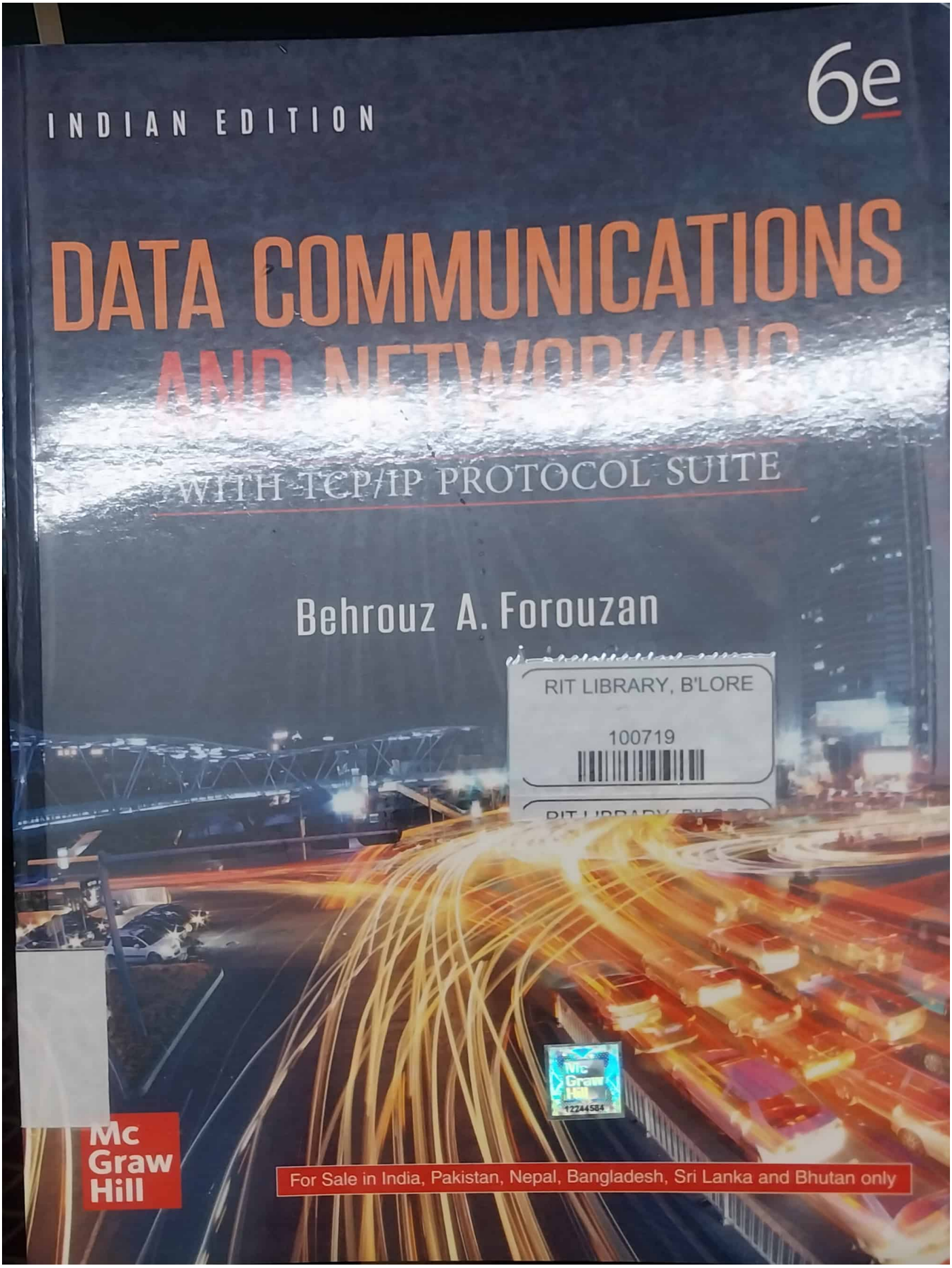

INDIAN EDITION
6e
DATA COMMUNICATIONS
AND NETWORKING
WITH TCP/IP PROTOCOL SUITE
Behrouz A. Forouzan
RIT LIBRARY, B'LORE
100719
Mc Graw Hill
For Sale in India, Pakistan, Nepal, Bangladesh, Sri Lanka and Bhutan only

CONTENTS

Preface xix

Trademark xxv

Chapter 1 *Introduction* 1

1.1 DATA COMMUNICATIONS 2
- 1.1.1 Components 2
- 1.1.2 Message 3
- 1.1.3 Data Flow 4

1.2 NETWORKS 5
- 1.2.1 Network Criteria 5
- 1.2.2 Physical Structures 5

1.3 NETWORK TYPES 8
- 1.3.1 Local Area Network 8
- 1.3.2 Wide Area Network (WAN) 8
- 1.3.3 The Internet 10
- 1.3.4 Accessing the Internet 12

1.4 PROTOCOL LAYERING 13
- 1.4.1 Scenarios 13
- 1.4.2 Principles of Protocol Layering 16
- 1.4.3 Logical Connections 16

1.5 TCP/IP PROTOCOL SUITE 17
- 1.5.1 Layered Architecture 17
- 1.5.2 Brief Description of Layers 18
- 1.5.3 Description of Each Layer 20

1.6 THE OSI MODEL 21
- 1.6.1 OSI versus TCP/IP 21
- 1.6.2 Lack of OSI Model's Success 22

1.7 END-OF-CHAPTER MATERIALS 23
- 1.7.1 Recommended Reading 23
- 1.7.2 Key Terms 23
- 1.7.3 Summary 23

1.8 PRACTICE SET 24
- 1.8.1 Quizzes 24
- 1.8.2 Questions 24
- 1.8.3 Problems 26

Chapter 2 *Physical Layer 29*

2.1 SIGNALS 31
2.1.1 Analog Signals 31
2.1.2 Digital Signals 33
2.2 SIGNAL IMPAIRMENT 35
2.2.1 Attenuation and Amplification 35
2.2.2 Distortion 35
2.2.3 Data Rate Limits 36
2.2.4 Performance 38
2.3 DIGITAL TRANSMISSION 40
2.3.1 Digital-to-Digital Conversion 40
2.3.2 Analog-to-Digital Conversion 41
2.4 ANALOG TRANSMISSION 42
2.4.1 Digital-to-Analog Conversion 42
2.4.2 Analog-to-Analog Conversion 45
2.5 MULTIPLEXING 47
2.5.1 Frequency-Division Multiplexing 48
2.5.2 Time-Division Multiplexing 48
2.6 TRANSMISSION MEDIA 49
2.6.1 Guided Media 50
2.6.2 Unguided Media: Wireless 53
2.7 END-OF-CHAPTER MATERIALS 55
2.7.1 Recommended Reading 55
2.7.2 Key Terms 55
2.7.3 Summary 55
2.8 PRACTICE SET 56
2.8.1 Quizzes 56
2.8.2 Questions 56
2.8.3 Problems 58

Chapter 3 *Data-Link Layer 63*

3.1 INTRODUCTION 64
3.1.1 Nodes and Links 65
3.1.2 Two Types of Links 65
3.1.3 Two Sublayers 66
3.2 DATA-LINK CONTROL 66
3.2.1 Framing 66
3.2.2 Error Control 70
3.2.3 Two DLC Protocols 80
3.3 MEDIA ACCESS PROTOCOLS 88
3.3.1 Random Access 88
3.3.2 Controlled Access 101

3.4 LINK-LAYER ADDRESSING 104
3.4.1 Three Types of Addresses 106
3.4.2 Address Resolution Protocol (ARP) 107
3.5 END-OF-CHAPTER MATERIALS 107
3.5.1 Recommended Reading 107
3.5.2 Key Terms 107
3.5.3 Summary 108
3.6 PRACTICE SET 108
3.6.1 Quizzes 108
3.6.2 Questions 109
3.6.3 Problems 110

Chapter 4 *Local Area Networks: LANs 115*

4.1 ETHERNET 116
4.1.1 Standard Ethernet (10 Mbps) 117
4.1.2 Fast Ethernet (100 Mbps) 121
4.1.3 Gigabit Ethernet (1000 Mbps) 123
4.1.4 10 Gigabit Ethernet 126
4.2 WIFI, IEEE 802.11 PROJECT 126
4.2.1 Architecture 127
4.2.2 MAC Sublayer 128
4.2.3 Addressing Mechanism 133
4.2.4 Physical Layer 135
4.3 BLUETOOTH 138
4.3.1 Architecture 138
4.3.2 Bluetooth Layers 140
4.4 END-OF-CHAPTER MATERIALS 145
4.4.1 Recommended Reading 145
4.4.2 Key Terms 145
4.4.3 Summary 146
4.5 PRACTICE SET 146
4.5.1 Quizzes 146
4.5.2 Questions 146
4.5.3 Problems 147

Chapter 5 *Wide Area Networks: WANs 149*

5.1 TELEPHONE NETWORKS 150
5.1.1 Major Components 150
5.1.2 LATAs 151
5.1.3 Signaling 152
5.1.4 Services Provided by Telephone Networks 155
5.1.5 Dial-Up Service 156
5.1.6 Digital Subscriber Line (DSL) 158

8.2 ROUTING ALGORITHMS 288
8.2.1 Distance-Vector Routing 288
8.2.2 Link-State Routing 294
8.2.3 Path-Vector Routing 297
8.3 UNICAST ROUTING PROTOCOLS 301
8.3.1 Internet Structure 301
8.3.2 Routing Information Protocol (RIP) 303
8.3.3 Open Shortest Path First (OSPF) 308
8.3.4 Border Gateway Protocol Version 4 (BGP4) 313
8.4 MULTICAST ROUTING 322
8.4.1 Unicasting 322
8.4.2 Multicasting 323
8.4.3 Distance Vector Multicast Routing Protocol 324
8.4.4 Multicast Open Shortest Path First 327
8.4.5 Protocol Independent Multicast (PIM) 327
8.5 IGMP 331
8.5.1 Messages 331
8.5.2 Propagation of Membership Information 332
8.5.3 Encapsulation 333
8.6 END-OF-CHAPTER MATERIALS 333
8.6.1 Recommended Reading 333
8.6.2 Key Terms 333
8.6.3 Summary 334
8.7 PRACTICE SET 335
8.7.1 Quizzes 335
8.7.2 Questions 335
8.7.3 Problems 337

Chapter 9 *Transport Layer 341*

9.1 TRANSPORT-LAYER SERVICES 342
9.1.1 Process-to-Process Communication 342
9.1.2 Addressing: Port Numbers 343
9.1.3 Encapsulation and Decapsulation 345
9.1.4 Multiplexing and Demultiplexing 346
9.1.5 Flow Control 346
9.1.6 Error Control 349
9.1.7 Combination of Flow and Error Control 350
9.1.8 Congestion Control 352
9.1.9 Connectionless and Connection-Oriented Protocols 352
9.2 TRANSPORT-LAYER PROTOCOLS 356
9.2.1 Services 356
9.2.2 Port Numbers 357

9.3 USER DATAGRAM PROTOCOL (UDP) 358
9.3.1 UDP Services 359
9.3.2 UDP Applications 362
9.4 TRANSMISSION CONTROL PROTOCOL 363
9.4.1 TCP Services 364
9.4.2 TCP Features 367
9.4.3 Segment 368
9.4.4 A TCP Connection 371
9.4.5 State Transition Diagram 378
9.4.6 Windows in TCP 380
9.4.7 Flow Control 383
9.4.8 Error Control 389
9.4.9 TCP Congestion Control 398
9.4.10 TCP Timers 408
9.4.11 Options 412
9.5 SCTP 412
9.5.1 SCTP Services 412
9.5.2 SCTP Features 414
9.5.3 Packet Format 416
9.5.4 An SCTP Association 418
9.5.5 Flow Control 421
9.5.6 Error Control 423
9.6 END-OF-CHAPTER MATERIALS 427
9.6.1 Recommended Reading 427
9.6.2 Key Terms 427
9.6.3 Summary 428
9.7 PRACTICE SET 429
9.7.1 Quizzes 429
9.7.2 Questions 429
9.7.3 Problems 432

Chapter 10 *Application Layer 437*

10.1 INTRODUCTION 438
10.1.1 Providing Services 439
10.1.2 Application-Layer Paradigms 440
10.2 CLIENT/SERVER PARADIGM 443
10.2.1 Application Programming Interface 443
10.2.2 Using Services of the Transport Layer 447
10.3 STANDARD APPLICATIONS 448
10.3.1 World Wide Web and HTTP 449
10.3.2 FTP 464
10.3.3 Electronic Mail 468
10.3.4 TELNET 481

10.3.5 Secure Shell (SSH) 484
10.3.6 Domain Name System (DNS) 486
10.4 PEER-TO-PEER PARADIGM 498
10.4.1 P2P Networks 498
10.4.2 Distributed Hash Table (DHT) 500
10.4.3 Chord 503
10.4.4 Pastry 510
10.4.5 Kademlia 515
10.4.6 A Popular P2P Network: BitTorrent 518
10.5 SOCKET INTERFACE PROGRAMMING 521
10.5.1 Data Structure for Socket 521
10.5.2 Header Files 522
10.5.3 Iterative Communication Using UDP 522
10.5.4 Communication Using TCP 528
10.6 END-OF-CHAPTER MATERIALS 535
10.6.1 Recommended Reading 535
10.6.2 Key Terms 536
10.6.3 Summary 536
10.7 PRACTICE SET 537
10.7.1 Quizzes 537
10.7.2 Questions 537
10.7.3 Problems 539

Chapter 11 *Multimedia 543*

11.1 COMPRESSION 544
11.1.1 Lossless Compression 544
11.1.2 Lossy Compression 554
11.2 MULTIMEDIA DATA 560
11.2.1 Text 560
11.2.2 Image 560
11.2.3 Video 564
11.2.4 Audio 566
11.3 MULTIMEDIA IN THE INTERNET 568
11.3.1 Streaming Stored Audio/Video 568
11.3.2 Streaming Live Audio/Video 571
11.3.3 Real-Time Interactive Audio/Video 572
11.4 REAL-TIME INTERACTIVE PROTOCOLS 577
11.4.1 Rationale for New Protocols 578
11.4.2 RTP 581
11.4.3 RTCP 583
11.4.4 Session Initialization Protocol (SIP) 587
11.4.5 H.323 594

11.5 END-OF-CHAPTER MATERIALS 597
11.5.1 Recommended Reading 597
11.5.2 Key Terms 597
11.5.3 Summary 597
11.6 PRACTICE SET 598
11.6.1 Quizzes 598
11.6.2 Questions 598
11.6.3 Problems 600

Chapter 12 *Network Management 605*

12.1 INTRODUCTION 606
12.1.1 Configuration Management 606
12.1.2 Fault Management 608
12.1.3 Performance Management 609
12.1.4 Security Management 609
12.1.5 Accounting Management 610
12.2 SNMP 610
12.2.1 Managers and Agents 611
12.2.2 Management Components 611
12.2.3 An Overview 613
12.2.4 SMI 614
12.2.5 MIB 618
12.2.6 SNMP Operation 622
12.3 ASN.1 627
12.3.1 Language Basics 628
12.3.2 Data Types 629
12.3.3 Encoding 632
12.4 END-OF-CHAPTER MATERIALS 632
12.4.1 Recommended Reading 632
12.4.2 Key Terms 632
12.4.3 Summary 632
12.5 PRACTICE SET 633
12.5.1 Quizzes 633
12.5.2 Questions 633
12.5.3 Problems 634

Chapter 13 *Cryptography and Network Security 637*

13.1 INTRODUCTION 638
13.1.1 Security Goals 638
13.1.2 Attacks 639
13.1.3 Services and Techniques 641
13.2 CONFIDENTIALITY 641
13.2.1 Symmetric-Key Ciphers 641
13.2.2 Asymmetric-Key Ciphers 653

13.3 OTHER ASPECTS OF SECURITY 658
13.3.1 Message Integrity 658
13.3.2 Message Authentication 659
13.3.3 Digital Signature 660
13.3.4 Entity Authentication 666
13.3.5 Key Management 668
13.4 NETWORK-LAYER SECURITY 674
13.4.1 Two Modes 675
13.4.2 Two Security Protocols 676
13.4.3 Services Provided by IPSec 680
13.4.4 Security Association 680
13.4.5 Internet Key Exchange (IKE) 684
13.4.6 Virtual Private Network (VPN) 684
13.5 TRANSPORT-LAYER SECURITY 685
13.5.1 SSL Architecture 686
13.5.2 Four Protocols 689
13.6 APPLICATION-LAYER SECURITY 691
13.6.1 E-mail Security 691
13.6.2 Pretty Good Privacy (PGP) 693
13.6.3 S/MIME 698
13.7 FIREWALLS 702
13.7.1 Packet-Filter Firewall 703
13.7.2 Proxy Firewall 704
13.8 END-OF-CHAPTER MATERIALS 705
13.8.1 Recommended Reading 705
13.8.2 Key Terms 705
13.8.3 Summary 706
13.9 PRACTICE SET 707
13.9.1 Quizzes 707
13.9.2 Questions 707
13.9.3 Problems 709

Appendices

Appendix A *Unicode* 713
Appendix B *Positional Numbering System* 719
Appendix C *HTML, CSS, XML, and XSL* 727
Appendix D *A Touch of Probability* 737
Appendix E *Checksum* 743
Appendix F *Acronyms* 751

Glossary 761
References 805
Index 811

PREFACE

Welcome to the sixth edition of *Data Communications and Networking with TCP/IP Protocol Suite*. We are living in an information age, and information is distributed faster that ever using the Internet, which works based on the topics discussed in this book.

Features

Although the main goal of this book is to teach the principles of networking, it is designed to teach these principles using the following features:

TCP/IP Protocol Suite

This book is designed to teach the principles of networking by using the TCP/IP protocol suite. Teaching these principles using protocol layering is beneficial because these principles are repeated and better understood in relation to each layer. For example, *addressing* is an issue that is applied to several layers of the TCP/IP protocol suite. Another example is *framing and packetizing*, which is repeated in several layers, but each layer treats the principle differently.

Bottom-Up Approach

This book uses a bottom-up approach. Each layer in the TCP/IP protocol suite is built on the services provided by the layer below. We learn how bits are moving at the physical layer (first layer) before learning how some programs exchange messages at the application layer (fifth layer).

Organization

The book is made up of 13 chapters, six appendices, a list of references, and a glossary.

Chapter 1: Introduction

This chapter is an introduction to *Data Communications and Networking with TCP/IP Protocol Suite*. It defines the concept of protocol layering and gives a brief description of the TCP/IP protocol suite and the OSI model.

Chapter 2: Physical Layer

This chapter describes the first layer of the TCP/IP protocol suite: the physical layer. It explains the relationship between data and signals and describes both analog and digital signals. It also discusses multiplexing to benefit from the available bandwidth. Finally, it goes below the physical layer and discusses the transmission media.

Chapter 3: Data-Link Layer

This chapter discusses the data-link layer, the second layer in the TCP/IP protocol suite. It shows that the data-link layer is made up of two sublayers: medial link control and data link control. It also discusses link-layer addressing.

Chapter 4: Local Area Networks: LANs

This chapter discusses the local area networks (LANs) that use only the first two layers of the TCP/IP protocol suite. It describes both wired LANs (Ethernet) and wireless LANs (WiFi and Bluetooth).

Chapter 5: Wide Area Networks: WANs

This chapter discusses the wide area networks (WANs) that also use only the first two layers of the TCP/IP protocol suite. It describes several WANs, including the telephone network, cable network, cellular telephony, and satellite networks.

Chapter 6: Connecting Devices and Virtual LANs

This chapter discusses the connecting devices such as hubs, link-layer switches, and routers. It also describes virtual LANs.

Chapter 7: Network Layer: Data Transfer

This chapter discuses the first duty of the network layer: data transfer. It explains the service in this duty such as packetizing, routing, error control, flow control, congestion control, and quality of services. It then describes the concept of packet switching. It also describe network-layer performance. The main goal is to introduce the two versions of the network layer in the Internet: IPv4 and IPv6.

Chapter 8: Network Layer: Routing Packets

This chapter discusses the second duty of the network layer: routing of packets. It discusses unicast routing protocols such as distance vector routing, link-state routing, and path-vector routing. It also discuss multicast routing and protocols.

Chapter 9: Transport Layer

This chapter discusses the transport layer. It first describes the services expected from a transfer-layer protocol. It then describes a simple transport layer protocol UDP. Finally, it describes a more sophisticated protocol TCP. Finally, it describes SCTP, a transport-layer protocol that uses association.

Chapter 10: Application Layer

This chapter discusses the application layer, the highest level in the TCP/IP protocol suite. It shows how this layer uses client/server programs. It then introduces some applications such as the Web, file transfer, and e-mail. Finally, the chapter discusses some peer-to-peer applications. It finally shows how application programs can be created using the C-language.

Chapter 11: Multimedia

This chapter discusses multimedia. It shows how compression is used in multimedia. It then defines the elements of multimedia such as text, image, video, and audio. It then describes how multimedia is used in the Internet.

Chapter 12: Network Management

This chapter introduces network management and discusses five general areas used in network management. It also defines the Simple Network Management Protocol (SNMP) that is used in the Internet, which is based on Simple Management Information (SMI).

Chapter 13: Cryptography and Network Security

This chapter briefly discusses the concept of security goals including confidentiality, integrity, and availability. It then describes how these goals can be achieved using message integrity, message authentication, digital signature, and entity authentication. The chapter then describes how these goals can be achieved using security in the transport layer and application layer.

Appendix A

This appendix discusses Unicode, the coding system used in communication.

Appendix B

This appendix discusses the positional numbering system and how the system uses numbers in different bases.

Appendix C

This appendix discusses mark-up languages such as HTML, CSS, XML, and XSL, which are used in data communications and networking.

Appendix D

This appendix gives a touch of probability that can be useful in understanding some networking protocols.

Appendix E

This appendix discusses checksum.

Appendix F

This appendix gives the list of acronyms used in the book for quick reference.

References

The book contains a list of references for further reading.

Glossary

The Glossary provides definitions for all key terms from the text and other important terminology.

Pedagogy

Several pedagogical features of this text are designed to make it particularly easy for students to understand data communications and networking.

Visual Approach

The book presents highly technical subject matter without complex formulas by using a balance of text and figures. More than 500 figures accompanying the text provide a visual and intuitive opportunity for understanding the material. Figures are particularly important in explaining networking concepts. For many students, these concepts are more easily grasped visually than verbally.

Highlighted Points

The book repeats important concepts in boxes for quick reference and immediate attention.

Examples and Applications

Whenever appropriate, examples illustrate the concepts introduced in the text. Also, some real-life applications provided throughout each chapter help motivate students.

End-of-Chapter Materials

Each chapter ends with a set of materials that includes the following:

Key Terms

The new terms used in each chapter are listed at the end of the chapter, and their definitions are included in the glossary.

Summary

Each chapter ends with a summary of the material covered by that chapter. The summary glues the important materials together to be seen in one shot.

Recommended Reading

This section gives a brief list of references relative to the chapter. The references can be used to quickly find the corresponding literature in the reference section at the end of the book.

Practice Set

Each chapter includes a practice set designed to reinforce salient concepts and encourage students to apply them. It consists of questions, and problems.

Audience

This book is written for both an academic and a professional audience. It can be used as a self-study guide for interested professionals. As a textbook, it can be used for a one-semester or one-quarter course. It is designed for the last year of undergraduate study or the first year of graduate study. Although some problems at the end of the chapters require some knowledge of probability, only general mathematical knowledge taught in the first year of college is needed to study the text.

Instruction Resources

The book contains complete instruction resources that can be downloaded from the book web site https://connect.mheducation.com. They include:

Presentations

The site includes a set of colorful and animated PowerPoint presentations for teaching the course.

Solution to Practice Sets

Solutions to all questions and problems are provided at https://connect.mheducation.com for the use of professors who teach the course.

Quizzes

There are quizzes at the end of each chapter that can be taken by the students. Students are encouraged to take the quizzes to test their general understanding of the materials presented in the corresponding chapter.

Student Solution Manual

Student solution manual is also given for all the chapters that can be referred by the instructors.

Student Resources

The book contains complete student resources that can be downloaded from **mheducation.co.in**. It include:

Student Solution Manual

Student solution manual for all chapters is provided at mheducation.co.in for the use of students.

Acknowledgments

It is obvious that the development of a book of this scope needs the support of many people. I would like to acknowledge the contributions from peer reviewers to the development of the book. These reviewers are:

Azad Azadmanesh, University of Nebraska–Omaha
Maurice Dosso, Mt. Sierra College
John Doyle, Indiana University
Meng Han, Kennesaw State University
Tamer Omar, Cal Poly Pomona
Pat Smith, Oklahoma Christian University
Lawrence Teitelman, Queens College, City University of New York
Zhanyang Zhang, City University of New York

Special thanks go to the staff of McGraw-Hill. Beth Bettcher, the portfolio manager, proved how a proficient publisher can make the impossible, possible. Beth Baugh, the product developer, gave help whenever I needed it. Jane Mohr, the project manager, guided us through the production process with enormous enthusiasm. I also thank Sandeep Rawat, the full-service project manager, and David Hash, the cover designer.

Behrouz A. Forouzan
Los Angeles, CA
January 2021

TRADEMARK

CHAPTER 1

Introduction

Data communications and networking have changed the way we do business and the way we live. The largest computer network, the Internet, has billions of users in the world who use wired and wireless transmission media to connect small and large computers.

Data communications and networking are not only used in business and personal communication but have found many political and social applications. People are able how to communicate with others all over the world to express their social and political opinions and problems. Communities are not isolated any more.

But before we ask how quickly we can get hooked up, we need to know how networks operate, what types of technologies are available, and which design best fills which set of needs.

This chapter paves the way for the rest of the book. It is divided into six sections.

- The first section introduces data communications and defines its components and the types of data exchanged.
- The second section introduces networks and defines their criteria and structures.
- The third section discusses different types of networks: LANs, WANs, and internetworks (internets). It also introduces the Internet, the largest internet in the world.
- The fourth section introduces protocol layering and its principles.
- The fifth section introduces the TCP/IP protocol suite and gives a brief description of each layer.
- The sixth section gives a brief historical description of the OSI model and compares it with the TCP/IP protocol suite.

1.1 DATA COMMUNICATIONS

When we communicate, we are sharing information or data. This sharing can be local or remote. Local communication usually occurs face to face, while remote communication takes place over a distance. The word ***data*** refers to information presented in whatever form is agreed upon by the parties creating and using it.

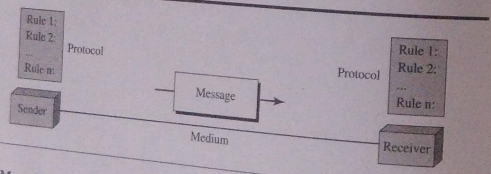
Data communications is the exchange of data between two devices via some form of transmission medium such as a wire cable. For data communications to occur, the communicating devices must be part of a communications system made up of a combination of hardware (physical equipment) and software (programs). The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter.

1. **Delivery.** The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.
2. **Accuracy.** The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.
3. **Timeliness.** The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called *real-time* transmission.
4. **Jitter.** Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 30 ms. If some of the packets arrive with a 30-ms delay and others with a 40-ms delay, the video will have an uneven quality.

1.1.1 Components

A data communications system has five components (see Figure 1.1).

Figure 1.1 *Five components of a data communications system*



1. **Message.** The **message** is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.
2. **Sender.** The **sender** is the device that sends the data message. It can be a computer, a telephone handset, a video camera, and so on.

3. **Receiver.** The **receiver** is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.
4. **Transmission medium.** The **transmission medium** is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.
5. **Protocol.** A **protocol** is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not able to communicate, just as a person speaking French cannot be understood by a person who speaks only Japanese.

1.1.2 Message

Information today comes in different forms such as text, numbers, images, audio, and video.

Text

In data communications, text is represented as a bit pattern, a sequence of bits (0s or 1s). Different sets of bit patterns have been designed to represent text symbols. Each set is called a **code**, and the process of representing symbols is coding. Today, the prevalent coding system is **Unicode**, which uses 32 bits to represent a symbol or character used in any language in the world (see Appendix A).

Numbers

Numbers are also represented by bit patterns. However, a code such as Unicode is not used to represent numbers; a number is directly converted to a binary number to simplify mathematical operations (see Appendix B).

Images

Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The number of pixels depends on the *resolution*. For example, an image can be divided into 1000 pixels or 10,000 pixels. In the second case, there is a better representation of the image (better resolution), but more memory is needed to store the image.

After an image is divided into pixels, each pixel is assigned a bit pattern. The size and the value of the pattern depend on the image. For an image made up of only black and white dots (e.g., a chessboard), a 1-bit pattern is enough to represent a pixel.

If an image is not made up of pure white and pure black pixels, you can increase the size of the bit pattern to include the gray scale. For example, to show four levels of gray scale, you can use 2-bit patterns. A black pixel can be represented by 00, a dark gray pixel by 01, a light gray pixel by 10, and a white pixel by 11.

There are several methods to represent color images. One method is called **RGB**, so called because each color is made up of a combination of three primary colors: *r*ed, green, and *b*lue. The intensity of each color is measured, and a bit pattern is assigned to it. Another method is called **YCM**, in which a color is made up of a combination of three other primary colors: *y*ellow, cyan, and magenta.

Audio

Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete. Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal. Later in the book we learn how to change sound or music to a digital or an analog signal.

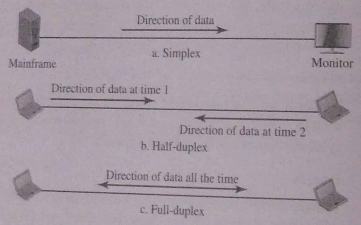
Video

Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion.

1.1.3 Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex as shown in Figure 1.2.

Figure 1.2 *Data flow (simplex, half-duplex, and full-duplex)*



Simplex

In **simplex mode**, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive (see Figure 1.2a). Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output.

Half-Duplex

In **half-duplex mode**, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa (see Figure 1.2b). Walkie-talkies and CB (citizens band) radios are both half-duplex systems.

Full-Duplex

In **full-duplex mode**, both stations can transmit and receive simultaneously (see Figure 1.2c). The full-duplex mode is like a two-way street with traffic flowing in both directions at the same time. In full-duplex mode, signals going in one direction share the capacity of the link with signals going in the other direction. One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.

1.2 NETWORKS

A **network** is the interconnection of a set of devices capable of communication. In this definition, a device can be a **host**, such as a large computer, desktop, laptop, workstation, cellular phone, or security system. A device in this definition can also be a **connecting device** such as a router that connects the network to other networks, a switch that connects devices together, or a modem (modulator-demodulator) that changes the form of data.

1.2.1 Network Criteria

A network must be able to meet a certain number of criteria. The most important of these are ***performance***, ***reliability***, and ***security***.

Performance

Performance can be measured in many ways, including transit time and response time. Transit time is the amount of time required for a message to travel from one device to another. Response time is the elapsed time between an inquiry and a response. The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software.

Reliability

In addition to accuracy of delivery, network **reliability** is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

Security

Network **security** issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

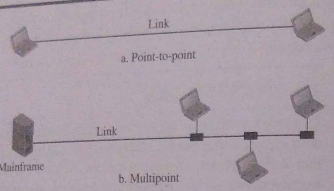
1.2.2 Physical Structures

Before discussing networks, we need to define some network attributes.

Type of Connection

A network is two or more devices connected through links. A link is a communications pathway that transfers data from one device to another. For communication to occur, two devices must be connected in some way to the same link at the same time. There are two possible types of connections: *point-to-point* and *multipoint* (see Figure 1.3 on next page).

Figure 1.3 *Types of connections: point-to-point and multipoint*



Point-to-Point

A **point-to-point connection** provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices.

Multipoint

A **multipoint** (also called **multidrop**) **connection** is one in which more than two devices share a single link. In a multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a *spatially shared* connection. If users must take turns, it is a *timeshared* connection.

Physical Topology

The term ***physical topology*** refers to the way in which a network is laid out physically. Two or more devices connect to a link; two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called **nodes**) to one another. There are four basic topologies possible: *mesh*, *star*, *bus*, and *ring*.

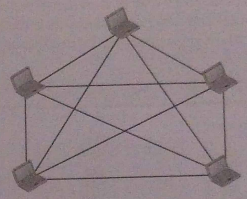
Mesh Topology

In a **mesh topology**, every device has a dedicated point-to-point link to every other device. The term *dedicated* means that the link carries traffic only between the two devices it connects. To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to $n - 1$ nodes, node 2 must be connected to $n - 1$ nodes, and finally node n must be connected to $n - 1$ nodes. We need $n(n - 1)$ physical links. To accommodate that many links, every device on the network must have $n - 1$ input/output (I/O) ports (see Figure 1.4 on next page) to be connected to the other $n - 1$ stations.

Star Topology

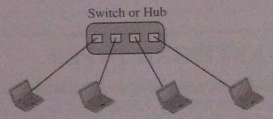
In a **star topology**, each device has a dedicated point-to-point link only to a central controller, usually called a **hub**. The devices are not directly linked to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends

Figure 1.4 *A fully connected mesh topology (five devices)*



the data to the controller, which then relays the data to the other connected device (see Figure 1.5).

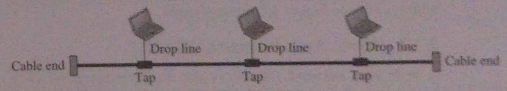
Figure 1.5 *A star topology connecting four stations*



Bus Topology

The preceding topology examples all describe point-to-point connections. A **bus topology**, on the other hand, is multipoint. One long cable acts as a **backbone** to link all the devices in a network (see Figure 1.6).

Figure 1.6 *A bus topology connecting three stations*

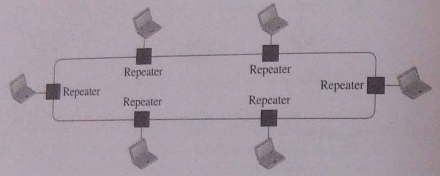


Nodes are connected to the bus cable by drop lines and taps. A *drop line* is a connection running between the device and the main cable. A *tap* is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core.

Ring Topology

In a **ring topology**, each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater, which regenerates the bits and passes them along (see Figure 1.7).

Figure 1.7 *A ring topology connecting six stations*



1.3 NETWORK TYPES

Now we discuss different types of networks: LANs and WANs.

1.3.1 Local Area Network

A **local area network** (**LAN**) is usually privately owned and connects some hosts in a single office, building, or campus.

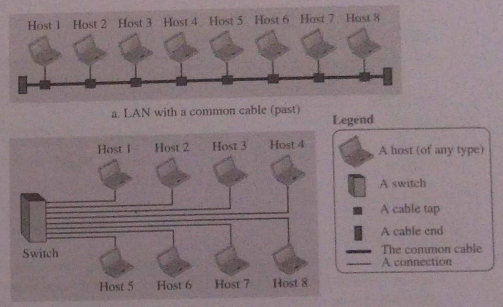
Each host in a LAN has an identifier, which is an address that uniquely defines the host in the LAN. A packet sent by a host to another host carries both the source host's and the destination host's addresses.

When LANs were used in isolation (which is rare today), they were designed to allow resources to be shared between the hosts. As we will see shortly, LANs today are connected to each other and to WANs (discussed next) to create communication at a wider level (see Figure 1.8 on next page).

1.3.2 Wide Area Network (WAN)

A **wide area network (WAN)** is also an interconnection of devices capable of communication. However, there are some differences between a LAN and a WAN. A LAN is normally limited in size, spanning an office, a building, or a campus; a WAN has a wider geographical span, spanning a town, a state, a country, or even the world. A LAN interconnects hosts; a WAN interconnects connecting devices such as switches, routers, or modems. We see two distinct examples of WANs today: point-to-point WANs and switched WANs.

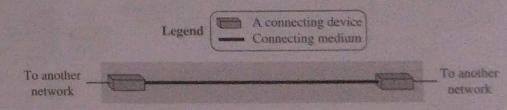
Figure 1.8 *An isolated LAN in the past and today*



Point-to-Point WAN

A point-to-point WAN is a network that connects two communicating devices through a transmission medium (cable or air). Figure 1.9 shows an example of a point-to-point WAN.

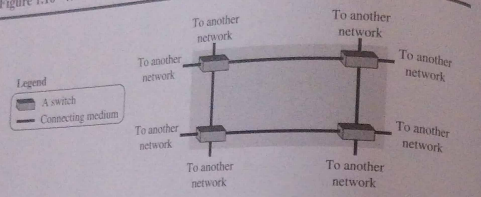
Figure 1.9 *A point-to-point WAN*



Switched WAN

A switched WAN is a network with more than two ends. It is used in the backbone of a global communications network today. Figure 1.10 shows an example of a switched WAN.

Figure 1.10 *A switched WAN*



Internetwork

Today, it is very rare to see a LAN or a WAN in isolation; they are connected to one another. When two or more networks are connected, they make an **internetwork**, or **internet**. As an example, assume that an organization has two offices, one on the east coast and the other on the west coast. Each office has a LAN that allows all employees in the office to communicate with each other. To make the communication between employees at different offices possible, the management leases a point-to-point dedicated WAN from a service provider, such as a telephone company, and connects the two LANs. Now the company has an internetwork, or a private internet (with lowercase *i*). Communication between offices is now possible. Figure 1.11 shows this internet.

Figure 1.11 *An internetwork made of two LANs and one point-to-point WAN*

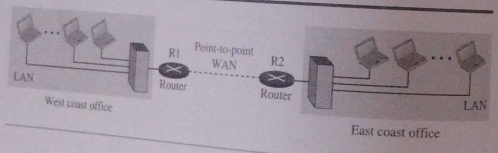


Figure 1.12 shows another internet with several LANs and WANs connected. One of the WANs is a switched WAN with four switches.

1.3.3 The Internet

As we discussed before, an *internet* (note the lowercase *i*) is two or more networks that can communicate with each other. The most notable internet is called the **Internet** (uppercase *I*) and is composed of thousands of interconnected networks. Figure 1.13 shows a conceptual (not geographical) view of the Internet.

Figure 1.12 *A heterogeneous internetwork made of four WANs and two LANs*

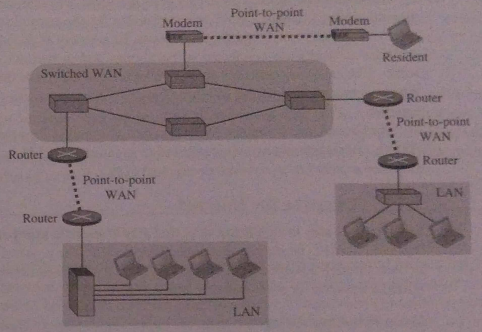
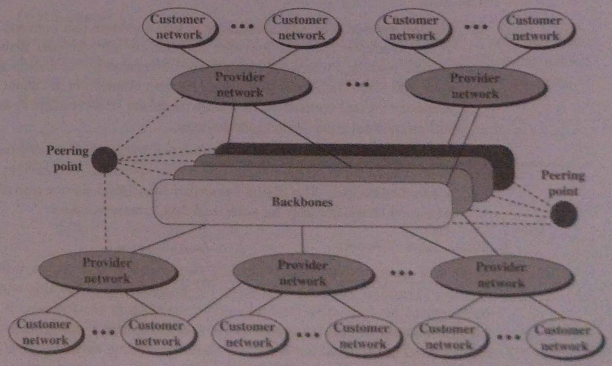


Figure 1.13 *The Internet today*



The figure shows the Internet as several backbones, provider networks, and customer networks. At the top level, the *backbones* are large networks owned by some communication companies. The backbone networks are connected through some complex switching systems, called *peering points*. At the second level, there are smaller networks, called *provider networks*, that use the services of the backbones for a fee. The provider networks are connected to backbones and sometimes to other provider networks. The *customer networks* are networks at the edge of the Internet that actually use the services provided by the Internet. They pay fees to provider networks for receiving services.

Backbones and provider networks are also called **Internet Service Providers (ISPs)**. The backbones are often referred to as international ISPs; the provider networks are often referred to as national or regional ISPs.

1.3.4 Accessing the Internet

The Internet today is an internetwork that allows any user to become part of it. The user, however, needs to be physically connected to an ISP. The physical connection is normally done through a point-to-point WAN (such as a telephone network, a cable network, a wireless network, or other types of networks).

Using Telephone Networks

Today most residences and small businesses have telephone service, which means they are connected to a telephone network. Because most telephone networks have already connected themselves to the Internet, one option for residences and small businesses to connect to the Internet is to change the voice line between the residence or business and the telephone center to a point-to-point WAN. This can be done in two ways.

- ***Dial-up service.*** The first solution is to add a modem that converts data to voice to the telephone line. The software installed on the computer dials the ISP and imitates making a telephone connection. Unfortunately, the dial-up service is very slow, and when the line is used for an Internet connection, it cannot be used for a telephone (voice) connection. It is only useful for small residences and businesses with occasional connection to the Internet.
- ***DSL Service.*** Since the advent of the Internet, some telephone companies have upgraded their telephone lines to provide higher-speed Internet services to residences or small businesses. The digital subscriber line (DSL) service also allows the line to be used simultaneously for voice and data communications.

Using Cable Networks

More and more residents over the last two decades have begun using cable TV services instead of antennas to receive TV broadcasting. The cable companies have been upgrading their cable networks and connecting to the Internet. A residence or a small business can be connected to the Internet by using this service. It provides a higher-speed connection, but the speed varies depending on the number of neighbors that use the same cable.

Using Wireless Networks

Wireless connectivity has recently become increasingly popular. A household or a small business can use a combination of wireless and wired connections to access the Internet. With the growing wireless WAN access, a household or a small business can be connected to the Internet through a wireless WAN.

Direct Connection to the Internet

A large organization or a large corporation can itself become a local ISP and be connected to the Internet. This can be done if the organization or the corporation leases a high-speed WAN from a carrier provider and connects itself to a regional ISP. For example, a large university with several campuses can create an internetwork and then connect the internetwork to the Internet.

1.4 PROTOCOL LAYERING

We defined the term *protocol* before. In data communications and networking, a protocol defines the rules that both the sender and receiver and all intermediate devices need to follow to be able to communicate effectively. When communication is simple, we may need only one simple protocol; when the communication is complex, we may need to divide the task between different layers, in which case we need a protocol at each layer, or **protocol layering**.

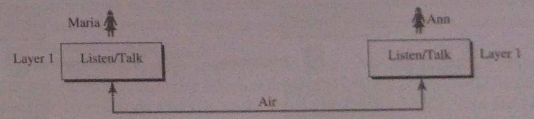
1.4.1 Scenarios

Let us develop two simple scenarios to better understand the need for protocol layering.

First Scenario

In the first scenario, communication is so simple that it can occur in only one layer. Assume Maria and Ann are neighbors with a lot of common ideas. Communication between Maria and Ann takes place in one layer, face to face, in the same language, as shown in Figure 1.14.

Figure 1.14 *A single-layer protocol*



Even in this simple scenario, we can see that a set of rules needs to be followed. First, Maria and Ann know that they should greet each other when they meet. Second, they know that they should confine their vocabulary to the level of their friendship.

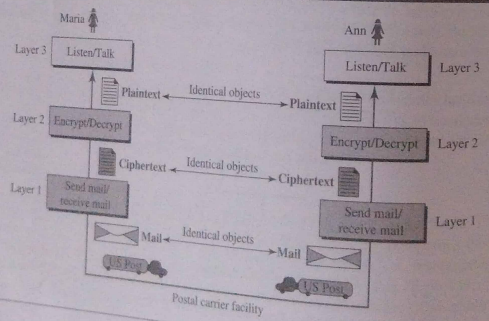
Third, each party knows that she should refrain from speaking when the other party is speaking. Fourth, each party knows that the conversation should be a dialog, not a monolog: Both should have the opportunity to talk about the issue. Fifth, they should exchange some nice words when they leave.

We can see that the protocol used by Maria and Ann is different from the communication between a professor and the students in a lecture hall. The communication in the second case is mostly monolog; the professor talks most of the time unless a student has a question, a situation in which the protocol dictates that she should raise her hand and wait for permission to speak. In this case, the communication is normally very formal and limited to the subject being taught.

Second Scenario

In the second scenario, we assume that Ann is offered a higher-level position in her company but needs to move to another branch located in a city very far from Maria. The two friends still want to continue their communication and exchange ideas because they have come up with an innovative project to start a new business when they both retire. They decide to continue their conversations using regular mail through the post office. However, they do not want their ideas to be revealed to other people if the letters are intercepted. They agree on an encryption/decryption technique. The sender of the letter encrypts it to make it unreadable by an intruder; the receiver of the letter decrypts it to get the original letter. We discuss the encryption/decryption methods later in the book, but for the moment we assume that Maria and Ann use one technique to make it hard to decrypt the letter if one does not have the key for doing so. Now we can say that the communication between Maria and Ann takes place in three layers, as shown in Figure 1.15. We assume that Ann and Maria each have three machines (or robots) that can perform the task at each layer.

Figure 1.15 *A three-layer protocol*



Let us assume that Maria sends the first letter to Ann. Maria talks to the machine at the third layer as though the machine is Ann and is listening to her. The third-layer machine listens to what Maria says and creates the plaintext (a letter in English), which is passed to the second-layer machine. The second-layer machine takes the plaintext, encrypts it, and creates the ciphertext, which is passed to the first-layer machine. The first-layer machine, presumably a robot, takes the ciphertext, puts it in an envelope, adds the sender and receiver addresses, and mails it.

At Ann's side, the first-layer machine picks up the letter from Ann's mailbox, recognizing the letter from Maria by the sender address. The machine takes out the ciphertext from the envelope and delivers it to the second-layer machine. The second-layer machine decrypts the message, creates the plaintext, and passes the plaintext to the third-layer machine. The third-layer machine takes the plaintext and reads it as though Maria is speaking.

Protocol layering enables us to divide a complex task into several smaller and simpler tasks. For example, in Figure 1.15, we could have used only one machine to do the job of all three machines. However, if Maria and Ann decide that the encryption/decryption done by the machine is not enough to protect their secrecy, they have to change the whole machine. In the present situation, they need to change only the second-layer machine; the other two can remain the same. This is referred to as *modularity*. Modularity in this case means independent layers. A layer (module) can be defined as a black box with inputs and outputs, without concern about how inputs are changed to outputs. If two machines provide the same outputs when given the same inputs, they can replace each other. For example, Ann and Maria can buy the second-layer machine from two different manufacturers. As long as the two machines create the same ciphertext from the same plaintext and vice versa, they do the job.

One of the advantages of protocol layering is that it allows us to separate the services from the implementation. A layer needs to be able to receive a set of services from the lower layer and to give the services to the upper layer; we don't care about how the layer is implemented. For example, Maria may decide not to buy the machine (robot) for the first layer; she can do the job herself. As long as Maria can do the tasks provided by the first layer, in both directions, the communications system works.

Another advantage of protocol layering, which cannot be seen in our simple examples, but reveals itself when we discuss protocol layering in the Internet, is that communication does not always use only two end systems; there are intermediate systems that need only some layers, but not all layers. If we did not use protocol layering, we would have to make each intermediate system as complex as the end systems, which makes the whole system more expensive.

Is there any disadvantage to protocol layering? One can argue that having a single layer makes the job easier. There is no need for each layer to provide a service to the upper layer and give service to the lower layer. For example, Ann and Maria could find or build one machine that could do all three tasks. However, as mentioned above, if one day they found that their code was broken, each would have to replace the whole machine with a new one instead of just changing the machine in the second layer.

1.4.2 Principles of Protocol Layering

Let us discuss the two principles of protocol layering.

First Principle

The first principle dictates that if we want bidirectional communication, we need to make each layer so that it is able to perform two opposite tasks, one in each direction. For example, the third-layer task is to listen (in one direction) and *talk* (in the other direction). The second layer needs to be able to encrypt and decrypt. The first layer needs to send and receive mail.

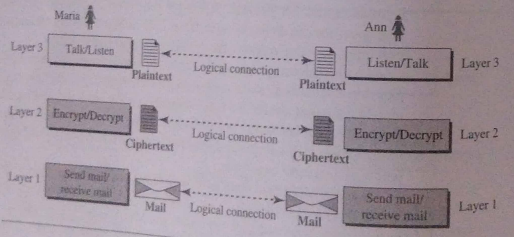
Second Principle

The second important principle that we need to follow in protocol layering is that the two objects under each layer at both sites should be identical. For example, the object under the third layer at both sites should be a plaintext letter. The object under the second layer at both sites should be a ciphertext letter. The object under the first layer at both sites should be a piece of mail.

1.4.3 Logical Connections

After following the above two principles, we can think about logical connections between each layer as shown in Figure 1.16. This means that we have layer-to-layer communication. Maria and Ann can think that there is a logical (imaginary) connection at each layer through which they can send the object created from that layer. We will see that the concept of logical connection will help us better understand the task of layering we encounter in data communications and networking.

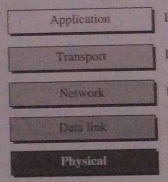
Figure 1.16 *Logical connections between peer layers*



1.5 TCP/IP PROTOCOL SUITE

Now that we know about the concept of protocol layering and the logical connections between layers in our second scenario, we can introduce the **Transmission Control Protocol/Internet Protocol (TCP/IP).** TCP/IP is a protocol suite (a set of protocols organized in different layers) used in the Internet today. It is a hierarchical protocol made up of interactive modules, each of which provides a specific functionality. The term *hierarchical* means that each upper-level protocol is supported by the services provided by one or more lower-level protocols. The **TCP/IP protocol suite** is defined as five layers as shown in Figure 1.17.

Figure 1.17 *Layers in the TCP/IP protocol suite*



1.5.1 Layered Architecture

To show how the layers in the TCP/IP protocol suite are involved in communication between two hosts, we assume that we want to use the suite in a small internet made up of three LANs (links), each with a link-layer switch. We also assume that the links are connected by one router, as shown in Figure 1.18 (on next page).

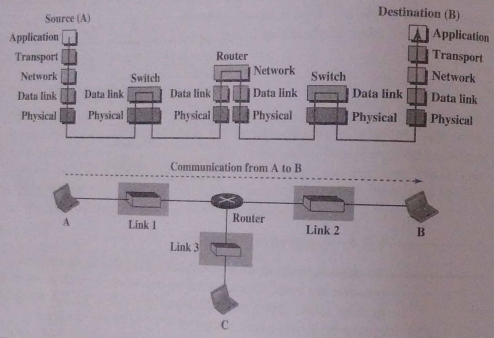
Let us assume that computer A communicates with computer B. As Figure 1.18 shows, we have five communicating devices in this communication: source host (computer A), the link-layer switch in link 1, the router, the link-layer switch in link 2, and the destination host (computer B). Each device is involved with a set of layers depending on the role of the device in the internet. The two hosts are involved in all five layers. The source host needs to create a message in the application layer and send it down the layers so that it is physically sent to the destination host. The destination host needs to receive the communication at the physical layer and then deliver it through the other layers to the application layer.

The router is involved only in three layers; there is no transport or application layer in a router as long as the router is used only for routing. Although a router is always involved in one network layer, it is involved in n combinations of link and physical layers in which n is the number of links the router is connected to. The reason is that each link

may use its own data-link or physical protocol. For example, in Figure 1.18, the router is involved in three links, but the message sent from source computer A to destination computer B is involved in two links. Each link may be using different link-layer and physical-layer protocols; the router needs to receive a packet from link 1 based on one pair of protocols and deliver it to link 2 based on another pair of protocols.

A link-layer switch in a link, however, is involved only in two layers, data-link and physical protocols. Although each switch in Figure 1.18 has two different connections, the connections are in the same link, which uses only one set of protocols. This means that, unlike a router, a link-layer switch is involved only in one data-link and one physical layer.

Figure 1.18 *Communication through an internet*

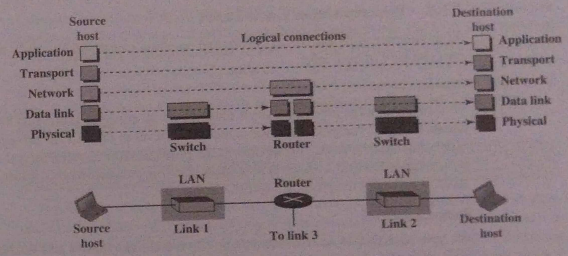


1.5.2 Brief Description of Layers

We now briefly discuss the functions and duties of layers in the TCP/IP protocol suite. Each layer is discussed in detail in a separate chapter of the book. To better understand the duties of each layer, we need to think about the logical connections between the layers. Figure 1.19 shows the logical connections in our simple internet.

Using logical connections makes it easier for us to think about the duty of each layer. As Figure 1.19 shows, the duty of the application, transport, and network layers is end-to-end. However, the duty of the data-link and physical layers is hop-to-hop, in which a hop is a host or router. In other words, the domain of duty of the top three layers is the internet, and the domain of duty of the two lower layers is the link.

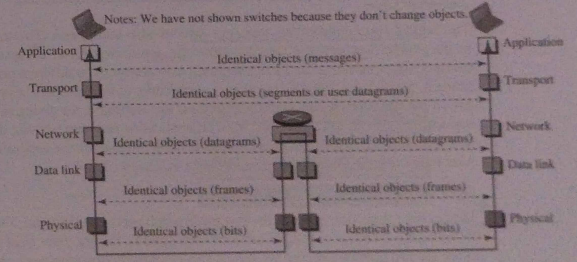
Figure 1.19 *Logical connections between layers of the TCP/IP protocol suite*



Another way of thinking about the logical connections is to think about the data unit created from each layer. In the top three layers, the data unit (packets) should not be changed by any router or link-layer switch. In the bottom two layers, the packet created by the host is changed only by the routers, not by the link-layer switches.

Figure 1.20 shows the second principle discussed previously for protocol layering. We show the identical objects below each layer related to each device.

Figure 1.20 *Identical objects in the TCP/IP protocol suite*



Note that, although the logical connection at the network layer is between the two hosts, we can only say that identical objects exist between two hops in this case because a router may fragment the packet at the network layer and send more packets than

received. (See Chapter 4 for a discussion of fragmentation.) Note that the link between two hops does not change the object.

1.5.3 Description of Each Layer

After understanding the concept of logical communication, we are ready to briefly discuss the duty of each layer.

Physical Layer

We can say that the physical layer is responsible for carrying individual bits in a frame across the link. The physical layer is the lowest level in the TCP/IP protocol suite. The communication between two devices at the physical layer is still a logical communication because there is another hidden layer, the transmission media, under the physical layer. We discuss the physical layer in Chapter 2.

Data-Link Layer

We have seen that an *internet* is made up of several links (LANs and WANs) connected by routers. When the next link to travel is determined by the router, the data-link layer is responsible for taking the datagram and moving it across the link. We discuss the data-link layer in Chapters 3, 4, 5, and 6.

Network Layer

The network layer is responsible for creating a connection between the source computer and the destination computer. The communication at the network layer is host-to-host. However, because there can be several routers from the source to the destination, the routers in the path are responsible for choosing the best route for each packet. We discuss the network layer in Chapters 7 and 8.

Transport Layer

The logical connection at the transport layer is also end-to-end. The transport layer at the source host gets the message from the application layer; encapsulates it in a transport-layer packet (called a *segment* or a *user datagram* in different protocols); and sends it, through the logical (imaginary) connection, to the transport layer at the destination host. In other words, the transport layer is responsible for giving services to the application layer: to get a message from an application program running on the source host and deliver it to the corresponding application program on the destination host. We need to say that we discuss the transport layer in Chapter 9.

Application Layer

The logical connection between the two application layers is end-to-end. The two application layers exchange *messages* between each other as though there were a bridge between the two layers. However, we should know that the communication is done through all the layers. Communication at the application layer is between two *processes* (two programs running at this layer). To communicate, a process sends a request to the other process and receives a response. Process-to-process communication is the duty of the application layer. We discuss the application layer in Chapter 10.

1.6 THE OSI MODEL

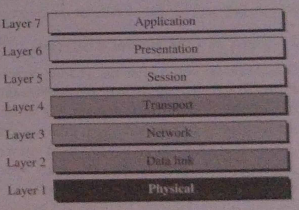
Although, when speaking of the Internet, everyone talks about the TCP/IP protocol suite, it is not the only suite of protocols defined. Established in 1947, the **International Organization for Standardization (ISO)** is a multinational body dedicated to worldwide agreement on international standards. Almost three-fourths of the countries in the world are represented in the ISO. An ISO standard that covers all aspects of network communications is the **Open Systems Interconnection (OSI)** model. It was first introduced in the late 1970s.

ISO is the organization; OSI is the model.

An *open system* is a set of protocols that allows any two different systems to communicate regardless of their underlying architecture. The purpose of the OSI model is to show how to facilitate communication between different systems without requiring changes to the logic of the underlying hardware and software. The OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust, and interoperable. The OSI model was intended to be the basis for the creation of the protocols in the OSI stack.

The OSI model is a layered framework for the design of network systems that allows communication between all types of computer systems. It consists of seven separate but related layers, each of which defines a part of the process of moving information across a network (see Figure 1.21).

Figure 1.21 *The OSI model*

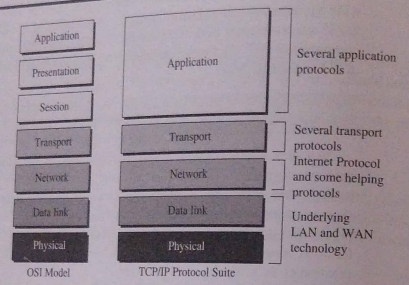


1.6.1 OSI versus TCP/IP

When we compare the two models, we find that two layers, session and presentation, are missing from the TCP/IP protocol suite. These two layers were not added to the TCP/IP protocol suite after the publication of the OSI model. The application layer in the suite

is usually considered to be the combination of three layers in the OSI model, as shown in Figure 1.22.

Figure 1.22 *TCP/IP and OSI model*



Two reasons were mentioned for this decision. First, TCP/IP has more than one transport-layer protocol. Some of the functionalities of the session layer are available in some of the transport-layer protocols. Second, the application layer is not only one piece of software. Many applications can be developed at this layer. If some of the functionalities mentioned in the session and presentation layers are needed for a particular application, they can be included in the development of that piece of software.

1.6.2 Lack of OSI Model's Success

The OSI model appeared after the TCP/IP protocol suite. Most experts were at first excited and thought that the TCP/IP protocol would be fully replaced by the OSI model. This did not happen for several reasons, but we describe only three, which are agreed upon by all experts in the field. First, OSI was completed when TCP/IP was fully in place and a lot of time and money had been spent on the suite; changing it would cost a lot. Second, some layers in the OSI model were never fully defined. For example, although the services provided by the presentation and the session layers were listed in the document, actual protocols for these two layers were not fully defined, nor were they fully described, and the corresponding software was not fully developed. Third, when OSI was implemented by an organization in a different application, it did not show a high enough level of performance to entice the Internet authority to switch from the TCP/IP protocol suite to the OSI model.

1.7 END-OF-CHAPTER MATERIALS

1.7.1 Recommended Reading

For more details about subjects discussed in this chapter, we recommend the following books, papers, and Requests for Comments (RFCs). The items enclosed in brackets refer to the reference list at the end of the book.

Books and Papers

Several books and papers give a thorough coverage about the materials discussed in this chapter: [Seg 98], [Lei et al. 98], [Kle 04], [Cer 89], and [Jen et al. 86].

Requests for Comments

Two RFCs in particular discuss the TCP/IP suite: RFC 791 (IP) and RFC 817 (TCP). In future chapters we list different RFCs related to each protocol in each layer.

1.7.2 Key Terms

audio
backbone
bus topology
code
connecting device
data
data communications
full-duplex mode
half-duplex mode
host
hub
image
International Organization for Standardization (ISO)
internet
Internet
Internet Service Provider (ISP)
internetwork
local area network (LAN)
mesh topology
message
multipoint or multidrop connection
network
node
Open System Interconnection (OSI)
performance
physical topology
point-to-point connection
protocol
protocol layering
receiver
reliability
RGB
ring topology
security
sender
simplex mode
star topology
TCP/IP protocol suite
Transmission Control Protocol/ Internet Protocol (TCP/IP)
transmission medium
Unicode
video
wide area network (WAN)
YCM

1.7.3 Summary

Data communications are the transfer of data from one device to another via some form of transmission medium. A data communications system must transmit data to the correct destination in an accurate and timely manner. The five components that make up a data communications system are the message, sender, receiver, medium, and

protocol. Text, numbers, images, audio, and video are different forms of information. Data flow between two devices can occur in one of three ways: simplex, half-duplex, or full-duplex.

A network is a set of communication devices connected by media links. In a point-to-point connection, two and only two devices are connected by a dedicated link. In a multipoint connection, three or more devices share a link. Topology refers to the physical or logical arrangement of a network. Devices may be arranged in a mesh, star, bus, or ring topology.

A network can be categorized as a local area network or a wide area network. A LAN is a data communication system within a building, plant, or campus, or between nearby buildings. A WAN is a data communication system spanning states, countries, or the whole world. An internet is a network of networks. The Internet is a collection of many separate networks.

TCP/IP is a hierarchical protocol suite made up of five layers: physical, data-link, network, transport, and application. The physical layer coordinates the functions required to transmit a bit stream over a physical medium. The data-link layer is responsible for delivering data units from one station to the next without errors. The network layer is responsible for the source-to-destination delivery of a packet across multiple network links. The transport layer is responsible for the process-to-process delivery of the entire message. The application layer enables the users to access the network.

Four levels of addresses are used in an internet following the TCP/IP protocols: physical (link) addresses, logical (IP) addresses, port addresses, and specific addresses. The physical address, also known as the link address, is the address of a node as defined by its LAN or WAN. The IP address uniquely defines a host on the Internet. The port address identifies a process on a host. A specific address is a user-friendly address.

Another model that defines protocol layering is the Open Systems Interconnection (OSI) model. Two layers in the OSI model, session and presentation, are missing from the TCP/IP protocol suite. These two layers were not added to the TCP/IP protocol suite after the publication of the OSI model. The application layer in the suite is usually considered to be the combination of three layers in the OSI model. The OSI model did not replace the TCP/IP protocol suite because it was completed when TCP/IP was fully in place and because some layers in the OSI model were never fully defined.

1.8 PRACTICE SET

1.8.1 Quizzes

A set of interactive quizzes for this chapter can be found on the book website. It is strongly recommended that the students take the quizzes to check their understanding of the materials before continuing with the practice set.

1.8.2 Questions

Q1-1. Identify the five components of a data communications system.

Q1-2. What are the three criteria necessary for an effective and efficient network?

Q1-3. What are the advantages of a multipoint connection over a point-to-point connection?

Q1-4. What are the two types of line configuration?

Q1-5. Categorize the four basic topologies in terms of line configuration.

Q1-6. What is the difference between half-duplex and full-duplex transmission modes?

Q1-7. Name the four basic network topologies, and cite an advantage of each type.

Q1-8. For n devices in a network, what is the number of cable links required for a mesh, ring, bus, and star topology?

Q1-9. What are some of the factors that determine whether a communications system is a LAN or WAN?

Q1-10. What is an internet? What is the Internet?

Q1-11. Why are protocols needed?

Q1-12. In a LAN with a link-layer switch (Figure 1.8b), host 1 wants to send a message to host 3. Because communication is through the link-layer switch, does the switch need to have an address? Explain.

Q1-13. How many point-to-point WANs are needed to connect n LANs if each LAN should be able to directly communicate with any other LAN?

Q1-14. When a resident uses a dial-up or DLS service to connect to the Internet, what is the role of the telephone company?

Q1-15. What is the first principle we discussed in this chapter for protocol layering that needs to be followed to make the communication bidirectional?

Q1-16. Which layers of the TCP/IP protocol suite are involved in a link-layer switch?

Q1-17. A router connects three links (networks). How many of each of the following layers can the router be involved with?
a. physical layer **b.** data-link layer **c.** network layer

Q1-18. In the TCP/IP protocol suite, what are the identical objects at the sender and the receiver sites when we think about the logical connection at the application layer?

Q1-19. A host communicates with another host using the TCP/IP protocol suite. What is the unit of data sent or received at each of the following layers?
a. application layer **b.** network layer **c.** data-link layer

Q1-20. Which of the following data units is encapsulated in a frame?
a. a user datagram **b.** a datagram **c.** a segment

Q1-21. Which of the following data units has an application-layer message plus the header from layer 4?
a. a frame **b.** a user datagram **c.** a bit

Q1-22. List some application-layer protocols mentioned in this chapter.

Q1-23. If a port number is 16 bits (2 bytes), what is the minimum header size at the transport layer of the TCP/IP protocol suite?

Q1-24. What are the types of addresses (identifiers) used in each of the following layers?
a. application layer **b.** network layer **c.** data-link layer

Q1-25. Assume we want to connect two isolated hosts together to let each host communicate with the other. Do we need a link-layer switch between the two? Explain.

Q1-26. If there is a single path between the source host and the destination host, do we need a router between the two hosts?

1.8.3 Problems

P1-1. What is the maximum number of characters or symbols that can be represented by Unicode?

P1-2. A color image uses 16 bits to represent a pixel. What is the maximum number of different colors that can be represented?

P1-3. Assume six devices are arranged in a mesh topology. How many cables are needed? How many ports are needed for each device?

P1-4. For each of the following four networks, discuss the consequences if a connection fails.

a. Five devices arranged in a mesh topology
b. Five devices arranged in a star topology (not counting the hub)
c. Five devices arranged in a bus topology
d. Five devices arranged in a ring topology

P1-5. In the ring topology in Figure 1.7, what happens if one of the stations is unplugged?

P1-6. In the bus topology in Figure 1.6, what happens if one of the stations is unplugged?

P1-7. When a party makes a local telephone call to another party, is this a point-to-point or multipoint connection? Explain your answer.

P1-8. Compare the telephone network and the Internet. What are the similarities? What are the differences?

P1-9. Answer the following questions about Figure 1.15 when the communication is from Maria to Ann:

a. What is the service provided by layer 2 to layer 3 at Maria's site?
b. What is the service provided by layer 2 to layer 3 at Ann's site?

P1-10. Assume that the number of hosts connected to the Internet at year 2010 is 500 million. If the number of hosts increases only 20 percent per year, what is the number of hosts in year 2020?

P1-11. Assume a system uses five protocol layers. If the application program creates a message of 100 bytes and each layer (including the fifth and the first) adds a header of 10 bytes to the data unit, what is the efficiency (the ratio of application-layer bytes to the number of bytes transmitted) of the system?

P1-12. Match the following to one or more layers of the TCP/IP protocol suite:

a. route determination
b. connection to transmission media
c. providing services for the end user

P1-13. Match the following to one or more layers of the TCP/IP protocol suite:

a. creating user datagrams
b. responsibility for handling frames between adjacent nodes

P1-14. Assume that a private internet requires that the messages at the application layer be encrypted and decrypted for security purposes. If we need to add some information about the encryption/decryption process (such as the algorithms used in the process), does it mean that we are adding one layer to the TCP/IP protocol suite? Redraw the TCP/IP layers (Figure 1.17b) if you think so.

P1-15. Protocol layering can be found in many aspects of our lives such as air travelling. Imagine you make a round trip to spend some time on vacation at a resort. You need to go through some processes at your city airport before flying. You also need to go through some processes when you arrive at the resort airport. Show the protocol layering for the round trip using some layers such as baggage checking/claiming, boarding/unboarding, takeoff/landing.

P1-16. The presentation of data is becoming more and more important in today's Internet. Some people argue that the TCP/IP protocol suite needs to add a new layer to take care of the presentation of data (see Appendix C). If this new layer is added in the future, where should its position be in the suite? Redraw Figure 1.17 to include this layer.