terminology for networking concepts, multiple groups each attempt to create their own terminology. Researchers cling to scientifically precise terminology. Corporate marketing groups often associate a product with a generic technical term or invent new terms merely to distinguish their products or services from those of competitors. Thus, technical terms are easily confused with the names of popular products. To add further confusion, professionals sometimes use a technical term from one technology when referring to an analogous feature of another technology. Consequently, in addition to a large set of terms and acronyms that contains many synonyms, networking jargon contains terms that are often abbreviated, misused, or associated with products.

1.3 The Five Key Aspects Of Networking

To master the complexity in networking, it is important to gain a broad background that includes five key aspects of the subject:

- Network Applications And Network Programming
- Data Communications
- Packet Switching And Networking Technologies
- Internetworking With TCP/IP
- Additional Networking Concepts And Technologies

1.3.1 Network Applications And Network Programming

The network services and facilities that users invoke are each provided by application software — an application program on one computer communicates across a network with an application program running on another computer. Network application services span a wide range that includes email, file transfer, web browsing, voice telephone calls, distributed databases, and audio and video teleconferencing. Although each application offers a specific service with its own form of user interface, all applications can communicate over a single, shared network. The availability of a unified underlying network that supports all applications makes a programmer's job much easier because a programmer only needs to learn about one interface to the network and one basic set of functions — the same set of functions are used in all application programs that communicate over a network.

As we will see, it is possible to understand network applications, and even possible to write code that communicates over a network, without understanding the hardware and software technologies that are used to transfer data from one application to another. It may seem that once a programmer masters the interface, no further knowledge of networking is needed. However, network programming is analogous to conventional programming. Although a conventional programmer can create applications without understanding compilers, operating systems, or computer architecture, knowledge of the underlying systems can help a programmer create more reliable, correct, and efficient programs. Similarly, knowledge of the underlying network system allows a programmer to write better code. The point can be summarized:

> *A programmer who understands the underlying network mechanisms and technologies can write network applications that are more reliable, correct, and efficient.*

1.3.2 Data Communications

The term *data communications* refers to the study of low-level mechanisms and technologies used to send information across a physical communication medium, such as a wire, radio wave, or light beam. Data communications is primarily the domain of Electrical Engineering, which studies how to design and construct a wide range of communication systems. Data communications focuses on ways to use physical phenomena to transfer information. Thus, many of the basic ideas are derived from the properties of matter and energy that have been studied by physicists. For example, we will see that the optical fibers used for high-speed data transfer rely on the properties of light and its reflection at a boundary between two types of matter.

Because it deals with physical concepts, data communications may seem somewhat irrelevant to our understanding of networking. In particular, because many of the terms and concepts refer to physical phenomena, the subject may only seem useful for engineers who design low-level transmission facilities. For example, modulation techniques that use physical forms of energy, such as electromagnetic radiation, to carry information appear to be irrelevant to the design and use of protocols. However, we will see that several key concepts that arise from data communications influence the design of many protocol layers. In the case of modulation, the concept of bandwidth relates directly to network throughput.

As a specific case, data communications introduces the notion of multiplexing that allows information from multiple sources to be combined for transmission across a shared medium and later separated for delivery to multiple destinations. We will see that multiplexing is not restricted to physical transmission — most protocols incorporate some form of multiplexing. Similarly, the concept of encryption introduced in data communications forms the basis of most network security. Thus, we can summarize the importance:

> *Although it deals with many low-level details, data communications provides a foundation of concepts on which the rest of networking is built.*

1.3.3 Packet Switching And Networking Technologies

In the 1960s, a new concept revolutionized data communications: packet switching. Early communication networks had evolved from telegraph and telephone systems that connected a physical pair of wires between two parties to form a communication circuit. Although mechanical connection of wires was being replaced by electronic switches, the underlying paradigm remained the same: form a circuit and then send information across the circuit. Packet switching changed networking in a fundamental way, and provided the basis for the modern Internet: instead of forming a dedicated circuit, packet switching allows multiple senders to transmit data over a shared network. Packet switching builds on the same fundamental data communications mechanisms as the phone system, but uses the underlying mechanisms in a new way. Packet switching divides data into small blocks, called packets, and includes an identification of the intended recipient in each packet. Devices throughout the network each have information about how to reach each possible destination. When a packet arrives at one of the devices, the device chooses a path over which to send the packet so the packet eventually reaches the correct destination.

In theory, packet switching is straightforward. However, many designs are possible, depending on the answers to basic questions. How should a destination be identified, and how can a sender find the identification of a destination? How large should a packet be? How can a network recognize the end of one packet and the beginning of another packet? If many computers are sending over a network, how can they coordinate to insure that each receives a fair opportunity to send? How can packet switching be adapted to wireless networks? How can networking technologies be designed to meet various requirements for speed, distance, and economic cost? Many answers have been proposed, and many packet switching technologies have been created. In fact, when one studies packet switching networks, a fundamental conclusion can be drawn:

> *Because each network technology is created to meet various requirements for speed, distance, and economic cost, many packet switching technologies exist. Technologies differ in details such as the size of packets and the method used to identify a recipient.*

1.3.4 Internetworking With TCP/IP

In the 1970s, another revolution in computer networking arose: the concept of an Internet. Many researchers who investigated packet switching looked for a single packet switching technology that could handle all needs. In 1973, Vinton Cerf and Robert Kahn observed that no single packet switching technology would ever satisfy all needs, especially because it would be possible to build low-capacity technologies for homes or offices at extremely low cost. The solution, they suggested, was to stop trying to find a single best solution, and instead, explore interconnecting many packet switching technologies into a functioning whole. They proposed that a set of standards be developed for such an interconnection, and the resulting standards became known as the *TCP/IP Internet Protocol Suite* (usually abbreviated *TCP/IP*). The concept, now known as *internetworking*, is extremely powerful. It provides the basis of the global Internet, and forms an important part of the study of computer networking.

One of the primary reasons for the success of TCP/IP standards lies in their tolerance of heterogeneity. Instead of attempting to dictate details about packet switching technologies, such as packet sizes or the method used to identify a destination, TCP/IP takes a virtualization approach that defines a network-independent packet and a network-independent identification scheme, and then specifies how the virtual packets are mapped onto each possible underlying network.

Interestingly, TCP/IP's ability to tolerate new packet switching networks is a major motivation for the continual evolution of packet switching technologies. As the Internet grows, computers become more powerful and applications send more data, especially graphic images and video. To accommodate increases in use, engineers invent new technologies that can transmit more data and process more packets in a given time. As they are invented, new technologies are incorporated into the Internet along with extant technologies. That is, because the Internet tolerates heterogeneity, engineers can experiment with new networking technologies without disrupting the existing networks. To summarize:

> *The Internet is formed by interconnecting multiple packet switching networks. Internetworking is substantially more powerful than a single networking technology because the approach permits new technologies to be incorporated at any time without requiring the replacement of old technologies.*

1.4 Public And Private Parts Of The Internet

Although it functions as a single communication system, the Internet consists of parts that are owned and operated by individuals or organizations. To help clarify ownership and purpose, the networking industry uses the terms *public network* and *private network*.

1.4.1 Public Network

A *public network* is run as a service that is available to subscribers. Any individual or corporation who pays the subscription fee can use the network. A company that offers communication service is known as a *service provider*. The concept of a service provider is quite broad, and extends beyond *Internet Service Providers* (*ISPs*). In fact, the terminology originated with companies that offered analog voice telephone service. To summarize:

> *A public network is owned by a service provider, and offers service to any individual or organization that pays the subscription fee.*

It is important to understand that the term *public* refers to the general availability of service, not to the data transferred. In particular, many public networks follow strict government regulations that require the provider to protect communication from unintended snooping. The point is:

> *The term* public *means a service is available to the general public; data transferred across a public network is not revealed to outsiders.*

1.4.2 Private Network

A *private network* is controlled by one particular group. Although it may seem straightforward, the distinction between public and private parts of the Internet can be subtle because control does not always imply ownership. For example, if a company leases a data circuit from a provider and then restricts use of the circuit to company traffic, the circuit becomes part of the company's private network. The point is:

A network is said to be private *if use of the network is restricted to one group. A private network can include circuits leased from a provider.*

Networking equipment vendors divide private networks into four categories:

- Consumer
- Small Office / Home Office (SOHO)
- Small-To-Medium Business (SMB)
- Large Enterprise

Because the categories relate to sales and marketing, the terminology is loosely defined. Although it is possible to give a qualitative description of each type, one cannot find an exact definition. Thus, the paragraphs below provide a broad characterization of size and purpose rather than detailed measures.

Consumer. One of the least expensive forms of private network consists of a LAN owned by an individual — if an individual purchases an inexpensive LAN switch and uses the switch to attach a printer to a PC, the individual has created a private network. Similarly, a wireless router constitutes a private network that a consumer might purchase and install.

Small Office / Home Office (SOHO). A SOHO network is slightly larger than a consumer network. A typical SOHO network connects two or more computers, one or more printers, a router that connects to the Internet, and possibly other devices, such as a cash register. Most SOHO installations include a battery-backup power supply and other mechanisms that allow them to operate without interruption.

Small-To-Medium Business (*SMB*). An SMB network can connect many computers in multiple offices in a building, and can also include computers in a production facility (e.g., in a shipping department). Often an SMB network contains multiple Layer-2 switches interconnected by routers, uses a broadband Internet connection, and may include wireless access points.

Large Enterprise. A large enterprise network provides the IT infrastructure needed for a major corporation. A typical large enterprise network connects several geographic sites with multiple buildings at each site, uses many Layer-2 switches and routers, and has two or more high-speed Internet connections. Enterprise networks usually include both wired and wireless technologies.

To summarize:

A private network can serve an individual consumer, a small office, a small-to-medium business, or a large enterprise.

1.5 Networks, Interoperability, And Standards

Communication always involves at least two entities, one that sends information and another that receives it. In fact, we will see that most packet switching communication systems contain intermediate entities (i.e., devices that forward packets). The important point to note is that for communication to be successful, all entities in a network must agree on how information will be represented and communicated. Communication agreements involve many details. For example, when two entities communicate over a wired network, both sides must agree on the voltages to be used, the exact way that electrical signals are used to represent data, procedures used to initiate and conduct communication, and the format of messages.

We use the term *interoperability* to refer to the ability of two entities to communicate, and say that if two entities can communicate without any misunderstandings, they *interoperate* correctly. To insure that all communicating parties agree on details and follow the same set of rules, an exact set of specifications is written down. To summarize:

> *Communication involves multiple entities that must agree on details ranging from the electrical voltage used to the format and meaning of messages. To insure that entities can interoperate correctly, rules for all aspects of communication are written down.*

Following diplomatic terminology, we use the term *communication protocol*, *network protocol*, or *protocol* to refer to a specification for network communication. A given protocol specifies low-level details, such as the type of radio transmission used in a wireless network, or describes a high-level mechanism such as the messages that two application programs exchange. We said that a protocol can define a procedure to be followed during an exchange. One of the most important aspects of a protocol concerns situations in which an error or unexpected condition occurs. Thus, a protocol usually explains the appropriate action to take for each possible abnormal condition (e.g., a response is expected, but no response arrives). To summarize:

> *A communication protocol specifies the details for one aspect of computer communication, including actions to be taken when errors or unexpected situations arise. A given protocol can specify low-level details, such as the voltage and signals to be used, or high-level items, such as the format of messages that application programs exchange.*

1.6 Protocol Suites And Layering Models

A set of protocols must be constructed carefully to ensure that the resulting communication system is both complete and efficient. To avoid duplication of effort, for example, each protocol should handle a part of communication not handled by other protocols. How can one guarantee that protocols will work well together? The answer lies in an overall design plan: instead of creating each protocol in isolation, protocols are designed in complete, cooperative sets called *suites* or *families*. Each protocol in a suite handles one aspect of communication; together, the protocols in a suite cover all aspects of communication, including hardware failures and other exceptional conditions. Furthermore, the entire suite is designed to allow the protocols to work together efficiently.

The fundamental abstraction used to collect protocols into a unified whole is known as a *layering model*. In essence, a layering model describes how all aspects of a communication problem can be partitioned into pieces that work together. Each piece is known as a *layer*; the terminology arises because protocols in a suite are organized into a linear sequence. Dividing protocols into layers helps both protocol designers and implementors manage the complexity by allowing them to concentrate on one aspect of communication at a given time.

Figure 1.1 illustrates the concept by showing the layering model used with the Internet protocols. The visual appearance of figures used to illustrate layering has led to the colloquial term *stack*. The term is used to refer to the protocol software on a computer, as in “does that computer run the TCP/IP stack?”

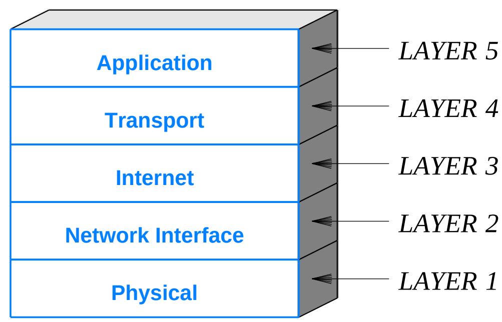

Figure 1.1 The layering model used with the Internet protocols (TCP/IP).

Later chapters will help us understand layering by explaining protocols in detail. For now, it is sufficient to learn the purpose of each layer and how protocols are used for communication. The next sections summarize the role of the layers; a later section examines how data passes through layers when computers communicate.

Layer 1: Physical

Protocols in the *Physical* layer specify details about the underlying transmission medium and the associated hardware. All specifications related to electrical properties, radio frequencies, and signals belong in layer 1.

Layer 2: Network Interface†

Protocols in the *Network Interface* layer specify details about communication between higher layers of protocols, which are usually implemented in software, and the underlying network, which is implemented in hardware. Specifications about network addresses and the maximum packet size that a network can support, protocols used to access the underlying medium, and hardware addressing belong in layer 2.

Layer 3: Internet

Protocols in the *Internet* layer form the fundamental basis for the Internet. Layer 3 protocols specify communication between two computers across the Internet (i.e., across multiple interconnected networks). The Internet addressing structure, the format of Internet packets, the method for dividing a large Internet packet into smaller packets for transmission, and mechanisms for reporting errors belong in layer 3.

Layer 4: Transport

Protocols in the *Transport* layer provide for communication from an application program on one computer to an application program on another. Specifications that control the maximum rate a receiver can accept data, mechanisms to avoid network congestion, and techniques to insure that all data is received in the correct order belong in layer 4.

Layer 5: Application

Protocols in the top layer of the TCP/IP stack specify how a pair of applications interact when they communicate. Layer 5 protocols specify details about the format and meaning of messages that applications can exchange as well as procedures to be followed during communication. Specifications for email exchange, file transfer, web browsing, telephone services, and video teleconferencing all belong in layer 5.

†Some publications use the term *Data Link* in place of *Network Interface*. In a later section, we will see that ambiguity can arise because another layering model uses *Data Link* for layer 2.

1.7 How Data Passes Through Layers

Layering is not merely an abstract concept that helps one understand protocols. Instead, protocol implementations follow the layering model by passing the output from a protocol in one layer to the input of a protocol in the next layer. Furthermore, to achieve efficiency, rather than copy an entire packet, a pair of protocols in adjacent layers pass a pointer to the packet. Thus, data passes between layers efficiently.

To understand how protocols operate, consider two computers connected to a network. Figure 1.2 illustrates layered protocols on the two computers. As the figure shows, each computer contains a set of layered protocols. When an application sends data, the data is placed in a packet, and the outgoing packet passes down through each layer of protocols. Once it has passed through all layers of protocols on the sending computer, the packet leaves the computer and is transmitted across the underlying physical network†. When it reaches the receiving computer, the packet passes up through the layers of protocols. If the application on the receiving computer sends a response, the process is reversed. That is, a response passes down through the layers on its way out, and up through the layers on the computer that receives the response.

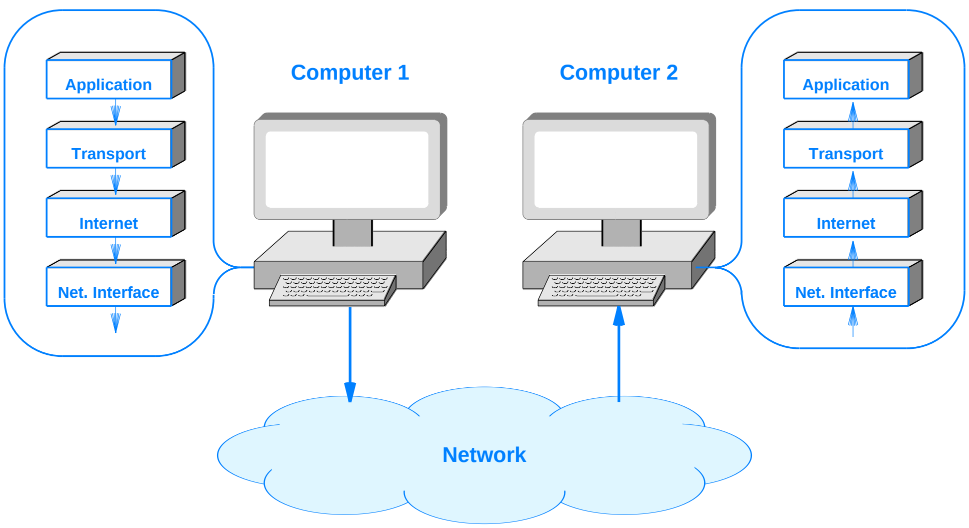

Figure 1.2 Illustration of how data passes among protocol layers when computers communicate across a network. Each computer has a set of layered protocols, and data passes through each layer.

†The figure shows only one network. When we study Internet architecture, we will learn about intermediate devices called *routers* and see how layered protocols operate in an Internet.

1.8 Headers And Layers

We will learn that each layer of protocol software performs computations that insure the messages arrive as expected. To perform such computation, protocol software on the two machines must exchange information. To do so, each layer on the sending computer prepends extra information onto the packet; the corresponding protocol layer on the receiving computer removes and uses the extra information.

Additional information added by a protocol is known as a *header*. To understand how headers appear, think of a packet traveling across the network between the two computers in Figure 1.2. Headers are added by protocol software as the data passes down through the layers on the sending computer. That is, the Transport layer prepends a header, and then the Internet layer prepends a header, and so on. Thus, if we observe a packet traversing the network, the headers will appear in the order that Figure 1.3 illustrates.

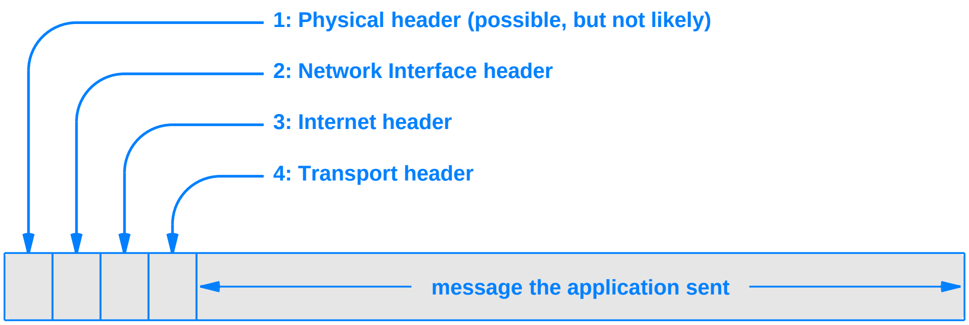

Figure 1.3 The nested protocol headers that appear on a packet as the packet travels across a network between two computers. In the diagram, the beginning of the packet (the first bit sent over the underlying network) is shown on the left.

Although the figure shows headers as the same size, in practice headers are not of uniform size, and a physical layer header is optional. We will understand the reason for the size disparities when we examine header contents. Similarly, we will see that the physical layer usually specifies how signals are used to transmit data. Thus, one does not expect to find a Physical layer header.

1.9 ISO and the OSI Seven Layer Reference Model

At the same time the Internet protocols were being developed, two large standards bodies jointly formed an alternative reference model. They also created a set of internetworking protocols. The organizations are:

- International Organization for Standardization (ISO)
- International Telecommunications Union, Telecommunication Standardization Sector (ITU-T)†,

The ISO layering model is known as the *Open Systems Interconnection Seven-Layer Reference Model*. Confusion arises in terminology because the acronym for the protocols, OSI, and the acronym for the organization, ISO, are similar. One is likely to find references to both the *OSI seven-layer model* and to the *ISO seven-layer model*. Figure 1.4 illustrates the seven layers in the model.

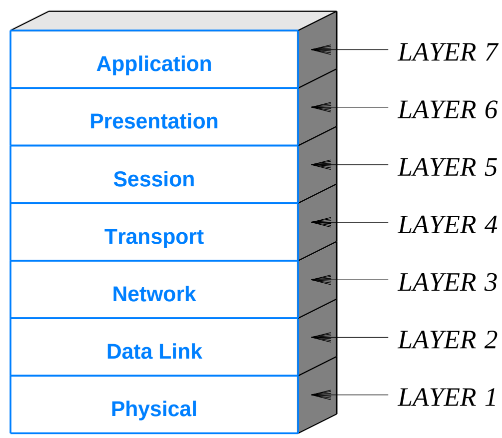

Figure 1.4 The OSI seven-layer model standardized by ISO.

1.10 The Inside Scoop

Like most standards organizations, ISO and the ITU use a process that accommodates as many viewpoints as possible when creating a standard. As a result, some standards can appear to have been designed by a committee making political compromises rather than by engineers and scientists. The seven-layer reference model is controversial. It did indeed start as a political compromise. Furthermore, the model and the OSI protocols were designed as competitors for the Internet protocols.

ISO and the ITU are huge standards bodies that handle the world-wide telephone system and other global standards. The Internet protocols and reference model were

†When the standard was first created, the ITU was known as the *Consultative Committee for International Telephone and Telegraph* (*CCITT*).

created by a small group of about a dozen researchers. It is easy to see why the standards organizations might be confident that they could dictate a set of protocols and everyone would switch away from protocols designed by researchers. At one point, even the U.S. government was convinced that TCP/IP should be replaced by OSI protocols.

Eventually, it became clear that TCP/IP technology was technically superior to OSI, and in a matter of a few years, efforts to develop and deploy OSI protocols were terminated. Standards bodies were left with the seven-layer model, which did not include an Internet layer. Consequently, for many years, advocates for the seven-layer model have tried to stretch the definitions to match TCP/IP. They argue that layer three could be considered an Internet layer and that a few support protocols might be placed into layers five and six. Perhaps the most humorous part of the story is that many engineers still refer to applications as *layer 7 protocols*, even when they know that layers five and six are unfilled and unnecessary.

1.11 Remainder Of The Text

The text is divided into five major parts. After a brief introduction, chapters in the first part introduce network applications and network programming. Readers who have access to a computer are encouraged to build and use application programs that use the Internet while they read the text. The remaining four parts explain how the underlying technologies work. The second part describes data communications and the transmission of information. It explains how electrical and electromagnetic energy can be used to carry information across wires or through the air, and shows how data is transmitted.

The third part of the text focuses on packet switching and packet technologies. It explains why computer networks use packets, describes the general format of packets, examines how packets are encoded for transmission, and shows how each packet is forwarded across a network to its destination. The third part of text also introduces basic categories of computer networks, such as Local Area Networks (LANs) and Wide Area Networks (WANs). It characterizes the properties of each category and discusses example technologies.

The fourth part of the text covers internetworking and the associated TCP/IP Internet Protocol Suite. The text explains the structure of the Internet and the TCP/IP protocols. It explains the IP addressing scheme, and describes the mapping between Internet addresses and underlying hardware addresses. It also discusses Internet routing and routing protocols. The fourth part includes a description of several fundamental concepts, including: encapsulation, fragmentation, congestion and flow control, virtual connections, address translation, bootstrapping, IPv6, and various support protocols.

The fifth part of the text covers a variety of remaining topics that pertain to the network as a whole instead of individual parts. After a chapter on network performance, chapters cover emerging technologies, network security, and network management.

1.12 Summary

The large set of technologies, products, and interconnection schemes make networking a complex subject. There are five key aspects: network applications and network programming, data communications, packet switching and networking technologies, internetworking with TCP/IP, and topics that apply across layers, such as security and network management.

Because multiple entities are involved in communication, they must agree on details, including electrical characteristics such as voltage as well as the format and meaning of all messages. To insure interoperability, each entity is constructed to obey a set of communication protocols that specify all details needed for communication. To insure that protocols work together and handle all aspects of communication, an entire set of protocols is designed at the same time. The central abstraction around which protocols are built is called a *layering model*. Layering helps reduce complexity by allowing an engineer to focus on one aspect of communication at a given time without worrying about other aspects. The TCP/IP protocols used in the Internet follow a five-layer reference model; the phone companies and International Standards Organization proposed a seven-layer reference model.

EXERCISES

1.1 Provide reasons for Internet growth in recent years.

1.2 List ten industries that depend on computer networking.

1.3 According to the text, is it possible to develop Internet applications without understanding the architecture of the Internet and the technologies? Support your answer.

1.4 To what aspects of networking does *data communications* refer?

1.5 What is packet-switching, and why is packet switching relevant to the Internet?

1.6 Provide a brief history of the Internet describing when and how it was started.

1.7 What is interoperability, and why is it especially important in the Internet?

1.8 What is a communication protocol? Conceptually, what two aspects of communication does a protocol specify?

1.9 What is a protocol suite, and what is the advantage of a suite?

1.10 Describe the TCP/IP layering model, and explain how it was derived.

1.11 List the layers in the TCP/IP model, and give a brief explanation of each.

1.12 Explain how headers are added and removed as data passes through a layered model.

1.13 List major standardization organizations that create standards for data communications and computer networking.

1.14 Give a brief explain of the layers in the ISO Open System Interconnection model.