M.Sc. (IT) Previous Year
MIT-03

DATA COMMUNICATION AND COMPUTER NETWORKING

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय – भोपाल
MADHYA PRADESH BHOJ (OPEN) UNIVERSITY – BHOPAL

COURSE WRITERS

Rajneesh Agrawal, Senior Scientists, Department of Information Technology, Government of India
Units: (1.0-1.6, 1.7, 2, 3.0-3.8, 4.0-4.2, 5.0-5.2, 5.3.1, 5.4-5.9.1)

Jyoti Singh, Senior Lecturer, Jaipuria institute of management, Noida
Units: (4.3-4.5.3, 4.6-4.7, 5.10)

Gagan Varshney, Professer & HOD, IMS Engineering College, Ghaziabad
Units: (1.7.1-1.12, 3.8.1-3.9.1, 3.10-3.16, 4.5.4, 4.8-4.13, 5.3, 5.3.2-5.3.3, 5.10.1-5.10.2, 5.11-5.18)

Information contained in this book has been published by VIKAS® Publishing House Pvt. Ltd. and has been obtained by its Authors from sources believed to be reliable and are correct to the best of their knowledge. However, the Madhya Pradesh Bhoj (Open) University, Bhopal, Publisher and its Authors shall in no event be liable for any errors, omissions or damages arising out of use of this information and specifically disclaim any implied warranties or merchantability or fitness for any particular use.

Published by Registrar, MP Bhoj (Open) University, Bhopal in 2020

VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Data Communication and Computer Networking

Syllabi	Mapping in Book
UNIT - I **Fundamentals of Data Communication :** Analog Versus Digital, Fundamentals of Data Transmission Communication Modes Transmission System, Synchronous System, Communication Channnels-Modems : Classification of Modems Modem Based on Range Modems Based on Line Modems Based on Operation Mode Modems Based on Synchronization Modems Based on Modulation. Data Transmission Protocol : Protocols An Overview of Networking The Role of Computer Networks in Development, Transmission Media : Introduction Transmission Concepts and Terms Master Sites Interconnection to Telephone.	**Unit-1:** Unit-1: Fundamentals of Data Communication, Communication Channels and Data Transmission Protocol **(Pages 3-44)**
Unit - II **Local Area Network :** Local Area Network Baseband Versus Broadband LAN Hardware LAN Operating Systems. Implementing LAN : Implementation of LAN Using Fiber-Optic Cables Implementation of LAN Using Wireless Technology Fast LANs Non-standard LANs. Extending LAN : Transmission Concepts and Terms, Master sites, Interconnection to Telephone.	**Unit-2:** Local Area Network, Implementing and Extending LAN **(Pages 45-95)**
Unit - III **Data Transmission Network :** Telephone Networks, Wan Technologies. TCP/IP and the Internet : History of Internet, Internet2 Internet Services Standards for TCP/IP and the Internet RFCS and the TCP/IP Standardization Process. Network Architectures and OSI, Network Architectures Layering the Communications Process The Need for Layered Solutions Open Systems Interconnection (OSI) Model. Routing and Congestion Control : Routing Concepts Routing in Wide Area Networks Hop-By-Hop versus Source Routing Congestion Control, Deadlocks. Queueing Theory: Basic Design Techniques : Basic Concepts, Queueing Model and Factors Traffic Theory Lost Call Rate.	**Unit-3**: Data Transmission Network, TCP/IP and OSI Model **(Pages 97-168)**
Unit - IV **Wide Area Network :** Introduction Network Using WAN and Network Services Communication Protocols Over WAN. Transmission Control Protocol/Internet Protocol (TCP/IP) : LAN Protocol and OSI TCP/IP Protocol Data Transmission by TCP and Ethernet Data Encapsulation Data Routing TCP/IP Services and Application Protocols. Data Link Layer Address : Physical Address. Naming, Addressing, and Routing : Network Layer Addresses Subnet Address Resolution Protocol (ARP) Domain Name System (DNS).	**Unit-4:** Wide Area Network, TCP/IP and Data Link Layer Addressing **(Pages 169-251)**
Unit - V **Broadband Network:** Local Loop Technologies Asymmetric Digital Subscriber Line (ADSL) High Bit-Rate Digital Subscriber Line (HDSL), Line Coding Techniques Wireless Local Loop (WLL). Security : Basic Requirements of Network Security, Security Levels Data Security Invalid Access/ Possibility of Eavesdropping Firewalls (Access Control) Encryption Security Against Remote Access. Electronic Mail and other Internet Services : Electronic Mail E-Mail Addresses Format of A Mail Message Some Important Features of E-Mail Services Available on The Internet, Electronics Commerce and EDI, Electronic Commerce Internet: A Tool for Electronic Commerce Electronic Data Interchange (EDI), Implementing EDI - Major Considerations User Characteristics and Electronic Commerce , Issues in Electronic Commerce.	**Unit-5:** Broadband Network and Internet Services **(Pages 253-324)**

CONTENTS

INTRODUCTION

NOTES

Computers have brought about major changes in all spheres of life. Today, it is extremely difficult to imagine the world without computers. Computers help us, communicate using modems, telephone and Wi-Fi facilities and it seems as if you are communicating directly with each other. Internet links are computer networks across the world so that users can share resources and also communicate with each other. The advances in telecommunication technologies have made it possible for computers to interact with each other to provide services to all lines of business. The study of data communication and computer networks becomes essential to know more about computing techniques and communication technologies. It is now inevitable for everybody, from high-tech company professionals to elementary students, to have a good insight of the principles of data communication to become aware of how it can be used in the growth of networking, and thereafter Internet, which has influenced almost every aspect of life. Conveniences like ATM bank services, Internet, video conferencing, wireless telephony and electronic mail could not have been possible without data communication and computer networks. Communication facilities available with an organization or with an individual, measure the level of standard for them.

This book is directly linked to the various aspects of data communication and computer networking vis-à-vis related emerging trends in network-centric information technology. It attempts to provide students a framework of data communication and computer networks, fundamental concepts which have already made a very important place in the life of engineers, managers, professionals and individuals. It does not boast about making you a computer network expert or technician but promises to focus on the fundamental understanding of the various concepts involved in modern data communication and computer networks. Data communication and computer networks form are an extremely exciting field of study, and therefore, a good amount of emphasis has been put to prepare objective questions also so that the reader may enjoy the exercises in the most effective way to test their understanding of the concepts.

This book provides a good learning platform for students who need to be skilled in the area of data communication and computer networks without going into the elaborate details of computer programming. It explains the underlying concepts of data communication and computer networks so that students may visualize communication systems from the hardware level right up to the application level. This book covers most of the currently relevant areas of data communication and computer networks.

This book, *Data Communication and Computer Networking*, follows the SIM format wherein each Unit begins with an Introduction to the topic followed by an outline of the 'Objectives'. The detailed content is then presented in a simple and an organized manner, interspersed with 'Check Your Progress' questions to test the understanding of the students. A 'Summary' along with a list of 'Key Terms' and a set of 'Self-Assessment Questions and Exercises' is also provided at the end of each unit for effective recapitulation.

UNIT 1 FUNDAMENTALS OF DATA COMMUNICATION, COMMUNICATION CHANNELS AND DATA TRANSMISSION PROTOCOL

Structure

1.0 INTRODUCTION

In this unit, you will learn about the data communication, various channels for data transmission and transmission

Data Communication is a process of exchanging data or information. In case of computer networks this exchange is done between two devices over a transmission medium. This process involves a communication system which is made up of hardware and software. The hardware part involves the sender and receiver devices and the intermediate devices through which the data passes. The software part involves certain rules which specify what is to be communicated, how it is to be communicated and when. It is also called as a Protocol.

Data transmission and data reception (or, more broadly, data communication or digital communications) is the transfer and reception of data (a digital bit stream or a digitized analog signal) over a point-to-point or point-to-multipoint communication channel. Examples of such channels are copper wires, optical fibers, wireless communication channels, storage media and computer buses. The data are represented as an electromagnetic signal, such as an electrical voltage, radio wave, microwave, or infrared signal.

In data communication terminology, a transmission medium is a physical path between the transmitter and the receiver, i.e., it is the channel through which data is sent from one place to another. A communication channel that is used to carry the data in the form of bits from the sender to the receiver using LAN is known as Transmission media. Here, the transmission of data can be done using electromagnetic signals. In data communication, transmission media acts as a physical lane among transmitter & receiver. The form of bits varies based on the type of network like for copper-based network; the bits are in electrical signals form whereas, in a fiber network, the bits are in the light signals form.

1.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Understand the fundamentals of data communication
- Discuss the methods of data communication
- Explain communication channels-modems
- Interpret data transmission protocol
- Understand the modem of data transmission
- Define modem and discuss its types
- Discuss the various data transmission protocols
- Define transmission media
- Explain the types

1.2 METHODS OF DATA TRANSMISSION

Data transmission can be divided into parallel and serial data transmission.

(i) Parallel Transmission

One or more bytes of data are sent over two or more wires. Each wire transmits one digit of binary code. Therefore, sending one byte (8 bits) of data requires 8 wires as shown in Figure 1.1. In this type of transmission, it is necessary to detect where each byte of data is separated from the next. Normally, this detection is made on elapsed time base. The interface of a printer with PC is a good example for this case.

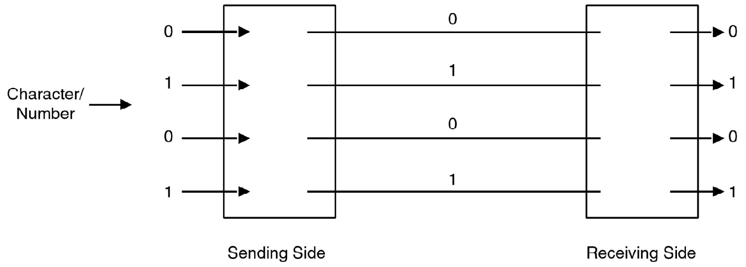


Fig. 1.1 *Parallel Transmission System*

Two key issues occur in parallel transfer. The wire itself is the first issue. Minimum of nine wires (eight for data bits, one for circuit ground) are required. Many times extra wires are needed to control the flow of data across the interface. The other issue is with the nature of the bits or voltages itself. When there is change in the state of the bit/voltage from a one to zero, or vice versa, it happens at the rate of nanoseconds (one billionth of a second). A crucial part of the data transfer is the abruptness itself. The changes that occur slowly, i.e., between zero and one are not accepted as data. The electrical properties (capacitance and inductance) of a longer cable limit the suddenness with which a bit changes from zero to one, and corruption of data or loss becomes probable. Due to this, inherent speed in parallel transfer of data creates problems while transmitting over longer cables. Hence, its usage is limited to selected peripheral devices, such as printers used in close proximity to the computer, or that operates at high speed.

(ii) Serial Transmission

Data is sent over a single wire as shown in Figure 1.2. Therefore, sending one byte does not require 8 wires. These are sent one after the other. In this transmission, it is necessary to detect where each bit is separated from the next and also where each block is separated from the next. Normally, the former is detected based on elapsed time, the latter using one of a variety of so called synchronous system that will follow next; for example, the RS–232C interface and Ethernet can be cited in this category. Serial transmission is suitable for long distance data transmission because it is less costly and more resistant to noise. Therefore, almost all transmission lines for data communication systems are serial transmission lines.

Transmitting eight individual bits one after the other involves eight times more time than transmitting them all at the same time parallely. This speed lim it does not prove to be significant for several applications. Compared to the internal speed of the microprocessors, serial peripheral mechanisms are slower. They involve long, mechanical processes which restrict their speed: the speeds of their print-heads often limit the printers; the frequency limitations of the telephone lines affect the modems and time consuming rotational speed limits the disk drives. The speed that is built within the process of parallel data transfer is a waste on such peripheral mechanisms. The serial method thus, sacrifices a part of the speed while sufficiently servicing the peripheral devices. In such situations, the sacrifice in speed is unimportant compared to the added transmission range and reliability.

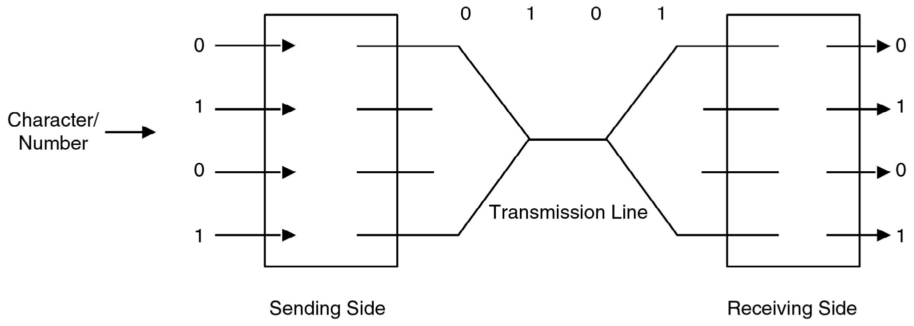


Fig. 1.2 *Serial Transmission*

1.2.1 Analog versus Digital

Data communication and networks deal with data or information transmission. Data can be represented in many ways such as a human voice, a bunch of numbers, images, text and sounds, etc. There are two ways to communicate, display, store or manipulate information, as follows:

- Analog
- Digital

In the analog form of electronic communication, information is represented as a continuous electromagnetic wave form as shown in Figure 1.3. Digital communication represents information in binary form through a series of discrete pulses as shown in Figure 1.4.

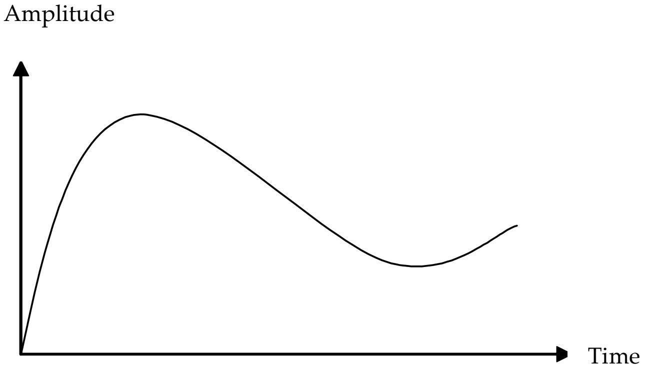


Fig. 1.3 *Representation of Analog Signals*

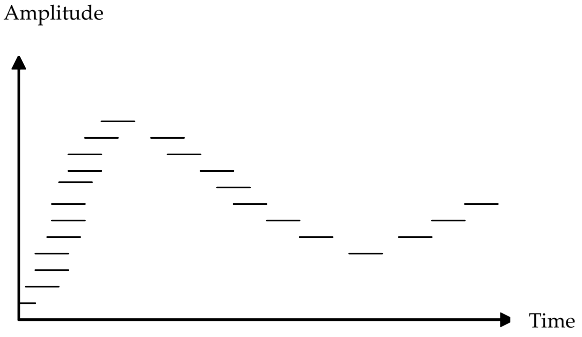


Fig. 1.4 *Representation of Digital Signals*

NOTES

Analog Signal

Analog is best explained by the transmission of such signals as human speech or sound, over an electrified copper wire. In its native form, human speech is an oscillatory disturbance in the air as shown in Figure 1.3, which varies in terms of its volume, or power (amplitude) and its pitch or tone (frequency). Analog signals are therefore defined as continuous electrical signals varying in time as shown in Figure 1.4. Analogous variations in radio or electrical waves are created in order to transmit the analog information signal for video or audio or both over a network from a transmitter (TV station or CATV source) to a receiver (TV set, computer connected with antenna). At the receiving end, an approximation (analog) of the original information is presented. Information that is analog in its native form (image and audio) can vary continuously in terms of intensity (brightness or volume) and frequency (color or tone) as shown in Figures 1.3 and 1.4. These variations in the native information stream are translated, in an analog electrical network, into variations in the frequency and amplitude of the carrier signal. In other words, the carrier signal is modulated (varied) in order to create an analog of the original information stream.

The electromagnetic sinusoidal waveform or sine wave as shown in Figure 1.5 can be varied in amplitude at a fixed frequency, using Amplitude Modulation (AM). Alternatively, the frequency of the sine wave can be varied at constant amplitude, using Frequency Modulation (FM). Additionally, both amplitude and frequency can be modulated simultaneously. Figures 1.6 and 1.7 represent a sinusoidal waveform in amplitude and frequency form. The example of analog signal in the field of data communication is telephone voice signal in which the intensity of the voice causes electric current variations. At the receiving end, the signal is reproduced in the same proportion.

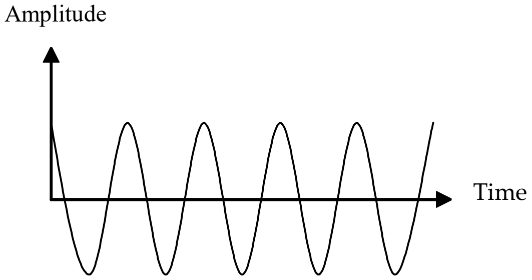


Fig. 1.5 *Waveform in the Form of Sine Wave*

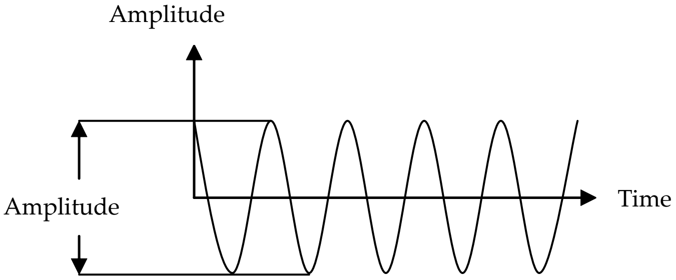


Fig. 1.6 *Amplitude*

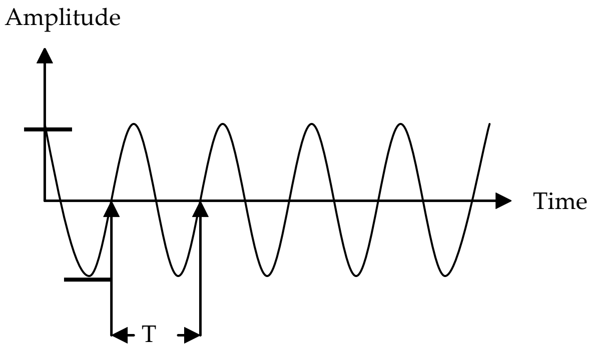


Fig. 1.7 *Frequency Representation*

Voice: A voice grade channel is approximately 4,000 Hz, or 4 kHz. Approximately 3.3 kHz (200 Hz to 3,500 Hz) is used for the voice signal itself. The remaining bandwidth is used for the purpose of network signalling and control in order to maintain separation between information channels. While human speech transmission and reception encompasses a much wider range of frequencies, 3.3 kHz is considered to be quite satisfactory and cost-effective. Band-limiting filters are used in carrier networks to constrain the amount of bandwidth provided for a voice application.

Video: A CATV video channel is approximately 6 MHz. Approximately, 4.5 MHz is used for information transmission, while the balance is used for guard bands to separate the various adjacent channels using the common, analog coaxial cable system.

Digital Signal

Computers are digital in nature. Computers communicate, store and process information in binary form, i.e., in the combination of 1s and 0s, which has specific meaning in computer language. A binary digit (bit) is an individual 1 or 0. Multiple bit streams are used in a computer network. The computer systems communicate in binary mode through variations in electrical voltage. The digital signals that are non-continuous change in individual steps consisting of digits or pulses with discrete values or levels. The value of each pulse is uniform but there is an abrupt change from one digit to the next. They have two amplitude levels, which are specified as one of two possibilities like 1 or 0, high or low, true or false and so on. In other words, the digital signalling, in an electrical network, involves a signal which varies in voltage to represent one of two discrete and well-defined states as depicted in Figure 1.8, such as either a positive (+) voltage and a null or zero (0) voltage (unipolar) or a positive (+) or a negative (–) voltage (bipolar).

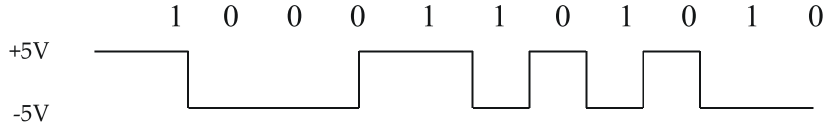


Fig. 1.8 *Binary Representation Forming Digital Signal*

1.2.2 Data Transmission

For transmission across a network, data has to be transformed into electromagnetic signals. Both, data and signals can be either of analog type or digital type. A signal is termed periodic if it has a continuously repeating pattern. Therefore, the data and signals are two essential building blocks of any computer network. Signals are the electric or electromagnetic encoding of data specifically used for data transmission. A digital signal is a composite signal with an infinite bandwidth.

Signals

Information exchange is an essential part of communication. It may be exchange of information among users or equipment in the communication system. In the communication context, signalling refers to the exchange of information between components required to provide and maintain data communication service. In case of PSTN (Public Switched Telephone Network), signalling between a telephone user and the telephone network may include dialling digits, providing dial tone, accessing a voice mailbox and sending a call-waiting tone etc. Looking at networking, perspectives, it is transmission of service information such as addresses, type of service etc., between nodes and/or terminals of a network. In other words, it is a process of exchanging and generating information between components of a telecommunications system to establish, release, or monitor connections (call handling functions) and to control related network and system operations (other functions).

Signalling System7 (SS7)

Signalling System 7 (SS7) is the protocol designed for public switched telephone system for providing services and setting up calls.The various value-added features such as providing intelligence to PSTN services come under the service of SS7. Earlier the same physical path was used for both the call-control signalling and the actual connected call. This is called in-band signalling technique. This method of signalling was inefficient and replaced by out-of-band or common-channel signalling techniques. Out-of-band signalling performs its job by utilizing two networks in one. As we know that in PSTN, our voice and data is carried over circuit-switched network. It provides a physical path between the destination and source. The other one is the signalling network, which carries the call control traffic. It is a packet-switched network using a common channel switching protocol.

Functions of SS7

- It controls the network.
- The SS7 network sets up and tears down the call.
- It handles all the routing decisions and supports all telephony services including Local Number Portability (LNP), remote network management, called ID and forwarding.

In order to accomplish the above functions, SS7 uses voice switches, which are known as Service Switching Points (SSPs). They handle the SS7 control network as well as the user circuit-switched network. Basically, the SS7 control network tells the switching office which paths to establish over the circuit-switched network. SSPs also query Service Control Point (SCP) databases using packet switches called Signal Transfer Points (STPs). The STPs route SS7 control packets across the signaling network. The concept of SSP, STP and SCP has been illustrated in Figure 1.9.

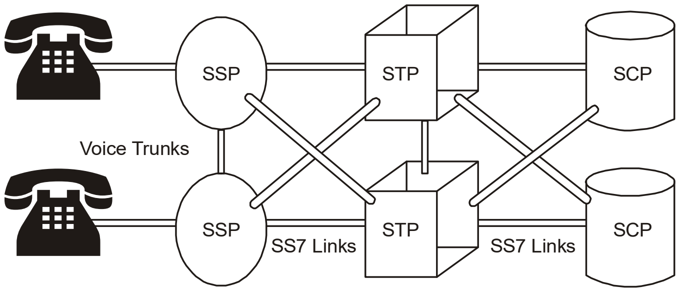


Fig. 1.9 *SS7 Signaling Points*

1.2.3 Comparison of Analog and Digital Data Transmission

Digital signals are identified through bit interval and bit rate. The bit interval is the time occupied by a single bit and the bit rate is the number of bit intervals per second which is expressed in bits per second or bps. Although analog voice and video can be converted into digital, and digital data can be converted to analog, even then, each format has its own advantages.

Advantages of Analog Transmission

The following are the advantages of analog transmission:

- Analog transmission offers advantages in the transmission of analog information. Additionally, it is more bandwidth-conservative and is widely available.
- Analog has an inherent advantage as voice, image and video are analog in nature. Therefore, the process of transmission of such information is relatively straightforward in an analog format, whereas conversion to a digital bit stream requires conversion equipment. Such equipment increase cost, are susceptible to failure, and can negatively affect the quality of the signal through the conversion process, itself.
- More bandwidth is consumed by a raw information stream in digital than in analog form. This is particularly evident in CATV transmission, where 50 or more analog channels are routinely provided over a single coaxial cable system. Without the application of compression techniques on the same cable system, only a few digital channels could be supported.
- Finally, analog transmission systems are already in place, worldwide. Interconnection of these systems is very common and all standards are well established. As the majority of network traffic is voice and as the vast majority of voice terminals are analog devices, therefore, voice communication largely depends on analog networks. Conversion to digital networks would require expensive, wholesale conversion of such terminal equipment.

Advantages of Digital Transmission

The following are the advantages of digital transmission:

- **Digital Data:** When it comes to the transmission of binary computer data, the advantage is with digital transmission. The equipment required for converting digital data to an analog format and sending the digital bit streams over an analog network can be expensive, susceptible to failure, and can create errors in the information.
- **Compression:** It is relatively easy to compress digital data, thus the efficiency of transmission increases. As a result, image, video, voice and data information can be transmitted in substantial volumes using relatively little raw bandwidth.
- **Security:** Digital systems offer better security while analog systems offer some measure of security through the scrambling of several frequencies. Scrambling is fairly simple to defeat. Digital information, on the other hand, can be encrypted to create the appearance of a single, pseudo-random bit stream. Thereby, the true meaning of, sets of bits and individual bits or the total bit stream that cannot be determined without the key that unlocks the encryption algorithm that has been employed.
- **Quality:** Digital transmission offers improved error performance (quality) as compared to analog. This is due to the devices that boost the signal at periodic intervals in the transmission system in order to overcome the effects of attenuation. Additionally, digital networks deal more effectively with noise, which is always present in transmission networks.
- **Cost:** The cost of the computer components required in transmission and digital conversion has dropped considerably. At the same time the reliability and ruggedness of those components has increased over the years.
- **Upgradability:** It is relatively simple to upgrade digital networks as the comprise of computer (digital) components. Such upgrades can enhance functionality improve error performance and increase bandwidth. Some upgrades can be effected remotely over a network, eliminating the need to dispatch expensive technicians for that purpose.

1.2.4 Commuincation Modes

From the viewpoint of transmission, communication modes can be classified into the following three types:

(i) Simplex

In this communication mode, data is always transmitted only in one direction. TV broadcasting is an example of such kind of communication mode. The information flows in one direction across the circuit, with no capability to support a response in the other direction. Simplex transmission generally involves dedicated circuits as shown in Figure 1.10. Simplex circuits are analogous to escalators, doorbells, fire alarms and security systems.

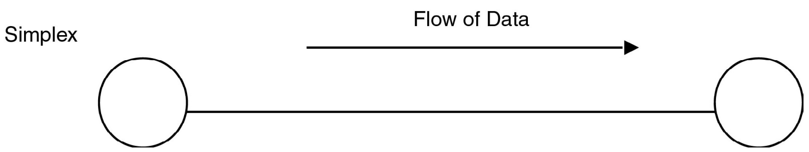


Fig. 1.10 Simplex

(ii) Half Duplex

In this mode data is transmitted in one direction at a time, for example, a walkie-talkie. This is generally used for relatively low-speed transmission, usually involving two-wire, analog circuits as shown in Figure 1.11. Due to switching of communication direction, data transmission in this mode requires more time and processes than under full duplex mode. Examples of half duplex application include line printers, polling of buffers and modem communications (many modems can support full duplex also).

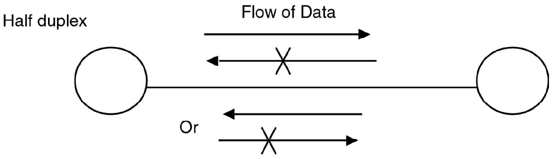


Fig. 1.11 Half Duplex

(iii) Full Duplex

In a full duplex mode data can be transmitted in both directions at the same time. In general, four wires, as shown in Figure 1.12, are required for full duplex transmission. Full duplex typically requires two simplex circuits, one operating in each direction. All wide-band and broadband circuits are full duplex in nature, as they contain most of the multichannel circuits. More typical examples of full duplex applications include channel links between host processors, channel links between controllers/concentrators and hosts, and other applications involving the interconnection of substantial computing systems. Services, such as Frame Relay, SMDS (Switched Multimegabit Data Service) and ATM (Asynchronous Transfer Mode) are based on full duplex transmission.

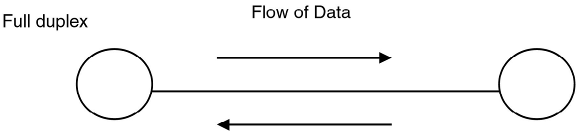


Fig. 1.12 Full Duplex

1.2.5 Synchronous System

The receiving equipment cannot detect where the transmitted data begins or ends, if data does not include any sign indicating the separation of data items. In serial transmission, sending equipment converts each of the characters into a bit string and sends them sequentially over the transmission line. To receive correct information, the receiving equipment must be able to read the value of each bit and also determine which bits are the beginning and end of each character.

For this reason, receiving equipment must synchronize with sending equipment during reception of data. Synchronization refers to correct detection by receiving equipment at the beginning and end of data that was sent from sending equipment. Systems employed for this detection are called synchronization systems. Synchronous systems can be classified into three categories:

- Asynchronous systems (start-stop).
- Character synchronous systems (SYN synchronous).
- Flag Synchronous systems.

(i) Asynchronous System

Asynchronous or character framed transmission as shown in Figure 1.13, is a method that grew out of telegraphy. From Latin and Greek, it translates as 'not together with time'. In other words, it is *not* synchronous. Asynchronous transmission is a start-stop method of transmission in which a sign bit is added to the beginning and end of each character (8 bits) in order to detect the separation of data items.

Stop Bit	1	2	3	4	5	6	7	Parity Bit	Start Bit

Fig. 1.13 *Asynchronous System*

These sign bits are called a *start bit* and *stop bit*, respectively. The start bit alerts the receiving terminal to the transmission of something worthy of its attention and a stop bit informs the receiving terminal that the transmission of that set of information be finished. Additionally, asynchronous transmission adds a parity checking bit for relatively poor error control. The framing of the data with these three or four bits of control information yields an overhead, or inefficiency, factor of 20 per cent to 30 per cent.

In almost all cases, a PC and a modem exchange data asynchronously. The length of the start and stop bit can be specified through PC communication software. Generally, a start bit is 1-bit long. The length of a stop bit can be selected as 1, 1.5 or 2 bits. Normally, the stop bit is also 1-bit long.

(ii) Character Synchronous System

Figure 1.14 shows a character synchronous (SYN synchronous) system. With this system, special characters are added to the beginning of a data block to allow detection of separation of data items. These special characters are called SYN characters. The character string of the SYN character is 00010110. Upon receipt of this character, receiving equipment determines that all succeeding data consists of data bits. It then receives each succeeding 8 bits as one character. Normally, the sending equipment sends 2 or more SYN characters before sending data to ensure synchronization with receiving equipment. The receiving equipment remains attentive to SYN characters at all times so that it can receive bit strings other than SYN characters as data. String containing 00010110 (the same as the SYN character) cannot be used to transmit data.

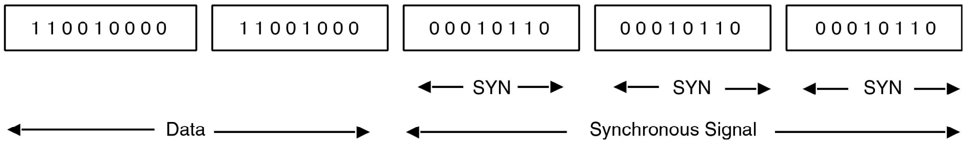


Fig. 1.14 *Character Synchronous System*

(iii) Flag Synchronous System

Figure 1.15 shows a flag synchronous system. Within this system, a special bit string is sent before and after each data block to allow detection of separation of data items. This string also continues to be sent when no data is being sent over the transmission line. This string is called a flag and consists of 01111110. The receiving equipment considers bit strings as data if they are not flags of this format.

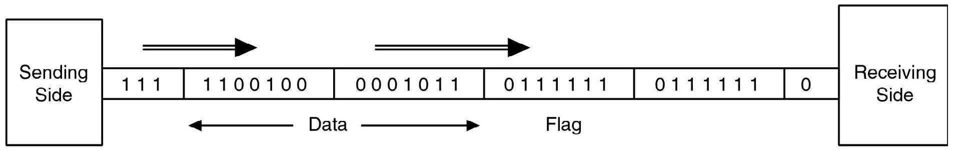


Fig. 1.15 *Flag Synchronous System*

With this system, data of a desired bit length can be sent. It may seem impossible to send the same data bit string as the flag, i.e., 01111110. A method called transparency can be used to send this flag as data.

Transparency means that receiving equipment will use the same format as sending equipment to receive any data in its original format. To send the same bit string as the 01111110 flag string as data with a flag synchronous system, a technique called zero bit insertion or bit stuffing is used. This technique is explained below:

If the sending equipment detects a bit string 11111 (5 consecutive 1s) in the data, it inserts a 0 bit at the end to send 111110. If the receiving equipment detects a bit string 111110, it deletes the 0 at the end. Although the flag 01111110 contains a bit string 11111, the sending equipment sends the flag as is without inserting a 0.

When this technique is used, the data string 01111110 is converted into 011111010. Therefore, the same bit string, cannot appear in the data as the flag.

Check Your Progress

1. What are two key issues in parallel transfer?
2. What are the ways of data representation?
3. How the data is represented in computers?
4. What is signalling?
5. Define signalling system.

1.3 COMMUNICATION CHANNELS

Channel bandwidth may be simply defined as the size of the range of frequencies that can be transmitted through a channel. In other words, we may define it as the volume of information per unit time that a computer, person, or transmission medium can handle. It is measured in Hertz (Hz). Bandwidth is expressed as data speed in bits per second (bps) in digital systems while as the difference between highest frequency to lowest frequency in analog system. Bandwidth determines how fast data flows on a given transmission path. It is determined as the amount of data transmitted or received per unit time. As it has already been explained in noise that low bandwidth signal produces less internal noise compared to high bandwidth signal; therefore this is preferred. However, in this case, we have to sacrifice data transmission speed. Therefore, a trade-off based on the performance requirements is required to be determined.

Bandwidth is dependent on the variety and physical characteristics of the transmission media, the amount of noise in the communication channel, the method of data encoding, etc.

Channel data transmission rate (bit rate)

The highest number of bits that transmits in unit time through the physical transmission media determines the channel data transmission rate. The unit of channel data transmission rate is bits per second (bps). In 1924, H. Nyquist gave the maximum rate of data of a noiseless communication channel. Further, C. Shannon extended the work of Nyquist and proposed a data rate for random noise.

Nyquist stated that if an arbitrary signal has been run through a low pass filter of bandwidth H, the filtered signal can be reconstructed by sampling the signal twice the frequency of the signal. Mathematically,

$$\textit{Maximum data rate} = 2H \, Log_2 W/Second$$

where, *W* represents the number of discrete levels in the signal

The above is a case of a noiseless channel. If random (internal) noise is present, the situation deteriorates rapidly. As we have already explained that SNR is given by a quantity 10 $\log_{10}$ S/N dB.

Therefore, Shannon stated that the maximum data rate of a noisy channel whose bandwidth is H Hz and whose signal to noise ratio S/N is given by

$$\textit{Maximum data rate} = H \log_2(1+S/N)$$

Channel capacity

Channel capacity is the amount of information passed through a communication link or transmission channel in unit time. It is measured in bits per second.

Transmission time

It is the time taken by a signal to pass over a communication link or transmission media. Transmission time is measured in seconds. It is calculated by dividing the maximum number of bits in a message by data rate. The data rate is measured in bits per second (bps). It is also given as the packet length divided by the channel capacity.

Propagation time (channel latency)

It is the time taken by a signal or information to propagate or pass from the source to the destination over a communication link or transmission media. Propagation time is deduced by dividing the length of the communication channel or distance from the source to the destination by the speed of signal propagation. The propagation speed of electromagnetic signals is normally taken as the speed of light. The characteristics of the mediums, the speed of signal propagation and transmission distance are the major factors influencing channel latency.

Throughput

Throughput may be defined as the number of bits, characters, or blocks passing through a data communication system over a period of time.

$$\text{Throughput} = \frac{\textit{Packet length in bits}}{\textit{Transmission time + Propagation on time}}$$

Channel utilization

The traction of the channels data rate that is used for the transmission of data is known as channel utilization. From the throughput it is observed that the propagation time and transmission time are two different parameters which are dependent upon the path length and packet length, respectively (number of bits in a message).

Hence, $\text{Channel Utilization} = \frac{a}{1+a}$

where, *a* is given as the ratio of propagation time and transmission time and is known as bit length.

We may now consider an example to understand the above concepts. Suppose, a channel data transmission rate is 10 Mbps and time taken by one bit to transmit through channel 10^{-7} seconds. The signal propagation speed in the medium is 2×10^{8} m/s.

The transmission rate is 1bit /10^{-7} seconds that is equal to 107 bit per second (bps).

Therefore, bit length will be equal to 2×10^{8} m/s/10^{7} bps which is equal to 20 meters.

1.4 MODEMS

The square waves or digital signals are composed of wide spectrum and are prone to attenuation in the signal strength and distortion due to different frequency components of the signal. These signal impairment effects are not suitable for baseband (DC) signalling for higher speed and long distances. They are suitable only for slow speeds and over short distances. The data communication also seeks to communicate over large distances. Hence, another technique called *AC signalling* is employed in which a continuous wave called *sine wave* is used. A sine wave is characterised by frequency, amplitude and phase. Any one of the characteristics of a sine wave is modulated in accordance with the information so that the information can be transmitted over large distances in which sine wave acts as a carrier for information.

A device that accomplishes the above function in which it accepts a series of bits in the form of 0 and 1 as input and produces a modulated carrier as output at the transmitting end and a reverse operation at the receiving end is called a *modem*. In other words, a modem is an electrical component that can connect another modem over an analog telephony network. When two modems are connected, they can send each other a two-way stream of digital bits. A computer sends information to another computer located at a remote location using modems. The modem receives digital information from the computer, translates it to an analog signal using digital-to-analog converter unit (DAC) and sends the analog information to the PSTN. On the other side, when receiving data from the network, an analog-to-digital converter (ADC) unit is being used to retrieve the data. It is important to know that DAC/ADC units are noisy units and are thus limitations on the performance of the modem.

The data communication techniques were developed based on the existing telephone network so that no extra expenditure may be incurred on infrastructure. It was the voice communication that had necessitated the communication between remote computers and computing devices using the existing telephone network for voice communication. Most of the telephone lines were installed for voice communication, and therefore, they were able to transmit only analog information. On the other hand, the computers and related computing devices were based on digital signal in the form of pulses or 0 and 1. Therefore, to use the existing telephone lines or the analog medium, a device that may convert digital signal into analog and vice versa was needed. This device is known as the modem and stands for modulator demodulator. It performs the function of modulating and demodulating a signal.

A modem, therefore, receives serial binary data as its input. It modulates some of the characteristics of a sine wave like amplitude, frequency or phase generated by it in accordance with the input signal so that the binary signal may be transmitted over long distances. A reverse procedure takes place at the receiving end where the received signal is demodulated to retrieve the binary signal as the output of the modem which can be inputted to the digital device at the receiving end for further processing. In other words, the modem changes the analog information into digital pulses at the computer or the digital device at the receiving side of the communication link or channel.

Conventionally, modems were devised for communication between a host computer and data terminals. Subsequently, they were also deployed to communicate between remote computers and computing devices. As they were used to communicate between remote digital devices, their data transmission rates were also subjected to increase from 300 bps to 28.8 Kbps. The modem technologies were also upgraded to involve data compression techniques. However, they increased the additional burden of error detection and error correction to maintain reliability.

Therefore, the modem can be considered as a peripheral device for computers to enable two remote computers to communicate over standard telephone lines. Modems are developed in different shapes and sizes for various types of applications and needs. The word modem stands for modulator/

demodulator and performs the conversion of digital signals to analog signals (modulation) and vice versa (demodulation) as shown in Figure 1.16.

In order to establish interoperability among different types of modems from different manufacturers, standards for modem interface were developed. Modems are deployed to perform various types of functions. Some of them are used in voice and text mail systems, facsimiles, etc., and others are attached or assimilated into mobile phones or laptops making data transmission possible from any location to any other. In future, modems may be utilized for other types of applications. Modem speeds are still around 28/56 Kbps and further increase in speed will be possible only on digital phone technology, like ISDN and fibre optic lines. Some of the new applications are videophones in which simultaneous communication of voice and data are performed.

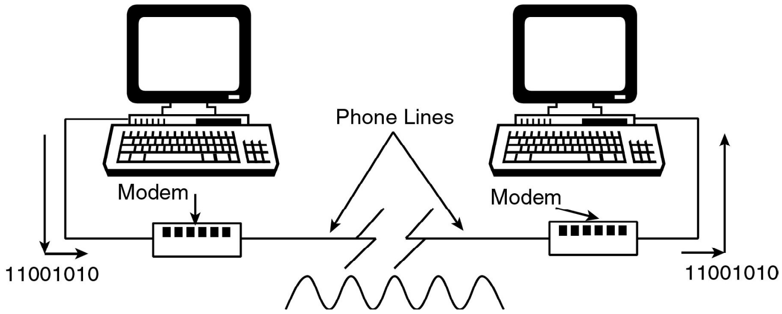


Fig. 1.16 *Connecting Two Computers via Modems*

Modems continuously generate a carrier signal to send information so that the information may be delivered from one location to another remote location. The information to be transmitted is superimposed on the carrier signal. In this manner, the transmitted information varies or modulates this carrier signal. The terms *baud* and *bps* used to measure the data rate are very popular with this technology and are continually used interchangeably. However, they are not the same at all.

The number of pulses transmitted in a second characterizes the carrier signal in which each pulse is called a baud. The bps stands for bits per second and indicates the number of bits that can be transmitted during one pulse (one baud). Similarly, kbps stands for kilo bits per second.

Therefore, bps = baud × number of bits per baud.

The baud and bps often create confusion because early modems were based on 1 bit per baud and used to transmit only 1 bit per baud. In such a case, for example, a 2400 baud modem will also transmit 2400 bps. However, because of the need of higher speeds, modems are designed to have more number of bits per baud.

The difference between baud and bps can be understood from this analogy. Bit rate means the number of bits (0 or 1) transmitted during one second of time. The number of changes in signal per unit of time to represent the bits is called the modem's data rate. This rate is expressed in terms of *baud*. A signal unit may have one or more than one bits. Therefore, baud signifies the number of times per second the line condition can switch from 1 to 0. Baud rate and bit rate, which are expressed in bits per second, are not similar, as number of bits may be transmitted by the modem through the channel in each signal change (some bits can be send as one symbol). The relation between bit rate and baud is that bit rate is equal to baud rate multiplied by the number of bits representing each signal unit. Bit rate is always more than or equal to baud rate because baud rate determines the bandwidth required to transmit the signal. The signal may be in the form of pieces or block that may contain bits. A fewer bandwidth is required to move these signal units with large bits for an efficient system. To understand the relation between bit and baud rate, we consider an analogy of car, passengers and highway with signal units, bits and bandwidth respectively.

A car has a capacity of carrying a maximum of five passengers at a time. Suppose a highway may support only 1000 cars per unit time without congestion, when each car on the highway carries five passengers, it is considered that the highway is capable of providing services without congestion. Thus, it is thought that the highway provides an excellent service. In another scenario, when all these 5000 passengers wish to go in separate cars, they require 5000 cars while the highway can only support 1000 cars at a time. The services offered get deteriorated because the highway's capacity is meant only for 1000 cars. It does not bother as to whether these 1000 cars are carrying 1000 passengers or 5000 passengers or more. To support more cars, the highway needs to be widened. Similarly, the number of bauds determines the bandwidth.

1.4.1 Classification of Modems

Modems are classified based on these characteristics:

- *Range:* Short Haul, Voice Grade (VG), Wide band.
- *Line Type:* They use dial-up, leased or private circuits.
- *Operation Mode:* Half Duplex, Full Duplex, Simplex are based on the direction of flow of information.
- *Synchronization:* Asynchronous, Synchronous.
- *Modulation:* Based on modulation techniques such as AM, FM/FSK, PM.
- *Transmission Media:* Radio, Optical, Dial-up.

1.4.2 Modems Based on Range

(a) Short Haul

Short haul modems are widely deployed over private lines and are not part of a public system. These are economical solutions to systems of short ranges up to 15 km. Short haul modems can also be used on an end-to-end length of the direct connection longer than 15 km, when both ends of the line called local loops are served by the same exchange in the telephone system. They are distance-sensitive, because signal attenuation happens as the signal travels through the line. The transmission rate must be reduced to ensure consistent and error-free transmission on longer distances.

There are two main types of short haul modems:

Analog Modems: They use a simple modulation method. Sophisticated devices for error control or equalizers are not employed. They operate at a maximum rate of 9600 bps, but there are some, which support higher rates up to 64,000 bps.

Line drivers: They increase the strength of the digital signal. Unlike conventional modems, they do not transmit carrier signals to the communication channel. Line drivers are inexpensive, tiny and do not have a power supply. Power supply to the line driver is provided through the RS232 connector of the terminal.

(b) Voice Grade (VG)

Voice grade modems have a frequency range of moderate to high data rate. They have no limitation about distance and are therefore used for long distances. These modems are expensive and involve expertise in their maintenance and tuning.

Communication channels are leased lines and dial-up and uses telephone network for data transmission on a dedicated or dialed connection.

Wideband

Wideband modems find their use in large-volume telephone line multiplexing dedicated for computer-to-computer links. They provide high data rates.

1.4.3 Modems Based on Line

(a) Leased Line

Leased, private or dedicated lines comprising four wires are for the exclusive use of leased line modems. It has two pairs out of which it uses either pair for a simple point-to-point connection or several connection on a multi-drop network for polling or a contention system. In a telephone network, their transmission characteristics generally ensure to provide certain specifications. In another case, if the link includes any radio transmission, its quality may be as variable as that of a switched line.

(b) Dial-up

Dial-up modems are used for point-to-point connections on the PSTN by any combination of manual or automatic dialling or answering. The quality of the circuit may vary from carriers to carriers.

It uses two-wire links. A four-wire line can be used as a pair of two-wire lines for transmitting and receiving. In this case, the signals in the two directions are kept totally separate to avoid interference.

1.4.4 Modems Based on Operation Mode

(a) Half Duplex

As its name indicates, the signal can be passed in either direction, but not in both simultaneously. The concept of half duplex is explained in Unit 1.

Echo-suppressors are available in a telephone network which allow transmission in only one direction. Because of this reason, the channel acts as half duplex. Echo suppressors are slowly being replaced by echo cancelers, which are theoretically full-duplex devices. When a modem is connected to a two-wire line, its output impedance cannot be matched exactly to the input impedance of the line which causes some of its transmitted signal to always reflect back. For this reason half- duplex receiver are disabled when their local transmitter is operative. Half-duplex modems can work in full-duplex mode.

(b) Full Duplex

Full duplex means that signals can be passed in either direction, simultaneously, as explained earlier.

Full duplex requires two-wire lines where two simultaneous signals in opposite directions flow. This necessitates the ability of circuits to separate a receive signal from the reflection of the transmitted signal. This is accomplished by either FDM in which the signals in the two directions occupy different frequency bands and are separated by filtering, or by echo canceling. In full duplex, modems may provide full data rate in both directions or reduced data rate in either direction. Modems that provide a low-speed reverse channel are sometimes called split-speed or asymmetric modems.

(c) Simplex

Simplex transmission allows transmission in one direction only. A remote modem for a telemetering system can be considered as an example of simplex transmission.

Echo Suppressor and Echo Canceler

A 2-wire circuit and the trunk, which is a 4-wire circuit, forms a local loop. At the junction, echoes occur and the person speaking on the telephone hears his own words after a short delay. Echo suppressors are used to eliminate the problem of echoes. They are installed on lines longer than 2000 km. In case of short lines the effect of echoes are too fast to detect them. An echo suppressor differentiates between human speech coming from one end and echoes produced at junction because of impedance mismatch of the connection. It then suppresses the echoes going in the reverse direction. The device compares the levels at its two input ports. It inserts an attenuator in the return path of the talking end to suppress echo and vice versa.

Echo suppressors improve the quality of telephone lines. It however, prevents full-duplex data transmission, which would otherwise be possible, even over a 2-wire local loop. Sometimes, in a 2-wire local loop, a part of the bandwidth is allocated in the forward direction and other in the reverse direction. It also results in delay because of switching time in either direction even if the half-duplex transmission is adequate. Moreover, they are designed to reverse upon detecting human speech and not digital data.

In case of data, if it does not detect a specific tone, it shuts down and remains as long as the carrier is present. Echo suppressors are slowly being replaced by echo cancelers, which allow a certain amount of double-talking and do not require capture time for any one talker to assume control of the connection.

1.4.5 Modems Based on Synchronization

(a) Asynchronous Modems

Most modems that operate in slow and moderate rates, up to 1800 bps, are asynchronous. Asynchronous modems operate in FSK (Frequency Shift Keying) modulation. Two frequencies for transmission and another two for receiving are

NOTES

used. Asynchronous modems can be connected with different modes in a communication media. They may use:

- 2-wire or 4-wire interface
- Switched lines or leased lines
- Interface to call unit/automatic answer, when dialing-up

(b) Synchronous Modems

Synchronous modems operate in the audio range, at rates up to 28.8 kbps in audio lines. They are used in telephones systems. The usual modulation methods are phase modulation and mix of phase and amplitude modulation at rates higher than 4800 bps. Equalizers are used in synchronous modems to offset the mismatch of telephone lines.

These equalizers can be classified into three main groups:

Fixed/Statistical equalizers: They offset the signal according to the average of the known attenuation in each frequency. They are used to operate at low rates in a dial-up line.

Manually adjusted equalizers: They are tuned to optimal performance for a given line. They need to be tuned frequently when the line is of a low quality and changes its parameters frequently.

Automatic equalizers: They are tuned automatically depending on the line quality when the connection is established.

The operation of synchronous modems is similar to that of asynchronous modems.

1.4.6 Modems Based on Transmission Medium

In addition to dedicated wires, modems are also used with other media, including RF transmission, glass fibres and conventional telephone connections. Therefore, modems can also be classified based on the medium used. For example, RF transmission and glass fibres. Basically, three types of modems are in use:

(a) Radio Modems

These can be used to send data across a pair of glass fibres using a radio frequency signal.

(b) Optical Modems

These can be used to send data across a pair of glass fibres using light. Such modems use an entirely different technology than modems that operate over dedicated wires.

(c) Dial-up Modems

Dial-up modems contain a circuitry that mimics that of a telephone. That is, modems can simulate lifting the handset, dialing, or hanging up the telephone. Second, a telephone system is designed to carry sound, a dialup modem uses a carrier that is an audible tone. Third, although they send all the data through a single voice channel, a pair of dial-up modems offer full duplex communication. That is, a single telephone connection between two dial-up modems usually allows data to flow in both directions.

(d) Cable Modems

It is a modem meant to operate over cable TV lines and to facilitate your desktop PC to connect with the Internet. Cable modems are devices that enable high-speed access to the Internet via a cable television network. While similar in some respects to a traditional analog modem, a cable modem is significantly more powerful and capable of delivering data about 500 times faster.

The coaxial cable used in cable TV has greater bandwidth than a telephone line. This allows it to achieve extremely fast access to the Internet. Another advantage is that millions of homes already have a connection for cable TV. It has tried to overcome many technical difficulties. In fact, the coaxial cable used in cable TV is simplex in nature while an Internet connection requires a duplex type connection where data needs to flow from the client to the server. Cable modems which can offer speeds up to 2 mbps are available in the market.

A cable modem needs two connections. One connection goes to the cable wall outlet and the other to a PC or to a set-top box for a TV set. Although a cable modem does modulation between analog and digital signals, it is a much more complex device than a telephone modem. It can be an external device or integrated within a computer or set-top box. Typically, the cable modem attaches a standard 10BASE-T Ethernet card in the computer.

Figure 1.17 explains the general concept of data communication between two computers. Communication between two computers is accomplished with the help of modems. In the figure each computer is connected with a modem that converts digital signal output of the computer to analog output so that it may be transmitted through transmission media that may be wired or wireless. At the receiving end the modem converts analog signal output of the transmission media into digital input so that the computer could understand it.

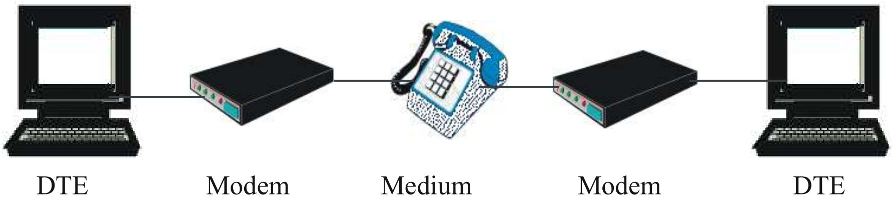


Fig. 1.17 *Data Communication between Two Computers*

1.4.7 Amplitude Modulation

Amplitude Modulation (AM) refers to the modulation of the amplitude of the carrier as analog sine wave. It occurs when a signal to be modulated is applied to a carrier frequency. The carrier frequency may be a radio wave or light wave. The amplitude of carrier wave changes according to the amplitude of the modulating signal and the frequency of carrier remains unchanged. Basically, the AM signal represents a sum of three sin waves of different frequencies. These are $f_c - f_m$, f_c, $f_c + f_m$. The sin wave with frequency f_c possesses the same amplitude as the un-modulated carrier. The other two waves of equal amplitudes and different frequencies as $f_c - f_m$, $f_c + f_m$ are called as lower and upper side band respectively. The amplitude of the lower band and upper band, which are equal, are proportional to the amplitude of the modulating signal. It is clear from the above that the bandwidth is equal to $2f_m$, which is deduced as the difference between the upper band and lower band frequencies. Thus, the AM signal consists of the carrier signal, plus upper and lower side bands. This is known as Double Side Band - Amplitude Modulation (DSB-AM), or more commonly referred to as AM. This is shown in Fig. 1.18. As the information to be transmitted is contained in the sidebands, hence the carrier frequency is transmitted at a relatively low level to avoid the additional circuitry at the receiving end to generate the carrier frequency for demodulation. This type of transmission is known as Double Side Band - Suppressed Carrier (DSB-SC).

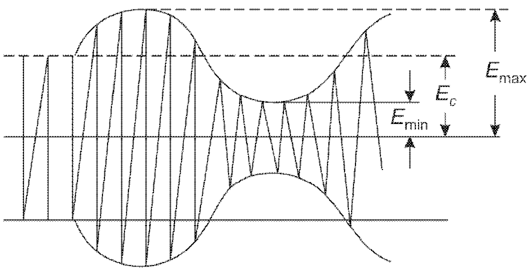


Fig. 1.18 *Amplitude Modulation of a Carrier Wave*

It is also possible to transmit a single side band. The advantage is a reduction in analog bandwidth needed to transmit the signal. This type of modulation is known as Single Side Band - Suppressed Carrier (SSB-SC) and is ideal for Frequency Division Multiplexing (FDM). Another type of analog modulation is known as Vestigial Side Band modulation. This is almost like Single Side band, except that the carrier frequency is preserved and one of the side bands is eliminated through filtering. Vestigial Side band transmission is usually found in television broadcasting. Amplitude modulation is rarely used individually as it is highly sensitive to the impacts of attenuation and line noise. The modulating index is given as:

$$m = E_{\max} - E_c/E_c$$

The above Equation, we may derive the following equation for modulating index m:

$$m = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}}$$

Angle Modulation

In angle modulation, carrier is being reproduced as follows:

$$f_c = B \sin \acute{E}_c t + \varphi_2$$

In the equation, there is an argument of sin as $\omega_c t + \varphi_2$ which can be varied in accordance with equation 1 and thus producing either frequency or phase modulation. In either case, the amplitude of the carrier remains unchanged with incremental change in $\omega_c t + \varphi_2$.

Frequency Modulation

Frequency Modulation refers to the modulation of the frequency of the analog sine wave as shown in Fig. 1.19 where the instantaneous frequency of the carrier is varied in proportion of the variation of the modulated carrier with respect to the frequency of the instantaneous amplitude of the modulating signal. It may be said in a simple word that it occurs when the frequency of a carrier is varied in proportion to the amplitude of input signal. Unlike AM, the frequency of the carrier signal is varied. Frequency variations are more immune to noise and therefore FM modulation is considered more immune to noise than AM. This leads to the overall improvement in signal-to-noise ratio of the communications system. As in FM only frequency is varied, therefore the amplitude of the modulated signal remains constant, which gives rise to constant power. However, the frequency-modulated system imposes a constraint in terms of bandwidth requirement. They require a greater bandwidth to transmit FM signal than the AM. The modulating index for FM is given as below:

$$^{2} = f_p/f_m, \text{ where}$$

^{2} = Modulation index, f_m = frequency of the modulating signal and f_p = peak frequency deviation

From the Fig. 1.19, it is inferred that the amplitude of the modulated signal always remains constant, irrespective of frequency and amplitude of modulating signal. It means that the modulating signal adds no power to the carrier in frequency modulation unlike to amplitude modulation. FM produces an infinite number of side bands spaced by the modulation frequency, f_m that is not in case of AM. Therefore, AM considered a linear process whereas FM as a nonlinear process. It is necessary to transmit all side bands to reproduce a distortion free signal. Ideally, the bandwidth of the modulated signal is infinite in this case. In general the determination of the frequency content of an FM waveform is complicated, but when ^{2} is small, the bandwidth of the FM signal is $2f_m$. On the other hand when ^{2} is large, the bandwidth is determined (empirically) as $2f_m\,(1 + ^{2})$.

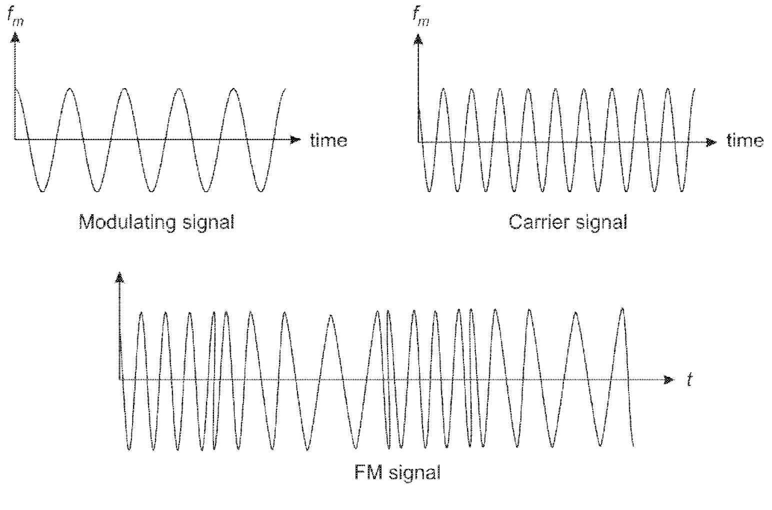


Fig. 1.19 *Frequency Modulation*

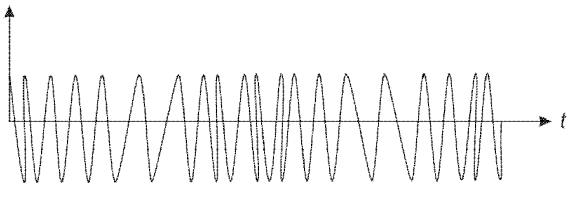


Fig. 1.20 *Phase Modulation*

Phase Modulation

Phase Modulation (PM) is like frequency modulation. Unlike, variation in the frequency of the carrier wave as in FM, the phase of the carrier wave is varied. In PM, the instantaneous amplitude of the modulating signal varies the phase of the carrier proportionately. Modulating index for PM is given as ² = “φ, where “φ is the peak phase deviation in radians. As in the case of angular modulation argument of sinusoidal is varied and therefore we will have the same resultant signal properties for frequency and phase modulation. A distinction in this case can be made only by direct comparison of the signal with the modulating signal wave, Figure 1.20.

Phase modulation and frequency modulation are interchangeable. Phase modulation is obtained by selecting the frequency response of the modulator so that its output voltage will be proportional to integration of the modulating signal. When it is differentiated, it gives frequency modulation. Bandwidth and power issues are same as that of the frequency modulation.

1.5 DATA TRANSMISSION PROCTOCOL

In telecommunication technology, *communication protocol* is defined as the characteristic set of standard norms and rules used for connection, communication, data representation, data transfer, signalling, authentication and error detection that help in sending data or information through a specific communication channel. Basically, it follows standard rules so that the network systems work properly. These protocol rules govern the syntax, semantics and synchronization of communication and are implemented by hardware, software, or a combination of both. It also defines the working behaviour of a hardware connection.

Nowadays, telecommunication technology has *connectionless* networking system to communicate between the endpoints of networks for sending messages from one end point to another. The device configured at one end of the network transmits data to any connected recipient even without confirming that the recipient is there to receive the data. Problems may be encountered during transmission and the data may have to be sent several times. This is not the case with connection-oriented protocols. The network administrators avoid the use of connectionless protocols because it is not easy to filter malicious packets using a firewall. While TCP/IP is a connection-oriented protocol, Internet Protocol (IP) and User Datagram Protocol (UDP) are connectionless protocols.

A wide range of communication protocols were defined by authentic and standard organizations throughout the world. The most common and well-known protocol suite is TCP/IP, which is termed as the base of Internetworking communications. The IP exchanges information between routers and helps the routers to select the proper path for network traffic, whereas the TCP ensures that the data packets are smoothly and reliably transmitted across the network without any error. LAN and WAN are considered as critical protocols in network communications. The LAN protocols are authentic for the physical and data link layers of communication over other specified LAN media, i.e., Ethernet wires and wireless radio waves. The WAN protocol is authentic for the three lowest layers of the OSI (Open System Interconnection) model and helps in communication with the help of other wide-area media, i.e., fibre optic and copper cables. The OSI model protocols for data communication perform the communication functions over one or more layers of the seven layers of the OSI model.

Protocols can be grouped into different suites according to their technical functions. A protocol can define one or multiple protocol suites; for example, the Gigabit Ethernet protocol IEEE 802.3z which is basically a LAN protocol is also used for MAN communications.

The following are the characteristics of communication protocols:

- They help detect the physical connection (connection or connectionless) and the existence of the end-points or nodes.
- They negotiate with different connected nodes.
- They check for how to start and end a message.
- They help in formatting a message.
- They identify corrupted messages and help in error correction.
- They terminate the session or connection.

Common communication protocols

The following are the common communication protocols:

- IP (Internet Protocol)
- UDP (User Datagram Protocol)
- TCP (Transmission Control Protocol)
- DHCP (Dynamic Host Configuration Protocol)
- HTTP (Hypertext Transfer Protocol)
- FTP (File Transfer Protocol)
- Telnet (Telnet Remote Protocol)
- SSH (Secure Shell Remote Protocol)
- SMTP (Simple Mail Transfer Protocol)
- IMAP (Internet Message Access Protocol)

Types of network communication protocols

The network communication protocols also have standard sets of rules that govern the communication process between computers which are connected to or defined on a network. These are in the form of basic guidelines which help to regulate the access method, physical topologies, types of cabling and speed of data transfer on a network. The following are the common network protocols:

- Ethernet
- Local Talk
- Token Ring
- FDDI
- ATM

1.5.1 Protocols: Overview of Networking, Role of Computer Networks in Development

In the mainframe and minicomputer environment, each user is connected to the main system through a dumb terminal that is unable to perform any of its own processing tasks. In this computing environment, processing and memory are centralized. However, this type of computerization has its merits but the major disadvantage is that the system could get easily overloaded as the number of users, and consequently, terminals, increase. Second, most of the information is centralized to one group of people, the systems professionals, rather than the end-users. This type of centralized processing system differs from the distributed processing system used by LANs. In a distributed processing system, most of the processing is done in the memory of individual PCs or workstations besides sharing expensive computer resources like software, disk files, printers and plotters, etc.

There may arise a question as to why PCs cannot be connected together in a point-to-point manner. The point-to-point scheme provides separate communication channels for each pair of computers. When more than two computers need to communicate with one another, the number of connections grow quickly as the number of computers increase. Figure 1.21 illustrates that two computers need only one connection, three computers need three connections and four computers need six connections.

Figure 1.21 also illustrates that the total number of connections grow more rapidly than the total number of computers. Mathematically, the number of connections needed for N computers is proportional to the square of N:

Point-to-point connections required = $(N^2–N)/2$

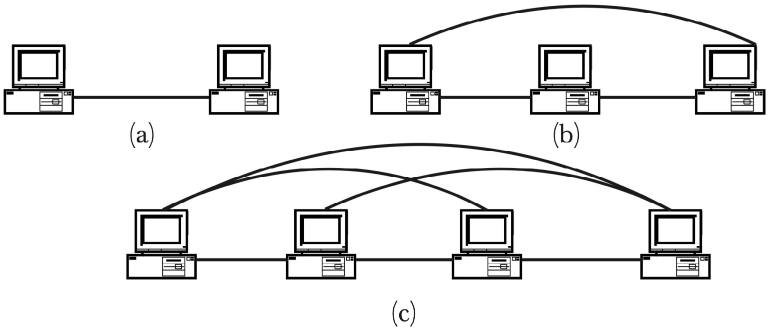


Fig. 1.21 (a), (b), (c) *Number of Connections for 2, 3, 4 Computers, Respectively*

Adding the N^{th} computer requires N–1 new connections, which becomes a very expensive option. Moreover, many connections may follow the same physical path. Figure 1.22 shows a point-to-point connection for five computers located at two different locations, say, ground and first floor of a building.

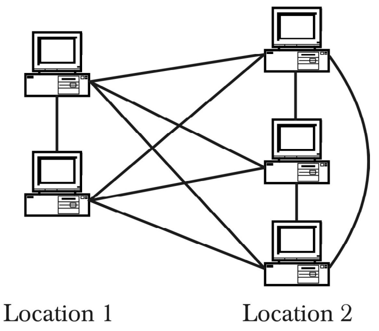


Fig. 1.22 *Five PCs at Two Different Locations*

As there are five PCs, therefore, ten connections will be required for point-to-point connection. Out of these ten connections, six pass through the same location, thereby making point-to-point connection an expensive one. By increasing one PC in the above configuration, at location 2 as shown in Figure 1.22, the total number of connections will increase to fifteen. Out of these connections, eight connections will pass through the same area.

Definition

Privately owned networks offer consistent, fast paced communication channels which are optimized to connect information processing tools in a restricted geographical area. These are known as Local Area Networks (LANs).

A shared, local (restricted-distance) packet network for computer communication is a form of LAN. A common medium is used by LAN to link peripherals and computers so that the user can share access to databases, files, host computers, peripherals and applications.

LANs, in addition to linking the computer equipment available in a particular premises, also provides a connection to other networks either through a computer, which is attached to both networks, or through a dedicated device called a *gateway*. The main users of LANs include business organizations, research and development groups in science and engineering, industry and educational institution. The electronic or paperless office concept is possible with LANs.

LANs offer raw bandwidth of 1 Mbps to 100 Mbps or more, although actual throughput often is much less. LANs are limited to a maximum distance of only a few miles or kilometers, although they may be extended through the use of bridges, routers, and other devices. Data is transmitted in packet format, with packet sizes ranging up to 1500 bytes and more. Mostly, IEEE develops LAN specifications, although ANSI and other standard bodies are also involved.

Check Your Progress

6. What are three categories of synchronous systems?
7. Define channel bandwidth.
8. What is channel capacity?
9. What is a modem?
10. List the various common network protocols.
11. Define LAN.

1.6 TRANSMISSION MEDIA

The data signal travels through this medium. There are two general categories—bounded (guided) and unbounded (unguided) media. Twisted pair, coaxial cable and fibre optic cables are all bounded media. Data signals travel within the boundaries of the transmission media. On the other hand, microwave and satellite transmissions, both travel through the air, which has no boundaries, hence called un-bounded transmission.

Guided Transmission Media

Bounded media or wired transmission systems employ physical media, which are tangible. Also known as conducted systems, wired media generally employ metallic or glass conductors which serve to conduct, some form of electromagnetic energy. For example, twisted pair and coaxial cable systems conduct electrical energy, employing a copper medium. Fibre optic systems conduct light or optical energy, generally using a glass conductor. The term bounded or guided media means that the signal is contained within an enclosed physical path. It also refers to the fact that some form of insulation, cladding and shield is used to bind the signal within the core medium. This improves the signal strength over a distance and in the process enhances the performance of the transmission system. Fibre and coaxial optical cable and twisted pair (both shielded and unshielded), systems fall into this category.

Twisted Pair Wires

Figure 1.23 shows a pair of copper wires twisted together and wrapped with a plastic coating as a twisted pair and which has a diameter of 0.4–0.8.The error rate of transmission and the electrical noise is reduced by the twisting. Each conductor is separately insulated by some low-smoke and fire-retardant substance. Teflon(r) flouropolymer resin, polyvinyl chloride and polyethylene are some of the substances used for insulation purposes.

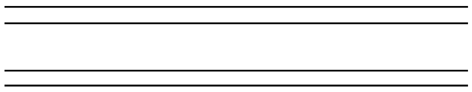


Fig. 1.23 *Two Wires Open Lines*

The twisting process serves to improve the performance of the medium by containing the electromagnetic field within the pair. Thereby, the radiation of electromagnetic energy is reduced and the strength of the signal within the wire is improved over a distance. This reduction of radiated energy also serves to minimize the impact on adjacent pairs in a multiple cable configuration. This is especially important in high-bandwidth applications as higher frequency signals tend to lose power more rapidly over a distance. Additionally, the radiated electromagnetic field tends to be greater at higher frequencies, impacting adjacent pairs to a greater extent. Generally, more twists per foot means a better performance of the wire. These are popular for telephone networks. The energy flow is in guided media. For the last eight decades, until satellite and microwave radio communications were developed, telecommunications exclusively used metallic wires. The copper wire has developed into an established technology which is strong and cost-effective. In certain applications, nickel and/(or) even aluminum metallic gold plated copper, copper alloy and copper covered steel, conductors are employed.

The maximum transmission speed is limited in this case. The copper conductor that carries analog data can be used to carry digital data also in association with modems. A modem is a device that changes analog signals into digital signals and vice versa. In this category, data rate is restricted to approximately 28 Kbps.The use of better modulation and coding schemes led to the introduction of Integrated Services Digital Network (ISDN) along with an increased data rate of 128 Kbps. Local Area Networks (LANs) also use twisted pairs. These networks were also upgraded to support high bit rate real time multimedia. In Asymmetric Digital Subscriber Lines (ADSL) technology, a new technique was introduced which intended to use two copper loops at a data rate of 1.544 Mbps. This data rate is developed as per the user direction in the network and data rates upto 600 Kbps from the user to the network.

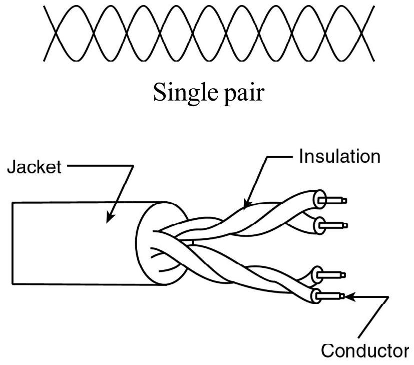
There are two categories of twisted pair cables—with and without shielding.

In Figure 1.24, an Unshielded Twisted Pair (UTP) is shown as a copper medium which was first used in telephone systems by Alexander Graham Bell and is now being utilized more and more for transmitting data. It is being frequently used for horizontal wiring. It states the link between the end in the communication closet and the outlet which is further restricted to 90 metres. A communication closet is universal to every application working over the media and is independent of the type of media.

The suggested connectors and media for horizontal wiring are discussed as follows:

- 150 Ohms Shielded Twisted Pair (STP) contains 2 pairs (IBM connector or RJ45).
- 100 Ohm UTP contains 4 pairs and 8-pin modular connector (ISDN).
- 62.5/125 contains multi-mode fibre.
- 50 Ohm coaxial (thin)-IEEE10BASE2, standard BNC connector.

A UTP cable contains 2 to 4200 twisted pairs. Flexibility, cost-effective media and usability of both data communication and voice are the biggest advantages of UTP. On the other hand, the major disadvantage of UTP is the fact that the bandwidth is limited. This limits long distance transmission with low error rates.



Single pair

Fig. 1.24 *Unshielded Twisted Pair (UTP)*

Shielded Copper or STP

UTP and Shielded Twisted Pair (STP) differ from each other in the metallic shield or screen which surrounds the pairs, which may or may not be twisted. As illustrated in Figure 1.25, the pairs can be individually shielded. A single shield can surround a cable containing multiple pairs or both techniques can be employed in tandem. The shield itself is made of copper, aluminium or steel. The shield which is electrically grounded, is in the form of a woven meshe or a metallic foil. Although less effective, the shield sometimes is in the form of nickel and/(or) gold plating of the individual conductors.

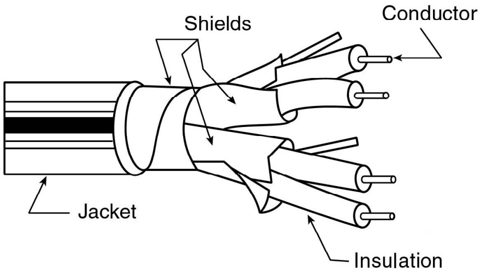


Fig. 1.25 *Shielded Twisted Pair (STP) Configuration*

The advantage of shielded copper is that performance is enhanced because both electromagnetic interference and emissions are reduced. If emissions are reduced, then the electromagnetic field is confined within the conductor. This maintains the signal strength. In other words, signal loss is reduced. Moreover, a reduction in emissions ensures that high-frequency signals do not interfere with adjacent cables or pairs. The shielding process ensures immunity from interference as it reflects the electromagnetic noise from such outside sources as radio systems, wires, cables and electric motors.

Juxtaposed with shielded copper, the shielded twisted pair has many disadvantages. Since the raw cost of acquisition is greater it is more expensive to produce the medium. Moreover, the shield's additional weight makes it difficult to deploy. Therefore, the cost of deployment increases even further. Even the shield's electrical grounding requires more effort and time.

General Properties of Twisted Pair Cables

The following are the general properties of twisted pair cables:

Gauge: It is a measure of the thickness of the conductor. A medium performs better if the wire is thick. This is because thick wires offer less resistance which in turn ensures a strong signal over a given distance. Thicker wires also have the advantage of greater break strength. The gauge numbers are retrogressive. In other words, the larger the number, the smaller the conductor.

Configuration: In a single pair configuration, the pair of wires is enclosed in a jacket or sheath, made of teflon, polyvinyl chloride or polyethylene. Usually, multiple pairs are so bundled as to minimize deployment costs associated with connecting multiple devices (for example, modems, data terminals and KTS or Key Telephone System telephone sets or electronic PBX or Private Branch eXchange) at a single workstation.

Bandwidth: The effective capacity of twisted pair cable depends on several factors, including the spacing of the amplifiers (repeaters), the length of the circuit, and the gauge of the conductor. You must also recognize that a high-bandwidth (high frequency) application may cause interference with other signals on other pairs in close proximity.

Error Performance: Signal quality is invariably important, more so in relation to data transmission. Twisted pair is susceptible to the impacts of outside interference. A wire that is lightly insulated acts as an antennae. It therefore absorbs errant signals. It follows then, that twisted pairs are susceptible to the impact of outside interference. Potential sources of Electromagnetic Interference (EMI) include electric motors, radio transmissions and fluorescent light boxes. As transmission frequency increases, the error performance of copper degrades significantly with signal attenuation increasing approximately as the square root of frequency.

Distance: Twisted pair is distance-limited. As the distance between network elements increases, attenuation (signal loss) increases and quality decreases at a given frequency. As bandwidth increases, the carrier frequency increases, attenuation becomes more of an issue, and amplifiers (repeaters) must be spaced more closely.

Security: Insecurity is an inherent feature of twisted pair. Placing taps on UTP is a simple exercise. Moreover, the radiated energy is easily intercepted through the use of inductive coils or antennae, without the requirement for placement of a physical tap.

Cost: The rearrangement, deployment and acquisition costs of UTP are very low, at least in inside wire applications. In high-capacity, long-distance applications, such as interoffice trunking, however, the relative cost is very high, due to the requirements for trenching or boring, conduit placement, and splicing of large, multipair cables. Additionally, there are finite limits to the capacity and other performance characteristics of UTP, regardless of the inventiveness of technologists. Hence, the popularity of alternatives such as microwave and fibre-optic cable.

Applications: UTP’s low cost performance has increased its application in short-haul distribution systems or inside wire applications. Current and continuing applications include the local loop, inside wire and cable and terminal-to-LAN. UTP no longer is deployed in long haul or outside the premises transmission systems.

The application of shielded copper is limited to inside applications due to its additional cost. Specifically, it is generally limited to applications in high-noise environments. It is also deployed where high-frequency signals are transmitted and there is concern about either distance performance or interference with adjacent pairs. Examples include LANs and image transmission.

Coaxial Cable

The core factor that limits a twisted pair cable is due to the skin effect. The flow of the current in the wires is likely to flow only on the wire's outer surface as the frequency of the transmitted signal raises, thus, less of the available cross-section is used. The electrical resistance of the wires is increased for signals of higher frequency which leads to higher attenuation. Further, significant signal power is lost due to the effects of radiation at higher frequencies. Thus, another kind of transmission medium can be used for applications that require higher frequencies. Both these effects are minimized by coaxial cable.

A coaxial cable, as shown in Figure 1.26 is a robust shielded copper wire two-conductor cable in which a solid centre conductor runs concentrically (coaxial) inside a solid outer circular conductor. This forms an electromagnetic shield around the former that serves to greatly improve signal integrity and strength. The two conductors are separated by insulation. A layer of such dielectric (non-conductive) material as Teflon or PVC, protects the entire cable.

The coaxial cable comes under the category of bounded media and is still an effective medium to use in data communication. For better performance the coaxial cable contains shields which make it costly. Cable television uses coaxial cables. LANs function over coaxial cable to the 10BASE5, 10BASE2 and 10BASET specifications.Generally, a coaxial cable allows longer distance transmission instead of twisted pair cable at a higher data rate. However, this is costly.

There are two types of coaxial cables:

(i) **Baseband:** It transmits a single signal at a time at very high speed. The signal on baseband cable must be amplified at a specified distance. It is used for local area networks.

(ii) **Broadband:** It can transmit many simultaneous signals using different frequencies.

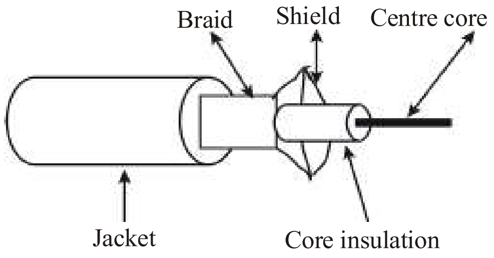


Fig. 1.26 *Coaxial Cable Configuration*

General Properties of Coaxial Cable

The following are the general properties of coaxial cable:

Gauge: The gauge of coaxial cable is thicker than the twisted pair. While this increases the available bandwidth and increases the distance of transmission, it also increases the cost. Traditional coaxial cable is quite thick, heavy and bulky of which Ethernet LAN 10BASE5 is an example. Ethernet LAN 10BASE2 is of much lesser dimensions but offers less in terms of performance.

Configuration: Coaxial cables comprise of a two-conductor wire which is single, with an outer shield (conductor) made of solid metal and a centre conductor. At times, stranded or braided metal is employed. Twin axial cables contain two such configurations within a single cable sheath. The centre conductor carries the carrier signal and the outer conductor is generally used for electrical grounding. Coaxial cable connectivity can be extended through the use of twisted pair with a BALUN (BALanced/UNbalanced) connector serving to accomplish the interface.

Bandwidth: The effective capacity of coaxial cable depends on several factors. These include the spacing of amplifiers, the length of the circuit, the gauge of the centre conductor and other intermediate devices. The available bandwidth over coaxial cable is very significant, hence it is used in high capacity applications, such as image and data transmission.

Error Performance: Coaxial cable performs exceptionally well due to the outer shielding. As a result, it is often used in data applications.

Distance: Coaxial cable is not so limited as UTP, although amplifiers or other intermediate devices must be used to extend high frequency transmissions over distances of any significance.

Security: Coaxial cable is inherently quite secure. It is relatively difficult to place physical taps on coaxial cables. Radiation of energy is also minimal. Hence, its interception is not easy.

Cost: The acquisition, deployment and rearrangement costs of coaxial cables are very high, compared with UTP. In high capacity data applications, however, that cost is often outweighed by its positive performance characteristics.

Applications: Coaxial cable's superior performance characteristics make it the favoured medium in many short hauls, bandwidth-intensive data applications. Current and continuing applications include LAN backbone, host-to-host, host-to-peripheral and CATV.

Optical Fibre

You have seen in the previous section, that the geometry of coaxial cable significantly reduces the various limiting effects and the maximum signal frequency. Hence, the information rate that can be transmitted using a solid conductor, although very high, is limited. This is also the case for twisted lines. An optical fibre is different from the transmission media. The transmitted information is carried through a beam of light which is fluctuating in a glass fibre instead of a wire or an electrical signal. This type of transmission has become strong support for digital networks owing to its high capacity and other factors favourable for digital communication.

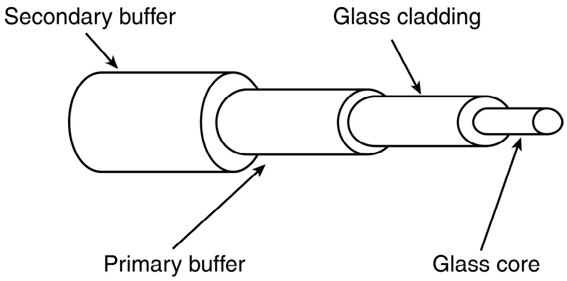


Fig. 1.27 *Fibre Optic Cable — General View*

Fibre optic transmission systems are opto-electric in nature. In other words, a combination of optical and electrical electromagnetic energy is involved. The signal originates as an electrical signal, which is translated into an optical signal, the optical signal is subsequently reconverted into an electrical signal at the receiving end. Figure 1.27 shows a clean, thin glass fibre reflecting light internally as the transmission carries light with encoded data.Fibres can bend without breaking with the help of a plastic jacket. Light Emitting Diode (LED) or laser injected light for transmission into the fibre. Receivers that are light sensitive translate light back into data at the other end.

The optical fibre consists of a number of substructures as shown in Figure 1.28. In this case, the core is made of glass. The glass core carrying the light is encircled by a glass cladding which has lower refractive index. Thus, blending the light and confining it to the core. A substrate layer of glass encircled the core thus, adding to the diameter and the power of the fibre.This layer of glass, however, does not carry light. The mechanical protections cover the secondary buffer coating and primary buffer coating.

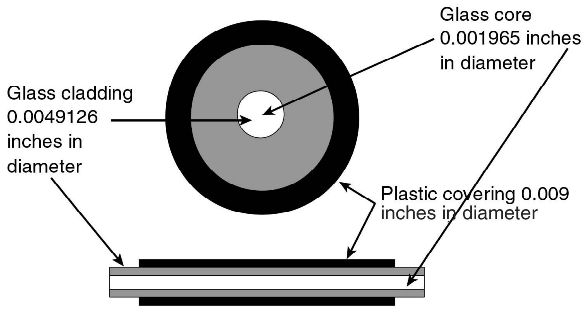


Fig. 1.28 *Glass Fibre Optic Cable, Side View and Cross Section*

The light pulse travels down the centre core of the glass fibre. Surrounding the inner core is a layer of glass cladding, with a slightly different refractive index. The cladding serves to reflect the light waves back into the inner core. Surrounding the cladding is a layer of protective plastic coating that seals the cable and provides mechanical protection. This is shown in Figure 1.22. Typically, multiple fibres are housed in a single sheath, which may be heavily armoured.

Light propagates along the optical fibre core in one of the following ways as given below depending on the type and width of core material used.

1.7 TRANSMISSION CONCEPTS AND TERMS

Before discussing the different types of transmission medium, it is necessary to know the basic concepts and terminologies associated with the transmission of a signal. In the transmission of data, the range of carrier frequencies depends on the nature of the medium and the requirements of the applications supported. Therefore, frequency spectrum may be defined as the range of frequencies being supported by a particular transmission medium. The actual range of frequencies supporting a given communication is known as a pass band. These are given in the Table 1.1.

***Table 1.1** Frequency Spectrum*

Name of Band	*Frequency Range*	*Wavelength*	*Usage*
Audible	20 Hz–20 kHz	>100Km	Voice
Extremely/Very Low Frequency (ELF/VLF) Radio	3 kHz–30 kHz	100–10 Km	Radio Navigation, Weather, Submarine Communications
Low Frequency (LF) Radio	30 Hz–300 kHz	10–1 Km	Radio Navigation, Maritime Communications
Medium Frequency (MF) Radio	300 kHz–3 MHz	1 Km–100 m	Radio Navigation, AM Radio
High Frequency (HF)	3 MHz–30 MHz	100–10 m	Citizens Band (CB) Radio
Very High Frequency (VHF) Radio	30 MHz–300 MHz	10–1 m	Amateur (HAM) Radio, VHF TV, FM Radio
Ultra High Frequency (UHF)	300MHz–3GHz	1 m–10 cm	Microwave, Satellite, UHF TV
Super High Frequency (SHF) Radio	3 GHz–30 GHz	10–1 cm	Microwave, Satellite
Extremely High Frequency (EHF) Radio	30 GHz–300 GHz	1 cm–.1 mm	Microwave, Satellite
Infrared Light	103–105 GHz	300–3µ	Infrared
Visible Light	1013–1015 GHz	1–.3µ	Fiber Optics
X-Rays	1015–1018 GHz	103–107µ	N/A
Gamma and Cosmic Rays	>1018 GHz	<017µ	N/A

The symbols used have the following meanings:

K (Kilo)	=	1,000,	M (Mega)	=	1,000,000 (1 million),
G (Giga)	=	1,000,000 (1 billion)	T (Tera)	=	1,000,000,000 (1 trillion)
cm	=	centimetre (1/100 metre)	mm	=	millimetre (1/1,000 metre)
µ	=	micron (1/1,000,000 metre)			

(a) Bandwidth

In a very general way, bandwidth may be defined as the range of frequencies assigned to a channel. In other words, we may say that bandwidth is the difference expressed in hertz, between the highest and the lowest frequencies of a band. The higher the bandwidth, the more will be the data transmission rate or throughput. It should be noted that bandwidth and data transmission rates are very closely interrelated. Clearly, any transmission system becomes more attractive if the available bandwidth is greater, introduced errors are fewer, and the maximum distance between various network elements (amplifiers, repeaters and antennae) is greater.

(b) Distances

The higher frequency signals offer greater bandwidth. They also suffer to a greater extent from signal attenuation than lower frequencies. This results in more errors in transmission, unless the amplifiers/repeaters are placed closely together. It clearly demonstrates the close and direct relationship between bandwidth, distance, and error performance.

Bandwidth, in this context, refers to the raw amount of bandwidth the medium supports. Error performance refers to the number or percentage of errors which are introduced in the process of transmission. Distance refers to the minimum and maximum spatial separation between devices over a link, in the context of a complete, end-to-end circuit.

(c) Propagation Delay

Propagation delay refers to the length of time required for a signal to travel from transmitter to receiver across a transmission system. While electromagnetic energy travels at roughly the speed of light (30,000 km per second) in free space, the speed of propagation for a twisted pair or coaxial cable is a fraction of this figure. The nature of the transmission system will have considerable impact on the level of propagation delay. In other words, the total length of the circuit directly influences the length of time it takes for the signal to reach the receiver.

(d) Security

Security, in the context of transmission systems, addresses the protection of data from interception as it transverses the network. Particularly in the case of data networking, it is also important that access to a remote system and the data resident on it be limited to authorized users; therefore, some method of authentication must be employed in order to verify that the access request is legitimate and authentic.

(e) Resistance to Environmental Conditions

Resistance to environmental conditions applies especially to wired systems. Twisted pair, coaxial and fibre optic cables are manipulated physically as they are deployed and reconfigured. Clearly, each has certain physical limits to the amount of bending and twisting (flex strength) it can tolerate as well as the amount of weight or longitudinal stress it can support (tensile strength) without breaking (break strength). Fibre optic cables are notoriously susceptible to breaking. Cables hung from poles expand and contract with changes in ambient temperature; while glass fibre optic cables expand and contract relatively less, and twisted pair copper wires are more expansive.

Resistance to environmental conditions also applies to airwave systems since reflective dishes, antennae and other devices used in microwave, satellite and infrared technologies must be mounted securely to deal with wind and other forces of nature. Additionally, the towers, walls and roofs on which they are mounted must be constructed and braced properly to withstand such forces.

(f) Physical Dimensions

The physical dimensions of a transmission system have to be considered as well. This is especially true, once again, in the case of wired systems. Certainly, the sheer weight of a cable system must be taken into account as one attempts to deploy it effectively. The bulk (diameter) of the cable is important, as conduit and

NOTES

raceway space is often at a premium. The physical dimensions of airwave systems must also be looked into, as the size and weight of the reflective dish and mounting system (e.g., bracket and tower) may require support.

(g) Cost and Ease of Installation

Cost is a concern in the selection of an appropriate transmission medium. It includes the cost of acquisition, deployment, operation and maintenance (O&M), and upgradation or replacement. It is noteworthy to compare the costs of deployment of wired versus wireless media.

Wired transmission systems require a right-of-way and this should be secured. Wired transmission involves a cost component in the form of infrastructure. This includes digging of trenches and boring of holes under streets so that cables can be pulled and poles may be mounted. In addition, amplifiers or repeaters may be placed. Such costs are not trivial. Unlike wired systems, wireless systems require secured right-of-way and antennae. It may be inferred that the deployment of wired systems is more costly.

1.7.1 Extending LAN - Master Site and Interconnection to Telephone

There are numerous methods to extend LAN which includes wired options and wireless options. The 'LAN Extenders' use the existing network or inexpensive telephone cables to extend an Ethernet network beyond 100m without using an optical fiber cable system. A Network Extender device is used to extend a LAN beyond 100m, while up to 2000m the inexpensive RJ-11 telephone or RJ-45 network (copper) cables are used.

A Wide Area Network (WAN) is a telecommunications network that typically extends over a large geographic area for extending the computer networking. WANs are often established with leased telecommunication circuits.

WANs are specifically used to connect LANs (Local Area Networks) and other types of networks all together such that the users and computers of one geographical location can easily communicate with users and computers of other geographical locations, as shown in the Figure 1.29. You will learn about LAN extension and cellular networks in unit 2.

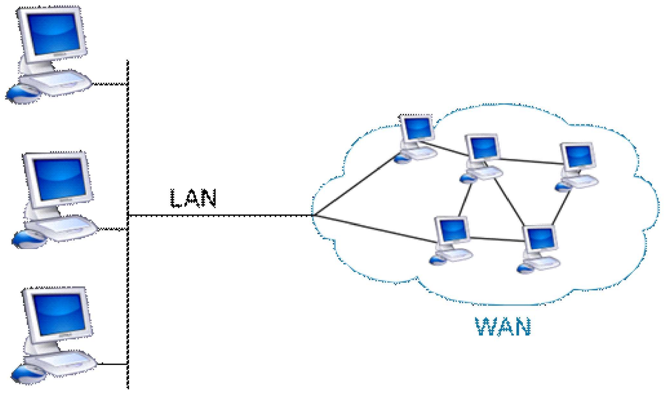


Fig. 1.29 *WAN Connecting LANs*

NOTES

Check Your Progress

12. Which transmission is called un-bounded transmission?
13. Which are two types of coaxial cables?
14. What is a pass band?

1.8 ANSWERS TO ‘CHECK YOUR PROGRESS’

1. Two key issues occur in parallel transfer. The wire itself is the first issue. Minimum of nine wires (eight for data bits, one for circuit ground) are required. Many times extra wires are needed to control the flow of data across the interface. The other issue is with the nature of the bits or voltages itself. When there is change in the state of the bit/voltage from a one to zero, or vice versa, it happens at the rate of nanoseconds (one billionth of a second). A crucial part of the data transfer is the abruptness

10. The following are the common network protocols:
    - Ethernet
    - Local Talk
    - Token Ring
    - FDDI
    - ATM
11. Privately owned networks offer consistent, fast paced communication channels which are optimized to connect information processing tools in a restricted geographical area. These are known as Local Area Networks (LANs).
12. Mmicrowave and satellite transmissions, both travel through the air, which has no boundaries, hence called un-bounded transmission.
13. There are two types of coaxial cables:

    (i) **Baseband:** It transmits a single signal at a time at very high speed. The signal on baseband cable must be amplified at a specified distance. It is used for local area networks.

    (ii) **Broadband:** It can transmit many simultaneous signals using different frequencies.
14. The actual range of frequencies supporting a given communication is known as a pass band.

1.9 SUMMARY

- Data transmission can be divided into parallel and serial data transmission.
- A voice grade channel is approximately 4,000 Hz, or 4 kHz. Approximately 3.3 kHz (200 Hz to 3,500 Hz) is used for the voice signal itself.
- CATV video channel is approximately 6 MHz CATV video channel is approximately 6 MHz Approximately, 4.5 MHz is used for information transmission,
- Computers are digital in nature. Computers communicate, store and process information in binary form, i.e., in the combination of 1s and 0s.
- For transmission across a network, data has to be transformed into electromagnetic signals. Both, data and signals can be either of analog type or digital type.
- Information exchange is an essential part of communication. It may be exchange of information among users or equipment in the communication system.
- Signalling System 7 (SS7) is the protocol designed for public switched telephone system for providing services and setting up calls.
- Digital systems offer better security while analog systems offer some measure of security through the scrambling of several frequencies.

- In a full duplex mode data can be transmitted in both directions at the same time.
- Channel bandwidth may be simply defined as the size of the range of frequencies that can be transmitted through a channel.
- Conventionally, modems were devised for communication between a host computer and data terminals.
- Modems continuously generate a carrier signal to send information so that the information may be delivered from one location to another remote location.
- Short haul modems are widely deployed over private lines and are not part of a public system.
- Voice grade modems have a frequency range of moderate to high data rate.
- Dial-up modems are used for point-to-point connections on the PSTN by any combination of manual or automatic dialing or answering. The quality of the circuit may vary from carriers to carriers.
- Synchronous modems operate in the audio range, at rates up to 28.8 kbps in audio lines
- Dial-up modems contain a circuitry that mimics that of a telephone. That is, modems can simulate lifting the handset, dialing, or hanging up the telephone.
- In the mainframe and minicomputer environment, each user is connected to the main system through a dumb terminal that is unable to perform any of its own processing tasks.
- Privately owned networks offer consistent, fast paced communication channels which are optimized to connect information processing tools in a restricted geographical area. These are known as Local Area Networks (LANs).
- There are two categories of twisted pair cables—with and without shielding.
  (i) Unshielded Twisted Pair (UTP)
  (ii) Shielded Twisted Pair (STP))
- The core factor that limits a twisted pair cable is due to the skin effect. The flow of the current in the wires is likely to flow only on the wire's outer surface as the frequency of the transmitted signal raises, thus, less of the available cross-section is used.

1.10 KEY TERMS

- **Analog Signal:** It is any continuous signal where the time varying feature of the signal represents some other time varying quantity.
- **Signalling**: It refers to the exchange of information between components required to provide and maintain data communication service.
- **Signalling System 7**: Signalling System 7 (SS7) is the protocol designed for public switched telephone system for providing services and setting up calls.

- **Half Duplex:** In half duplex mode, data is transmitted in one direction at a time, for example, a walkie-talkie.
- **Full Duplex**: In a full duplex mode, data can be transmitted in both directions at the same time.
- **Bandwidth:** It is defined as the range of frequencies assigned to a channel.
- **Propagation Delay**: It refers to the length of time required for a signal to travel from transmitter to receiver across a transmission system.

1.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the different communication modes from the view point of transmission?
2. What are the functions of SS7?
3. What is a modem? Describe its function.
4. What are the different types of modems?
5. Write the characteristics of communication protocols.
6. What advantages do coaxial cables offer over twisted pair cables?
7. What are the suggested connectors and media for horizontal wiring?
8. What do you mean by parallel and serial transmission?

Long-Answer Questions

1. Describe the advantage of analog transmission.
2. Explain the significance of digital signals in data communication.
3. What are the different types of modem? Explain.
4. Describe simplex, half duplex and full duplex mode of communication with the help of a diagram.
5. Compare fibre optical cable with UTP cable when used as transmission media in LANs.
6. Write a detailed note on:
   (i) Twisted pair cable
   (ii) Coaxial cable
   (iii) Optical fibre cable

1.12 FURTHER READING

Forouzan, Behrouz A. *Data Communications and Networking*. New Delhi: Tata McGraw-Hill, 2004.

Stallings, William and Richard Van Slyke. *Business Data Communications*. New Jersey: Prentice Hall, 1998.

NOTES

Black, Uyless. *Computer Networks*. New Jersey: Prentice Hall, 1993.

Stallings, William. *Data and Computer Communications*. New Jersey: Prentice Hall, 1996.

Tanenbaum, Andrew S. *Computer Networks*. New Jersey: Prentice Hall PTR, 2002.

Stallings, William. *Data and Computer Communications*. NJ: Prentice-Hal, 1996.

UNIT 2 LOCAL AREA NETWORK, IMPLEMENTING AND EXTENDING LAN

NOTES

Structure

2.0 INTRODUCTION

A Local Area Network (LAN) is a computer network that interconnects computers within a limited area such as a residence, school, laboratory, university campus or office building. LAN is a group of computers or other devices interconnected within a single, limited area, typically via Ethernet or Wi-Fi. By contrast, a wide area network (WAN) not only covers a larger geographic distance, but also generally involves leased telecommunication circuits. Ethernet and Wi-Fi are the two most common technologies in use for local area networks. Historical network technologies include ARCNET, Token Ring, and AppleTalk.

LAN networking requires Ethernet cables and Layer 2 switches along with devices that can connect and communicate using Ethernet. Larger LANs often include Layer 3 switches or routers to streamline traffic flows. A LAN enables users to connect to internal servers, websites and other LANs that belong to the same WAN. Ethernet and Wi-Fi are the two primary ways to enable LAN connections.

LAN implementation uses several network devices, like a router, switch, network cable (UTP cable), and PC or laptop as user. Some of features that used in this LAN network are website blocking, (Dynamic Host Configuration Protocol) DHCP, Hotspot, and bandwidth.

In this unit, you will learn about local area network, LAN operating systems, implementing LAN, fast LANs, non-standard LANs, extending LAN, transmission concepts and terms, master sites, and interconnection to telephone.

NOTES

2.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Define local area network
- Explain LAN operating systems
- Understand the implementation of LAN using optical fibre cable and wireless technology
- Discuss fast LANs, and non-standard LANs
- Understand how to extend LAN
- Understand transmission concepts and terms
- Illustrate master sites
- Analyze interconnection to telephone

2.2 LOCAL AREA NETWORK

Local Area Network technology connects people and machines within a site. A Local Area Network (LAN) is a network that is restricted to a relatively small area as shown in Figure 2.1. Local Area Networks (LANs) can be defined as privately owned networks offering reliable high speed communication channels that are optimized to connect information processing equipment in a small and restricted geographical area, namely, an office, a building, a complex of buildings, a school or a campus.

A LAN is a form of local (limited-distance), shared packet network for computer communications. LANs interconnect computers and peripherals over a common medium so that users are able to share access to host computers, databases, files, applications, and peripherals. They can also provide a connection to other networks either through a computer, which is attached to both networks, or through a dedicated device called a gateway.

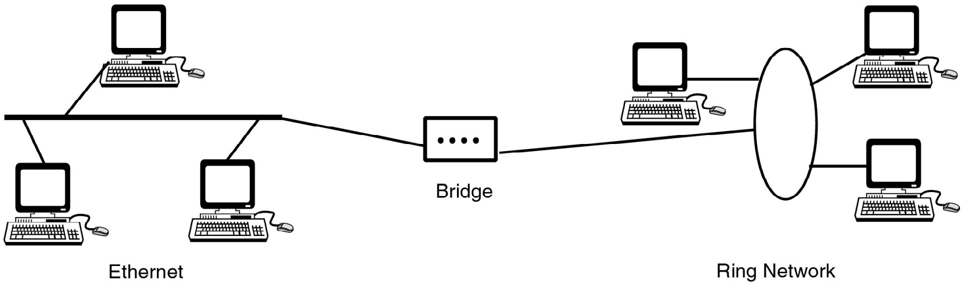


Fig. 2.1 *Local Area Network (LAN)*

The components used by LANs can be categorized into hardware, cabling standards, and protocols. Various LAN protocols are Ethernet, Token Ring: TCP/

IP, SMB, NetBIOS and NetBeui, IPX/SPX, Fibre Distributed Data Interchange (FDDI) and Asynchronous Transfer Mode (ATM).

LANs are used almost exclusively for data communication over relatively short distances such as within an office, office building or campus environment. LANs allow multiple workstations to share access to multiple host computers, other workstations, printers and other peripherals, and connections to other networks. LANs are also being utilized for imaging applications. They are being used for video and voice communication as well, although currently on a very limited basis.

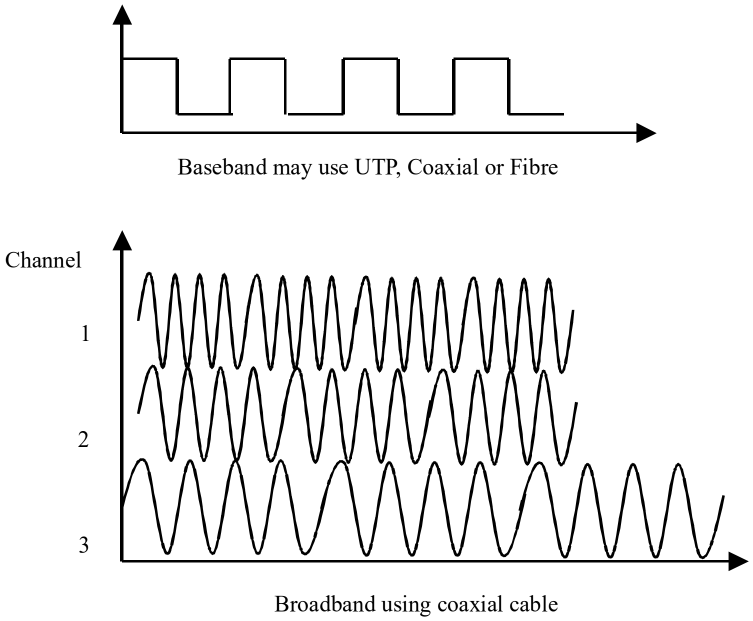
LAN applications include communication between the workstation and host computers, other workstations and servers. The servers may allow sharing of resources. Resources could be information, data files, e-mail, voice mail, software, hardware (hard disk, printer, fax, etc.) and other networks.

LAN benefits include the fact that a high-speed transmission system can be shared among multiple devices in support of large number of active terminals and a large number of active applications in the form of a multi-user, multitasking computer network. LAN-connected workstations realize the benefit of decentralized access to very substantial centralized processors, perhaps in the form of mainframe host computer and storage capabilities (information repositories). Additionally, the current technology allows multiple LANs to be inter-networked through the use of LAN switches, routers, etc.

Disadvantages of LANs include concern for security of files and accounts.

2.2.1 Broadband versus Baseband

There exist two LAN transmission options, baseband and broadband. Baseband LANs, which is the most prevalent by far, is a single-channel system that supports a single transmission at any given time. Broadband LANs support multiple transmissions via multiple frequency channels.



Baseband may use UTP, Coaxial or Fibre

Broadband using coaxial cable

Fig. 2.2 Broadband versus Baseband

Broadband LANs

Broadband LANs are multichannel, analog LANs as shown in Figure 2.2. They are typically based on coaxial cable as the transmission medium, although fibre optic cable is also used. Individual channels offer bandwidth of 1 to 5 Mbps, with 20 to 30 channels typically supported. Aggregate bandwidth is as much as 500 MHz. Its characteristics are:

- Stations connected via RF modems, i.e., radio modems accomplish the digital-to-analog conversion process, providing the transmitting device access to an analog channel.
- Digital signal modulated onto RF carrier (analog).
- Channel allocation based on FDM.
- Head-End for bidirectional transmission.

Advantages

- Greater bandwidth
- Data, voice and video can be accommodated on broadband channel
- Greater distances

Disadvantages

- High cost, requires modems
- Lack of well-developed standards
- Cable design
- Alignment and maintenance

Some broadband LANs are referred to as 10Broadband36 where 10 stands for 10 Mbps, Broadband for multichannel and 36 for 3600 metres maximum separation between devices.

Baseband LANs

Baseband LAN is a single channel connection, supporting a single communication at a time as shown in Figure 2.2. They are digital in nature. Total bandwidth of 1 to 100 Mbps is provided over coaxial cable, UTP, STP, or fibre optic cable. Distance limitations depend on the medium employed and the specifics of the LAN protocol. Baseband LAN physical topologies include the ring, bus, tree, and star topologies.

Baseband LANs are, by far, the most popular and the most highly standardized. Ethernet, Token Passing, Token Ring and FDDI LANs are all baseband. They are intended only for data, as data communication is, after all, the primary reason for the existence of LANs. The characteristics of this system may be summarized as follows:

- No need of modems—low cost installation
- Bidirectional propagation of signal
- Unmodulated digital signal
- Single channel
- Stations connected via T connectors

Advantages

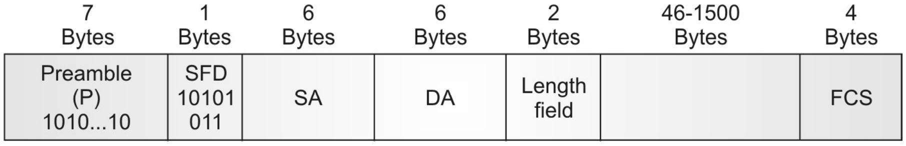
- Simplicity
- Low cost
- Ease of installation and maintenance
- High rates

Disadvantages

- Limited distances
- Data and voice only

2.2.2 LAN Hardware and LAN Operating System

The attached devices to a cable in any topology are referred nodes, hosts or stations. In addition to them, LANs may make use of other devices to control physical access to the shared medium, to extend the maximum reach of the LAN, and to switch traffic. Such hardware is in the form of NIC/NIU, transceivers, MAU, hubs, bridges, routers, and gateways.



7 Bytes	1 Bytes	6 Bytes	6 Bytes	2 Bytes	46-1500 Bytes	4 Bytes
Preamble (P) 1010...10	SFD 10101 011	SA	DA	Length field		FCS

Fig. 2.3 *Frame Format for IEEE 802.3*

Network Interface Card (NIC)

This is also known as Network Interface Unit (NIU). NIC is a hardware card to provide physical access from a node to the LAN medium. The NIC can be fitted into the expansion slot of a PC, or it can exist as a separate box. A standalone, multiport NIC can serve a number of devices, thereby providing an additional level of contention control. A standard IEEE NIC are encoded with a unique hardware coded logical address. Transceivers are embedded in NIC/NIU and MAU. MAU (Media Access Unit, or Multistation Access Unit) are standalone devices that contain NIC in support of one or more nodes.

LAN Operating Systems

A LAN Operating System, or Network Operating System (NOS), is software that provides the network with multi-user, multitasking capabilities and supports communications and resource sharing. Thus, a LAN operating system is used to provide the basic framework for the operation of the LAN. The operating system does not confine to one environment but it has distributed modules across the LAN environment. They may be located in server and client. LAN architecture is an example of client server architecture in which a client is a user interface like a computer with several applications like word processing, spreadsheet or database. The client and server are on the same computer network or LAN. The client requests a server for certain services. The servers are usually a multi-port computer with large memory and enabling multiple clients to share their resources along with performing certain functions independently. Servers are database engines capable of processing client requests for information. Servers also manage the data.

In addition to supporting multitasking and multi-user access, LAN operating systems provide for recognition of users based on passwords, user IDs, and terminal IDs. On the basis of such information, LAN operating systems can manage security using access privileges. Additionally, a LAN operating system provides multi-protocol routing, as well as directory and message services. DOS-based LAN operating systems include Novell NetWare and Sun Microsystems' TOPS/DOS. OS/2 and UNIX-based LAN operating systems include Banyan VINES, IBM LAN Server, Microsoft LAN Manager, and Novell SFT NetWare.

Check Your Progress

1. What is a Local Area Network (LAN)?
2. What are the components used by LANs?
3. Name the two LAN transmission options.
4. Define baseband LAN.
5. What do you understand by NIC?
6. What is a LAN operating system?

2.3 IMPLEMENTATION OF LAN USING FIBER-OPTIC CABLES

The shared medium for LANs includes most of the transmission media discussed in previously. Although coaxial cable was the original medium and still is used widely in various configurations, twisted-pair has recently become the medium of choice in many environments. Fiber optic cable is used widely as a backbone technology. Wireless LANs generally are limited to special radio technologies, although infrared technology is used in certain applications. Microwave and infrared systems are used to connect LANs and LAN segments in a campus environment. Satellite rarely is used in any way, as propagation delay renders it generally unsatisfactory for interactive communications.

Implementation of LAN using Coaxial Cable

Coaxial cable is still a popular transmission media for LAN. The coaxial cable is considered as the foremost transmission medium to implant LANs. However, the coaxial cable is considered an expensive option in terms of its cost, deployment, reconfiguration and maintenance but due to its excellent performance characteristics, it is the most widely deployed transmission media. The advantages of coaxial cable include high bandwidth of the order of 500 MHz and more, good error performance and no distance limitation. Additionally, security is high and durability is excellent. On the other hand, the costs of acquisition, deployment, and reconfiguration are high. The disadvantages of coaxial cable have been mitigated to a large extent through the development of new coaxial designs. The variations of coaxial designs are thick net and thin net. In the following sections both baseband and broadband versions of Ethernet/IEEE 802.3 are explained.

NOTES

10Base5 (Thick Net/Yellow Ethernet)

This uses traditional thick baseband coaxial cable in bus topology to connect multiple computers as shown in Fig. 2.4. This single transmission line is called a segment. A coaxial cable 10mm in diametre, known as a thick coaxial cable is used as a transmission line. A terminator is connected at each end of the cable. Note that proper data communication cannot be assured if even one of these terminators is missing or not properly connected.

A transceiver is used to connect a coaxial cable and terminals. A transceiver cable (also referred to as an AUI (Attachment Unit Interface) cable) is used to connect a transceiver and the NIC. The maximum length of this cable is 50 metre. Up to 100 transceivers can be connected to each segment. The minimum allowable distance between transceivers is 2.5 metre.

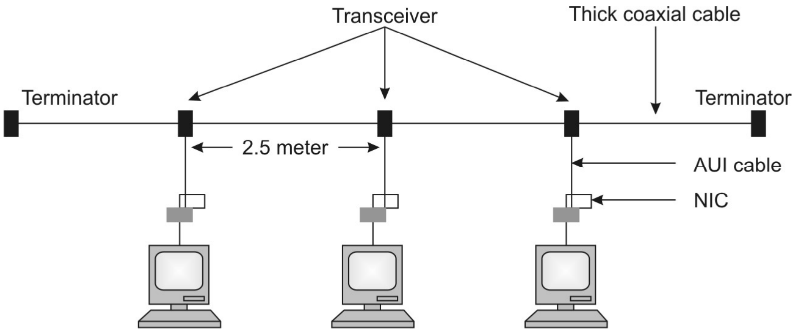


Fig. 2.4 *Hardware Configuration of 10Base5*

10Base stands for 10Mbps and baseband transmission system. The 5 of 10Base5 signify a maximum of 500-metre segment length. The segment may be extended upto 1500 metres by using repeaters.

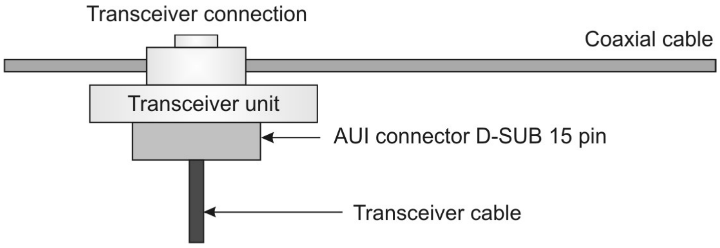


Fig. 2.5 *Transceiver*

Transceiver

The transceiver exchanges data signals handled by the NIC and electric signals sent over a transmission line. A 15-pin D-SUB connector shown in Fig. 2.5 is used to connect transceivers and transceiver cables. Multiport transceiver supports more than one NIC.

10Base2 (Thin Net/Black Ethernet)

This is also known as 10Base2, uses coaxial of thinner gauge of 5mm in diametre and bus topology as in the case of 10Base5 so that multiple computers can be connected to a single transmission line as shown in Fig. 2.6. Primarily it was used in office environments. The thinner cable is less costly to acquire and deploy,

NOTES

although its performance is less in terms of transmission distance. Because of its cost it is sometimes called as cheapnet. 10Base2 signifies in the same manner as 10Bases5 except 2 is signified here as 200 metres maximum segment length (actually 185 metres).

A BNC (Bayonet Neil Connector) or a T-connector is used to connect a cable and terminals or terminators. Note also that the NIC for 10Base2 can be connected directly to a T-connector because this NIC has a built-in transceiver. Only up to 30 nodes per segment can be connected. The minimum allowable distance is 0.5 metre between consecutive connections.

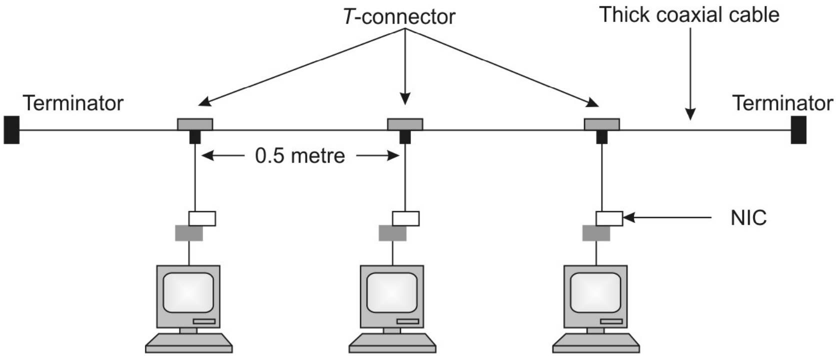


Fig. 2.6 *Hardware Configuration 10Base2*

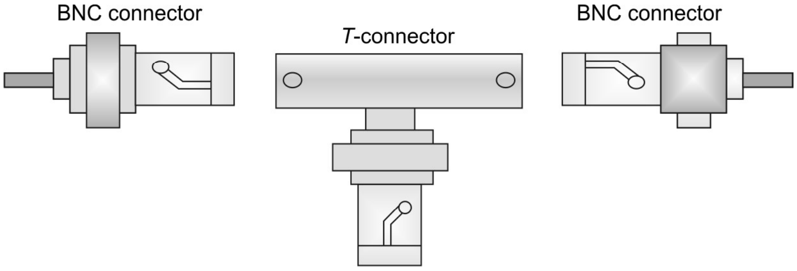


Fig. 2.7 *BNC/T Connector*

BNC/T-connector

The Fig. 2.7 shows the BNC connector and T-connector. These are simple connectors that cannot exchange data. An NIC and T-connector must be directly connected.

Implementation of LAN using Twisted Pair

Recently twisted pair has become very popular as a LAN medium. Although its performance characteristics are less appealing, its low cost and high availability certainly are attractive. Unshielded Twisted-Pair (UTP) cable's performance is very good at low data rates. LAN uses the same cable which telephone sets use. This enables to deploy Category 5 UTP pairs to each jack to ensure that voice and data terminals share a single wiring system. Additionally, UTP are also capable of working at very high data rates up to 100 Mbps but for short distances. Shielded twisted-pair (STP) sometimes is used in LAN technology. STPs are more useful in the environment susceptible to EMI/RFI because UTP data transmission might cause interference on adjacent pairs.

NOTES

Implementations of different LANs using UTP Cables are explained below.

10BASET (Twisted pair Ethernet)

This uses Cat 3, 4, or 5 UTP. 10BaseT translates to 10 Mbps, Baseband, Twisted pair. 10BaseT actually is a wire hub that serves as a multiport repeater, as well as a central point of interconnection. Figure 2.8 shows the hardware configuration of 10BaseT. A star type topology is used. The device shown at the center is known as hub. A connector known as RJ45 is connected at each end of the cable. The hub has multiple ports, each of which is connected to node using NIC via the UTP cable. Each of the NIC for the 10BaseT has a built-in transceiver as do those of the 10Base2. The maximum distance between the 10BaseT hub and the attached device is 100 metres/segment.

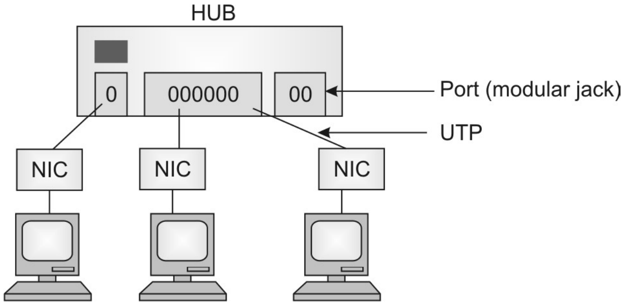


Fig. 2.8 *Hardware Configuration 10BaseT*

Hub

Hubs receive signals through one port and send them through all other ports as shown in Fig. 2.9. That is, a LAN configures with hubs physically falls under the category of a star type topology. However, logically, it falls under the category of a bus type topology. Commercially available hubs normally have eight or sixteen ports.

CDDI (Cable Distributed Data Interface)

CDDI is also known as TPDDI (Twisted-pair Distributed Data Interface). CDDI employs Cat 5 UTP as an inexpensive means of connecting workstations and peripherals to FDDI fibre optic backbone LANs. A transmission rates up to 100 Mbps may be achieved in this scheme. The maximum allowable distance between hub and the device is not specified categorically but generally it is kept less than 20metres. Cat 3 UTP often is used for 4 Mbps token ring LANs, cat 4 UTP has a bandwidth of 20 MHz and commonly is used for 16 Mbps token ring LANs.

Ethernet Expansion

The maximum allowable length of a segment for the 10Base5 is 500 metre. Upto 100nodes can be connected to a segment. That is upto 100 transceivers can be connected. Repeaters are used to connect terminals that are separated from one another beyond the distance specified above, or to connect more than a limited number of nodes. Each repeater has two ports so that it receives signals through one of these ports and sends them through the other port after amplification. Figure

2.10 shows an example of the 10Base5 LAN. Upto 1024 nodes can be connected to a LAN expanded by using repeaters.

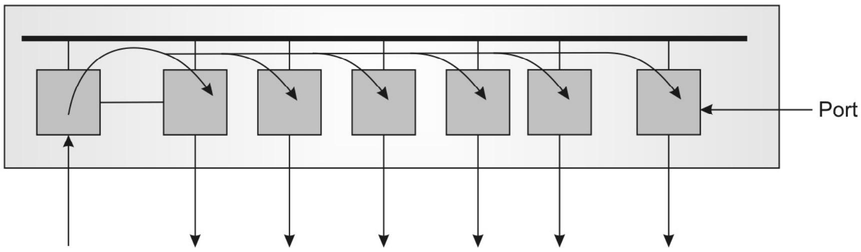


Fig. 2.9 Hub Configuration

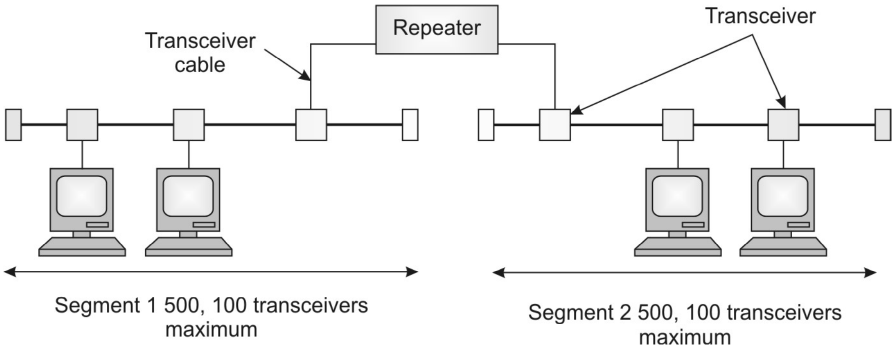

Repeater
Transceiver cable
Transceiver
Segment 1 500, 100 transceivers maximum
Segment 2 500, 100 transceivers maximum

Fig. 2.10 10Base5 Ethernet Expansion

5-4-3 Rule

The number of repeaters that can be used is limited. In the case of data communication between terminals, data can be sent through only up to four repeaters. Therefore, maximum five segments can be provided between terminals. Of these five segments, only up to three can be connected to terminals and are called as populated. The other two are link segments as shown in Fig. 2.11. This limitation is called the 5-4-3 rule (five segments, four repeaters and three populated segments). Those segments that are not or cannot be connected to any terminals are called link segments.

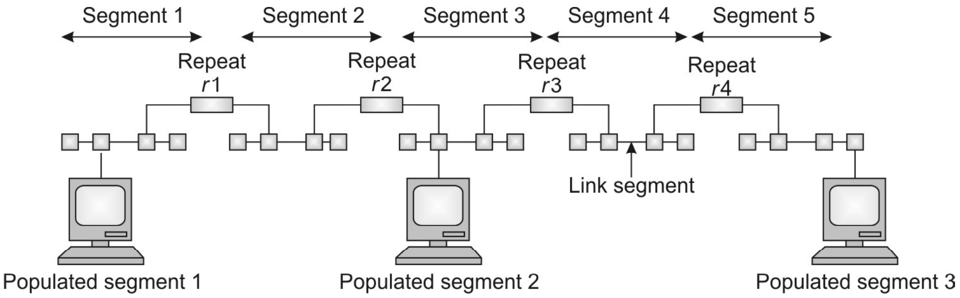


Fig. 2.11 Example of Maximum Configuration

10Base2 Expansion

The maximum allowable length of 10Base2 is 185 metre. Up to 30 can be connected to segment. Repeaters are used as with the 10Base5 to connect terminals that are separated from one another by more than the above distance or to connect more nodes than the above limit. The limitations to the number of repeaters that can be used are the same as those for the 10Base5. As with the 10Base5 (5-4-3 rule), up to 1024 nodes can be connected to a LAN expanded by using repeaters. This is shown in Fig. 2.12.

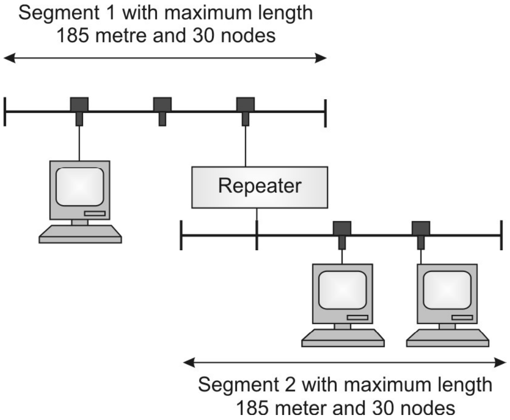


Fig. 2.12 *10Base2 Expansion*

10BaseT Expansion

The maximum allowable length of a cable between a hub and terminal is 100 metre in this case. Cascade connection is employed to connect a hub and terminal that are separated from one another by more than the above distance. This connection is also applied when more ports than available are required. The numbers of hubs that can be connected are limited up to four only.

In a cascade connection the hub at downstream from other must have a port to be used exclusively for cascade connection. This port is called an uplink port as shown in Fig. 2.13. Almost all hubs have such a port. However, if the hubs are connected without uplink ports, a special cable, called a cross cable, is used to connect the hub ports. Hub can be considered a kind of repeater with multiple ports (multiport repeater). This is why the number of repeaters and hubs that can be used is the same.

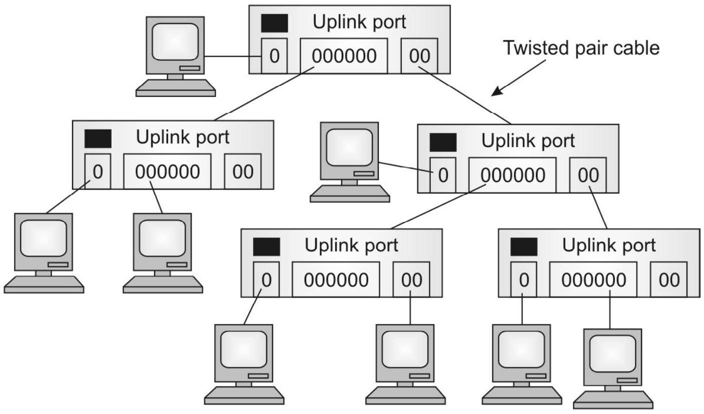


Fig. 2.13 *10BaseT Expansion*

Stackable Hub

The numbers of hubs are limited to four only for cascade connection. Sometimes more ports are required. In this case stackable hubs are used. Each stackable hub has a special interface designed to connect it to another stackable hub as shown in Fig. 2.14. Connecting stackable hubs through this interface allows these hubs to be treated as a single hub. This interface is product dependent. The maximum number of stackable hubs that can be connected varies depending upon the product used.

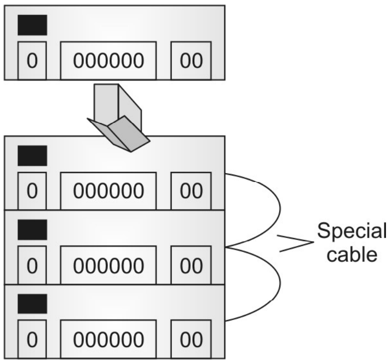
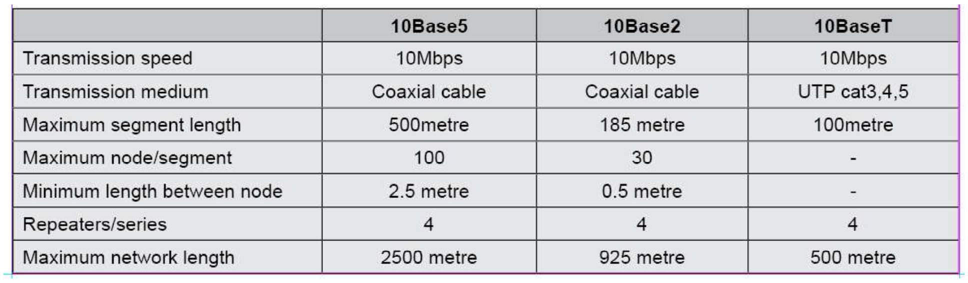


Fig. 2.14 *Stackable Hub*

Table 2.1 *Ethernet Specifications*



	10Base5	10Base2	10BaseT
Transmission speed	10Mbps	10Mbps	10Mbps
Transmission medium	Coaxial cable	Coaxial cable	UTP cat3,4,5
Maximum segment length	500metre	185 metre	100metre
Maximum node/segment	100	30	-
Minimum length between node	2.5 metre	0.5 metre	-
Repeaters/series	4	4	4
Maximum network length	2500 metre	925 metre	500 metre

Ethernet Specification

The above Table 2.1 shows the major Ethernet specifications described so far. Note that the maximum network length means the maximum allowable distance between nodes.

Implementation of LAN using Fibre-optic Cable

Due to its outstanding performance characteristics, fibre-optic cable is also used in LANs. FDDI (Fibre Distributed Data Interface) is the current LAN standard (IEEE and ANSI) for such a network. FDDI can be extended to the desktop, either directly or through the use of twisted-pair in a CDDI application.

- FDDI is the standard (ANSI X3T9-5; IEEE 802.2) for a fibre optic, token-passing ring LAN
- High bandwidth - 100 Mbps with full duplex interfaces
- Excellent error performance in the range of 10^{-14}
- Fibre is capable of transmitting data to very long distances
- Separation for multimode fibre can be as much as 1.2 miles (2 km)

- 37.2 miles (62 km) over single mode fibre
- Excellent security
- The maximum frame size is 4500B capable of accommodating the native frame sizes of all standard LAN networks
- Costly acquisition, deployment and reconfiguration.
- Protection of optical fibre is very important since it is extremely fragile.
- The fragility of fibre acts as a deterrent to the application of FDDI. FDDI specifications ask for a dual countr-rotating ring as shown in Fig. 2.15, which provides a measure of redundancy. If the primary ring fails, a Dual Attached Station (DAS) or a Dual Attached Concentrator (DAC) can still communicate with any other device by transmitting through the secondary ring.

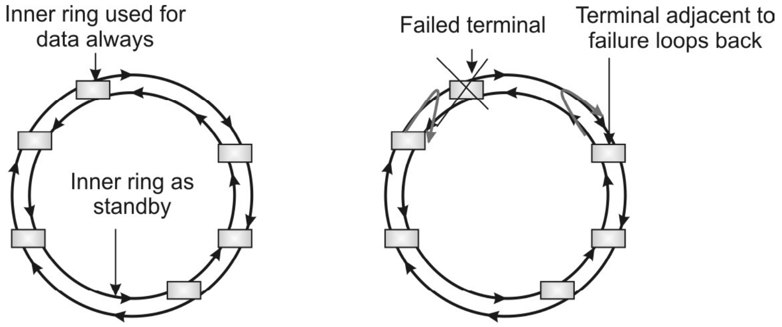

Inner ring used for data always
Inner ring as standby
Failed terminal
Terminal adjacent to failure loops back

***Fig. 2.15** FDDI dual counter-rotating ring*

2.4 IMPLEMENTATION OF LAN USING WIRELESS TECHNOLOGY

Wireless LANs offer the obvious advantage of avoidance of cabling costs, which can be especially important in a dynamic environment where there is frequent reconfiguration of the workplace. Additionally, wireless LANs provide LAN capabilities in temporary quarters, where costly cabling would have to be abandoned. Each workstation is fitted with a low power radio antenna. The antenna is then connected to other hub antennae and to the servers, peripherals, and hosts via cabled connections, which also connect multiple hub antennae for transmission between rooms, floors, and buildings. In order to serve multiple workstations, spread-spectrum radio technology is used to make effective use of limited bandwidth. Spread spectrum involves scattering packets of a data stream across a range of frequencies, rather than using a single transmission frequency. A side benefit of spread-spectrum is increased security, as the signal is virtually impossible to intercept.

- Wireless LAN is a relatively immature technology.
- Acquisition costs are not particularly low when compared to wired LANs, although reconfiguration costs are virtually non-existent.
- Frequency range—900 MHz, 2 GHz, and 5 GHz bands.

- A hub antennae is located at a central point from where line-of-sight can be established with the various terminal antennae.
- Bandwidth of a wireless radio LAN—4 Mbps
- The effective throughput is more in the range of 1 to 2 Mbps per hub.
- The infrared transmission technique can also be used. PDA (Personal Digital Assistant) make widespread use of infrared to establish links with workstations and other PDA for data transfer. Enhanced infrared technology recently has been demonstrated at speeds of 1.5, 4, and even 155 Mbps.
- Error performance and security are issues of some significance.

Some wireless LANs also use direct sequence transmission, which means that a signal is sent simultaneously over several frequencies, and therefore increase its chances of getting through to the access hub. Figure 2.2Hardware 16 shows an example of hardware configuration.

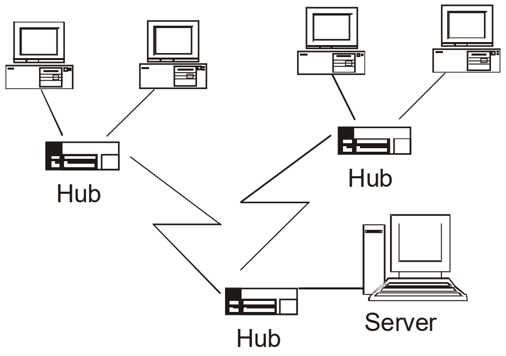


Fig. 2.16 *Wireless LAN Configuration*

WLAN is mostly a mix of wire and wireless media having an access point or wireless router that is connected to a wired network via a coaxial cable, universal serial bus (USB), or Ethernet connection. The devices you want to connect to the wireless LAN must have a peripheral component to communicate with the access point.

IEEE 802.11a and IEEE 802.11b are wireless network standards with a data rate of only 2Mbps and 11 Mbps respectively. They have a distance limitation up to 100 feet from the access point router. This uses 2.4 GHz band. IEEE 802.11g allows speeds up to 54Mbps. This wireless technology is known as IEEE 802.11g, Wireless-G, or 54g, and continues to use the 2.4 GHz band.

Exponential growth rates in mobile communication systems, enhanced awareness in society and deregulation of former monopolized markets have paved the way for the easy use of mobile communication systems as well as a new set of issues, techniques and solutions. Digital Cellular networks are the wireless extensions of traditional PSTN or ISDN networks and allow for nationwide or even worldwide seamless roaming with the same mobile phone. Cellular networks have traditionally been the preserves of voice. However, data traffic is continuously growing. Most of the mobile phones have the ability to send and receive short text messages, and an increasing number now incorporate more advanced Internet capabilities such as World Wide Web onto mobile and wireless devices. New applications and new mobile networks will bring ubiquitous multimedia computing to the radios, Personal Digital Assistants (PDAs), laptops and mobile phones. These device may also get converge and many more functions will be available on one device only.

This unit attempts to introduce the field of mobile communications and digital data transfer. It requires an understanding of communication and networking which have been adequately provided in the previous chapters. There are two different kinds of mobility. If you ask a question about the appearance of a computer after 20 years or more you will not get an accurate answer but many people may predict that most computers will certainly be portable. Therefor, a user will have to access the network without any wires i.e. wireless. Keeping on view these features in mind, we can say that there are two different kinds of mobility. These are user mobility and device portability. In case of user mobility a user can access to the same or similar telecommunication services at different places. Unlike user mobility, device portability allows the movement of device with or without user. The network and device architecture ensures the commencement of communication. The term wireless is used in the context of device portability.

A number of mobile and wireless devices are available in different forms depending upon various applications. Some of these devices are sensor, embedded controllers, pager, mobile phones, personal digital assistant (PDA), palmtop, notebook, etc. The availability of low cost microprocessors and digital switching made the wireless communication popular among the masses.

Cellular Radio Definition

Cellular radio has another popular names as cellular mobile or mobile phone. Radio is basically a device, which uses receiver and transmitter. Wireless communication can also be carried out without using radio. The interference caused by high power line to radio transmission is also an example of wireless communication though it is termed as noise. Inductive and conductive circuits and devices can communicate wirelessly for limited distance with less reliability and implementation problem. Therefore these techniques can not be termed as radio transmission.

The term radio may be defined as consisting of modulation and radiation of the signal. Modulation techniques have been discussed in details. Therefore, a transmitter and an antenna are used to modulate and radiate the modulated signal within radio spectrum as shown in Figure 2.17. On the hand at the receiving end, an antenna and a receiver is required to demodulate the signal. If the transmitting and receiving end are receiving and transmitting, a transceiver (consisting of transmitter and receiver operation) is employed. In telephone system as we know that a voice with bandwidth of approximately 4 kHz modulates the current of a telephone line. Wireless can be defined as the radio transmission and reception of signals by means of high frequency electrical waves without a connecting wire.

VLF	LF	MF	HF	VHF	UHF	SHF	EHF	Infra-red	Visible light

10kHz 100kHz 1MHz 10MHz 100MHz 1GHz 10GHz

VLF: Very low frequency LF: Low frequency MF : Medium frequency
VHF: Very high Frequency HF: High frequency UHF: Ultra high frequency
SHF: Super high frequency EHF: Eextremely high frequency

Fig. 2.17 *Radio Spectrum*

A cellular system is the communications systems that divide a geographic region into sections, called cells, each having its own dedicated frequency. The frequency of a cell may be reused after the interference zone. Now we can define cellular radio or cellular mobile or mobile phone as a communication system that consists of a combination of radio transmission and Public Switched Telephone Network (PSTN) to permit telephone communication to and from mobile subscribers within a specified area. This requires a cellular architecture, which has been described in the subsequent sections. For digital communications, several competing cellular systems exist. These are Global System for Mobile Communication (GSM), CDMA, etc.

Basic Wireless Principles

A cellular mobile communications system consists of a large number of low power wireless transmitters to create cells. These cells cover a certain area and typically called as base station. Depending upon the power level the size of cells can be decided. In this way, the radii of a cell may vary from ten of meters to tens of kilometers in a building to a city respectively. It will also depend upon the subscriber density and demand within a particular area. The shape of cell may not be a perfect hexagon or circle and depends upon the environment. When a mobile user travels from cell to cell, their conversations are "handed off" between cells in order to maintain seamless service. Cells can be added as per the demands based upon the user density or newly created areas. Channels (frequencies) used in one cell can be reused in another cell after some distance and therefore it uses Space Division Multiplexing (SDM). Frequencies for communication may vary from very high frequency (VHF) to – microwave range. Regulation bodies of the concerned countries regulate these. The signal may be analog or digital with amplitude, frequency and phase modulation. The multiplexing and access techniques are space division multiplexing (SDM), frequency division multiplexing (FDM), time division multiplexing (TDM), and code division multiplexing (CDM).

The advantages of mobile communication may be looked into higher capacity, higher number of users, less transmission power needed, more robust, decentralized base station deals with interference, transmission area etc. The disadvantages are fixed network needed for the base stations, handover (changing from one cell to another) necessary, interference with other cells such as co-channel, adjacent-channel.

It is now evident that cellular networks are essential for wireless transmission. We ought to know about the cellular concept, frequency reuse, channel allocation, call setup, location management, cell handoffs, optimizations in terms of power control and cell capacity and implementations of GSM, GPRS, 3G etc. The important issues on wireless communication are cell sizing, frequency reuse planning and channel allocation strategies.

GSM Cellular Radio System

In the beginning around 1980s, analog cellular telephone systems were developing in Europe and each country were developing its own system and thus making them confined within their country boundaries. Later on, the need for an European public land mobile system were realized. In 1982 the Conference of European Posts and Telegraphs (CEPT) formed a study group called the Groupe Spécial Mobile (GSM) to develop a pan-European standard with the objectives of providing good subjective speech quality, support for international roaming etc. The proposed system was expected to meet certain criteria as mentioned below:

- Good subjective speech quality
- Low terminal and service cost
- Support for international roaming
- Ability to support handheld terminals
- Support for a range of new services and facilities
- Spectral efficiency
- ISDN compatibility

Subsequently, European Telecommunication Standards Institute (ETSI) published phase I of the GSM specifications in 1990 by 1993 there were 36 GSM networks were deployed in 22 countries and gradually over 200 GSM networks including DCS1800 and PCS1900 around the world. With the entry of North America, GSM systems now exist on every continent, and GSM is now known as Global System for Mobile communications. The analog cellular systems like AMPS in the United States and TACS in the United Kingdom were replaced as digital system keeping in view the advantages of digital system in compression algorithms and digital signal processing all over the world

In brief, we may now say that GSM is a digital mobile communications system based on European standard which has been defined within the framework of the European Telecommunications Standards Institute (ETSI), and in the meantime has been adopted by 396 network operators in 150 countries. It was designed to be compatible with ISDN systems and the services provided by GSM are a subset of the standard ISDN services (speech is the most basic).

General Features of GSM

GSM (Global System for Mobile Communications) is a second-generation (2G) digital mobile telephones standard using a combination Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) to share the bandwidth among as many subscribers as possible.

- GSM provides only 9.6 kbps data connection. Increase in data rates can be achieved when GSM changes into a radio service based on wide band code division multiple access, and not TDMA.

- GSM digitizes and compresses voice data, then sends it down a channel with two other streams of user data, each in its own time slot. It operates at either the 900, 1800 or 1,900MHz frequency bands.
- The uplink and down link frequencies for GSM are different and therefore a channel has a pair of frequencies 80 MHz apart. The separation between uplink and downlink frequencies is called duplex distance.
- In a channel the separation between adjacent carrier frequencies is known as channel separation which is 200 kHz in case of GSM.
- The services supported by GSM are telephony, fax and SMS, call forwarding, caller ID, call waiting and the like.
- GSM supports data at rates up to 9.6 kbps on POTS (Plain Old Telephone Service), ISDN, Packet Switched Public Data Networks, and Circuit Switched Public Data Networks.
- The access methods and protocols for GSM may be from X.25 or X.32.
- Being a digital system, GSM does not require a modem between subscriber and GSM network. However, an audio modem is required inside the GSM network to establish connection with POTS.

GSM Subscriber Services

There are three basic types of services offered through GSM. These are telephony or teleservices, data or bearer services and supplementary services. Bearer service provides the necessary data or control signal for different entities and interfaces. An interface is an access point between two entities, which will be described subsequently under GSM architecture. Telephony services comprise of voice services between different subscribers. The supplementary services are used to enhance the features of bearer and teleservices. In this way GSM provides the following services to subscribers:

- **Dual-tone multifrequency (DTMF)**— DTMF signals are tone signals used for various control purposes via the telephone network and are different from dial pulses. This can pass through the entire channels to the subscriber and therefore offer itself to various schemes for remote control after the connection is established. One example is the remote control of an answering machine.
- **Short message services (SMS)—** SMS is a message consisting of a maximum of 160 alphanumeric characters. SMS can be sent to or from a mobile station even when the subscriber's mobile station is powered off or has left the coverage area. In this case the message is stored and offered back to the subscriber when the mobile station is powered on or the subscriber has come back to the coverage area of the network.
- **Facsimile group III**— GSM supports CCITT Group 3 facsimile. A special fax converter in the GSM system enables a fax connected with GSM to communicate with any analog fax in the network.
- **Call forwarding—** This is a supplementary service, which enables a subscriber to forward incoming calls to another number on his own convenience.

The other services are cell broadcast, voice mail, fax mail, barring of outgoing and incoming calls conditionally, call hold, call waiting, conferencing, closed user groups etc.

GSM consists of many subsystems, such as the mobile station (MS), the base station subsystem (BSS), the network and switching subsystem (NSS) and the operation subsystem (OSS). Mobile station (MS), the base station subsystem (BSS) together forms a radio subsystem. Figure 2.18 shows the various is parts of GSM architecture, which has been further, explained in Figure 2.18.

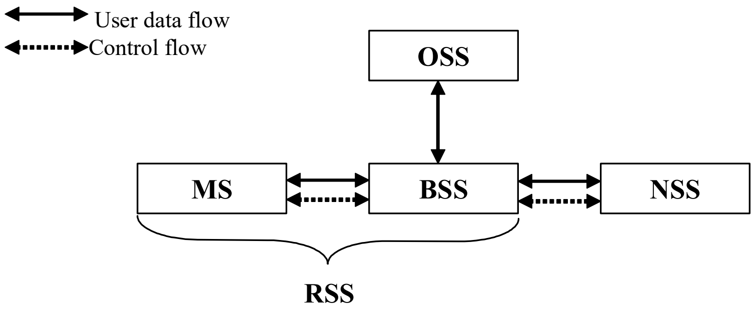


Fig. 2.18 An Overview of GSM Architecture

Architecture of the GSM network

The generic GSM network architecture is shown in Figure 2.19, which is composed of three subsystems as the radio subsystem (RSS), the network and switching subsystem (NSS) and the operation subsystem (OSS). The subscriber carries the Mobile Station, which is part of RSS.

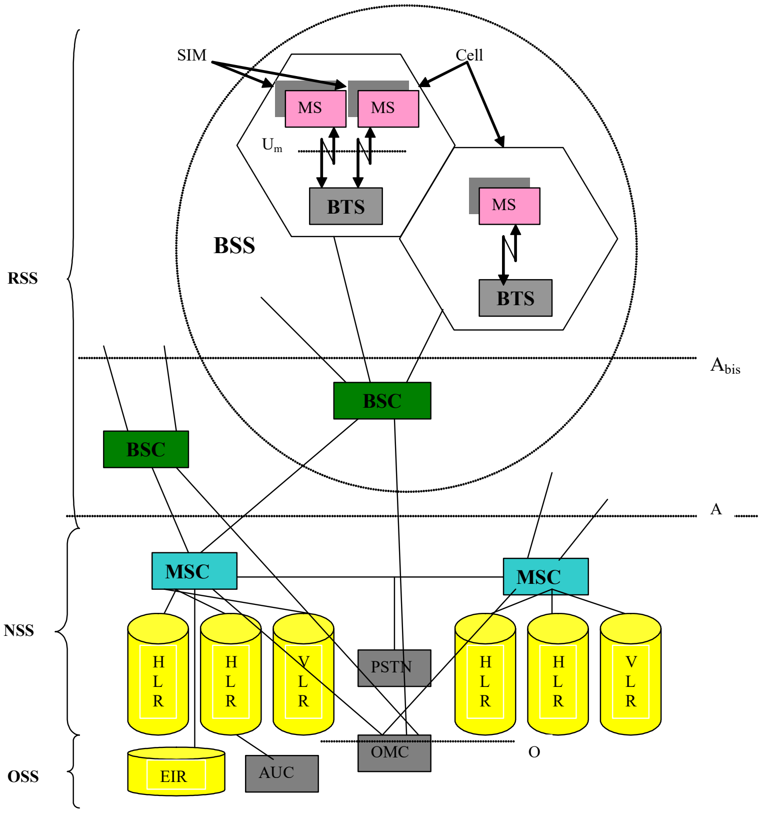


Fig. 2.19 Architecture of Mobile Communication System

NOTES

The RSS is basically consisting of radio specific equipment such as mobile station (MS), base station subsystem (BSS) to control the radio link. The connection between RSS and NSS is established with A interface based on circuit switched PCM-30 system with 2.048 Mbit/s date rate. The chief components of RSS are BSS, cellular layout and base station controller (BSC).

Radio Substation (RSS)

The RSS consists of the components that are necessary in order to allocate the radio resources to the individual subscribers. It principally consists of the mobile terminals (mobile phone or mobile station, MS) and the base station subsystem (BSS).

Base Station Subsystem (BSS)

The Mobile Station and the Base Station Subsystem communicate across the U_m interface, also known as the air interface or radio link. A BSS is controlled by a BSC as shown in Figure 2.19. A BSS maintains radio connections to an MS, coding/decoding of voice and data rate adaptation to/from the wireless network part. There may be many BSS in a GSM network and each BSS contains several MS, base transceiver station (BTS) and a base station controller (BSC) along with the cellular layout. These communicate across a standardized Abis interface, allowing operation between components made by different suppliers.

Mobile Station (MS)

As shown in Figure 2.19 that MS is basically mobile equipment which comprises of all user equipment and software needed for mobile communication and a smart card called the Subscriber Identity Module (SIM). Figure 2.20 (the handset) shows the MS, which contains a SIM card in the form of a very small chip inside the equipment.

Fig. 2.20 *Mobile Station/Cellular Phone*

The Subscriber Identity Module (SIM) contains all subscriber information necessary for identifying GSM subscriber. Broadly, it holds a subscriber’s International Mobile Subscriber Identity (IMSI), authentication key and algorithm. SIM is independent of the device or handset in which it is being used because an MS can be identified via International Mobile Equipment Identity (IMEI). As soon as the SIM is inserted into handset, it becomes immediately programmed for use. Therefore, it can be inserted into any handset. If you have forgotten to carry your handset but are carrying your SIM card, it can be inserted in any borrowed phone for use. Without SIM a handset can access only emergency services. Advances in memory and processing capacity has enabled SIM cards to be programmed to

NOTES

display custom menus for personalized services and therefore makes it different from conventional cellular phones. Typically mobile stations have transmitted power from 2W to 1W depending upon the cell size. If cell size is smaller, the transmitted power will be less.

Cellular Layout

Cells are the basic constituents of a cellular layout with cell sites. Cell site is defined as the location where base station and antennas are placed. A cell is simply represented by simple hexagon. The size of a cell may vary from ten of meters to tens of kilometers in a building to a city respectively. Factors for determining cell size basically look for numbers of users to be supported and multiplexing and transmission technologies. The size of cells, in case of Global System for Mobile Communication (GSM) and Personal Communication Service (PCS) are much smaller in the range of 10 Kms. There is a provision to provide umbrella cell, which is a large cell that includes several smaller cells. This avoids frequent handoffs for fast moving traffic.

Figure 2.21 represents the cells and cell sites within cellular layout. Cells are further split into sectors or individual areas to make them more efficient and to let them to carry more calls.

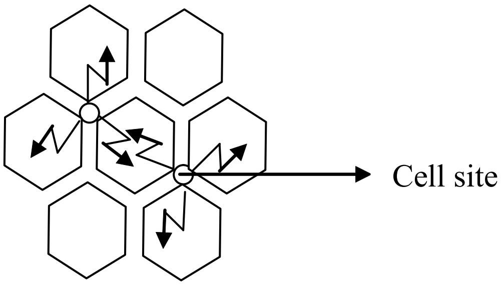


Fig. 2.21 *Cells and Cell Site within Cellular Layout*

As shown in the Figure 2.21 that antenna transmits inward to each cell and therefore covers a portion of each cell instead of the whole cell. The portions covered by the antenna are called sector.

Base Transceiver Stations (BTS) and Antennas

A base transceiver station is responsible to communicate with hosts in its cell by means of passing all calls coming in and going out of a cell site. The BTS or Base Transceiver Station is also called an RBS or Remote Base station. Each base transceiver station covers a certain transmission area, which is a cell, and it is allocated a portion of the total number of channels available. A group of nearby base transceiver stations forms a network and uses all available channels mutually. A network of radio BTS and antennas covers a large geographical area to cater large numbers of subscribers in the form of cells for analog and digital mobile communication and need to be rugged, reliable, portable with minimum cost. BTS are housed with all radio equipment such as antennas, signal processors, amplifiers etc to handle the radio-link protocols with the mobile station and for radio transmission.

Placement of base transceiver station is also an important issue. Figure 2.21 shows a cell site, which may considered as edge-excited cell. Base station may also be placed near center of cell and known as center excited cell. Figure 2.22 shows a centrally excited cell. The actual cell is the blue and red hexagon, with the towers at the corners, as has been depicted in Figure 2.22.

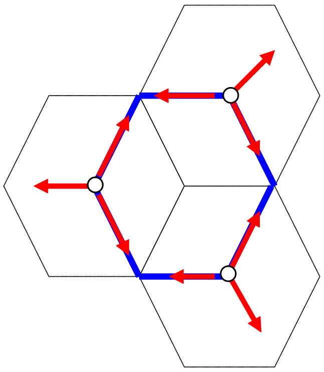


***Fig. 2.22** Centrally Excited Cell*

Frequency selection for each base transceiver station is a very important factor. Incorrect selection of frequency may generate interference with neighboring cells. The transceivers outside the interference range of other transceiver may reuse the frequencies being used by other transceiver.

Antenna always transmits inward to each cell and area served depends on topography, population, and traffic. Though theoretically these radiates equally in all directions however in reality, they have directive effects and therefore sectored antennas are used. Mobile stations (MS) communicate only via the BTS and a BTS is connected to a mobile station (MS) via U_m interface and A_{bis} interface with BSC as depicted in Figure 2.19. The U_m interface basically consists of FDMA, TDMA CDMA etc. required for wireless transmission. On the other side a BTS is connected to BSC via A_{bis} interface consisted of 16 or 64 kbits/s connections. As it is already mentioned that GSM utilizes Time and Frequency Division Multiple Access (TDMA/FDMA) to divide up the bandwidth among as many users as possible. The FDMA involves the division up to the maximum of 25MHz bandwidth into 124 carrier frequencies spaced 200kHz apart. One or more carrier frequencies are then assigned to each BTS depending upon the traffic. FDMA channels are further divided in time with a burst period of approximately 0.577ms, using a TDMA technique. The burst period is a fundamental unit of time in this TDMA technique. Subsequently, eight burst periods are grouped into a TDMA frame which forms the basic unit for the definition of logical channels. One physical channel is one burst period per TDMA frame.

A base transceiver station may have its own hierarchy with picocells covering building interiors, microcells covering selected outdoor areas, and macrocells providing more extensive coverage to wider areas as shown in Figure 2.23. As demand increases, more channels are needed and therefore number of base stations is increased and transmitter power is decreased correspondingly to avoid interference. Each BTS is connected to a BSC through fixed links where all traffic gets gathered first for the onward journey to MSC.

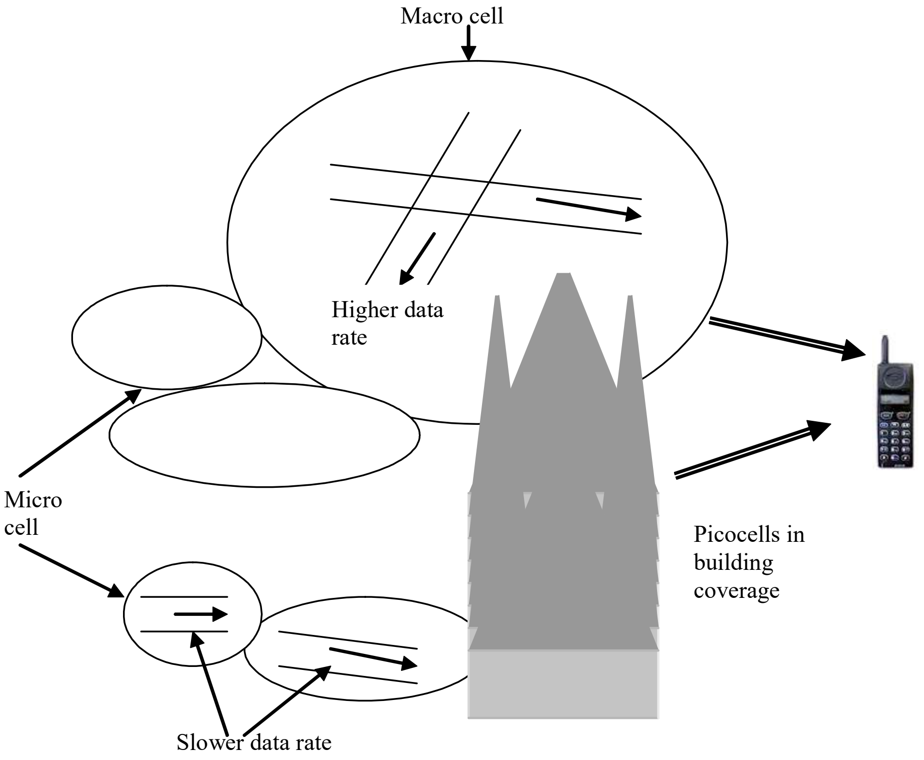


Fig. 2.23 BTS Own Cell Hierarchy

Base Station Controllers (BSC)

BSC is also a part of RSS and basically manages several base transceiver stations (BTS) at a time by providing a link between wireless devices such as cell phones, and the PSTN. It is the connection between the mobile station and the mobile services switching center. Base station controller is nothing but a high capacity switch. This switch performs control and management functions such as handover from one BTS to another within the BSS, management of radio network resources and handling of cell configuration data, control of radio frequency power levels, setting of transceiver configurations and frequencies for each cell etc. The number of BSC may vary from cell to cell depending on the complexity and capacity of a carrier's system. It also multiplexes the radio channels onto the fixed network connections. BSC also enables compression of traffic emerging from mobile phones even further by using a transcoder/rate adaptation unit which carries out encoding and speech decoding and rate adaptation for transmitting data. This an additional advantage of BSC as the voice coders in the handset have already compressed voice and data and puts the traffic into a format the Mobile Switch can understand. Figure 2.19 illustrates the position of BSC in GSM architecture.

Network and Switching Subsystem

The network switching subsystem (NSS) constitutes the fixed network component of the mobile radio telephone service network at one end and between the mobile radio telephone service network and other public networks on the other hand. It is composed of the Mobile Services Switching Center (MSC), the Home Location Register (HLR) and the Visitor Location Register (VLR).

Mobile Switching Center (MSC)

The other popular name for MSC is mobile switch (MS) and mobile telecommunications switching office (MTSO).

The MSC coordinates call set-up to and from GSM users. Each mobile switch manages dozens to scores of cell sites. Each MSC is connected to a base station at one end and to other MSCs and PSTN or ISDN on the other hand through fixed links. In this case each MSC acts as a local switching exchange that handles switching of mobile subscriber from one base station to another and locating the current cell of a mobile subscriber. Besides, it provides all the functionality such as registration, authentication, location updating, handover, and call routing to a roaming subscriber. MSCs thus form a fixed backbone for a mobile communication. An additional MSC may also be provided to establish link with other fixed networks. It is termed as Gateway MSC. MSC provides these services in conjunction with several functional entities, which together form the NSS. Therefore, the main tasks of MSC are entrusted upon as interworking functions (IWF), mobility management operations, data service units (DSU), SS7. Large systems may have two or more MSCs.

As shown in Figure 2.19 BSCs interact with a Mobile Services Switching Center (MSC), which is a high capacity ISDN switch, through an interface A. It basically gathers traffic from dozens of cells and passes it on to the public switched telephone network (PSTN). The mobile switch controls the entire network by interacting with distant databases and the public switched telephone network or PSTN. MSC also performs the administrative functions in the form of checking the validity of a user's account before letting a call go through, delivers subscriber services like Caller ID, and pages the mobile when a call comes in. It also handles all signaling needed for call set up, call release and handover of calls to other MSCs by using standard signaling system 7 (SS7).

Home location register and visitor location registers are the two chief databases. These interact with MSC for performing signaling and administrative functions and provide the call-routing and roaming capabilities of GSM.

Home Location Register (HLR)

The Home Location Register (HLR) which has its main task as association with MSC. The HLR consists of a database, which contains subscriber's all administrative information along with the current location in GSM network. The information includes subscriber's international mobile equipment identity (IMEI) number, directory number, current city, last visited area and the class of service subscriber has. The VLR temporarily contains administrative information that is relative to whatever mobile is currently in its area.

The Home Location Register (HLR) is an integral part of mobile communication, which keeps a profile of a user or a subscriber. HLR keeps the current location of each mobile that belongs to the MSC to which it is interacting. The subscriber's home service provider maintains it. Home Location Register performs the functions such as delivery of calls, information and messages to subscribers at their current locations by using user profile information. In case of a roaming user, the user's data is maintained at one location which makes database administrator functions easy. A roaming subscriber means that he is coming under different MSCs and the servicing MSC just collects information about the user from his home MSC. This facilitates in reducing administrative errors and duplication expenses.

The HLR maintains user information in the form of static and dynamic information. The static information is the International Mobile Subscriber Identity (IMSI), account status, service subscription information authentication key and options etc. The dynamic information is the current location area of the mobile subscriber which is the identity of the currently serving Visitor Location Register (VLR) to enable the routing of mobile-terminated calls. As soon as a mobile user leaves its current location area the information in the HLR is updated so that the mobile user can be localized in the GSM network. The HLR handles SS7 transactions with both Mobile Switching Centers (MSCs) and VLR nodes.

Visitor Location Register (VLR)

Visiting Location Register (VLR) as its name implies maintains temporary user information (such as current location area) falling under the other MSCs being visited by the subscriber or mobile user but whose Home Location Register (HLR) is elsewhere. VLR therefore controls and manages call request from subscriber who is out of the area covered by their home system and currently located in the geographical area controlled by the VLR. This is achieved by copying all relevant information for the subscriber from the its HLR. VLR also provides interfacing with PSTN as per the requirement. Therefore, its main tasks are association with MSC, IMSI, TMSI and roaming. This mechanism of maintaining VLR avoids frequent HLR updates and long distance signaling of subscriber information. When a call is initiated from the outside the subscriber home area, the MSC of that area VLR contacts the appropriate MSC using SS7 signaling which in turn relays the HLR information to the VLR. The VLR sends routing information back to the MSC under which the mobile subscriber is currently visiting and this creates a temporary record for the subscriber in that VLR.

In nutshell we can say that both the HLR and VLR work together to provide local connection as well as roaming outside the local service area.

Operation Subsystem (OSS)

The Operations and Maintenance Center oversees the all-important functions for proper operation and setup of the network and therefore provides Telecommunication Management Network (TMN). It also provides interface to NSS via O-Interface that may be X.25 Interface. The following parts have been defined:

Operation and Maintenance Center

Figure 2.25 shows the position of OMC in GSM network. It allows monitoring and controlling of the system as well as modify the configuration of the elements of the system. It provides all necessary information for controlling point of view and uses an interface O as shown in Figure 2.19. O interface is SS7 with X.25. It manages and controls the traffic load of the BSS. Basically, the management functions include traffic monitoring, status report of network elements, subscriber and security management. The subscriber and security management is accomplished through accounting and billing.

The Authentication Center (AuC)

The Authentication Center (AuC) is considered a subsystem of the HLR as shown in Figure 2.19 and is used for authentication and security by generating authentication algorithms, cryptographic codes etc. It is a secured database to protect the subscriber identity and data transmission as mobile station and interfaces are vulnerable from security point of view. It generates and sends a randomly generated number to the mobile for correct reply back. The SIM inside the mobile generates another number with the aid of the its encryption key and the received number from AuC. The MSC will proceed the call only when it receives the expected number back from mobile. In this manner a call is authenticated and therefore AuC is responsible for maintaining all data needed to authenticate a call and to then encrypt both voice traffic and signaling messages.

The Equipment Identification Register (EIR)

EIR fulfills the security and authentication requirement of GSM. It is a protected database for the subscriber and equipment identification number (IMEI) that contains a list of all valid mobile equipment on the network, where each mobile station is identified by its International Mobile Equipment Identity (IMEI). Each SIM card has a secret key for authentication and encryption over the radio channel. This made available to EIR. GSM network checks the type and serial number of a mobile device through EIR database and determines whether or not to offer any service. It also monitors and stops the use of the stolen mobile when the owner reports about theft. It also maintains a database of malfunctioning devices.

GSM System Radio Interfaces

As has been shown in the previous sections and Figure 2.19 that GSM architecture consists of a number of entities to handle the different requirements of call processing at different stages. A mobile station, which digitizes and encodes voice, requires an access to a BTS controlling a cell within cellular outlay. This BTS further gets connected to BSC, which is in turn connected to MSC to send the call outside the cellular outlay on MS either in the same network or on different network. The connection between different entities is achieved by using interfaces at each stage because the requirement of data at each stage is different from another stage and different equipment from any manufacturer will work together. This is a standardized method for passing information back and forth and can be a mechanical or electrical link connecting equipment together. We may now discuss the various interfaces one by one in the following sections.

The Radio Interface U_m

U_m is the radio link between a mobile station and a base station. The mobile station uses many techniques to create physical channels through FDM (Frequency Division Multiplexing) and TDM (Time Division Multiplexing). It uses FDMA/TDMA for accessing the cellular network where cells containing BTS are arranged using SDM techniques. GSM 900 operates on a frequency range of 890-915 MHz for uplink and 935-960 MHz for down link. It uses different frequencies for uplink and downlink to avoid interference and different power requirement at MS and BTS.

Frequency Division Multiplex (FDM) is used to divide the available frequency band in GSM. The available bandwidth is therefore 25 MHz, which is into 124 carrier frequencies, spaced 200 kHz apart using FDMA technique. Time Division Multiplex Structure (TDM) provides physical TDM channels, time slots and TDMA frame. The fundamental unit of time in this TDMA technique is called a burst period. Typically, GSM has many burst types such as normal burst, access burst, synchronization burst, frequency correction burst and dummy burst.

In the normal burst case, a burst period lasts 15/26 milliseconds (ms) or approximately 577 ms. Figure 2.24 describes GSM time slots for normal burst of 577 ms which includes 30.5 ms for guard space. Guard space is provided to avoid overlap with other bursts, different path delays and delay because of the transmitter on-off operation. Slots hold individual call information within the frame, that is, the multiplexed pieces of each conversation as well as signaling and control data.

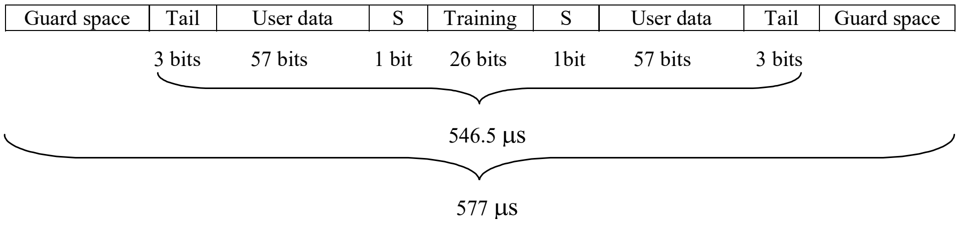


Fig. 2.24 GSM Time Slots for Normal Burst

Such types of eight burst periods are grouped into a TDMA frame of time duration of 4.615ms. This is the TDMA scheme. As has already been described that the available bandwidth of 25 MHz is divided into 124 carrier frequencies, spaced 200 kHz apart. Each 200 kHz carrier frequency contains eight such types of TDMA frames, each of 4.615 ms time duration that gives the concept of a channel. The repetition of one particular time slots every 4.615 ms makes up one basic channel. A base station may have one or more carrier frequencies. Each time slot consists of a voice channel. Channel comprises of a pair of radio frequencies for transmission and reception separately in cellular radio. Therefore, a channel is a dedicated time slot within a data or bit stream, which repeats after a certain period of time. Channels can further be divided into dedicated channels and common channels. Both the dedicated and common channels are allocated to a mobile station but the latter is used by mobile stations in idle mode.

A normal burst contains a training packet of 26 bits surrounded on both sides by two packets of 57 bits each of user data and two packets of 1bit each of stealing bit. Stealing bit is denoted as S in Figure 2.24. Thereafter, 3 tail bits are added on each side. The purpose of 26-bit training sequence is to reconstruct the original signal by comparing it with the received pattern. Training sequence is of a known pattern.

Logical Channels and Frame Structure

It is explained earlier that when a slot repeated every 4.615 ms constitute a physical channel which may be split into several (logical) channels. TDMA is used to split carrier frequency of 200 kHz into 8 time slots. Figure 2.25 explains a TDMA frame. These slots are known as 8 logical channels. A logical channel is therefore defined by its frequency and the TDMA frame time slot number. GSM 900 has 124 physical full duplex channels or 248 physical half-duplex channels. As there are 8 time slots for each physical half-duplex channel, hence 248 channel will have a total of 1984 logical half-duplex channels. In a cellular system, a cell can only use one seventh of the total number of frequencies therefore 283 (1984/7) logical half-duplex channels per cell are used effectively. The logical channel can be divided in traffic channels used for user data and signaling channels reserved for network management messages.

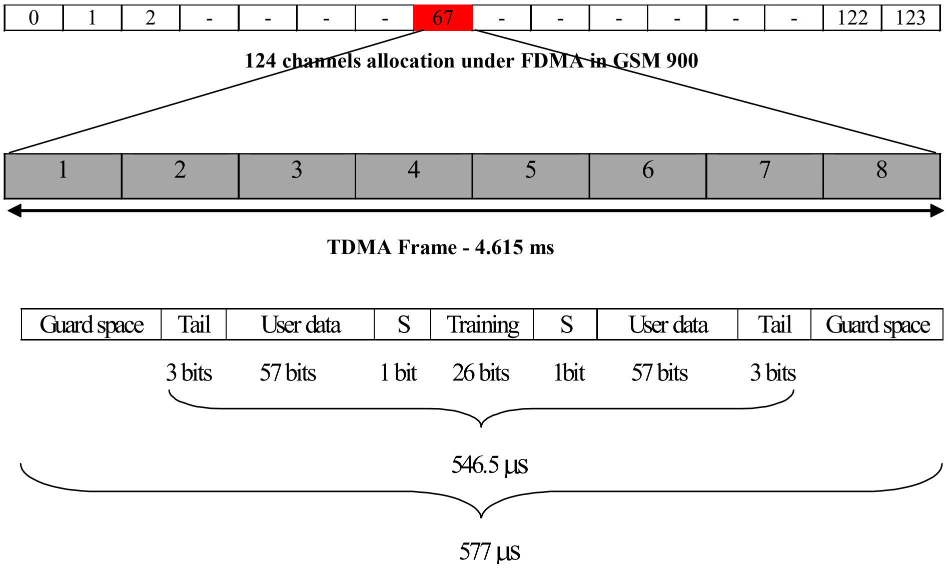


Fig. 2.25 TDMA Frames

A traffic channel (TCH) is defined for speech and data at the rates of 9.6kb/s, 4,8kpbs and 2.4kbps. It comprises of group of 26 TDMA frames which includes 24 frames for traffic, 1 frame for the Slow Associated Control Channel (SACCH) and remaining 1 frame is currently unused. The length of these 26 TDMA frames are kept 120 ms. This TCH frame structure defines the duration of a burst period. The half-duplex distance in time is 3 burst periods. Besides, these full-rate TCH, there are also half-rate TCH available to make the capacity of the system double

Control Channels (CCHs)

Control channels as its name implies are basically used to control the logical channels and takes different form depending upon the task assigned to these channels. It basically controls medium access, allocation of traffic channels or mobility management. Depending upon the task performed by these channels, they are categorized in three categories such as broadcast control channels (BCCH), common control channels (CCCH) and dedicated control channels (DCCH). These can be accessed both by idle mode and dedicated mode mobiles.

Broadcast Control Channel (BCCH)

BCCH is a unidirectional downlink point-to-multi-point-signaling channel from BTS to MS. It is used by BTS to broadcast control information to all MS in that particular cell about cell identifier, options available within this cell (frequency hopping) and the frequency available in the cell and cells in its surrounding. It has frequency correction channel (FCCH) and synchronization channel (SCH) as subchannels which are used to correct frequency and to synchronize the mobile to the time slot structure of a cell by defining the boundaries of burst periods and the time slot numbering respectively. Every cell in a GSM network broadcasts exactly one FCCH and one SCH.

Common Control Channel (CCCH)

CCCH is a bi-directional point-to-multi-point-signaling channel that exchanges the signaling information for network access management and transports information regarding connection setup between MS and BTS. BTS uses paging channel (PCH) to find out the location of MS by paging prior to downlink packet transfer. Whenever a mobile wishes to establish a call, it uses Random Access Channel (RACH) where mobile uses Slotted Aloha channel to request access to the network. Access grant channel is a downlink only, which replies to a random access channel and allocate an standalone dedicated channel (SDCCH) to a mobile for signaling with a low data rate of 782 bps in order to obtain a dedicated control channel for subsequent signaling.

Dedicated Control Channel (DCCH)

These are bi-directional and are multiplexed on to a standard channel for registration, location updating and authentication in order to set up a call or TCH. Each SDCCH and TCH is associated with slow associated dedicated control channels (SACCH) for control and supervisory signals associated with the traffic channels. GSM also uses Fast associated dedicated control channels (FACCH) which captures time slots from the TCH and are used for control requirements such as handover where MS and BTS need to be transferred large amount of data.

GSM specifies a multiplexing scheme to integrate several frames where a periodic pattern of 26 slots occurs in all TDM frames with a TCH. The combination of these frames is called traffic multiframe. Out of the 26 frames, 24 are used for traffic, one is used for the Slow Associated Control Channel (SACCH) and one is currently unused. It is already mentioned that the duration of one TDMA frame is 4.615 ms and therefore duration of a multiframe will be 120 ms. Likewise, control multiframe comprises of 51 TDMA frame with a duration of 235.4 ms. These multiframe with appropriate combination generates logical channel hierarchy such as superframe and hyperframe. A superframe is generated with the combination of 26 and 51 multiframes. 2048 superframes constitute a hyperframe.

Protocols

GSM has three functional layers. These are physical, data link and layer three in correspondence with OSI model. Figure 2.26 shows the functional layers of GSM. In OSI model, the lower three layers usually terminate in the same node but it is not true in case of GSM. In GSM the functionality is spread over distinct functional entities with standardized interfaces between them. For instance, the RR part of layer three is spread over the MS, BTS, BSC, and MSC. The three layers provide connectivity to BSSs, MSCs, and across MSCs.

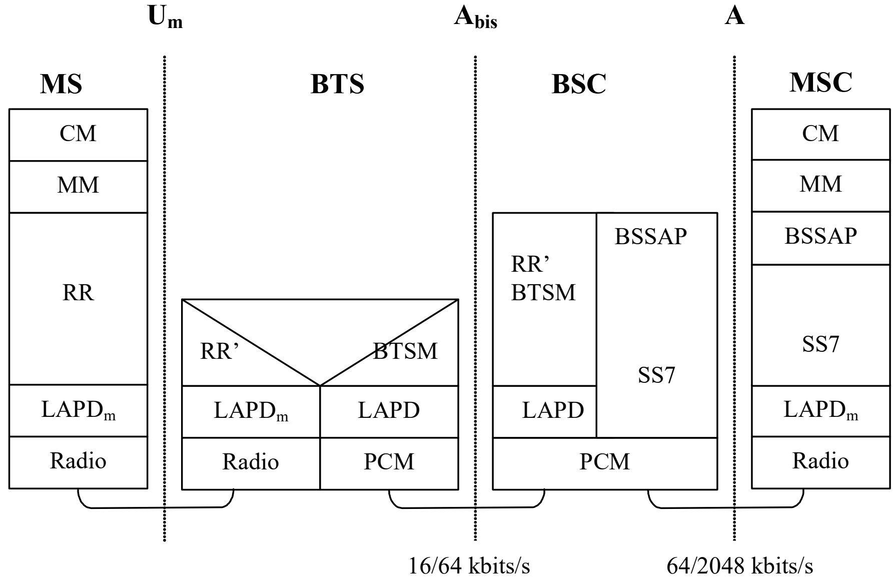


Fig. 2.26 Protocol Architecture for GSM

Physical Layer

The layer as shown as radio in Figure 2.26 is the lowest layer which provides transfer of bit streams over the physical radio links through U_m interface. It handles all radio specific functions such as creation of bursts, multiplexing of bursts into TDMA frame, synchronization with BTS, channel coding, error detection and correction and quality control on the downlink. The digital modulation and security related issues such as encryption of digital data are carried over the radio interface between MS and BTS.

Data Link Layer

It is required to introduce layer two as data link layer for signaling between different entities in a GSM network. A protocol for $LAPD_m$ is defined at layer two based on an adaptation of ISDN link access procedure for D channel (LAPD). Unlike LAPD, it does not require synchronization or checksumming for error detection, which is handled at physical layer. $LAPD_m$ provides a reliable dedicated signaling link between the MS and BSS on the air interface. The communication on A_{bis} interface between BTS and BSC is established by using the standard LAPD. A reliable data link service is provided between BSC and MSC through Message Transfer Part (MTP) of SS7. Layer 2 exploits MTP1 of SS7 for communication between different MSCs, from MSC to HLR and AUC and also connecting with PSTN via GMSC (Gateway MSC).

Layer 3

The layer three chiefly comprises of radio resource management (RR), mobility management (MM) and call control management (CM). Location update, authentication, and Temporary Mobile Subscriber Identity (TMSI) reallocation are the functions of mobility management. CM performs establishment, maintenance, and termination of a circuit- switched call. Other supporting-Supplementary Service (SS) support, Short Message Service (SMS) support

Radio Resource Management (RR)— The radio resource management sublayer (RR) terminates at the BSS and is used to establish physical connections over the radio for call-related signaling and traffic channels between the MS and BSS. RR manages and provides control functions for establishment, operation and release of a dedicated radio channel connection between MS and various BSCs for the duration of the call. This protocol also provides stable uninterrupted communications path between the MSC and MS over which signaling and user data can be conveyed. The RR' layer is the part of the RR layer is implemented in the BTS to provide functions between the BTS and BSC.

The RR tasks such as transferring the RR information to the BSC, which are not managed by the RR' protocol in the BTS, are handled by the Base Transceiver Station Management (BTSM). The BSSs have a LAPDm and RR layer to talk to the MS, but use a separate stack to communicate with MSCs. This stack consists of a Message Transfer Part (MTP) of SS7 and BSSAP and SCCP sublayer, which replace the RR layer on the MS. The BSSAP and SCCP sublayer implement call, resource, and signaling management and messaging between the BSS and MSC. The BSS Management Application Part (BSSMAP) protocols provide RR messages between the BSC and MSC. Signaling Connection Control Part (SCCP) is used above layer 2 between BSCs and MSCs and between MSCs and different databases such as HLR, AuC, etc.

The **handover or handoff** is also responsibility of the RR layers. The BSC, BTS, and MS control most of the functions, though some are performed by the MSC in particular for inter-MSC handoffs. The MSC or BSS uses signal strength measurements and cell congestion information to determine when a handoff should occur. Handoff notifications are sent to respective VLRs, which in turn forward them to HLRs.

Mobility Management (MM)— The mobility management sublayer (MM), on top of the RR, is terminated at the MSC and is used to establish, maintain, and release connections between the MS and the network MSC. It also takes care of maintaining the location data, in addition to the authentication and ciphering procedures.

The Communication Management (CM)— On top of MM lies the connection management sublayer (CM). CM protocol controls end-to-end call establishment and initiates calls setting up at the subscriber's request. Its functions are divided in three categories. These are call control, which manages the circuit-oriented services, supplementary services management, which allows modifications and checking of the supplementary services configuration and SMS, which provides point-to-point short message services.

There are some more protocols used in GSM network. These are Transaction Capabilities Application Part (TCAP) protocol and Mobile Application Part (MAP) protocol. TCAP sits above SSCP and supports transactions between 2 nodes of network to manage transaction on an end-to-end basis. MAP is used between MSC, VLR, HLR, and AuC in form of query and response messages.

Call Setup

The Home Location Register (HLR) and Visitor Location Register (VLR), together with the MSC, provide the call routing and roaming capabilities of GSM where a subscriber can roam nationally and even internationally. This is not possible in the fixed network, where a terminal is semi-permanently wired to a central office. Localization and calling system in GSM always knows where a subscriber is and his phone number remains valid worldwide. It requires frequent updates of the subscriber's whereabouts. HLR provides the current location and VLR keeps the track of MS and informs the HLR about the current location of MS and in turn HLR sends all subscribers data needed to new VLR. When a mobile subscriber roams into a new location area i.e. new VLR in terms of GSM technology, this VLR automatically determines that it must update the HLR with the new location information, which it does using an SS7 Location Update Request Message. The HLR responds with a message that informs the VLR whether the subscriber should be provided service in the new location.

This location update and call set up requires several numbers:

- **Mobile Subscriber ISDN (MSISDN) Number -** The MSISDN is the number that callers use to reach a mobile subscriber. This consists of country code (such as 91 for India), the national subscriber destination code and the subscriber number. The national subscriber destination (NSD) is the address of the GSM provider.
- **International Mobile Subscriber Identity (IMSI) Number –** This is a unique identification number allocated to each mobile subscriber independently of MS in the GSM system used internally and cannot be dialed. It sits inside SIM card, which can be carried over anywhere and can be used in any MS. Subscriber mobility is provided through mapping the subscriber to the SIM card rather than the terminal. The IMSI is made up of three parts. These are the mobile country code (MCC) consisting of three digits, the Mobile Network Code (MNC) consisting of two digits and the Mobile Subscriber Identity Number (MSIN) with up to 10 digits.
- **Temporary Mobile Subscriber Identity (TMSI) –** TMSI has valid temporarily and local significance only in the area handled by the VLR.

It is used in place of the IMSI for the definite identification and addressing of the mobile station and nobody can determine the identity of the subscriber by listening to the radio channel. It resides on the SIM card and VLR only and a VLR changes it regularly. It is not the part of HLR. GSM uses the 4 byte TMSI for local subscriber identification.

- **Mobile Station Roaming Number (MSRN) -** MSRN is a temporary location-dependent ISDN number assigned by the locally responsible VLR to each mobile station in its area. The calls are routed to the MS by using the MSRN and MSRN generated in VLR is passed from the HLR to the MSC on request. The MSRN contains the current visitor country code (VCC), the visitor national destination code (VNDC), the identification of the current MSC along with the subscriber number (SN).

GSM calls may be classified into two types. These are mobile terminated call (MTC) which may get originated either from MS or fixed line phone. Another is mobile originated call (MOC). The addresses as explained above are required in both the cases. When a subscriber either from a fixed line or MS calls to an MS by dialing GSM mobile subscriber's MSISDN, PSTN routes mobile terminated calls to the Gateway MSC which requests routing information for the caller from HLR and VLR by using the information in the MSISDN. The gateway MSC query the HLR based on the MSISDN to obtain routing information required routing the call to the subscribers' current location and therefore containing a table linking MSISDNs to their corresponding HLR. Once the address of the destination HLR is determined, the gateway MSC routes the call to the MSC in which the caller is currently roaming by sending a Routing Information Request to the HLR. HLR maps the MSISDN to the IMSI after having received the Routing Information Request to establish the identity of the subscriber and the current location of subscriber with the help of the current VLR address. The current VLR also provides the Mobile Station Roaming Number (MSRN) to HLR on query through which the mobile subscriber can be contacted and remains valid for the call duration. The HLR then identifies the correct MSC by identifying a temporary ID given to the mobile by the VLR for the purpose of anonymity.

In case of mobile originating call a MS requests BSS to set up a call with PSTN. BSS check ups with the MSC and connects the call.

Handover

Hand-over becomes necessary when mobile moves from area of one BSC into another area of the same or into another BSC. Handover involves a number of procedures depending upon the location. These are defined for each of the following cases:

- **Intra-cell handover—** It involves the transfer of connections from one channel to another channel on the same BTS.
- **Inter-cell, intra-BSC handover**— It involves the transfer of the connection from one BTS to another BTS on the same BSC.
- **Inter-BSC, intra-MSC handover—** The connection is transferred between BTSs belonging to two different BSCs within one MSC.

- **Inter MSC handover—** The connection is required to transfer to a BTS between two cells within another MSC.

As explained previously in RR under layer 3 that a BSC takes the decision to perform a handover. Either BSC or MS may assist the handover. In the case of BSC assisted handover, BSC monitors the signal level of the mobile and handover occurs if signal level falls below threshold. This increases load on BSC because it monitors signal level of each mobile and determines target BSC for handover. During a connection, BTS gets reports from the MS received signal level for all the BTSs it can receive using each SACCH frame after 480 ms. The reports about the received signal strength are usually forwarded directly to the BSC by BTS where it is analyzed using BTSM protocol and based on analysis, the BSC initiates the handover procedure. Thus MS immediately follows the command sent by the BTS to switch over to the new BTS and starts transmitting on the new channel. This completes the handover and release of the previous channel occupied by the MS.

In case of mobile assisted handover each BSC periodically transmits beacon to mobile. When mobile hears a stronger beacon from a new BSC, it sends an acknowledgement and changes routing tables to make new BSC its default gateway. New BSC acknowledges the MS and begins to route call at its new destination. In the areas controlled by other MSC's, the call handover is handled similar to mobile assisted case with additional HLR/VLR effort and local call may become long-distance call.

Check Your Progress

7. Which device is used to connect a coaxial cable and terminals?
8. What is the minimum allowable distance between consecutive connections?
9. Define radio.
10. Define GSM (Global System for Mobile Communications).

2.5 EXTENDLING LAN : TRANSMISSION CONCEPTS AND TERMS

Apart from the components needed by the conventional wired LAN, a wireless LAN needs additional components. They are the transmitters and receivers at Radio Frequency (RF) or InfraRed (IR). The RF transmitter and receivers need antennas to perform two-way communication. Usually a trial installation is carried out before the actual implementation. Hubs, bridges, network operating system, servers, and other components are function exactly as on a wired LAN.

Mobile Clients

Mobile clients are portable computing devices that act as clients. The following are some of the portable mobile systems.

- Laptop computers: Laptop PCs with two-way communication facility (transceiver)

- Palmtops or Personal Digital Assistants (PDAs) with communication capability
- Portable FAX
- Cellular phones

Special Units

For network management and efficient communication, a wireless LAN needs additional equipments. They are:

Communication units: These units communicate within the network and also with other networks.

Data collecting units: These units collect data from other systems.

Security units: These units take care of the network security.

Transceivers: A transceiver is a half-duplex device. It performs transmission and reception of data within a wireless LAN. It can transmit in one direction at a time.

Portable bridges: A portable bridge can support the Internet working functions. Two wireless LANs can communicate with each other using a bridge. It can be a transceiver or a satellite port or other communication unit that provides a bridge service.

Working of Wireless LANs

Wireless LANs use electromagnetic waves (radio or infrared technology) to communicate information from one point to another without relying on any physical connection. Radio waves are often referred to as radio carriers because they simply perform the function of delivering energy to a remote receiver. The data being transmitted is superimposed on the radio carrier so that it can be accurately extracted at the receiving end. This is generally referred to as modulation of the carrier by the information being transmitted. Once the data is superimposed (modulated) onto the radio carrier, the radio signal occupies more than a single frequency, since the frequency or bit rate of the modulating information adds to the carrier. Multiple radio carriers can exist in the same space at the same time without interfering with each other if the radio waves are transmitted on different radio frequencies. To extract data, a radio receiver tunes into one radio frequency while rejecting all other frequencies. In a typical wireless LAN configuration, a transmitter/receiver (transceiver) device, called an access point, connects to the wired network from a fixed location, using standard cabling. The access point receives, buffers, and retransmits data between the wireless LAN and the wired network infrastructure. A single access point can support a small group of users and can function within a range of less than one hundred to several hundred feet. The access point (or the antenna attached to the access point) is usually mounted high but may be mounted essentially anywhere that is practical as long as the desired radio coverage is obtained.

End-users access the wireless LAN through wireless LAN adapters, which are implemented as add-on cards in notebook or palmtop computers, as cards in desktop computers, or integrated within hand-held computers. Wireless LAN adapters provide an interface between the client Network Operating System (NOS) and the airwaves via an antenna. The nature of the wireless connection is transparent to the NOS.

NOTES

Transmission Media

Wireless LANs may use either radio wave technology or infrared technology (optical) for transmission of data. Each technology comes with its own set of advantages and limitations. The properties of these two technologies are discussed here.

Radio Wave Technologies

Radio waves propagate freely on air. They are used for many applications. Radio broadcast, television, telephony, and defence applications use radio waves. The band used for a specific application is highly significant and cannot be used for other applications. There are national and international agreements in the selection of a specific band for an application. Radio wave transmission and reception require highly sophisticated circuitry. Both the transmitter and the receiver must work within a short band. The following are the problems associated with radio frequency transmission.

Path loss: Signal-to-Noise Ratio (SNR) is defined as the ratio of power of the received signal to power of the noise in the received signal. The performance of the communication system is good if this factor is improved. But the design will be more complex if this parameter is to be improved. Increasing the transmitting power or reducing the distance between the transmitter and receiver can improve the SNR.

Adjacent channel interference: Interference is another phenomenon that affects the radio frequency transmission, when the same frequency band is allocated to two adjacent transceivers, resulting in interference. Hence, interference occurs when one useful signal is mixed up with another signal. This problem can be avoided by dividing the available band into sub-bands and allotting different bands to adjacent transceivers.

Multipath: Another problem with radio wave transmission is the multipath. A receiver at any point can get two types of signal from the transmitter. One is the direct signal and the other is the reflected signal. Every object reflects the radio wave. Hence, the receiver can get multiple reflected signals through various paths. The signal strength is additive at certain points and out of phase at some other points. Hence, the receiver can get peak power at some points and minimum power at some other points. This phenomenon is known as frequency selective fading. By employing two antennas at quarter wavelength separation, this problem can be solved.

Narrowband Technology

A narrowband radio system transmits and receives user information on a specific radio frequency. Narrowband transmission uses single frequency modulation, set up mostly in the 5.8 GHz band. The biggest advantage of narrowband systems is high throughput because they do not have the overhead involved with broadband systems. RadioLAN is an example of a system with narrowband technology.

Undesirable crosstalk between communications channels is avoided by carefully coordinating different users on different channel frequencies. In a radio system, privacy and non-interference are accomplished by the use of separate radio frequencies. The radio receiver filters out all radio signals except the one to which it is tuned. From the customer's point of view, one drawback with narrowband technology is that the user must obtain a license for the usage of the specific frequency.

Direct Sequence Spread Spectrum (DSSS) technology

Most wireless LAN systems use this technology in which more bandwidth is consumed compared to narrowband transmission. With direct sequence spread spectrum, the transmission signal is spread over an allowed band (for example, 25 MHz). A random binary string called a spreading code is used to modulate the transmitted signal. The data bits are mapped to a pattern of 'chips' at the source and mapped back into a bitstream at the destination. The number of chips that represent a bit is called the spreading ratio. The higher the spreading ratio, the more the signal is resistant to interference, at the expense of increased bandwidth. The Federal Communication Commission for international radio transmission recommends that the spreading ratio must be more than ten. Most products have a spreading ratio of less than twenty and the new IEEE 802.11 standard requires a spreading ratio of eleven. The transmitter and the receiver must be synchronized with the same spreading code. To an unintended receiver, DSSS appears as low-power wideband noise and is rejected (ignored) by most narrowband receivers.

Frequency-Hopping Spread Spectrum (FHSS) technology

This technique splits the band into many small sub-channels each of 1 MHz band. The signal then hops from sub-channel to sub-channel transmitting short bursts of data on each sub-channel for a set period of time called dwell time. The hopping sequence must be synchronized at the sender and the receiver, otherwise the whole information will be lost. Frequency hopping is less susceptible to interference because the frequency is constantly shifting. This makes frequency-hopping systems extremely difficult to intercept. This feature gives frequency-hopping systems a high degree of security. To jam a frequency hopping system, the whole band must be jammed. These features are very attractive to agencies involved with law enforcement or the military. To an unintended receiver, FHSS appears to be a short-duration impulse noise.

Infrared technology

A technology, little used in commercial wireless LANs, is infrared. Infrared has extremely high frequency, higher than radio wave range. They are in the frequency range of 10^{14} Hz and higher. Infrared technology is used in optical fibres, TV remote control, CD players, etc. IR systems are simple in design and therefore inexpensive. They use the same signal frequencies used on fibre optic links. IR systems detect only the amplitude of the signal and so interference is greatly reduced.

Characteristics of Infrared Transmission

Infrared systems need special infrared emitters and infrared detectors. Infrared transmission is performed in two ways. The first method uses the direct modulation and the second uses carrier modulation. The direct modulation scheme is described below as wireless LANs use only direct modulation scheme.

Direct Modulation

Direct modulation, often referred as on-off keying, is widely used in optical fibre systems. A light source, usually a Light Emitting Device (LED) is directly switched on by a binary 1 and switched off by a binary 0. The direct modulation system is similar to the one shown in Figure 2.27. The source *bit stream* is encoded, using a standard encoding technique prior to modulation. The encoded data is then modulated, using a modulator. Pulse position modulation or a similar modulation technique is employed to reduce the power requirements. Modulated signal is then fed to the LED device. At the receiving side, an optical *band-pass filter* is used to select the required band that contains the transmitted signal component. The photo-detector produces electrical signal, which is in the modulated form. A demodulator extracts the encoded data from this and a decoder recovers the data in the original form. Direct modulation is commonly used within a room or a small area where the transmitter and the receiver are in the line of sight.

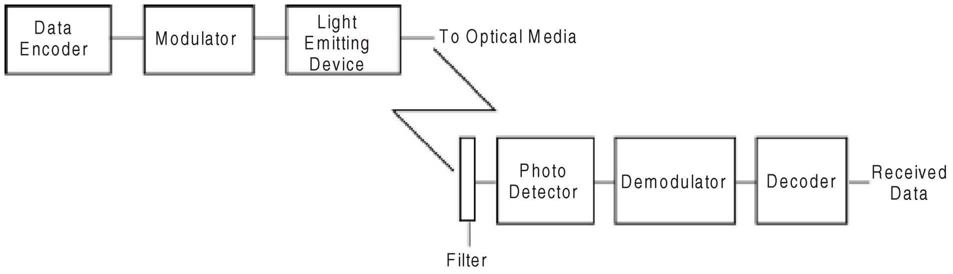


Fig. 2.27 *Optical Transmission System*

Operating Modes

Infrared links can be used in two different modes. They are direct (point-to-point) mode and diffuse (omnidirectional) modes. In a point-to-point mode, the light emitter is directly pointed to the detector. Hence, low power emitters or less-sensitive photo detectors can be used. This mode of operation is adequate for providing a direct wireless link between two portable devices. Direct systems give a range of a couple of kilometres and can be used outdoors. It also offers the highest bandwidth and throughput. High performance-directed IR is impractical for mobile users and is therefore used only to implement fixed sub-networks.

In the diffuse mode, the infrared light from the source is optically diffused to scatter the light to a wide area. Thus, this mode is suitable for broadcast operation. Omnidirectional IR systems provide very limited range and typically reduces the coverage range to 30 to 60 feet, and are occasionally used in specific wireless LAN applications. All the detectors within the room can receive the signal from one transmitter; each with varying phase. The phase variation is due to the variations of path length between the transmitter and the receiver. Multiple reflections of light also cause phase variation. This phenomenon is known as multipath dispersion. This problem does not affect the communication process much in a typical room environment. Signal rate up to 1 Mbps can be satisfactorily achievable. Beyond this rate, intersymbol interference causes the major problem.

Benefits and Drawbacks

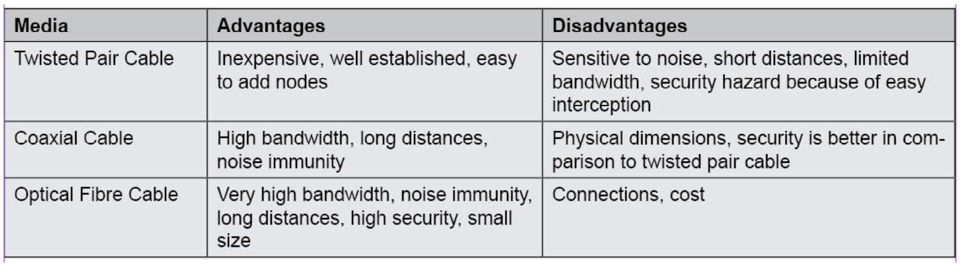
Benefits: IR systems are not bandwidth limited and thus can achieve transmission speeds greater than the other systems. Infrared transmission operates in the light spectrum and does not require a license from the Federal Communications Commission (FCC) to operate. IR technology was initially very popular because of its high data rates and relatively cheap price.

Drawbacks: The transmission spectrum of infrared system is shared with the sunlight and other sources, such as fluorescent lights. If there is enough interference from these sources, it can render the LAN useless. IR systems require an unobstructed line-of-sight. IR signals cannot penetrate opaque objects.

Wireless Transmission

Wireless transmission systems do not make use of a physical conductor, or guide to bind the signal. In this case, data are transmitted using electromagnetic waves. Therefore, they are also known as unguided or unbounded systems. Energy travels through the air rather than copper or glass. Hence the term radiated often is applied to wireless transmission. Finally, such systems employ electromagnetic energy in the form of radio or light waves that are transmitted and received across space, and are referred to as airwave systems. The transmission systems addressed under this category include microwave, satellite and infrared. There are different techniques to convert the data suitable for this mode of communication. Radio waves can travel through walls and through an entire building. They can travel for long distances using satellite communication or short distance using wireless communication. Radio waves need attention and caution when this technology is used for delivery of real time applications like multimedia contents because radio links are susceptible to fading, interference, random delays, etc.

Table 2.2 *Bounded Media Comparison Chart*



Media	Advantages	Disadvantages
Twisted Pair Cable	Inexpensive, well established, easy to add nodes	Sensitive to noise, short distances, limited bandwidth, security hazard because of easy interception
Coaxial Cable	High bandwidth, long distances, noise immunity	Physical dimensions, security is better in comparison to twisted pair cable
Optical Fibre Cable	Very high bandwidth, noise immunity, long distances, high security, small size	Connections, cost

2.5.1 Radio

It is a technique in which data is transmitted using radio waves and therefore energy travels through the air rather than copper or glass. Conceptually radio, TV, cellular phones, etc., uses radio transmission in one form or another. The radio waves can travel through walls and through an entire building. Depending upon the frequency, they can travel long distance or short distance. Satellite relay is the one example of long distance communication. Therefore, each frequency range is divided into different bands, which has a specific range of frequencies in the radio frequency (RF) spectrum. The RF is divided in different ranges starting from very low frequencies (VLF) to extremely high frequencies (EHF). Figure 2.28 shows each band with a defined upper and lower frequency limit.

NOTES

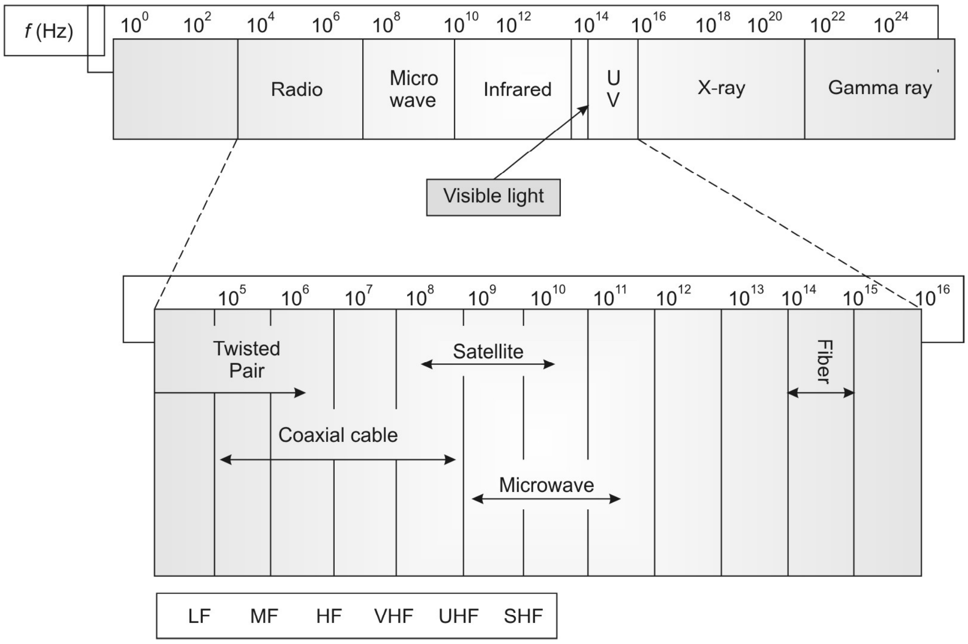


Fig. 2.28 *Radio Frequency Range and Types of Transmission Media*

Two transmitters cannot share the same frequency band because of the mutual interference and therefore band usage is regulated. The International Telecommunication Union (ITU) regulates international use of the radio spectrum. Domestic use of the radio spectrum is regulated by national agencies such as Wireless Planning and Coordination (WPC) in India. WPC assigns each transmission source a band of operation, a transmitter radiation pattern, and a maximum transmitter power. The Table 2.2 shows the bands and frequency ranges. Omni directional or directional antennas are used to broadcast radio waves depending upon band. The transceiver unit, which is consisted of transmitter and receiver along with the antenna, determines the power of RF signal. Other characteristics of radio waves is that in vacuum all electromagnetic waves or radio waves travel at the same speed, i.e., at the speed of light which is equal to 3×10^8 metre per seconds. In any medium this speed gets reduced and also becomes frequency dependent. In case of copper the speed of light becomes approximately two thirds of the speed of light. The basic features of the radio waves are as follows:

- they are easy to generate
- they have the same velocity in vacuum
- they may traverse long distances
- they are omni directional
- they can penetrate building easily so they find extensive use in communication both indoor and outdoor
- they are frequency dependent. At low frequency they can pass through obstacles. However, the power falls off sharply with distance from the source because power is inversely proportional to cube of the distance from the source. At HF they travel in straight lines and bounce off obstacles.

NOTES

2.5.2 Very Low Frequency (VLF)

The VLF method takes advantage of electromagnetic radiation generated in the low frequency band of 3-30 kHz by powerful radio transmitters used in long-range communications and navigational systems. At large distances from the source, the electromagnetic field is planar and horizontal and the electric component E lies in a vertical plane perpendicular to the H component in the direction of propagation and follow the ground. AM uses VLF band. This band of frequencies cannot be used for data transfer because they offer relatively low bandwidth.

2.5.3 Microwave Transmission

Microwave transmission is a form of radio transmission which uses extremely high frequencies. All the specified frequency ranges are in the GHz range and the wavelength in the millimetre range. Since these types of high frequency signals are prone to attenuation, hence, amplification is required after a specific distance. The radio beams are highly focussed in order to increase the transmission distance of the signals. The transmit antenna is centred in a concave metallic dish which focuses the radio beam with maximum effect, as illustrated in Fig. 2.29. Similarly the receiver dish is also concave in nature which collects the maximum amount of incoming signal.

It is a point-to-point transmission system, instead of a broadcast system. Also each antenna must be within the line of sight of the next antenna. Due to the curvature of the earth, the microwave signal hops are limited to a maximum of 80 km.

General Properties of Microwave Transmission

Configuration: Microwave radio consists of an antennae at the center of a reflective dish which is attached to the structure such as a tower or a building. Cables connect the antennae to the actual equipment.

Bandwidth: Bandwidth in excess of 6 Gbps is common in microwave transmission.

Error Performance: Assuming proper design, digital microwave performs well in this regard. However, environmental interferences such as, precipitation, haze, smog and smoke create troubles for high frequency transmission, yet microwave performs much better in this regard.

Distance: At higher frequencies, microwave is distance limited, which can be overcome through complex arrays of antennae incorporating spatial diversity in order to collect more signals.

Security: As is the case with all radio communication systems, microwave is inherently insecure, which can be improved through encryption.

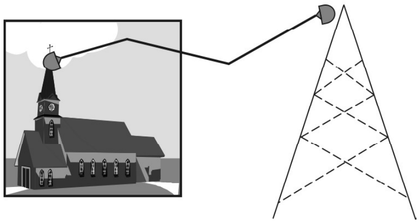
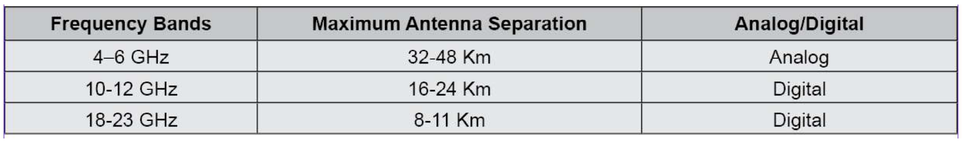


Fig. 2.29 *Point-to-point Microwave*

Table 2.3 *Microwave Frequency Bands*



Frequency Bands	Maximum Antenna Separation	Analog/Digital
4–6 GHz	32-48 Km	Analog
10-12 GHz	16-24 Km	Digital
18-23 GHz	8-11 Km	Digital

Cost: Even though the acquisition, deployment and rearrangement cost can be very high, yet, microwave compares very favourably with cabled systems, which require right-of-way, trenching and conduit and splicing.

Applications: Microwave was originally used for long haul voice and data communication since it was found to be the most attractive alternative to cable system. However the recent upsurge of fibre optic communication system is currently used in this regard. Contemporary applications include private networks, interconnections of cellular radio switches and an alternative of cabled communication system in difficult terrain.

2.5.4 Satellite Communication

Satellite radio is a non-terrestrial microwave transmission system utilizing a space relay station. Satellites have proved invaluable in extending the reach of voice, data, and video communications around the globe and into the most remote regions of the world. Exotic applications such as the Global Positioning System (GPS) would have been unthinkable without the benefit of satellites. Contemporary satellite communications systems involve a satellite relay station that is launched into a geostationary, geosynchronous, or geostatic orbit. Such satellites are called geostationary satellite. Such an orbit is approximately 36,000 kms above the equator as depicted in Fig. 2.30. At that altitude and in an equatorial orbital slot, the satellite revolves around the earth with the same speed as of that the speed of revolution of earth and maintains its relative position over the same spot of the earth’s surface. Consequently, transmit and receive earth stations can be pointed reliably at the satellite for communications purposes.

The popularity of satellite communications has placed great demands on the international regulators to manage and allocate available frequencies, as well as the limited number of orbital slots available for satellite positioning are managed at national, regional and international levels. Generally speaking, geostationary satellites are positioned approximately 2 apart in order to minimise interference from adjacent satellites using overlapping frequencies. Such high frequency signals are especially susceptible to attenuation in the atmosphere. Therefore, in case of satellite communication two different frequencies are used as carrier frequencies to avoid interference between incoming and outgoing signals. These can be listed as follows.

Uplink frequency: It is the frequency used to transmit signal from earth station to satellite. Table 2.3 shows the higher of the two frequencies that is used for the uplink. The uplink signal can be tailored stronger and therefore can better deal with atmospheric distortion. The antenna at transmitting side is centered in a concave, reflective dish that serves to focus the radio beam, with maximum effect, on the receiving satellite antenna. The receiving antenna, similarly, is centered in a concave metal dish, which serves to collect the maximum amount of incoming signal.

NOTES

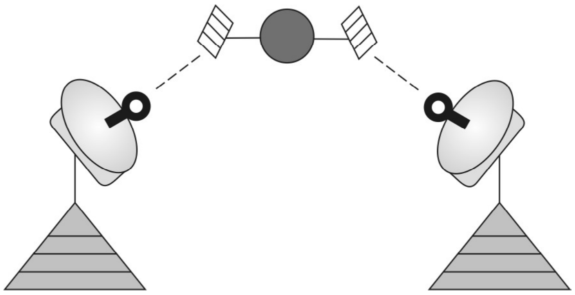


Fig. 2.30 *Satellites in Geostationary Earth Orbit*

Downlink frequency: It is the frequency used to transmit the signal from satellite to earth station. In other words, the downlink transmission is focused on a particular footprint, or area of coverage. The lower frequency, used for the downlink, can better penetrate the earth's atmosphere and electromagnetic field, which can act to bend the incoming signal much as light bends when entering a pool of water.

Broadcast: The wide footprint of a satellite radio system allows a signal to be broadcast over a wide area. Thereby any number (theoretically an infinite number) of terrestrial antennae can receive the signal, more or less simultaneously. In this manner, satellites can serve a point-to-multipoint network requirement through a single uplink station and multiple downlink stations. Recently, satellites have been developed which can serve a mesh network requirement, whereby each terrestrial site can communicate directly with any other site. Previously, all such communications were required to travel through a centralized site, known as a head end. Such a mesh network, of course, imposes an additional level of difficulty on the network in terms of management of the flow and direction of traffic.

General Properties of Satellite Communication

Configuration: Satellite communication systems consist of antennae and reflective dishes, much as in terrestrial microwave. The dish serves to focus the signal from a transmitting antenna to a receiving antenna. The send/receive dishes that make up the earth segment are of varying sizes, depending on power levels and frequency bands. They generally are mounted on a tripod or other type of brace, which is anchored to the earth, pad or roof, or attached to a structure such as building. Cables connect the antennae to the actual transmit/receive equipment. The terrestrial antennae support a single frequency band for example, C-band, Ku-band or Ka-band. The higher the frequency bands the smaller the possible size of the dish. Therefore, while C-band TV dishes tend to be rather large, Ku-band DBS (Direct Broadcast Satellite) TV dishes tend to be very small. The space segment dishes are mounted on a satellite, of course. The satellite can support multiple transmit/receive dishes, depending on the various frequencies which it employs to support various applications, and depending on whether it covers an entire footprint or divides the footprint into smaller areas of coverage through the use of more tightly focused spot beams. Satellite repeaters are in the form of number of transponders. The transponders accept the weak incoming signals, boost them, shift from the uplink to the downlink frequencies, and transmit the information to the earth stations.

Bandwidth: Satellites can support multiple transponders and, therefore, substantial bandwidth, with each transponder generally providing increments in bandwidth.

Error Performance: Satellite transmission is susceptible to environmental interference, particularly at frequencies above 20 GHz. Sunspots and other types of electromagnetic interference affect satellite and microwave transmission. Additionally, some satellite frequency bands, for example, C-band needs careful frequency management. As a result of these factors, satellite transmission often requires rather extensive error detection and correction capabilities.

Distance: Satellite is not considered to be distance limited as the signal largely travels through the vacuum of space. Further each signal travels approximately 36,000 kms in each direction.

Propagation Delay: Geostationary satellites, by virtue of their high orbital altitude, impose rather significant propagation delay on the signal. Hence, highly interactive voice, data, and video applications are not effectively supported via two-way satellite communications.

Security: As is the case with all microwave and other radio systems, satellite transmission is inherently not secure. Satellite transmission is especially vulnerable to interception, as the signal is broadcast over the entire area of the footprint. Therefore, the unauthorized user must know only the satellite and associated frequency range being employed. Security must be imposed through encryption (scrambling) of the signal.

Cost: The acquisition, deployment, and rearrangement costs of the space segment of satellite systems can be quite high in several millions dollars. However, the satellite can be shared by a large number of users, with each user perhaps connecting a large number of sites. As a result, satellite networks often compare very favorably with cabled systems or microwave systems for many point-to-multipoint applications.

Applications: Satellite applications are many and increasing rapidly as the traditional voice and data services have been augmented. Traditional international voice and data services have been supplanted to a considerable extent by submarine fibre optic cable system. Traditional, applications include international voice and data, remote voice and data, television and radio broadcast, maritime navigation, videoconferencing, inventory management and control through VSATs, disaster recovery and paging. More recent and emerging applications include air navigation, Global Positioning Systems (GPS), mobile voice and data because of Low Earth Orbit Satellites (LEOs), Advanced Traffic Management Systems (ATMS), Direct Broadcast Satellite (DBS) TV, Integrated Digital Services Network (ISDN), interactive Television, and interactive multimedia.

Very Small Aperture Terminals (VSATs): VSATs or Very Small Aperture Terminals are a breed of satellite system involving terrestrial dishes of very small diameter (aperture). Operating in the C-band and Ku-band, VSATs are digital and are designed primarily to support data communications on a point-to-multipoint basis for large private networks in applications such as retail inventory management and credit verification and authorisation. While some newer systems also support mesh networks and voice communications, they are unusual at this time. Bandwidth is in channel increments of 56/64 Kbps, generally up to an aggregate bandwidth of 1.544 Mbps.

2.6 MASTER SITES AND INTER CONNECTION TO TELEPHONE

NOTES

In telecommunications, the term 'Interconnection' is the physical linking of a carrier's network with equipment or facilities not belonging to that network. The term may refer to a connection between a carrier's facilities and the equipment belonging to its customer, or to a connection between two or more carriers. In the United States regulatory law, interconnection is specifically defined as, "The linking of two or more networks for the mutual exchange of traffic". One of the key tools used by regulators in the field of telecommunications markets is to impose interconnection requirements on dominant carriers.

Currently the standard electrical connector for interconnection in the world is the registered jack family of standards, especially RJ11 (Registered Jack 11). This was introduced by the Bell System in the 1970s, following a 1976 Federal Communications Commission (FCC) order. Since then, it has gained popularity worldwide, and is a de facto international standard. A Registered Jack (RJ) is a standardized telecommunication network interface for connecting voice and data equipment to a service provided by a local exchange carrier or long distance carrier. Registration interfaces were first defined in the Universal Service Ordering Code (USOC) system of the Bell System in the United States for complying with the registration program for customer-supplied telephone equipment mandated by the Federal Communications Commission (FCC) in the 1970s. The specification includes physical construction, wiring, and signal semantics. Accordingly, registered jacks are primarily named by the letters RJ, followed by two digits that express the type. Additional letter suffixes indicate minor variations. For example, RJ11, RJ14, and RJ25 are the most commonly used interfaces for telephone connections for one-, two-, and three-line services, respectively.

The communications infrastructure of the Internet consists of its hardware components and a system of software layers that control various aspects of the architecture. As with any computer network, the Internet physically consists of routers, media (such as, cabling and radio links), repeaters, modems, etc. The Internet carries many applications and services, most prominently the World Wide Web (WWW), including social media, electronic mail, mobile applications, Internet telephony, file sharing, and streaming media services.

Network Switching Subsystem (NSS) or GSM (Global System for Mobile Communications) core network is the component of a GSM system that carries out call out and mobility management functions for mobile phones roaming on the network of base stations. It is owned and deployed by mobile phone operators and allows mobile devices to communicate with each other and telephones in the wider Public Switched Telephone Network (PSTN).

Wireless communication is the transfer of information between two or more points that do not use an electrical conductor as a medium by which to perform the transfer. The most common wireless technologies use radio waves. With radio waves, intended distances can be short, such as a few meters for Bluetooth or as far as millions of kilometres for deep-space radio communications. It encompasses various types of fixed, mobile, and portable applications, including two-way radios, Cellular Telephones, Personal Digital Assistants (PDAs), and wireless networking. Other examples of applications of radio wireless technology include GPS units, wireless computer mouse, keyboards and headsets, headphones, radio receivers, satellite television, broadcast television and cordless telephones.

Cellular Phones

The development of cellular phones is recent one. This is also known as mobile phone and as its name implies it is designed for mobile users who need to make telephone calls from different locations when they are usually away from home or office. The rapid development in hardware technology helps in designing such kind of portable telephone sets so that user may carry it within their office bag or pocket during movement. Cellular phone uses radio frequencies to establish access to a nearby cell site which is an access point for cellular calls. The cellular phone regularly communicates with the nearest cell site to inform the network that it is connected.

Cell Site

This may be defined as a circular geographical area that handles cellular phones within its defined physical boundary. A cellular network as shown in Fig. 2.31 is considered consisting of overlapped cells so that a larger area with low probability of call dropping may be provided. This overlapping structure helps in keeping the call intact as a user moves location from one cell site to another. In this case, the call is transferred to the nearest cell site responsible for that physical area. Cellular telephones are suitable for larger geographical areas including remote sites. It saves the cost of copper wire and efforts in laying the same in densely populated areas. Each cell site shown in Fig. 2.31 is connected to a master site, which acts as an access point for a particular cellular network. Master site furnishes an interconnection to the regular telephone network. Calls handled by each cell site are relayed back to the master cell site, which then relays it to the telephone network as shown in Fig. 2.32.

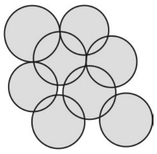


Fig. 2.31 *Cellular Networks Consisting of Individual Cells*

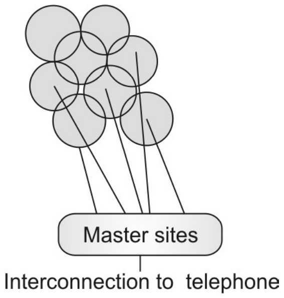


Fig. 2.32 *Cellular Network Connections to Telephone Exchange*

The forward cell can reuse frequencies used in the previous cell. This helps in sharing the same frequency band. Many calls can be handled by one frequency especially where digital phones are used.

NOTES

Satellite Cellular Telephone

This works on the same principle as cellular phones but uses Low Earth Orbiting (LEO) satellites. The advantages of satellite cellular phone may be seen in its capability to cover much wider geographical area. This is particularly a good technology in mountainous terrain and at sea. Unlike to cellular phones, satellite cellular phone requires a large number of cells and their accurate positioning to avoid blind spots. Blind spots are the spaces where no cell overlapping or cell is present therefore no call can be made at such spots.

Check Your Progress

11. What is a PSTN or public switched telephone network?
12. What is Path loss signal-to-noise ratio (SNR)?
13. What do you understand by the term interconnection?

2.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A Local Area Network (LAN) is a network that is restricted to a relatively small area.
2. The components used by LANs can be categorized into hardware, cabling standards, and protocols.
3. The two LAN transmission options, baseband and broadband. Baseband LANs, which is the most prevalent by far, is a single-channel system that supports a single transmission at any given time. Broadband LANs support multiple transmissions via multiple frequency channels.
4. Baseband LAN is a single channel connection, supporting a single communication at a time.
5. NIC is also known as Network Interface Unit (NIU). NIC is a hardware card to provide physical access from a node to the LAN medium. The NIC can be fitted into the expansion slot of a PC, or it can exist as a separate box.
6. A LAN Operating System, or Network Operating System (NOS), is software that provides the network with multi-user, multitasking capabilities and supports communications and resource sharing.
7. A transceiver is used to connect a coaxial cable and terminals. A transceiver cable (also referred to as an AUI (Attachment Unit Interface) cable) is used to connect a transceiver and the NIC.
8. The minimum allowable distance is 0.5 metre between consecutive connections.
9. The term radio may be defined as consisting of modulation and radiation of the signal.
10. GSM (Global System for Mobile Communications) is a second-generation (2G) digital mobile telephones standard using a combination Time Division

NOTES

Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) to share the bandwidth among as many subscribers as possible.

11. PSTN or public switched telephone network relates to the public telephone network. It is based on circuit-switched connection and can be compared to the Internet terms, referring to a public IP network based on a packet-switched connection.
12. Path loss signal-to-noise ratio (SNR) is defined as the ratio of power of the received signal to power of the noise in the received signal. The performance of the communication system is good if this factor is improved.
13. The term 'Interconnection' is the physical linking of a carrier's network with equipment or facilities not belonging to that network. The term may refer to a connection between a carrier's facilities and the equipment belonging to its customer, or to a connection between two or more carriers.

2.8 SUMMARY

- Local Area Network technology connects people and machines within a site. A Local Area Network (LAN) is a network that is restricted to a relatively small area.
- A LAN is a form of local (limited-distance), shared packet network for computer communications.
- LANs are used almost exclusively for data communication over relatively short distances such as within an office, office building or campus environment.
- There are two LAN transmission options, baseband and broadband.
- Broadband LANs are multichannel, analog LANs. They are typically based on coaxial cable as the transmission medium, although fibre optic cable is also used.
- Baseband LAN is a single channel connection, supporting a single communication at a time. They are digital in nature.
- NIC is a hardware card to provide physical access from a node to the LAN medium. The NIC can be fitted into the expansion slot of a PC, or it can exist as a separate box.
- A LAN Operating System, or Network Operating System (NOS), is software that provides the network with multi-user, multitasking capabilities and supports communications and resource sharing.
- The transceiver exchanges data signals handled by the NIC and electric signals sent over a transmission line.
- Hubs receive signals through one port and send them through all other ports.
- The maximum allowable length of a segment for the 10Base5 is 500 metre.
- The number of repeaters that can be used is limited. In the case of data communication between terminals, data can be sent through only up to four repeaters.

NOTES

- Wireless LANs offer the obvious advantage of avoidance of cabling costs, which can be especially important in a dynamic environment where there is frequent reconfiguration of the workplace.
- WLAN is mostly a mix of wire and wireless media having an access point or wireless router that is connected to a wired network via a coaxial cable, universal serial bus (USB), or Ethernet connection.
- A cellular system is the communications systems that divide a geographic region into sections, called cells, each having its own dedicated frequency.
- In a channel the separation between adjacent carrier frequencies is known as channel separation which is 200 kHz in case of GSM.
- The generic GSM network architecture is composed of three subsystems as the radio subsystem (RSS), the network and switching subsystem (NSS) and the operation subsystem (OSS).
- Cells are the basic constituents of a cellular layout with cell sites. Cell site is defined as the location where base station and antennas are placed.
- A base transceiver station is responsible to communicate with hosts in its cell by means of passing all calls coming in and going out of a cell site.
- The network switching subsystem (NSS) constitutes the fixed network component of the mobile radio telephone service network at one end and between the mobile radio telephone service network and other public networks on the other hand.
- The Authentication Center (AuC) is considered a subsystem of the HLR.
- EIR fulfills the security and authentication requirement of GSM. It is a protected database for the subscriber and equipment identification number (IMEI) that contains a list of all valid mobile equipment on the network
- It is required to introduce layer two as data link layer for signaling between different entities in a GSM network.
- Hand-over becomes necessary when mobile moves from area of one BSC into another area of the same or into another BSC.
- In telecommunications, the term 'Interconnection' is the physical linking of a carrier's network with equipment or facilities not belonging to that network.
- The communications infrastructure of the Internet consists of its hardware components and a system of software layers that control various aspects of the architecture.

2.9 KEY TERMS

- **Local Area Network (LAN):** A form of local (limited-distance), shared packet network for computer communications**.**
- **Baseband LAN:** It is a single-channel system that supports a single transmission at any given time.
- **Broadband LAN:** It supports multiple transmissions via multiple frequency channels.

- **Network Interface Card (NIC):** It is a hardware card to provide physical access from a node to the LAN medium.
- **LAN Operating Systems**: It is software that provides the network with multi-user, multitasking capabilities and supports communications and resource sharing.
- **Transceiver:** It is used to exchange data signals handled by the NIC and electric signals sent over a transmission line.
- **The Subscriber Identity Module (SIM):** It contains all subscriber information necessary for identifying GSM subscriber.
- **Common Control Channel (CCCH):** It is a bi-directional point-to-multi-point-signaling channel that exchanges the signaling information for network access management and transports information regarding connection setup between MS and BTS.
- **Public Switched Data Network (PSDN):** It is a network that is accessible to the public. It assists packet-switched data as well as PSTN.
- **Microwave Transmission:** It is a form of radio transmission which uses extremely high frequencies. All the specified frequency ranges are in the GHz range and the wavelength in the millimeter range.
- **Uplink Frequency:** It is the frequency used to transmit signal from earth station to satellite.
- **Downlink Frequency:** It is the frequency used to transmit the signal from satellite to earth station.
- **Cell Site:** It is defined as a circular geographical area **that handles cellular phones within its defined physical boundary.**

2.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Give the advantages as well as disadvantages of broadband LAN.
2. What are broadband LANs? Write its characteristics.
3. Which is the foremost and most popular transmission media for LAN. Discuss.
4. What is a transceiver? How does it works?
5. What are the basic wireless principles?
6. Define hub.
7. State the 5-4-3 rule.
8. Why fibre-optic cable is used in LANs?
9. What is a WLAN?
10. Write a note on cellular layout in data communication.
11. What are Base Station Controllers (BSC)?

12. What are the advantages and disadvantages of telephone networks?
13. What are the benefits and drawbacks of infra-red technology?
14. What are the general properties of microwave transmission?
15. Discuss the general properties of infrared transmission.

Long-Answer Questions

1. Briefly describe LAN Operating Systems.
2. Explain the10Base5 (Thick Net/Yellow Ethernet) in detail.
3. Describe the implementation of LAN using optical fiber cable pair.
4. Analyze the basic types of services offered through GSM.
5. Explain the architecture of the GSM network with appropriate diagrams.
6. Discuss the radio interface (Um) in detail.
7. Explain the OSI model used in GSM.
8. Write the detailed note on:
   (i) Infrared technology
   (ii) Microwave transmission
   (iii) Satellite transmission
9. What do you understand by interconnection to telephone? Explain.

2.11 FURTHER READING

Forouzan, Behrouz A. *Data Communications and Networking*. New Delhi: Tata McGraw-Hill, 2004.

Stallings, William and Richard Van Slyke. *Business Data Communications*. New Jersey: Prentice Hall, 1998.

Black, Uyless. *Computer Networks*. New Jersey: Prentice Hall, 1993.

Stallings, William. *Data and Computer Communications*. New Jersey: Prentice Hall, 1996.

Tanenbaum, Andrew S. *Computer Networks*. New Jersey: Prentice Hall PTR, 2002.

Stallings, William. *Data and Computer Communications*. NJ: Prentice-Hal, 1996.

UNIT 3 DATA TRANSMISSION NETWORK, TCP/IP AND OSI MODEL

NOTES

Structure

3.0 INTRODUCTION

Data transmission is the transfer of data from one digital device to another. This transfer occurs via point-to-point data streams or channels. These channels may previously have been in the form of copper wires but are now much more likely to be part of a wireless network. The effectiveness of data transmission relies heavily on the amplitude and transmission speed of the carrier channel. Network congestion, latency, server health, and insufficient infrastructure can decrease data transmission rate.

Network architecture is the design of a computer network. It is a framework for the specification of a network's physical components and their functional organization and configuration, its operational principles and procedures, as well as communication protocols used.

In telecommunication, the specification of a network architecture may also include a detailed description of products and services delivered via a communications network, as well as detailed rate and billing structures under which services are compensated.

The OSI Model (Open Systems Interconnection Model) is a conceptual framework used to describe the functions of a networking system. The OSI model characterizes computing functions into a universal set of rules and requirements in order to support interoperability between different products and software.

Queueing theory is the mathematical study of waiting lines, or queues. A queueing model is constructed so that queue lengths and waiting time can be predicted. Queueing theory is generally considered a branch of operations research because the results are often used when making business decisions about the resources needed to provide a service.

In this unit, you will lean about data transmission network, telephone networks, WAN technologies, network architectures and OSI model, routing and congestion control and queuing theory.

3.1 UNIT OBJECTIVES

After going through this unit you will be able to

- Understand data transmission system
- Understand the telephone networks
- Explain the OSI model for telephone networks
- Comprehend WAN technologies
- Explain TCP/IP model
- Discuss the services of internet
- Explain standards for TCP/IP
- Discuss open systems interconnection (OSI) model
- Understand the need for OSI model
- Interpret routing and congestion control
- Define routing and its need.
- Discuss the strategies for routing
- Explain general principles of congestion control
- Define deadlock
- Explain queuing theory and its models

3.2 DATA TRANSMISSION NETWORK

From the advent of the human race, wanting to correspond gave way to the progress of various methods and practices on the basis of situations and technology that is available. The initial types of communication were signs, gestures and writings that were illustrated on the caves, walls, etc. When language was developed usage of symbols, papyrus and paper made it easy to capture communication for future use. The continuous desire for communication past the physical boundaries encouraged man to use diverse practices. A few of these practices were formed on the basis of using gestures during storytelling, sound and animation.

Claude Elwood Shannon in the year 1948, worked for the Bell Telephone Company in the United States of America. Figure 3.1 displays the model of communication he proposed. This has become the basis of explanation of communication since then.

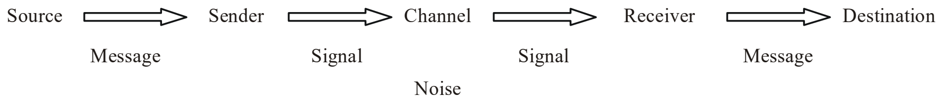


Fig. 3.1 *Shannon's Model of Communication*

The model applied, is based on oral communication between two people, is as follows:

Source	—	The brain
Message	—	The idea, thought
Sender	—	The transmitting device, the mouth
Channel	—	The medium the message travels over: air
Receiver	—	The receiving device: the ear
Destination	—	The brain

In any form of communication, the message is affected by the message as it moves across from the sender to the receiver in the channel. Data communications is about transmitting information between two locations. The transmission broadly involves sending and receiving the information. Information is thus, sent between machines connected with each other by physical wires or radio links. The machines may be transmitter, telephone, computer, etc.

Many jobs that were done on a centralized computer based on time sharing can now be done on standalone Personal Computers (PCs). Large number of dispersed users can share database located at a central place or at remote locations in an N-tier environment. This is the reason why the growth of data communication facilities is taking place together with the use of PCs so that a computer communication facility can be established in network form.

A data communication system is a computer system that collects data from remote locations through data transmission circuits, then outputs processed data to remote locations. A data communication system consists of data terminal equipment, a data communication circuit and an information processed unit. A data communication circuit transmits information input from the data terminal

NOTES

equipment to the remote informationprocessing unit or transmits processing results to the data terminal equipment. The information processing unit processes the data. This is shown in Figure 3.2. Some examples of data transmission circuits are telephone network, leased line, ISDN, packet switched network, frame relay, cell relay, etc.

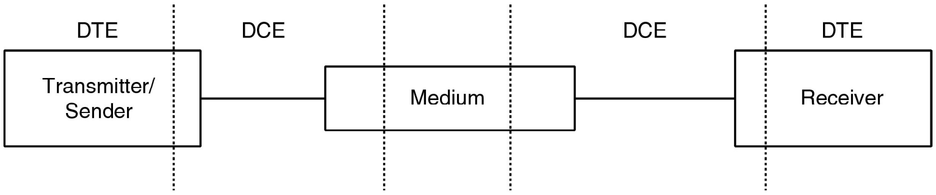


Fig. 3.2 *Data Communication Systems*

In data communication system, data is transmitted from terminals to the information processing unit through data communication circuits. These are two types of data transmission methods that are used to transmit data from its origin to the information processing. These are as follows:

1. **Offline:** Computers are not connected by communication circuits. Data is transmitted between a terminal and information processing unit through a magnetic tape and magnetic disk packs.
2. **Online:** Computers are connected by communication circuits. Data can be instantly transmitted between a terminal and information processing unit.

Components of Data Communication System

The data communication system consists of the following:

1. **Transmitter or Sender of Data:** These may be terminals, computers and mainframes, etc.
2. **Medium:** The medium, through which the data is transmitted, can be cables, Radio Frequency (RF) wave, microwave, fibre optics, infrared, etc.
3. **Receiver:** As the name implies, it is the device, which receives the data transmitted. These are printers, terminals, mainframes, computers, cell phone, etc.

In the Figure 3.2, the transmitter — medium interface and medium receiver interface have been shown by dotted lines. The transmitter may be a device which transmits signal in such a format that is not compatible with the medium. Similarly, medium provides signal in unacceptable format to receiver. Hence, the signal from transmitter to medium and medium to transmitter require conversion of signal from one form to another as per the requirement.

3.2.1 Data Communication Equipment

Data circuit terminating equipment is also known as Data Communication Equipment (DCE). DCE is the equipment that interfaces the source with the medium and vice versa. DCE includes modems, DSUs and CSUs and Front End Processors (FEPs). Each device is located at both ends of a communication circuit.

DCE works as follows:

- In a data station, the equipment that performs functions, such as signal conversion and coding, at the network end of the line between the Data Terminal Equipment (DTE) and the line and that may be a separate or an integral part of the DTE or of intermediate equipment.
- The interfacing equipment that may be required to couple the DTE into a transmission circuit or channel and from a transmission circuit or channel into the DTE.
- DCE is a device that communicates with a DTE device in RS-232C communications.
- Usually, the DTE device is the terminal or computer and the DCE is a modem.
- When two devices that are both DTE or both DCE that must be connected together without a modem or a similar media translator between them, a NULL modem must be used.

3.2.2 Data Terminal Equipment

Data Terminal Equipment (DTE) is the equipment which is a data communication system terminal that inputs and outputs data. In general, data terminals have a human machine interface. A typical example of DTE is an Automated Teller Machine (ATM) at a bank. In other words, DTE is the computer transmitting and receiving equipment, including a wide variety of dumb terminals (terminals without embedded intelligence in the form of programmed logic), intelligent terminals and in the form of host computers, such as mainframes and minicomputers.

DTE works as follows:

- An end instrument that converts user information into signals for transmission or reconverts the received signals into user information.
- The functional unit of a data station that serves as a data source or a data sink and provides for the data communication control function to be performed in accordance with link protocol.
- The DTE may be a single piece of equipment or an interconnected subsystem of multiple pieces of equipment that perform all the required functions necessary to permit users to communicate.
- A user interacts with the DTE or the DTE may be the user. The DTE interacts with the DCE.
- Usually, the DTE device is the terminal or computer and the DCE is a modem.

3.2.3 Communication Software

Now, in wider sense we may understand that a transmitter or sender may be a terminal (computer) responsible for with communication and application software controls the terminal and processes data. There may be more than one terminal connected to the sender.

Communications software is generally embedded in the computer operating system. Alternatively, it can take the form of a systems task under the control of the computer's operating system. The role of communications software is to assist the operating system in managing local and remote terminal access to host resources, to manage security and to perform certain check point activities.

NOTES

Figure 3.3 explains more clearly the components of data communication system with devices as DTE, DCE and medium. DTE, DCE and medium as depicted in Figure 3.2 have been replaced by computer with communication software, Modulator DEModulator (MODEM) and telephone line, respectively.

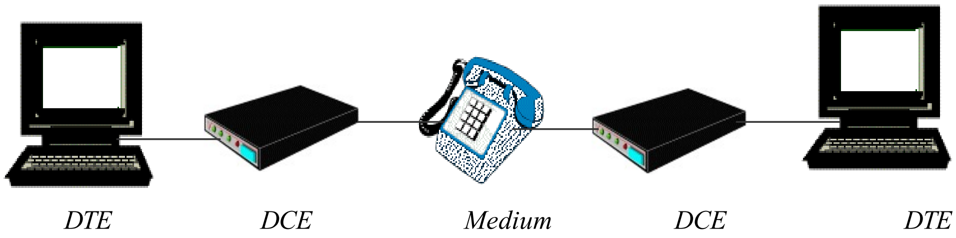


Fig. 3.3 *Data Communication System with Interfaces*

3.3 TELEPHONE NETWORKS

The earliest electronic network is the telephone system. This is shown in Figure 3.4. This telephone network commonly uses analog technology that was quite different from digital technology used in the computer-based networks. The advantages of digital technology over the analog technology in terms of economics and services forced the telephone industry to move rapidly to install fiber and digital networks. The telephone network transmits analog signals and hence a modem is required whenever a computer or terminal is connected to the telephone line as shown in Figure 3.4. The modem then converts digital data from a computer to an analog signal that can be transmitted via a telecommunication line and converts the analog signal received to computer data.

3.3.1 Dial up Telephone Networks

The telephone network consists of the subscriber’s line, switchboards, and trunk lines as shown in Figure 3.4. Each subscriber line has an address i.e. telephone number. When a caller transmits a dial signal to the switchboard, the switchboard connects the caller’s subscriber line to that of the receiver, enabling communication. The trunk line between the caller and the receiver is occupied until either discontinues the communication.

When the telephone system is to connect with a network, it becomes necessary to dial the telephone number to select the target device on the network as shown in Figure 3.4. A device called Network Control Unit (NCU) performs this, and most of the available modems, include this NCU.

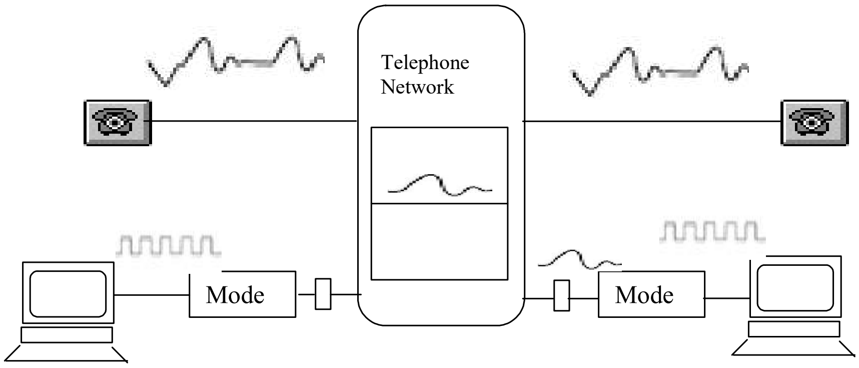


Fig. 3.4 Telephone Network

NOTES

The Computer Communication System – An example

The Computer communication system is an example of a system using the telephone Network as shown in Figure. The system is used to send and receive mail, connection to Internet, if the account is TCP/IP, post messages on a Bulletin Board system by accessing the host computer system of a ISP through telephone network.

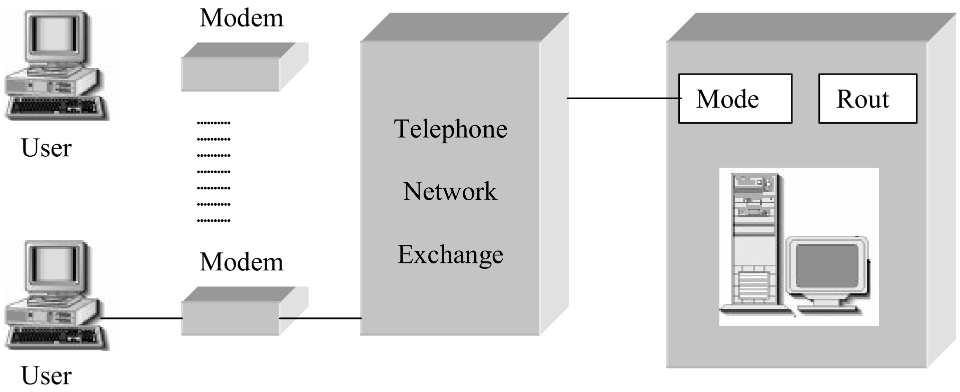


Fig. 3.5 An Example

3.3.2 Advantages and Disadvantages of Telephone Networks

Advantages

- It is circuit-switching network, therefore, any receiver can be selected and there is virtually no transmission delay.
- As it is widely spread therefore it is available at a low price.

Disadvantages

- It requires a long time for connection. A dial-up operation is necessary before the line can be connected to the receiver. This dial-up time is too long to use in data communication systems.
- It has low transmission speed.
- The line quality is not sufficient for data transmission, and is therefore not appropriate for high-speed data transmission because telephone lines were originally developed for audio communication.

3.3.3 Telephone Network Standards

The V Series Recommendations from the ITU-T include the most commonly used modem standards and other telephone network standards. Prior to the ITU-T standards, the American Telephone and Telegraph Company and the Bell System offered its own standards (Bell 103 and Bell 212A) at very low transfer rates. Another set of standards, the Microcom Networking Protocol, or MNP Class 1 through Class 10 (there is no Class 8), has gained some currency, but the development of an international set of standards means these will most likely prevail and continue to be extended. (Some modems offer both MNP and ITU-T standards.)

In general, when modems handshake, they agree on the highest standard transfer rate that both can achieve.

3.3.4 Leased Lines

A computer can be connected permanently to the Internet using leased lines as shown in Figure 3.6 in addition to a modem and router. These lines are based on speed of the connection, installation cost, and recurring monthly charges.

An example of usage of leased line is a system in which only one terminal is connected to the host computer. Though multiple computers/terminals using multiplexing can be connected to one system via a single leased line. It uses FDM for an analog leased line or TDM method for digital leased line. DSU (Digital Service Unit) unit is used instead of modem for digital line. Leased lines may also be used to connect LANs.

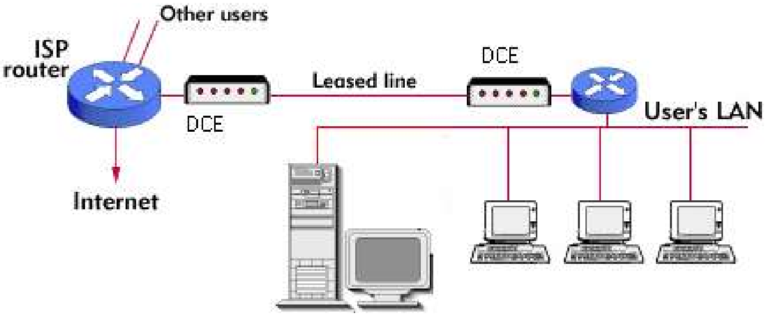


Fig. 3.6 *Leased Line Configuration*

Telephone networks are intended to transmit analog signals. It uses layer 1 (Physical Layer) of the OSI model as shown in Figure 3.7. It is the service to provide physical media. Therefore, telephone networks can carry any type of protocol data. The data transmission speed depends on the performance of the modem and quality of the line.

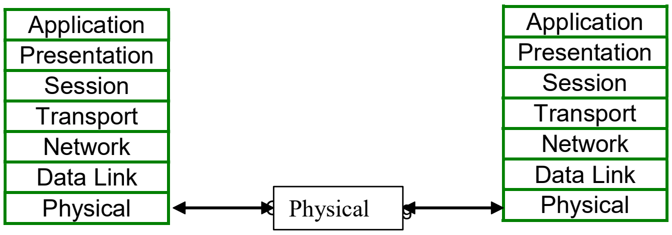


Fig. 3.7 *OSI Model for Telephone Network*

3.3.5 Public Switched Telephone Network (PSTN)

NOTES

PSTN or Public Switched Telephone Network relates to the public telephone network. It is based on circuit-switched connection and can be compared to the Internet terms, referring to a public IP network based on a packet-switched connection. The term PSTN was initially used for fixed-line analog telephone system but nowadays, due to the advancement in technology, it is also referred for digital circuit- switched telephone network including both mobile and fixed. ITU-T technical standard and an addressing rule (telephone number) E.163/E.164 are followed by the PSTN.

PSTN is the global compilation of interconnects made for assisting circuit-switched voice communication. The conventional Plain Old Telephone Service (POTS) is provided by PSTN to dwellers and to various enterprises.

Some of the DSL, VoIP and other Internet-based network technologies also make use of some parts of PSTN .

Almost 64 Kbps bandwidth is supported by the basic PSTN network link. The PSTN lines, in case of residences, are in the form of copper cables transferring the data in such a bandwidth. The dial-up modems make use of 56 Kbps of the total bandwidth while joined to the phone line. The Signalling System#7 (SS7) signalling protocol is used by the PSTN.

The evolution PSTN has gone from analog technology to digital technology. In analog technology, the data delivery is directly based on the accessible data. Contrary to that, the digital technology Involves sending data after it has been manipulated into the digital format. DSL, ISDN, FTTX and cable modem systems are some forms of digital PSTN.

PSTN require 64kbps channel as the vital digital circuit which also known as digital signaling 0/DS0. SS7 is used as a communication protocol between telephone exchanges by which the calls are routed to the destination. Being a circuit switch-base communication protocol, SS7 includes all the accessible resources which are used by a single dedicated call connection.

A limitation to the PSTN capacity is DS0, as it is a natural Time Division Multiplexing (TDM) that implies that every call data mix with one another that is time-based. In PSTN, the delivery is usually done through multiplexing of various DS0 together with DS1 for optimizing the transmission. DS1 can be sub-divided into two parts, namely, 24 DS0, also called as T1 that is located in North America or Japan and 32 DS0, also called as E1 that is in most of the other countries. Both T1 and E1 are known as the Transmission type. The hardware of PSTN can handle only one Transmission type due to which it always requires a hardware base that needs to be bought on the transmission plan.

ISDN and other non-PSTN services have comparatively more speed and acquire features due to which they are more preferable for using the Internet. For instance, wile using a non-PSTN service like ISDN or DSL, voice and data can be used simultaneously with the use of only one line instead of getting another phone line for accessing Internet which is the case with other services.

3.3.6 PSDN

Public Switched Data Network (PSDN) is a network that is accessible to the public. It assists packet-switched data as well as PSTN.

Earlier PSDN was termed as PSS (Packet Switch Stream) that was a X.25-based packet-switched network. The basic purpose of PSS was to present leased-line connections between LANs and also the Internet with the help of PVCs (Permanent Virtual Circuits). Now, as technology is advancing day by day, PSDN is not only limited to frame relay and ATM (Asynchronous Transfer Mode) that are as providers of PVCs, but also extended to various other packet switching methods like IP, GPRS, etc.

By watching the working of PSDN, one may consider it to be a replica of the data networks such as ISDN (Integrated Services Digital Network), ADSL (Asymmetric Digital Subscriber Line), SDSL (Symmetric Digital Subscriber Line) and VDSL (Very-high bitrate DSL). However, a closer study of PSDN shows that it is a lot more than these. The PSTN circuit switched network is used by ISDN whereas, DSL is point-to-point circuit mode communication services imposed over the PSTN local loop copper wires, commonly used for entry to a network of packet switched broadband IP.

3.3.7 ISDN: Broadband Communications

ISDN which is short for Integrated Services Digital Network is a set of CCITT/ ITU standards used for digital transmission over ordinary telephone copper wire and other media. This technology uses ISDN adapters in place of modems and provides very fast speed up. ISDN requires adapters at both ends of the transmission.

In reality, a widespread network with the potential to deliver at high data rates is required to deliver multimedia. Currently, ISDN is implemented in the form of the narrow band. This is the best medium available for access and delivery. Many in the industry consider ISDN as the tool for promoting multimedia, a channel through which multimedia will gain acceptance. The governments of various countries are coming out with plans and policies to implement ISDN as soon as possible.

Integrated Services Digital Network in concept is the integration of both analog or voice data together with digital data over the same network. ISDN integrates these on a medium that is designed for analog transmission. However, broadband ISDN (BISDN) will extend the integration of both services throughout the rest of the end-to-end path through fibre optic and radio media. Broadband ISDN will comprise frame relay service for high-speed data capable of being sent in large bursts, the Synchronous Optical Network (SONET) and the Fibre Distributed Data Interface (FDDI). BISDN will support transmission from 2 Mbps and much higher but unspecified rates.

Definition of ISDN

ISDN is a network architecture in which digital technology is used to convey information from multiple networks to the end-user. This information is end-to-end digital.

Features

- Offers point-to-point delivery.
- Network access and network interconnection for multimedia.
- Different data rates from 64 Kbps up to 2 Mbps are commercially available which can meet many needs for transporting multimedia and is four to many times more than today's analogue modems.
- Call set-up times are under one second. ISDN can dramatically speed up transfer of information over the Internet or over a remote LAN connection, especially rich media like graphics, audio or video or applications that normally run at LAN speeds.
- ISDN will be the feeder network for broadband ISDN based on ATM standards.

Although ISDN could be cheaper, particularly in the case of widespread use, it is likely to be cheaper than ATM connections and more widespread in availability for a long time. It is, therefore, an important tool in bringing multimedia applications to a wide range of users.

There are two forms of ISDN service: narrow band and broad band.

Narrow band ISDN

Narrow band ISDN is digital service where the transport speeds are 1.544 Mbps (T1) or less. Narrow band ISDN provides for the following services:

- **Circuit Switched Voice** — Circuit switched voice service is a digital voice service that offers many of the capabilities of a business. It is centred over a 4-wire ISDN Digital Subscriber Line (DSL).
- **Circuit Switched Data** — Circuit switched data service provides end-to-end digital service to pass data or video information over the public network. ISDN uses out-of-band signalling to establish and maintain data connections, which require special processing.
- **Low Speed Packet** — ISDN lines are equipped with a packet connection that is used to manage ISDN connections. This monitoring capability is provided by using the D channel on a DSL. The D channel is a 16 Kbps X.25 connection that is also capable of passing low speed packet while also relaying call processing information.
- **High Speed Packet** — ISDN lines are also equipped with two B channels. Each B channel is a 64Kbps channel that can be used for circuit switched voice, circuit switched data, or high-speed packet service. To provision high-speed packet service one or two of the 64 Kbps B channels are connected (permanent virtual circuit) to the packet network thus providing a 64 Kbps X.25 connection.

Broadband ISDN service

Broadband ISDN Service is a digital service in excess of 1.544 Mbps. This digital service can be in the form of Frame Relay, SMDS, or ATM. Broadband ISDN is the service of the future. The higher speeds offered are required to support the

NOTES

many applications of the Information Super Highway. The range of speeds for the Broadband ISDN services usually range from 25 Mbps up to the Gigabit range. The two speeds that are most often discussed are OC 1 that is 155 Mbps and OC 3 that is 622 Mbps. The speeds in the Broadband are made possible by the high quality of the digital facilities in place on the network. The early data protocols such as X.25 required extensive overhead to insure the delivery of data. Error correction and flow control were performed at a number of intermittent points along the way of a data connection. The new digital facilities and the introduction of fibre optics have eliminated this need up to a maximum extent. High-speed broadband services rely for the most part on the upper layer protocols to perform these functions on an end-to-end basis.

B(bearer) channels and a D (delta) channel. Voice, data and other services are carried by B channels while control and signalling information is carried by the D channel.

Basic Rate Access or BRA offers an ISDN user simultaneous access to two 64 Kbps data channels using the existing twisted pair copper telephone cable as shown in Figure 3.8.

- Each data channel is considered a B-channel and is capable of carrying voice or data. D-channel is another channel that operates at 16 Kbps and is used to signal between user devices and the ISDN. Therefore, 144 Kbps is the total data rate of BRA. The term 2B+D arises from the two B-channels and the single signaling channel. BRA is also referred to as I.420 following the CCITT recommendation. Basic rate ISDN is meant for low capacity usage as is needed by small businesses.
- **Basic Rate Access (BRA)** provides an ISDN user with simultaneous access to two 64 Kbps data channels using the existing twisted pair copper telephone cable as shown in Figure 3.8. Each data channel is referred to as a B-channel and can carry voice or data. Another channel, the D-channel, operates at 16 Kbps and is used for signaling between user devices and the ISDN. The total data rate of BRA is therefore 144 Kbps. The two B-channels and the single signaling channel give rise to the term 2B+D. BRA is also referred to as I.420, after the CCITT recommendation. Basic rate ISDN is intended for low capacity usage, such as that required for small businesses.

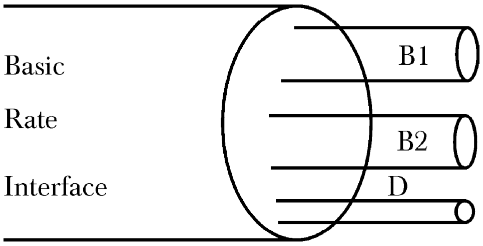


Fig. 3.8 *Basic Rate Interface*

- **Primary rate access** service provides up to 30 independent 64 Kbps B channels and a separate 64 Kbps D channel to carry the signaling. This basically provides digital access via a T1 line as shown in Figure

NOTES

3.9. A T1 line provides a 1.544 bandwidth. This bandwidth is divided into twenty-four 64Kb channels. The ISDN PRI service uses 23 of the T1 channels to provide B channel access and uses the 24th channel for signaling purposes. ISDN call control procedures use packet messages to initiate, monitor, and release connections. In a BRI connection these messages are routed via the D channel. On the PRI service the / connection/call control messages are routed over the 24th channel which is the D channel in this instance. The total data rate of PRA is 1.544 Mbps. Primary rate access is often referred to as 23B+D because of the number of B-channels and D-channels, or I.421 because of the CCITT recommendation from which it is taken. This form of access is primarily intended for use in situations which require a large transmission capacity, such as when organizations make voice and data calls through an Integrated Services PBX.

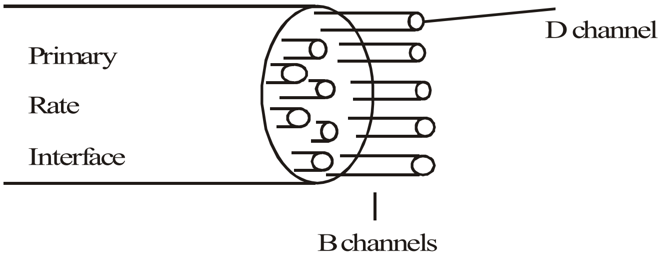


Fig. 3.9 Primary Rate Interface

3.3.8 ISDN Standards

Products for ISDN technology from different vendors even with similar features and options may create some compatibility issues. CCITT after good deliberations over the years published the first significant ISDN standards in a number of red binders in 1984 and they were simply known as the Red Book standards. The group subsequently met four years later which culminated in the publication of the 1988 Blue Book standards. These international publications were the foundation for the evolving ISDN national standards. The CCITT eventually was reformed into the group, which is now called the ITU–T. The standards used to define ISDN make use of the OSI reference model with the first three layers of this OSI reference model.

The two standard ISDN connectors are used as follows:

(i) To access basic rate ISDN, an RJ-45 type plug and socket (like a telephone plug) is used through unshielded twisted pair cable.

(ii) To access primary rate ISDN a coaxial cable is used.

The ISDN passive bus whose maximum length can be 1 km is a cable in user premises. It allows the attachment of up to eight devices to the basic rate ISDN interface. As only two B-channels are available, only two of the eight devices can communicate at any one time. Therefore, each device is required to compete for access to the passive bus.

The equipment available for ISDN includes the following:

- Video conference PC cards
- Gateways or bridges for LAN access (of which some are based on PC cards or stand alone boxes)
- Terminal adapters
- ISDN internal computer terminal adapter cards

All the issues related to API standard for internal computer adapters can be avoided or dealt with using an external ISDN terminal adapter. Most serial ports on a PC had limited speed of about 19.2 Kbps till now which is why this approach was not feasible. Of late, internal PC cards have come in which work asynchronously up to 115 Kbps. When used along with an appropriate external terminal adapter, this could have multimedia applications working.

ISDN is accessed through one of two services, named by the CCITT as Basic Rate Interface (BRI) and Primary Rate Interface (PRI). BRI provides 144 Kbps using the existing twisted pair copper telephone cable.

BRI includes 2B channels and 1-D channel. This may be written as 2B+D. B channel (bearer) provides 64 Kbps data transmission and can carry voice or data. D channel (Delta) operates at 16 Kbps and is used for control, i.e. for signalling between user devices and the ISDN. Therefore, the total data rate of BRA is therefore 144 Kbps. Audio digitized using pulse code modulation (PCM).

PRI also called '30B+D' owing to the number of B-channels and D-channels, is capable of carrying thirty independent data/voice channels of 64 Kbps each. Its structure consists of a 64 Kbps D-channel for singling between devices and the network as well as a 64kbps channel for synchronizing and monitoring. 2.048 Mbps is the total data rate of PRI.

It is also referred to as I.421 as per the recommendation of CCITT from which it is taken. This type of access is mainly meant for use in situations requiring a large transmission capacity such as when organizations make voice and data calls using an Integrated Services PBX.

ISDN internetworking equipment

ISDN can be used by many different internetworking devices as follows:

- **Terminal Adapters (TAs)** —— These are external devices that help in connecting X.21 and other conventional data interface to an ISDN circuit. This allows non-ISDN equipment to use the ISDN. Terminal adapters are used by internetworking manufacturers without an approved native ISDN interface for their devices.

  A demerit of this solution is that all information from the D-channel does not pass through the TA. Therefore, full advantage of ISDN facilities cannot be taken by the non-ISDN equipment such as Calling Line Identification.

- **ISDN Bridges** —— Being rather simple, bridging is amongst the most popular and commonly used methods of linking LANs. One major problem faced with ISDN bridging is the control of its use of the ISDN network. Bridges are simple to set up and use as they forward broadcasts and similar data by default. This implies that over ISDN, calls will be made to convey non-essential data which might prove to be costly in the long run.

This can be avoided if bridges are configured in such a way that broadcasts from particular addresses are blocked and certain protocols are understood. However, the major plus point of bridges, that is, simplicity, is lost. Bridges are appropriate for backing up ISDN.

- **ISDN Routers** ——A much more effective technique of utilizing ISDN for LAN networking is routing. It is the approach adopted by all networking vendors. Data is transmitted over the ISDN network only when it is actually required. In other words, unlike bridges, only necessary broadcast messages are sent to ensure efficient and effective use of bandwidth. It is possible to simplify the configuration. Unnecessary traffic is blocked out using filters.

Merits

(i)	High quality	-	ISDN connections are digital pipes with low error rate
(ii)	Flexible	-	ISDN connections can be established between two locations at any time provided the locations have ISDN which is like a configured leased line. It offers an almost transparent and quick call set-up. Therefore, for most users, the nature of dial-up is transparent.
(iii)	Economical	-	Rent is payed for ISDN just as in a telephone call. The cost of using ISDN is similar to that of the telephone service. It is quite cost-effective when it comes to intermittent LAN to LAN connectivity.
(iv)	Widely available	-	ISDN is now available widely following government initiatives in various countries.

3.3.9 Internet Service Providers (ISPs)

Internet Service Provider (ISP) is a company that access internet services. This service provider provides a software package in which you get registration with the providing services. Once you registered with username, password and dialing phone number, you can access ISP by paying the monthly fee. This software package is equipped with modem that is connected with internet services. Good ISPs have their own leased-line provided by telecommunication providers. Some of the largest and popular ISPs are At&T WorldNet, MCI, IBM Global Network, UUNet, PSINet, Netcom etc. It is sometimes known as internet access provider. There are 183 ISPs in India. The Table 3.1 shows the list of ISPs having all India license:

Table 3.1 ISPs in India

BSNL	RPGInfotech	Gateway systems	RailTel Corporation	i2i Enterprise
CMC	Sifi	ERNET India	GTL	Tata Power Broadband
Essel	VSNL	Jumpp India	Bharti Infotel	RailTel Corporation
Astro India Network	Primus Telecommunication India	Siti Cable Network	World Phone Internet Services	Escorts Communication
Reliance	L&T Finanace	In2Cable (India) Reliance	Spectra Net Reach	Estel Communication

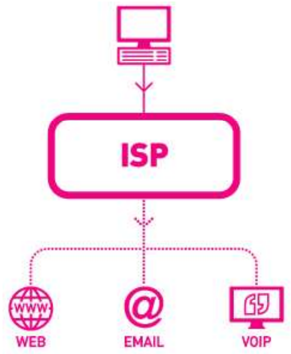


Fig. 3.10 *Services of ISP*

In the Figure 3.10, ISP provides web, Email and VoIP etc. as main services. ISP includes domain name registration and hosting, internet transmit, dial-up or DSL access, lease-line and collocation. You can take your domain name, secured website and high- availability web servers with this facility. Suppose a firewall is implemented with two separate Ethernet interfaces. The following figure shows how two ISPs are connected with Internet.

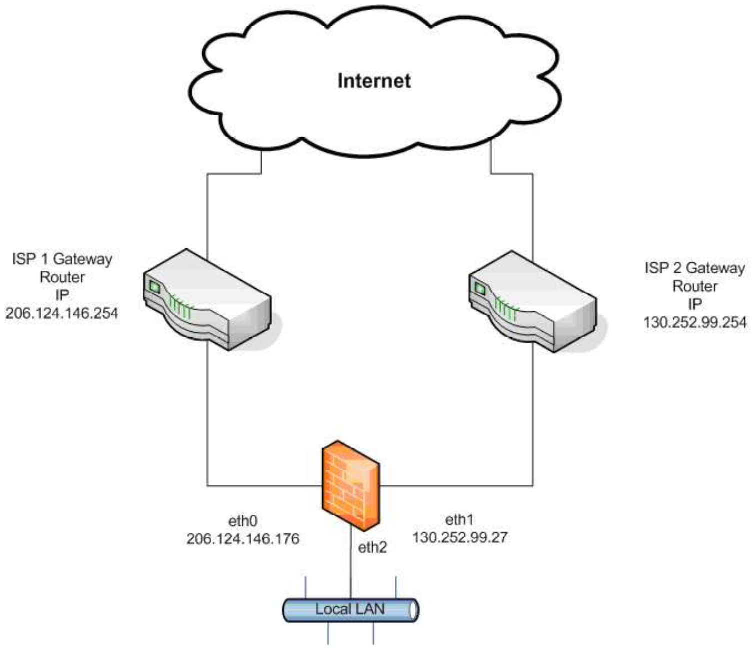


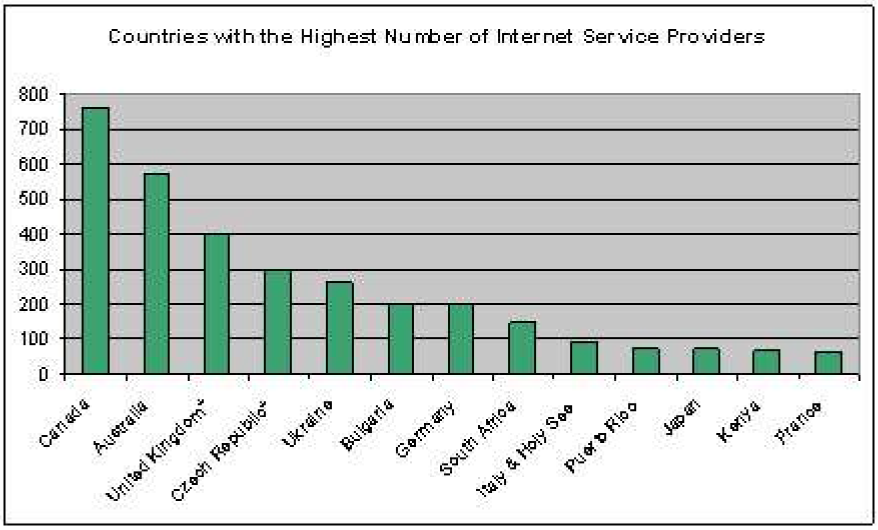
Fig. 3.11 *Two ISPs Connected with Internet*

NOTES

In the Figure 3.11, the following explanation can be analysed as follows:

- The Ethernet eth0 connects to ISP1. The IP address of eth0 is 206.124.146.176 and ISP's gateway router has address as 206.124.146.254.
- The Ethernet eth1 connects to ISP2. The IP address of eth1 is 130.252.99.27 and ISP's gateway router has address as 130.252.99.254.
- The Ethernet eth2 connects to local LAN.

The following graph shows the internet service providers in world:



Function of ISP

Commercial ISPs easily access and communicate with individual or various organizations across net. They are facilities-based carriers, for example, telephone and cable companies. The interconnected routers are assembled with ISP known as autonomous system (AS). ISP operates AS to information providers via Google and Yahoo search engines. They exchange traffic networking from other network. This process is called peering. The networks are connected to Internet Exchange (IX).

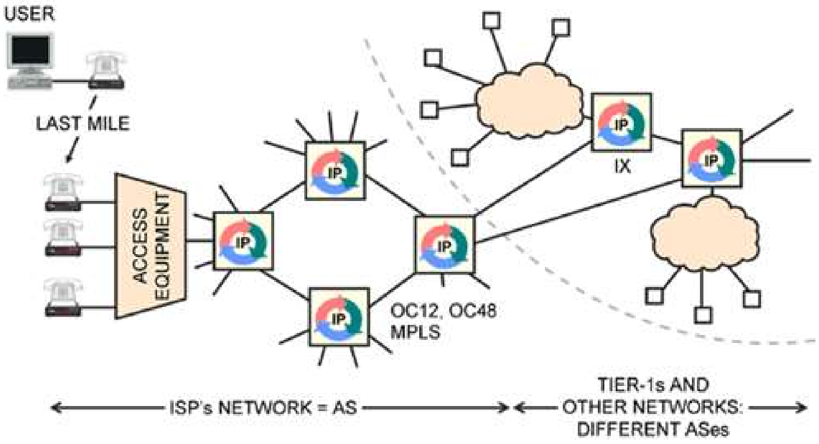


Fig. 3.12 *ISP Network*

In the Figure 3.12, ISP interconnects with IX providing Tier-1 and other networks. The Tier-1 network provides the largest service with reference to ISP. Peering is settlement free therefore no money transaction is done between ISP and commercial business houses.

Check Your Progress

1. What are two types of data transmission methods?
2. Write the components of data communication system.
3. What do you understand by the term DCE?
4. Which is the earliest electronic network?
5. State the components of telephone networks.
6. Define PSTN.
7. Which are two forms of ISDN services?

3.4 WAN TECHNOLOGIES

This technology connects sites that are in diverse locations. Wide Area Networks (WANs) connect larger geographic areas, such as New Delhi, India, or the world. The geographical limit of WAN is unlimited. Dedicated transoceanic cabling or satellite uplinks may be used to connect this type of network. Hence, a WAN may be defined as a data communications network covering a relatively broad geographical area to connect LANs together between different cities with the help of transmission facilities provided by common carriers, such as telephone companies. WAN technologies operate at the lower three layers of the OSI reference model. These are the physical data link and network layers.

Figure 3.13 explains the WAN, which connects many LAN together. It also uses switching technology provided by local exchange and long distance carrier.

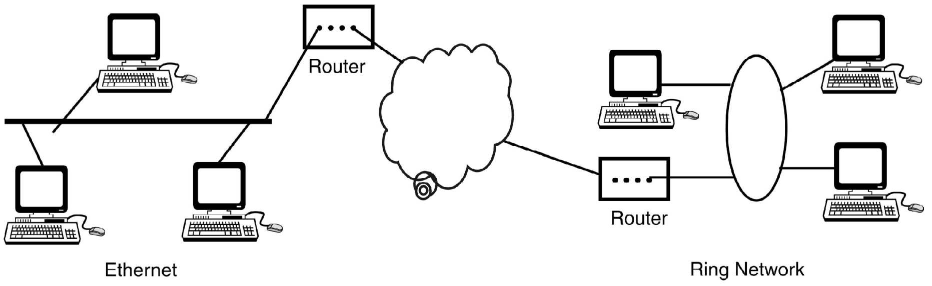


Fig. 3.13 *Wide Area Network (WAN)*

Packet switching technologies such as Asynchronous Transfer Mode (ATM), Switched Multimegabit Data Service (SMDS), Frame Relay and X.25 are used to implement WAN along with statistical multiplexing to allow devices to use and share these circuits.

The difference between MAN and WAN may be understood only from the services being used by them. WAN uses both the local and long distance carrier while MAN uses only local carrier. Hardware and protocols are same as in case of MAN.

There is a lot of confusion between LAN technology and WAN technology. The answer lies in how data is switched. Switching techniques are described subsequently in this chapter. It is the LAN/(WAN) integration that makes the network work. After all, people and machines not only need to be accessible locally, but from different sites as well.

NOTES

3.5 HISTORY OF INTERNET

The Internet, WWW and Information Super Highway are terms which have deep impact in the lives of millions of people all over the world. The widespread impact of Internet across the globe could not be possible without the development of Transmission Control Protocol/Internet Protocol (TCP/IP). This is the protocol suite developed specifically for the Internet. The information technology revolution could not have been achieved without this vast network of networks. It has become a fundamental part of the lives of millions of people all over the world. All the aforesaid services, basically, provide us the necessary backbone for information sharing in organizations and within common interest groups. That information may be in several forms. It can be notes and documents, data to be processed by another computer, files sent to colleagues, and even more exotic forms of data.

During late 1960s and 1970s, organizations were inundated with many different LAN and WAN technologies such as packet switching technology, collision-detection local area networks, hierarchical enterprise networks, and many other excellent technologies. The major drawbacks of all these technologies were that they could not communicate with each other without expensive deployment of communications devices. These were not only expensive, but also put users at the mercy of the monopoly of the vendor they were dealing with. Consequently, multiple networking models were available as a result of the research and development efforts made by many interest groups. This paved the way for development of another aspect of networking known as *protocol layering*. This permits communication between applications. A full range of architectural models were recommended and implemented by various computer manufacturers and research teams. As a result of this know-how, today any user group can find an architectural model and a physical network that are suitable to their specific needs. This includes cheap asynchronous lines with no other error recovery than a bit-per-bit parity function, through full-function wide area networks (private or public) with reliable protocols such as private SNA networks or public packet switching networks to high-speed but limited-distance local area networks.

It is now evident that organizations or users are using different network technologies to connect computers over the network. The desire of sharing more and more information among homogeneous or heterogeneous interest groups motivated the researchers to devise a technology whereby one group of users could extend its information system to another group who had a different network technology and different network protocols. This necessity was recognized in early 70s by a researchers' group in USA, who hit upon a new principle popularly known as internetworking. Other organizations, such as ITU-T (formerly CCITT) and ISO, also became involved in this area of interconnecting networks. All were trying to define a set of protocols, layered in a well-defined suite, so that applications are able to communicate with each other, regardless of the operating systems and underlying network technology.

NOTES

Internetworks

The availability of different operating systems, hardware platforms and the geographical dispersion of computing resources necessitated the need of networking in such a manner that computers of all sizes could communicate with each other, regardless of the vendor, the operating system, the hardware platform, or geographical proximity. Therefore, we may say that internetworking is a scheme for interconnecting multiple networks of dissimilar technologies. To interconnect multiple networks of dissimilar technologies, both additional hardware and software should be used. This additional hardware is positioned between networks and software on each attached computer. This system of interconnected networks is called an internetwork or an Internet.

To develop standards for internetworking, the US Defense Advanced Research Projects Agency (DARPA) funded research projects. ARPAnet, a project of DARPA, introduced the world of networking with protocol suite concepts such as layering, well before ISO's initiative in this direction. DARPA continued its research for an internetworking protocol suite. This may be seen in the early NCP (Network Control Program) host-to-host protocol to the TCP/IP protocol suite, which took its current form around 1978. DARPA was well known for its pioneering of packet switching over radio networks and satellite channels and ARPAnet was declared an operational network with responsibility of administering it to Defense Communications Agency (DCA) in 1975. TCP/IP had not yet been developed.

ARPAnet was basically a network based on leased lines connected by special switching nodes, known as Internet Message Processors (IMP). Many researchers were involved in TCP/IP research by 1979. This motivated DARPA to form an informal committee to coordinate and guide the design of the communication protocols and architecture. The committee was called the Internet Control and Configuration Board (ICCB).

The first real implementation of the Internet was when DARPA converted the machines of its research network ARPAnet to use the new TCP/IP protocols. After this transition which started in 1980 and finished in 1983, DARPA demanded that all computers willing to connect to its ARPAnet must use TCP/IP. The US military adopted the TCP/IP as standard protocol in 1983 and recommended that all networks connected to the ARPAnet conform to the new standards.

The success of ARPAnet was more than the expectations of its own founders and TCP/IP internetworking became widespread. As a result, new wide area networks (WAN) were created in the USA and connected to ARPAnet using TCP/IP protocol. In turn, other networks in the rest of the world, not necessarily based on the TCP/IP protocols, were added to the set of interconnected networks. Computing facilities all over North America, Europe, Japan, and other parts of the world are currently connected to the Internet via their own sub-networks, constituting the world’s largest network. In 1990, ARPAnet was eliminated, and the Internet was declared as the formal global network.

DARPA also funded a project to develop TCP/IP protocols for Berkeley UNIX on the VAX and to distribute the developed codes free of charge with their UNIX operating system. The first release of the Berkeley Software Distribution (BSD) to include the TCP/IP protocol set was made available in 1983 (4.2BSD). This led to the spread of TCP/IP among universities and research centers and has become the standard communications subsystem for all UNIX connectivity. There are many updated versions of BSD code available. These are 4.3BSD (1986), 4.3BSD Tahoe (1988), 4.3BSD Reno (1990) and 4.4BSD (1993).

Some examples of the different networks that have played key roles in this development are described below:

Internet

The word Internet is an acronym of the word ‘internetwork’ or ‘interconnected network’. Therefore, it can be said that the Internet is not a single network, but a collection of networks. The commonality between them in order to communicate with each other is TCP/IP. The Internet consists of the following groups of networks:

(a) **Backbones:** These are large networks that exist primarily to interconnect other networks. Some examples of backbones are NSFNET in the USA, EBONE in Europe and large commercial backbones.

(b) **Regional networks:** These connect, for example, universities and colleges. ERNET (Education and Research Network) is an example in the Indian context.

(c) **Commercial networks:** They provide access to the backbones to subscribers, and networks owned by commercial organizations for internal use and also have connections to the Internet. Mainly, Internet Service Providers come into this category.

(d) **Local networks:** These are campus-wide university networks.

The networks connect users to the Internet using special devices that are called gateways or routers. These devices provide connection and protocol conversion of dissimilar networks to the Internet. Gateways or routers are responsible for routing data around the global network until they reach their ultimate destination as shown in Figure 3.14. The delivery of data to its final destination takes place based on some routing table maintained by router or gateways. These are mentioned at various places in this book as these are the fundamental devices to connect similar or dissimilar networks together.

Over time, TCP/IP defined several protocol sets for the exchange of routing information. Each set pertains to a different historic phase in the evolution of architecture of the Internet backbone.

NOTES

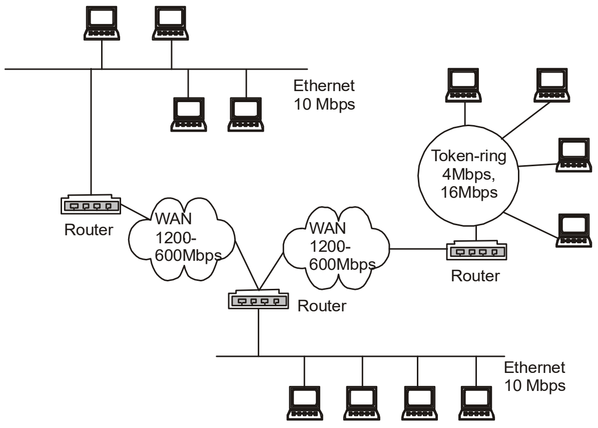


Fig. 3.14 *Local Area Networks Connected to the Internet via Gateways or Routers*

ARPAnet

ARPAnet was built by DARPA as described earlier. This initiated the packet switching technology in the world of networking and therefore is sometimes referred to as the 'grand-daddy of packet networks'. The ARPAnet was established in the late 1960s for the US Department of Defense with the aim to accommodate research equipment on packet switching technology besides allowing resource sharing for the Department's contractors. This network includes research centres, some government locations and military bases. It soon became popular with researchers for collaboration through electronic mail and other services. ARPAnet marks the beginning of Internet. ARPAnet provided interconnection of various packet-switching nodes (PSN) located across continental USA and Western Europe using 56 Kbps leased lines. ARPAnet provided connection to minicomputers running a protocol known as 1822 (after the number of a report describing it) and dedicated it to the packet-switching task. Each PSN had at least two connections to other PSNs (to allow alternate routing in case of circuit failure) and up to 22 ports for user computer connections. Later on, DARPA replaced the 1822 packet switching technology with the CCITT X.25 standard. The increase in data traffic made 56 Kbps capacity of the lines insufficient. ARPAnet has now been replaced with new technologies as backbone for the research side of the connected Internet.

Internet2

The success of the Internet and the consequent frequent congestion of the existing backbones has led the research community to look for alternatives. The university community, together with government and industry partners, and encouraged by the funding agencies, has started the Internet2 project. Internet2 has the following objectives:

(a) To create a high bandwidth; leading-edge network capability for the research community in the US.

(b) To enable a new generation of applications and communication technologies to fully exploit the capabilities of broadband networks.

(c) To rapidly transfer newly developed technologies to all levels of education and to the broader Internet community, both in the U.S. and abroad.

Internet services

The Internet is becoming a necessary tool rather than a convenient tool in society. It has proved its utility in all walks of life, such as education, economy, and socio-political arenas. This is because of the presence of extensive networks with wide information sources, commercial vehicles, Internet and provision of applications and information to carry out useful tasks. Daily users of the Internet are allowed to access these applications to reach other users, which was not possible few years ago. Today they can be accessed on his/her terminal in a very short time. Moreover, they are not required to know the details of the technology underlying the Internet. This is the major reason behind the popularity of Internet among laymen. The information available on the Internet is making them more confident about their area of working and without which they feel their productivity and profitability of their businesses would be seriously affected.

Following is a summary of the most widespread applications on the Internet today:

World Wide Web (WWW)

The World Wide Web is also known as the Web, WWW or W3. It is a global system of hypertext and multimedia services. WWW is a client-server model based on TCP/IP protocols and consists of browsers as clients and Web servers as servers. Web servers use HTTP (HyperText Transmission Protocol) and HTML (HyperText Markup Language) to make the WWW hypertext and multimedia services available to clients over the Internet. WWW supports hypertext to access several Internet protocols on a single interface. Hypertext or Hypermedia system allows interactive access to collections of documents. These documents can hold text (hypertext), graphics, sound, animations and video. These documents are linked together and may be seen as non-distributed and distributed. In non-distributed documents, all documents are stored locally (like CD-ROM). In distributed style, all documents are stored on remote servers.

Internet supports various protocols and network services. This includes e-mail, FTP, Gopher, Telnet, Usenet News. In addition to these, the World Wide Web has its own protocol.

The WWW provides a single interface for accessing all these protocols. This creates a convenient and user-friendly environment. It is no longer necessary to be conversant in these protocols. The web gathers together these protocols into a single system. Because of this feature, and the web's ability to work with multimedia and advanced programming languages, the WWW is the fastest-growing component of the Internet.

The operation of the web relies primarily on hypertext. HyperText is a document that contains links (pointers) to other documents. A button represents these links. A single hypertext document can contain links to many documents. In the context of the Web, button or graphics may serve as links to other documents, images, video, and sound.

A page represents each document. The initial page for individual or organization is called a Home Page**.** The page can contain many different types of information and must specify content, type of content, location and links. These pages are formatted with HTML rather than fixed WYSIWYG (What You See Is What You Get) representation (e.g., Word). With HTML, tags are placed within the text to accomplish document formatting, visual features such as font size, italics and bold, and the creation of hypertext links. Graphics may also be incorporated into an HTML document. The HTML is an evolving language, with new tags being added as each upgrade of the language is developed and released.

The web provides a vast array of experiences including multimedia presentations, real-time collaboration, interactive pages, radio and television broadcasts, and the automatic push of information to a client computer. Newer programming languages such as Java and JavaScript are extending the capabilities of the Web.

E-mail

Electronic mail, or e-mail, allows computer users locally and worldwide to exchange messages. E-mail users have an electronic mailbox into which incoming mail is dropped. Messages sent through e-mail can arrive within a matter of seconds. The user accesses these mails with a mail reader program, called mail user interface that is usually associated with computer account. One user may have different electronic mailboxes. The electronic mailbox is identified by an e-mail address and is given a user's account ID. This is not always true because on non-networked multi-user computer, e-mail address is just account ID.

Mail delivery among networked computers is more complicated. In this case, mail must identify computer as well as mailbox. Syntactically, e-mail address is composed of computer name and mailbox name, for example, user_id@domain.

E-mail message format contains header and body. Header includes delivery information and body carries message part. The header and body are separated by a blank line. An e-mail message can only be transmitted in form of 7-bit ASCII (American Standard Code for Information Interchange) data. ASCII is a 7-bit code, resulting in a maximum of 128 characters. The data in e-mail could not contain arbitrary binary values, e.g., executable program. There are techniques for encoding binary data so that it may be transported.

A powerful aspect of e-mail is the option to send electronic files to a person's e-mail address. Non-ASCII files, known as binary files, may be attached to e-mail messages. These files are referred to as MIME (Multipurpose Internet Mail Extensions) attachments. MIME extends and automates encoding mechanisms and was developed to help e-mail software handle a variety of file types. It allows inclusion of separate components, i.e., programs, pictures, audio clips in a single mail message. The sending program identifies the components so that the receiving program can automatically extract and inform mail recipients. Many e-mail programs, including Eudora and Netscape Messenger, offer the ability to read files written in HTML, which is itself of MIME type.

E-mail communication is actually a two-part process. The user composes mail with an e-mail interface program. This mail transfer program delivers mail to the destination and waits for the mail to be placed in outgoing message queues. SMTP (Simple Mail Transfer Protocol) is a standard application protocol for delivery of mail from source to destination. It provides reliable delivery of messages using TCP and message exchange between client and server including e-mail address lookup and e-mail address verification.

The e-mail can be considered as an electronic version of paper-based office memo, which is quick and much cheaper than a written communication. Because e-mail is encoded in an electronic medium, fast, automatic processing in the form of sorting and reply is possible. It allows quick, asynchronous communication across the Internet. Asynchronous communication consists of asynchronous characters as output at a rate that is independently generated by the transmitter. The asynchronous characters are actually self-synchronized because they are framed by Start and Stop bits that delineate the character. E-mail is the most widely used Internet service in the world. The best feature of the mail is its quick and reliable delivery of messages since it is contained as short data.

Telnet

A popular utility provided by TCP/IP is the TELNET. Telnet is a virtual terminal emulation facility that allows a user to connect to a remote system as if the user’s terminal was hard wired to that remote system. This works on client server architecture. There is one telnet server hosting various files and databases to share a client machine that accesses these resources. Telnet is a program that allows logging into computers on the Internet and using online databases, library catalogs, chat services, and more. For a computer to work on telnet the basic need is that its address should be known. This can consist of word (rag.gov.in) or numbers (140.147.254.3). The operation of this service is very simple. It requires just typing of the word telnet and then the address. Telnet is available on the WWW. Probably the most common web-based resources available through telnet are library catalogs. A link to a telnet resource may look like any other link, but it will launch a telnet session to make the connection. A telnet program must be installed on local computer and configured to Web browser, in order to work.

FTP (File Transfer Protocol)

The file transfer facilities are usually provided for by a mechanism known as the File Transfer Protocol (FTP). It is a simple featured ‘file moving’ utility that allows a record oriented (one record at a time) transfer, a block transfer (which moves chunk of a file) or an image transfer. To transfer a file, the user invokes the host, FTP utility specifies file name, type (if necessary), remote destination.

This is both a program and a method used for transfering files between computers on the Internet. Anonymous FTP is an option that allows users to transfer files from thousands of host computers on the Internet to their personal computer accounts. File transfer is quite rapid. FTP sites contain books, articles, softwares games, images, sounds, multimedia, course work, data sets, etc.

Archive

A computer site stores a large amount of public domain information, shareware software and many types of documentation. Archive functions as a catalogue of FTP sites. Archive is a program that searches all the FTP sites on the Internet, which are available on its master list, and stores the filenames in a central database. This database is available for users to search. When a user contacts an archive site and enters a search string, archive searches the database and returns a list of all files that contain that string. The list provides with the address to contact and the directories where the files are stored.

NOTES

Some archives are heavily used and must be supported by multiple sites, which are often located far apart. Each site should ideally have identical information available, therefore, they are mirrors of each other. When one site gets a new file, it must be mirrored to the other sites, usually by using FTP.

SMTP

It stands for Simple Mail Transfer Protocol. SMTP is a defined standard for e-mail over the TCP/IP protocol and therefore is widely used on the Internet. A utility that is somewhat popular provides a mechanism by which a user can specify a destination address, a particular path to follow (if desired), and a message. All e-mail servers use this protocol for message delivery and receipt.

E-mail discussion groups

One of the benefits of the Internet is the opportunity it offers to people worldwide to communicate via e-mail. The Internet is home to a large community of individuals who carry out active discussions organized around topic-oriented forums distributed by e-mail. There are all kinds of discussion groups, such as sports, politics, software, troubleshooting etc. where users post their queries in the form of e-mail and others reply.

Network information server

News, Gopher, and WWW are special servers for information, which require a client software package for access. They can be thought of as network databases. Each server is powerful with some interconnectivity. Each server requires a client application to allow the user to access information. The main purpose of these client/servers is to help a user navigate the Internet to find information and files.

- **News –** News constitutes broad topics called news groups, to which people can post or respond to posts. News is available via Usenet, Internet, and some commercial services. Almost nobody carries all the news groups. User access is through a news reader application that accesses a news server. There are many variations of news readers. Usenet News is a global electronic bulletin board system in which millions of computer users exchange information on a vast range of topics. There are thousands of Usenet newsgroups in existence. While many are academic in nature, numerous newsgroups are organized around recreational topics. Serious computer related work takes place in Usenet discussions. A few e-mail discussion groups also exist as Usenet newsgroups.
- **Gopher –** Before the advent of the WWW, Gopher was the document access protocol of choice. Gopher is a menu utility mainly for text-based documents on the Internet, which uses FTP for retrieving files from archive sites. Gopher also uses a search utility called Veronica for aiding users in finding files in the Gopher archive sites. Veronica can do keyword searches whereas Archie can only search for file names.

IRC, chat and instant messaging

IRC is the Internet Relay Chat service in which participants around the world can ‘talk’ to each other by typing in real time on hundreds of channels. These channels are usually based on a particular topic. To have access to IRC, an IRC software program is essential. This program connects the user to an IRC server and allows him to visit IRC channels.

Chat programs are now common on the Web. They are sometimes included as a feature of a web site, where users can log into the chat room to exchange comments and information about the topics addressed on the site.

A variation of chat is the phenomenon of instant messaging. With instant messaging, a user on the Web can contact another user currently logged in and type a conversation.

There are other services available on Internet like FAQ (Frequently Asked Questions), RFC (Request for Comments), FYI (For Your Information), MUD (Multi User Dimension) and WAIS (Wide Area Information Server).

MUDs are multi-user virtual reality games based on simulated worlds. Traditionally text based, graphical MUDs now exist. There are MUDs of all kinds on the Internet, and many can be joined free of charge. MUDs are accessible by Telnet.

WAIS provides information lookup services to libraries and databases on the Internet. A simple WAIS client allows the user to select databases to search from a list. The user then provides keywords to search for, and the WAIS client allows the user to view any matches found. This is cumbersome once the list of databases grows into thousands. Screen after screen of database names scroll by one after another.

3.5.1 Standards for TCP/IP and the Internet

The group of people who were responsible for monitoring and reviewing the progress made in the effort to develop TCP/IP initiated by US Department of Defense was known as the Internet Activities Board (IAB). Gradually, the IAB evolved from a DARPA-specific research group into an autonomous organization. Its members chaired smaller groups called Internet Task Forces (ITFs). Each ITF was required to deal with different aspects of the evolution of TCP/IP and the Internet. In 1989, the IAB was reorganized. Two subsidiary groups were created, viz., the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF). The former was assigned the task of developing the Internet standards, and the latter was made responsible for research and development. In 1992, the Internet Society (ISOC) was formed as the standardizing body for the Internet community and the IAB was renamed as the Internet Architecture Board (IAB). This group itself relies on the Internet Engineering Task Force (IETF) for issuing new standards, and on the Internet Assigned Numbers Authority (IANA) for coordinating values shared among multiple protocols.

3.5.2 RFCS and TCP/IP Standardization Process

Within the IETF, subsidiary working groups are formed to look after the specific aspect of the overall Internet protocol suite. There are groups dedicated to network management, security, and routing, among other interests. The RFC is a series of technical papers commonly referred to as Request for Comments (RFCs). This is responsible for reviewing and publishing new standards documents. RFC series are traditionally referenced using numbers in a chronological order based on gradual development and is publicly available on the Internet to both workgroup members and the general public for discussion point of view.

The Internet Standards Process is described in RFC2026. The Internet Standards Process — Revision3 is concerned with all protocols, procedures, and conventions that are used in or by the Internet, whether or not they are part of the TCP/IP protocol suite. The objectives of the Internet Standards Process are to achieve technical excellence, prior implementation and testing, clear, concise, and easily understood documentation, openness and fairness and timeliness.

The process of standardization includes submission of the new specification to the IESG for technical discussion and feasibility, and also for publication as an Internet draft document. This should take no shorter than two weeks and no longer than six months. Once the IESG reaches a positive conclusion, it issues a last-call notification to allow the specification to be reviewed by the whole Internet community. After the final approval by the IESG, an Internet draft is recommended to the Internet Engineering Taskforce (IETF), another subsidiary of the IAB, for inclusion into the standards track and for publication as a RFC. It may also be revised over time or phased out when better solutions are found. If the IESG does not approve of a new specification after, or if a document has remained unchanged within six months of submission, it will be removed from the Internet drafts directory.

Check Your Progress

8. Define World Wide Web (WWW).
9. How is e-mail useful?
10. What do you understand by the term TELNET?
11. What is the function of archive?

3.6 NETWORK ARCHITECTURES

TCP/IP

TCP/IP stands for Transmission Control Protocol/Internet Protocol. It was developed with the objective to specify a suite of protocols capable of providing transparent communications interoperability services between computers of all sizes, regardless of the hardware or operating system platforms supporting them. Over the years, TCP/IP has become the most widespread among today's protocols. One reason for TCP/IP's popularity is the public availability of its protocols' specifications. In this sense, TCP/IP can justifiably be considered an open system. Most users rely on TCP/IP for the purpose of file transfers, electronic mail (e-mail), and remote login services.

Network Architectures

Network architecture defines the communications products and services, which ensure that various components work together. In the early days of data communication systems, the majority of communications were between the DTE and the host computer. Therefore, transmission control procedures were alone enough as communication protocols. However, recent computer systems link with other systems to form a network which result in a situation where different protocols serving for different purposes is required. Hence, the network architecture represents a systemization of the various kinds of protocols needed to build a network.

Computer manufacturers have developed different protocols as needed. This means that each type of computer needed to support different protocols. This necessitated large development and maintenance costs. All computer manufacturers, as shown in Table 3.2, worked together to standardize and systemize protocols to link their models and reduce the development and maintenance costs thereby. This was how each manufacturer built his own network architecture.

Since the concept of the network architecture was first introduced, connecting computers of the same manufacturer has become easier. However, from a user's perspective, the ideal form of network architecture is one which enables machines of all manufacturers to connect to each other. Therefore, the need of standardization of network architecture arose.

Table 3.2 *Network Architecture by Vendor*

Manufacturer	Network architecture
IBM	System Network Architecture (SNA)
DEC	Digital Network Architecture (DEC)
Borroughs	Borroughs Network Architecture (BNA)
UNIV AC	Distributed Communication Architecture (DCA)
Toshiba	Advanced Network System Architecture (ANSA)
NEC	Distributed Information Processing Architecture (DINA)
Honeywell	Distributed System Environment (DSE)

The following are ways to achieve connection between different manufacturers:

- **Protocol Converters:** These are devices that translate from one native protocol into another, for example, from ASCII to IBM SNA/SDLC
- **Gateways:** These are hardware/software combinations that connect devices running different native protocols. In addition to protocol conversion, gateways provide a gateway connection between incompatible networks. Examples include Ethernet-to-Token Ring gateways, X.25-to-Frame Relay gateways, and T-carrier-to-E-Carrier International Gateway Facilities (IGFs).

In addition to the above, Protocol Analyzers are available as diagnostic tools for displaying and analysing communications protocols. Analysers allow technicians, engineers and managers to test the performance of the network to ensure that the systems and the network are functioning according to specifications.

LAN managers, for instance, use protocol analysers to perform network maintenance and troubleshooting and to plan network upgrades and expansions.

Example of TCP/IP Operations

TCP/IP Layers and Protocols

TCP/IP defines a suite of communications and applications protocols in layer structure, with each layer handling distinct communication services. TCP/IP defines a four-layer model as shown in Figure 3.15 consisting of Application, host-to-host, Internet, and Network Access layers. This architecture is based on three sets of interdependent processes, namely, application-specific processes, host-specific processes, and network-specific processes.

Application Layer
(application-specific processes)

Host to Host Layer
(Host-specific processes)

Internet Layer
(routing processes)

Network Access Layer
(network-specific processes)

Fig. 3.15 *TCP/IP Communication Architecture*

The following are examples of concerns that each of these processes should handle:

Application-specific processes: TCP/IP defines the External Data Representation (XDR) protocol to provide an agreement between the data syntax running between the different platforms.

Host-specific processes: It is the responsibility of the host-specific process to establish, maintain, and release a connection on behalf of an application without losing track of other logical connections on multiuser/multitasking operating systems. Therefore, it ensures that data integrity is maintained without confusing the identity of the communicating applications.

Network-specific processes: These are processes that concerns with the delivery of data to the transmission medium and route data across networks until it reaches its ultimate destination.

The correspondence between the TCP/IP and OSI model is shown in Figure 3.21. From Figure 3.15, the relationship between the two figures may be established. Layer 5, 6 and 7 corresponds to application layer (4th layer) of TCP/IP communication architecture. In a similar manner layer 4 and 3 of OSI can be related with the host-to-host layer and Internet layer of TCP/IP suite, respectively. The physical layer and data link layer of OSI is similar to the network layer of TCP/IP.

3.6.1 Layering the Communications Process

Open Systems Interconnection (OSI) was set up as an international standard for network architecture. The International Organization for Standardization (ISO) took the initiative in setting up OSI.

Layering the Communication Process

OSI has two meanings. It refers to:

(i) Protocols that are authorized by ISO

(ii) OSI basic reference model

OSI reference model divides the required functions of the network architecture into several layers and defines the function of each layer. Layering the communications process means breaking down the communication process into smaller and easier to handle interdependent categories, with each solving an important and somehow distinct aspect of the data exchange process. The objective of this detail is to develop an understanding of the complexity and sophistication that this technology has achieved, in addition to developing the concept for the inner workings of the various components that contribute to the data communications process.

Physical data encoding

Information exchanged between two computers is physically carried by means of electrical signals assuming certain coding methods. These codings can be characterized by changing voltage levels, current levels, frequency of transmission, phase changes, or any combination of these physical aspects of electrical activity. For two computers to reliably exchange data, they must have a compatible implementation of encoding and interpreting data carrying electrical signals. Over time, network vendors defined different standards for encoding data on the wire. Figure 3.16 shows one such standard, namely, bipolar data encoding.

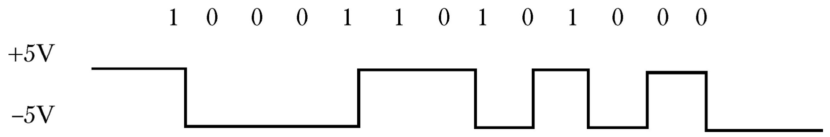


Fig. 3.16 Bipolar Data Encoding

In bipolar encoding, binary data is simply represented by the actual signal level, in which a binary 1 is encoded using a fixed voltage level (for example, +5 volts) and a binary 0 is encoded using a negative voltage level (for example, –5 volts).

Transmission media

This deals with the type of media used (fibre, copper, wireless, and so on), which is dictated by the desirable bandwidth, immunity to noise, and attenuation properties. These factors affect the maximum-allowable media length while still achieving a desirable level of guaranteed data transmission.

Data flow control

Data communications processes allocate memory resources, commonly known as communication buffers, for the sake of transmission and reception of data. A computer that is in the process of receiving data runs the risk of losing data when its communication buffers exhaust. This can be avoided by employing a data flow control mechanism as shown in Figure 3.17. For proper data flow control, the receiving process must send a ‘step sending’ signal to the sending computer, if it is unable to cope up with the rate at which data is being transmitted by the sending computer. Later, when data communication buffer is available, the receiving computer sends 'resume sending' signal.

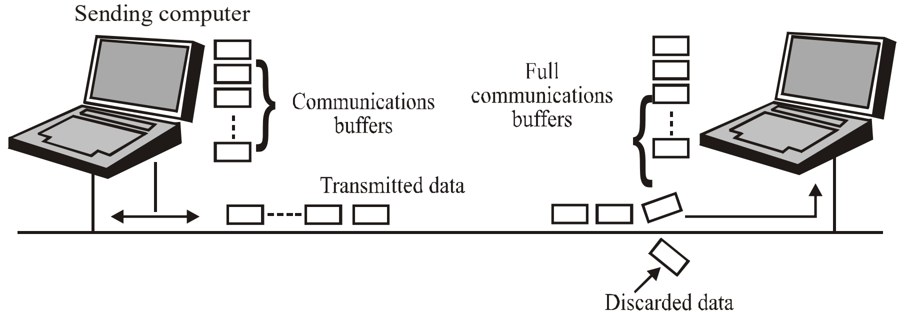


Fig. 3.17 *Data Flow Control Mechanism*

Data frame format

Information exchange between computers, communication processes need to have following for accomplishing these aspects of the exchange process:

- The receiving computer must be capable of distinguishing between an information carrying signal and mere noise.
- There should be a detection mechanism to detect whether the information carrying signal is intended for itself or some other computer on the network, or a broadcast (a message that is intended for all computers on the network).
- The receiving end should be able to recognize the end of data train intended for receiver before it engages itself to recover data from the medium.
- The receiving end after completion of receiving of information, must also be capable of dealing with and recognizing the corruption, if any, introduced in the information due to noise or electromagnetic interference.

To accommodate the above requirements, data is delivered in well-defined packages called data frames as shown in Figure 3.18. This frame belongs to the Ethernet packet format and has been explained earlier in the unit on Local Area Network. The receiving end compares the contents of this data frame. If the comparison is favourable, the contents of the Information field are submitted for processing. Otherwise, the entire frame is discarded. It is important to realize that the primary concern of the receiving process is the reliable recovery of the information embedded in the frame.

7	1	6	6	2	46-1500	4	Byte
Preamble (P) 1010.10	SED 10101011	SA	DA	L		FCS	

Fig. 3.18 *Frame Format for IEEE 802.3*

NOTES

Routing

With the growth of network size, traffic also grows affecting the overall network in performance and responses. To manage a situation like that, network specialists break the network into multiple networks, interconnected by specialized devices that include routers, bridges, brouters and switches (refer Figure 3.19).

The routing approach requires implementation of various processes in cooperation, both in routers and workstations with the sole objective of delivering the data, intelligently to the final destination. Such exchange of data can take place between any two workstations, within or without the same network.

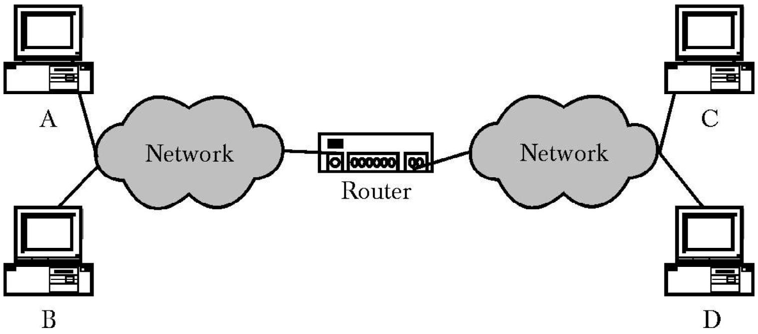


Fig. 3.19 Router connecting Two Networks

The network address and the complete address

In addition to the data link address, which should be guaranteed to be unique for each workstation on a particular physical network, all workstations must have a higher-level address in common. This is known as the network address. The network address is very similar in function and purpose to the concept of a street name. A street name is common to all residences located on that street.

Unlike data link addresses, which are mostly hardwired on the network interface card, network addresses are software configurable. It should also be noted that the data structure and rules of assigning network addresses vary from one networking technology to another.

Inter-process dialogue control

When two applications engage in the exchange of data, they have established a session between them. Consequently, a need arises to control the flow and the direction of data flow between them for the duration of the session. Depending on the nature of the involved applications, the dialogue type might have to be set to full duplex, half duplex, or simplex mode of communication. Even after setting the applicable communications mode, applications might require that the dialogue itself be arbitrated. For example, in the case of half duplex communications, it is important that somehow applications know when to talk and for how long.

Session recovery

Another application-oriented concern is the capability to reliably recover from failures at a minimum cost. This can be achieved by providing a check mechanism which enables the resumption of activities since the last checkpoint. As an example,

NOTES

consider the case of invoking a file transfer application to have five files transferred from point A to point B on the network. Unless a proper check mechanism is made to take care of the process, a failure of some sort during the transfer process might require the retransmission of all five files, regardless of where in the process failure took place. Check pointing circumvents this requirement by retransmitting only the affected files, saving time and bandwidth.

Presentation problems

Whenever two or more communicating applications run on different platforms, another concern arises about the differences in the syntax of the data they exchange. Resolving these differences requires an additional process. Good examples of presentation problems are the existing incompatibilities between the ASCII and EBCDIC standards of character encoding, terminal emulation incompatibilities, and incompatibilities due to data encryption techniques.

3.7 NEED FOR LAYERED SOLUTIONS AND OPEN SYSTEMS INTERCONNECTION (OSI)

Layering involves breaking the communication process into different categories and dealing with them according to the steps to which they belong. Categorization must take into account the interdependency of some processes relative to others. At least three advantages could be achieved by using the layered approach, including the following:

- **Specialization:** Solution developers can specialize in one or the other category of problems, which, given the rate at which the technology is advancing, is more affordable than an approach based on integrating all problems into one category
- **Minimal cost:** Using the layered approach, it is easier for vendors to introduce changes to, or even replace, an entire layer, while leaving others intact.
- **Freedom of choice:** As you will see later, users benefit from layering because it provides them the freedom to implement networks that can be tailored to meet their needs.

Network Design and Problem of Communication between Layers

Network design experts came up with the hierarchical network design to help in developing a topology in between discrete layers. For example, while routers with medium speed can connect buildings within each campus, high-speed WAN routers can carry traffic across the enterprise. WAN backbone and switches can connect user devices and servers within the buildings.

The ***physical layer*** determines the type of network design exclusively designed for the physical layer and connected to higher levels such as data link, network, session, transport, presentation and application layers. It also determines whether data transfer uses simplex, half-duplex or full duplex modes of communication.

In the ***data link layer***, the network is generally designed as Ethernet, ARCNET and Token Ring. The following Table 3.3 shows the content of Ethernet Address using data frames in a data link layer:

Table 3.3 *Content of Ethernet Address*

64 bits	48 bits	48 bits	16 bits	368-12000 bits	32 bits
Preamble	Destination Address	Source Address	Frame Type	Data Frame Redundancy Checks(CRC)	Cyclic

In this layer, the network is designed according to the content of data frame to detect and correct data corruption in the network communication channel.

The ***network layer*** delivers units of data as individual packets. The network designers design the protocols used for routing data.

The ***Transport Layer*** delivers data within a host computer and then hands the data over to the transport layer. The source transport layer carries a virtual conversation with the destination transport layer. The network is designed on a hop-to-hop basis.

Table 3.4 *Network Design of Peers Process among Discrete Layers*

Application Layer Layer	← Application Layer →	Application
Presentation Layer Layer	← Presentation Layer →	Presentation
Session Layer	← Session Layer →	Session Layer
Transport Layer	← Transport Layer →	Transport Layer
Network Layer	← Network Layer →	Network Layer
Data Link Layer	← Data Link Layer →	Data Link Layer
Physical Layer	← Physical Layer →	Physical Layer

The Table 3.4 shows how dotted lines among the corresponding layers (in each host) indicate a virtual conversation of different layers. Network designers often recommend a mesh topology to meet the availability requirement because routers are connected to single-link delay between two sites. This layer is responsible for getting data from or sending data to each network that manages flow control on an end-to-end basis.

A ***session layer network*** is designed to occasionally merge the session and transport layers. Therefore, network designers design the hierarchical topology for this layer.

The ***presentation layer*** handles printers, video displays and file formats. So, hierarchical topology is a suitable network design for this layer.

The ***application layer*** deals with network-wide applications which include electronic mail and distributed databases. Generally, API forms a session layer upon an application layer. It provides file services, network printer services and mail services. The network designs included in this layer are, Novell's Netware, Banyan's VINES and Artisoft's LANtastic. Figure 3.20 shows the interconnection between the layers of OSI.

NOTES

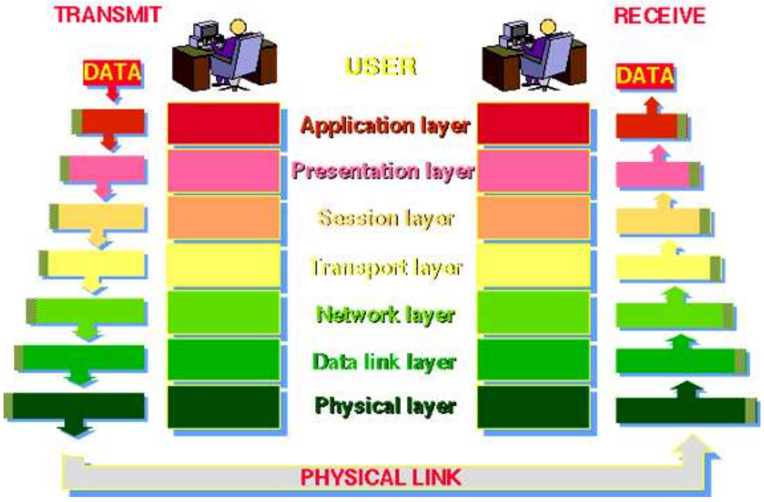


Fig. 3.20 *Interconnection between Layers of OSI*

3.7.1 Open Systems Interconnection (OSI) Model

The OSI model of data communication was developed in 1984 by the International Standardization Organization (ISO). OSI specifies a seven-layer model as shown in Figure 3.20. In addition to forming the basis of the ongoing development of OSI's own protocols, the model is used by the industry as the frame of reference when describing protocol architectures and functional characteristics.

The International Standard Organization (ISO), in an effort to encourage open networks, developed an open systems interconnect reference model. The model logically groups the functions and sets rules, called protocols, necessary to establish and conduct communication between two or more parties. The model consists of seven functions, often referred to as layers as shown in Figure 3.21.

The last three layers are mainly concerned with the organization of terminal software and are not directly the concern of communications engineers. The transport layer is the one which links the communication processes to this software-oriented protocols.

The basic philosophy of the seven-layer model is that each layer may be defined independently of every other layer. Thus, from the user point of view, interchange takes effect across each operation and passes down through the layers of the model until data interchange is affected through the physical connection.

The top layer is used by the transmitting device where data is placed into a packet under a header.

The protocol data unit that consists of data and header, are handled by each of the successive lower layers as data flows across the network to the receiving node. Data flows through the layer model and each of the successive higher layer strip off the header information.

Another alternative standards approach was being led by the CCITT (Consultative Committee on International Telephony and Telegraphy) and the ISO (International Organization for Standardization) parallely to the development of TCP/IP by DARPA. The CCITT has now become the ITU-T (International Telecommunications Union-Telecommunication Standardization Sector).

The outcome of this joint attempt was the creation of the OSI (Open Systems Interconnect) reference model (ISO 7498). This outlines a seven-layer model of data communication with the bottom layer comprising physical and transport layers and the application protocols forming the upper layers. Each layer of the model is responsible for specific functions. The operation of a network protocol stack is understood on the basis of this model (figures 3.20 and 3.21). It is also used as a reference tool to compare network stack.

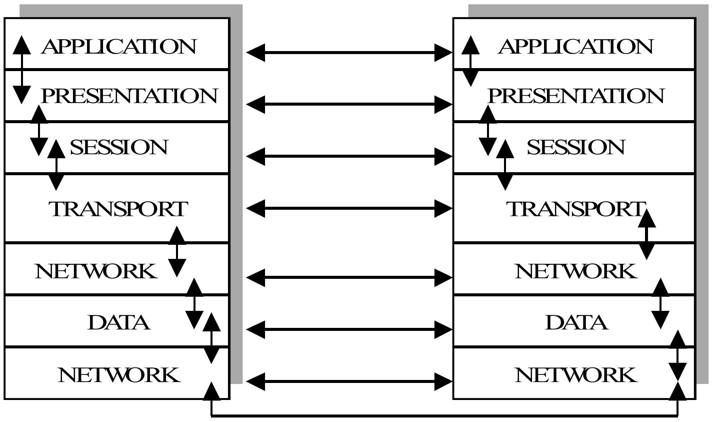


Fig. 3.21 *The OSI Reference Model Implementations*

Each layer provides some of the functions to the layer above it in return for the functions provided by the layer below it. In this fashion, messages are transmitted vertically through the stack from one layer to the other. Logically, each layer communicates directly with its peer layer on the other nodes.

3.7.2 Layered Architecture of OSI

The Physical Layer (Layer 1)

This layer describes the physical media over which the bit stream is to be transmitted. It tells about the electrical and mechanical aspects of data transmission to a physical medium that includes setting up, maintaining and disconnecting physical links apart from trans-mitting data. It is primarily concerned with moving bits from one node to next over the physical link.

It accepts data from the Data link layer in bit streams for subsequent transmission over the physical medium. At this layer, the mechanical (connector type), electrical (voltage levels), functional (ping assignments), and procedural (handshake) characteristics are defined. RS-232C/D is an example of a physical layer definition.

The Data Link Layer (Layer 2)

It takes the bits that are received by the physical layer and detects errors. This ensures the proper sequence of transmitted data by establishing an error-free communication path over the physical channel between network nodes. Framing messages for trans-mission, checking integrity of received messages and managing access to the channel and its use are its main work. Hence, this layer is responsible for the reliable transfer of data across the physical link. Its responsibilities include such functions as data flow control, data frame formatting, error detection, and link management.

NOTES

The Network Layer (Layer 3)

The network layer sets up appreciate paths between various nodes and to do this it uses a software that handles PDUs to transport them to the final destination. The Internet Protocol (IP) operates at this layer. It is mainly responsible for providing routing services across the Internet. It also shields the above layers from details about the underlying network (the network topology and road map) and the routing technology that might have been deployed to connect different networks together. In addition to routing, this layer is responsible for establishing and maintaining the connection.

The next three layers are task oriented and have to do with the operations performed by the user rather than with the network.

The Transport Layer (Layer 4)

This layer guarantees the orderly and reliable delivery of data between end systems. Data is received from session conrol layer and transported to network control layer. The two protocols used here include transmission control protocol or TCP and OSI transport protocol or TP's five levels. The transport layer also performs additional functions such as data multiplexing and de-multiplexing. This layer divides up a transmitting message into packets and reassembles them at the receiving end.

The Session Layer (Layer 5)

The session layer is responsible for establishing, maintaining, and arbitrating the dialogues between communicating applications. It is also responsible for the orderly recovery from failures by implementing appropriate check pointing mechanisms.

The Presentation Layer (Layer 6)

Formatting and displaying of data, received by terminals and printers are functions performed by the presentation layer. It is concerned with differences in the data syntax used by communicating applications. This layer is responsible for remedying those differences by resorting to mechanisms that transform the local syntax (specific to the platform in question) to a common one for the purpose of data exchange. For example, it performs conversion between ASCII and EBCDIC character codes, does data compression and encrypts data if necessary.

The Application Layer (Layer 7)

The application layer provides support services for user and application tasks. It determines how the user is using the data network. It allows the user to use the network. For example, it provides network-based services to the end user. Examples of network services are distributed databases, electronic mail, resource sharing, file transfers, remote file access and network management. This layer defines the nature of the task to be performed.

OSI Protocol of Different Layers

Before going on to the OSI protocol of different layers, let us first define a protocol. A protocol is a set of conventions that governs the format and control of the interaction that takes place among functional units.

The OSI architectural model was developed by ISO (International Organization for Standardization).

The model comprises seven layers— the first four layers are referred to as the **lower layers** whereas the last three layers are referred to as the **upper layers**.

Responsibilities of the Seven Layers

Application Layer: Layer seven is the highest layer which interprets data and may also indulge in encryption or decryption.

Applications using the network learn the technique of sending requests, specifying filenames and responding to requests. At this layer, PDU or Protocol Data Unit is called data and this layer not only sends requests to the presentation layer but also interfaces directly with the application processes. It also performs common application services for these processes such as virtual terminal and virtual file protocols as well as job transfer and manipulation protocols.

Presentation Layer: The sixth layer helps determine the representation of data by computers [ASCII, GIF.]. As in the seventh layer, PDU is referred to as data in this layer.

This layer provides a response to the service requests received from the Application Layer and also sends service requests to the fifth layer called the Session Layer. Conversion of an EBCDIC-coded text file to an ASCII-coded file is an example of a presentation service. Within the end-user systems, the Presentation Layer relieves the layer above it of concern regarding syntactical differences in the representation of data.

Session Layer: Layer number 5 establishes a communication session, provides security and authentication. NetBIOS is a layer 5 protocol. The Session Layer responds to the service requests from the layer above it and sends service requests to the fourth layer which is the Transport Layer. This layer manages the dialogue between end-user application processes. It facilitates either half-duplex or duplex operation and establishes procedures related to adjournment, restart, check-pointing and termination.

Transport Layer: The fourth layer provides transfer correctness, data recovery and flow control. TCP is a layer 4 protocol. In this layer, PDU is referred to as a segment. This layer responds to service requests coming from the fifth layer and sends service requests to the third layer, that is, the Network Layer. The Transport Layer is mainly responsible for ensuring transparency in the transfer of data between end-users. Therefore, it relieves the upper layers of any concerns or issues regarding provision of cost-effective and reliable data transfer.

Network Layer: Layer three is concerned with assigning addresses and packet forwarding techniques. Here, PDU is called a packet. This layer is responsible for responding to service requests from the fourth layer, that is, the Transport Layer and issues service requests to the layer below, that is, the Data Link Layer. The Network Layer provides the functional and procedural means of transferring data sequences of variable length from a source to a destination through one or more networks, but at the same time, maintains the quality of service requested by the fourth layer, that is, the Transport Layer. Functions performed by the Network Layer include flow control, network routing, error control and segmentation/desegmentation.

Data Link layer: The second layer is concerned with transmitting frames over the Net [start/stop flags, additional bit/byte stuffing, checksum and CRC] and Frame format. Examples of layer 2 protocols are ATM (Asynchronous Transfer Mode), CAN (Controller Area Network) bus, StarLAN, HDLC (High-level Data Link Control), ADCCP (Advanced Data Communication Control Procedures) and LocalTalk (a protocol used with Apple computer systems).

Different data link layer specifications define different network and protocol characteristics. The second layer is subdivided as follows:

Media Access Control (MAC): Controls access and encodes data for the Physical Layer into a signalling format that is valid.

Logical Link Control (LLC): Provides the Network Layer with the link to the network. PDU is referred to as a frame at this layer.

This layer responds to service requests from the third layer and issues service requests to the first layer. This layer provides the procedural and functional techniques for transferring data between network entities and detecting as well as rectifying errors that may take place at the first layer, that is, the Physical Layer.

Note: HDLC and ADCCP are examples of data link protocols for point-to-point and packet-switched networks; and LLC is an example of a data link protocol for LANs.

Physical Layer: This is the first layer of the model which defines the physical and electrical implementation of the bus. In other words, it defines the hardware and signal-level implementation of the bus. It is also concerned with network cabling, data transmission encoding, types of connectors, physical data rates and maximum transmission distances. At this layer, information is placed on the physical network medium.

Examples of a physical layer specification include RS-232 and RS422. Here, a PDU is referred to as a bit. This layer fulfils the service requests received from the Data Link Layer. The important responsibilities of the Physical Layer are as follows:

(a) It establishes and terminates a connection to a communication medium

(b) It participates in the process wherein the communication resources are effectively shared among multiple users, e.g., contention resolution and flow control.

(c) It facilitates conversion between the digital data representation in user equipment and the corresponding signals transmitted over a communication channel.

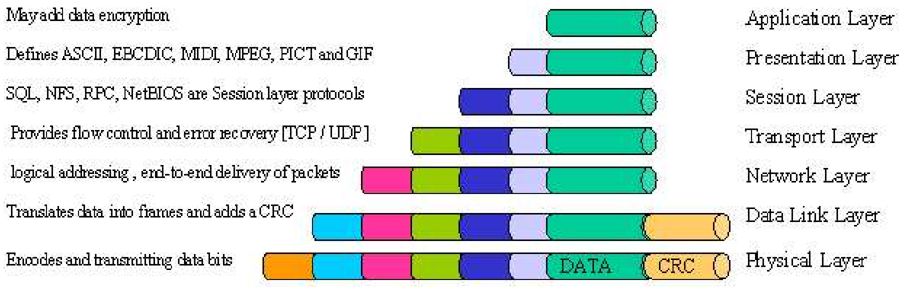

May add data encryption
Application Layer
Defines ASCII, EBCDIC, MIDI, MPEG, PICT and GIF
Presentation Layer
SQL, NFS, RPC, NetBIOS are Session layer protocols
Session Layer
Provides flow control and error recovery [TCP / UDP]
Transport Layer
logical addressing , end-to-end delivery of packets
Network Layer
Translates data into frames and adds a CRC
Data Link Layer
Encodes and transmitting data bits
DATA
CRC
Physical Layer

Fig. 3.22 *OSI Protocol Stack*

NOTES

Check Your Progress

12. What does TCP/IP stands for?
13. Who took the initiative in setting up OSI?
14. What does the physical layer determines?
15. Which layer is the first layer of the OSI model?

3.8 ROUTING CONCEPTS

The IP addressing mechanism requires hosts and networks so that a host on the network can transmit and receive IP packets. The host could be a workstation or a router. Routing refers to the process of moving data from one host computer to another by selecting the shortest and most reliable path intelligently. This is the path or route over which data is sent to its ultimate destination. IP routing protocol makes the distinction between hosts and gateways. A host is the end system to which data is ultimately delivered. An IP gateway, on the other hand, is the router that accomplishes the act of routing data between two networks. A router can be a specialized device supporting multiple interfaces, with connections to a different network as shown in Figure 3.23 or a computer multiple interfaces (commonly called a multihomed host) with routing services running in that computer.

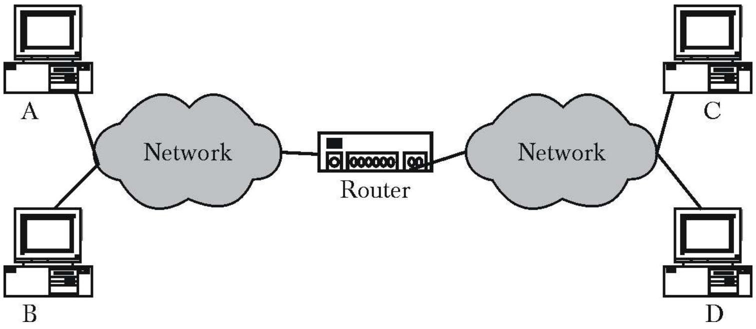


Fig. 3.23 *IP Router Providing Services between Two Networks*

By OSI norms and standards, a gateway is not only a router but also a connectivity device that provides translation services between two completely hybrid networks. For example, a gateway (not a router) is needed to connect a TCP/IP network to an AppleTalk network.

It is important to know that both hosts and IP routers (gateway) perform routing functions and therefore, compatible implementations of the IP protocol are necessary at both ends. In other words, datagrams are submitted either to a host that shares the same physical network with the originating host or to a default gateway for further routing across the network. As such, the IP on a particular host is responsible for routing packets that originate on that host only, fulfilling local needs for routing. A *gateway*, on the other hand, is responsible for routing all traffic regardless of its originator (as long as the TTL field is valid).

A default gateway is a router that a host is configured to trust for routing traffic to remote systems across the network. However, the trusted router must be attached to the same network as the trusting host. A router on a remote network cannot be used for providing the functionality of the default gateway.

Bridging and routing are different in a distinct way. While bridging occurs at the data link layer or Layer 2 of the OSI reference model, routing takes place at the network layer or Layer 3 of the OSI reference model. Routing algorithms are used to determine the optimal routing paths along a network.

Characteristics of Routers

Primarily, the packet switching network intends to deliver data packets transmitted from one host to another over a network. The routing protocols and algorithm tend to provide optimal routes over the network for communicating hosts so that the network could be used effectively and efficiently. In order to guide a data packet to arrive to its intended destination, the packet switching network attempts to provide a best path as well as alternate path through itself to such data packets. To deliver the packets from source host to destination host, there exist few key design issues for such networks so that the network could select a route across the network between end nodes with the following characteristics:

- **Correctness:** It is the responsibility of the routers along with routing protocols and the associated routing algorithm in a packet-switched network, that the routers run among themselves to provide the correct routing decisions to the next router in the way of the destination end. The correctness function of a router should provide a valid route, visible path and safe links. Route validity ensures that if a route exists for a destination, a usable route should also exist for that route in the network. If this condition is not satisfied, the users will experience a failure of end-to-end connectivity as data packets are forwarded along non-existent paths. The visible path or path visibility ensures that the details of an existing path between two nodes should be propagated by the routing protocol. A lack of path visibility will prevent two connected nodes from learning routes between them. Link safety ensures route availability without taking into consideration the order in which routing messages are exchanged.

  Briefly, the objective of 'correctness' is to ensure correct routing so that the packets reach their end points.

- **Simplicity:** It implies the routing protocol's ability to reduce the path computation complexity. In other words, the routing should provide simple methods to compute paths to the destinations so that the overhead is as low as possible. As the complexity of the routing algorithms increases, the overhead also increases.

- **Robustness:** It is the property of a network which defines the expectancy of a network to run continuously for many years adhering well with changes in physical topology and traffic pattern. Due to the change in topology and traffic pattern, there may be some local failure or overload in the network. Robustness is, therefore, the capability of the network to route the packets to the destination through some routes in case of hardware and software failures.

NOTES

- **Stability:** It describes the ability to tackle the changing conditions of the network without influencing network performance. At the time of the change in network conditions, the network should not be too lethargic nor should it become unstable or oscillatory. Briefly, the routing algorithms should provide stability under all possible circumstances.
- **Fairness:** Fairness and optimality are correlated. If you try to improve one, the other will deteriorate. Fairness refers to equal priority given to all nodes or hosts on the network for transmitting their packets; for example, to increase the performance of a network in terms of throughput, the performance criteria are set in such a manner that network policy provides higher priority to the transfer of packets between nearby hosts compared to transfer between remote hosts. Hence, this appears to be unfair to the host that wants to transfer information to remote hosts. This is generally done on a first-come first-serve basis.
- **Optimality:** The routing algorithms should provide optimal throughput and least mean packet delays.
- **Efficiency:** The routing algorithm keeps on adding some processing overhead at each node including transmission overhead. Such overhead tends to deteriorate network efficiency.

Performance Criteria

Primarily, a network is designed to share information among remote hosts over the network efficiently with least cost. Therefore, network design calls for meeting some performance criteria. These are:

(i) Selection of route

(ii) Distance (minimum number of hops)

(iii) Least cost

(iv) Throughput

The selection of a route to the destination is determined differently. The simplest way is to choose the route which passes through the least number of nodes. Such routes are called *minimum hop routes*. The advantage of a minimum hop route is that it involves least network resources. Another performance criterion is the least cost routing in which each route or link is associated with a cost and a least cost path is chosen for delivery of the packets. Between the minimum number of hops and least cost routing, the algorithm used for determining the minimum path is straightforward and takes almost the same processing time as least cost. However, the least cost routing is more flexible than minimum hop; the least cost algorithm is therefore used for determining the route. Another important performance criterion is the throughput. It defines the ability of the network to clear the number of packets per second from the network.

3.8.1 Strategies for Routing

It is one of the simplest routing techniques in which a permanent path from one node to another node is determined with the help of the least cost algorithm. In this algorithm, a least cost route is configured permanently. The fixed route is changed only when there is some change in the topology of the network. Obviously, the link costs are not calculated based on the dynamic variables, such as the traffic pattern. However, the link may be designed based on the expected traffic.

Figure 3.24 shows the implementation of fixed routing. At the network operation center, a central matrix showing a least cost path from one to node to the other nodes is created and stored. Table 3.5 illustrates destination pair of nodes for every source, the identity of the next node on the network. It is found to be quite cumbersome to identify all nodes for each pair of nodes. Therefore, only the next hop or node for each pair of node is identified. In Figure 10.1, let us consider a link X-A and the remaining networking from A to Y as R_1. In the next step, the least cost value of A-Y link is determined and defined as R_2. If $R_1 > R_2$ then the X-Y route can be improved by using R_2. If $R_1 < R_{2,}$ then R_2 is not the least cost value of the network from A to Y. Therefore, $R_1 = R_2$. This provides a way to identify the next node, not the entire network along the network.

In Figure 3.24, in reaching node 6 from node 1, a packet has to pass through either node 2, 3 or 4. Again, the route to node 6 from node 2 goes via node 5. Similarly, the route to node 6 from node 3 goes via node 4. The complete route from X to Y is either through node 1, 2, 5 and 6 or 1, 4 and 6. Out of these two routes, the least cost route is determined. According to the performance criteria, the least cost route is considered better than the minimum hop route. Therefore, it may not be true that a route through node 1, 4 and 6 will be the least cost route as compared to a route through nodes 1, 2, 5. An intuitive conclusion says that links emanating from node 4 will have much more weight than links emanating from other nodes. Therefore, the route passing through node 1, 2, 5 and 6 is considered the optimal path from X to Y.

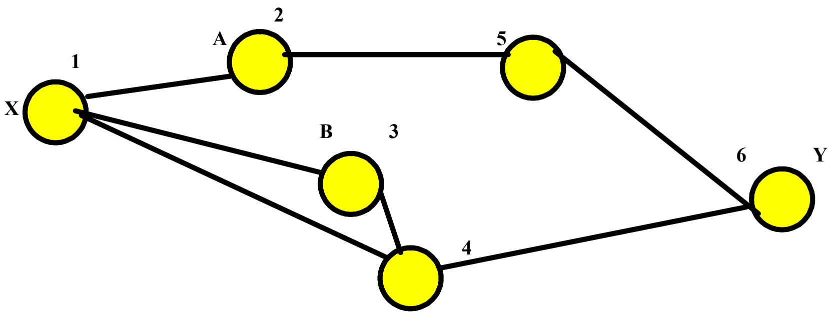


Fig. 3.24 *Implementation of the Fixed Routing*

From Figure 3.24, the matrix at each node and the overall matrix may be determined and stored. Table 3.5 shows these matrices.

Table 3.5 *Destination Nodes for Source Nodes*

Node 1 Directory		Node 2 Directory		Node 3 Directory		Node 4 Directory		Node 5 Directory		Node 6 Directory	
Destination	Next node	Destination	Next node	Destination	Next node	Destination	Next node	Destination	Next node	Destination	Next node
2	2	1	1	1	1	1	1	1	2	1	5
3	3	3	1	2	1	2	1	2	2	2	5
4	4	4	1	4	4	3	3	3	2	3	4
5	2	5	5	5	1	5	6	4	6	4	4
6	2	6	5	6	4	6	6	6	6	5	5

**Central Routing Directory
from Node**

		1	2	3	4	5	6
To node	1	-	1	1	1	2	5
	2	2	-	1	1	2	5
	3	3	1	-	3	2	4
	4	4	1	4	-	6	4
	5	2	5	1	6	-	5
	6	2	5	4	6	6	-

NOTES

Fixed routing does not differentiate between datagram and virtual circuits. All packets between two end-points follow the same route. The major advantage of fixed routing is the simplicity that performs well in a balanced stable load. Its major drawback is its rigid structure and lack of flexibility. In case of link failure or congestion, its response is abysmally poor.

Random Routing

Random routing offers the same simplicity and robustness as provided by the flooding algorithm. However, its major advantage is that it provides far less traffic load than flooding. In random flooding, a node chooses one outgoing path randomly for retransmission of the incoming packet. The incoming link through which the incoming packet arrives is excluded for retransmission. Second, selection of the outgoing link from a particular node can be random or round robin. Random algorithm is refined with probability calculation in which a probability to each outgoing link for selecting the link is calculated based on data rate, or on fixed link costs. Similar to the flooding algorithm, random algorithm also does not require network details because route selection is random. It does not take into consideration the minimum hop or least-cost factors. This enables the network to carry a higher than optimum traffic load, although not nearly as high as for flooding.

The major advantage of this routing is found when a network is highly interconnected because this algorithm uses alternative routes excellently. The probability calculation enables the packet to opt for the least queued link for retransmission.

Adaptive routing

Almost all packet switching networks use some sort of adaptive routing technique. The adaptive routing algorithm is capable of changing routing decisions to reflect changes in network conditions, such as topology and traffic pattern. Routers automatically update routing information when changes are made to the network configuration. The routing details are obtained from adjacent routers or from all routers. It uses the distance, number of hops and estimated transit time as optimization parameters. It is convenient, as it does not involve human intervention in case of changes to the network configuration. Its disadvantage, however, is that the overhead required to send configuration change information can be a heavy burden. This is also known as dynamic routing. In adaptive routing, the following conditions influence routing decisions:

(i) **Failure:** If a node or link breaks up or fails, it is immediately taken out from the route.

(ii) **Congestion:** If a part of the network is heavily congested, the data packet should be avoided to choose such congested links. It should take a path around the congested area rather than through the area of congestion.

Some of the disadvantages of adaptive routing are as follows:

(i) The routing decision is more complex and tends to increase the processing burden on the network nodes.

(ii) An adaptive strategy proactiveness may cause congestion-producing oscillation.

The adaptive algorithm can be further classified as follows:

(i) **Centralized:** A central node in the network collects all details concerning the network topology, traffic pattern and other nodes. Subsequently, the details are forwarded to the routers in the network. The advantage of this is that only one node keeps the details of the network without engaging much resource. However, failure of the central node leads to the failure of the entire network.

(ii) **Isolated:** This refers to the node that decides the routing decision without seeking details from other nodes. Isolation may lead to the selection of a congested route resulting in delay. Popular examples of this routing algorithm are hot potato and backward learning.

(iii) **Distributed:** The node takes decisions after receiving the information from its neighboring nodes.

3.8.2 Shortest Path Routing

Shortest path routing (SPR) is a form of routing which attempts to send packets of data over a network in such a way that the path taken from the sending computer to the recipient computer is minimized. The path can be measured in either physical distance or in the number of hops. This form of routing uses a non-adaptive routing algorithm.

To solve single-source shortest path routing problem, Dijkstra’s algorithm is used. Dijkstra’s algorithm is a graph search algorithm that solves the shortest path problem for a graph with nonnegative edge path costs to produce a shortest path tree. For a given source vertex (node) in the graph, the algorithm finds the path with the lowest cost, i.e., the shortest path between the vertex and every other vertex. It is also used to find costs of shortest paths from a single vertex to a single destination vertex by ending the algorithm once the shortest path to the destination vertex is determined. Dijkstra’s algorithm can be used to find the shortest route between one city and all other cities. Consequently, the shortest path first is widely used in network routing protocols.

The Dijkstra algorithm uses the following steps:

1. A network graph is built to identify source and destination nodes. Thereafter, a matrix, known as the ‘adjacency matrix’ is created in which a coordinate is used to indicate weight. If no direct link exists between two nodes, the weight becomes infinity.

NOTES

2. A status record including the predecessor field (indicating the previous node), length field (giving the sum of the weights from the source to that node) and label field (indicating the status of node whether it is settled or tentative) for each node on the network is constructed in the router.
3. The parameters of the status record for all nodes are initialized to set their label to tentative and their length to infinity.
4. In the next step, a T-node is created. If T-node is a source node, the router changes source node label to permanent which never changes again.
5. The status record for all tentative nodes directly linked to the source T-node are updated.
6. The router selects a tentative node whose weight is lowest. That node becomes the destination T-node.
7. If the new T-node is not the intended destination, the router repeats the step 5 above. If this node is the intended node, the router extracts its previous node from the status record and does this until it arrives at the destination node.

Let us find the best route between routers A and E. It may not always be as easy as it sounds. In certain complicated cases, the best route can only be found using algorithms.

1. The source node (A) has been selected to be the T-node. Its label is therefore permanent (permanent nodes are shown with filled circles and T-nodes with the -> symbol).

Let us consider a simple example to understand the Dijkstra algorithm to identify the best path with least cost. Figure 3.25 is a simple example to illustrate how to reach point B from point A in a weighted graph using the Dijkstra algorithm. The weighted graph has 5 nodes and 6 links with their weights. The weight may be according to the distance, time, cost or any type of appropriate weight.

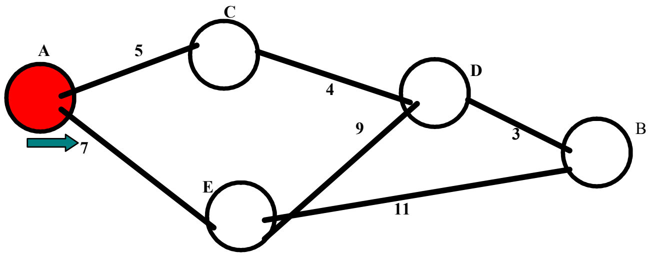


Fig. 3.25 *An Example of Dijkstra Algorithm*

As per these rules of the Dijkstra algorithm, Figure 3.25 is constructed as a network graph with each link having weight. Thereafter, a status table is structured with a row for each node in which the number of columns depends on the map and represents an iteration of the algorithm. Each column will contain the distance from the starting location during that iteration. A * is put in each column when it is assumed that the distance is definite. There exist four possible routes between node A and node B. They are ACDB, AEB, ACDEB and AEDB. Obviously, among these routes, ACDB with weight 12 is the best route due to the smallest weight. Table 3.6 illustrates this simple example. The actual distances from node A are given in the columns on the extreme right.

Table 3.6 *An Iteration of the Dijkstra Algorithm (Figure 3.25)*

Nodes	Values				
A					0*
C	5*				5*
E	7	7*			7*
D	∞	9	9*		9*
B	∞	∞	18	12*	12*

The following steps are used to build Table 3.6:

1. Choose an initial location and obtain all distances to adjacent nodes to fill in the appropriate box of the table. In case of nodes which cannot be reached, fill the box with infinity. In the Figure 3.25, the beginning of the map is taken as node A and the adjacent nodes are C and E. In this example, the source node A is the T-node with a permanent label. The permanent nodes and T-nodes are depicted with filled circles and with the -> symbol respectively.
2. Mark the lowest number in the box with *. A node marked with a * in its row is called the T-node. The initial location itself is always settled or permanent with a distance 0*. In Table 3.6, the node C with a distance 6 is the shortest route because other routes to node C will take a route through E, which has larger distance than direct route to C. The status record of tentative nodes C and E that are directly attached to T-node has been changed. The node C has been selected as T-node because C has less weight and its label has changed to permanent. Figure 3.26 shows this situation.

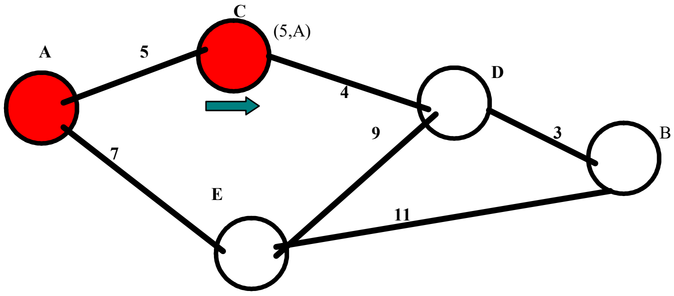


Fig. 3.26 *An Example of Dijkstra Algorithm*

3. The next column is filled by obtaining the adjacent nodes to the T-node C. So far nodes A and C are permanent. Node D is adjacent to node C and node E is adjacent to A. Now their distances are computed from the starting point through the permanent node(s) to which they are adjacent. Their values are filled in the appropriate box of the table. The distance of node E from node A is 7 which is filled in the box. The distance from node A to C is 5 and from node C to D is 4. This provides a value of 9 for node D in the second column. In case, a permanent node is found to be adjacent to more than one permanent node, the least distance is filled in the table.

   Alternatively, Figure 3.27 illustrates the status records of the tentative node that is directly linked to T-node D with a value of 9 and therefore the node is selected as T-node and its label has been changed to permanent.

NOTES

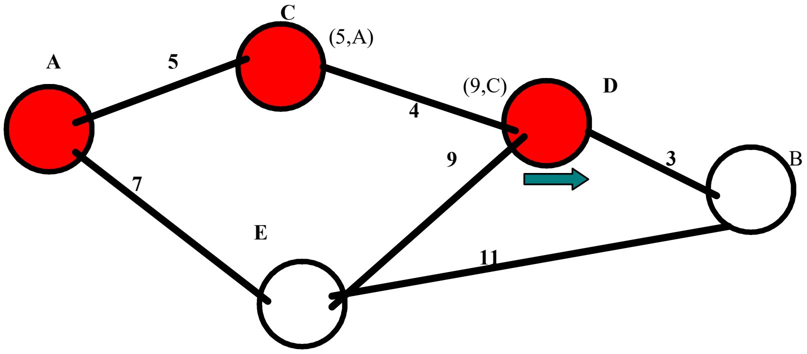


Fig. 3.27 *An Example of Dijkstra Algorithm*

4. Step 2 and 3 is repeated unless all nodes are covered. Now nodes A, C and E are considered permanent. Nodes adjacent to E are B and D and their distances are 18 and 12 which are filled in the respective columns of the table. As node B has 18 in the next column but it has changed to 12 in the next box. Therefore, from location A to location B the shortest value is 12, therefore ACDB is chosen as the shortest path.

   Alternatively, there exist no tentative nodes, the next node B is determined as T-node because node B has the least weight. Now a route is identified. The previous node of B is node D and the previous node of D is node C, while C's previous node is node A. Therefore, the best route is ACDB with total weight 12 which is 5+4+3.

The Dijkstra algorithm works well but it is quite a complicated algorithm and requires a substantially long time for routers to process it at the cost of the efficiency of the network. Secondly, if one router computes wrong information and passes it on to other routers, all routing decisions will become ineffective.

3.8.3 Flooding in Hop

Flooding is yet another static algorithm wherein each packet that comes in is forwarded to every outgoing line except the one from where it came on the router. Thus, it generates an infinite number of duplicate packets. To control the number of packets so generated, a measure namely hop counter is applied. In this method, the header of each packet is decremented at each hop and the packet is discarded till the counter reaches zero. If the source host knows the path from the source to the destination, it initializes the hop counter to the length of the path from the source to the destination. If the sender does not have an idea of the path length, he initializes the counter to the full diameter of the subnet.

Alternatively, a track of the packets flooding the communication link is stored so that they cannot be sent out a second time. The source router attaches a sequence number to each packet received from its hosts. Each router then requires a list per source router indicating which sequence numbers originating at that source have been avoiding any incoming packet on the list. Each list is incremented by a counter, k. This is indicative of the fact that all sequence numbers through k have been looked at. This prevents the list from growing unnecessarily.

NOTES

Figure 3.28 explains the concept of flooding in which a packet is sent from a node to all links emanating from that particular node except the incoming link to that node from which the data packet arrives to the node. A data packet is to be transmitted from node 1 to node 6. The data packet is transmitted from node 1 to all outgoing links to node 2, node 5, node 3 and node 4. The packet from node 2 is transmitted to node 5. Node 5 transmits the packet to node 6 and node 1. Similarly, the same packet will also reach from node 4 to node 6. The packets so delivered have unique identifiers like source node and sequence number (or virtual circuit number and sequence number). According to the unique identifier, node 6 knows how to discard all packets except the first packet.

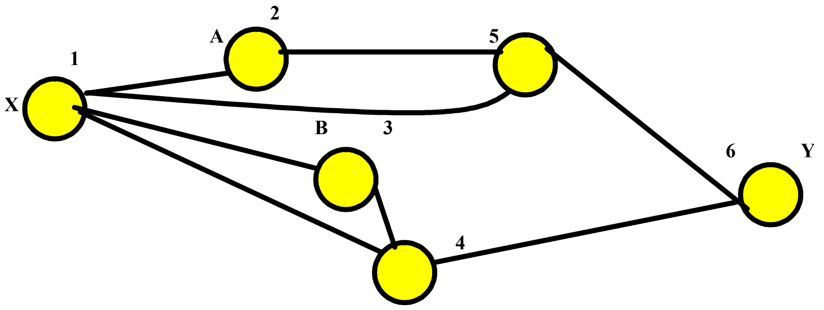


Fig. 3.28 *Implementation of the Fixed Routing*

If Figure 3.28 is observed carefully the retransmission of the packets from one node to other nodes grows incessantly. To prevent this exponential growth of data packets due to retransmission, each node is enabled to remember the identity of those which have already been retransmitted.

The flooding algorithm has three significant characteristics:

(i) Because all possible routes between two end-points are attempted, a data packet will always arrive at the destination irrespective of any breakage in the network, provided a valid path exists between the two end-points. This exhibits the robustness of this algorithm and its applications especially in military-related networks.

(ii) Due to the reason that all paths are attempted, at least a copy of the packet follows a minimum hop path which may be utilized for setting up a route to virtual circuit.

(iii) Each and every node directly or indirectly connected to the source node is visited. This helps in disseminating or broadcasting routing information to all nodes.

Selective Flooding: Selective flooding, which is slightly more practical, is a variation of flooding. Every incoming packet is not forwarded to each line. Instead, incoming packets are forwarded only to those lines going approximately in the right direction.

Flooding is a broadcast protocol. This protocol is used to deliver the message to all nodes in a network. It is considered as the basic mechanism to propagate control messages. Flooding is primarily used in routing. It is based on the performance of path discovery in on-demand routing protocols that are commonly used in multi-hop wireless networks. In mobile ad-hoc networks, the smaller end-to-end delays of shorter routes rarely compensates for their reduced route lifetime, and hop-count is a lousy metric. The flooding schemes are classified into three categories based on the following conditions retained by each node:

- There is no need of sending information to neighbors
- One-hop knowledge of neighbors is necessary
- Two-hop or more hop knowledge of neighbors is also necessary

Flooding most often occurs when a large number of packets (droplets in a stream of data) are flowing through the network that regular data cannot send at a normal speed. Generally, it is a packet of SYN-ACK (synpackets). The initialization of connections between two hosts settled in TCP/IP requires a set of back and forth responses, for example, 'Hi, are you there?' and the reply comes as 'Yes, I am here.' And the next question is asked as 'Are you ready to receive data?' which is replied as 'Yes, I can.' The theory follows as 'the faster the flood, the slower the network'. They get control usually by malware viruses or trojans and often cannot be traced. Innocent flooding can occur when a router is given a circular route to some of the hosts on the network, in which the router asks for the response from a certain host and another router and passes the request to router 1, which then passes it again to router 2, which is sent to router 1. The router 2 uses protocols to test for and close internal loops in a network, which will most often stop flooding. Flooding comes under the category of nonadaptive routing algorithm. This is the simplest way of routing and requires each node in the network to broadcast a packet upon receiving it for the first time. Flooding is widely used in Wireless Sensor Network (WSN) by many applications. The algorithms used for WSN need to be distributed and depend on localized information. The flooding algorithms need to be simple in both computation and communication process. To get efficiency in Internetworking services, flooding algorithms reduce unnecessary redundant transmissions and save energy. When a router receives a multicast packet for a group, it determines the packet status and then it is forwarded to all interfaces except the incoming interface. Routers only need to store recently seen packets. The simplest technique for delivering multicast datagrams to all routers in an internetwork is to implement a flooding algorithm. The flooding procedure begins when a router receives a packet that is addressed to a multicast group. The router employs a protocol mechanism to determine whether this is the first time that it has seen this particular packet or whether it has seen the packet before. If it is the first reception of the packet, the packet is forwarded on all interfaces except the one on which it arrived, guaranteeing that the multicast packet reaches all routers in the internetwork. If the router has seen the packet before, it is simply discarded. A flooding algorithm is very simple to implement, since a router does not have to maintain a routing table and only needs to keep track of the most recently seen packets. However, flooding does not scale for Internet-wide applications since it generates a large number of duplicate packets and uses all available paths across the internetwork instead. This determines that packets have been flooded.

The basic function of flooding is to spread queries across a network, by forwarding queries to all neighboring peers. This simplest type of flooding is known as pure flooding. Let take an example, in which a network is composed by 7 peers (Peer 0, Peer 1, Peer 2, Peer 3, Peer 4, Peer 5, Peer 6 and Peer 7). Peer 0 sends a query to its adjacent neighbor in step 1. Peer 5 is constitutes the relevant resource. Peer 2 and Peer 3 receive the query in Step 2 and forward the query to the neighbors except Peer 0, where the query is coming from. Peer 2 and Peer 3 again receive the query as they did in Step 3. These two peers get an error message so that the query is not being processed. Peer 1, Peer 4 and Peer 6 to process the query same as Peer 2 and Peer 3 did in the last step. Peer 5 has only one neighbor, i.e., Peer 2 and it is the only peer that sends a query therefore Peer 5 has no option but to send the query. Thus, it sends a successful query hit message. It then forwards the result message to Peer 2 which is the last peer in the trail of the query. In Step 4, three events take place. The first event reveals that Peer 2 transfers the query result of Peer 5 to the query issuer, Peer 0. The second event reveals that peer 1 and Peer 6 send the error message. The third event reveals that both do not forward queries. Flooding is a simple blind routing method that uses query forwarding to make all peers accessible to the query. The produced query result is then transferred to the query issuer according to the message trail. However, this whole method exploits a lot of bandwidth to ensure the robustness.

A broadcasting method is used for a data packet that is destined for multiple hosts. Broadcasts can occur at the data link layer and the network layer. Data-link broadcasts are sent to all hosts attached to a particular physical network. Network layer broadcasts are sent to all hosts attached to a particular logical network. The Transmission Control Protocol/Internet Protocol (TCP/IP) supports the following types of broadcast packets:

All Ones: By setting the broadcast address to all ones (255.255.255.255), all hosts on the network receive the broadcast.

Network: By setting the broadcast address to a specific network number in the network portion of the IP address and setting all ones in the host portion of the broadcast address, all hosts on the specified network receive the broadcast. For example, when a broadcast packet is sent with the broadcast address of 131.108.255.255, all hosts on network number 131.108 receive the broadcast.

Subnet: By setting the broadcast address to a specific network number and a specific subnet number, all hosts on the specified subnet receive the broadcast. For example, when a broadcast packet is set with the broadcast address of 131.108.4.255, all hosts on subnet 4 of network 131.108 will receive the broadcast.

3.9 CONGESTION CONTROL

Congestion is a global issue, involving the behaviour of all the hosts, all the routers, the store-and-forward processing within the routers, etc. Congestion is caused if the input traffic rate exceeds the capacity of the output lines, which occurs if the routers are too slow to perform bookkeeping tasks, such as queuing buffers, updating tables, etc., and if the routers' buffer is too limited. When excessive packets are rushed to the node or a partial part of the network, the performance of the network marketedly decreases and this process is termed as *congestion*. If the number of packets is dumped into the subnet the traffic increases the network performance. At this stage, the network is also no longer able to cope with the high rate at which the losing packets are last, resulting in a complete collapse in performance. As a result, no packets are delivered. To overcome this problem, congestion control algorithms were introduced. Congestion control is the process of maintaining the number of arranged packets in a network below a certain level at which network performance is degraded. Congestion control makes that subnet capable of carrying the offered traffic. Summing the relative delay measurements over a period of data flow gives an indication of the level of queuing at the bottleneck. If the sum of relative delays over an interval is 0, no additional congestion or queuing is present in the network at the end of the interval with respect to the beginning. Similarly, if the successful data packets are summed up from the beginning of a session, then any point of summation is equal to zero, since, the entire data contained in the links are not in the network queues. The congestion control algorithm involves as relative delays from the beginning of a session and then updating the measurements at intervals equal to the amount of time to transmit a window of data and receive the corresponding ACKs. The sum of relative delay is then translated into the equivalent number of packets queued at the bottleneck represented by the sum of relative delays. In other words, the algorithm attempts to maintain the following condition:

$$N_{t_i} = n$$

where

$$N_{t_i} = N_{t_{i-1}} + M_{W_{i-1}}$$

In the above equation, N_{ti} is the total number of packets queued at the bottleneck from the beginning of the connection until t_i; n is the desired number of packets sent and received as per session, to be queued at the bottleneck. $M_{Wi\text{-}1}$ is the additional amount of queuing introduced over the previous window $W_{i\text{-}1}$; and $N_{t1} = M_{W0}$.

Applications which perform congestion control make more efficient use of the network and should generally see better performance because of it. Congestion control algorithms prevent the network from entering congestive collapse. Congestive collapse is a situation where the network links are heavily utilized, adversely affecting productivity. The network will soon begin to require applications to perform congestion control and those applications which do not perform congestion control will be harshly penalized by the network, probably in the form of preferentially dropping their packets during times of congestion. The assumption in which statistical multiplexing can be used to improve the link utilization is that the users do not reach their peak rate values simultaneously, but if the traffic demands are stochastic and cannot be predicted, then congestion is unavoidable. Whenever the total input rate is greater than the output link capacity, congestion occurs. When the network becomes congested, the queue lengths might become very large in a short time, resulting in buffer overflows.

Congestion is caused by the shortage of buffer space. The problem can be solved when the cost of memory becomes cheap enough to allow large memory. Larger buffers are useful only for very short term congestions and will cause undesirable long delays. The long queue and long delay introduced by a large memory is undesirable for many applications.

Congestion is caused by slow links. This problem can be solved by introducing high-speed links. However, this is not always the case as sometimes increase in link bandwidth can aggravate the congestion problem because higher speed links may make the network unbalanced. If two sources begin to send data to the destination at their peak rate, congestion will occur at the switch. Higher speed links can make the congestion condition in the switch even worse.

Another reason for congestion is slow processors. However, faster processors will transmit more data per unit time. If several nodes begin to transmit to one destination simultaneously at their peak rate, the target will soon be overwhelmed. Congestion is a dynamic problem and any static solutions are not sufficient to solve the problem. All other issues, such as buffer shortage, slow link, slow processors are symptoms and are not caused by congestion. Proper congestion management mechanisms are more important than ever.

Adding more memory may help till a point, but if routers have an infinite amount of memory, congestion gets worse. Flow control relates to the point-to-point traffic between a given sender and a given receiver. A situation requiring flow control refers to the fiber optic network with a capacity of 1000 gigabits/sec on which a supercomputer was trying to transfer a file to a personal computer at 1Gbps. A situation requiring congestion control is to store-and-forward network with 1-Mbps lines and 1000 large minicomputers, in which half of the part tries to transfer files at 100 kbps to the other half. Figure 3.29 describes flowchart in which the basic algorithm is used for congestion control.

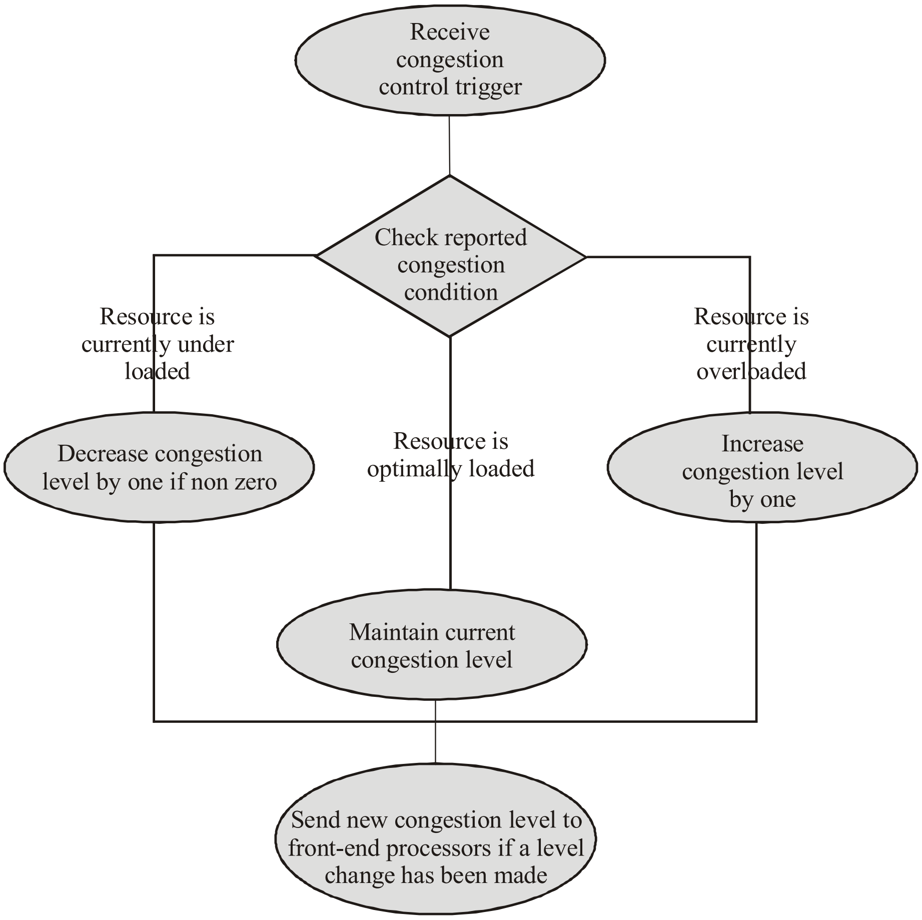

Receive congestion control trigger
Check reported congestion condition
Resource is currently under loaded
Resource is currently overloaded
Decrease congestion level by one if non zero
Resource is optimally loaded
Increase congestion level by one
Maintain current congestion level
Send new congestion level to front-end processors if a level change has been made

Fig. 3.29 *Flowchart of Congestion Control Algorithm*

Figure 3.29 shows that the congestion control task receives a congestion trigger if any of the resources being monitored crosses a congestion threshold, i.e., congestion has either risen above a threshold or fallen below a threshold. New congestion is sent to the front-end processor, if a level change has been made. On the basis of the reported congestion condition, the congestion control task applies a congestion level in the following way:

- **Resource overloaded:** If the monitored resource is overloaded, the traffic in the system needs to be reduced. Thus, the congestion control task increases the throttling of traffic by increasing the congestion level by one.
- **Resource optimally loaded:** If the monitored resource is loaded just right, i.e., to its optimum usage, the congestion level needs to be maintained. Therefore, no change is made to the congestion control level. Hence, no congestion control action is taken.
- **Resource underloaded:** If the monitored resource is underloaded, the system can handle more traffic than is being currently offered. Thus, the congestion control task decreases the throttling of traffic by decreasing the congestion level by one.

If the congestion level has changed in the previous step, the task needs to implement the level change by asking the front-end processors to increase or decrease traffic blocking. The front-end processors take the appropriate action, which will result in change in traffic load. Changes in load handled by the system will finally result in further congestion triggers, thus bringing us back to the starting point.

3.9.1 General Principles of Congestion Control

Growing demand of computer usage requires efficient ways of managing network traffic to avoid or limit congestion in cases where increases in bandwidth is not desirable or possible. It is generally accepted that network congestion control problem remains a critical issue and a high priority, given the growing size, demand and speed (bandwidth) of increasingly integrated services network. ATM is also influenced with the performance of a vast majority of congestion control schemes proposed for the solution of the available bit rate (ABR) problem that not has been proven analytically. In part, due to lack of a structured approach and a strong theoretical foundation in stabilizing controlled systems. The proposed schemes are developed using intuition and simple non-linear designs. Using simulation, these simple schemes are demonstrated to be robust in variety of scenarios. Since these are designed with significant non-linearity, based mostly on intuition, for example, two-phase slow start and congestion avoidance dynamic windows, binary feedback, etc., refers to the analysis of a closed loop behaviour, which is difficult, if possible, even for single control loop networks. From a control theory point of view, all solutions to problems in complex systems, such as computer networks, can be divided into the following groups:

- **Open loop:** The solutions solve the problem by good design, in essence, to make sure the problem does not occur in the first place. Tools include factors such as when to accept new traffic, when to discard packets and which packets to discard, and how to schedule packets at various points in the network. It is helpful in making decisions without regard to the current state of the network.
- **Closed loop:** The solutions are based on the concept of a feedback loop which consists of various parts. In the first part, the system is used to detect when and where congestion occurs. The second part passes the information

to places where actions can be taken and the third part adjusts the system operation to correct the problem.

NOTES

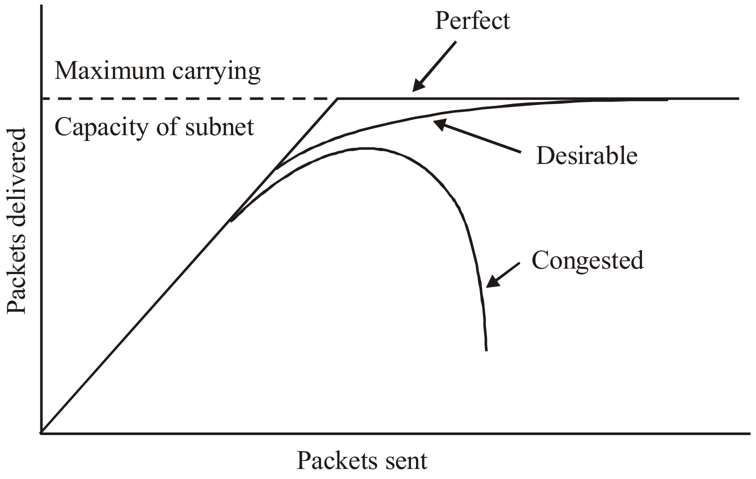


Fig. 3.30 *Congestion Graph*

Figure 3.30 shows that if too much traffic is offered, congestion sets in and the performance degrades sharply. Only perfect data packets are transmitted across the network. In Figure 3.30 congested data packets are tilted, which implies that these data packets are not going to be transmitted. The chief metrics for monitoring the subnet for congestion can be of the following types:

- The percentage of all packets discarded for lack of buffer space
- The average queue lengths
- The number of packets that time out and are retransmitted
- The average packet delay
- The standard deviation of packet delay

The monitored congestion information is propagated in the following way:

- The router detecting the congestion sends a separate warning packet to the traffic source.
- A bit or field can be reserved in each packet. When a router detects a congested state, it fills in the field in all outgoing packets to warn the neighbours.
- Hosts or routers send probe packets out periodically to explicitly ask about congestion and to route traffic around problem areas.

When congestion takes places, buffers get full, so packets are discarded leading to more retransmissions and less packets delivered to their destinations. Adding memory might help, but then the queues get longer leading to more time-outs and retransmissions. Congestion thus tends to feed upon itself and become worse, leading to collapse of the system. Congestion control involves insuring that the subnet is able to carry the offered load. This is a global issue that involves the behaviour of all hosts and routers. In contrast, flow control is related to the point-to-point traffic between a sender and a receiver, ensuring that the sender is not overloading the receiver. There are many policies on different layers that affect

NOTES

congestion. An important issue here is the setting for timers, if they are set too low extra retransmissions occurs. For this adaptive setting of timers is required. Sliding window protocols using selective repeat give less retransmissions than using the 'Go-Back-N' control. Using piggybacking for ACKs reduces the number of packets, but adds an extra timer involving a chance of extra retransmissions. Using a smaller window size reduces the data rate and thus helps fight congestion. When virtual circuits are used it is easy to deny new connections in case congestion is near. A routing algorithm can help avoid congestion by spreading traffic over all possible routes, instead of selecting the best one. A sequence of virtual circuits is set up from the source through one or more gateways to the destination. Each gateway maintains tables telling which virtual circuits pass through it, where they are to be routed, and what the new virtual circuit number is. Once data packets begin flowing, each gateway relays incoming packets, converting between packet formats and virtual circuit (VC) numbers as needed. As all packets must transverse the same sequence of gateways, they arrive in order.

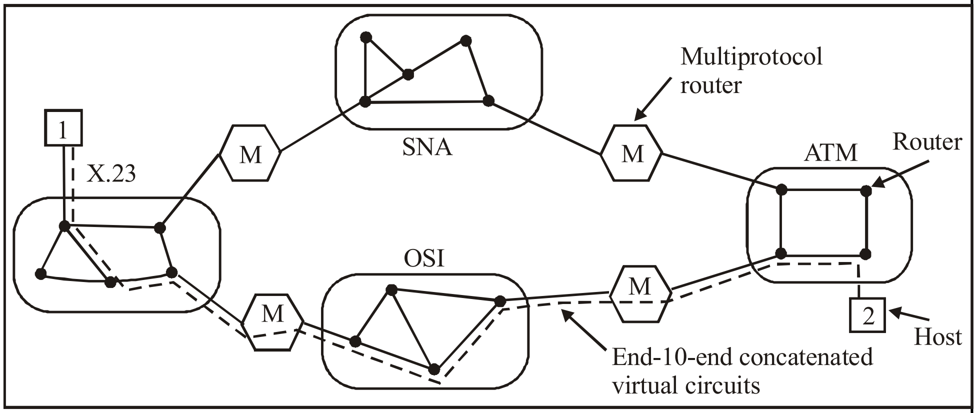


Fig. 3.31 *Concatenating Virtual Circuit*

This scheme works best when all the networks have roughly the same properties. It can not be used if one of the subnets does not offer VCs but only datagrams.

Check Your Progress

16. What is the purpose of a router?
17. Name the different types of routing algorithms.
18. Give one characteristic of flooding algorithm.

3.10 DEADLOCKS

A deadlock is a situation in which some processes wait for each other's actions indefinitely. In real life deadlocks can arise when two processes wait for phone calls from one another or when persons crossing a narrow bridge in opposite directions meet in the middle of the bridge. Deadlock is more serious than indefinite postponement or starvation because it affects more than one job. Because resources are tied up in deadlocks, the entire system is affected.

Processes involved in a deadlock remain blocked permanently which affects the throughput, resource efficiency and the performance of the operating system. A deadlock can bring the system to standstill.

Operating system handles only deadlocks caused by sharing of resources in the system. Such deadlocks arise when some conditions concerning resource requests and resource allocations are held simultaneously.

Deadlock detection mechanism detects a deadlock by checking whether all conditions necessary for a deadlock hold simultaneously. The deadlock prevention and deadlock avoidance ensure that deadlocks cannot occur, by not allowing the conditions for deadlocks to hold simultaneously.

The most common example for deadlock is a traffic jam. In the example (refer Figure 3.32) shown below, there is no proper solution to a deadlock; no one can move forward until someone moves out of the way, but no one can move out of the way until either someone advances or a rear of a line moves back. Only then can the deadlock be resolved.

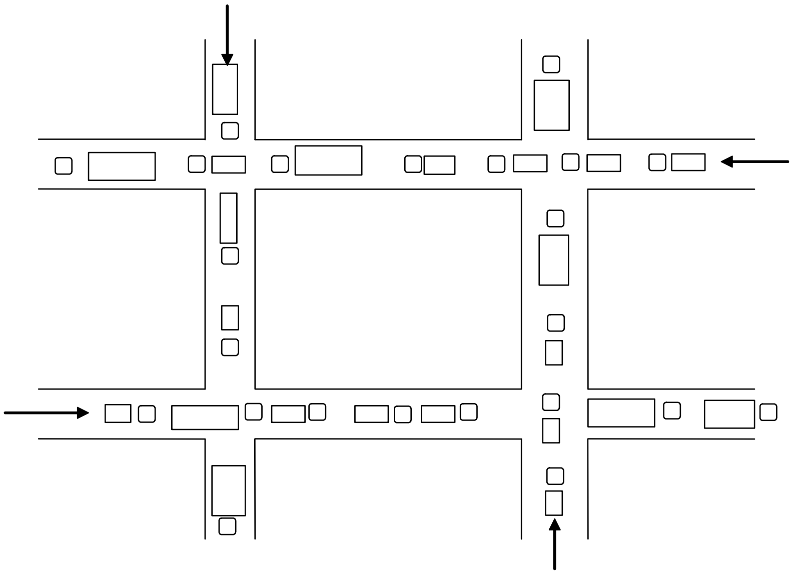


Fig. 3.32 *A Classic Case of Traffic Deadlock*

Thus, we can say that a deadlock refers to a situation in which two or more competing actions are waiting for the other to finish and thus neither ever does this. For example, consider about the two trains approaching each other at a crossing. In this situation, both the trains stop and none of them can restart until the other has gone.

In computer science, deadlock refers to a specific condition when two or more processes are each waiting for another to release a resource or more than two processes are waiting for resources in a circular chain.

Following are the examples of deadlock:

- The occurrence of deadlocks is common in the multiprocessing system. The reason for this is that in the multiprocessing system, several processes have to share a specific type of mutually exclusive resource known as a *software,* or *soft,* lock. There often exists a *hardware lock* (or *hard lock*) in computers that intend for the *time sharing* and/or *real time* markets. This lock provides an *exclusive access* to processes. This leads to a forced serialization. Deadlocks create troubles as we lack a general solution to this problem.

- Think of two people drawing diagrams with only one pencil and one ruler between them. If one person possesses the pencil and the other possesses the ruler, this would lead to a deadlock if the person having the pencil needs the ruler and vice versa. As it is not possible to satisfy both the requests, a deadlock is inevitable.
- In case of telecommunications, deadlock is a little more complex. Here, deadlock occurs when neither of the processes meets the condition for moving to another state and each communication channel is empty. The second condition is ignored in case of other systems but is very important in the context of telecommunications.
- We may consider an example of a deadlock in database products. Client applications using the database may need an exclusive access to a table. To acquire such an access, a *lock* may be demanded by the applications. Think of a client application holding a lock on a table and attempting to obtain the lock on a second table which is already held by a second client application. This may lead to deadlock if the second application tries to obtain the lock possessed by the first application. However, this particular type of deadlock is easily prevented, for example by using an *all-or-none* resource allocation algorithm.

When a Deadlock Occurs?

A deadlock occurs when the following four conditions are met:

- **Mutual Exclusion:** Each resource is allocated to only one process at any given point of time.
- **Hold and Wait:** The previously granted resources are not released by processes.
- **No Preemption:** The previously granted resources are not taken away from the processes which hold them.
- **Circular Wait:** There exists a chain of two or more processes. These processes should exist in such a way that each process in the chain holds a resource requested by the next process in the chain; there must exist a set (P0, P1, P2, P3,……., Pn) of waiting processes such that P0 is waiting for a resource that is held by P1, P1 is waiting for a resource that is held by P2, ... Pn–1is waiting for a resource that is held by Pn, and Pn is waiting for the a resource that is held by P0 (refer Figure 3.33).

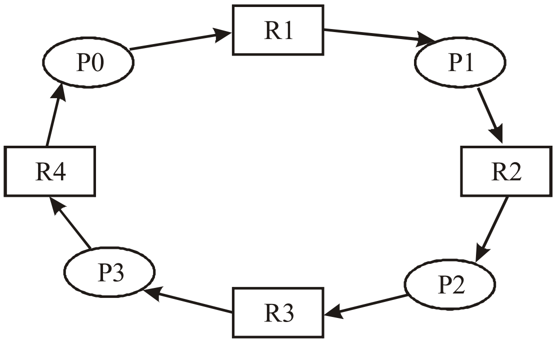


Fig. 3.33 Circular Wait

Basics of Resource Allocation Graph or RAG

Three events concerning resource allocation can occur in a system: request for a resource, allocation of a resource and release of a resource.

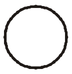
A request can occur when some process Pi makes a request for a resource Ri. If ri is currently allocated to some process Pk, process Pi gets blocked on an allocation event ri. In effect, Pi waits for process Pk so that ri is released. A release event does the task to free ri by Pk. Table 3.7 shows the function of request allocation and release of resources.

Table 3.7 *Request Allocation and Release of Resources*

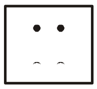
Request	A process requests a resource through a system call. If the resource is free, the kernel allocates it to the process immediately; otherwise, it changes the state of the process to block.
Allocation	The process becomes the holder of the resource allocated to it. The resource state information gets updated and the process state changes to ready.
Release	A process releases a resource through a system call. If some processes are blocked on the allocation event for the resource, the kernel uses some tie-breaking rule, for example FCFS allocation, to decide which process should be allocated the resource

Symbols used in RAG

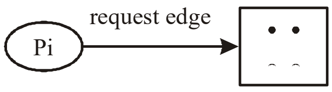
1. Process



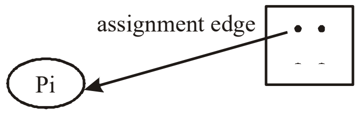
2. Resource type with four instances



3. Pi requests instances of Rj



4. Process Pi is holding an instance of Rj



Basic Facts

1. If a graph contains no cycles, there will be no deadlock.
2. If a graph contains a cycle
   - If only one instance per resource type, then a deadlock will occers.
   - If several instances per resource type, possibility of a deadlock.

Possibility 1 for a Deadlock

The possibility 1 for a deadlock is shown in Figure 3.34.

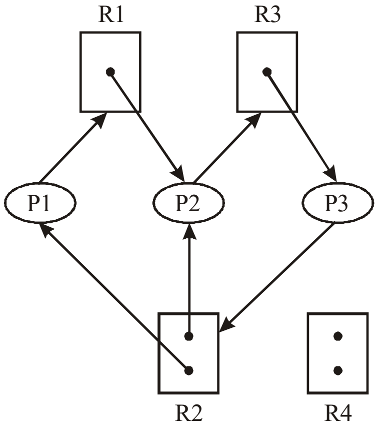


Fig. 3.34 Possibility1 for a Deadlock

E = {P1 → R1, P2 → R3, R1 → P2, R2 → P2, R2 → P1, R3 → P3}

Two possibility for cycles are as follows:

- P1 → R1 → P2 → R3 → P3 → R2 → P1
- P2 → R3 → P3 → R2 → P2

Possibility 2 Cycles but No Deadlock

Two possibility cycles but for no deadlock is shown in Figure 3.35.

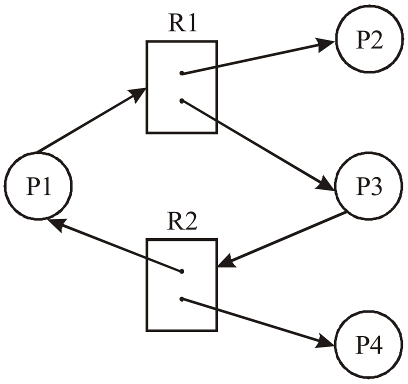


Fig. 3.35 Possibility 2 for a Deadlock

Cycle is P1 → R1 → P3 → R2 →P1.

No deadlock, observe that process P4 may release its instance of resource type R2. That resource can then be allocated to P3 breaking this cycle.

3.10.1 Deadlock Conditions

NOTES

Two processes, A and B (applications) need the same file resources F1 and F2 to complete their task. Initially, file F1 is allocated to process A, and F2 is allocated to process B. Suppose each of the processes starts execution in this state of resource allocation. Now, during the execution, suppose process A requests file F2 which is in use by B, and process B requests file F1 which is in use by A. That is, now process A is waiting for file F2 that is held by process B, and process B is also waiting for file F1 that is held by process A. The processes are in a circular wait state. The chances that such a circular wait will occur can be negated by adopting the strategy of allocating all or none of the resources to each of the processes. That is, if all of the resources needed for a process are available, they are all allocated before the start of execution; otherwise, no resources are allocated and the process has to wait. This negates the *hold* and *wait* condition and hence prevents circular wait.

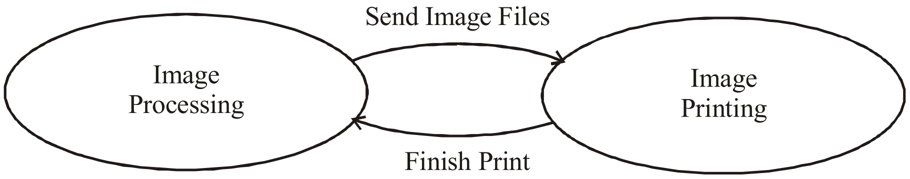


Fig. 3.36 *Cicular Wait May Occur between Image Processing and Printing System*

For example:

Consider the photo processing and printing system given in Figure 3.36. The image processing system sends images in lots of ten for printing to the image printing system and waits for *Finish-Print* message from the image printer. The image printing system accepts images send by the image processing system one by one, and when it receives ten images, it prints them all in a single sheet of paper to minimize the waste of printer stationery. The image printing system sends the *Finish-Print* message back to the image processing system after printing every ten images. Suppose the image processing system had only nine images to be printed in a single sheet and it sends them all and waits for the *Finish-Print* message. The image printer receives the nine images and then waits for the tenth one before starting printing. The image processing system also waits for the *Finish-Print* message for it to send the next lot of images for printing. This is an example of a circular wait (deadlocked state) in which the image printer is waiting for the event (of sending the tenth image) from the image processor and the image processor is waiting for the *Finish-Print* message from the image printer to send the next lot of images.

Suppose the image processor sends the last image indicator of a lot. Then the image printer can start printing with whatever number of images it has received. In this way the cause of circular wait is eliminated.

3.11 QUEUEING THEORY: BASIC DESIGN TECHNIQUES

Queueing theory is the mathematical study of waiting lines or queues. A queueing model is constructed so that queue lengths and waiting time can be predicted. Queueing theory was first defined by Agner Krarup Erlang when he created models to describe the system of Copenhagen Telephone Exchange company, a Danish company. The standard approach and applications include telecommunication, traffic engineering, computer networks, routing and congestion, etc. In addition, the queueing theory has been specifically used in industrial engineering and project management.

Queuing theory deals with problems which involve queuing or waiting, for example in computers it refers to waiting for a response.

Fundamentally, all queuing systems can be categorised into individual sub-systems consisting of entities queuing for some activity, as shown in the following Figure 3.36.

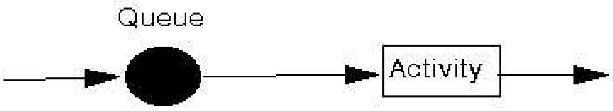


Fig. 3.36 *Sub-System of Queuing*

Another example includes the theory of traffic signals which typically focuses on the estimation of delays and queue lengths that result from the adoption of a signal control strategy at individual intersections, as well as on a sequence of intersections. In the physical layer and DLL (Data Link Layer) analysis, the simplified assumptions of queuing theory can be used for conducting significant analysis, since realistic assumptions include the theoretical analysis.

Computer Scientists and Mathematicians have researched and studied queues based on computer networking technology and formulated their first applications for telephone exchanges, known as the Erlang's loss formula. It was recognised that single queues and networks of queues could be used as performance models of computer systems. In recent times, developments and advancements in most of the computer systems typically structured as the queueing networks. However, various performance analysis is synonymous with queueing theory.

Queueing theory was first used in the late 1960s to model time sharing computer systems. Specifically, single queues were used to examine and evaluate the allocation strategies for the computer CPUs (Central Processing Units). Analysis of the single queues define a qualitative as well as quantitative ability of some aspects of the operating system and disk management system design. Though, in the beginning, these single queue models were not capable to represent systems of interacting computing devices. Subsequent developments in queueing theory consequently studied the interaction between service centres or queues—queueing networks. Recent computer systems are defined as a set of loosely coupled hardware components through which routing packets or transactions get circulated. The success of queueing theory as a performance modelling paradigm depends on the data communication systems of computers.

A queueing system can be completely described by:

(*i*) The input (arrival pattern).

(*ii*) The service mechanism (service pattern).

(*iii*) The queue discipline.

(*iv*) Customer's behaviour.

The Input (Arrival Pattern)

Input describes the way in which the customers arrive and join the system. Generally, customers arrive in a more or less random fashion, which is not possible to predict. Thus, the arrival pattern can be described in terms of probabilities, and consequently, the probability distribution for inter-arrival times (the time between two successive arrivals) must be defined. We deal with those queueing systems in which the customers arrive in Poisson fashion. The mean arrival rate is denoted by λ.

The Service Mechanism

This means, the arrangement of service facility to serve customers. If there is an infinite number of servers, then all the customers are served instantaneously on arrival and there will be no queue.

If the number of servers is finite then the customers are served according to a specific order, with service time as a constant or random variable. Distribution of service time that is important in practice is the *negative exponential distribution*. The mean service rate is denoted by *m*.

The Queue Discipline

It is a rule according to which the customers are selected for service when a queue has been formed. The most common disciplines are:

- First Come First Served (FCFS).
- First In First Out (FIFO).
- Last In First Out (LIFO).
- Selection for Service In Random Order (SIRO).

There are various other disciplines according to which a customer is served in preference over the others. Under priority discipline, the service is of two types, namely pre emptive and non-pre emptive. In pre-emptive system, the high priority customers are given service over the low priority customers; in non-pre-emptive system, a customer of low priority is serviced before a customer of high priority. In the case of parallel channels 'fastest server rule' is adopted.

Customer's Behaviour

The customers generally behave in the following four ways:

(*i*) ***Balking:*** A customer who leaves the queue because the queue is too long and he has no time to wait or does not have sufficient waiting space.

(*ii*) ***Reneging:*** This occurs when a waiting customer leaves the queue due to impatience.

(*iii*) ***Priorities:*** In certain applications some customers are served before others, regardless of their arrival. These customers have priority over others.

(*iv*) ***Jockeying:*** Customers may jockey from one waiting line to another. This is most common in a supermarket.

Transient and Steady States: A system is said to be in a *transient state* when its operating characteristics are dependent on time.

A steady state system is the one in which the behaviour of the system is independent of time. Let $P_n(t)$ denote the probability that there are *n* customers in the system, at time *t*. Then in steady state,

$$\lim_{t \to \infty} p_n(t) = p_n \quad \text{(Independent of } t\text{)}$$

$$\Rightarrow \qquad \frac{dp_n(t)}{dt} = \frac{dp_n}{dt}$$

$$\Rightarrow \qquad \lim_{t \to \infty} p_n'(t) = 0$$

Traffic Intensity **(*or Utilization Factor*):** An important measure of a simple queue is its traffic intensity given by,

$$\text{Traffic Intensity } \rho = \frac{\text{Mean arrival rate}}{\text{Mean service rate}} = \frac{\lambda}{\mu}$$

The unit of traffic intensity is *Erlang*.

Kendall's notation for representing Queueing Models

Generally, queueing model may be completely specified in the symbol form (*a*/*b*/*c*): (*d*/*e*) where,

a = Probability law for the arrival (inter-arrival) time.

b = Probability law according to which the customers are being served.

c = Number of channels (or service stations).

d = Capacity of the system, i.e., the maximum number allowed in the system (in service and waiting).

e = Queue discipline.

3.11.1 Classification of Queueing models

The queueing models are classified as follows:

Model I: (M/M/1): (∞ /FCFS): This denotes Poisson arrival (exponential inter-arrival), Poisson departure (exponential service time), Single server, Infinite capacity and First come first served service discipline. The letter *M* is used due to Markovian property of exponential process.

Model II: (M/M/1): (N/FCFS): In this model, the capacity of the system is limited (finite), say *N*. Obviously, the number of arrivals will not exceed the number *N* in any case.

Model III: Multiservice Model (M/M/S):(∞/FCFS): This model takes the number of service channels as *S*.

Model IV: (M/M/S): (N/FCFS): This model is essentially the same as model II, except the maximum number of customers in the system is limited to *N,* where, (*N* > *S*).

Check Your Progress

19. What is congestion?
20. What is the use of congestion control algorithms?
21. What type of deadlock does an operating system handle?
22. What are the components of queuing system?
23. What are the most common queue discipline?

3.12 ANSWERS TO 'CHECK YOUR PROGRESS'

1. There are two types of data transmission methods that are used to transmit data from its origin to the information processing. These are as follows:
   a. **Offline:** Computers are not connected by communication circuits. Data is transmitted between a terminal and information processing unit through a magnetic tape and magnetic disk packs.
   b. **Online:** Computers are connected by communication circuits. Data can be instantly transmitted between a terminal and information processing unit.
2. The data communication system consists of the following:
   a. Transmitter or Sender of Data: These may be terminals, computers and mainframes, etc.
   b. Medium: The medium, through which the data is transmitted, can be cables, Radio Frequency (RF) wave, microwave, fibre optics, infrared, etc.
   c. Receiver: As the name implies, it is the device, which receives the data transmitted. These are printers, terminals, mainframes, computers, cell phone, etc.
3. DCE is the equipment that interfaces the source with the medium and vice versa. DCE includes modems, DSUs and CSUs and Front End Processors (FEPs).
4. The earliest electronic network is the telephone system.

5. The telephone network consists of the subscriber's line, switchboards, and trunk lines.
6. PSTN is the global compilation of interconnects made for assisting circuit-switched voice communication. The conventional Plain Old Telephone Service (POTS) is provided by PSTN to dwellers and to various enterprises.
7. There are two forms of ISDN service: narrow band and broad band.
8. The World Wide Web is also known as the Web, WWW or W3. It is a global system of hypertext and multimedia services.
9. Electronic mail, or e-mail, allows computer users locally and worldwide to exchange messages. E-mail users have an electronic mailbox into which incoming mail is dropped.
10. Telnet is a virtual terminal emulation facility that allows a user to connect to a remote system as if the user's terminal was hard wired to that remote system.
11. Archive is a program that searches all the FTP sites on the Internet, which are available on its master list, and stores the filenames in a central database. This database is available for users to search.
12. TCP/IP stands for Transmission Control Protocol/Internet Protocol. It was developed with the objective to specify a suite of protocols capable of providing transparent communications interoperability services.
13. The International Organization for Standardization (ISO) took the initiative in setting up OSI.
14. The physical layer determines the type of network design exclusively designed for the physical layer and connected to higher levels such as data link, network, session, transport, presentation and application layers.
15. Physical Layer is the first layer of the model which defines the physical and electrical implementation of the bus. In other words, it defines the hardware and signal-level implementation of the bus.
16. A router is used to manage network traffic and find the best route for sending packets.
17. There are two types of routing algorithm known as adaptive and non-adaptive.
18. An important characteristic of the flooding algorithm is that each and every node, directly or indirectly connected to the source node, is visited. This helps in disseminating or broadcasting routing information to all node.
19. Congestion is a global issue, involving the behavior of all the hosts, all the routers, the store-and-forward processing within the routers, etc.
20. Congestion control algorithms prevent the network from entering congestive collapse.
21. Operating system handles only deadlocks caused by sharing of resources in the system.

22. A queueing system can be completely described by:
    (*i*) The input (arrival pattern).
    (*ii*) The service mechanism (service pattern).
    (*iii*) The queue discipline.
    (*iv*) Customer's behaviour.
23. The most common queue disciplien are:
    (i) FCFS
    (ii) FIFO
    (iii) LIFO
    (iv) SIRO

3.13 SUMMARY

- A data communication system is a computer system that collects data from remote locations through data transmission circuits, then outputs processed data to remote locations.
- A data communication system is a computer system that collects data from remote locations through data transmission circuits, then outputs processed data to remote locations.
- Data circuit terminating equipment is also known as Data Communication Equipment (DCE). DCE is the equipment that interfaces the source with the medium and vice versa.
- DCE is a device that communicates with a DTE device in RS-232C communications.
- Communications software is generally embedded in the computer operating system.
- The earliest electronic network is the telephone system.
- The telephone network consists of the subscriber's line, switchboards, and trunk lines.
- A computer can be connected permanently to the Internet using leased lines.
- PSTN or Public Switched Telephone Network relates to the public telephone network.
- PSTN require 64kbps channel as the vital digital circuit which also known as digital signaling 0/DS0.
- Public Switched Data Network (PSDN) is a network that is accessible to the public. It assists packet-switched data as well as PSTN.
- ISDN which is short for Integrated Services Digital Network is a set of CCITT/ITU standards used for digital transmission over ordinary telephone copper wire and other media.
- ISDN is a network architecture in which digital technology is used to convey information from multiple networks to the end-user. This information is end-to-end digital.

- Broadband ISDN Service is a digital service in excess of 1.544 Mbps. This digital service can be in the form of Frame Relay, SMDS, or ATM. Broadband ISDN is the service of the future.
- Products for ISDN technology from different vendors even with similar features and options may create some compatibility issues.
- Internet Service Provider (ISP) is a company that access internet services. This service provider provides a software package in which you get registration with the providing services.
- Commercial ISPs easily access and communicate with individual or various organizations across net.
- Wide Area Networks (WANs) connect larger geographic areas, such as New Delhi, India, or the world.
- The Internet, WWW and Information Super Highway are terms which have deep impact in the lives of millions of people all over the world.
- The word Internet is an acronym of the word 'internetwork' or 'interconnected network'. Therefore, it can be said that the Internet is not a single network, but a collection of networks.
- ARPA net was built by DARPA as described earlier. This initiated the packet switching technology in the world of networking and therefore is sometimes referred to as the 'grand-daddy of packet networks'.
- The Internet is becoming a necessary tool rather than a convenient tool in society. It has proved its utility in all walks of life, such as education, economy, and socio-political arenas.
- Electronic mail, or e-mail, allows computer users locally and worldwide to exchange messages.
- Telnet is a virtual terminal emulation facility that allows a user to connect to a remote system as if the user's terminal was hard wired to that remote system. The file transfer facilities are usually provided for by a mechanism known as the File Transfer Protocol (FTP).
- SMTP is a defined standard for e-mail over the TCP/IP protocol and therefore is widely used on the Internet.
- Routing refers to the process of selecting the shortest and most reliable path intelligently over which to send data to its ultimate destination.
- Network architecture defines the communications products and services, which ensure that various components work together.
- OSI reference model divides the required functions of the network architecture into several layers and defines the function of each layer.
- Data communications processes allocate memory resources, commonly known as communication buffers, for the sake of transmission and reception of data.
- The physical layer determines the type of network design exclusively designed for the physical layer and connected to higher levels such as data link, network, session, transport, presentation and application layers.

- A session layer network is designed to occasionally merge the session and transport layers. Therefore, network designers design the hierarchical topology for this layer.
- The OSI model of data communication was developed in 1984 by the International Standardization Organization (ISO).
- It is the responsibility of the routers along with routing protocols and the associated routing algorithm in a packet-switched network that the routers run among themselves to provide the correct routing decisions to the next router in the way of the destination end.
- Congestion is a global issue, involving the behaviour of all the hosts, all the routers, the store-and-forward processing within the routers, etc.
- Congestion is caused by the shortage of buffer space.
- Flow control relates to the point-to-point traffic between a given sender and a given receiver.
- When congestion takes places, buffers get full, so packets are discarded leading to more retransmissions and less packets delivered to their destinations.
- A deadlock is a situation in which some processes wait for each other's actions indefinitely.
- Processes involved in a deadlock remain blocked permanently which affects the throughput, resource efficiency and the performance of the operating system.
- In computer science, deadlock refers to a specific condition when two or more processes are each waiting for another to release a resource or more than two processes are waiting for resources in a circular chain.
- Queueing theory is the mathematical study of waiting lines or queues.
- Queueing theory is the mathematical study of waiting lines or queues. A queueing model is constructed so that queue lengths and waiting time can be predicted.
- A queueing system can be completely described by:
  (*i*) The input (arrival pattern).
  (*ii*) The service mechanism (service pattern).
  (*iii*) The queue discipline.
  (*iv*) Customer's behaviour.
- The most common disciplines are:
  - o First Come First Served (FCFS).
  - o First In First Out (FIFO).
  - o Last In First Out (LIFO).
  - o Selection for Service In Random Order (SIRO).

3.14 KEY TERMS

- **Receiver:** It is the device, which receives the data transmitted. These are printers, terminals, mainframes, computers, cell phone, etc.
- **Public Switched Telephone Network (PSTN):** It is the global compilation of interconnects made for assisting circuit-switched voice communication.
- **Public Switched Data Network (PSDN)**: It is a network that is accessible to the public. It assists packet-switched data as well as PSTN.
- **ISDN:** It is a network architecture in which digital technology is used to convey information from multiple networks to the end-user. This information is end-to-end digital.
- **Wide Area Networks WAN:** It is defined as a data communications network covering a relatively broad geographical area to connect LANs together.
- **World Wide Web (WWW):** It is a global system of hypertext and multimedia services. WWW is a client-server model based on TCP/IP protocols and consists of browsers as clients and Web servers as servers.
- **Telnet**: It is a virtual terminal emulation facility that allows a user to connect to a remote system as if the user's terminal was hard wired to that remote system.
- **Routing:** It refers to the process of selecting the shortest and most reliable path intelligently over which to send data to its ultimate destination.
- **Deadlock:** It is a situation in which some processes wait for each other's actions indefinitely.
- **Queueing theory**: It is the mathematical study of waiting lines or queues.
- **Processing Delay:** It refers to the time from the packet's arrival at the network layer of the node until it is assigned to an outgoing queue (if routed at the IP layer).
- **Queuing Delay:** It refers to the specific time from when the packet arrives to the queue until it is transmitted.
- **Propagation Delay:** It refers to the time from when the last bit is transmitted until the last bit is received at the destination node.

3.15 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Discuss the process of digital data transmission.
2. What are the objective of Internet2 project?
3. What is the utility of internet?
4. What are the advantages of telephone networks?

5. List the various layers of OSI.
6. State the main advantages of TCP/IP?
7. What is the purpose of a gateway?
8. What are the design goals of routing algorithms?
9. Write the features of RAG algorithm.
10. What are the common queue discipline?
11. Discuss the varius queuing model.

Long- Answer Questions

1. Explain the concept of network architecture.
2. Write a detailed note on data communication system.
3. Describe the various advantages achieved by the layered approach.
4. Briefly describe the various layers of the OSI model.
5. Explain the responsibilities of the seven layers.
6. What are the general principles of congestion control?
7. Elaborate the various policies that effect congestion.
8. Illustrate all the methods used for handling deadlocks.
9. What do you understand by queuing theory? Explain.

3.16 FURTHER READING

Forouzan, Behrouz A. *Data Communications and Networking*. New Delhi: Tata McGraw-Hill, 2004.

Stallings, William and Richard Van Slyke. *Business Data Communications*. New Jersey: Prentice Hall, 1998.

Black, Uyless. *Computer Networks*. New Jersey: Prentice Hall, 1993.

Stallings, William. *Data and Computer Communications*. New Jersey: Prentice Hall, 1996.

Tanenbaum, Andrew S. *Computer Networks*. New Jersey: Prentice Hall PTR, 2002.

Stallings, William. *Data and Computer Communications*. NJ: Prentice-rHal, 1996.

UNIT 4 WIDE AREA NETWORK, TCP/IP AND DATA LINK LAYER ADDRESSING

NOTES

Structure

4.0 INTRODUCTION

A wide area network (WAN) is a telecommunications network that extends over a large geographic area for the primary purpose of computer networking. Wide area networks are often established with leased telecommunication circuits. The textbook definition of a WAN is a computer network spanning regions, countries, or even the world. However, in terms of the application of communication protocols and concepts, it may be best to view WANs as computer networking technologies used to transmit data over long distances, and between different networks.

The Internet protocol suite, commonly known as TCP/IP, is the set of communications protocols used in the Internet and similar computer networks. The current foundational protocols in the suite are the Transmission Control Protocol (TCP) and the Internet Protocol (IP).

The Internet protocol suite provides end-to-end data communication specifying how data should be packetized, addressed, transmitted, routed, and received. This functionality is organized into four abstraction layers, which classify all related protocols according to each protocol's scope of networking. From lowest to highest, the layers are the link layer, containing communication methods for data that remains within a single network segment (link); the internet layer, providing internetworking between independent networks; the transport layer, handling host-to-host communication; and the application layer, providing process-to-process data exchange for applications.

The network layer and data link layer are responsible for delivering the data from the source device or sender, to the destination device or receiver. Protocols at both layers contain source and destination addresses, but their addresses have different purposes.The network layer, or Layer 3, logical address contains information required to deliver the IP packet from the source device to the destination device. A Layer 3 IP address has two parts, the network prefix and the host part. The network prefix is used by routers to forward the packet to the proper network. The host part is used by the last router in the path to deliver the packet to the destination device.

In this unit you will learn about Wide Area Network WAN, Transmission Control Protocol/Internet Protocol (TCP/IP) data transmission by TCP and Ethernet, data encapsulation, data routing, and IP Addressing.

4.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Define Wide Area Network
- Explain transmission control protocol/Internet protocol (TCP/IP)
- Understand data transmission by TCP and Ethernet
- Discuss data encapsulation and data routing
- Explain TCP/IP services and application protocols
- Understand and explain IP addressing
- Discuss network layer addresses and subnetting
- Understand Address Resolution Protocol (ARP)
- Elaborate on Domain Name System (DNS).

4.2 WIDE AREA NETWORK

This technology connects sites that are in diverse locations. Wide area networks (WANs) connect larger geographic areas, such as New Delhi, India or the world. WAN has no geographical limits. Dedicated transoceanic cabling or satellite uplinks may be used to connect this type of network. Hence, a WAN may be defined as a data communications network that covers a relatively broad geographic area to connect LANs together between different cities with the help of transmission facilities provided by common carriers, such as telephone companies. WAN technologies function at the lower three layers of the OSI reference model. These are the physical layer, the data link layer and the network layer.

Figure 4.1 illustrates the system of WAN, which connects many LANs together. It also uses switching technology provided by local exchange and long distance carrier.

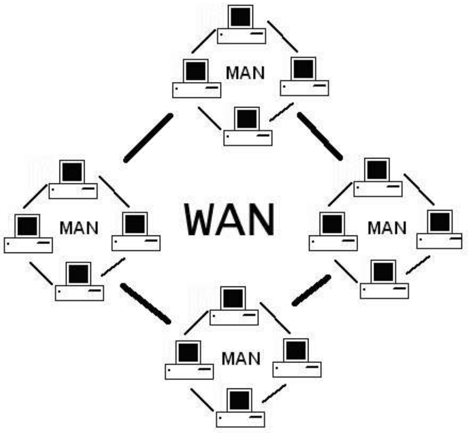


Fig. 4.1 *Wide Area Network*

Packet switching technologies such as asynchronous transfer mode (ATM), frame relay, switched multimegabit data service (SMDS), and X.25 are used to implement WAN along with statistical multiplexing to enable devices to share these circuits.

The difference between MAN and WAN may be understood only from the services being used by them. WAN uses both the local and long distance carrier while MAN uses only a local carrier. Hardware and protocols are same as in case of MAN.

The answer to the confusion between LAN and WAN technologies lies in how data is switched. It is the integration of LAN and WAN that makes the network work. After all, people and machines not only need to be accessible locally, but from different sites as well.

A network is accomplished using three basic components:

- Hardware
- Protocols (software)
- Applications (useful software)

Each of these comprises several layers. In the domain of computer design and networking the concept of layers is important. Each layer protects the layer below from the layer above so that each layer changes with minimal impact on the upper layers. This protection, in some cases, is so proficient that an application may not come to know that it is running on a different hardware. The OSI network model defines seven layers.

The role of computer networks in development is multi-faceted. A computer, along with the necessary networking infrastructure, is required to be connected with either LAN or WAN or Internet or all, so as to play a greater role in e-governance, telemedicine, e-education, e-business, etc. The Internet (internetworking) has become a potent tool for education, productivity and enlightenment. It can improve the quality of life at a relatively low cost. The Government of India had set up ERNET in 1986 to provide TCP/IP connections for education and research communities in India. ERNET established the first TCP/IP computer network in India and offered services like e-mail, Internet surfing, FTP, Telnet. Subsequently, the government liberalized the policies relating with Internet and its backbone. The liberalized policies encouraged many private players like DISHNET, Mantra online, JAIN TV, etc. and other government organizations like NIC, VSNL and MTNL to enter this field to bring the Internet to common people.

The major network infrastructure available in the country has two types of WAN:

1. Terrestrial WAN
2. VSAT WAN

Following are different options for setting up the Intranet, education portal or e-commerce, etc.

1. Leased line
2. Dial Up connection
3. VSAT
4. Radio Link

The role of Internet in development can be seen in the areas of education, economic productivity, health care, democracy and human rights and quality of life besides several others. In education, this contribution is by way of shared databases, organization of conferences, circulation of papers and discussion, collaborative research and writing. Web-based registration, online digital library privileges, other online learning facilities like virtual class rooms and information regarding courses and so forth. Economic productivity may be increased as Internet runs over telephone infrastructure at relatively marginal costs, providing increased economic benefit. Internet enables global communication with suppliers and customers etc. This can help to open global markets in developing nations. In this manner, Internet has facilitated the opening of e-commerce. Internet is being effectively utilized in the health sector. The rapid growth of Internet and related areas like switched leased lines, terrestrial and satellite packet radio and videoconferencing, etc has led to the development of telemedicine. By providing people living under dictatorships with outside information and ideas the Internet encourages democracy. The Internet also enables people to share their ideas and coordinate political activities within their nations. Internet can force transparency in the administration and therefore may be considered a catalyst to encourage human rights in a wider sense. The environment is under stress everywhere, in terms of pollution and limited resources for energy. Internet enables us to substitute communication for transportation and thereby reduce pollution and save energy and time in the larger interests of mankind.

WAN is the acronym for Wide Area Network and refers to a network used to connect different equipment from remote areas. Normally, network services are provided by a Common Carrier of, for example, a telephone company. Users can use services on rent basis. Available services include telephone network, leased line, packet switched network, X.25, ISDN, frame relay and cell relay.

NOTES

LANs can be extended to a wider area but it can not be extended arbitrarily far or to handle arbitrarily many computers. There is a distance limitations even with extensions. Therefore, other technologies for larger networks are needed. The networks can be broadly divided into three categories namely:

- LAN for a single building
- Metropolitan Area Network (MAN)— single city and
- WAN— country, continent and planet

WAN is composed a number of autonomous computer that are distributed over a large geographical area as shown in Figure 4.1. LAN can be extended across large distances using Satellite Bridge but still this can not accommodate arbitrarily many computers. WAN must be scalable to long distances **and** many computers. Therefore, network must replace shared medium with packet switches to span long distances or many computers. Each switch moves an entire packet from one connection to another. This mechanism is called packet switching. These switches are nothing but a small computer with network interfaces, memory and program dedicated to packet switching function. These packet switches may connect to computers and to other packet switches, typically high-speed connections to other packet switches, lower speed to computers as shown in Figure 4.2. These packet switches can be linked together to form WANs. WANs need not be symmetric or have regular connections, i.e., each switch may connect to one or more other switches and one or more computers.

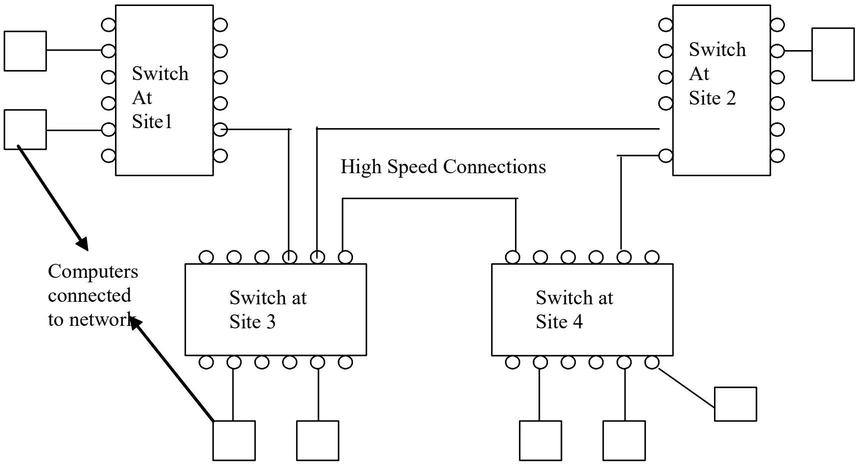


Fig. 4.2 *Packet Switches as Building Blocks to Make a WAN*

Data delivery from one computer to another is accomplished through store-and-forward technology. Packet switch store*s* incoming packet and forwards the packet to another switch or computer that has internal memory. Therefore this can hold packet in queue if outgoing connection is busy.

Difference between WAN and LAN

- With LAN additional expanses are rarely required once it is installed. With WAN, users must continue to pay a communication cost to their contracted common carrier.

- WAN is generally slower in transmission speed. Requesting the same level of speed as with LAN leads to a substantial increase in communication costs.
- The Satellite Bridge can extend LAN across large distances while in case of the WAN, it spans over a wide geographical area.
- LAN still cannot accommodate arbitrarily many computers, WAN must be scalable to long distances **and** many computers.

Network Using WAN and Network Services

The major objective of network design is to select the network service and to determine the transmission speed for the system. Following are the typical examples of network using WAN and network services:

Host to Terminal Connection

The host to terminal connection is a conventional type of connection between a main frame and dumb terminals. This connection is widely used for routine work. Fixed message length, for example, makes estimation of traffic easy. As terminals are relatively slow, telephone network, low speed leased line, packet switching network, ISDN and so on are mainly used. A terminal controller (TC) may be used to integrate two or more terminals for connection with a high-speed line. The TC shown in the Figure 4.3 is used to integrate two or more terminals for connection with a single WAN line.

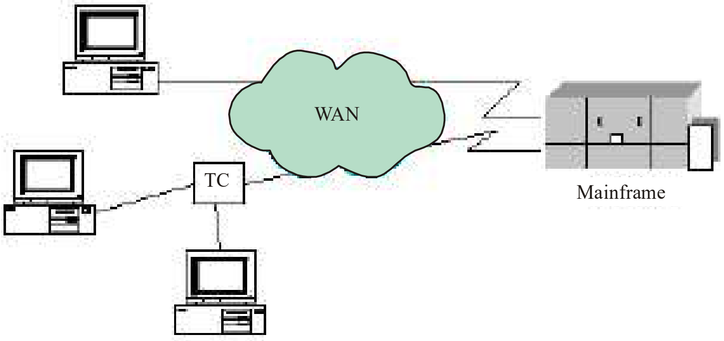


Fig. 4.3 *Host to Terminal Connection*

LAN to LAN Connection

The configuration shown in Figure 4.4 is used to connect LANs that are remote from one another. There is a large difference in transmission speed between LAN and WAN. In addition, using a high-speed line can become substantially expensive. Therefore, proper arrangements must often be made to reduce traffic within the WAN. Leased line, ISDN, and frame relay are mainly used for this connection.

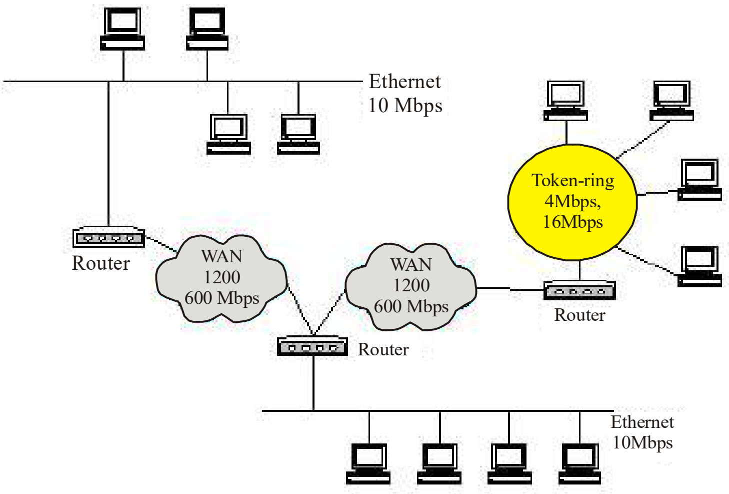


Fig. 4.4 LAN to LAN Connection

Remote LAN connection

The configuration shown in Figure 4.5 is mainly used to connect a remote LAN and PC. Due to the limited number of terminals to be connected and also due to restrictions on allowable expense, the types of WAN that can be used are also limited. In general ISDN is frequently used. Some servers are exclusively designed for the access purpose.

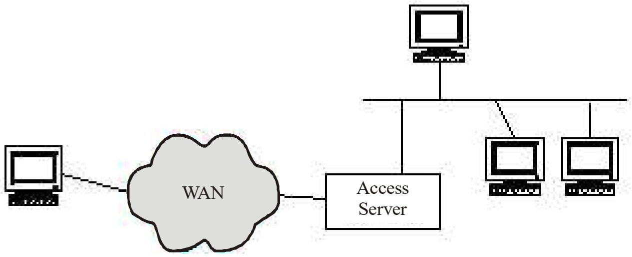


Fig. 4.5 Remote LAN Connection

Router Concepts

A router has two functions as follows:

- **Forwarding Function**— It is a function that allows selection of the appropriate route based on IP header information and sends packets through this route.
- **Filtering Function**— It is a function that allows dumping of invalid packets for a specific network instead of forwarding.

Forwarding Function

The most important processing in forwarding is to determine the route. In general, routing refers to the processing used for determining the next hop (IP address of next router). Here, we consider that the processing for sending packets to the host in the same network is also performed as an integral part of routing. Routing in the same network may be termed as local routing.

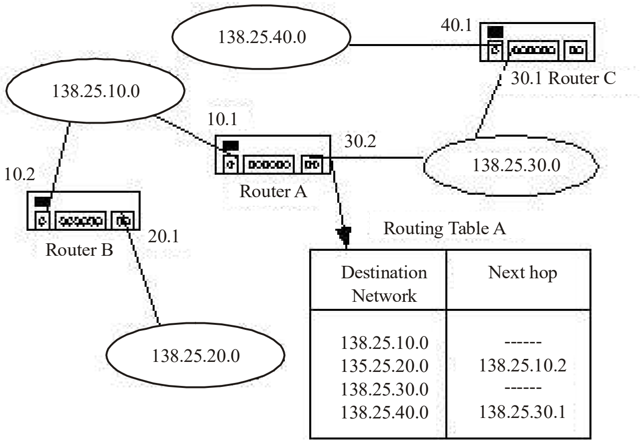
NOTES

The routing is simple if the destination address of the host is in the same subnet. However, if it is not in the same subnet the packet must first be sent to the boundary router. The boundary router references the routing table to determine the next hop. Routing table varies in configuration depending on the routing method used. In this section we will understand the typical hop-by-hop routing table.

The routing table possessed by a router includes combination of destination address and the next hop that corresponds to that address.

Figure 4.6 shows the routing table for router A (address 138.25.10.1). This table lists destination addresses for each local network, and not for each destination host. This table also includes as the next hop (the address of next router) to which the packet must be transferred. If no hops are included, this means that the destination network is directly connected to the router.

When router A receives a packet, it tracks this table to perform routing. For example, if the packets addressed to the host of network 138.25.40.0, then router A sends the packet to router C (138.25.30.1). Router C has a similar routing table so that it can perform routing.



Destination Network	Next hop
138.25.10.0	------
135.25.20.0	138.25.10.2
138.25.30.0	------
138.25.40.0	138.25.30.1

Fig. 4.6 *Forwarding Function*

Filtering Function

The second function available with a router is that it can filter packets to determine whether they can pass through it. For example, consider that there are two networks connected via router as shown in Figure 4.7. Now, we impose the following conditions:

(1) Hosts A and C cannot communicate using UDP packets

(2) Host B cannot use FTP for communication with host D

(3) Host D cannot communicate with host A

(4) Host D can communicate with host B only using TELNET

In this case the network administrator set the router as given below:

Exclusion List:

(Source host = A, Destination host = C, Protocol = UDP)

(Source host = C, Destination host = A, Protocol = UDP)

(Source host = B, Destination host = D, Port = FTP)

(Source host = D, Destination host = A)

Inclusion List:

(Source host = D, Destination host = B, Protocol = TELNET)

NOTES

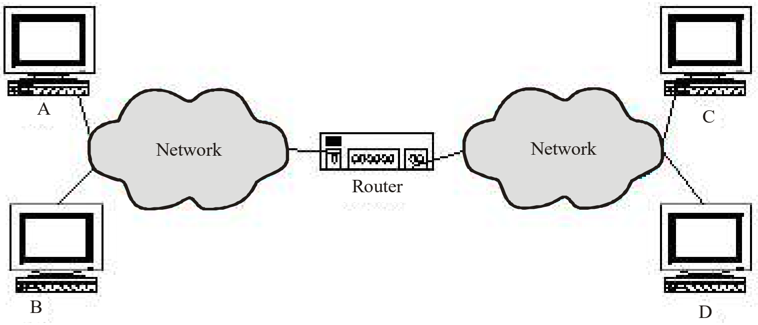


Fig. 4.7 Filtering Function

Hierarchical Routing

Because of the global nature of Internet system, it becomes more difficult to centralize the system management and operation. For this reason, the system must be hierarchical such that it is organized into multiple levels, with several group loops connected with one another at each level. Therefore, hierarchical routing is commonly used for such a system as shown in the Figure 4.8.

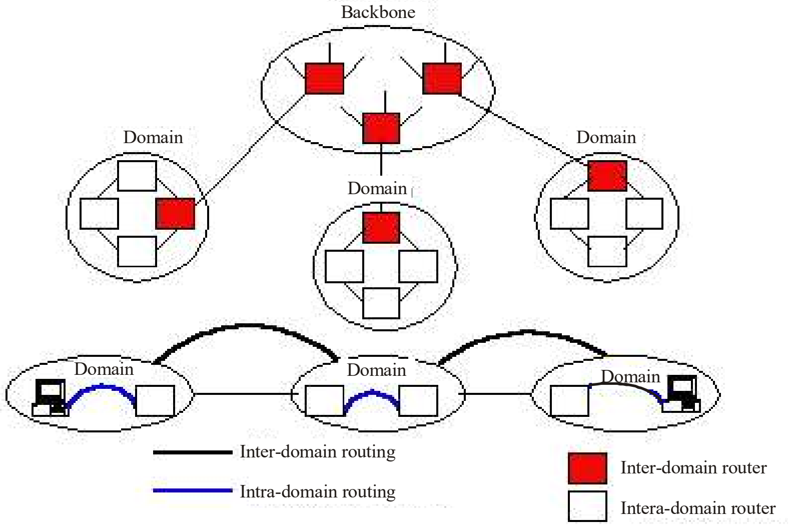


Fig. 4.8 Hierarchical Routing

- A set of networks interconnected by routers within a specific area using the same routing protocol is called domain.
- Two or more domains may be further combined to form a higher-order domain.
- A router within a specific domain is called intra-domain router. A router connecting domains is called inter-domain router.
- A network composed of inter-domain routers is called backbone.

Each domain, which is also called operation domain, is a point where the system operation is divided into plural organizations in charge of operation. Domains are determined according to the territory occupied by each organization.

Routing protocol in such an Internet system can be broadly divided into two types:

- Intra-domain routing
- Inter-domain routing.

Each of these protocols is hierarchically organized. For communication within a domain, only the former routing is used. However, both of them are used for communication between two or more domains.

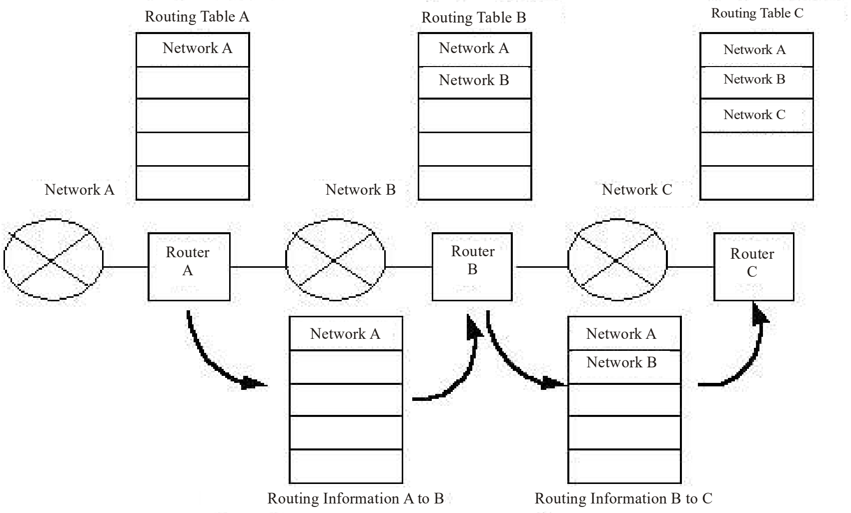
On the pages that follow, we will look at description of Routing information Protocol (RIP), Open Shortest Path First (OSPF), and IS-IS, that are intra-domain protocols. RIP and OSPF will be covered later in detail.

Two algorithms, Distance-Vector Protocol and Link-State Protocol, are available to update contents of routing tables.

Distance-Vector Protocol

Distance vector protocols are RIP and Interior Gateway Routing Protocol (IGPR).

Algorithm where each router exchanges its routing table with each of its neighbors. Each router will then merge the received routing tables with its own table, and then transmit the merged table to its neighbors. This is shown in Figure 4.9. This occurs dynamically after a fixed time interval by default, thus requiring significant link overhead.



Routing Information A to B *Routing Information B to C*

Fig. 4.9 *Routing Method-Distance—Vector Type*

There are problems, however, such as:

(1) If exchanging data among routers every 90 seconds for example, it takes 90 x 10 seconds that a router detects a problem in router 10 items ahead and the route cannot be changed during this period.

(2) Traffic increases since routing information is continually exchanged.

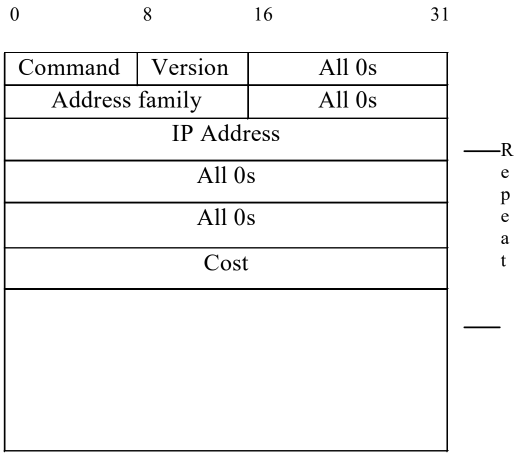
(3) There is a limit to the maximum amount of routing information (15 for RIP), and routing is not possible on networks where the number of hops exceeds this maximum.

(4) Cost data is only the number of hops, and so selecting the best path is difficult.

However, routing processing is simple, and it is used in small-scale networks in which the points mentioned above are not a problem.

RIP (Routing Information Protocol)

RIP is the most widely used routing protocol of distance-vector type today. RIP has been originally designed based on the routing protocol applied to XNS and PUP protocol systems of Xerox (RFC1058). The RIP packet format is shown in Figure 4.10.



Command: Request or response

Address family: Represents address types (protocols). For example, "2" for IP address

Address: Represents destination address. Any of the following types can be included:

- Host address
- Subnet address
- Network address
- Default route

Cost: Expressed by an integer between 1 and 5 to represent the cost of a route to reach the destination address

Fig. 4.10 *RIP Packet Format*

- RIP request is used, for example, by a router upon startup, to inquire of its neighbor router about route information to obtain routing information.
- RIP response includes a destination host address and cost information in the address part. Response is sent to the neighbor router in case of the following:

  (1) Receipt of RIP request

  (2) Regularly

  Response is sent every 30 seconds even if no RIP request is issued. All routers delete route information from their routing table if no route information is received within a specified period of time. This is intended to allow detection of fault of neighbor router.

  (3) In case of changes made to routing table contents

  If changes are made to the routing table because changes to the network configuration have been detected, information relating to these changes is sent to the neighbor router.

NOTES

Link-State Protocol

These are OSPF, IS-IS (Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol).

Algorithm where each router in the network learns the network topology then creates a routing table based on this topology. Each router will send an information of its links (Link-State) to its neighbor who will in turn propagate the information to its neighbors, etc. This occurs until all routers have built a topology of the network. Each router will then prune the topology, with itself as the root, choosing the least-cost-path to each router, then build a routing table based on the pruned topology as shown in Figure 4.11.

In link-state protocols, there are no restrictions in number of hops as in distance-vector protocols, and these are aimed at relatively large networks such as Internet backbones. The load on routers will be large however, since processing is complex.

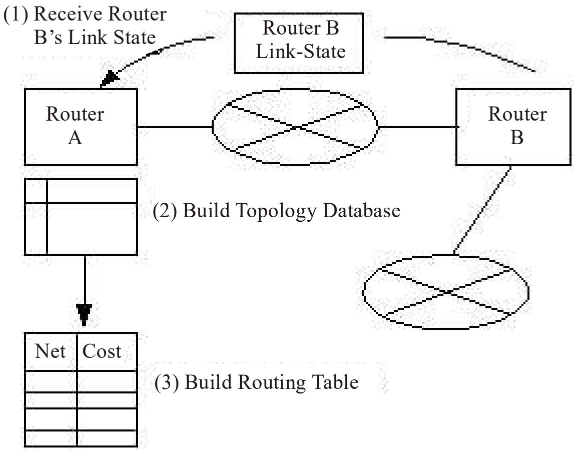


Fig. 4.11 Routing method – Link State type

OSPF (Open Shortest Path First)

OSPF (details in RFC1247) is a link-state type routing protocol developed for use in a large-scale network by eliminating the disadvantages of RIP. This is the only standardized inter-domain protocol for the Internet as of today, and offers the following features. The common part of OSPF packet format is shown in Figure 4.12.

- Compatible with hierarchical topology for network
- Allows use of subnet mask of variable length.
- Allows load distribution when two or more routes are available.
- Supports authorization method for improved security.

In OSPF, each domain is divided into several areas. Detailed configuration of each area can be hidden from other areas. Therefore, routers that belong to the same area have the same network configuration information while routers belonging to other areas have different configuration information. Because one area is composed of subnets with serially assigned addresses, external intervention is not necessary to manage the route to reach each address in that area. It is only necessary to manage the route to that area as an integral route to a series of those addresses.

NOTES

0 8 16 31

Version	Packet type	Packet Length
Router ID		
Area ID		
Check sum	Authentication type	
Authentication Data		
Individual Information Part		

Version:	OSPF Version
Packet Type:	The OSPF packet types 1. Hello 2. Database description 3. Link state request 4. Link state update 5. Link state acknowledgement
Packet length:	The length of the protocol packet in bytes
Router ID:	IP address of the router that has sent this packet
Area ID:	Area ID of the router that has sent this packet
Authentication type:	Authorized or not
Authentication Data:	Password etc.

Fig. 4.12 *OSPF Packet Format*

Network Model for OSPF

Routers can be classified into three types as follows. One router may play two or more roles. Also, routing information exchanged between these routers is called LSP (link state packet).

- **Domain border router—** This router exchanges route information with routers in other domains. Information thus obtained is included in an OSPF message and transferred to other routers in the same domain (domain to which domain border router belongs). This allows all routers in the same domain to know which domain border router can provide route information to a specific domain.
- **Internal router—** Internal router is a router having its links directly connected to a network within a specific area. That is, internal router does not have any direct links to a network in another area.
- **Area border router—** This router belongs to two or more areas and notifies the backbone of the outline of its own configuration information so that this outline information can be transferred to other area boundary routers.

The backbone consists of those networks not contained in any area, their attached routers, and those routers that belong to multiple areas.

To recapitulate what has been described above, OSPF is a hierarchical routing composed of intra-area routing, inter-area routing, inter-domain routing, and so on. This means that if a message needs to be sent from one area to another, this message will sequentially pass as shown in Figure 4.13 and also below:

Source host → Internal router → Area border router in the same area → Domain border router in the same domain → Destination domain border router → Destination area border router → — → → Destination internal router → Destination host

NOTES

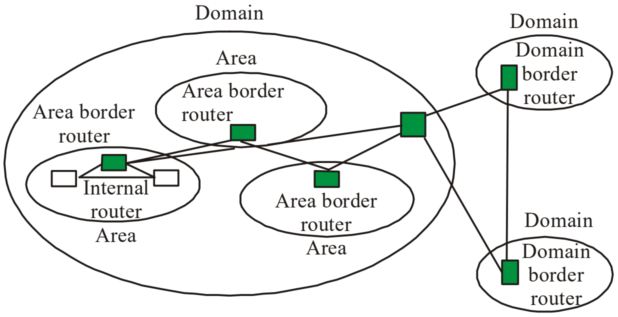


Fig. 4.13 *Network Model for OSPF*

Check Your Progress

1. What are the two types of network in frastructure in the country?
2. Define WAN.
3. What do you understand by RIP?
4. What is internal router?

4.3 TCP/IP OR COMMUNICATION PROTOCOL OVER WAN

The TCP/IP reference model is a network model used in Internet architecture. It has its beginnings back in the 1960s. The various design goals of TCP/IP are:

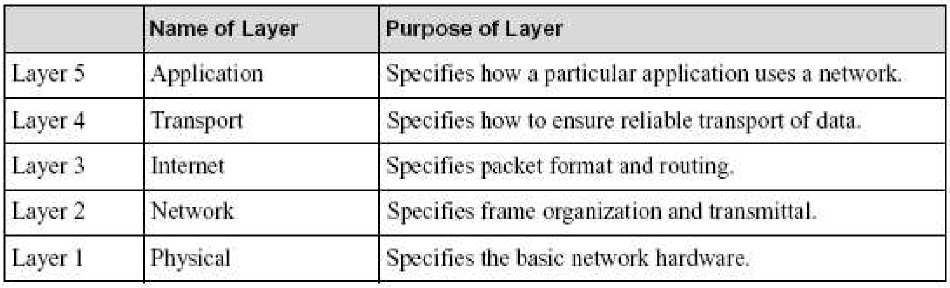
- Ability to connect multiple networks together effortlessly.
- Creation of a standardized concept of the communication mechanisms provided by each type of network.
- Ability for the connections to remain intact as long as the source and destination machines are functioning.
- A flexible architecture.

The TCP/IP protocol suite is so called because it specifies two of its most important protocols: Transmission Control Protocol (TCP) and Internet Protocol (IP).

TCP/IP Protocol Layers

Like most networking software, TCP/IP is modeled in conceptual layers. By dividing the communication software into layers, the protocol stack permits division of labor, ease of execution and code testing, and the ability to develop alternative layer implementations. Layers communicate with those above and below via concise interfaces. The TCP/IP protocol suite has five layers (Figures 4.14 and 4.15). The layers are as follows: physical, data link, network, transport and application.

NOTES



	Name of Layer	Purpose of Layer
Layer 5	Application	Specifies how a particular application uses a network.
Layer 4	Transport	Specifies how to ensure reliable transport of data.
Layer 3	Internet	Specifies packet format and routing.
Layer 2	Network	Specifies frame organization and transmittal.
Layer 1	Physical	Specifies the basic network hardware.

Fig. 4.14 *TCP/IP Five-Layer Reference Model*

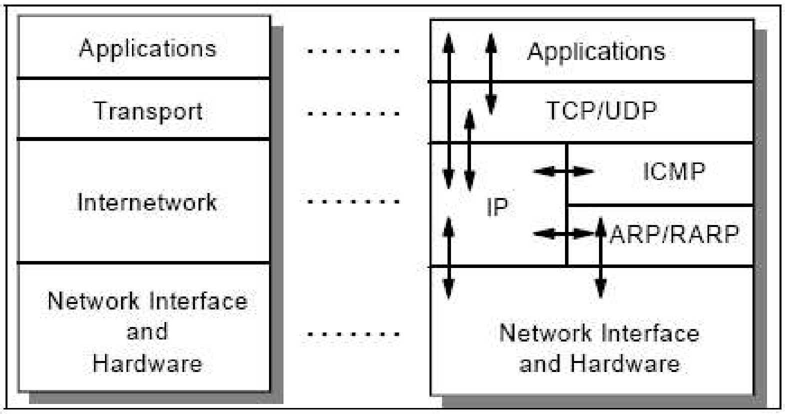


Fig. 4.15 *Another Depiction of the TCP/IP Five-Layer Reference Model*

These layers include:

- **Application layer:** The application layer is provided by the program that uses TCP/IP for communication. An application layer is a user process cooperating with another process, usually on a different host. Some common examples of standard networking services include Telnet and File Transfer Protocol (FTP). The interface between the application and transport layers is defined by port numbers and sockets.
- **Transport layer:** The transport layer manages end-to-end data transfer by delivering data from an application to its distant peer. Multiple applications can be supported simultaneously. In other words, this layer manages data transfer between networked applications.The most-used transport layer protocol is the Transmission Control Protocol (TCP), which provides connection-oriented reliable data delivery, duplicate data suppression, congestion control, and flow control. Another transport layer protocol is the User Datagram Protocol which provides connectionless, unreliable, best-effort service.
- **Internetwork layer:** The internetwork layer is also called the *internet layer* or the *network layer.* It provides the "virtual network" image of an internet. Internet Protocol (IP) is the most important protocol in this layer. It is a connectionless protocol. IP does *not* provide reliability, flow control or error handling. These functions must be provided at a higher level. IP provides a routing function that attempts to deliver transmitted messages to their destination. The basic unit of information

NOTES

transmitted across an IP network is called an *IP datagram*. Other internetwork-layer protocols are IP, ICMP, IGMP, ARP, and RARP.

- **Network interface layer:** The network interface layer is also called the *link layer* or the *data-link layer.* It is the interface to the actual network hardware. This interface may or may not provide reliable delivery, and may be packet or stream oriented. In fact, TCP/IP does not specify any protocol here, but can use almost any network interface available, which illustrates the flexibility of the IP layer. Examples are IEEE 802.2, X.25 (which is reliable in itself), ATM, FDDI, and even SNA 2.

TCP/IP Protocol Flow

Figure 4.16 provides the structure of TCP/IP protocol flow.

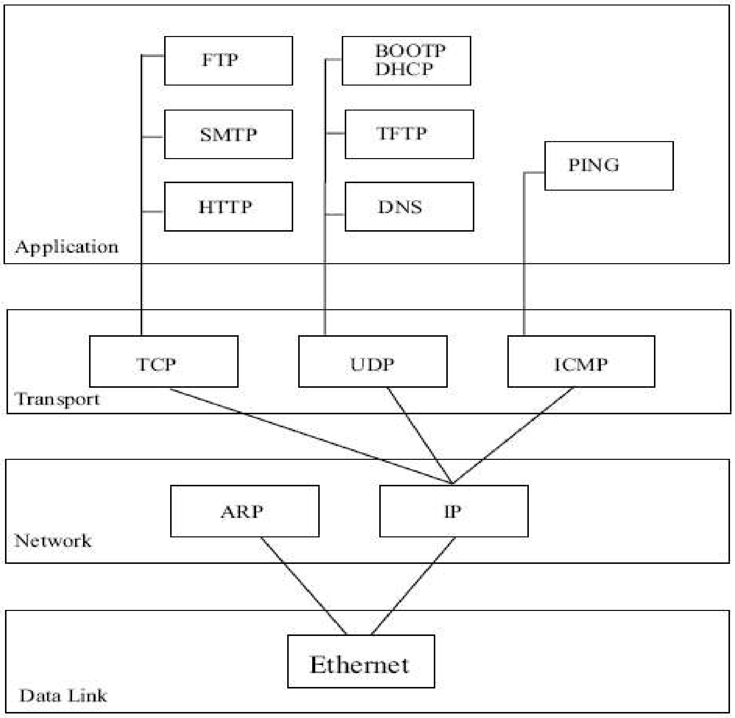


Fig. 4.16 *TCP/IP Protocol Flow*

Internet Protocol

The Internet Protocol (IP) is one of the most dominant protocols of the TCP/IP protocol suite and its main protocol is located at the network layer. The fundamental job of network layer is concerned with the delivery of data, from the source to the destination, between devices that may be on different networks. They are interconnected in an arbitrary manner: an *internetwork*. IP is the mechanism used for sending and communicating data from one device to another on TCP/IP networks.

The primary job of IP protocol is to deliver datagrams across an internetwork of connected networks.

Characteristics

Internet Protocol has proved to be a boon in incalculable ways. Of course, it has served the industry in manifold ways to accomplish the task because of it's unsurpassed characteristics. Let's take a look at the distinguishing attributes of the Internet Protocol which are as follows.

- **Universally-addressed:** In order to send data from point A to point B, IP first needs devices to set up a connection on how to send and receive data. It is also of paramount importance to confirm that devices are able to identify which device is 'point B'. Essentially, IP states precisely the addressing mechanism for the network and uses these addresses for data delivery purposes.
- **Underlying-protocol independent:** IP is designed to permit the transmission of data across any type of underlying network that is conducive to work with a TCP/IP stack irrespective of which of them instigates the proceedings. It includes provisions to allow it to adjust to the requirements of various lower-level protocols such as Ethernet or IEEE 802.11. IP can also run on the special data link protocols SLIP and PPP that were specially designed for it. An important example is IP's capability to fragment large blocks of data into smaller ones to match the size limits of physical networks, and then have the recipient rebuild the pieces again as needed.
- **Delivered connectionlessly:** IP is a *connectionless protocol*. This means that when A wants to send data to B, it doesn't first set up a connection to B and then send the data—it just makes the datagram and sends it.
- **Delivered unreliably:** IP is said to be an 'unreliable protocol'. It means that when datagrams are sent from device A to device B, device A just sends each one and then moves on to the next. IP doesn't keep track of the ones it has already sent . It does not provide reliability or service quality capabilities such as error protection for the data it sends, flow control or retransmission of lost datagrams.

  It is because of this reason that IP is called a *best-effort* protocol. It does what it can to get data to where it needs to go, but 'makes no guarantees' that the data will actually be delivered there.
- **Delivered without acknowledgments**: Because of its fallible nature, IP does not acknowledge for the delivery of data to the source. When device B receives a datagram from device A, it does not send back an acknowledgement to tell A that the datagram was received. There is always a question mark in the sender's mind regarding the delivery of data.

IP Functions

The primary job of IP functions is to add and manage IP addresses. With this as foundation, let us take a close look at four of its major functions.

Addressing: To deliver datagrams, IP includes a mechanism for host addressing. Since IP operates over internetworks, its system is designed to allow unique addressing of devices across large networks. It also contains a structure to facilitate the routing of datagrams to distant networks if that is required.

NOTES

Data encapsulation and formatting/packaging: As the TCP/IP network layer protocol, IP accepts data from the transport layer protocols UDP and TCP. It then encapsulates this data into an IP datagram using a special format prior to transmission.

Fragmentation and reassembly: IP datagrams are passed down to the data link layer for transmission on the local network. However, the maximum frame size of each physical/data-link network using IP may be different. For this reason, IP fragments IP datagrams into pieces so they can each be carried on the local network. The receiving device reassembles and restructures the whole IP datagram again.

Routing/indirect delivery: When an IP datagram is sent to a destination on the same local network, it is called *direct delivery*. This type of delivery is easy to perform using the network's underlying LAN/WLAN/WAN protocol. If the final destination is on a distant network not directly attached to the source, it is called *indirect delivery*. This is achieved by routing the datagram through intermediate devices called *routers*. IP accomplishes this with the support of the other protocols, including ICMP and TCP/IP gateway/routing protocols such as RIP and BGP.

Version 4 of the *Internet Protocol* is, in fact, the first version that was widely deployed and is the one in current widespread use.

IP Datagram

IP datagram is the rudimentary unit of information carried in the form of a packet in the IP layer, containing a source and destination address. This information is communicated across the network using Internet Protocol. It is a variable-length packet which comprises two parts: header and data (Figure 4.17). The header is 20–60 bytes in length and contains information critical to routing and delivery.

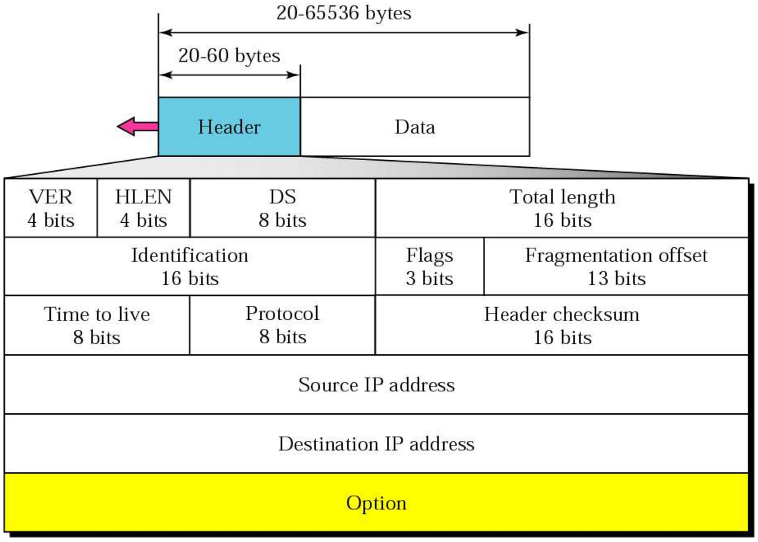


Fig. 4.17 *IP Datagram*

Version (VER): Identifies the version of IP used to generate the datagram. For IPv4, this is 4. This field ensures compatibility between devices that may be running different versions of IP. A device running an older version of IP will discard datagrams created by newer implementations, assuming that the older version may not be able to understand the newer datagram correctly.

Header length (HLEN): Specifies the length of the IP header, in 32-bit words. This includes the length of any options fields and padding. The normal value of this field when no options are used is 5 (5 32-bit words = 5 × 4 = 20 bytes).

Differentiated services (DS): Carries information to provide quality of service features, such as prioritized delivery, etc., for IP datagrams (Figure 4.18).

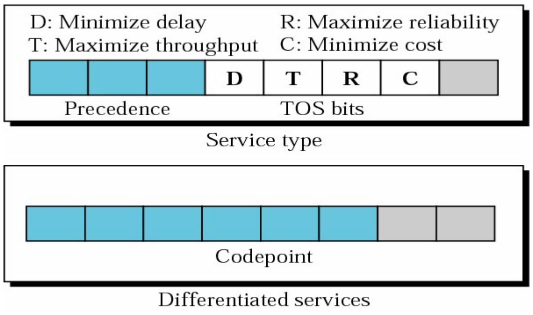
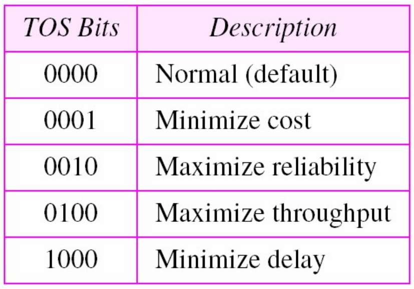
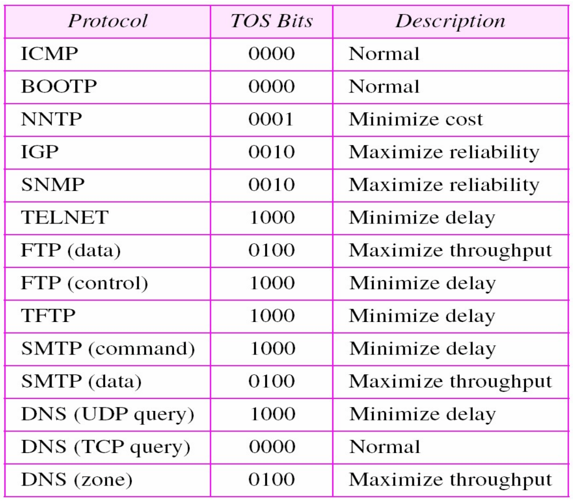


Fig. 4.18 *Differentiated Services*



TOS Bits	*Description*
0000	Normal (default)
0001	Minimize cost
0010	Maximize reliability
0100	Maximize throughput
1000	Minimize delay

Fig. 4.19 *Types of Service*



Protocol	*TOS Bits*	*Description*
ICMP	0000	Normal
BOOTP	0000	Normal
NNTP	0001	Minimize cost
IGP	0010	Maximize reliability
SNMP	0010	Maximize reliability
TELNET	1000	Minimize delay
FTP (data)	0100	Maximize throughput
FTP (control)	1000	Minimize delay
TFTP	1000	Minimize delay
SMTP (command)	1000	Minimize delay
SMTP (data)	0100	Maximize throughput
DNS (UDP query)	1000	Minimize delay
DNS (TCP query)	0000	Normal
DNS (zone)	0100	Maximize throughput

Fig. 4.20 *Default Types of Service*

Category	*Codepoint*	*Assigning Authority*
1	XXXXX0	Internet
2	XXXX11	Local
3	XXXX01	Temporary or experimental

Fig. 4.21 *Codepoint Values*

Total length (TL): Specifies the total length of an IP datagram, in bytes. Since this field is 16 bits wide, the maximum length of an IP datagram is 65,535 bytes.

Identification: Contains a 16-bit value common to each of the fragments belonging to a particular message; for datagrams originally sent unfragmented it is still filled in, so it can be used if the datagram must be fragmented by a router during delivery. Recipients can use this field to reassemble messages without mixing fragments from different messages. This is needed because fragments may reach the destination from multiple messages mixed together and IP datagrams can be received out of order from any device.

Flags: Three control flags, two of which are used to manage fragmentation and one that is reserved (Figure 4.22).

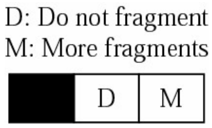


Fig. 4.22 *Flags*

Fragment offset: In a fragmented message, this field specifies the offset, or position, in the overall message where the data in this fragment goes. It is specified in units of 8 bytes (64 bits). The first fragment has an offset of 0.

Time to live (TTL): This field specifies how long the datagram is allowed to 'live' on the network, in terms of router hops. Each router reduces the value of the TTL field by one prior to transmitting it. If the TTL field drops to zero, it is assumed that the datagram has taken too long a route and is discarded.

Protocols: Figure 4.23 provides the structure for protocols.

Value	*Protocol*
1	ICMP
2	IGMP
6	TCP
17	UDP
89	OSPF

Fig. 4.23 *Protocol Structure*

Header checksum: A checksum is computed over the header to provide basic protection against corruption in the transmission. It is calculated by dividing the header bytes into words (a word is two bytes) and then adding them together. Each device receiving the datagram does the same checksum calculation and on a mismatch, discards the datagram as damaged.

Source address: The source address of the datagram is a 32-bit IP address . Intermediate devices, such as routers, handling the datagram do not put their address into this field. It is always the device that originally sent the datagram whose address comes here.

Destination address: The destination address is 32-bit IP address of the intended recipient of the datagram. Devices such as routers may be the intermediate targets of the datagram. This field always has the address of the ultimate destination.

Options: One or more options may be incorporated after the standard headers in certain IP datagrams. The header of the IP datagram constitutes two parts: a fixed part and a variable part. The variable part consists of the options that can be a maximum of 40 bytes.

Option format: Figure 4.24 provides the option format structure.

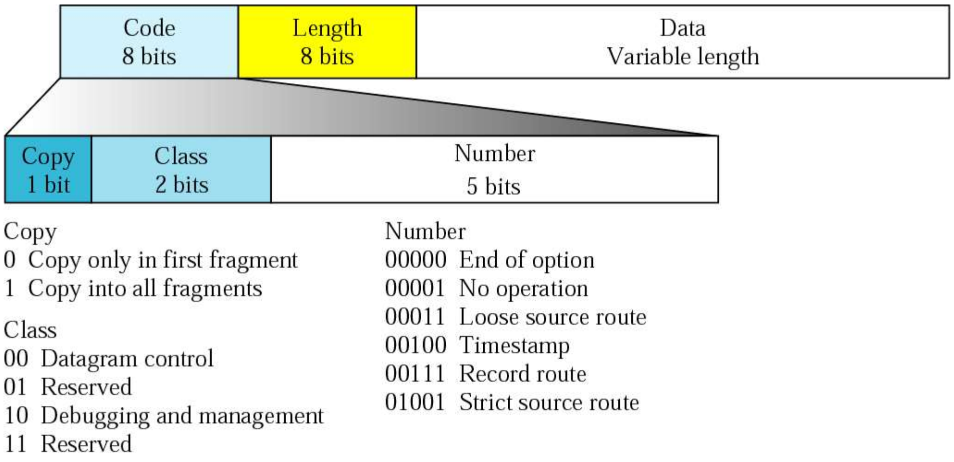


Fig. 4.24 *Option Format*

Categories of options: Figure 4.25 provides the categories of options structure.

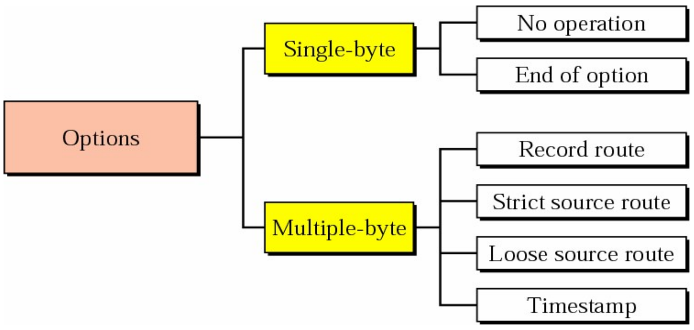


Fig. 4.25 *Categories of Options*

4.4 IP ADDRESSING

The Internet address or IP address is a unique identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet.

IP addresses are used by the IP protocol to uniquely identify a device on the Internet. IP datagrams are transmitted over a physical network attached to the host. Each IP datagram has a *source, IP address* and a *destination IP address*. To send a datagram to a certain IP destination, the *destination* IP address must be mapped to a physical address.

4.4.1 Characteristics of IP Addresses

- An IP address is a 32-bit address that uniquely and universally identifies the host or a router connected to the Internet.
- Each IP address is unique. No two devices on the Internet can have similar addresses.
- Each IP address comprises two elements: the prefix, which identifies the physical network, and the suffix, which identifies a computer on the network (Figure 4.26).

  IP address = <network number><host number>

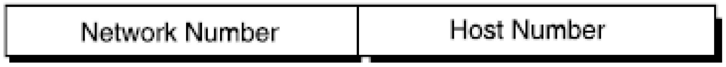


Fig. 4.26 *Two Level IP Address*

(i) Each IP address is 32-bit number represented in a dotted decimal form. For example, 128.11.3.31 is an IP address with 128.11 being the network number and 3.31 being the host number.

(ii) To make Internet address easier for people to comprehend and write, it is often expressed as four decimal numbers, each separated by a dot. This format is called 'dotted-decimal notation.' Dotted-decimal notation divides the 32-bit Internet address into four 8-bit fields and specifies the value of each field independently as a decimal number with the fields separated by dots (Figure 4.27).

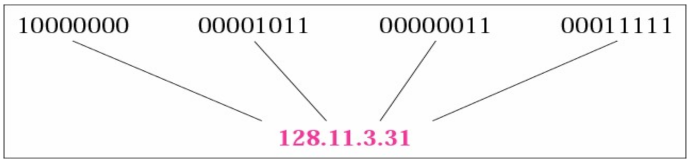


Fig. 4.27 *The Dotted-decimal Notation of the IP Address 128.11.3.31*

Classful Addressing

IP addresses, when started a few decades ago, were based on the concept of classes. Each class fixed a boundary between the network prefix and the host number at a different point within the 32-bit address.

There are five classes of IP addresses as shown in Figure 4.28.

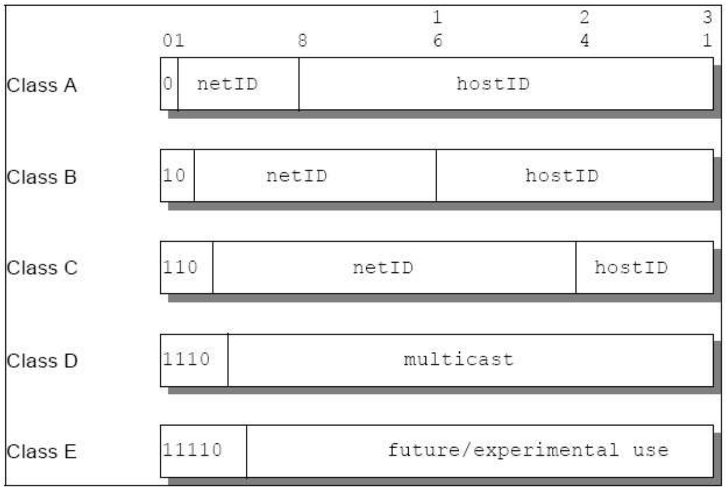


Fig. 4.28 *Five Classes of IP Addresses*

Class A addresses: These addresses use 7 bits for the <network> and 24 bits for the <host> portion of the IP address. This permits 2^7–2 (126) networks each with 2^{24}–2 (16777214) hosts—a total of more than 2 billion addresses.

Class B addresses: These addresses use 14 bits for the <network> and 16 bits for the <host> portion of the IP address. This allows for 2^{14}–2 (16382) networks each with 2^{16}–2 (65534) hosts—a total of more than 1 billion addresses.

Class C addresses: These addresses use 21 bits for the <network> and 8 bits for the <host> portion of the IP address. That allows for 2^{21}–2 (2097150) networks each with 2^8–2 (254) hosts—a total of more than half a billion addresses.

Class D addresses: These addresses are reserved for multicasting (a sort of broadcasting, but in a limited area, and only to hosts using the same Class D address).

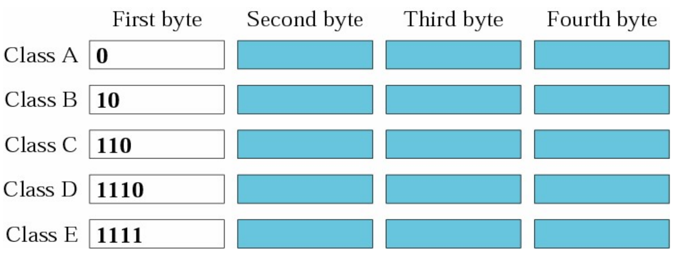
Class E addresses: These addresses are reserved for future use.

Addresses per class: The Class A address is more suitable for networks with an excessively colossal number of hosts (Figure 4.29). Class C addresses are apt for networks with a small number of hosts. This clearly indicates that medium-sized networks (those with more than 254 hosts or where there is an expectation of more than 254 hosts) must use Class B addresses.

Class	*Number of Addresses*	*Percentage*
A	2^{31} = 2,147,483,648	50%
B	2^{30} = 1,073,741,824	25%
C	2^{29} = 536,870,912	12.5%
D	2^{28} = 268,435,456	6.25%
E	2^{28} = 268,435,456	6.25%

Fig. 4.29 *Addresses per Class*

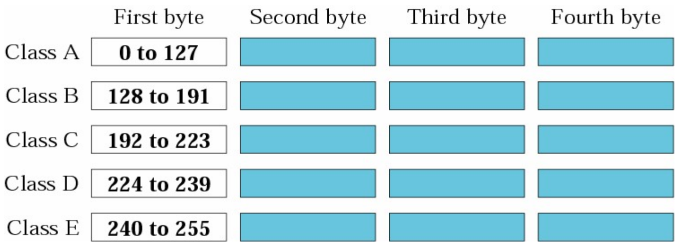
Finding the class in binary notation is shown in Figure 4.30.



	First byte	Second byte	Third byte	Fourth byte
Class A	**0**			
Class B	**10**			
Class C	**110**			
Class D	**1110**			
Class E	**1111**			

Fig. 4.30 Finding the Class in Binary Notation

Finding the class in decimal notation is shown in Figure 4.31.



	First byte	Second byte	Third byte	Fourth byte
Class A	**0 to 127**			
Class B	**128 to 191**			
Class C	**192 to 223**			
Class D	**224 to 239**			
Class E	**240 to 255**			

Fig. 4.31 *Finding the Class in Decimal Notation*

Find the class of each address:

- 00000001 00001011 00001011 11101111
- 11000001 10000011 00011011 11111111
- 227.12.14.87
- 193.14.56.22

Solution:

- The first bit is 0. This is a Class A address.
- The first 2 bits are 1; the third bit is 0. This is a Class C address
- The first byte is 227 (between 224 and 239); the Class is D.
- The first byte is 193 (between 192 and 223); the Class is C.

Figures 4.32, 4.33 and 4.34 depict blocks in classes A, B and C.

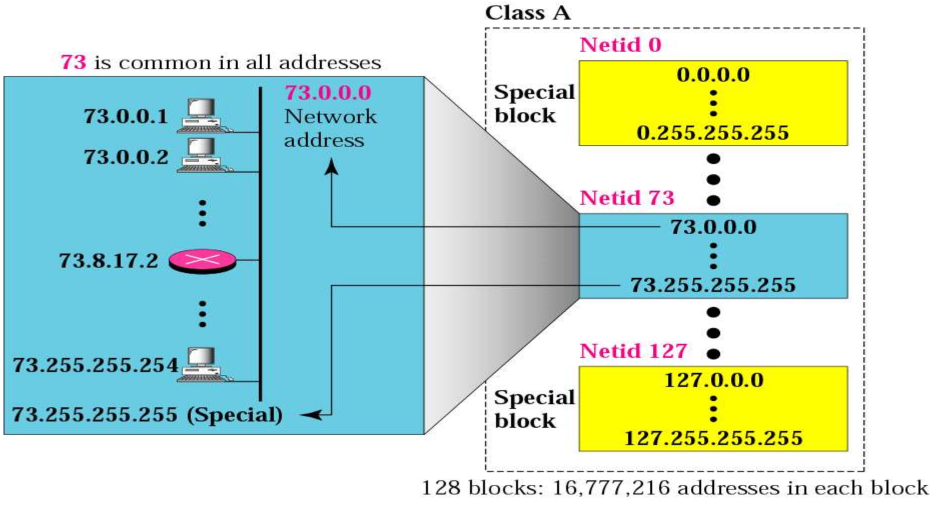


Fig. 4.32 *Blocks in Class A*

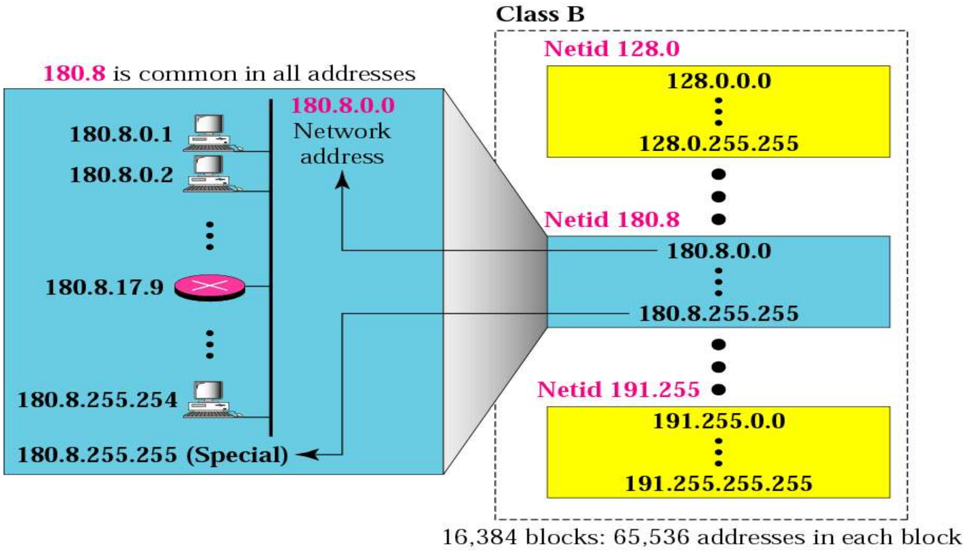


Fig. 4.33 Blocks in Class B

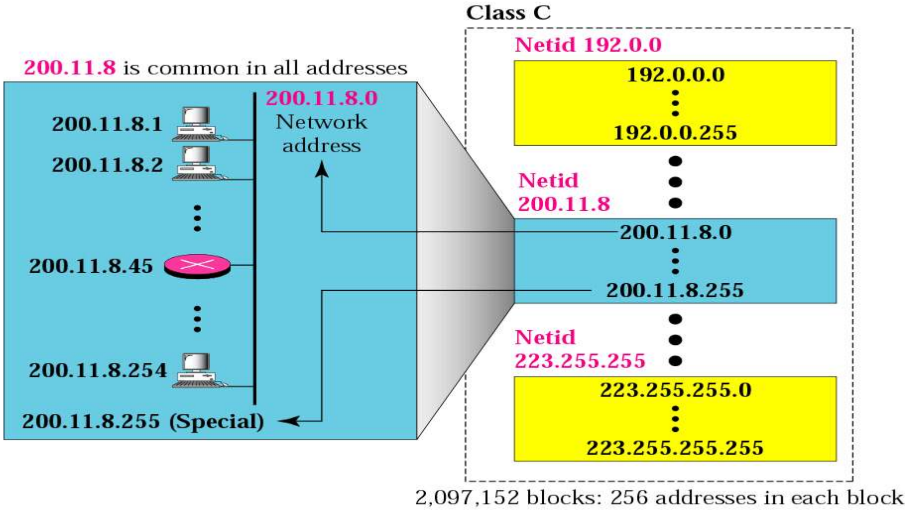


Fig. 4.34 Blocks in Class C

In classful addressing, the network address (the first address in the block) is assigned to the organization. The range of addresses can automatically be concluded from the network address.

Given the network address 17.0.0.0, find the class, the block, and the range of the addresses.

Solution:

The class is A because the first byte is between 0 and 127. The block has a netid of 17. The addresses range from 17.0.0.0 to 17.255.255.255.

Limitations of classful addressing: There are certain limitations of classful addressing which are mentioned below.

1. **Lack of internal address flexibility:** Massive organizations are assigned large, 'monolithic' blocks of addresses that do not match well

with the structure of their underlying internal networks.

2. **Inefficient use of address space:** Limited IP address space is wasted due to the existence of only three block sizes (classes A, B and C, repectively).
3. **Proliferation of router table entries:** With the growth of Internet, routers require more and more entries to handle the routing of IP datagrams, which causes performance problems for routers ultimately affecting the execution level of the task. Attempting to reduce inefficient address space allocation leads to even more router table entries.

These difficulties were addressed partially through subnet addressing, which provides more flexibility for the administrators of individual networks on an internet. Subnetting, however, doesn't really tackle the problems in general terms. Some of these issues remain due to the use of classes even with subnets.

Classless Addressing

Classful addressing resulted in efficient use of address space. The Class B address block contains a very large number of addresses (65,534) but a Class C block has only a relatively small number (254). There are many thousands of 'medium-sized' organizations who need more than 254 IP addresses, but a small percentage of these need 65,534 or anything even close to it. When setting up their networks, these companies and groups would tend to request Class B address blocks and not Class C blocks because they need more than 254, without considering how many of the 65,000-odd addresses they really would use and how may will go waste. The only solution to this would be to convince – or force – companies to use many smaller Class C blocks instead of 'wasting' the bulk of a Class B assignment. Many organizations resisted this due to the difficulty involved, and this caused the other main problem that subnetting didn't solve: the growth of Internet routing tables. Replacing one Class B network with 10 Class Cs would mean ten times as many entries for routers to keep the track of.

It was quite obvious that as long as there were only three sizes of networks, the allocation efficiency problem could never be properly rectified as it was permanent. The only solution left was to get rid of the classes completely, in favor of a *classless* addressing scheme. This system would solve both the main problems of 'classful' addressing namely inefficient use of address space, and the exponential growth of routing tables.

This system of classless addressing was developed in the early 1990s and formalized in 1993. The technology was called Classless Inter-Domain Routing (CIDR).

In classless addressing, when a host or router needs to be connected to the Internet, it is granted a block or range of addresses. Variable-length blocks are assigned that belong to no class. In this architecture, the entire address space (232 addresses) is divided into blocks of different sizes (Figure 4.35).

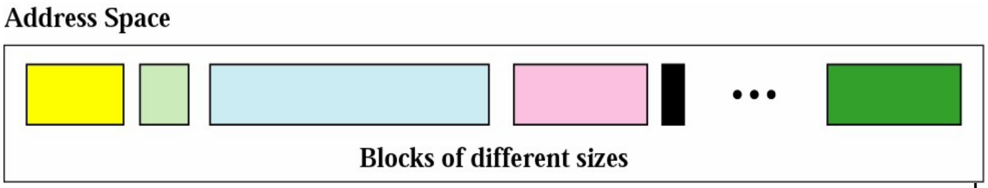


Fig. 4.35 *Address Space*

Conditions:

1. The addresses in the block must be contiguous.
2. The number of addresses in a block must be a power of 2.
3. The first address must be evenly divisible by the number of addresses.

A classless address is specified in CIDR or slash notation (Figure 4.36)

x.y.z.t/n

Fig. 4.36 *CIDR Notation*

where **x.y.z :** IP address; n, integer that tells us how many bits are used for the network ID

For example, consider the network specification 184.13.152.0/22. The '22' means this network has 22 bits for the network ID and 10 bits for the host ID. This is equivalent to specifying a network with an address of 184.13.152.0 and a subnet mask of 255.255.252.0 (Figure 4.37)

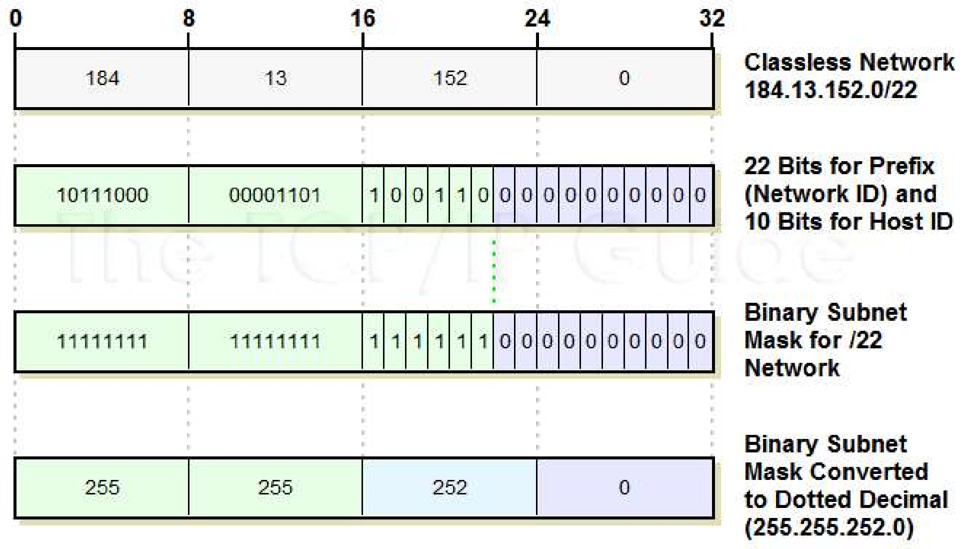
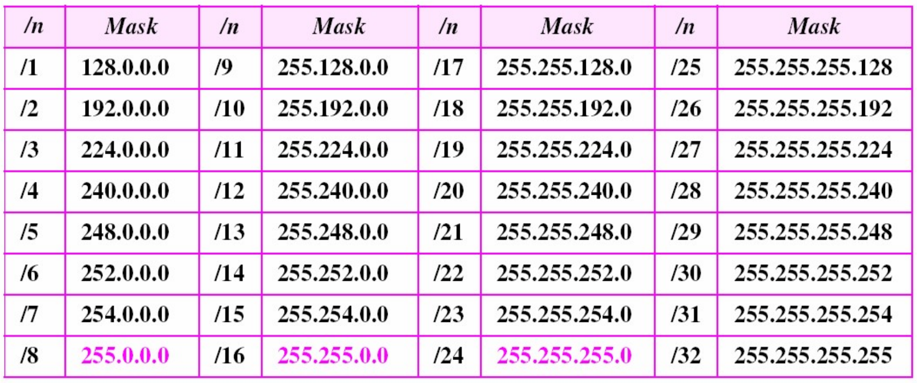

0
8
16
24
32
184
13
152
0
Classless Network 184.13.152.0/22
10111000
00001101
1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
22 Bits for Prefix (Network ID) and 10 Bits for Host ID
11111111
11111111
1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
Binary Subnet Mask for /22 Network
255
255
252
0
Binary Subnet Mask Converted to Dotted Decimal (255.255.252.0)
The TCP/IP Guide

Fig. 4.37 *Subnet Mask*

NOTES



/n	Mask	/n	Mask	/n	Mask	/n	Mask
/1	128.0.0.0	/9	255.128.0.0	/17	255.255.128.0	/25	255.255.255.128
/2	192.0.0.0	/10	255.192.0.0	/18	255.255.192.0	/26	255.255.255.192
/3	224.0.0.0	/11	255.224.0.0	/19	255.255.224.0	/27	255.255.255.224
/4	240.0.0.0	/12	255.240.0.0	/20	255.255.240.0	/28	255.255.255.240
/5	248.0.0.0	/13	255.248.0.0	/21	255.255.248.0	/29	255.255.255.248
/6	252.0.0.0	/14	255.252.0.0	/22	255.255.252.0	/30	255.255.255.252
/7	254.0.0.0	/15	255.254.0.0	/23	255.255.254.0	/31	255.255.255.254
/8	255.0.0.0	/16	255.255.0.0	/24	255.255.255.0	/32	255.255.255.255

Fig. 4.38 *Prefix Lengths*

- **First address**: The first address of the block can be found by setting the 32–*n* right most bits in binary notation to zero
- **Last address**: The last address of the block can be found by setting the 32–*n* right most bits in binary notation to one
- **Number of addresses**: 2^{32-n}

Example: What is the first, last and number of addresses in the block if one of the addresses is 205.16.37.39/28?

Solution:

First Address

The prefix length is 28, which means that we must keep the first 28 bits as it is and change the remaining bits (4) to 0s. The following shows the process:

Address in binary: 11001101 00010000 00100101 00100111

Keep the left 28 bits: 10100111 11000111 10101010 00100000

Result in CIDR notation: 205.16.37.32/28

Last Address

The prefix length is 28, which means that we must keep the first 28 bits as it is and change the remaining bits (4) to 1s. The following shows the process:

Address in binary: 11001101 00010000 00100101 00100111

Keep the left 28 bits: 10100111 11000111 10101010 00101111

Result in CIDR notation: 205.16.37.47/28

Number of Addresses : 2^{32-28} = 16

Under CIDR, all internet blocks can be of random size. Instead of having all networks use 8 (Class A), 16 (Class B) or 24 (Class C) bits for the network ID, we can have large networks with, say, 13 bits for the network ID (leaving 19 bits for the host ID), or very small ones that use 28 bits for the network ID (only 4 bits for the host ID).

- **Benefits of classless addressing**

  CIDR provides many advantages over the ‘classful’ addressing scheme, whether or not subnetting is used: With this as foundation, the benefits of classless addressing are as follows.

  o **Efficient address space allocation:** Under CIDR, addresses are allocated in sizes of any binary multiple instead of fixed-size blocks of low granularity. This ensures in minimum wastage of address space. So,

NOTES

a company that needs 5,000 addresses can be assigned a block of 8,190 instead of 65,534 as shown in Figure 4.39. In other words, the equivalent of a single Class B network can be shared amongst eight companies that each need 8,190 or fewer IP addresses.

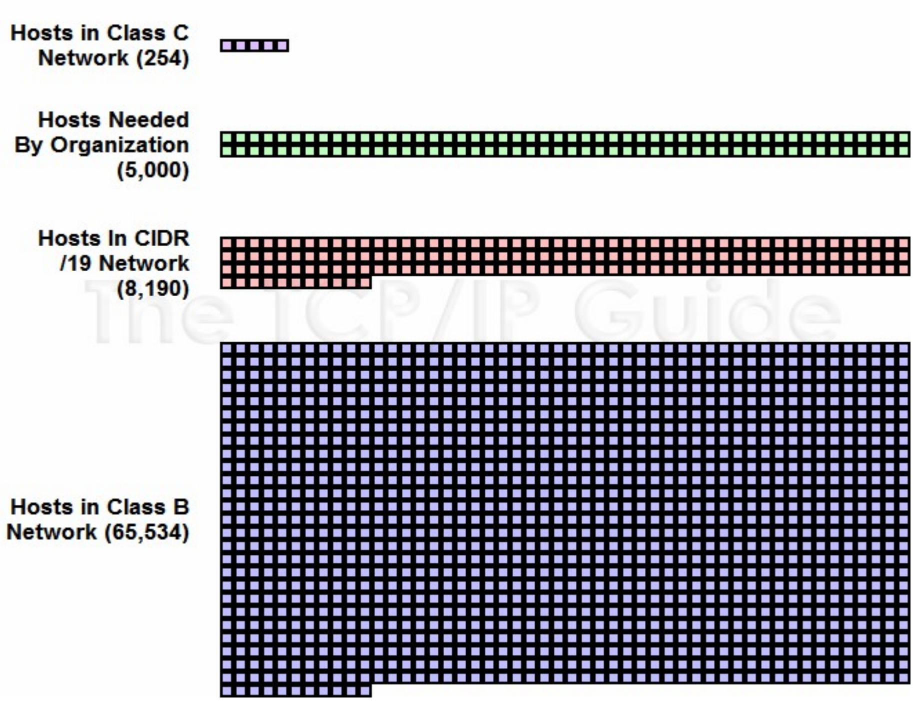


Fig. 4.39 *Efficient Address Space Allocation*

- o **Elimination of class imbalances:** Under CIDR, the classes (A, B and C) do not exist anymore , so there is no problem of imbalances in the use of addresses with some portions of the address space being widely used while others are neglected.
- o **Efficient routing entries:** CIDR's multiple-level hierarchical structure allows a small number of routing entries to represent a large number of networks. Network descriptions can be 'aggregated' and represented by a single entry. Since CIDR is hierarchical, the detail of lower-level, smaller networks can be hidden from routers that move traffic between large groups of networks.
- o **No separate subnetting method:** CIDR implements the concepts of subnetting within the internet itself. There is no separate subnetting method used. An organization can use the same method used on the Internet to subdivide its internal network into subnets of arbitrary complexity without the need for a separate subnetting mechanism.

4.4.2 Subnetting

Due to the exponential growth of the Internet, the principle of assigned IP addresses became too rigid to allow easy changes to local network configurations. Those changes might occur when:

- a new type of physical network is installed at a location.
- there is a growth of the number of hosts that require splitting the local network into two or more separate networks.

- growing distances require splitting a network into smaller networks, with gateways between them.

Adding an extra hierarchical level in the way IP addresses are currently interpreted forms the basic idea of subnetting. The concept of a network remains unchanged, but instead of having just 'hosts' within a network, a new two-level hierarchy is created: *subnets* and hosts. A three-level hierarchy is also created—there is a network, within the network is a subnet and there are hosts within the subnet. Each subnet is a subnetwork, and functions much the way a full network does in conventional classful addressing.

Thus, instead of an organization having to group all its hosts under that network in an unstructured manner, it can organize hosts into subnets that reflect the way internal networks are structured. These subnets fit within the network identifier assigned to the organization, just as all the 'unorganized' hosts used to.

Subnetting adds an additional level to the hierarchy of structures used in IP addressing. A three-level hierarchy is created for IP addresses (Figure 4.40). Now IP addresses must be broken into three elements instead of two. This is achieved by leaving the network ID alone. The host number part of the IP address is subdivided into a second network number and a host number. This second network is termed a subnetwork ID or subnet ID. The main network now consists of a number of subnets. These subnet ID bits are used to identify each subnet within the network. Hosts are assigned to the subnets in a manner that makes the most sense for that network.

The IP address is interpreted as:

```
<network number><subnet number><host number>
```

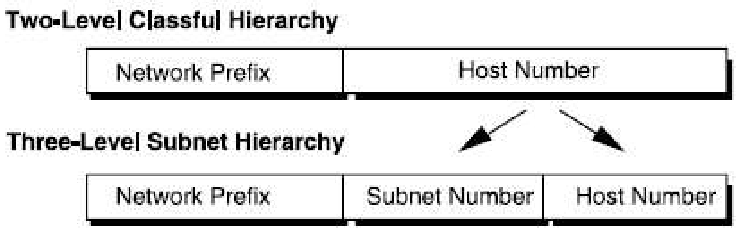


Fig. 4.40 *Three-Level Subnet Hierarchy*

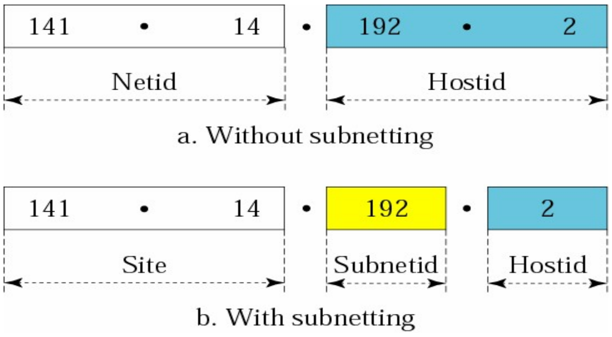


Fig. 4.41 *Addresses with and without Subnetting*

The combination of subnet number and host number is often termed the local address or the local portion of the IP address.

NOTES

In order to do subnetting, the host ID is split into subnet ID and host ID. In doing this, the size of the host ID part of the address gets reduced. In short, bits from the host ID are being 'stolen' to use for the subnet ID. Class A networks have 24 bits to split between the subnet ID and host ID: Class B networks have 16, and Class C networks only 8 (Figure 4.42).

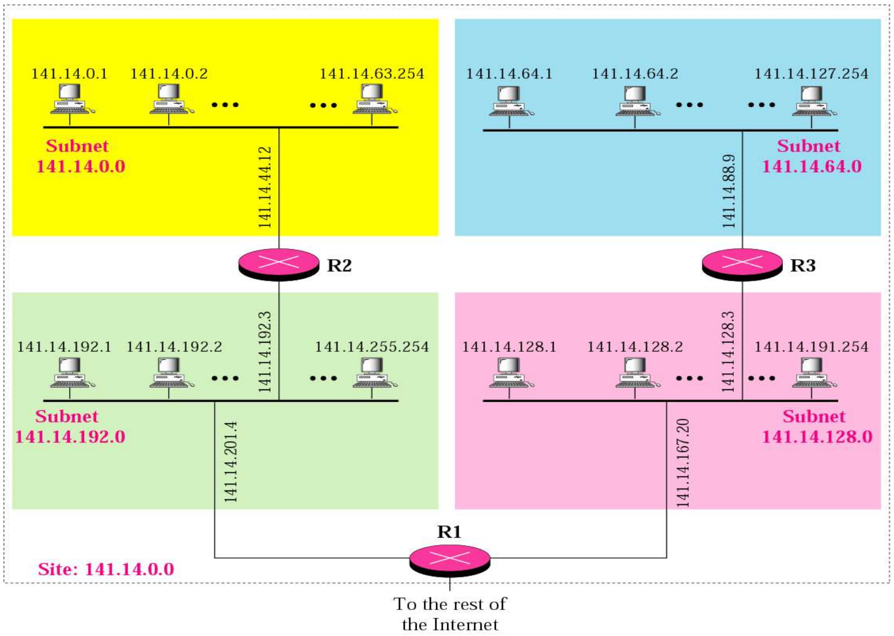


Fig. 4.42 *A Subnetted Network*

The division of the host ID that is a part of the IP address is used to define a subnet number and host number is chosen by the local administrator. Any number of bits in the host ID can be used to form the subnet.

The number of subnets is two to the power of the size of the subnet ID field. Similarly, the number of hosts per subnet is two to the power of the size of the host ID field (Figure 4.43).

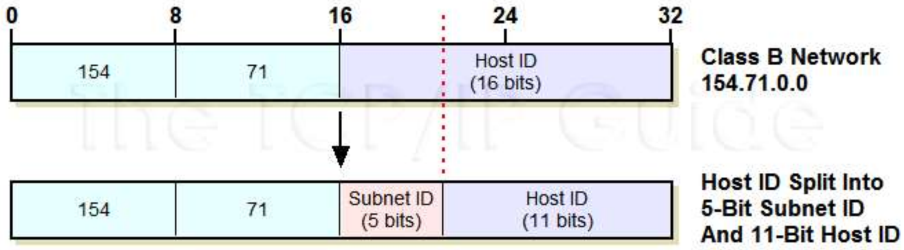


Fig. 4.43 *Division of Host-Id Bits into Subnet-Id and Host-Id*

In Class B network 154.71.0.0 16 bits are for the network ID (154.71) and 16 for the host ID. In regular 'classful' addressing there are no subnets and 65,534 total hosts. In order to subnet this network, the local administrator can decide to split these 16 bits. However, it is upon his discretion what suits best to the needs of the network. Any combination will work, as long as the total is 16: 1 bit for the subnet ID and 15 for the host ID, or 2 and 14, 3 and 13, and so on. The more the bits 'stolen' from the

NOTES

host ID for the subnet ID, more subnets are possible —but the fewer hosts will be available for each subnet.

Subnetting is assigned to an organization or at the most few organizations. Each organization is assigned one network number from the IPv4 address space. The organization is then free to assign a separate subnetwork number for each of its internal networks. This allows the organization to organize additional subnets without obtaining a new network number from the Internet.

- **Subnet mask**

In subnetting, it is necessary to communicate, which bits are for the subnet ID and which for the host ID, to devices that interpret IP addresses. A 32-bit binary number which provides this information to devices handling IP addresses is called a *subnet mask*.

The *subnet mask* is a 32-bit binary number that comes with an IP address. Like IP addresses, they are usually converted to dotted decimal notation for convenience. It is created in a way that it has a one bit for each corresponding bit of the IP address that is part of its network ID or subnet ID, and a zero for each bit of the IP address's host ID. The mask informs the TCP/IP devices about the bits in a IP address that belong to the network ID and subnet ID, and which are a part of the host ID.

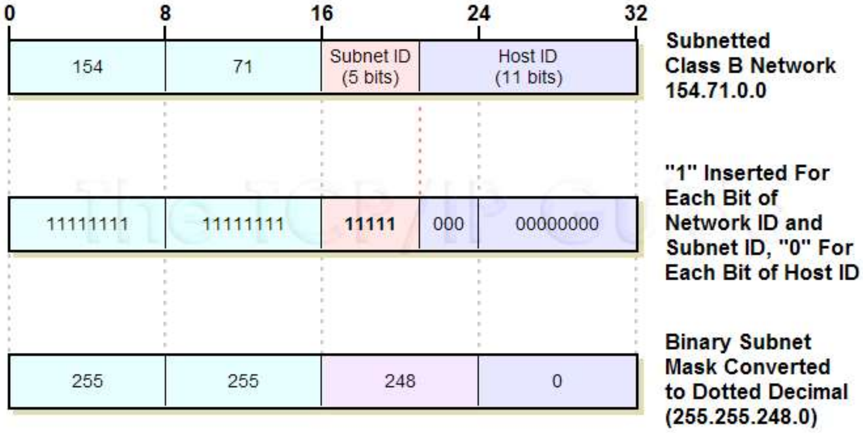

0
8
16
24
32
154
71
Subnet ID (5 bits)
Host ID (11 bits)
Subnetted Class B Network 154.71.0.0
11111111
11111111
11111
000
00000000
"1" Inserted For Each Bit of Network ID and Subnet ID, "0" For Each Bit of Host ID
255
255
248
0
Binary Subnet Mask Converted to Dotted Decimal (255.255.248.0)

Fig. 4.44 *Determining the Subnet Mask of a Subnetted Network*

Suppose there is a Class B network 154.71.0.0. It is decided to subnet this using 5 bits for the subnet ID and 11 bits for the host ID. In this case, the subnet mask will have 16 ones for the network portion (since this is Class B) followed by 5 ones for the subnet ID, and 11 zeroes for the host ID. That's '11111111 11111111 **11111**000 00000000' in binary, with the bits corresponding to the subnet ID highlighted. Converting to dotted decimal, the subnet mask would be 255.255.248.0.

- **Applying the subnet mask**

A mask is used to find out what subnet an IP address belongs to. Suppose there is a host on this network with an IP of 154.71.150.42. A router needs to find out which subnet this address is on. This is done by performing the masking operation as shown in Figures 4.45 and 4.46.

Determining the Subnet ID of an IP Address Through Subnet Masking				
Component Octet 1	Octet 1	Octet 2	Octet 3	Octet 4
IP Address 10011010(154)	10011010(154)	01000111 (71)	10010110 (150)	00101010 (42)
Subnet Mask 11111111(255)	11111111(255)	11111111 (255)	11111000 (248)	00000000 (0)
Result of AND Masking 10011010(154)	10011010(154)	01000111 (71)	10010000 (144)	00000000 (0)

Fig. 4.45 Applying the Subnet Mask

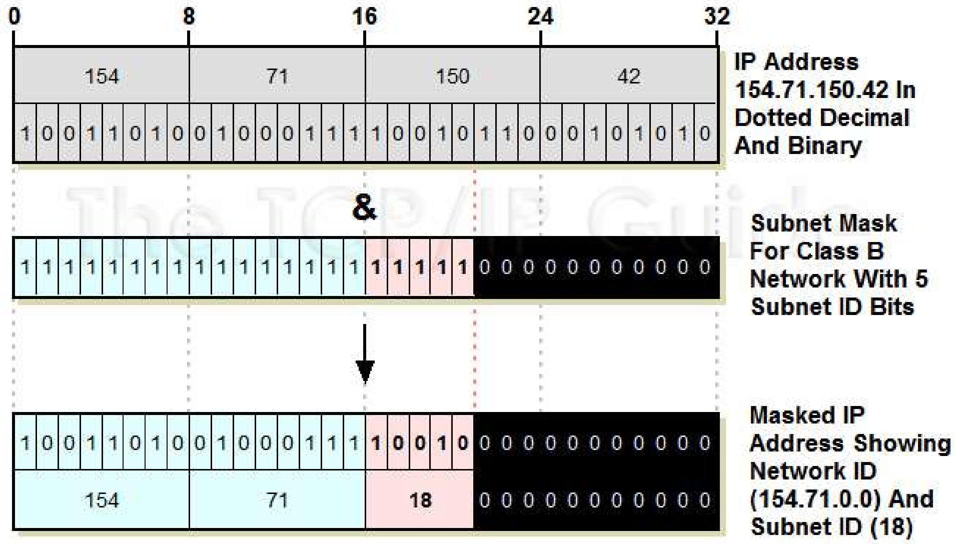


Fig. 4.46 Using a Subnet Mask

To use a subnet mask, a device performs a boolean *AND* operation between each bit of the subnet mask and each corresponding bit of an IP address. The resulting 32-bit number contains only the network ID and subnet ID of the address, with the host ID cleared to zero. This result, 154.71.144.0, is the IP address of the subnet to which 154.71.150.42 belongs.

Default subnet masks for Class A, Class B and Class C networks

IP Address Class	Total # Of Bits For Network ID / Host ID	Default Subnet Mask			
		First Octet	Second Octet	Third Octet	Fourth Octet
Class A	**8 / 24**	11111111 (255)	00000000 (0)	00000000 (0)	00000000 (0)
Class B	**16 / 16**	11111111 (255)	11111111 (255)	00000000 (0)	00000000 (0)
Class C	**24 / 8**	11111111 (255)	11111111 (255)	11111111 (255)	00000000 (0)

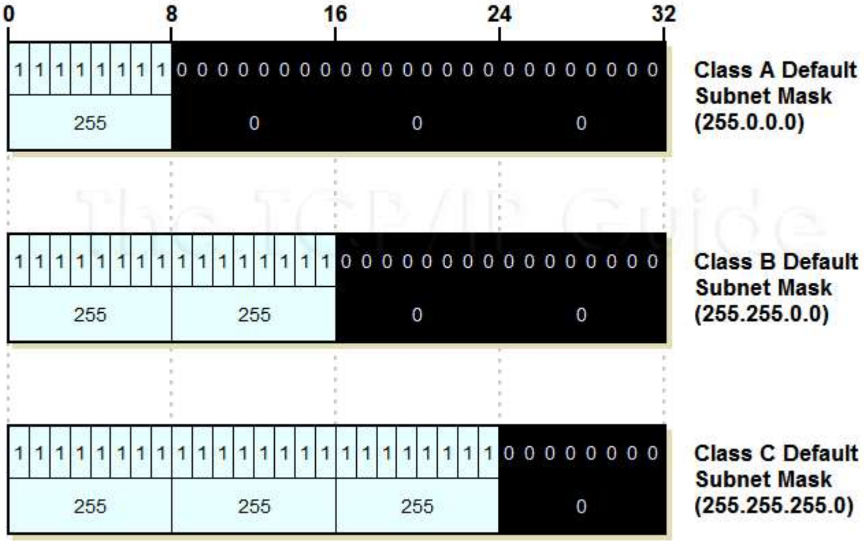

0
8
16
24
32
255
0
0
0
Class A Default Subnet Mask (255.0.0.0)
255
255
0
0
Class B Default Subnet Mask (255.255.0.0)
255
255
255
0
Class C Default Subnet Mask (255.255.255.0)

Fig. 4.47 *Default Subnet Masks for Class A, Class B and Class C Networks*

Since the local network administrator is able to customize his choice of dividing points between subnet ID and host ID to satisfy the requirements of the network, this is sometimes called *customized subnetting*. The subnet mask that is used to create a customized subnet is called a *custom subnet mask*. The custom subnet mask is used by network hardware to determine how the local network administrator has decided to divide the subnet ID from the host ID in the network.

- **Deciding how many subnet bits to use**

The crucial determining factor in customized subnetting is how to divide network ID into subnet ID and host ID. The local administrator has to decide on the number of bits to take from the host ID, portion of the IP address and put into the subnet ID. As mentioned earlier, the number of subnets possible on the network is two to the power of the number of bits used to express the subnet ID, and the number of hosts possible per subnet is two to the power of the number of bits left in the host ID minus two.

Thus, the decision of how many bits to use for subnet ID and host ID represents a fundamental trade-off in subnet addressing:

- o Each bit taken from the host ID for the subnet ID doubles the number of subnets that are possible in the network.
- o Each bit taken from the host ID for the subnet ID (approximately) reduces by half the number of hosts that are possible within each subnet on the network.

- **Subnetting bit allocation options**

  The above concept can be illustrated by the following example.

  Imagine that we begin with a Class B network with the network address 154.71.0.0. Since this is Class B, 16 bits are for the network ID (154.71) and 16 are for the host ID. In the default case, there are no subnets and 65,534 hosts total. To subnet the above network, there are a number of choices.

  - o We can decide to use 1 bit for the subnet ID and 15 bits for the host ID. If we do this, then the total number of subnets is 2^1 or 2: the first subnet is 0 and the second is 1. The number of hosts available for each subnet is $2^{15}–2$ or 32,766.
  - o We can use 2 bits for the subnet ID and 14 for the host ID. In this case, we double the number of subnets: we now have 2^2 or 4 subnets: 00, 01, 10 and 11 (subnets 0, 1, 2 and 3). But the number of hosts is now only $2^{14}–2$ or 16,382.
  - o We can use any other combination of bits that add up to 16, as long as they allow us at least 2 hosts per subnet: 4 and 12, 5 and 11, and so on.

  To divide the 'classful' host ID into subnet ID and host ID, bits is the key design decision in subnetting. The decision is primarily made keeping in mind the requirements of the network as foreseen by the network administrator. The network administrator must choose, based on the requirements for the number of subnets in the network, and the maximum number of hosts that need to be assigned to each subnet in the network. For example, suppose we have 10 total subnets for our Class B network. We need 4 bits to represent this, because 2^4 is 16 while 2^3 is only 8. This leaves 12 bits for the host ID, for a maximum of 4,094 hosts per subnet.

  However, suppose we have 20 subnets. If so, 4 bits for subnet ID won't suffice. Hence, we need 5 bits (2^5=32). This means in turn that we now have only 11 bits for the host ID, for a maximum of 2,046 hosts per subnet.

- **Practical example of subnetting**

  1. **Step 1:** For successful subnetting, the network administrator must start by taking cognizance of the present and future requirements of the network. The most important parameter to establish is the number of subnets required and the maximum number of hosts needed for each subnet.It is of paramount importance to keep in mind that numbers should be based not just on present needs but on the requirements in the near future.

NOTES

2. **Step 2:** Deciding How Many Bits to Use for the Subnet ID and Host ID.

   Each bit taken from the host ID for the subnet ID doubles the number of subnets that are possible in the network.

   Each bit taken from the host ID for the subnet ID (approximately) halves the number of hosts that are possible within each subnet on the network.

   There are six possible ways this decision can be made for a Class C network, as illustrated in Figure 4.48.

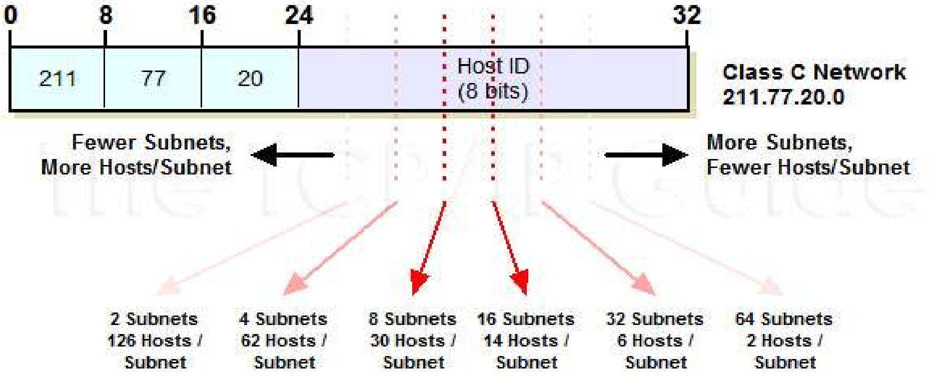


Fig. 4.48 *Six ways of Deciding How many Bits to use for the Subnet ID and Host ID*

The relationship between the bits and the number of subnets and hosts is as follows:

- o The number of subnets allowed in the network is two to the power of the number of subnet ID bits.
- o The number of hosts allowed per subnet is two to the power of the number of host ID bits, minus two.

Suppose there is a Class C network, base address 211.77.20.0, with a total of 7 subnets. The maximum number of hosts per subnet is 25.

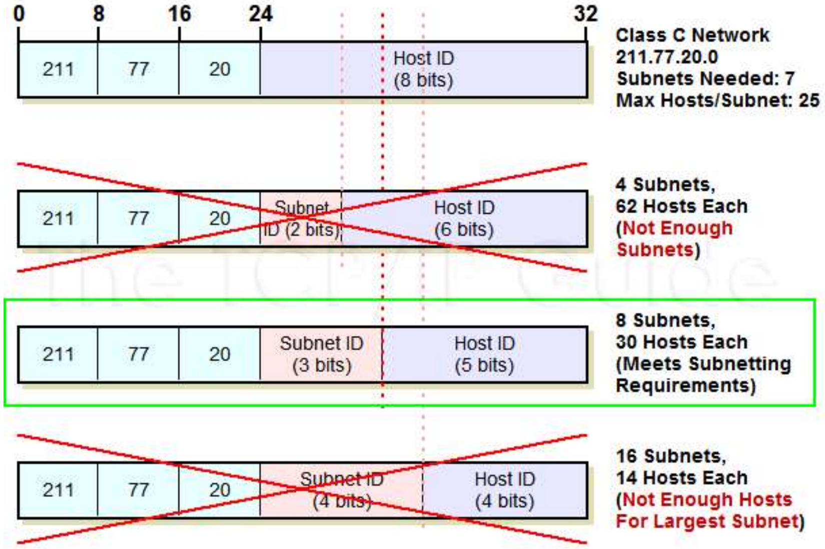


Fig. 4.49 *Deciding How many Bits to use for the Subnet ID and Host ID*

NOTES

If there is more than one combination of subnet ID and host ID sizes that meet requirements, the administrator should choose a 'middle-of-the-road' option that best foresees future growth requirements. If no combination meets the requirements, the requirements have to change.

3. **Step 3:** Determining the Custom Subnet Mask (Figure 4.50)

   Continuing with the above example

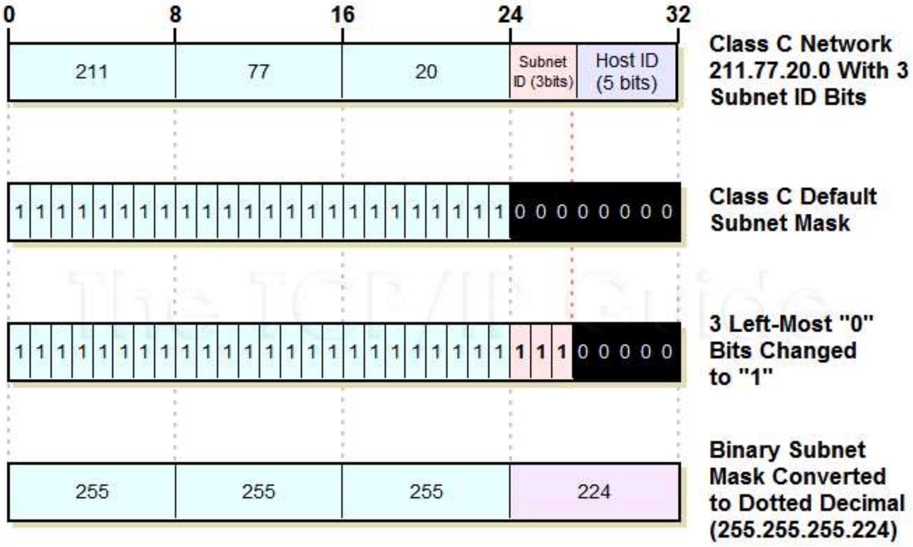


Fig. 4.50 *Determining the Custom Subnet Mask*

4. **Step 4:** Determining Subnet Identifiers and Subnet Addresses

   The network ID assigned to the network applies to the entire network. This includes all subnets and all hosts in all subnets. There is a unique identifier for each subnet within a network called the *subnet identifier* or *subnet ID*. This is to differentiate the subnet from the other subnets in the network (Figure 4.51).

   Recall our Class C network, 211.77.20.0. The network address in binary is:

   11010011 01001101 00010100 00000000

   We are subnetting using 3 bits for the subnet ID, leaving 5 bits for the host ID. Now let's see the network address with the subnet bits in bold:

   11010011 01001101 00010100 **000**00000

   Subnet #2 has a subnet ID of 2, or 010 in binary. To find its address we substitute "010" for the subnet ID bits, to give:

   11010011 01001101 00010100 **010**00000, which is 211.77.20.64 in binary.

NOTES

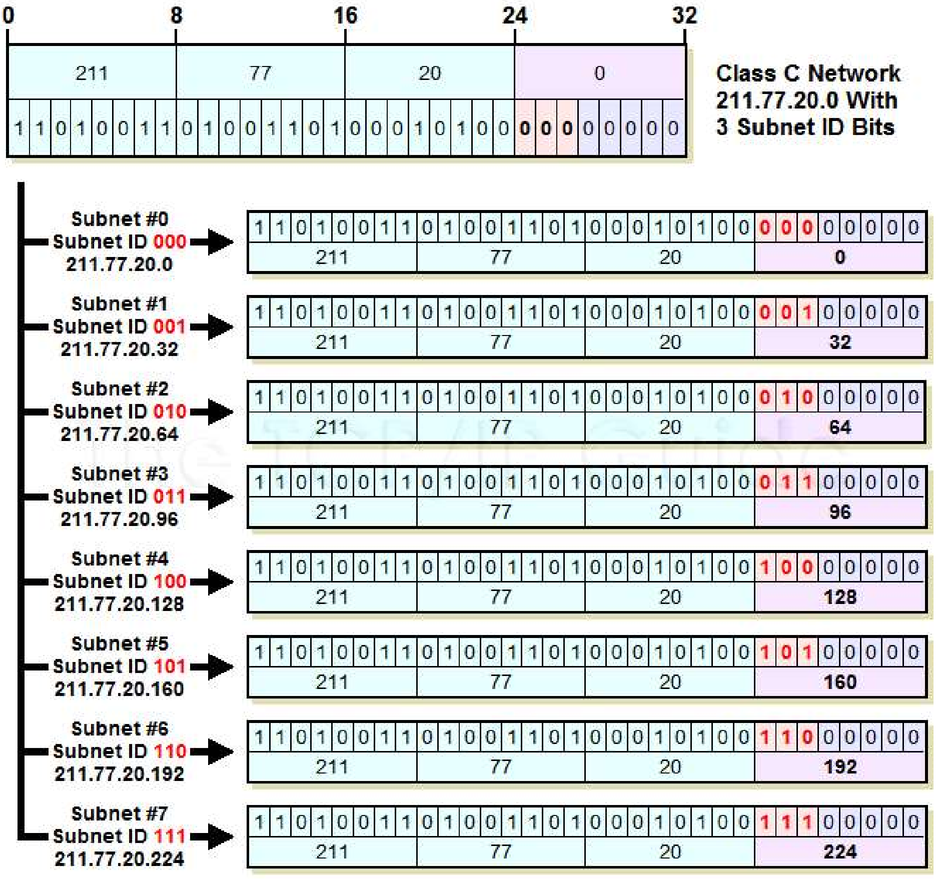


Fig. 4.51 Determining Subnet Identifiers and Subnet Addresses

5. **Step 5:** Determining Host Addresses for each Subnet

   Once the subnet addresses are known, these addresses can be used as the basis for assigning IP addresses to the individual hosts in each subnet.

   To continue with the above Class C example, 211.77.20.0, that was divided into 8 subnets using 3 subnet bits. The address appears as shown below with the subnet bits shown highlighted and the host ID bits shown highlighted and underlined.

   11010011 01001101 00010100 **00000000**

   The first subnet is subnet #0, which has all zeroes for those subnet bits, and thus, the same address as the network, as a whole: 211.77.20.0.

   1. The first host address has the number 1 for the host ID, or '00001' in binary. So, it is:

      11010011 01001101 00010100 **00000001**

      In decimal, this is 211.77.20.1.

   2. The second host address has the number 2 for the host ID, or '00010' in binary. Its binary value is:

      11010011 01001101 00010100 **00000010**

      In decimal, this is 211.77.20.2

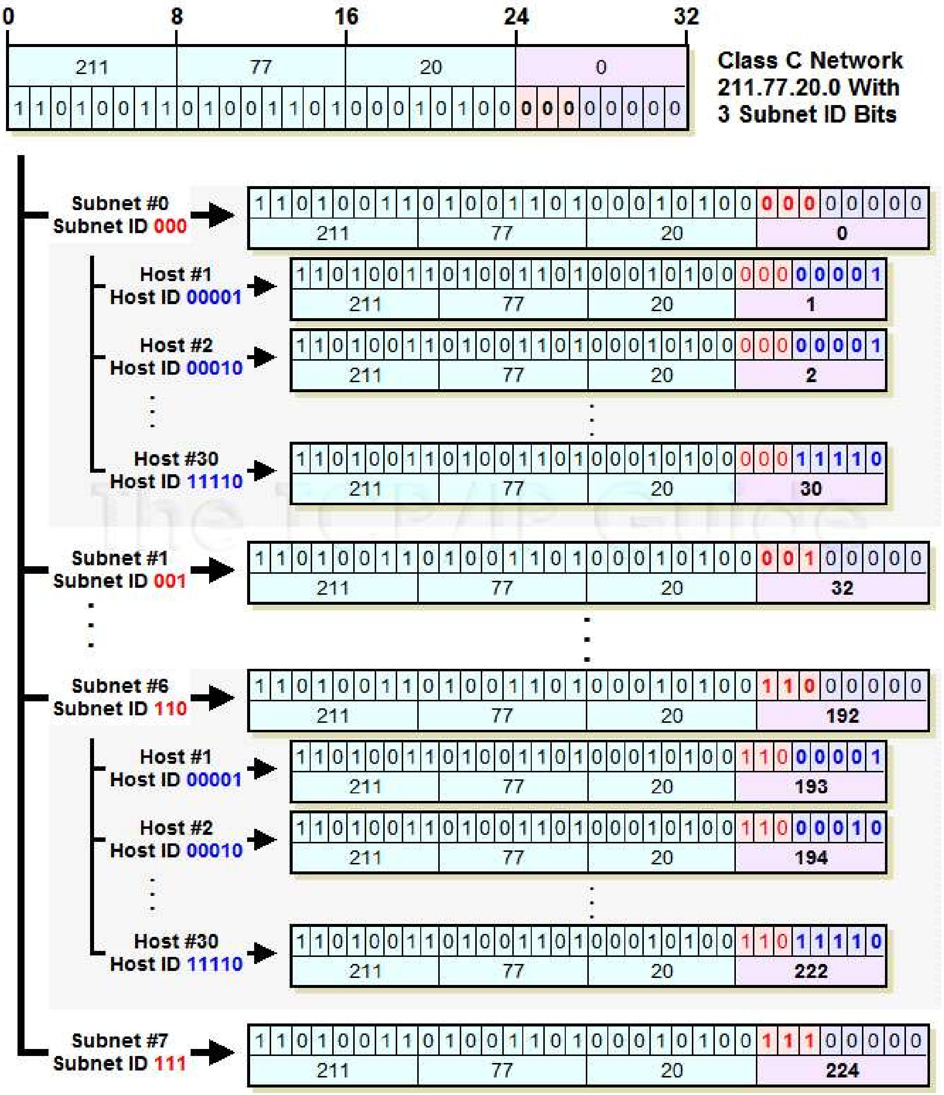


Fig. 4.52 Determining Host Addresses for each Subnet

- **Types of subnetting**

There are two types of subnetting: static and variable length.

1. **Static subnetting**: Static subnetting implies that all subnets obtained from the same network use the same subnet mask. The advantage of static subnetting is that it is simple to implement and easy to maintain. The only obvious disadvantage is that it might waste address space in small networks. Consider a network of four hosts using a subnet mask of 255.255.255.0. This allocation wastes 250 IP addresses. All hosts and routers are required to support static subnetting.
2. **Variable length subnetting**: In variable length subnetting or Variable Length Subnet Masks (VLSM), allocated subnets within the same network can use different subnet masks. A small subnet with only a few hosts can use a mask that accommodates this need. A subnet with many hosts requires a different subnet mask. The ability to assign subnet masks according to the needs of the individual subnets is useful in conserving network addresses. Variable length subnetting divides the network so that

each subnet contains sufficient addresses to support the required number of hosts.

An existing subnet can be split into two parts by adding another bit to the subnet portion of the subnet mask. Other subnets in the network are not affected by the change.

4.4.3 Supernetting

Supernetting allows the use of multiple IP networks on the same interface. Combination of multiple network addresses of the same Class into blocks is called Supernetting. If the IP networks are contiguous, you may be able to use a supernet. If the IP networks are not contiguous, you would need to use sub-interfaces.

A requirement for supernetting is that the network addresses be consecutive and that they fall on the correct boundaries. To combine two Class C networks, the first address' third octet must be evenly divisible by 2. If you would like to supernet 8 networks, the mask would be 255.255.248.0 and the first address' third octet needs to be evenly divisible by 8. For example, 198.41.15.0 and 198.41.16.0 could NOT be combined into a supernet, but you would be able to combine 198.41.18.0 and 198.41.19.0 into a supernet.

An IP address is a 32-bit number (4 bytes, called 'octets', separated by periods, commonly called 'dots.') Supernetting is most often used to combine Class C addresses (the first octet has values from 192 through 223). A single Class C IP network has 24 bits for the network portion of the IP address, and 8 bits for the host portion of the IP address. This gives a possibility of 256 hosts within a Class C IP network (2^8=256).

The subnet mask for a Class C IP network is normally 255.255.255.0. To use a supernet, the number of bits used for the subnet mask is reduced. For example, by using a 23 bit mask (255.255.254.0—23 bits for the network portion of the IP network, and 9 bits for the host portion), you effectively create a single IP network with 512 addresses. Supernetting, or combining blocks of IP networks, is the basis for most routing protocols currently used on the Internet.

For example: Consider two Class 'C' network numbers of 198.41.78.0 and 198.41.79.0.

The addresses pass the prerequisites. They are consecutive and the third octet of the first address is divisible by 2 (78 Mod 2 = 0). To further illustrate what is being done, let's look at the addresses in binary. The third octet of the first address (78) is 01001110. The second (79) is 01001111. The binaries are the same except for the last bit of the address (the 24th bit of the IP address). The 78 network is supernet 0 and the 79 network is supernet 1.

The subnet mask for this example supernet is 23 bits, or 255.255.254.0. All devices on the network MUST be using this subnet mask. Any device that is not using this subnet mask would be unreachable.

NOTES

Check Your Progress

5. What is internet protocol?
6. What do you understand by the term direct delivery?
7. Define IP datagram.
9. Define IP addressing.
9. Write the process of subnetting.
10. Define subnet mask.

4.5 OTHER NETWORK LAYER PROTOCOLS

The third layer is the network layer, also called the internet layer in the TCP/IP model, where internetworking protocols are defined, the most notable being the **Internet Protocol.** The main job performed here is the address resolution, or providing mappings between layer two and layer three addresses. This resolution can be done in either direction, and is represented by the two TCP/IP protocols namely, **ARP (Address Resolution Protocol)** and **RARP (Reverse Address Resolution Protocol)**. In TCP/IP, diagnostic, test and error-reporting functions at the network layer are performed by the **Internet Control Message Protocol (ICMP)**, which is like the Internet Protocol's 'administrative assistant'. The **Internet Group Management Protocol** (**IGMP**) is a communication protocol used to manage the membership of Internet Protocol multicast groups. IGMP is used by IP hosts and adjacent multicast routers to establish multicast group memberships.

ARP is a network-specific standard protocol. It is used for converting the higher-level IP addresses to physical network addresses.

RARP is used to find the logical address for a machine that only knows its physical address.

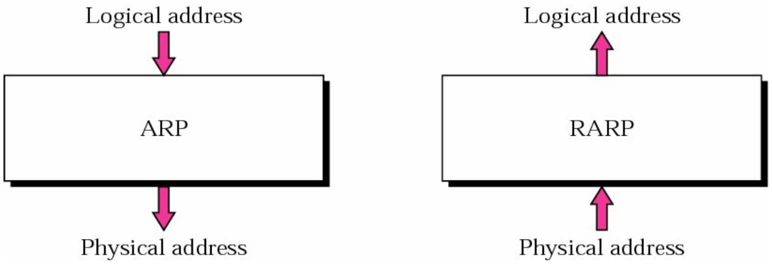


Fig. 4.53 *ARP and RARP*

Figure 4.54 shows the position of ARP and RARP in TCP/IP protocol suite.

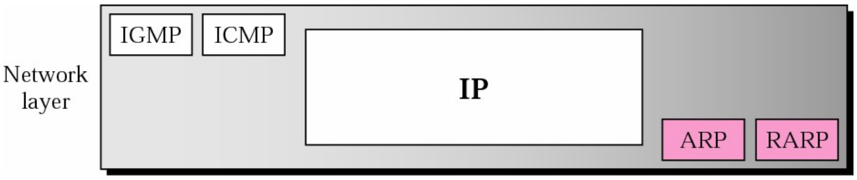


Fig. 4.54 *Position of ARP and RARP in TCP/IP Protocol Suite*

4.5.1 Address Resolution Protocol (ARP)

On a single physical network, individual devices are identified in the network by their physical hardware address. Higher-level protocols identify destination hosts by their symbolic address known as IP address. When such a high level protocol wants to send a datagram to destination host with IP address w.x.y.z, the device driver does not understand this address. The basic function of ARP is to translate the IP address to the physical address of the destination host. ARP uses a lookup table (sometimes referred to as the ARP cache) to carry out this translation (Figure 4.55).

In case the address is not found in the ARP cache, an ARP request is sent out in the network. An ARP request is a broadcast message with a special format sent in the network. An ARP reply is generated by a device on the network which recognizes its own IP address in the request. The reply contains the physical hardware address of the host and source route information (if the packet has crossed bridges on its path). The ARP cache of the requesting host is updated with this new information. All subsequent datagrams to this destination IP address can now be translated to a physical address by looking up the updated ARP cache, which is used by the device driver to send out the datagram in the network.

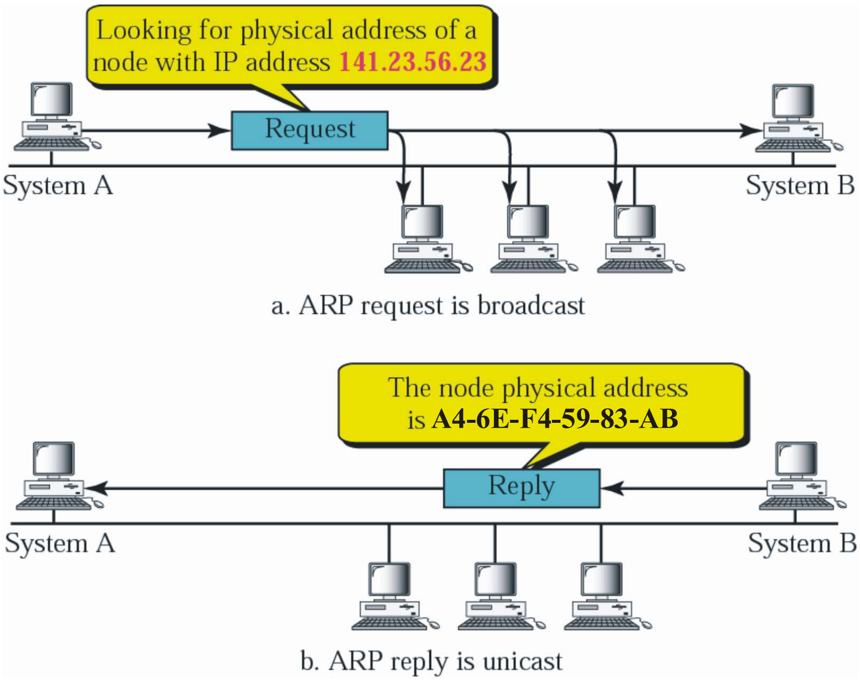

Looking for physical address of a node with IP address 141.23.56.23
Request
System A
System B
a. ARP request is broadcast
The node physical address is A4-6E-F4-59-83-AB
Reply
System A
System B
b. ARP reply is unicast

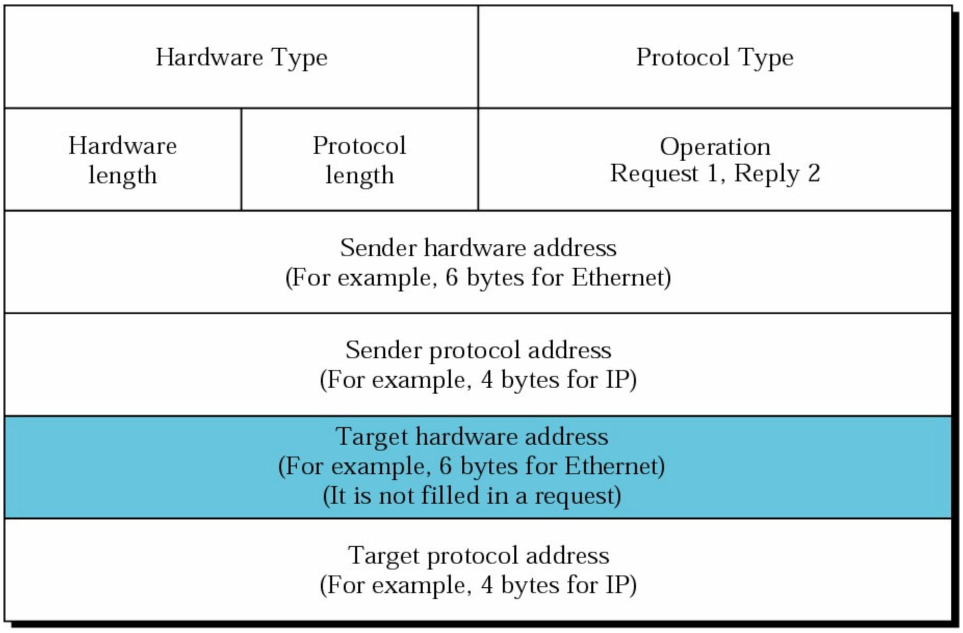
Fig. 4.55 *Address Resolution Protocol*

Working of ARP

- **ARP packet generation**

  If an application wants to send data to a certain IP destination address, the IP routing mechanism first determines the IP address of the next hop of the packet (it can be the destination host itself, or a router) and the hardware device on which it should be sent.

The ARP module tries to find the address in this ARP cache. If it finds the matching pair, it gives a corresponding 48-bit physical address back to the caller (the device driver), which then transmits the packet. If it does not find the pair in its table, it discards the packet (the assumption is that a higher-level protocol will retransmit) and generates a network broadcast of an ARP request.



Hardware Type		Protocol Type
Hardware length	Protocol length	Operation Request 1, Reply 2
Sender hardware address (For example, 6 bytes for Ethernet)		
Sender protocol address (For example, 4 bytes for IP)		
Target hardware address (For example, 6 bytes for Ethernet) (It is not filled in a request)		
Target protocol address (For example, 4 bytes for IP)		

Fig. 4.56 *ARP Packet*

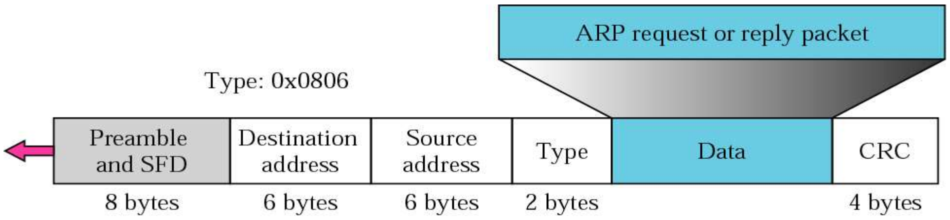


Fig. 4.57 *Encapsulation of ARP Packet*

ARP packet reception

On receipt of an ARP packet (either a broadcast request or a point-to-point reply), the recipient device driver passes the packet to the ARP module which treats it as shown in Figure 4.58.

NOTES

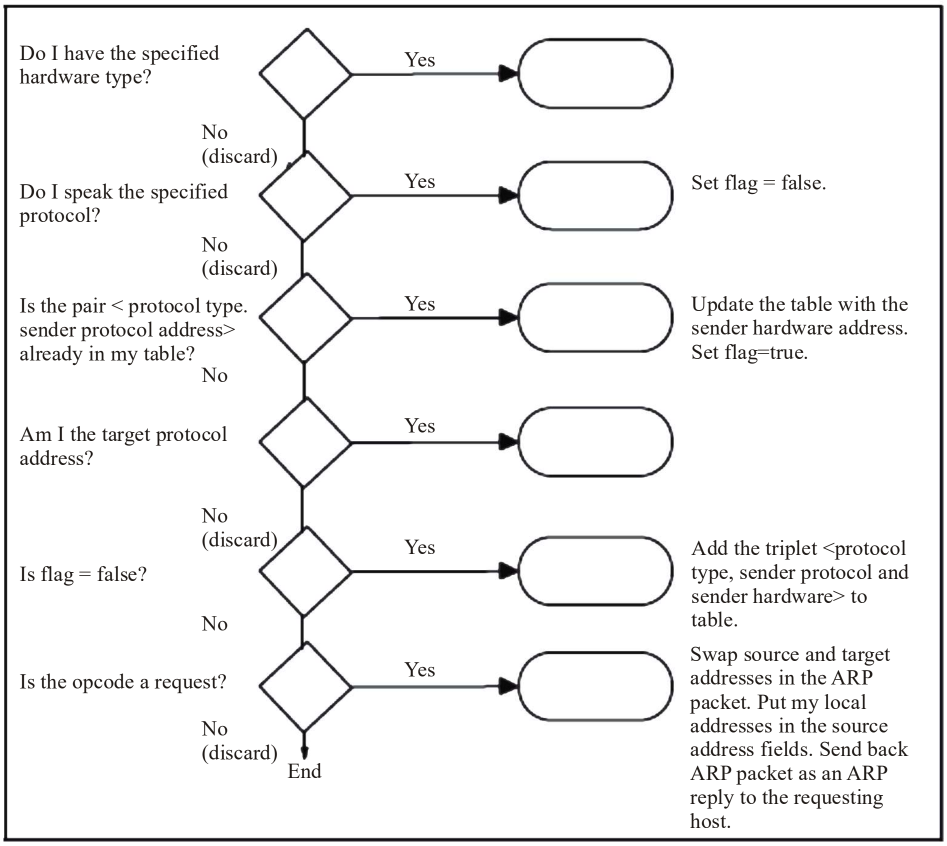

Do I have the specified hardware type?
Yes
No (discard)
Do I speak the specified protocol?
Yes
Set flag = false.
No (discard)
Is the pair < protocol type. sender protocol address> already in my table?
Yes
Update the table with the sender hardware address. Set flag=true.
No
Am I the target protocol address?
Yes
No (discard)
Is flag = false?
Yes
Add the triplet <protocol type, sender protocol and sender hardware> to table.
No
Is the opcode a request?
Yes
Swap source and target addresses in the ARP packet. Put my local addresses in the source address fields. Send back ARP packet as an ARP reply to the requesting host.
No (discard)
End

Fig. 4.58 *ARP Packet Reception*

The requesting host will receive this ARP reply, and will follow the same algorithm to treat it. As a result, its lookup table (ARP cache) will be updated with the hardware address of the recipient device. Subsequently, when a higher-level protocol wants to send a packet to that host, the ARP module will find the target hardware address and the packet will be sent to the corresponding host (Figure 4.59).

Example

A host with IP address 130.23.43.20 and physical address B2-34-55-10-22-10 has a packet to send to another host with IP address 130.23.43.25 and physical address A4-6E-F4-59-83-AB (which is unknown to the first host). The two hosts are on the same Ethernet network. Show the ARP request and reply packets encapsulated in Ethernet frames.

Solution:

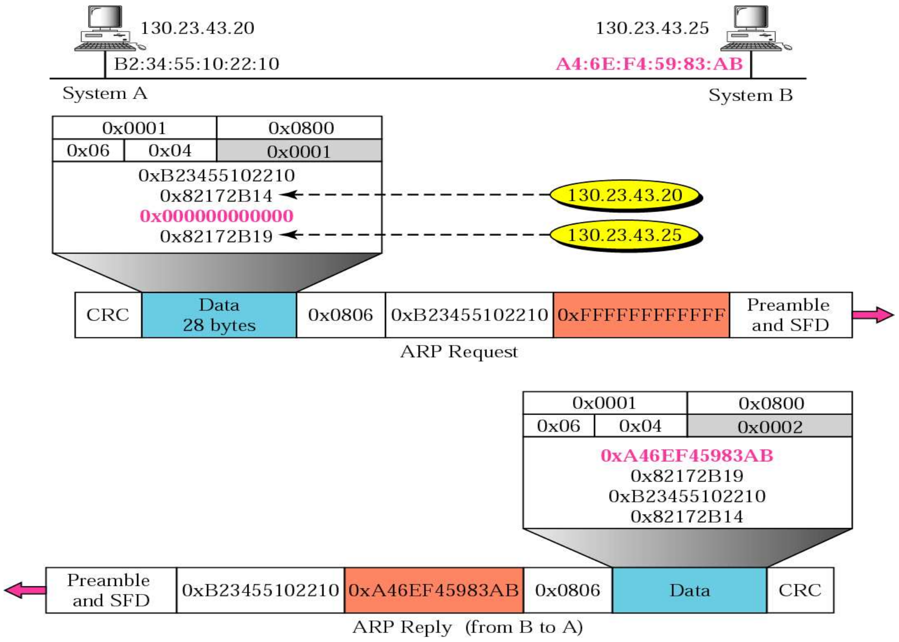


Fig. 4.59 *ARP Request and Reply Packets Encapsulated in Ethernet Frames*

Proxy ARP

Suppose there is an IP network that is divided into subnets and interconnected by routers. Considering the 'old' IP routing algorithm, no host on the network is aware of the existence of multiple physical networks. Consider hosts A and B, which are on different physical networks within the same IP network, and a router R between the two subnetworks as illustrated in Figure 4.60.

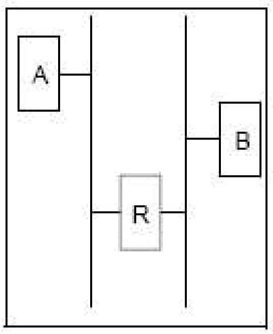


Fig. 4.60 *Hosts A and B and Router R*

When host A wants to send an IP datagram to host B, using the ARP protocol, it first has to determine the physical network address of host B. Assuming, host B to be on the local physical network, host A sends out a broadcast ARP request to find out the physical address of host B. Host B does not receive this broadcast, but router R does. Router R is aware of subnets. It runs the subnet version of the IP routing algorithm and is able to determine that the destination of the ARP request (from the target protocol address field) is on another physical network. The router R replies to the request in lieu of host B. Router R will specify its own address in the reply and the datagram will be delivered to Router R. The router will then forward such packets to the correct subnet.

NOTES

Host A updates its cache, and will send future IP packets for host B to the router R (Figure 4.61).

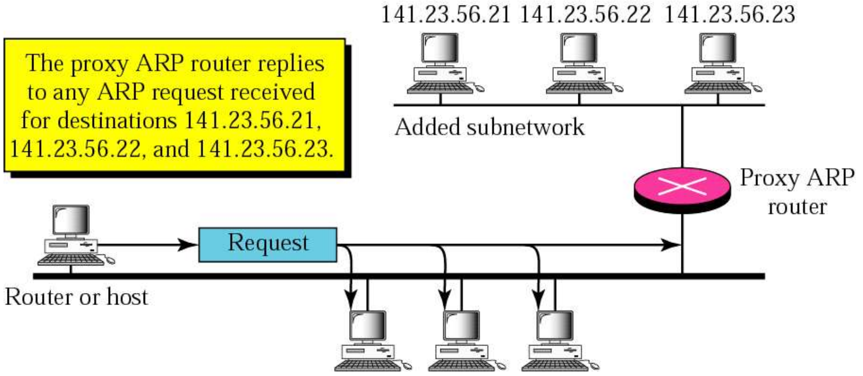
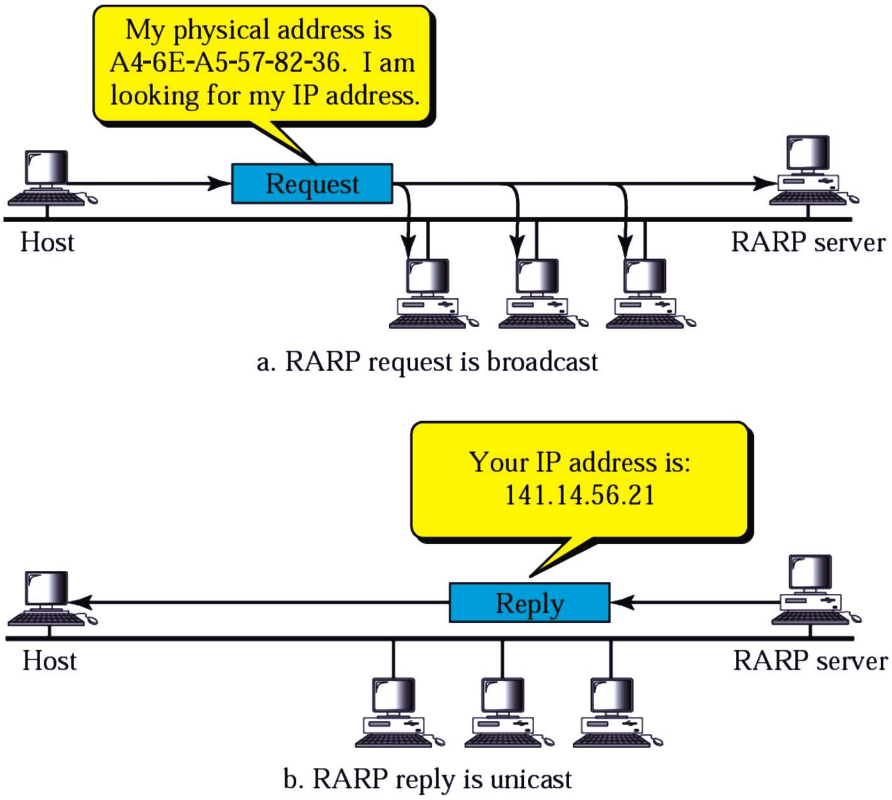

141.23.56.21 141.23.56.22 141.23.56.23
The proxy ARP router replies to any ARP request received for destinations 141.23.56.21, 141.23.56.22, and 141.23.56.23.
Added subnetwork
Proxy ARP router
Request
Router or host

Fig. 4.61 *Proxy ARP Router*

4.5.2 Reverse Address Resolution Protocol (RARP)

Some network hosts, such as diskless workstations, are not aware of their own IP address when they are booted. In order to determine their own IP address, they use a mechanism similar to ARP. This mechanism is called Reverse Address Resolution Protocol (RARP). The known parameter is the hardware address of the host and the IP address is the parameter to be determined. It differs more fundamentally from ARP in the fact that a RARP server must exist in the network (Figure 4.62). This server maintains a database of mappings from hardware address to protocol address and is preconfigured.


My physical address is A4-6E-A5-57-82-36. I am looking for my IP address.
Request
Host
RARP server
Your IP address is: 141.14.56.21
Reply
Host
RARP server

a. RARP request is broadcast

b. RARP reply is unicast

Fig. 4.62 *Reverse Address Resolution Protocol*

4.5.3 Encapsulation

The IGMP message is encapsulated in an IP datagram, which is encapsulated in a frame (Figure 4.63).

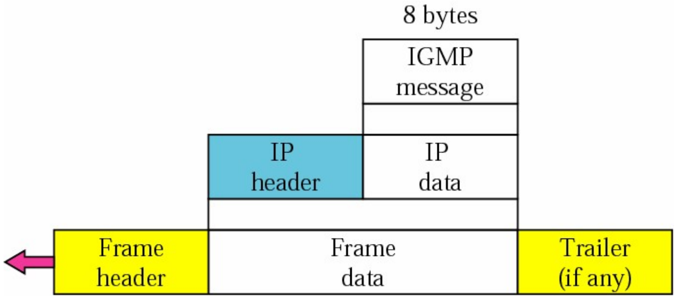


Fig. 4.63 *Encapsulation of IGMP Message*

- The IP packet that carries an IGMP packet has a value of 2 in its protocol field.
- The IP packet that carries an IGMP packet has a value of 1 in its TTL field.

4.5.4 Ethernet

The IEEE 802.3 or CSMA/CD protocol was based on the specification called Ethernet, formally developed by Xerox and later modified and accepted by IBM and DEC. The protocol is very simple. The station willing to transmit information would listen to the cables before transmitting anything. If the cable is busy, the station waits until it becomes idle. If two or more stations simultaneously start the transmission, the transmitted signals collide with each other. Under such circumstances, each transmitting station hears a collision message, waits for a random period, and repeats the transmission process again.

Depending on the type of the transmission media used, the Ethernet can be classified into the following categories. They are:

Thick Ethernet or 10base5

A 10 Mbps cable is like a yellow hose. The maximum length of the cable is 2.5 meters with direction markings. The direction markings help to find out where the cable goes.

The following are the characteristics of the thick Ethernet (10base5):

- Cable supports a maximum distance of 500 meters.
- It provides connectivity to a maximum of 1024 stations.
- Maximum distance covered by a network using thick Ethernet is 2.5 km.
- Maximum number of stations supported by the Ethernet is 1024.

A medium-sized network based on Ethernet is shown in Figure 4.64.

NOTES

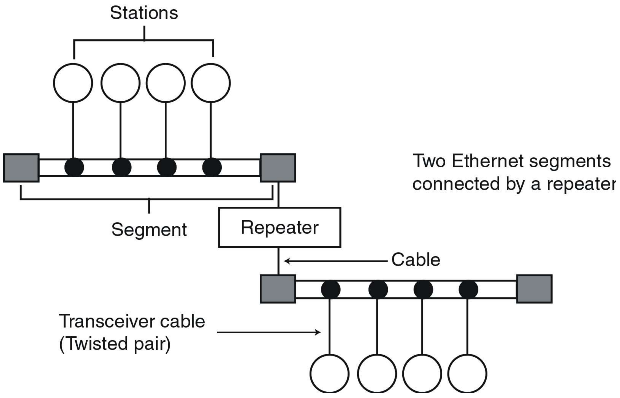

Stations
Two Ethernet segments connected by a repeater
Segment
Repeater
Cable
Transceiver cable (Twisted pair)

***Fig. 4.64** Baseband Ethernet*

All stations are connected to a coaxial cable. A group of stations connected to a cable forms a segment. A device called repeater is used to link two network segments, which are separated by long distance. Each station is connected to the Ethernet cable through a transceiver. A transceiver is a transmitter-receiver pair, which can extract or insert signals on a cable in one direction. A repeater consists of two transceivers. Hence, it is able to transmit and receive both directions.

Implementation of Ethernet

General implementation of an Ethernet is shown in Figure 4.65. The two significant layers that are implemented are the physical layer and the data link layer.

The physical layer performs the following functions:

(a) **Encoding the data:** This process generates the synchronization bits called preamble for the data frame at the transmitter. At the receiver, it removes the preamble from the received frame. The physical layer also performs the encoding and decoding of data.

(b) **Medium access:** The physical layer transmits and receives data by sensing the idle channel. In the event of a collision that has occurred on the channel during the transmission, the physical layer recognises it and intimates this to the data link layer.

The data link layer performs the following functions:

(a) **Data encapsulation:** Data encapsulation includes formation of frame, addressing, error detection.

(b) **Link management:** This includes allocation of channels, collision avoidance, error detection resolving collision.

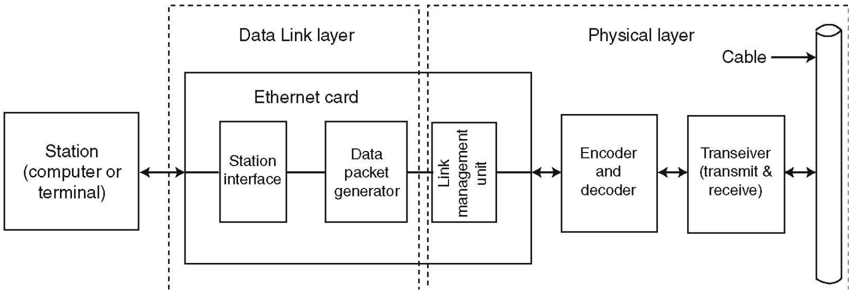
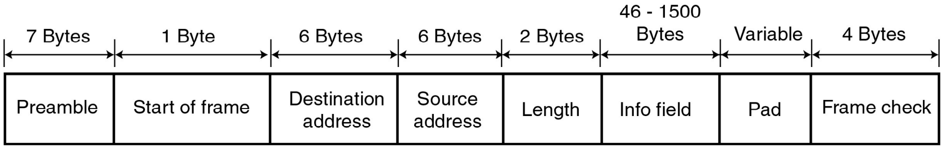


***Fig. 4.65** Ethernet Implementation*

The computer or station is connected to an Ethernet card. The Ethernet card consists of a station interface, data packet generator and a link management unit. The first two units form the data link layer. The output of the Ethernet card is connected to the data encoder/decoder, which in turn is connected to the transmission cable through a transceiver. The link management unit of the Ethernet card, data encoder/decoder, the transceiver and the transmission cable form the physical layer.

Ethernet frame format

The IEEE 802.3 Ethernet frame format is shown in Figure 4.66.



7 Bytes	1 Byte	6 Bytes	6 Bytes	2 Bytes	46 - 1500 Bytes	Variable	4 Bytes
Preamble	Start of frame	Destination address	Source address	Length	Info field	Pad	Frame check

Fig. 4.66 *Ethernet Frame Format*

Preamble: 7-byte synchronization pattern, consisting of alternative 0s and 1s is used for receiver synchronization.

Start of frame: A 1-byte word similar to the preamble ends with two consecutive 1 bits.

Destination address: A 6-bytes address specifies the station to which a packet is addressed. This address may be an individual address or a group address.

Source address: It is the address of the packet-originating station. Its size is equivalent to the size of the destination address.

Length: This field gives the length of the actual data bytes transmitted in the information field. Size of this field is 2 bytes.

Information field: The size of the information field is a variable. It must be a minimum of 46 bytes. In case it is less than 46 bytes, dummy frames (called 'pad') are included in place of the information field to make up the minimum length. The upper limit for the information field is 1500 bytes.

Frame check sequence: A 4-byte code used for the purpose of error-detection. It detects the presence of errors in the destination address, source address, length and information fields.

Cheaper Net or Thin Net (10base2)

For local area networks that do not require the capabilities of a complete Ethernet system, the IEEE 802.3 standard committee has created a new standard called thin net. The differences between the Ethernet and the cheaper net are given below.

Thick Ethernet	**Cheaper net or Thin net**
Maximum segment length is 500 metres.	Maximum length is upto 200 metres.
Maximum number of nodes per segment is 100.	Maximum number of nodes is 30.
Maximum number of stations per network is 1024.	Maximum stations per network is 1024.
Node spacing is 2.5 metres.	Node spacing is 0.5 metres.
Network cable diameter is 0.4 inches.	Cable diameter is 0.25 inches.
Cable is connected through a vampire tap.	BNC-T-connector is used to connect cables and N-series connector.

NOTES

StarLAN (10BaseT)

The third variation of IEEE 802.3 standard was a StarLAN. This particular standard was originally proposed by AT&T. Local area network switching, based on this standard operates at a data rate of 1 Mb per second. The twisted-pair cable already used in telephone lines could be used as a transmission media. Tree topology is used to configure a StarLAN. Each group of stations is connected to a local hub. Hubs are connected in the form of a tree. The root of the tree is the header hub. The configuration may contain upto 5 upward levels of hubs. The transmitted message first reaches the local hub. Then it is transmitted upward until it reaches the header hub. From there it is broadcasted down to all the stations on the network.

Optical Fibre CSMA/CD LAN (10BaseF)

Optical fibre version of CSMA/CD LAN has a number of advantages than the coaxial cable version of Ethernet. They have good immunity to the electromagnetic interference, low loss of power, high bandwidth and less weight, and high transmission security. Hubs in an optical fibre LAN are widely separated. However, it is expensive while considering the cost of couplers and terminators.

4.6 ROUTING PROTOCOLS

Internetworking involves connecting different physical networks. Providing connections between dissimilar networks is one of the basic functions provided by the IP. A system that performs this function is called an *IP router*. An IP router is a device that attaches to two or more physical networks and transfers datagrams between the networks.

A host sends data to a remote destination in the form of a datagram. The datagram travels from one router to another. The router forwards the datagram towards the final destination. Each router chooses the next device along the path to reach the destination. This next device is called the *next hop* device. The datagram travels till it reaches a router connected to the destination's LAN segment. The destination LAN segment differs from the one on which the system originally received the datagram, the intermediate host has *forwarded* (that is, routed) the IP datagram from one physical network to another.

To forward packets between network segments, it is the IP routing table in each device that is used. The basic table contains information about a router's locally connected networks. The configuration of the device can be extended to contain information detailing remote networks. This information provides a more complete view of the overall environment. Routing tables can be static or dynamic.

In a Static Routing Table, information is entered manually. The administrator enters the route for each destination into the table. As the Internet changes, static routing table is not automatically updated. A static routing table can be used in a small network that does not change very often.

NOTES

Dynamic Routing Table is updated periodically and automatically using one of dynamic routing protocols such as RIP, OSPF or BGP. All tables in the routers are updated automatically by dynamic routing protocols whenever there is a change in the Internet.

Common fields in a routing table (Figure 4.67):

- **Mask:** Defines mask applied for the entry
- **Network address:** Address of the destination host
- **Next-hop address:** Address of the next hop router to which packet is delivered
- **Interface:** Name of the interface
- **Flag:** This field defines five flags
  - *U(up):* Router is up and running
  - *G(Gateway):* Destination is in other network
  - *H(Host specific):* Entry in network address is a host specific address
  - *D(Added by redirection):* Routing information for this destination has been added to the routing table by a redirection message from ICMP
  - *M(Modified by redirection):* Routing information for this destination has been modified by a redirection message from ICMP
- **Reference count:** Number of users of the route at the moment
- **Use:** Shows the number of packets transmitted through this router for the corresponding destination.

Mask	Network address	Next-hop address	Interface	Flags	Reference count	Use
....	....	....	....	....	....	....

Fig. 4.67 *Routing Table*

A routing protocol is characterized as robust if it provides the ability to dynamically build and manage the information in the IP routing table. As changes occur in the network topology, the routing tables are updated with minimal or no manual intervention.

Autonomous Systems

An Autonomous System (AS) is defined as a logical portion of a larger IP network (Figure 4.68). An AS normally consists of an internetwork within an organization. It is administered by a single management authority.

NOTES

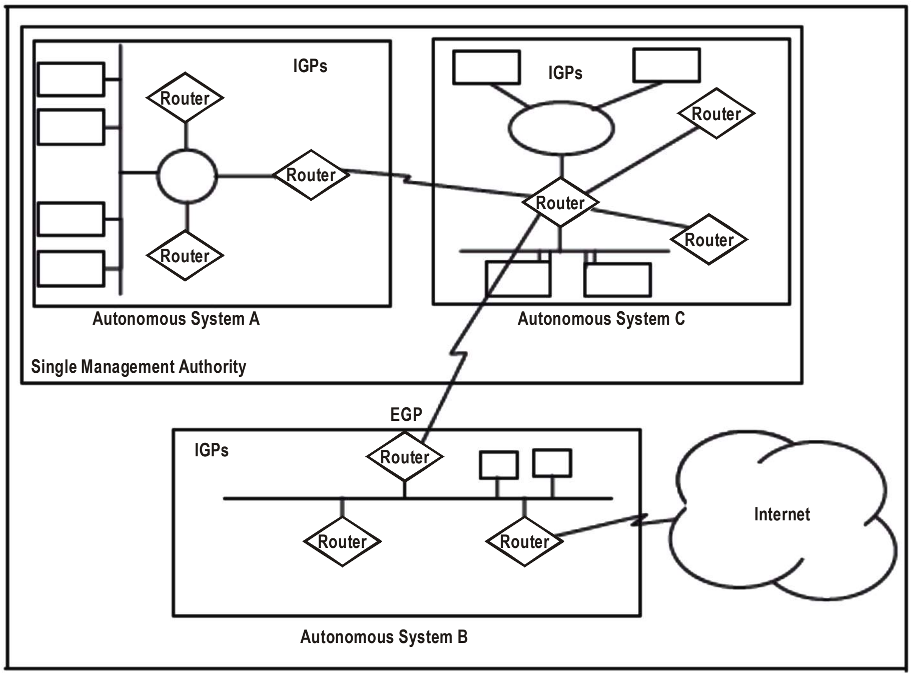


Fig. 4.68 *Autonomous Systems*

4.6.1 Routing Protocols

Some routing protocols are used to determine routing paths within an AS. Others are used to interconnect different autonomous systems:

- **Interior Gateway Protocols (IGPs)**: These are referred to as IntraDomain Protocols. Interior Gateway Protocols allow routers to transmit information within an AS. Examples of these protocols are Open Short Path First (OSPF) and Routing Information Protocol (RIP).
- **Exterior Gateway Protocols (EGPs)**: These are referred to as InterDomain Protocols. Exterior Gateway Protocols interconnect different autonomous systems. They allow the exchange of summary information between autonomous systems. An example of this type of routing protocol is Border Gateway Protocol (BGP) (Figure 4.69).

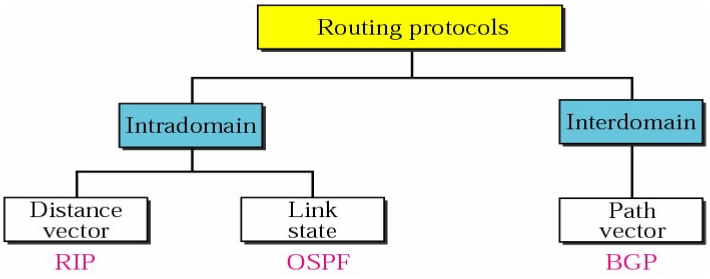


Fig. 4.69 *Types of Routing Protocols*

Unicasting

Communication between one source and one destination is known as unicast. There is a one to one relationship between source and destination. In this type of communication, both the source and destination addresses are unicast addresses (Figure 4.70).

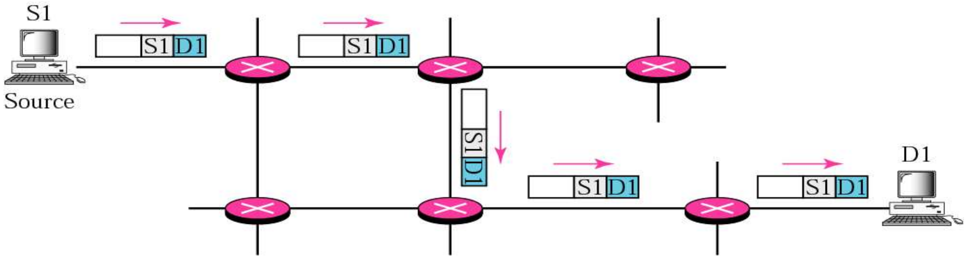


Fig. 4.70 *Unicasting*

NOTES

In unicasting, the router forwards the received packet through only one of its interfaces.

4.6.2 Types of IP Routing and IP Routing Algorithms

The main function of routing algorithms is to build and maintain the IP routing table on a device. There are two primary methods used to build the routing table.

- **Static routing**: Static routing uses fixed definitions representing paths through the network.
- **Dynamic routing**: In dynamic routing, algorithms routers can automatically determine and maintain knowledge of the paths through the network. This automatic discovery can use a number of currently available dynamic routing protocols. The protocols are differentiated on the basis of the way they determine and compute new routes to destination networks. They can be classified into four broad categories:
  - (i) Distance vector protocols
  - (ii) Link state protocols
  - (iii) Path vector protocols
  - (iv) Hybrid protocols

1. **Static Routing**

   Static routing is manually performed by the network administrator. It is the responsibility of the administrator to determine and broadcast routes through the network. These definitions are, then, manually programmed in every routing device in the network.

   The routers in the network do not communicate with each other about the changing topology of the network. After a device has been configured, it simply forwards packets to the predetermined ports.

   Static routing is relatively simple and easy to administer in a small non redundant network which does not change frequently. However, there are several disadvantages to this approach for maintaining IP routing tables:

   - Static routes require a considerable amount of co-ordination and maintenance in non-trivial network environments.

- In Static Routing, if the current operational state of the network changes, static routes cannot adapt to it. For example, if a destination subnetwork becomes unreachable, the change is not reflected in the routing table. The static routes pointing to the inaccessible network remain in the routing table. Traffic continues to be forwarded towards that destination. The network administrator has to manually make the change. Unless the network administrator updates the static routes to reflect the new topology, traffic is unable to use any alternate paths that may exist.

Normally, static routes are used only in simple network topologies. However, there are additional circumstances when static routing can prove to be of greater use. For example, static routes can be used:

- To define a default route manually. When the routing table does not contain a more specific route to the destination, a default route is used to forward traffic.
- To define a route that is not automatically advertised within a network.
- When it is undesirable to send routing advertisement traffic through lower-capacity WAN connections because of utilization or line tariffs.
- To define complex routing policies. Static routing can be used to create predetermined paths to a certain host. For example, static routes can be used to forward traffic to a specific host through a designated network path.
- To provide a more secure network environment. Since automatic update is not possible, the administrator has more control over the network. The administrator is aware of all subnetworks defined in the network. The administrator has the final authority over all communication permitted between these subnetworks.
- To provide more efficient resource utilization. Static Routing is more efficient because this method of managing a routing table requires no network bandwidth to advertise routes between neighbouring devices. It also uses less processor memory and CPU cycles to calculate network paths.

2. **Distance Vector Routing**

As per distance vector routing, the route with the minimum distance between two nodes is the route with the least cost. In this protocol, each node maintains a table of minimum distances to every node.

Distance vector algorithms are examples of dynamic routing protocols. In distance vector algorithms, each device in the network can automatically construct and maintain a local IP routing table.

There is a very simple principle behind distance vector routing. Each device in the network maintains a *distance vector table.* In the distance vector table, the device maintains the distance or cost from itself to a known destination. This value shows the attractiveness of the path. The least cost path is more desirable than a path with a higher value. The least cost path becomes the chosen path to reach the destination.

NOTES

The table is periodically advertised to each neighbouring router. Each router processes these advertisements to establish the best paths through the network and continually update their tables.

Distance vector algorithms are advantageous because they are simple, easy to implement and debug. They are very useful in small networks with limited redundancy. However, there are several disadvantages with this type of protocol, some of which are mentioned below.

- During an unfavourable condition, the length of time for every device in the network to produce an exact routing table is called the *convergence time*. This time can be excessive in large, complex internetworks using distance vector algorithms. While the routing tables are converging, networks are vulnerable to inconsistent routing behaviour. This can cause routing loops or other types of unstable packet forwarding.
- In order to reduce convergence time, a limit is often placed on the maximum number of hops contained in a single route. Paths that are valid but exceed this limit of hops cannot be used in distance vector networks.
- Devices periodically advertise their distance vector routing tables to neighbouring devices. They are sent even if no changes have been made to the contents of the table. This is not efficient because this can cause noticeable periods of increased utilization in reduced capacity environments.

RIP is a popular example of a distance vector routing protocol.

3. **Link State Routing**

Link State Routing algorithms are based on the principle of a *link state* to establish network topology. A link state is the description of an interface on a router, such as IP address, subnet mask, type of network and its relationship to neighbouring routers. The collection of these link states forms a link state database.

The following is the process used by link state algorithms to determine network topology.

(i) All other routing devices on the directly connected networks are identified by each router.

(ii) Each router maintains information about all directly connected network links and the associated cost of each link. This information is advertised to neighboring devices. This is carried out through the exchange of link state advertisements (LSAs) with other routers in the network.

(iii) Each router uses these advertisements to create a database detailing the current network topology. The topology database in each router is similar.

(iv) Using the Dijkstra Algorithm and the information in the topology database each router computes the most desirable routes to each destination network. This information is used to update the IP routing table.

NOTES

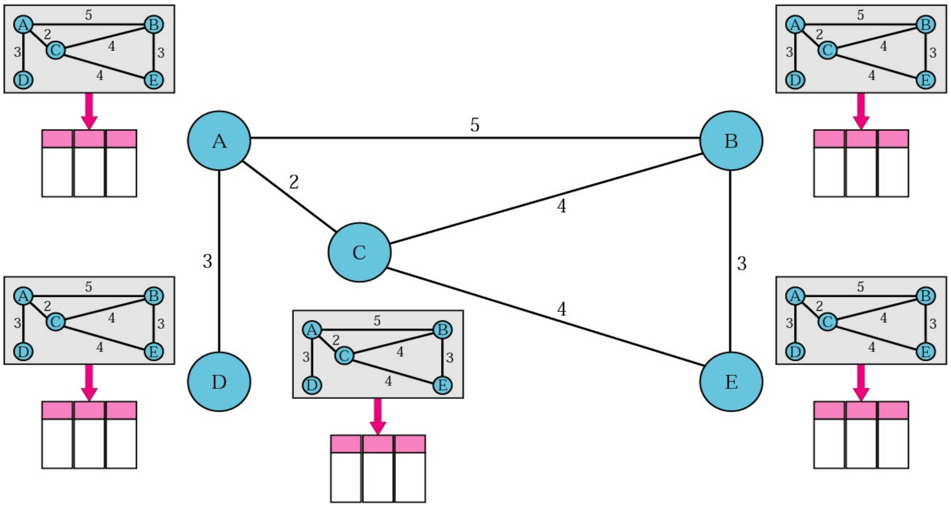


Fig. 4.71 *Link State Routing*

Figure 4.71 shows a simple domain with five nodes. Each node uses the same topology for creating a routing table, but the routing table for each node is unlike because the calculations are based on different interpretations of the topology.

4. **Dijkstra Algorithm (Shortest-Path First (SPF) Algorithm)**

The SPF algorithm is used to process the information in the topology database (Figure 4.72). The SPF algorithm provides a tree-representation of the network. The device running the SPF algorithm is the root of the tree. The output of the algorithm is the list of shortest-paths to each destination network.

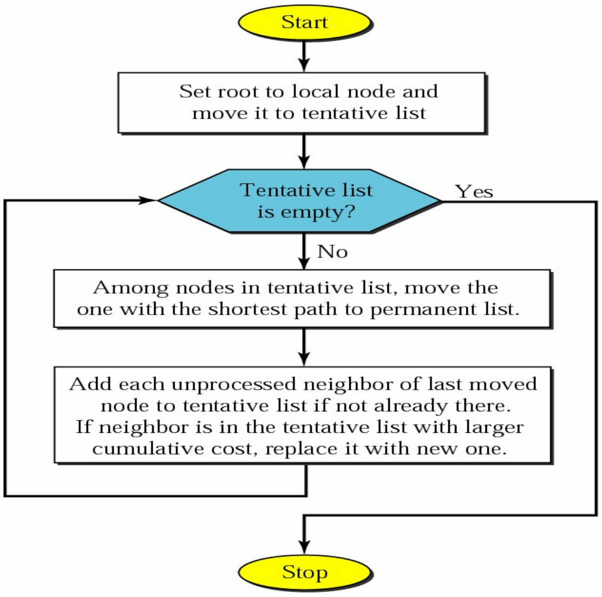
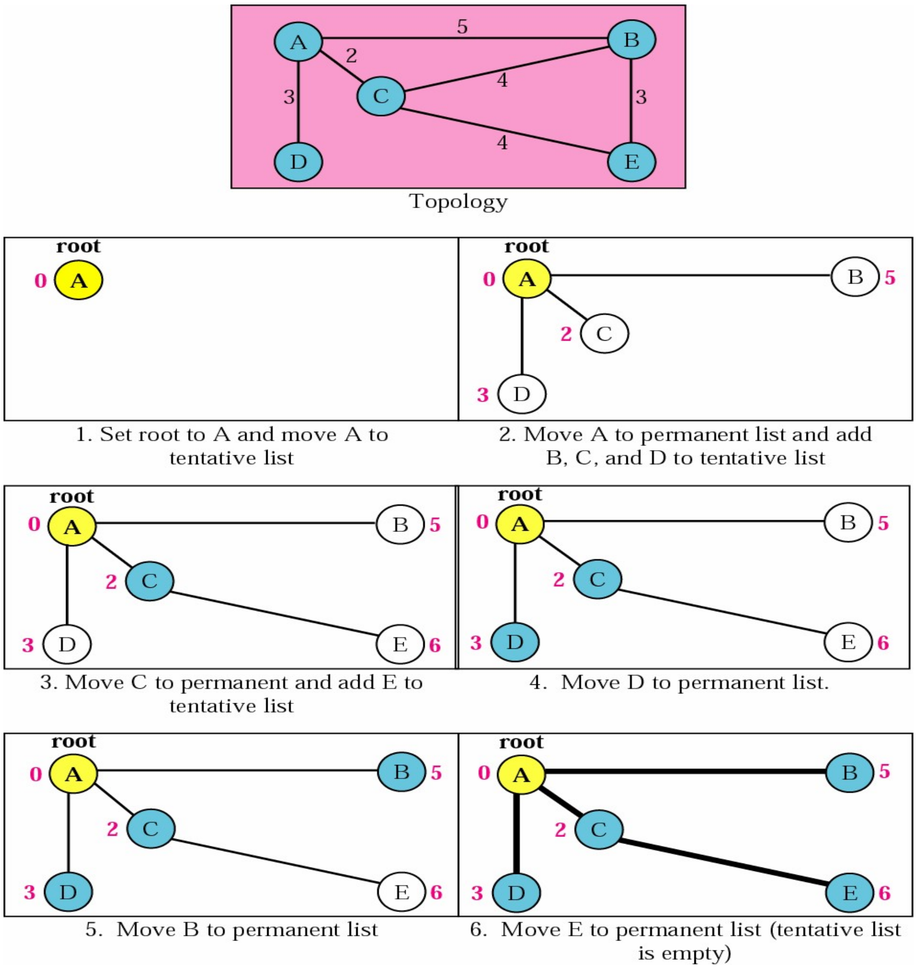

Start
Set root to local node and move it to tentative list
Tentative list is empty?
Yes
No
Among nodes in tentative list, move the one with the shortest path to permanent list.
Add each unprocessed neighbor of last moved node to tentative list if not already there. If neighbor is in the tentative list with larger cumulative cost, replace it with new one.
Stop

Fig. 4.72 *Dijkstra Algorithm*

NOTES

Figure 4.73 shows the process of the formation of a shortest path tree.



Topology

1. Set root to A and move A to tentative list

2. Move A to permanent list and add B, C, and D to tentative list

3. Move C to permanent and add E to tentative list

4. Move D to permanent list.

5. Move B to permanent list

6. Move E to permanent list (tentative list is empty)

Fig. 4.73 *Example of Formation of Shortest Path Tree*

Figure 4.74 shows a routing table for node A.

Node	*Cost*	*Next Router*
A	0	—
B	5	—
C	2	—
D	3	—
E	6	C

Fig. 4.74 *Routing Table for Node A*

The OSPF protocol is a popular example of a link state routing protocol.

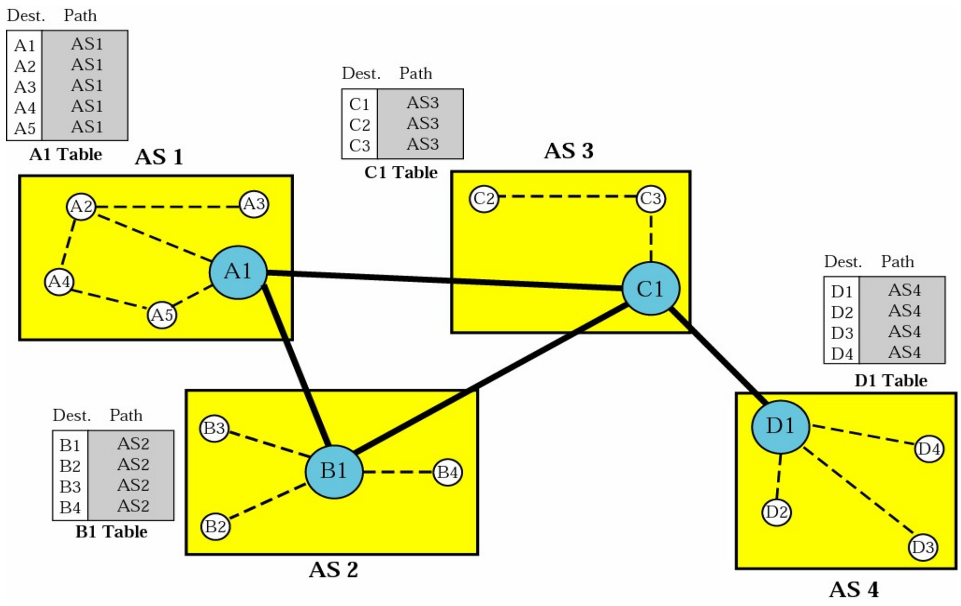
5. **Path Vector Routing**

This type of routing is mainly used in routing traffic between different Autonomous Systems. Each Autonomous System contains at least one node called the speaker node. This speaker node creates a routing table and advertises it to the speaker nodes in neighbouring autonomous systems. Only the path is advertised, the metrics of the node are not advertised.

A route is defined as a pairing between a destination and the attributes of the path to that destination, thus the name, path vector routing, where the routers receive a vector that contains paths to a set of destinations.

The path is expressed in terms of the domains (or confederations) traversed so far. A special path attribute that records the series of routing domains through which the reachability information has passed carries the path. The preferred path to reach the destination is the path represented by the smallest number of domains.

- **Initialization**



Dest.	Path
A1	AS1
A2	AS1
A3	AS1
A4	AS1
A5	AS1

A1 Table

Dest.	Path
C1	AS3
C2	AS3
C3	AS3

C1 Table

Dest.	Path
D1	AS4
D2	AS4
D3	AS4
D4	AS4

D1 Table

Dest.	Path
B1	AS2
B2	AS2
B3	AS2
B4	AS2

B1 Table

Fig. 4.75 *Path Vector Routing*

At the beginning, each speaker node knows only the paths inside its own autonomous system. Node A1 is the speaker node for AS1, B1 for AS2, C1 for AS3, and D1 for AS4. Each node creates an initial table that shows paths within the autonomous system. Node A1 creates an initial table that shows A1 to A5 are located in AS1 and can be reached through it. Node B1 creates an initial table that shows B1 to B4 are located in AS2 and can be reached through it (Figure 4.75).

- **Sharing**

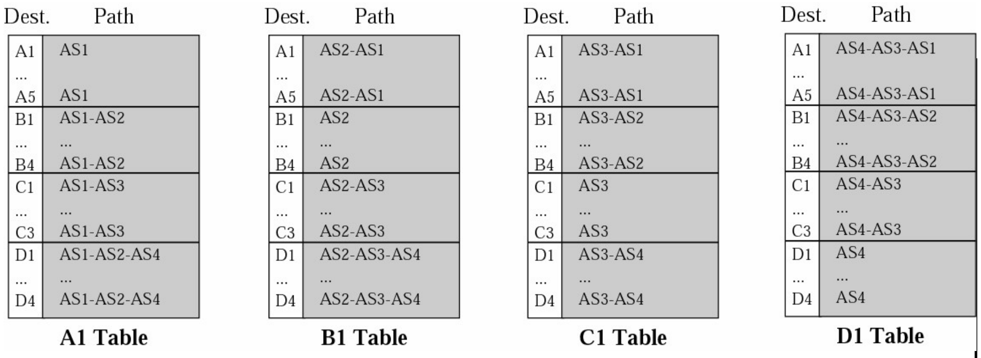
Each speaker node shares its initial tables with other speaker nodes. Node A1 shares its tables with Node B1and C1. Node C1 shares its tables with Node B1and D1 and A1. Node B1 shares its tables with Node A1and C1. Node D1 shares its tables with Node C1.

- **Updating**

A speaker node in the autonomous system updates its own routing table when it receives a two-column routing table from a neighbor on the different autonomous system. The table is updated by adding the nodes that are not in its routing table and adding its own autonomous system and autonomous system that sent the table. After completing the update process, each speaker node has a routing table that contains paths to

NOTES

reach nodes within the autonomous system and between different autonomous systems (Figure 4.76).



Dest.	Path
A1	AS1
...	
A5	AS1
B1	AS1-AS2
...	...
B4	AS1-AS2
C1	AS1-AS3
...	...
C3	AS1-AS3
D1	AS1-AS2-AS4
...	...
D4	AS1-AS2-AS4

A1 Table

Dest.	Path
A1	AS2-AS1
...	
A5	AS2-AS1
B1	AS2
...	...
B4	AS2
C1	AS2-AS3
...	...
C3	AS2-AS3
D1	AS2-AS3-AS4
...	...
D4	AS2-AS3-AS4

B1 Table

Dest.	Path
A1	AS3-AS1
...	
A5	AS3-AS1
B1	AS3-AS2
...	...
B4	AS3-AS2
C1	AS3
...	...
C3	AS3
D1	AS3-AS4
...	...
D4	AS3-AS4

C1 Table

Dest.	Path
A1	AS4-AS3-AS1
...	
A5	AS4-AS3-AS1
B1	AS4-AS3-AS2
...	...
B4	AS4-AS3-AS2
C1	AS4-AS3
...	...
C3	AS4-AS3
D1	AS4
...	...
D4	AS4

D1 Table

Fig. 4.76 *Updation in Path Vector Routing*

If router A1 receives a packet for node A3, it knows that the path is in AS1; but if it receives a packet for D1, it knows the packet has to go from AS1, to AS2 and then to AS3. If node D1 in AS4 receives a packet for node A2, it knows it should go through AS4, AS3 and AS1.

The major advantage of a path vector protocol is its flexibility. There are several other advantages of using a path vector protocol.

- Path Vector Protocol involves less complexity in computation than that of the Link State Protocol. To compute a path vector, a newly arrived route is evaluated and compared to the existing one, while to compute a link state, it is necessary to execute an SPF algorithm.
- In Path vector routing, it is not essential for all routing domains to have homogeneous policies for route selection. A routing domain may not be aware of the route selection policies used by other routing domains. The support for heterogeneous route selection policies has serious implications for computational complexity. In path vector protocol, each domain is allowed to select routes independently. This route selection is based only on local policies. However, little additional cost is incurred to accommodate heterogeneous route selection.
- It is more efficient because only the domains whose routes are affected by the changes have to recompute.
- Suppression of routing loops is implemented through the path attribute, in contrast to link state and distance vector, which use a globally-defined monotonically, thereby, increasing, metric for route selection. Therefore, different confederation definitions are accommodated because looping is avoided by the use of full path information.
- In Path vector routing, information is distributed after computation of route. Consequently, only routing information associated with the routes selected by a domain is distributed to adjacent domains.
- In Path vector routing, information can be selectively hidden.

The major disadvantages of path vector routing are:

1. When the network topology changes, only those routes which are affected by the changes are recomputed. This is more efficient than complete recomputation. However, because full path information is included in each

distance vector, the effect of a change in network topology can propagate farther than in traditional distance vector algorithms.

2. Unless the network topology is fully meshed or is able to appear so, routing loops can become an issue.

BGP is a popular example of a path vector routing protocol.

6. **Hybrid Routing**

The last category of routing protocols is hybrid protocols. In hybrid routing protocols, there is an attempt to combine the positive attributes of both distance vector and link state protocols. Hybrid protocols use metrics to assign a preference to a route as done in distance vector. However, the metrics are more precise than conventional distance vector protocols. In hybrid protocols, as in link state algorithms, routing updates are not periodic. The routes are updated only when the network topology changes. Networks that use hybrid protocols are likely to converge more quickly than networks using distance vector protocols. Finally, these protocols potentially reduce the costs of link state updates and distance vector advertisements.

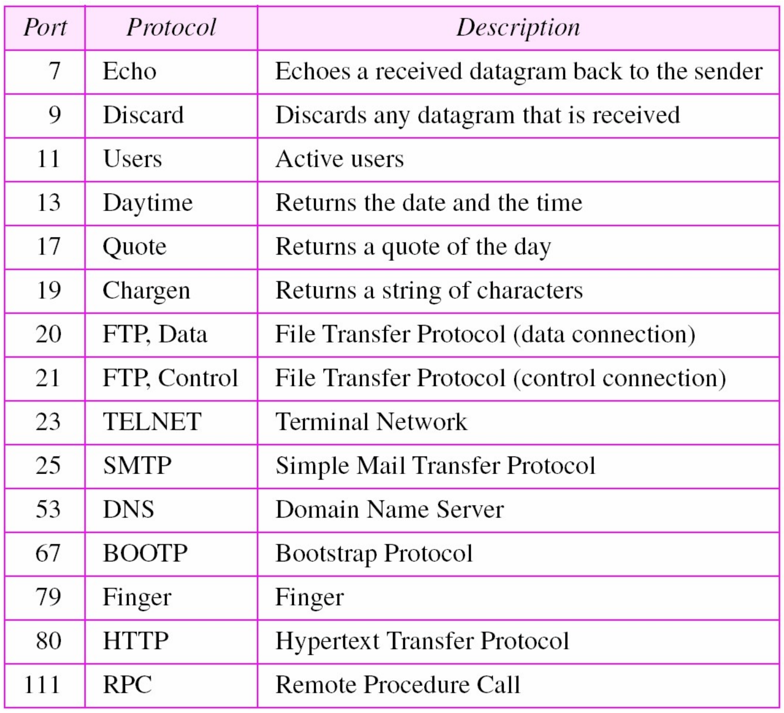
Although open hybrid protocols exist, this category is almost exclusively associated with the proprietary EIGRP algorithm developed by Cisco Systems, Inc.

4.7 TCP SERVICES AND APPLICATIONS

1. Process-to-process communication

TCP provides process-to-process communication using port numbers.

Table 4.1 *Ports Commonly used by TCP*



Port	*Protocol*	*Description*
7	Echo	Echoes a received datagram back to the sender
9	Discard	Discards any datagram that is received
11	Users	Active users
13	Daytime	Returns the date and the time
17	Quote	Returns a quote of the day
19	Chargen	Returns a string of characters
20	FTP, Data	File Transfer Protocol (data connection)
21	FTP, Control	File Transfer Protocol (control connection)
23	TELNET	Terminal Network
25	SMTP	Simple Mail Transfer Protocol
53	DNS	Domain Name Server
67	BOOTP	Bootstrap Protocol
79	Finger	Finger
80	HTTP	Hypertext Transfer Protocol
111	RPC	Remote Procedure Call

NOTES

2. Stream delivery service

TCP is a connection-oriented protocol that is accountable for realiable communication between end to end processes. It allows the sending process to deliver data as a stream of bytes and allows the receiving process to obtain data as a stream of bytes. TCP creates an environment in which two processes can be connected by an imaginary tube. The sending process produces (writes to) the stream of bytes and the receiving process consumes (reads from) them. The delivery process of the data would fail if the connection is not made or the connection is terminated on the either end (Figure 4.77).

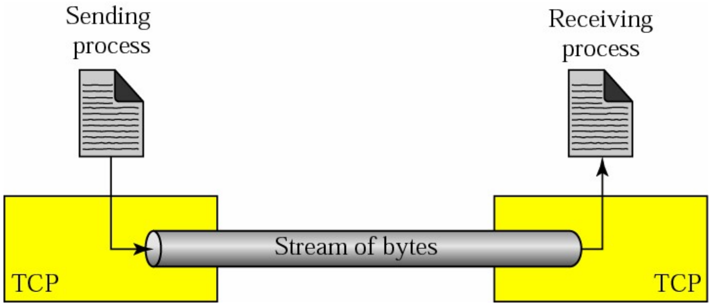


Fig. 4.77 Stream Delivery Service

Sending and receiving buffersThe sending and receiving processes may read or write data at varying speed. To counter this TCP needs buffer for storage, there are two buffers, the sending buffer and he receiving buffer, one for each direction. One way to implement a buffer is to use a circular array of 1-byte locations (Figure 4.78).

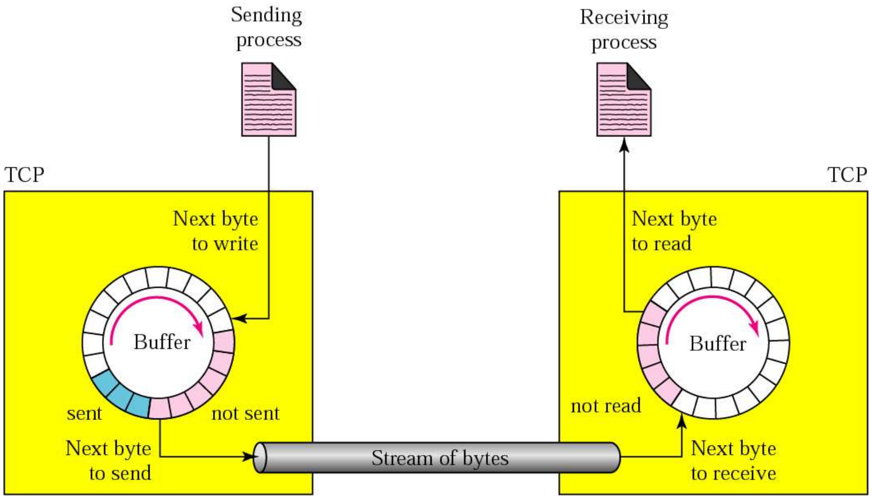


Fig. 4.78 Sending and Receiving Buffers

3. Segments

The IP layer, as a service provider for TCP, needs to send data in packets, not as stream of bytes. At the transport layer, TCP groups a number of bytes together into a packet called a segment. TCP adds a header to each segment and delivers the segment to the IP layer for transmission. The segments are encapsulated in IP datagrams and transmitted (Figure 4.79).

NOTES

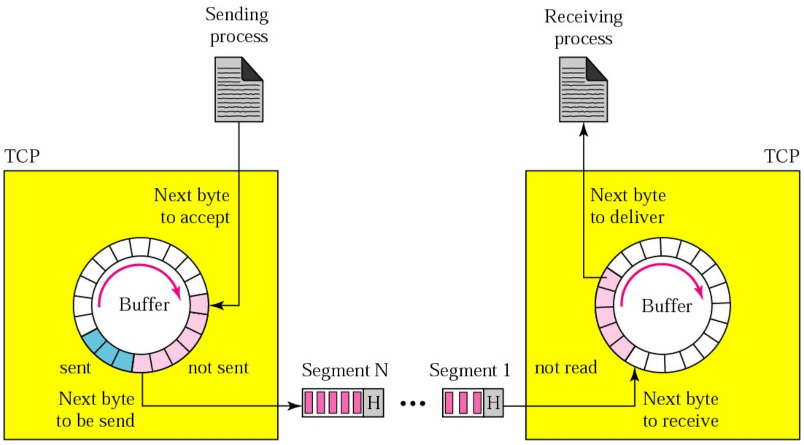


Fig. 4.79 Segments

4. Full duplex communication

TCP provides for concurrent data streams in both directions. Each TCP has sending and receiving buffer and segments move bi-directionally.

5. Connection-oriented service

TCP is a connection-oriented protocol. When a process at site A wants to send and receive data from another process at site B, a connection is established between the two and subsequently, data is exchanged in both directions and the connection is terminated. This is a virtual connection not a physical connection.

6. Reliable service

TCP guarantees reliable services. TCP assigns a sequence number to each byte transmitted, and expects a positive acknowledgment (ACK) from the receiving TCP layer. If the ACK is not received within a timeout interval, the data is retransmitted. Because the data is transmitted in blocks (TCP segments), only the sequence number of the first data byte in the segment is sent to the destination host.

4.7.1 TCP Features

To acknowledge safe delivery of packets, TCP provides well-organized, efficient and responsible mechanisms. It implements the following features to ensure the same.

1. Numbering system

- **Byte number**: TCP numbers the bytes of data being transferred to each connection. Numbering is independent in each direction. The numbering starts with a randomly generated number. A random number between 0 and $2^{32.}$ -1 is the number of the first byte.
- **Sequence number**: After the bytes have been numbered, TCP assigns a sequence number to each segment that is being sent. The value in the sequence number field of a segment defines the number of the first data byte contained in that segment.
- **Acknowledgement number**: The value of the acknowledgment field in a segment defines the number of the next byte a recipient expects to receive. The acknowledgment number is cumulative, which implies that the recipient

NOTES

takes the number of last byte it has received, safe and sound. Then the value is changed by adding one to it, and announces this as the acknowledgement number.

2. Flow control

Flow control is one of the important paraphernalias that TCP provides. The receiver of the data controls the amount of data that is being sent by the sender. This is done to prevent the receiver from being overwhelmed with data. In this manner, the sender restricts the data transmission to the recipient. The numbering system allows TCP to use a byte-oriented flow control.

3. Error control

To provide reliable service, TCP implements error control mechanism. This also ensure the authenticity and integrity of the receiving data. Error-control is byte oriented.

4. Congestion control

TCP takes into account congestion in network. The amount of data sent by a sender is not only controlled by the receiver, but is also determined by the level of congestion in the network.

4.7.2 TCP Segment

A packet in a TCP is called a segment (Figure 4.80).

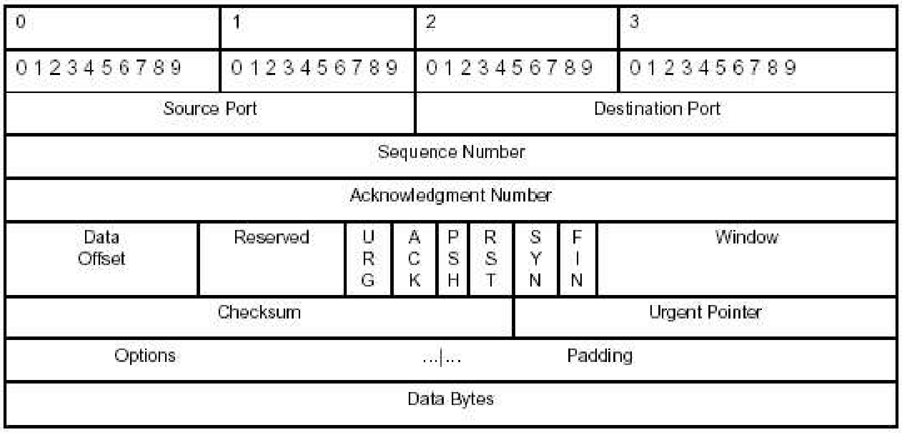


Fig. 4.80 *TCP Segment*

The various constituting terms are defined as follows:

Source port The 16-bit source port number, used by the receiver to reply.

Destination port The 16-bit destination port number.

Sequence number The sequence number of the first data byte in this segment. If the SYN control bit is set, the sequence number is the initial sequence number (n) and the first data byte is $n + 1$.

Acknowledgment number If the ACK control bit is set, this field contains the value of the next sequence number that the receiver is expecting to receive.

NOTES

Data offset The number of 32-bit words in the TCP header. It indicates where the data begins.

Reserved The six bits reserved for future use; must be zero.

URG This indicates that the urgent pointer field is significant in this segment.

ACK This indicates that the acknowledgment field is significant in this segment.

PSH Push function.

RST Resets the connection.

SYN Synchronizes the sequence numbers.

FIN No more data from the sender.

Window Used in ACK segments. It specifies the number of data bytes, beginning with the one indicated in the acknowledgment number field that the receiver (the sender of this segment) is willing to accept.

Checksum The 16-bit one's complement of the one's complement sum of all 16-bit words in a pseudo-header, the TCP header, and the TCP data. While computing the checksum, the checksum field itself is considered zero.

Urgent pointer Points to the first data octet following the urgent data. Only significant when the URG control bit is set.

Options Just as in the case of IP datagram options, options can be either:

– a single byte containing the option number;a variable length option.

4.7.3 A TCP Connection

TCP stands for Transmission Control Protocol. It is connection-oriented. A connection-oriented transport protocol establishes a virtual path between the source and destination. All of the segments belonging to a message are then sent over this virtual path. A connection-oriented transmission requires three phases: connection establishment, data transfer and connection termination.

1. Connection establishment

The connection establishment in TCP is called three-way handshaking (Figure 4.81).

The process starts with the server. The server program tells its TCP that it is ready to accept the connection. This is called a request for passive open. Although the server TCP is ready to accept any connection from any machine in the world, it cannot make this connection itself.

The client program issues a request for an active open. A client that wishes to connect to an open server tells its TCP that it needs to be connected to that particular server. TCP can now start the three-way handshaking process.

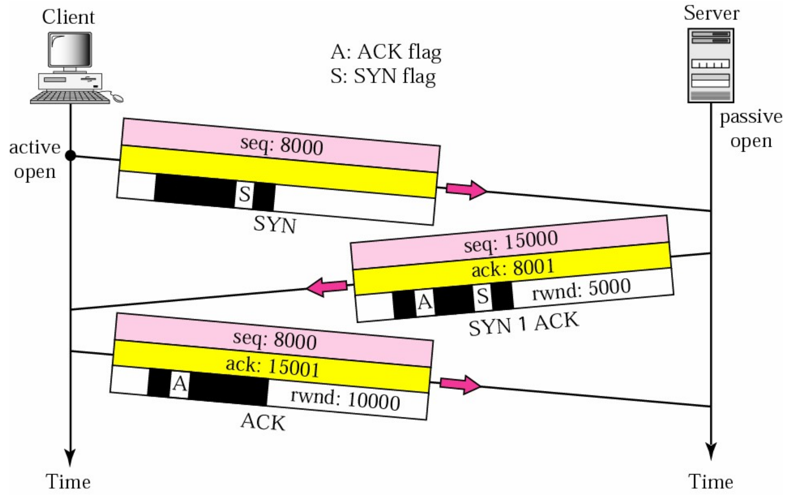


Fig. 4.81 Three-way Handshaking

- A SYN segment cannot carry data, but it consumes one sequence number.
- A SYN + ACK segment cannot carry data, but does consume one sequence number.
- An ACK segment, if carrying no data, consumes no sequence number.

2. Data transfer

After connection is established, bi-directional data transfer can take place. The client and server can both send data and acknowledgements. The sending TCP uses a buffer to store the stream of data coming from the sending application program (Figure 4.82). The sending TCP can select the segment size. The receiving TCP also buffers the data when they arrive and delivers them to the application program when the program is ready.

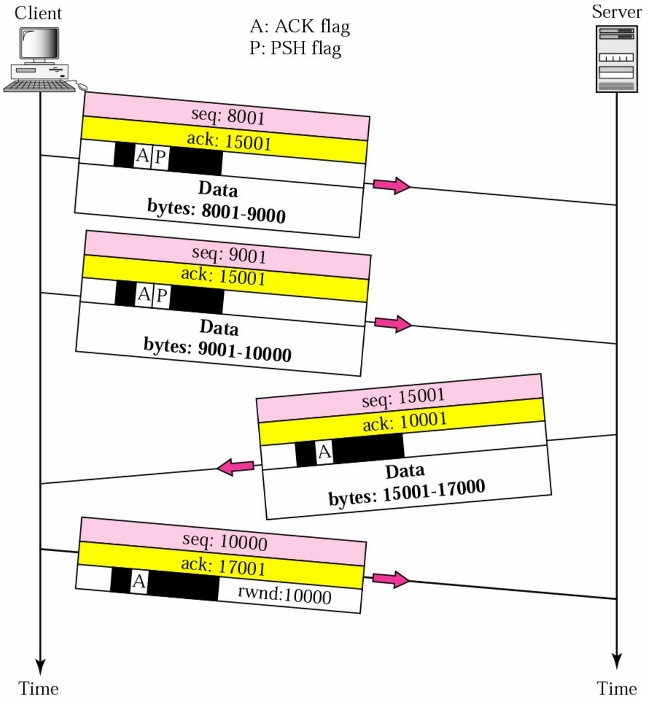


Fig. 4.82 Data Transfer

NOTES

3. Connection termination

Any of the two parties involved in exchanging data can close the connection, although the client initiates it (Figure 4.83).

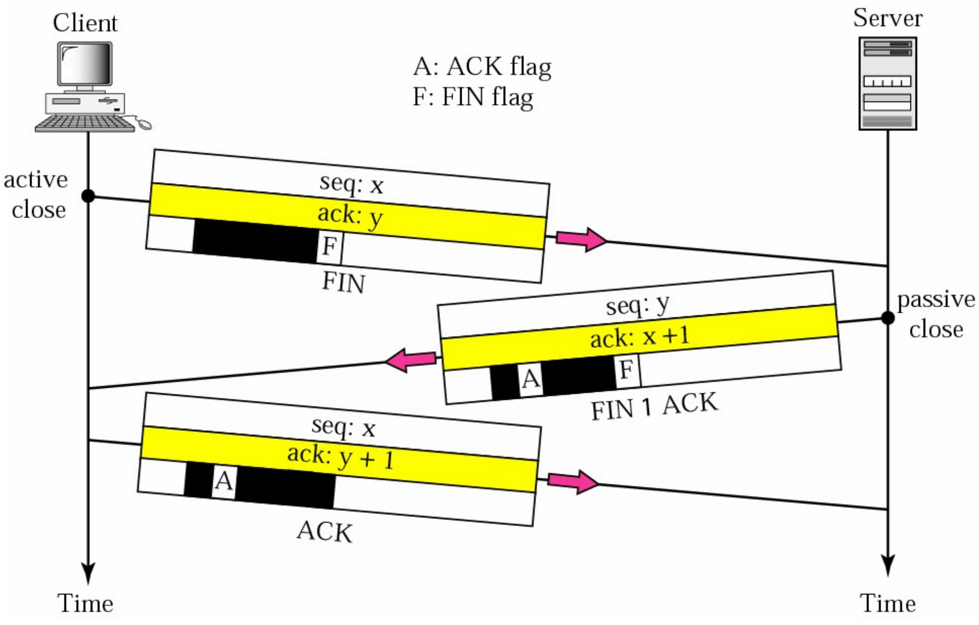


Fig. 4.83 *Connection Termination*

Closing the connection is done implicitly by sending a TCP segment with the FIN bit (no more data) set. Because the connection is full duplex (that is, there are two independent data streams, one in each direction), the FIN segment only closes the data transfer in one direction. The other process will now send the remaining data, it still has to transmit and also end with a TCP segment where the FIN bit is set. The connection is deleted (status information on both sides) after the data stream is closed in both directions.

- The FIN segment consumes one sequence number if it does not carry data.
- The FIN + ACK segment consumes one sequence number if it does not carry data.

4. Half close

In TCP, one end can stop sending data while still receiving data. This is called half close.

In Figure 4.84, the client half-closes the connection by sending a FIN segment in the middle of data transmission. The server accepts the half close by sending the ACK segment. The data transfer from server to client stops. The server, however, can still send data. When the server has sent all the processed data, it sends a FIN segment, which is acknowledged by an ACK from the client.

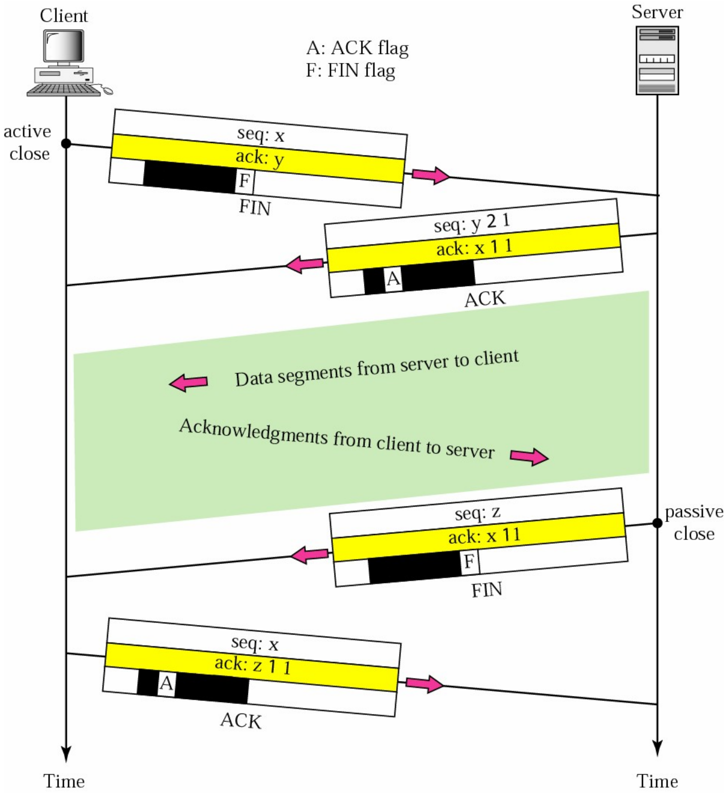


Fig. 4.84 *Half Close*

4.7.4 State Transition Diagram

Table 4.2 *States of TCP*

State	*Description*
CLOSED	There is no connection
LISTEN	Passive open received; waiting for SYN
SYN-SENT	SYN sent; waiting for ACK
SYN-RCVD	SYN+ACK sent; waiting for ACK
ESTABLISHED	Connection established; data transfer in progress
FIN-WAIT-1	First FIN sent; waiting for ACK
FIN-WAIT-2	ACK to first FIN received; waiting for second FIN
CLOSE-WAIT	First FIN received, ACK sent; waiting for application to close
TIME-WAIT	Second FIN received, ACK sent; waiting for 2MSL time-out
LAST-ACK	Second FIN sent; waiting for ACK
CLOSING	Both sides have decided to close simultaneously

NOTES

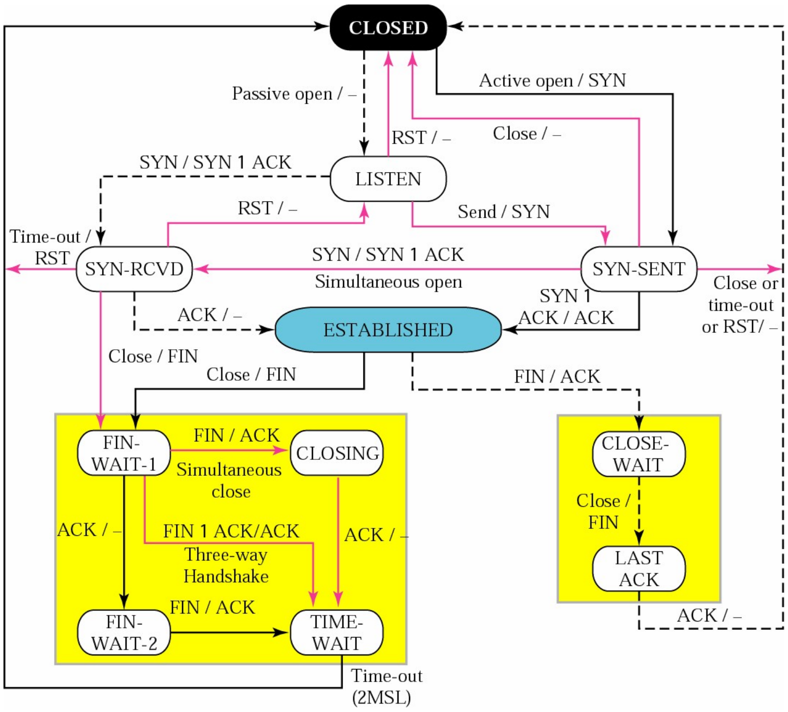


Fig. 4.85 *State Transition Diagram*

The two transitions leading to the ESTABLISHED state correspond to the opening of a connection, and the two transitions leading from the ESTABLISHED state are for the termination of a connection. The ESTABLISHED state is where the data transfer can occur between the two ends in both the directions.

If a connection is in the LISTEN state and a SYN segment arrives, the connection makes a transition to the SYN_RCVD state and takes the action of replying with an ACK+SYN segment. The client does an active open which causes its end of the connection to send a SYN segment to the server and to move to the SYN_SENT state. The arrival of the SYN+ACK segment causes the client to move to the ESTABLISHED state and to send an ack back to the server. When this ACK arrives the server finally moves to the ESTABLISHED state. In other words, we have just traced the THREE-WAY HANDSHAKE.

In the process of terminating a connection, the important thing to bear in mind is that the application process on both sides of the connection must independently close its half of the connection. Thus, on any one side, there are three combinations of transition that get a connection from the ESTABLISHED state to the CLOSED state:

- This side closes first:

  ESTABLISHED → FIN_WAIT_1 → FIN_WAIT_2 → TIME_WAIT → CLOSED.

- The other side closes first:

  ESTABLISHED → CLOSE_WAIT → LAST_ACK → CLOSED.

- Both sides close at the same time:

  ESTABLISHED → FIN_WAIT_1 → CLOSING → TIME_WAIT → CLOSED.

The main thing to recognize about connection teardown is that a connection in the TIME_WAIT state cannot move to the CLOSED state until it has waited for two times, the maximum amount of time an IP datagram might live in the Inter net. The reason for this is that while the local side of the connection has sent an ACK in response to the other side's FIN segment, it does not know that the ACK was successfully delivered. As a consequence, the other side might re-transmit its FIN segment, and this second FIN segment might be delayed in the network. If the connection were allowed to move directly to the CLOSED state, then another pair of application processes might come along and open the same connection, and the delayed FIN segment from the earlier incarnation of the connection would immediately initiate the termination of the later incarnation of that connection.

4.7.5 Flow Control

TCP carefully keeps track of the data it sends and what happens to it. This management of data is required to facilitate two key requirements of the protocol:

- **Reliability**: Ensuring that data that is sent, actually arrives at its destination, and if it fails, diagnozing and detecting the discrepancy and re-sends the data to the destination.
- **Data flow control**: Managing the rate at which data is sent, so that it does not overwhelm the receiving device.

To accomplish these tasks, the entire operation of the protocol is oriented around something called the *sliding window acknowledgement system*.

A basic technique for ensuring reliability in communications uses a rule that requires a device to send back an acknowledgement each time it successfully receives a transmission. If a transmission is not acknowledged after a period of time, the device is re-transmitted by its sender. This system is called *positive acknowledgement with retransmission* (*PAR*) (Figure 4.86). One drawback with this basic scheme is that the transmitter cannot send a second message until the first has been acknowledged.

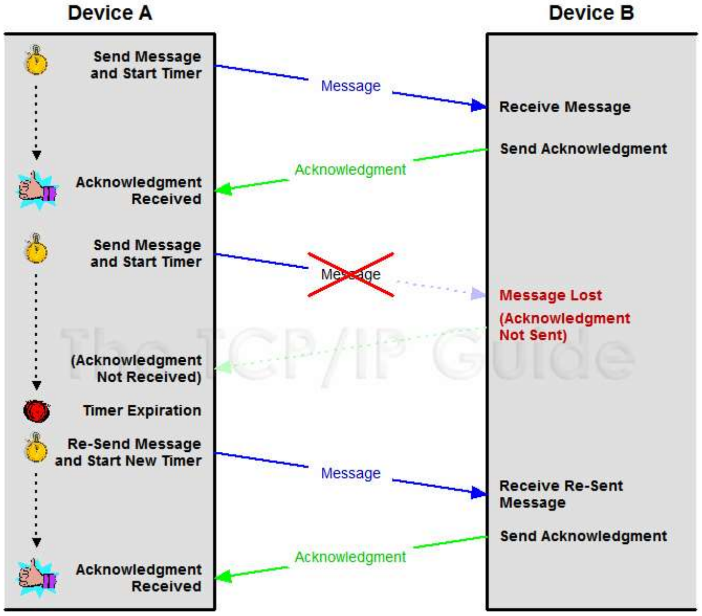


Fig. 4.86 *Positive Acknowledgement with Retransmission (PAR)*

The basic PAR reliability scheme can be enhanced by identifying each message to be sent, so multiple messages can be in transit at once. The use of a *send limit* allows the mechanism to also provide flow control capabilities, by allowing each device to control the rate at which the data is sent.

The TCP sliding window system is a variation on the enhanced PAR system, with changes made to support TCP's *stream orientation*. Each device keeps a track of the status of the byte stream it needs to transmit by dividing them into four conceptual categories:

1. **Bytes sent and acknowledged**: The earliest bytes in the stream will have been sent and acknowledged. These are basically 'accomplished' from the standpoint of the device sending data. For example, let's suppose that 31 bytes of data have already been sent and acknowledged. These would fall into Category #1.
2. **Bytes sent but not yet acknowledged**: These are certain bytes that the device has sent but for which it has not yet received an acknowledgment. The sender cannot consider these "accomplished" until they are acknowledged. Let's say there are 14 bytes here, in Category #2.
3. **Bytes not yet sent for which recipient is ready**: These are bytes that have not yet been sent, but for which the recipient has room, based on its most recent communication to the sender of how many bytes it is willing to handle at once. The sender will try to send these immediately (subject to certain algorithmic restrictions that we shall explore later). Suppose there are 6 bytes in Category #3.
4. **Bytes not yet sent for which recipient is not ready**: These are certain bytes further 'down the stream' which the sender is not yet allowed to send because the receiver is not ready. There are 44 bytes in Category #4.

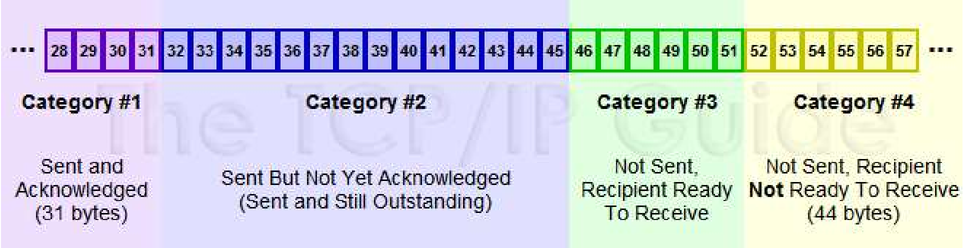

...
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
...
Category #1
Sent and Acknowledged (31 bytes)
Category #2
Sent But Not Yet Acknowledged (Sent and Still Outstanding)
Category #3
Not Sent, Recipient Ready To Receive
Category #4
Not Sent, Recipient Not Ready To Receive (44 bytes)

Fig. 4.87 *TCP Sliding Window System*

Once the devices are ready, it is now time for data transmission. The receiving device uses a similar parameters to differentiate between data received and acknowledged, not yet received but ready to receive, and not yet received and not yet ready to receive. In fact, both devices maintain a separate set of variables by segregating and keeping a track of the categories into which bytes fall in the stream; they are sending as well as the one they are receiving.

The sender and receiver must mutually agree on the sequence numbers to assign to the bytes in the stream. This is called *synchronization* and is done when the TCP connection is successfully established.

In our example, the byte ranges for the four categories are:

1. **Bytes sent and acknowledged**: Bytes 1–31.
2. **Bytes sent but not yet acknowledged**: Bytes 32–45.
3. **Bytes not yet sent for which recipient is ready**: Bytes 46–51.
4. **Bytes not yet sent for which recipient is not ready**: Bytes 52–95.

   - The Send Window and Usable Window

The key to the operation of the entire process is the number of bytes that the recipient is allowing the transmitter to have unacknowledged at one time. This is called the *send window*, or often, just the *window*. The window is what determines how many bytes the sender is allowed to transmit, and is equal to the sum of the number of bytes in Category #2 and Category #3. Thus, the dividing line between the last two categories (bytes not sent that recipient is ready for and ones it is *not* ready for) is determined by adding the window to the byte number of the first unacknowledged byte in the stream. In our example above, the first unacknowledged byte is #32. The total window size is 20.

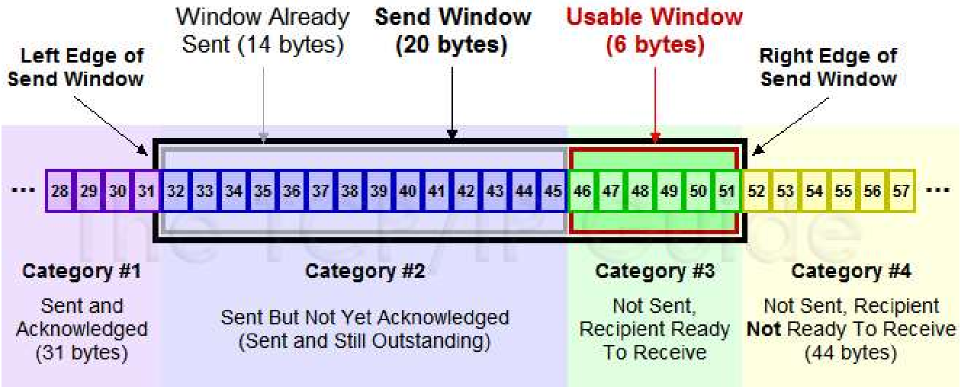

Window Already Sent (14 bytes)
Send Window (20 bytes)
Usable Window (6 bytes)
Left Edge of Send Window
Right Edge of Send Window
... 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 ...
Category #1
Sent and Acknowledged (31 bytes)
Category #2
Sent But Not Yet Acknowledged (Sent and Still Outstanding)
Category #3
Not Sent, Recipient Ready To Receive
Category #4
Not Sent, Recipient Not Ready To Receive (44 bytes)

Fig. 4.88 *The Send Window and Usable Window*

The term usable window is defined as the amount of data the transmitter is still allowed to send given the amount of data that is outstanding. It is, thus, exactly equal to the size of Category #3.

Let's suppose that in our example above, there is nothing stopping the sender from immediately transmitting the 6 bytes in the Category #3 (the usable window). When it does so, the 6 bytes will shift from Category #3 to Category #2. The byte ranges will now be as follows:

1. **Bytes sent and acknowledged**: Bytes 1–31.
2. **Bytes sent but not yet acknowledged**: Bytes 32–51.
3. **Bytes not yet sent for which recipient is ready**: None.
4. **Bytes not yet sent for which recipient is not ready**: Bytes 52–95.

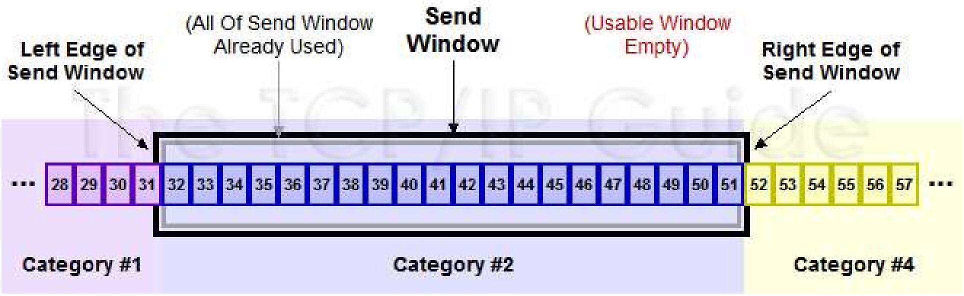


Fig. 4.89 *Sliding Window*

Some time later, the destination device sends back a message to the sender providing an acknowledgment. It will not specifically list out the bytes that have been acknowledged, because as we said before, doing this would be quite inefficient. Instead, it will acknowledge a *range* of bytes that represents the longest contiguous sequence of bytes received since the ones it had previously acknowledged.

For example, suppose the bytes already sent but not yet acknowledged at the start of the example (32–45) were transmitted in four different segments. These segments carried bytes 32–34, 35–36, 37–41 and 42–45, respectively. The first, second and fourth segments arrived, but the third did not. The receiver will send back an acknowledgement *only* for bytes 32–36 (32–34 and 35–36). It will hold bytes 42–45 but not acknowledge them, because this would imply receipt of bytes 37–41, which have not shown up yet. This is necessary because TCP is a *cumulative acknowledgement* system, which can only use a single number to acknowledge data, the number of the last contiguous byte in the stream successfully received. Let's also say the destination keeps the window size the same, at 20 bytes.

When the sending device receives this acknowledgment, it will be able to transfer some of the bytes from Category #2 to Category #1, since they have now been acknowledged. When it does so, something interesting will happen. Since five bytes have been acknowledged, and the window size did not change, the sender is allowed to send five more bytes. In effect, the window shifts, or *slides*, over to the right in the timeline. At the same time five bytes move from Category #2 to Category #1, five bytes move from Category #4 to Category #3, creating a new usable window for subsequent transmission. So, after receipt of the acknowledgement, the groups will look like this

1. **Bytes sent and acknowledged**: Bytes 1–36.
2. **Bytes sent but not yet acknowledged**: Bytes 37–51.
3. **Bytes not yet sent for which recipient is ready**: Bytes 52–56.
4. **Bytes not yet sent for which recipient is not ready**: Bytes 57–95.

For each time the data is transmitted, this process will take place as an acknowledgement is received, causing the window to slide across the entire stream to be transmitted. This is the TCP *sliding window* acknowledgement system. It is a very powerful technique, which allows TCP to easily acknowledge an arbitrary number of bytes using a single acknowledgement number, thus providing reliability to the byte-oriented protocol without spending time on an excessive number of acknowledgements. For simplicity, the example above leaves the window size constant, but in reality, it can be adjusted to allow a recipient to control the rate at which the data is sent, to enable efficient flow control and congestion handling during each transmission.

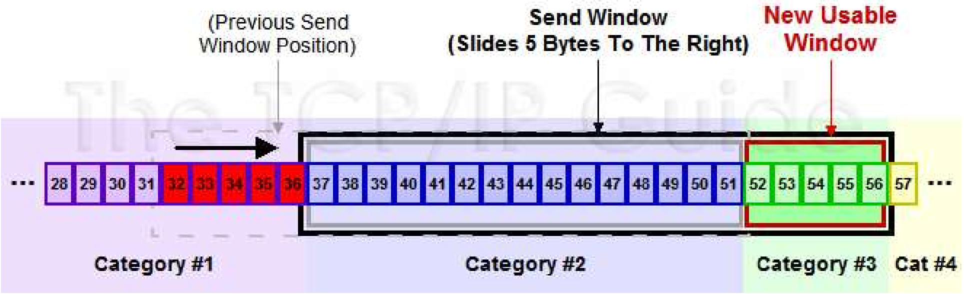


Fig. 4.90 *Send Window Sliding to the Right*

When a device gets an acknowledgement for a range of bytes, it knows that they have been successfully received by their destination. It moves them from the 'sent but unacknowledged' to the 'sent and acknowledged' category. This causes the send window to *slide* to the right, allowing the device to send more data.

TCP acknowledgements are *cumulative*, and tell a transmitter that all the bytes up to the sequence number indicated in the acknowledgment were received successfully. Thus, if bytes are received out of order, they cannot be acknowledged until all the preceding bytes are received. TCP includes a method for timing transmissions and retransmitting lost segments, if necessary.

The TCP sliding window system is used not just for ensuring reliability through acknowledgments and retransmissions—it is also the basis for TCP's *flow control* mechanism. By increasing or reducing the size of it's receive window, a device can raise or lower the rate at which its connection partner sends it data. In the case where a device becomes extremely busy, it can even reduce the receive window to zero, closing it; this will halt any further transmissions of data until the window is reopened.

4.7.6 Control

TCP provides reliability using error control, which detects corrupted, lost, out-of-order, and duplicated segments. Error control in TCP lets the receiver check the integrity of the received packet. This is precisely achieved through the use of the checksum, acknowledgment, and time-out.

- Checksum

Each segment includes a checksum field for checking the corrupt segment each time the data transmission takes place. If the segment is corrupt, it is discarded by the destination TCP and is considered as lost. TCP uses a 16-bit checksum that is mandatory in every segment.

- Acknowledgement

TCP issues acknowledgements to confirm the receipt of data segments. Control segments that carry no data but consume a sequence number are also acknowledged. ACK segments are never acknowledged.

- Retransmission

When a segment gets lost, corrupted or delayed, it is re-transmitted to the destination. A segment is transmitted on two occasions: when a transmission expires or when the sender receives three duplicate ACKs. No re-transmission timer is set for an ACK segment.

Retransmission after RTO: TCP maintains one re-transmission time-out (RTO) timer for all outstanding (sent, but not acknowledged) segments. When the timer matures, the earliest outstanding segment is re-transmitted. The value of RTO is dynamic in TCP and is updated based on round-trip time (RTT) of segments. An RTT is the time needed for a segment to reach a destination and for an acknowledgement to be received by the sender.

- Out-of-order segments

Data may arrive out of order and be temporarily stored by the receiving TCP, but TCP guarantees that no out-of-order segment is delivered to the process.

1. Normal operation

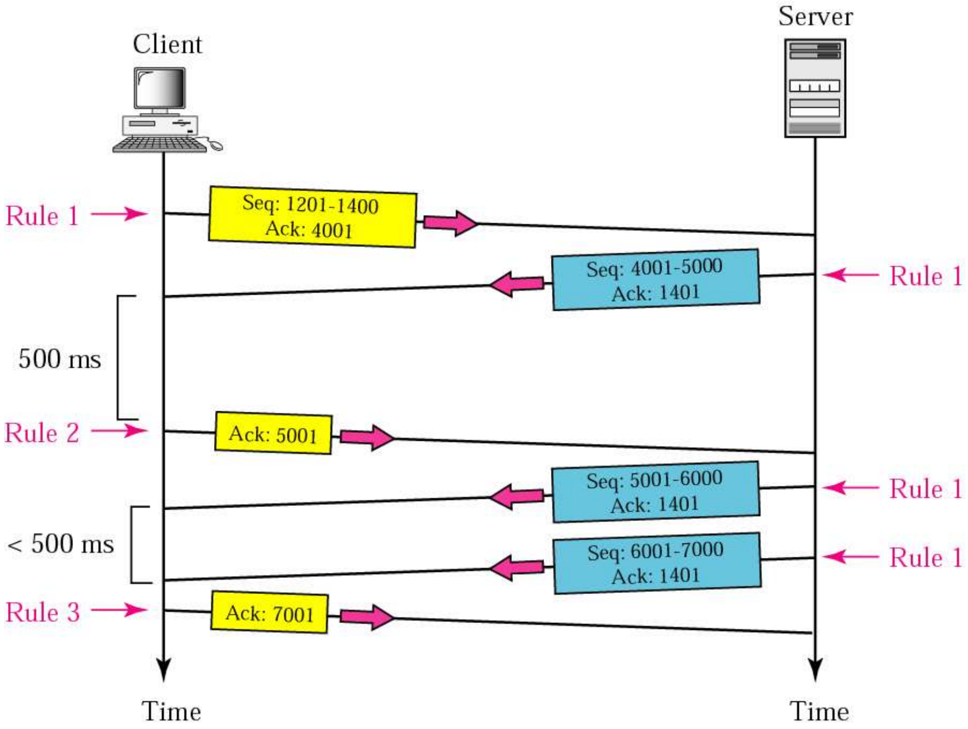


Fig. 4.91 *Normal Operation*

NOTES

2. Lost segment

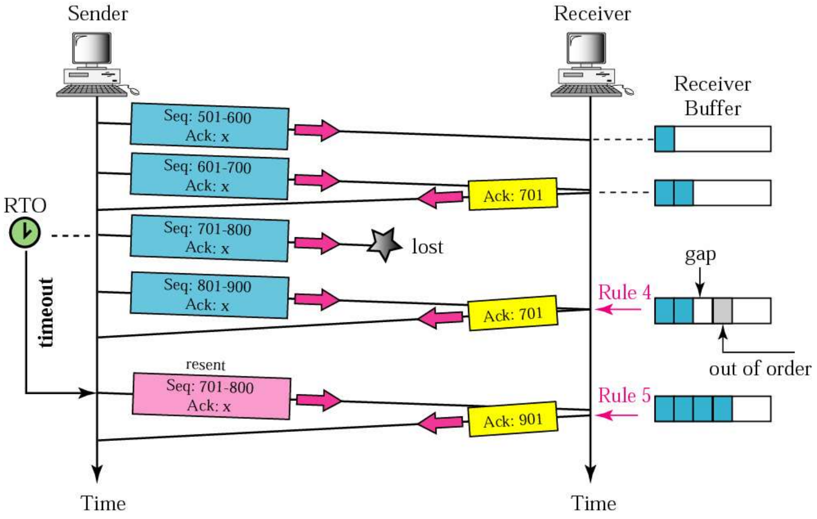


Fig. 4.92 *Lost Segment*

A lost segment and corrupted segment are regarded in the same perspective by the receiver. A lost segment is discarded somewhere in the network; a corrupted segment is discarded by the receiver. In the above example, the sender sends segments 1 and 2, which are acknowledged immediately by an ACK. Segment 3 is lost. The receiver receives segment 4, which is out of order. The receiver stores the data in the segment in its buffer but leaves a gap to indicate that there is no continuity in the data. The receiver immediately sends an acknowledgement to the sender, displaying the next byte it expects to be received. There is a timer for the earliest outstanding segment. When this timer matures, the sending TCP resends segment 3, which arrives and is acknowledged.

3. Fast retransmission

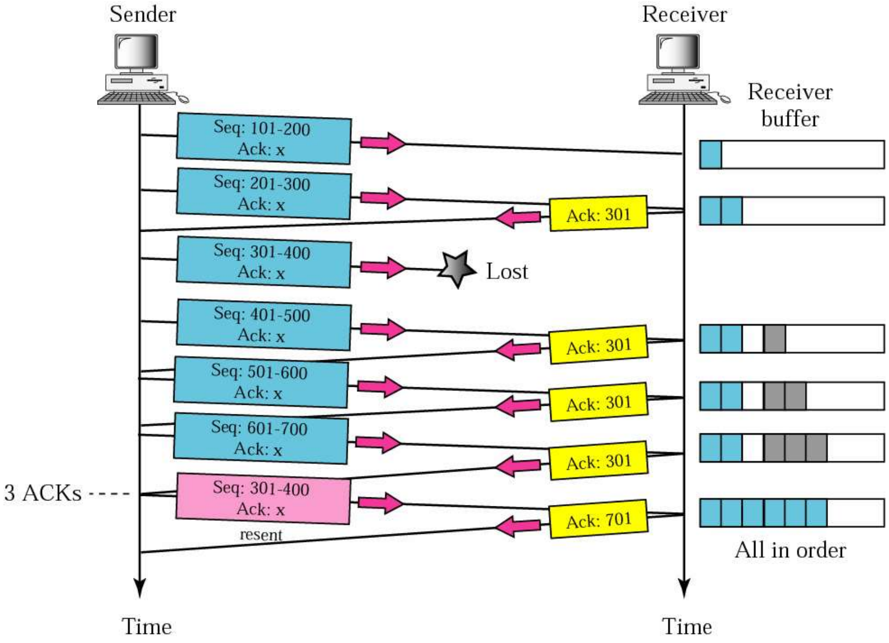


Fig. 4.93 *Fast Retransmission*

When the receiver receives the fourth, fifth and sixth segments, it triggers an acknowledgement. The sender receives four acknowledgements with the same value. Although the timer for segment 3 has not matured yet, the fast transmission requires that segment 3, the segment that is expected by all these acknowledgements, be resent immediately to all the destinations.

4. Lost acknowledgement

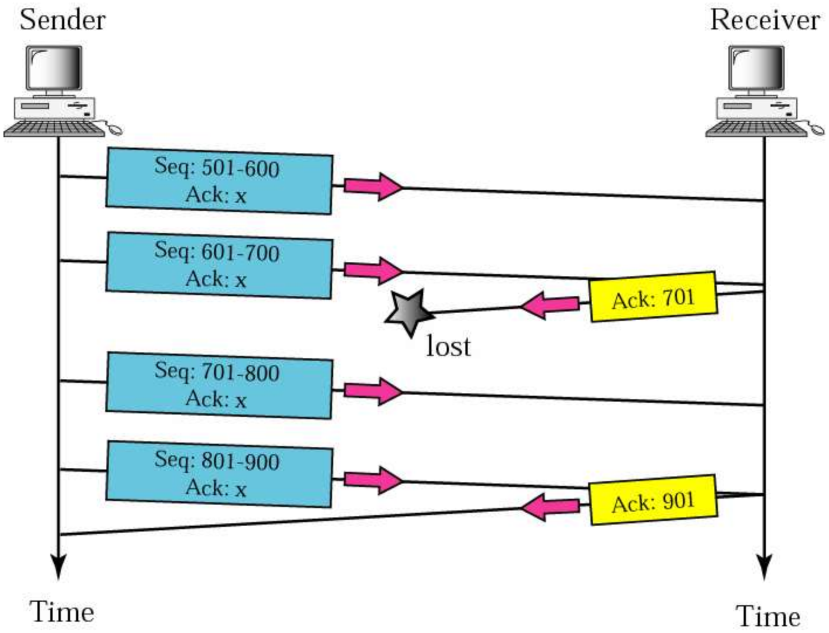


Fig. 4.94 *Lost Acknowledgement*

Lost acknowledgement corrected by resending a segment

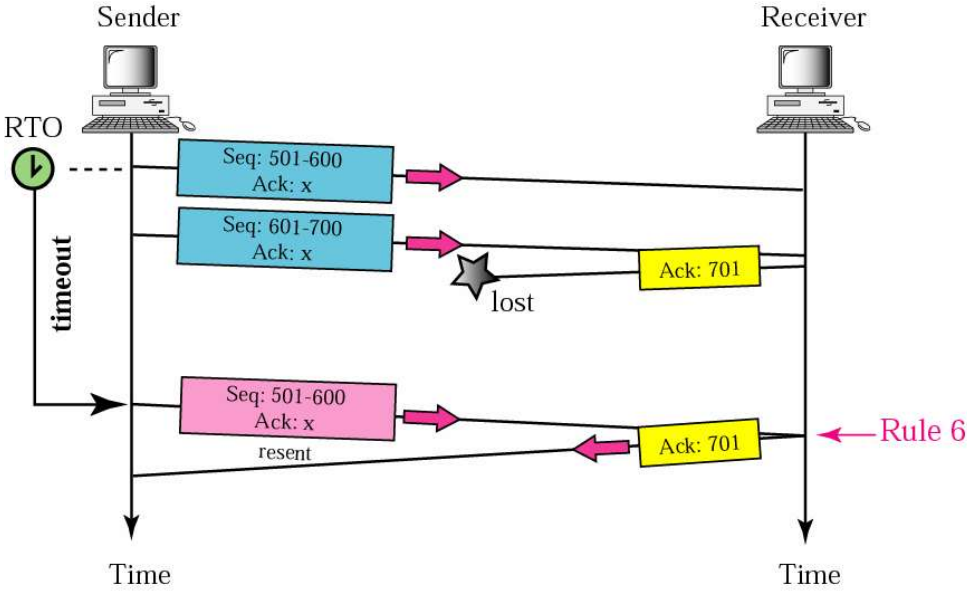


Fig. 4.95 *Deadlock*

Lost acknowledgements may create deadlock if they are not properly handled (Figure 4.95).

4.8 DOMAIN NAME SYSTEM (DNS)

NOTES

DNS is based on a hierarchical structure that enables same host names to be used unambiguously within different domains to simplify name space management. The name space describes the architecture of the names including rules for name creation, interpretation and the form of names. The Domain Name System (DNS) describes an architecture depending on domains or nodes. The domains are structured hierarchically according to their control of authority. The Internet is divided into more than 200 Top-Level Domains (TLDs). These domains are further partitioned into subdomains. The subdomains can be further partitioned and so on. Examples of TLD are countries like in, jp, us, ae, eg, etc. There are certain TLD, which comes under the category of generic TLD, and they are com, net, org, edu, int, etc.

The DNS hierarchical name architecture follows a directory structure and organizes names from most general types to most specific types. In this manner, the DNS name space allows names to be arranged into a hierarchy of domains looking as an inverted tree. In relation to computer terminology, it looks like the directory structure of a file system. Every standalone internetwork will define its own name space with unique hierarchical structure.

The domain name components are represented in English words separated by dots, for example, www.hotmail.com, www.yahoo.co.in, etc. Each name separated by dots is subdomains and are managed by a separate authorized server, for example, ".com" authorized server or servers are happened to manage all domains '*.com'.

The DNS name space defines an inverted tree type structures. Unlike real tree, the DNS tree grows from the top down. There are certain terminologies in relation to DNS tree that are defined below:

Root: The DNS tree grows from top to down, therefore, root occupies the top of the DNS name structure. However, it does not define any name and is considered null. The root domain is the parent of all the domains in the hierarchy.

Branch: It refers to any next closest part of DNS hierarchy and describes a domain with subdomains and objects within it. Like a real tree, all branches connect themselves to the root.

Leaf: Beneath the leaf, no object is defined and therefore it is an 'end object' in the structure. They are also referred as *interior nodes*, indicating that they occupy a position in the middle of the structure.

Top-Level Domains (TLDs): They come directly under the root of the tree and therefore referred as the highest-level domains. Other name is first-level domains. Similarly, the domains placed directly beneath the top-level domains are called the second level domains and so on. The TLDs are considered children domains. A peer at the same level in the hierarchy is known as sibling which defines that all the TLDs are siblings with root domain as the parent.

Subdomains: They are located directly below the second-level domains.

Conclusively, it may be understood that a domain is either a collection of objects, which represents a branch of the tree, or a specific leaf. **Thus, a** DNS name space is organized as a true topological tree with one parent only without any loops. The DNS name space is logical structure without having any relevance with physical locations of devices.

Naming in DNS

It involves DNS labels and label syntax rules in which each domain or node is described with a text label so as the domain may be identified within the structure. Syntax rules are:

Length: The character length may be of 0-63 characters. However, 1-20 character length is widely used.

Symbols: Name can be described with letters, numbers and the dash symbol ('–') only.

Case: They are not case-sensitive and lower and upper case for same label is equivalent. Every label needs to be unique within its parent domain but need not to be unique across domains.

Creating Domain Names

The individual domain within the domain name structure is uniquely created using the sequence of labels beginning from the root of the tree to the target domain from right to left separated by dots to provide a formal name to the domain. The root of the name space is defined with a zero-length or 'null' name by default. The DNS Name length is limited to 255 characters to describe a complete domain name.

The Domain Name may be either a Fully Qualified Domain Name (FQDN) or a Partially Qualified Domain Name (PQDN). A Fully Qualified Domain Name (FQDN) assigns full path of labels beginning from the root of the tree down to the target node to uniquely identify that node in the DNS name space. Unlike FQDN, the PQDN only describes a part of a domain name to provide a relative name for a particular context.

It is essential to have an authority structure to manage unique TLDs. Erstwhile the Network Information Center, now known, as the Internet Assigned Numbers Authority (IANA) is the central DNS authority for the Internet to create TLDs name. In some cases, IANA delegates their power for some of the TLDs to other organizations. Thus multiple authorities work in assigning and registering Domain Name. The authority for lower level domain is entrusted with the organization, which belongs to the second level domain. Conclusively, the DNS name space of the Internet is managed by several authorities arranged hierarchically in the similar manner as DNS name space.

Check Your Progress

11. What is an IP router?
12. How will you define an Autonomous System (AS)?
13. What is unicasting?
14. What is flow control?
15. What is the significance of Domain Name System (DNS)?

4.9 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. The major network infrastructure available in the country has two types of WAN:
    (i) Terrestrial WAN
    (ii) VSAT WAN
2. WAN is the acronym for Wide Area Network and refers to a network used to connect different equipment from remote areas. Normally, network services are provided by a Common Carrier of, for example, a telephone company.
3. RIP is the most widely used routing protocol of distance-vector type today. RIP has been originally designed based on the routing protocol applied to XNS and PUP protocol systems of Xerox (RFC1058).
4. Internal router is a router having its links directly connected to a network within a specific area. That is, internal router does not have any direct links to a network in another area.
5. The Internet Protocol (IP) is one of the most dominant protocols of the TCP/IP protocol suite and its main protocol is located at the network layer. The fundamental job of network layer is concerned with the delivery of data, from the source to the destination, between devices that may be on different networks.
6. When an IP datagram is sent to a destination on the same local network, it is called *direct delivery*.
7. IP datagram is the rudimentary unit of information carried in the form of a packet in the IP layer, containing a source and destination address. This information is communicated across the network using Internet Protocol.
8. The Internet address or IP address is a unique identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet.
9. In order to do subnetting, the host ID is split into subnet ID and host ID. In doing this, the size of the host ID part of the address gets reduced. In short, bits from the host ID are being 'stolen' to use for the subnet ID. Class A networks have 24 bits to split between the subnet ID and host ID: Class B networks have 16, and Class C networks only 8
10. The *subnet mask* is a 32-bit binary number that comes with an IP address. Like IP addresses, they are usually converted to dotted decimal notation for convenience. It is created in a way that it has a one bit for each corresponding bit of the IP address that is part of its network ID or subnet ID, and a zero for each bit of the IP address's host ID.
11. An IP router is a device that attaches to two or more physical networks and transfers datagrams between the networks.

12. An Autonomous System (AS) is defined as a logical portion of a larger IP network (Figure 4.68). An AS normally consists of an internetwork within an organization. It is administered by a single management authority.
13. Communication between one source and one destination is known as unicast. There is a one to one relationship between source and destination. In this type of communication, both the source and destination addresses are unicast addresses.
14. Flow control is one of the important paraphernalias that TCP provides. The receiver of the data controls the amount of data that is being sent by the sender. This is done to prevent the receiver from being overwhelmed with data.
15. DNS is based on a hierarchical structure that enables same host names to be used unambiguously within different domains to simplify name space management. The name space describes the architecture of the names including rules for name creation, interpretation and the form of names.

4.10 SUMMARY

- Wide Area Networks (WANs) connect larger geographic areas, such as New Delhi, India or the world.
- Packet switching technologies such as asynchronous transfer mode (ATM), frame relay, switched multimegabit data service (SMDS), and X.25 are used to implement WAN along with statistical multiplexing to enable devices to share these circuits.
- WAN is the acronym for Wide Area Network and refers to a network used to connect different equipment from remote areas.
- LANs can be extended to a wider area but it cannot be extended arbitrarily far or to handle arbitrarily many computers.
- WAN is composed a number of autonomous computer that are distributed over a large geographical area.
- The major objective of network design is to select the network service and to determine the transmission speed for the system.
- The routing is simple if the destination address of the host is in the same subnet.
- If a packet sent from the source host is addressed to a host in the same subnet, the IP layer of the source host must obtain the MAC address of the destination host.
- Default gateway consists of manually registering router IP addresses in the host.
- A set of networks interconnected by routers within a specific area using the same routing protocol is called domain.
- The TCP/IP reference model is a network model used in Internet architecture. It has its beginnings back in the 1960s.

NOTES

- The Internet Protocol (IP) is one of the most dominant protocols of the TCP/IP protocol suite and its main protocol is located at the network layer.
- The fundamental job of network layer is concerned with the delivery of data, from the source to the destination, between devices that may be on different networks.
- The primary job of IP functions is to add and manage IP addresses. With this as foundation, let us take a close look at four of its major functions.
- The Internet address or IP address is a unique identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet.
- IP addresses, when started a few decades ago, were based on the concept of classes. Each class fixed a boundary between the network prefix and the host number at a different point within the 32-bit address.
- Classful addressing resulted in efficient use of address space. The Class B address block contains a very large number of addresses (65,534) but a Class C block has only a relatively small number (254).
- In subnetting, it is necessary to communicate, which bits are for the subnet ID and which for the host ID, to devices that interpret IP addresses. A 32-bit binary number which provides this information to devices handling IP addresses is called a subnet mask.
- On a single physical network, individual devices are identified in the network by their physical hardware address.
- Internetworking involves connecting different physical networks. Providing connections between dissimilar networks is one of the basic functions provided by the IP. A system that performs this function is called an IP router. An IP router is a device that attaches to two or more physical networks and transfers datagrams between the network
- In hybrid routing protocols, there is an attempt to combine the positive attributes of both distance vector and link state protocols.
- Transmission Control Protocol (TCP) is a connection-oriented, reliable transport protocol for process-to-process delivery and end to end communication.
- To provide reliable service, TCP implements error control mechanism. This also ensure the authenticity and integrity of the receiving data. Error-control is byte oriented.
- DNS is based on a hierarchical structure that enables same host names to be used unambiguously within different domains to simplify name space management.
- The domain name components are represented in English words separated by dots, for example, www.hotmail.com, www.yahoo.co.in, etc.
- The individual domain within the domain name structure is uniquely created using the sequence of labels beginning from the root of the tree to the target domain from right to left separated by dots to provide a formal name to the domain

NOTES

4.11 KEY TERMS

- **WAN**: It is the acronym for Wide Area Network and refers to a network used to connect different equipment from remote areas.
- **Data Routing**: It refers to the process of selecting the shortest and the most reliable path intelligently over which to send data to its ultimate destination.
- **Static Routing**: This method uses fixed definitions representing paths through the network.
- **Dynamic Routing**: In dynamic routing, algorithms routers can automatically determine and maintain knowledge of the paths through the network.
- **IP datagram**: It is the rudimentary unit of information carried in the form of a packet in the IP layer.
- **Internet Address**: It is a unique identifier used in the IP layer of the TCP/IP protocol suite to identify each device connected to the Internet.
- **Subnet Mask**: It is a 32-bit binary number that comes with an IP address. Like IP addresses, they are usually converted to dotted decimal notation for convenience

4.12 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Write the difference between LAN and WAN.
2. What is distance vector protocol. What are the problems with it?
3. Write the various goals of TCP/IP.
4. State the characteristics of IP.
5. What is classful addressing?
6. What are the limitations of classful addressing?
7. How will you decide that how many subnet bits are required in subnetting? Discuss with the help of an example.
8. In which circumstances static routing is useful?
9. Write two key requirements of TCP.
10. What is DNS tree?
11. Write the process of naming in DNS.
12. How will you create a domain name?
13. Explain the concept of IP addresssing.

Long-Answer Questions

1. Explain router protocols in details.
2. Describe TCP/IP protocols in detail.
3. Explain the concept of IP addressing.
4. How many classes of IP addresses are there? Explain each class in detail.
5. Describe classless addressing in detail. Also, analyse the benefits of classless addressing.
6. Explain the working of ARP.
7. Explain the types of IP routing algorithms.
8. Write a detailed note on Domain Name System (DNS).

NOTES

4.13 FURTHER READING

Forouzan, Behrouz A. *Data Communications and Networking*. New Delhi: Tata McGraw-Hill, 2004.

Stallings, William and Richard Van Slyke. *Business Data Communications*. New Jersey: Prentice Hall, 1998.

Black, Uyless. *Computer Networks*. New Jersey: Prentice Hall, 1993.

Stallings, William. *Data and Computer Communications*. New Jersey: Prentice Hall, 1996.

Tanenbaum, Andrew S. *Computer Networks*. New Jersey: Prentice Hall PTR, 2002.

Stallings, William. *Data and Computer Communications*. NJ: Prentice-Hal, 1996.

UNIT 5 BROADBAND NETWORK AND INTERNET SERVICES

NOTES

Structure

5.0 INTRODUCTION

Broadband is the transmission of high-quality data of a wide bandwidth. Broadband connections include Wi-Fi, DSLS, fiber, and satellites. The term broadband commonly refers to high-speed Internet access that is always on and faster than the traditional dial-up access. Broadband includes various high-speed transmission technologies. DSL is a wireline transmission technology that transmits data faster over traditional copper telephone lines already installed to homes and businesses. DSL-based broadband provides transmission speeds ranging from several hundred Kbps to millions of bits per second (Mbps). Asymmetrical Digital Subscriber Line (ADSL) is used primarily by residential customers, such as Internet surfers, who receive a lot of data but do not send much. ADSL typically provides faster speed in the downstream direction than the upstream direction. Symmetrical Digital

Subscriber Line (SDSL) is used typically by businesses for services such as video conferencing, which need significant bandwidth both upstream and downstream.

Electronic mail (email or e-mail) is a method of exchanging messages ("mail") between people using electronic devices. Electronic mail (e-mail) is a computer-based application for the exchange of messages between users. A worldwide e-mail network allows people to exchange e-mail messages very quickly. E-mail is the electronic equivalent of a letter, but with advantages in timeliness and flexibility. While a letter will take from one day to a couple of weeks to be delivered, an e-mail is delivered to the intended recipient's mailbox almost instantaneously, usually in the multiple-second to subminute range. Internet Services allows us to access huge amount of information such as text, graphics, sound and software over the internet. Following diagram shows the four different categories of Internet Services.

E-commerce (Electronic Commerce) is the activity of electronically buying or selling of products on online services or over the Internet. E-commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems.

EDI (Electronic Data Interchange) is an electronic way of transferring business documents in an organization internally, between its various departments or externally with suppliers, customers, or any subsidiaries. In EDI, paper documents are replaced with electronic documents such as word documents, spreadsheets, etc.Many business documents can be exchanged using EDI, but the two most common are purchase orders and invoices. At a minimum, EDI replaces the mail preparation and handling associated with traditional business communication. However, the real power of EDI is that it standardizes the information communicated in business documents, which makes possible a "paperless" exchange.

In this unit, you will learn about broadband network, local loop technologies, Asymmetric Digital Subscriber Line (ADSL), High Bit-Rate Digital Subscriber Line (HDSL), network security, electronic mail and other internet services, like electronic commerce and EDI.

5.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Understand broadband network
- Analyze local loop technologies
- Comprehend Asymmetric Digital Subscriber Line (ADSL)
- Explain High Bit-Rate Digital Subscriber Line (HDSL)
- Comprehend line coding techniques
- Define Wireless Local Loop (WLL).
- Elaborate on network security and security levels
- Discuss electronic mail and other internet services

- Define electronic commerce
- Analyze internet as a tool for electronic commerce
- Explain Electronic Data Interchange (EDI)
- Understand the process of implementing EDI

5.2 BROADBAND NETWORK

Broadband LANs are multichannel, analog LANs. They are typically based on coaxial cable as the transmission medium, although fibre optic cable is also used. Individual channels offer bandwidth of 1 to 5 Mbps, with 20 to 30 channels typically supported. Aggregate bandwidth is as much as 500 MHz. Its characteristics are:

- Stations connected via RF modems, i.e., radio modems accomplish the digital-to-analog conversion process, providing the transmitting device access to an analog channel.
- Digital signal modulated onto RF carrier (analog).
- Channel allocation based on FDM.
- Head-End for bidirectional transmission.

Advantages

- Greater bandwidth
- Data, voice and video can be accommodated on broadband channel
- Greater distances

Disadvantages

- High cost, requires modems
- Lack of well-developed standards
- Cable design
- Alignment and maintenance

Some broadband LANs are referred to as 10Broadband36 where 10 stands for 10 Mbps, Broadband for multichannel and 36 for 3600 metres maximum separation between devices.

5.3 LOCAL LOOP TECHNOLOGIES

Local Loop (LL) is referred as an electronic circuit line from a subscriber's phone to the local exchange office termed as Local Central Office (LCO). The implementation of wires in the local loop technology is very tough for the operators, especially in the rural areas and also the remote areas due to lesser number of users which increases the cost of installation. Hence, the Wireless Local Loop (WLL) is used which is based on wireless links rather than the wired links made of copper wires for connecting the subscribers to the local central office. Public Switched Telephone Network (PSTN), Asymmetric Digital Subscriber Line (ADSL), High bit-rate Digital Subscriber Line (HDSL), Wireless Local Loop (WLL), Wireless Access Network Unit (WANU), etc. are examples of WLL.

Definition: A Local Loop (LL) is a physical connection from the end user site to a providers Point of Presence (POP). The local loop is provided in a number of ways depending on the type of provider. Local loops can be made of copper, fiber, coax, or wireless, and is installed to the demarcation point (Demarc).

Mostly, the local loop is installed by the Local Telephone Company (LTC), such as Incumbent Local Exchange Carrier (ILEC), but it is specifically installed by Competitive Local Exchange Carriers or CLEC's, fiber providers, or any other 3rd party providers. Local loops are also termed as the circuits, subscriber line, or physical link.

An Incumbent Local Exchange Carrier (ILEC) is a local telephone company, and a Competitive Local Exchange Carrier (CLEC) is a telephone company that competes with the already established local telephone business by providing its own network and switching.

Line Coding Techniques

Line coding is the process of converting digital data to digital signals. With the help of this technique a sequence of bits can be converted to a digital signal. At the sender side digital data are encoded into a digital signal and at the receiver side the digital data are recreated by decoding the digital signal.

Basically, the 'Line Coding', also called digital baseband modulation or digital baseband transmission, is a process carried out by a transmitter that converts data, in the form of binary digits, into a baseband digital signal that will represent the data on a transmission line. The transmission line in question could be a link between two devices in a computer network, or it could form part of a much larger telecommunications network. The receiver is responsible for converting the incoming line coded signal back into binary data.

There are different types of line coding techniques, ranging in complexity from very basic unipolar schemes in which the presence or absence of a voltage is used to represent a binary one or a binary zero, to highly sophisticate multilevel schemes in which different signal amplitudes are used, each representing a unique grouping of binary digits.

5.3.1 Asymmetric Digital Subsriber Line (ADSL)

In ADSL technology, there has been a new progress which intends to use two copper loops at a data rate of 1.544Mbps. This data rate is developed towards the user direction in the network and data rates upto 600Kbps from the user to network.

ADSL is widely used to connect most of the homes and small business subscribers to the Internet. ADSL divides the available frequencies in a telephone line by assumimg that most of the Internet subscribers are more inclined to download information from the Internet than upload. ADSL, therefore, provides the connection speed from the Internet to the subscribers three to four times faster than the connection from the subscribers back to the Internet. ADSL is considered as a distance-sensitive technology in which the quality of signal diminishes as the distance of the connection increases from home to the location of the service providers.

ADSL technology is capable of providing maximum download speed upto 8 Mbps at a distance of about 1.8 Km and upload speed of up to 640 Kbps. ADSL offers download speed in the range of 1-2 Mbps and upload speed in the range of 64-640 Kbps. Other versions of ADSL like ASDL2 and ASDL2+ are also availabe for higher downloadand upload speeds 12-24 Mbps and 1-3 Mbps respectively. The distance is not a limiting factor for telephone lines because telephone lines use repeaters to amplify voice signals while these repeaters are not compatible with DSL.

5.3.2 High Bit-Rate Digital Subscriber Line (HDSL)

High bit-rate Digital Subscriber Line (HDSL) is defined as a telecommunications protocol which was standardized in 1994. It was the first Digital Subscriber Line (DSL) technology which used a higher frequency spectrum over copper and twisted pair cables. HDSL was developed for transporting DS1 services at 1.544 Mbit/s or Mbps and 2.048 Mbit/s or Mbps over telephone local loops without requiring repeaters. Successor technology to HDSL includes HDSL2 (High bit-rate Digital Subscriber Line 2) and HDSL4 (High bit-rate Digital Subscriber Line 4), proprietary SDSL (Symmetric Digital Subscriber Line), and G.

High bit-rate Digital Subscriber Line 2 (HDSL2) is a standard developed by the American National Standards Institute (ANSI) Committee T1E1.4 and published in 2000 as ANSI T1.418-2000. Identical to its predecessor HDSL, HDSL2 provides a symmetric data rate of 1,544 Kbit/s or Kbps in both the upstream and downstream directions at a noise margin of 5-6 dB (deciBel). The modulation technique used in HDSL2 is TC-PAM (Trellis-Coded Pulse-Amplitude Modulation), which is also used in G. SHDSL. Spectral shaping is applied to increase compatibility with ADSL and HDSL2 on the same bundle of packet being transmitted. HDSL4 provides the similar bitrate as HDSL2, but uses four wires instead of two, in order to increase robustness. On a local loop, the reach of HDSL2 is approximately 9,000 feet (2.7 km), while that of HDSL4 is approximately 11,000 feet (3.4 km).

A Symmetric Digital Subscriber Line (SDSL) is referred as a Digital Subscriber Line (DSL) that transmits digital data over the copper wires of the telephone network, where the bandwidth in the downstream direction, from the network to the subscriber, is identical to the bandwidth in the upstream direction, from the subscriber to the network. This symmetric bandwidth is considered as the opposite of the asymmetric bandwidth recommended by Asymmetric Digital Subscriber Line (ADSL) technologies, where the upstream bandwidth is lower than the downstream bandwidth.

Single-pair High-speed Digital Subscriber Line (SHDSL) is referred as a specific type of Symmetric Digital Subscriber Line (SDSL), it is a data communications technology for even transmits and receives, i.e., symmetric data rate over copper telephone lines, which is faster than a conventional voiceband modem can provide. As opposed to other DSL technologies, SHDSL uses Trellis-Coded Pulse-Amplitude Modulation (TC-PAM). As a baseband transmission scheme, TC-PAM operates at frequencies that include those used by the analog voice Plain Old Telephone Service (POTS). Support of symmetric data rates made SHDSL a popular network for data transmission for Private Branch eXchange (PBX), Virtual Private Network (VPN), web hosting and other data services.

SHDSL features symmetrical data rates in both the upstream and downstream directions, from 192 Kbit/s Kbps to 2,312 Kbit/s Kbps of payload in 8 Kbit/s or Kbps increments for one pair and 384 Kbit/s Kbps to 4,624 Kbit/s Kbps in 16 Kbit/s Kbps increments for two pairs of wires. The reach varies according to the loop rate and noise conditions (more noise or higher rate means decreased reach) and may be up to 3,000 meters.

An optional extended SHDSL mode allows symmetric data rates up to 5,696 Kbit/s or Kbps on one pair, using the 32-TC-PAM modulation scheme specified in Annexes F and G. The SHDSL payload may be either 'Clear Channel' (unstructured), T1 (Transmission System 1) or E1 (E-Carrier 1), full rate or fractional, multiple ISDN (Integrated Services Digital Network) Basic Rate Interface (BRI), Asynchronous Transfer Mode (ATM) cells or Ethernet packets.

HDSL was developed for T1 service at 1.544 Mbit/s by the American National Standards Institute (ANSI) Committee T1E1.4 and published in February 1994 as ANSI Technical Report TR-28. This American variant uses two wire pairs at a rate of 784 kbit/s each, using the 2B1Q line code, which is also used in the American variant of the ISDN U interface. First products were developed in 1993. A European version of the standard for E1 service at 2.048 Mbit/s was published in February 1995 by the European Telecommunications Standards Institute (ETSI) as ETSI ETR 152. The first edition of ETR 152 specified the line code 2B1Q on either three pairs at 784 kbit/s each or two pairs at 1,168 kbit/s each. Second edition of ETR 152, published in June 1995, specified trellis coded carrierless amplitude/phase modulation (CAP) as an alternative modulation scheme, running on two pairs at 1,168 kbit/s each. Third version of ETR 152, published in december 1996, added the possibility of using a single CAP-modulated pair at 2,320 kbit/s. Later, an international HDSL standard was published by Study Group 15 of the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) on 26 August 1998 and adopted as recommendation ITU-T G.991.1 on 13 October 1998.

5.3.3 Wireless Local Loop (WLL)

Wireless Local Loop (WLL) is the use of a wireless communications link as the Last Mile / First Mile connection for delivering Plain Old Telephone Service (POTS) or Internet access (marketed under the term Broadband) to telecommunications users. Various types of WLL systems and technologies exist.

In addition, the various terms used for this type of access include Broadband Wireless Access (BWA), Radio In The Loop (RITL), Fixed-Radio Access (FRA), Fixed Wireless Access (FWA) and Metro Wireless (MW).

Fixed Wireless Terminal (FWT) units differ from conventional mobile terminal units operating within cellular networks, such as GSM (Global System for Mobile

Communications) in that a fixed wireless terminal or desk phone will be limited to an almost permanent location with almost no roaming abilities. WLL and FWT are generic terms for radio based telecommunications technologies and the respective devices which can be implemented using a number of different wireless and radio technologies.

Check Your Progress

1. What is a broadband network?
2. What do you understand by the term line coding ?
3. Write the use of ADSL.
4. How will you define High bit-rate Digital Subscriber Line (HDSL)?
5. What is the use of WLL?

5.4 NETWORK SECURITY

Network security is a broad topic with a multi-layered approach. It can be addressed at the data link, network and the application layers. The issues concerned are packet intrusion and encryption, IP packets and routing tables with their update versions, and host-level bugs that occur at the data link, network and the application layers respectively.

TCP/IP protocols are used globally irrespective of the nature of the organizations, whether they are general category organizations or security-specific sensitive organizations. The news or information about hacking of websites or portals by undesired people is very common nowadays. This shows that the TCP/IP protocols are susceptible to interception. This generates a need to ensure all round security for the network in an organization. The tasks of the network administrator have to be widened to include the overall security of the network. He has to ensure that all parts of this network are adequately protected and adequate measures of security have been implemented within a TCP/IP network. He should be aware of an effective security policy. He should also be able to pinpoint the main areas of risk that the network may face. These main areas of risk vary from network to network depending upon the functioning of the organization. There are, therefore, various security-related aspects which have direct implications for the network administrator alongwith the means to monitor the implemented measures of security effectively and to tackle the problem of breach of security if it happens.

5.4.1 Basic Requirements of Network Security

The main objective of the network is to share information amongst its users situated locally or remotely. Therefore, it is possible that undesired users can hack the network and prove to be harmful for the health of the network or the user. The network administrator must follow the following points to provide the network adequate security other than network-specific security as in the case of e-commerce, etc.

- Networks are designed to share information. Therefore, the network must be clearly configured to identify the shareable information and non-shareable information.

- The network should also clearly specify with whom the shareable information could be shared.
- An increase in the system security means a corresponding increase in the costs to the management. Therefore a compromising level between security and prices should be established as per the requirements of the network security system policy. This will largely depend upon the level of security needed for the network, the overall security requirements and the effective implementa-tion of the chosen level of security.
- Division of the responsibilities concerning the network's security must be clearly defined between the users and the system administrator.
- The requirements for security must be detailed within a network security policy of the organization that indicates the valuable data and their associated cost to the business.
- After defining the detailed network security policy and clearly identifying his responsibilities in the organization, the system administrator should be made responsible for ensuring that the security policy is effectively applied to the company environment, including the existing networking infrastructure.

5.4.2 Levels of Security

The US Department of Defense has listed different steps in the evolution of security levels. The first step in this direction was the trusted computer system evaluation criteria in December 1985, popularly termed as the Orange Book. In continuation with this level, another security level known as the trusted network interpretation of the trusted computer system evaluation criteria or the Red Book was described in July 1987. These security levels contain the security-related problems in the component or the modular form. Each level contains the specific security problem which is broken down into different divisions. Each of the divisions or classifications represents a security level defined in terms of the following general categories:

- User identification and authentication.
- The capability to monitor and audit system activity.
- Provision of discretionary access.
- Control of the reuse of resources.
- Identifying specific areas of possible attack.
- Provision of suitable counter measures.
- The level of system trusts, including systems architecture, design, implementation, transport, and trust of other hosts.

5.4.3 Data Security

Data security is concerned with the protection of data contained in a file or many files in a computer either as a standalone or on a network, from unauthorized interception.

In case of a postal system, a postcard as a carrier of information is open to all. It does not have any sort of security measures. An envelope is used to hide information from other people. An envelope acts as a means for security. Therefore, postcard and envelope have different purposes with respect to security. These two particular cases initiated similar actions to solve the security-related issues in case of data communication. E-mails are open to all as post cards. Following the envelope example in the postal system will enable users to secure at least some of their data.

The access protection provided by log on passwords is not a foolproof system and these may easily be bypassed. The bypassing can be done by booting from a diskette or connecting the stolen hard drive as a secondary one to another computer. In this manner, any vital data might easily be accessed. Consequently, encryption of the information seems to be the only effective way to protect data from being intercepted by unauthorized persons. The encryption must be developed to ensure reliable data security and that data is not decrypted without the right password or the right user. The main drawback of the password-based encryption includes the loss of password or registration of wrong passwords due to wrong spelling or some other human mistakes. In this case, it becomes impossible to restore the data. There are other rules to avoid in such situations.

The invalid access to the host can be prevented to a certain extent in the case of conventional host-to-terminal as the number of terminals connected is limited. The situation is entirely different in the case of Internet where access is allowed from any terminal connecting on a network. Therefore, this requires proper security measures. The following are the three types of security measures:

- Invalid access/Possibility of eavesdropping
- Firewall security
- Encryption (VPN Function)

5.4.4 Basic Techniques

Figure 5.1 shows a LAN connected to the Internet to allow access to the outside world. The terminals connected to the Internet have access to many servers to obtain a variety of information bases. But it also provides an opportunity to the undesirable person to execute commands on the servers and other computers. Thus, allowing users or terminals to access the server or other computers through websites to obtain desired information may create a number of problems. These problems can be as follows:

- Hackers may access someone else’s computer and servers without valid authentication to steal the information not meant for general purpose. The confidential information may reach wrong hands. Hacking shows that there is a security lapse in the particular system or network and the system or network is unstable and prone to tampering.
- Hacking may also lead to interception of the information from the network, which is lacking in security measures. The information may be tampered or altered from the actual contents.

NOTES

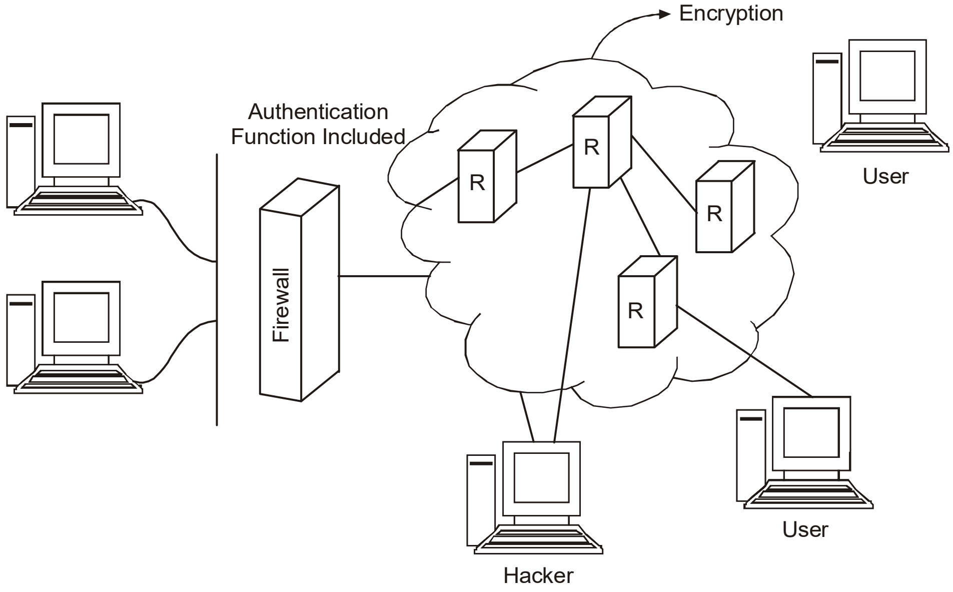


Fig. 5.1 *Unauthorized Access to LAN and Eavesdropping*

5.5 FIREWALLS

The Internet provides a two-way flow of traffic that may be undesirable in many organizations where some information is needed exclusively for the organization or for the Intranet. The Intranet is a TCP/IP network that is modelled after the Internet that only works within the organization. In order to delineate information meant only for the benefit of the organization or its Intranet and the other open to all or meant for the Internet, some sort of security measures are needed to control the two-way flow of traffic. A measure known as firewall is used for this purpose.

A firewall is a combination of software and hardware components that controls the traffic between a secure network (usually an office LAN) and an insecure network (usually the Internet), using rules defined by the system administrator. The firewall sits at the gateway of a network or sits at a connection between the two networks, and the entire traffic between two or more networks has to traverse the firewall. The firewall stops or allows the traffic based on the security policy as defined in rules' table.

The secure trusted network is said to be 'inside' the firewall. The insecure untrusted network is said to be 'outside' the firewall. The firewall's architecture has to be such that it would permit a certain amount of traffic to get through, otherwise it would be more of a 'stonewall', preventing access to the Internet, or sending of e-mails or any other information from either side of the firewall, thus turning into a self-defeating exercise.

The fact that it allows some traffic through provides a channel that could potentially be exploited, and could carry viruses.

However, principally, the philosophy behind firewall is:

- It exists to block traffic.
- It exists to permit traffic.

In brief, the basic aim of firewall is to provide only one entrance and exit to the network. There are two firewalls. One blocks the undesirable traffic, while the other allows traffic.

Network Architecture of a Firewall

The most important aspect of a firewall is that it is at the entry point of the networked system it protects. This means that the firewall is the first program that receives and handles incoming network traffic, and it is the last to handle outgoing traffic.

The logic is simple – a firewall must be positioned in a network to control all incoming and outgoing traffic. The internal network also needs to be structured and configured in such a way as to implement security policy of firewall to protect specific services running on the systems. The following are some examples of network structure to protect it from external threats using a firewall.

1. A router on dedicated connections to the Internet can be plugged into the firewall system as shown in Figure 5.2. This can also be provided with the help of a hub for full access servers outside the firewall as shown in Figure 5.3.

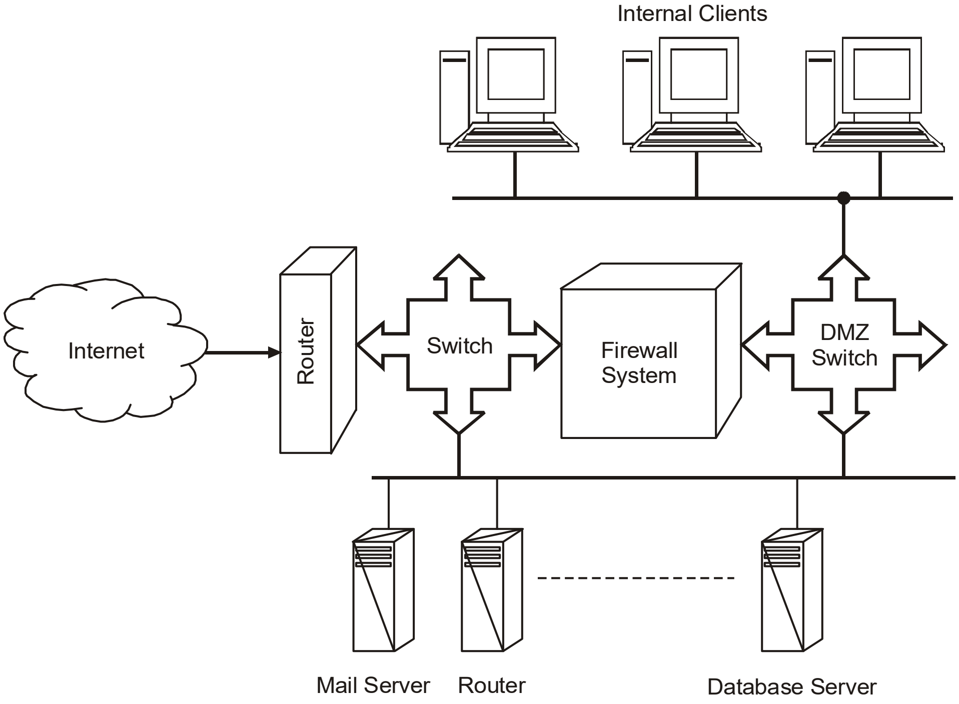


Fig. 5.2 *Router Connected to Firewalls on a Dedicated Connection*

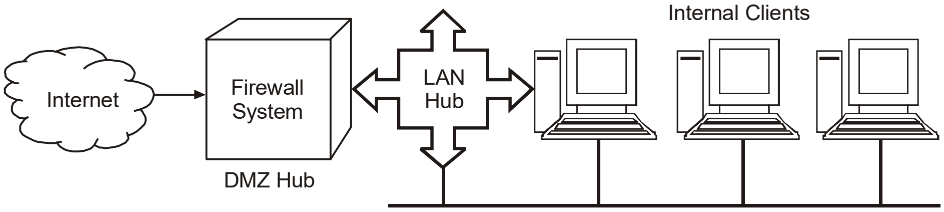


Fig. 5.3 *LAN Hub Connected to Firewalls*

NOTES

2. The router can be configured with some filtering rules. However, this router may be owned by ISP, therefore, ISP may be asked to put all desired control.
3. In a dial-up service like an ISDN line, a third network card is used to provide a filtered DMZ. This gives full control over the Internet services and still separates them from the regular network.
4. A proxy server can be used to monitor the traffic on the network and allow the users a limited number of services or some unwanted services may be blocked. This can be integrated with the firewall as shown in Figure 5.4.

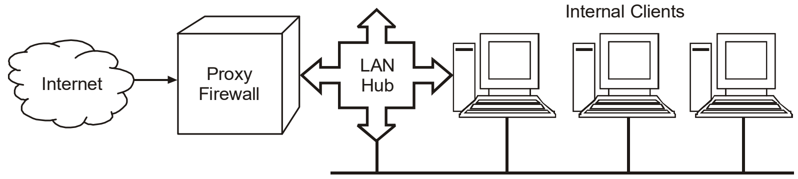


Fig. 5.4 *Proxy Server Connected with Firewall*

5. A proxy server on an organization's LAN connected with the firewall should have rules to only allow the proxy server to connect to the Internet for the services it is providing. This way the users can get to the Internet only through the proxy as shown in Figure 5.5.

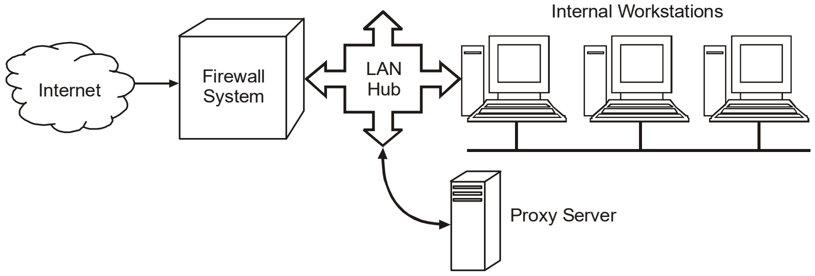


Fig. 5.5 *LAN Connected to the Internet via Proxy*

Check Your Progress

6. What is data security?
7. Write the three types of security measures.
8. Define the term firewall.
9. What is the basic aim of firewall?

5.6 DATA ENCRYPTION

Encryption hides your data from curious eyes. This is a method of encoding data to prevent unauthorized persons from viewing or modifying it. The main features of data encryption are:

- Prevents unwanted access to documents and e-mail messages.
- Even the strongest levels of encryption are very difficult to break.

Processes and Types of Encryption

The process of data encryption consists of certain steps. The data passes through a mathematical formula called an algorithm, which converts it into encrypted data called ciphertext. These algorithms create a key and then encapsulate the message with this key.

There are two types of encryptions – asymmetric and symmetric.

Asymmetric Encryption

In public key (asymmetric) encryption, two mathematically-related keys are used – one to encrypt the message and the other to decrypt it. These two keys combine to form a key pair. Asymmetric encryption provides both data encryption and validation of the communicating parties' identities and is considered more secure than symmetric encryption, but is computationally slower.

A public key encryption scheme has following six major parts:

(i) **Plaintext:** This is the text message to which an algorithm is applied.

(ii) **Encryption Algorithm:** It performs mathematical operations to conduct substitutions and transformations to the plaintext.

(iii) **Public and Private Keys:** These are a pair of keys where one is used for encryption and the other for decryption.

(iv) **Ciphertext:** This is the encrypted or scrambled message produced by applying the algorithm to the plaintext message using keys.

(v) **Decryption Algorithm:** This algorithm generates the ciphertext and the matching key to produce the plaintext.

Encryption Process

The asymmetric data encryption process has the following steps:

- The process of encryption begins by converting the text to a pre-hash code. This code is generated using a mathematical formula.
- This pre-hash code is encrypted by the software using the sender's private key.
- The private key is generated using the algorithm used by the software.
- The encrypted pre-hash code and the message are encrypted again using the sender's private key.
- The next step is for the sender of the message to retrieve the public key of the person for whom this information is intended.
- The sender encrypts the secret key with the recipient's public key, so that only the recipient can decrypt it with his/her private key, thus concluding the encryption process.

Decryption Process

The asymmetric data decryption process has the following steps:

- The recipient uses his/her private key to decrypt the secret key.
- The recipient uses his/her private key along with the secret key to decipher the encrypted pre-hash code and the encrypted message.

- The recipient then retrieves the sender's public key. This public key is used to decrypt the pre-hash code and to verify the sender's identity.
- The recipient generates a post-hash code from the message. If the post-hash code equals the pre-hash code, then this verifies that the message has not been changed enroute.

Symmetric Encryption

Private key encryption (symmetric) – also known as conventional or single-key encryption – is founded on a secret key shared by two communicating parties. It requires all parties that are communicating to share a common key. The secret key is used by the sending party to convert simple text to encrypted text, i.e., text that is enciphered using the secret key as the security component of the mathematical process. The receiving party then proceeds to decipher the encrypted material, using the same secret key that it shares. Examples of symmetric encryption systems would include the RSA RC4 algorithm (that furnishes the basis for Microsoft Point-to-Point Encryption (MPPE), Data Encryption Standard (DES), International Data Encryption Algorithm (IDEA), and the procedure now put forward by the US Government called 'Skipjack' Encryption Technology already utilized in the clipper chip.

An encryption scheme has five major parts:

(i) **Plaintext:** This is the text message to which an algorithm is applied.

(ii) **Encryption Algorithm:** It performs mathematical operations to conduct substitutions and transformations to the plaintext.

(iii) **Secret Key:** This is the input for the algorithm as the key dictates the encrypted outcome.

(iv) **Ciphertext:** This is the encrypted or scrambled message produced by applying the algorithm to the plaintext message using the secret key.

(v) **Decryption Algorithm:** This is the encryption algorithm in reverse. It uses the ciphertext and the secret key to derive the plaintext message.

When using this form of encryption, it is essential that the sender and the receiver have a way to exchange secret keys in a secure manner. If someone knows the secret key and can figure out the algorithm, communications will be insecure. There is also the need for a strong encryption algorithm. What this means is that if someone were to have a ciphertext and a corresponding plaintext message, they would be unable to determine the encryption algorithm. There are two methods of attacking conventional encryption – brute force and cryptanalysis. Brute force is just as it sounds; using a method (computer) to find all possible combinations and eventually determining the plaintext message. Cryptanalysis is a form of attack that strikes the characteristics of the algorithm to deduce a specific plaintext or the key used. One would then be able to figure out the plaintext for all past and future messages that continue to use this compromised setup.

NOTES

Encryption

Encryption Digital signature is not forged by other person. Once signer signs the document or message, it can not be forged. Signer can not replace the sign once message is signed. The two types of encryption are essentially used known as private key encryption and public key encryption.

Private Key Encryption: The private key encryption contains a secret key that is taken as code. This mechanism encrypts a packet of information if it passed across network to the other computer. The private key requires installing the key which is essentially the same as secret code. The code provides the key to decode the message. For example, a coded message A is substituted by C and B is substituted by D. Therefore, A becomes C and B becomes D. If a clue is given for the message that code is shifted by 2. The message can be decoded by your friend. A person wants to see the message, can not get the message, until and unless he knows the secret key. The financial, legal, ecommerce business transaction hindrance can be solved in the presence or absence of an authorized handwritten signature.

Public Key Encryption: The public key encryption uses private and public keys. The private key is restricted for the individual systems, whereas public key can be accessed by any system where message would be communicated securely with the individual system. Decoding for encrypted message is possible with public key that is provided by the individual system and its own private key. Basically, the key is based on hash value. The mechanisms are interrelated with each other that increase its popularity in transaction of digital cash, e-money transfer across net etc.

The worst kind of software failure you can have is when your computer refuses to start up because your operating system (Windows) refuses to work. The other type of failure you might experience is that an application refuses to run. Both of these types of failures can really stop you in your tracks and ruin your day. Internet hit by cable breakdown is to be considered as a major cause of irregular Internet accessing. Recently, Internet access in India and large parts of West Asia has been hit following breakdown of two undersea cables in the Mediterranean Sea. Internet access in India and large parts of West Asia including the UAE, Kuwait, and Saudi Arabia has been hit following breakdown of two undersea cables in the Mediterranean Sea. The disruption was caused by an anchoring ship that accidentally damaged Indian-owned Flag cable and SEA-ME-WE after being diverted from the Egyptian port of Alexandria due to bad weather. Repair teams have already set sail for the location to troubleshoot the problem and hence this target was termed as breakthrough process to breakdown the Internet. The following factors are responsible Internet breakdown to breakthrough list:

Registry Problems

Windows has a built-in database and control system to keep track of all of the software and critical information that lives on your PC. Usually they make comments under their breath. Your PC might not start up or an important program installed on your PC might not work.

- **Improper Installations:** If you are having problems with software, you can usually fix them by simply reinstalling your software. This is something we often put off doing because it is a pain, but it does work better than anything else.
- **Recent Software Installation:** When you install new software or hardware, internal changes are made to your operating system, and these changes can affect the balance of power on your computer. When you suspect problems like this then you can see and suggest for troubleshooting process if you can work through the list to try to find what went wrong.
- **Expired Software:** Many programs that users install on their PCs are often downloaded from the Internet. Many are trial versions. Programs like this run fine and offer all of the features of the commercial version until one day when they simply stop running. Software that needs to be activated may be fully functional for a few days and then may completely cease to functions until you activate it. Finally, subscription-based software, such as antivirus software might continue to work after it expires typically one year. The expired software creates problem to those system which are installed with the recent operating system.

5.7 AUTHENTICATION

Authentication

Authentication is any process by which one verifies that someone is who they claim they are. Basically, it involves a username and a password. It can also include any other method of demonstrating identity, such as a smart card, retina scan, voice recognition, or fingerprints. Basic authentication is the most basic form of authentication to Web applications. Both server and client authentication are required in Windows security.

Server Authentication: The server authentication is a part of client-server computing. Basically, SSL/TLS is used for authentication. A Web server acquires digital certificate from available server using Certification Authority (CA). CA is third party authority that issues digital certificates for authentication. The Digital Certificate (DC) authenticates the signature that is in fact digitally signed message. The DC uses SSL/TLS (Secured Socket Layer/ Transport Layer Security) in X.509 public key infrastructure that was defined by International Telecommunication Standardization Sector (ITU-T). If client connects to server using SSL/TLS both client and server follows strong cryptographic algorithm. Then the server sends X.509 certificate that contains the server's public key. The client then generates a 48-byte random number, a premaster secret key after encrypting the number used by the server's public key. The encrypted premaster secret key is sent to the server by client. After getting premaster secret key, the server decrypts the message using the private keys. Then both client-server shares the same premaster secret key which is basically symmetric key used to encrypt the message. Then they start communicating via generated keys. In this mechanism, only server knows the private key which decrypts the encrypted premaster secret key and then clients knows the message after sending the decrypted message by server. It proves that client is talking with correct server. This whole mechanism represents the complete scenario of authenticating the server.

NOTES

Client Authentication: In SSL/TSL, client authentication is not required instead it is optional. A client stays anonymous communicating between Web server and browser in B2B business transaction. Therefore, they use HTTP authentication methods. The HTTP authentication known as RFC 2617 represents the HTTP protocol in which client and server communicates between each other via HTTP protocol. It basically considers two factors as userid and password to authenticate the users/clients. Sometimes, userid might be user's email-id also. Both values are sent to authenticate without encryption and hence they are not considered as secure method of authentication in cryptography. In this mechanism, client sends Base64-encoded **userid** and **password** in HTTP header. If data is sent through SSL/TLS connection therefore, it is not altered or stolen during transmission. The malicious server can not disguise itself as genuine Web server and also not steal the password of user. For client authentication, SSL/TLS certificate is used to obtain an appropriate digital certificate before connecting to the server. A client generates the private key/public key pair to obtain the client certificate. The private key is kept as secret key and protected by passphrase. The passphrase works as password with added security. It is a sequence of word to control access to the system. The application does not maintain the database of userid and password. It verifies the certificate that is signed by trusted CA.

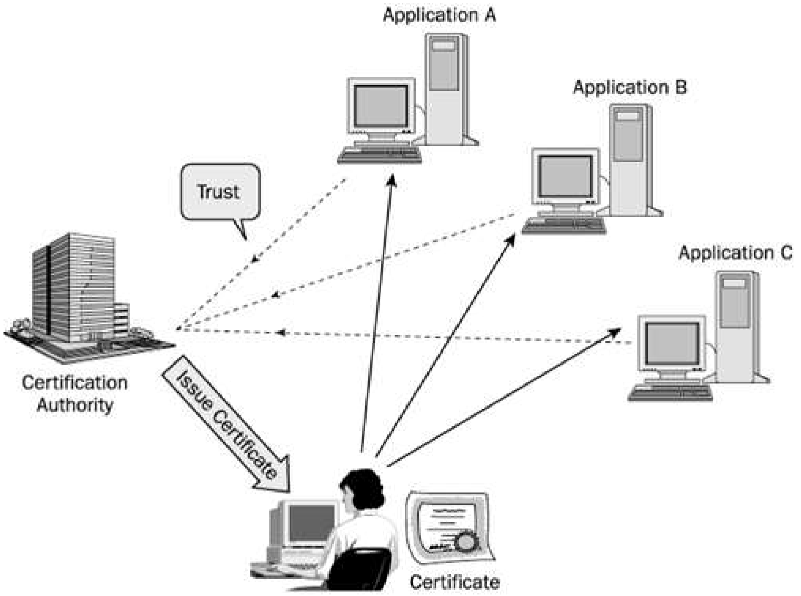
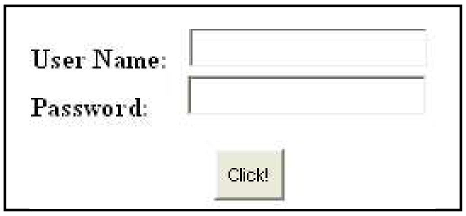


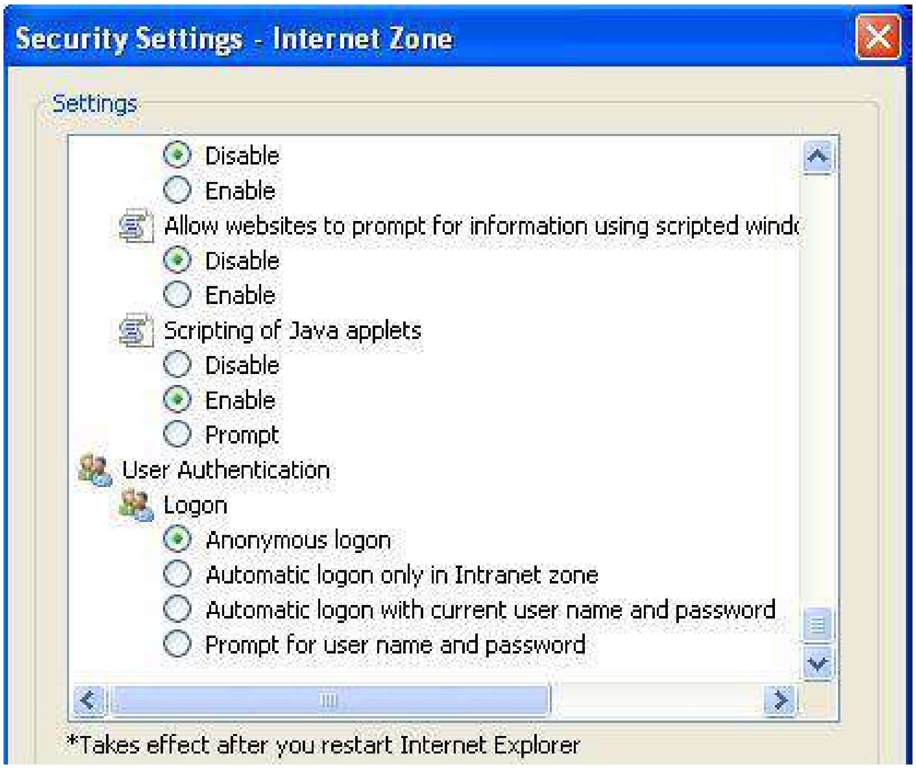
Fig. 5.6 *Uses of Client Certificates*

Figure 5.6 shows the complete scenario of using the client certificates. Let take an example, the customer manages ten passwords in which company **'XXX'** uses specific password to access the system and company **'YYY'** uses the service. Once certificate-based authentications are used by applications 'A', 'B' and 'C', the company issues CA where company trusts on legitimate user. In this way, client certificates are used to authenticate the message.

Authentication attaches date and time along with author of the signature and scrutinizes the contents when signature was being completed. It ensures that message is not altered. The message can be electronic documents, such as email, text file, spreadsheet etc. A person or information is authenticated on the computer by using various techniques. User name and password provides authentication. If user logs on the system unit or application, user name and password will be asked for checking authentication. Generally password authentication is provided to user in which two prime fields, such as **'User Name and Password'** are required to access the system. If the two requirements are not matched users are not allowed to access the system. If two requirements are not matched, users are not allowed to access the system.



The security setting for Internet user can be configured with '**Security Settings – Internet Zone'** tab.



Security Settings - Internet Zone

Settings

● Disable
○ Enable
Allow websites to prompt for information using scripted windc
● Disable
○ Enable
Scripting of Java applets
○ Disable
● Enable
○ Prompt
User Authentication
Logon
● Anonymous logon
○ Automatic logon only in Intranet zone
○ Automatic logon with current user name and password
○ Prompt for user name and password

*Takes effect after you restart Internet Explorer

5.8 VIRUSES

Viruses are frequently transmitted through e-mail attachments, peer--to--peer downloads, phishing and instant messages. Phishing refers to a fraudulent process in which user's credentials are easily grabbed. Among these, e-mail attachments carry and spread virus fast in an address book or a random combination of address book. If these viruses are not controlled quickly, the servers can disrupt the e-mail services for all systems. The functions of computer viruses are as follows:

- It deletes registry attaches, system files and log files. It also destroys the OS of system units.
- Many viruses slow down the computer performance.
- Viruses decline suddenly in speed and in loss of files or data.
- They cause to display unknown and uncommon messages or even playing a tune.
- They are responsible for the loss of hard disk partition.
- They release of files normally via e-mail.

To know the working and spreading of viruses on PC or even on the Internet, you must know the lifecycle of computer viruses. The lifecycle of computer virus is as follows:

- **Coding** → In this phase, the virus program is coded and developed.
- **Releasing** → Once the code is ready, it is spread to the system and network.
- **Spreading** → It is out forwarded through a simple e-mail.
- **Quarantining** → In this phase, the virus gets quarantined. This phase often happens when it validates the signature of the virus and develops an antivirus update.

Viruses spread from machine to machine and across network in a number of ways. The viruses are always trying to trigger and execute the malicious programs that intently spread the computer system. For example, a macro virus is booted with infected disk and spread the viruses to boot sector. Then it started to share the network drive or other media that exchanges infected files across Internet by downloading files and attachments form Internet. The transmission of viruses is possible by following ways:

- If system unit is booted from infected files.
- If programs are executed with an infected programs.
- The virus infiltration includes common routes, floppy disks and e-mail attachments.
- If pirated software and shareware are used in the system files.

Viruses are malicious and smart. Many a times, users are not able to realize that their system unit has got infected with viruses. The property of viruses is that they hide themselves among regular system files or as attachments and camouflage themselves as standard files. The following steps are usually taken if the system gets infected with the viruses:

- The Internet facility and LAN utilities must be disconnected for the some time.
- The OS must be installed within the system unit if the system has not been booted properly. If the system does not recognize the hard drive it means that it is infected with virus. It is better to boot from Windows to rescue the disk. The partition table of scandisk must be recovered using scandisk for a standard Windows program.

- A back-up of all important and critical data must be taken at regular intervals by using external devices, such as CD, USB, floppy disks or even flash memory etc.
- Antivirus software must be installed in the system and should be made to rweekend or at the end of the month. A good quality antivirus disinfects infected objects, quarantines infected objects and is able to delete Trojans and worms.
- The latest updates must be taken to remove the antivirus database. The infected computer is not included to download these updates.
- Scans disinfect the mails of a client's database and ensure the maliciousograms are not reactivated if messages are sent from one address to another across the network.
- Firewall security features must be installed in the system that prevent attacks from malicious and foreign programs.
- Delete and clean the corrupted applications and files. Try to reinstall the required applications in your system but be sure that the corresponding software is not pirated.

Antivirus Software(s)

Now you know the concept of worms, spyware and viruses. You need to install and run the antivirus programs to clean the virus and provides the security for Windows. It helps in protecting Windows from crashing. The antivirus software available in the market to deal with virus-related issues are as follows:

- Symantec Antivirus that is used to check the security of foreign programs and applications.
- Windows Vista AntivirusSpyware, AntivirusNorton Antivirus that is used to catch worms, rootkits, spywares, viruses, etc..
- Avast Antivirus.
- Kaspersky Antivirus that is used for HTTP traffic-checking and for providing a security wizard.

These antiviruses are useful for those types of viruses that are downloaded from the net or from email attachments. The most popular antivirus programs are Data Fellows F-Prot, EliaShim ViruSafe, ESaSS ThunderBYTE, IBM Antivirus, McAfee Scan, Microsoft Anti Virus, Symantec Norton Antivirus and S&S Dr Solomon's AVTK. The hard disks and drives must be scanned on a daily basis. Every week, hackers and malicious programmers release their virus programs across the Internet so it is better to keep the system updated with the latest antivirus software and programs. The updated user manual and help files must be provided to the users during the installation of expensive applications and the operating system. In fact, automatic updates to the list of antivirus and multithread detection are the standard features of an antivirus program.

Windows XP Firewall

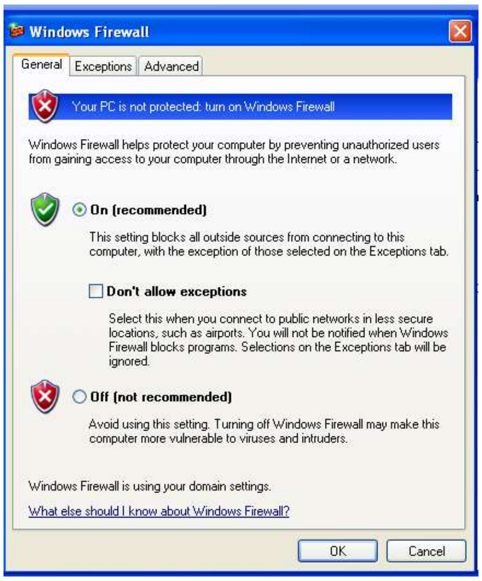
A security audit logging feature is provided with the Windows XP Firewall. Windows security is protected with firewall installation. Firewall provides a secure barrier for Windows security that protects your PC from outside and foreign world. It is configured with full-time Internet connection. Firewall software is considered as an effective means to protect the Internet from malfunctions and networked-based security threats. Information and services are essentially required for the organizations. Internet connectivity uses dialup capability and installed with the system unit to the Internet service provider. Connections network requires various types of software as well as operating system. Firewall is inserted between Internet and Internet-based attacks that provide a single choke point in which all other malfunctions are tracked. The characteristics of firewall are as follows:

- All Internet traffic must be passed via firewall. It the foreign accessing approaches to the local network.
- Only authorized traffic is to be allowed to pass.
- It itself is to be immune to be permutation.
- It filters traffic with the help of allotted IP address and also takes help of TCP port number.
- It hosts the server software, such as Web or mail service.
- It monitors security-related events.
- It provides a platform for the Internet Protocol Security (IPSec) that includes a network address translator, audits and alarms.

The limitations of firewall are as follows:

- Some of the complex types of attacks are not protected by firewalls.
- It can not protect virus-infected programs and not able to scan the incoming files, messages for viruses, emails, etc.
- It does not protect against threats.

The firewall is consisted of two systems, known as a bastion-host and a packet filtering router. Bastion-host is worked for authenticating services and performing the proxy functions. The configuration of firewall is assembled between two packet-filtering routers. In this setting, one approach comes between the bastion-host and the Internet work. This configuration is set with an isolated subnetwork, which provides three levels of defense to thwart intruders.



Your PC is not protected: turn on Windows Firewall

Windows Firewall helps protect your computer by preventing unauthorized users from gaining access to your computer through the Internet or a network.

● **On (recommended)**

This setting blocks all outside sources from connecting to this computer, with the exception of those selected on the Exceptions tab.

□ **Don't allow exceptions**

Select this when you connect to public networks in less secure locations, such as airports. You will not be notified when Windows Firewall blocks programs. Selections on the Exceptions tab will be ignored.

○ **Off (not recommended)**

Avoid using this setting. Turning off Windows Firewall may make this computer more vulnerable to viruses and intruders.

Windows Firewall is using your domain settings.

What else should I know about Windows Firewall?

You need to select the On radio button in Windows Firewall tab, because it blocks all outside and foreign sources from connecting to this computer.

NOTES

5.9 INTERNET PRIVACY AND SECURITY ATTACKS

Each individual has his or her own personal space. Breaching that privacy is considered to be unacceptable. This rule applies to the Internet or Web as well. There are innumerable Internet users all over the world who are constantly accessing data and uploading data. So the Internet privacy is to regulate what a person reveals about himself or herself and who can gain access to it. The Internet privacy is the desire or mandate of personal privacy concerning transactions or transmission of data via the Internet. It involves the exercise of control over the type and amount of information a person reveals about himself on the Internet and who may access such information. The term is often understood to mean universal Internet privacy, i.e., *every* user of the Internet possessing Internet privacy. Internet privacy forms a subset of computer privacy. People with only a casual concern for Internet privacy need not achieve total anonymity.

As people use computers for a variety of purposes, confidential information, confidential communications, and personal choices can be registered in a variety of ways. The *Internet privacy* is a broad term referring to the various concerns, technologies, and strategies for protecting information, communications, and choices that are meant to be private. In general, using the Internet often means giving up some measure of privacy. For people who wish to remain completely anonymous, the best approach is to use a public computer, such as those available at public libraries. Other steps to take when anonymity is the goal include clearing the cache and browsing history before leaving the computer. This is done in different ways depending on the browser used and refraining from entering any personal information or creating any user names or passwords.

For example, if you want to shop online, use social networking sites, play online games, or participate in forums, Internet privacy can become an issue in a number of ways. If your passwords are exposed, your identity can be fraudulently used or even stolen. If your words, photographs or products you have created are posted without your permission, your reputation and income can be damaged. If your contact information is passed around, you may be subject to spam. If your browsing history becomes public, people will know what you have been looking at online. Fortunately, taking certain precautions can reduce the privacy risks that you face.

Strong passwords that are kept secret are one way to safeguard your Internet privacy. Browser privacy settings, which control elements like storage of your browsing and download history and the acceptance of cookies, are there for you to alter to meet your preferences. The options differ with different browsers. Similarly, social networking sites have settings to allow you to control the level of privacy of various postings you may make. The default settings may be skewed towards the public exposure of information rather than towards Internet privacy. For sites such as forums, make sure you read the privacy terms before signing up.

Internet users obtain Internet access through an Internet Service Provider (ISP). All data transmitted to and from users must pass through the ISP. Thus, an ISP has the potential to observe users' activities on the Internet. In addition, search engines have the ability to track a user's searches. Personal information can be revealed through searches including search items used, the time of the search, etc. Search engines have claimed a necessity to retain such information in order to provide better services, protect against security pressure and protect against fraud.

Security Attacks and Counter Measures

ITU-T (International Telecommunication Union) Recommendation X.800, Security Architecture for OSI (Open Systems Interconnections) defines systematic way to define the requirements for security and characterize the approaches to satisfying those requirements. It effectively assesses the security needs of an organization and evaluates, and chooses various security products and policies. The OSI security architecture is a way of organizing the task of providing security. Furthermore, because this architecture was developed as an international standard, computer and communications vendors have developed security features for their products and services that relate to this structured definition of services and mechanisms. The OSI security architecture focuses on security attacks, mechanisms and services as follows:

- **Security Attack**: It refers to any action that compromises the security of information owned by an organization.
- **Security Mechanism**: It is a process (or a device incorporating such a process) that is designed to detect, prevent or recover from a security attack.
- **Security Service**: It refers to processing or communication service that enhances the security of the data processing systems and the information transfers of an organization. The services are intended to counter security attacks, and they make use of one or more security mechanisms to provide the service.

Threat is a potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm. That is, a threat is a possible danger that might exploit vulnerability.

Attack is an assault on system security that derives from an intelligent threat; that is, an intelligent act that is a deliberate attempt (especially in the sense of a method or technique) to evade security services and violate the security policy of a system.

Security Attacks

A security policy can be defined as the framework within which an organization establishes needed levels of information security to achieve the desired confidentiality goals. A policy is a statement of information values, protection responsibilities and organization commitment for a system. Prior to evaluating attacks against a system and deciding on appropriate mechanisms to repulse these threats, it is necessary to specify a security policy. A security policy that is sufficient for the data of one organization may not be sufficient for another organization.

Security Attacks

A graphical representation of the communication process and some of the attacks is given in Figure 5.7.

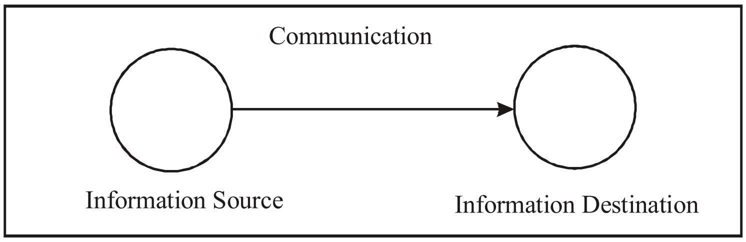


Fig. 5.7 Information Flow

Interruption

An interruption can be defined as a state where the asset of a system gets destroyed or becomes unavailable. This type of attack targets the source or the communication channel and prevents the information from reaching its intended target, for example, an attacker may cut the physical wire to prevent the information from reaching its destination. Another commonly used technique by the attacker is to overload the carrying media whereby the information gets dropped due to congestion. Attacks in this category attempt to perform a kind of Denial of Service (DoS). A graphical representation of interruption is given in Figure 5.8.

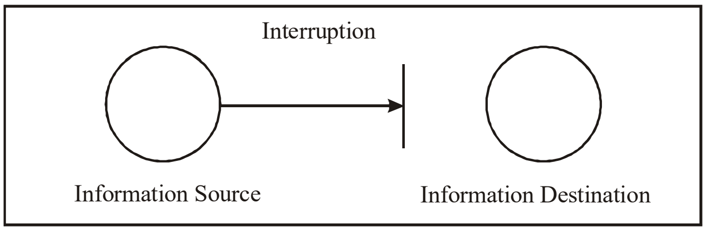


Fig. 5.8 Interruption

Interception

Interception happens when an unauthorized party gets access to the information by eavesdropping into the communication channel. Wiretapping is a good example of interception. A graphical representation of interception is given in Figure 5.9.

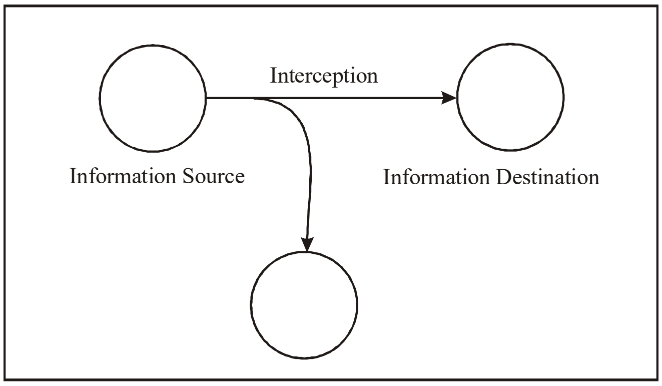


Fig. 5.9 Interception

Modification

In modification, the information is not only intercepted but modified by an unauthorized party while in transit from the source to the destination (e.g. by modifying the message content). A graphical representation of modification is given in Figure 5.10.

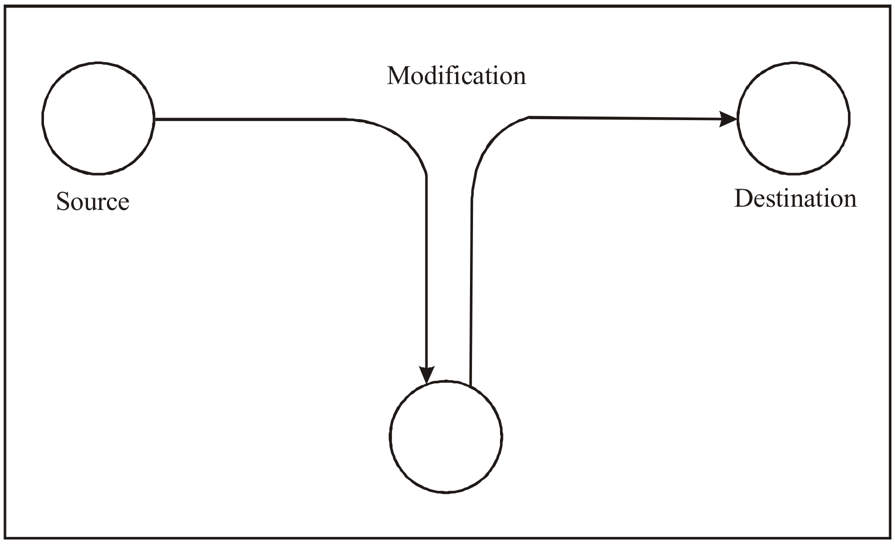


Fig. 5.10 Modification

Fabrication

Fabrication occurs when an attacker inserts forged objects into the system without the sender's knowledge or involvement. Fabrication can be categorized as follows:

- **Replaying:** When a previously intercepted entity is inserted, this process is called replaying. For example, replaying an authentication message.
- **Masquerading:** When the attacker pretends to be the legitimate source and inserts his/her desired information, the attack is called masquerading. For example, adding new records to a file or database.

A graphical representation of fabrication is given in Figure 5.11.

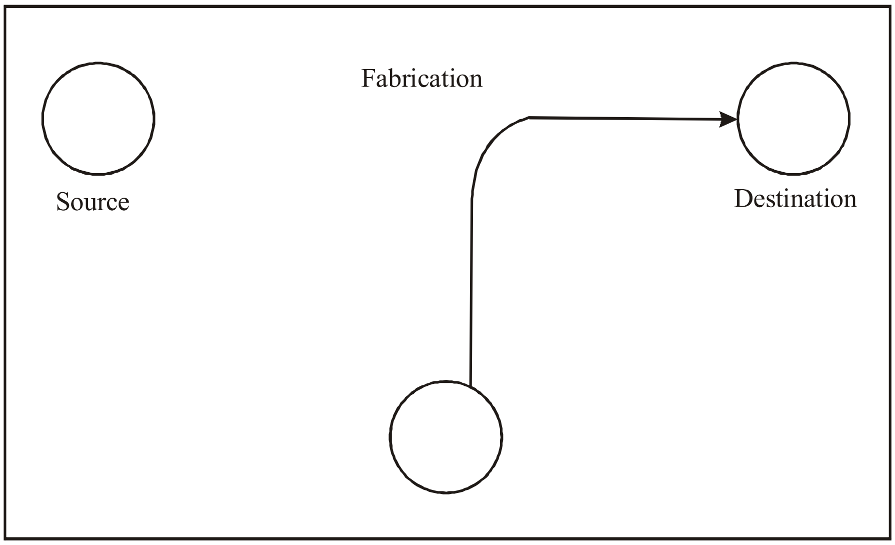


Fig. 5.11 Fabrication

Network Security and Counter Measures

Security management for networks is different for all kinds of situations. A small home or an office would only require basic security while large businesses will require high maintenance and advanced software and hardware to prevent malicious attacks from hacking and spamming.

Small Homes

- Use a basic firewall or a unified threat management system.
- For Windows users, a basic Antivirus software. An anti-spyware program would also be a good idea. There are many other types of antivirus or anti-spyware programs out there to be considered.
- When using a wireless connection, use a robust password. Also try to use the strongest security supported by your wireless devices, such as WPA2 with AES encryption.
- If using Wireless change the default SSID network name, also disable SSID Broadcast; as this function is unnecessary for home use.
- Enable MAC Address filtering to keep track of all home network MAC devices connecting to your router.
- Assign STATIC IP addresses to network devices.
- Disable ICMP ping on router.
- Review router or firewall logs to help identify abnormal network connections or traffic to the Internet.
- Use passwords for all accounts.
- Have multiple accounts per family member, using non-administrative accounts for day-to-day activities. Disable the guest account (Control Panel → Administrative Tools → Computer Management → Users).
- Raise awareness about information security to children.

Medium Businesses

- Use a fairly strong firewall or Unified Threat Management System
- Strong Antivirus software and Internet Security Software.
- For authentication, use strong passwords and change it on a bi-weekly/ monthly basis.
- When using a wireless connection, use a robust password.
- Raise awareness about physical security to employees.
- Use an optional network analyser or network monitor.
- An enlightened administrator or manager.

Large Businesses

- A strong firewall and proxy to keep unwanted people out.
- A strong Antivirus software package and Internet Security Software package.

NOTES

- For authentication, use strong passwords and change it on a weekly/bi-weekly basis.
- When using a wireless connection, use a robust password.
- Exercise physical security precautions to employees.
- Prepare a network analyser or network monitor and use it when needed.
- Implement physical security management like closed circuit television for entry areas and restricted zones.
- Security fencing to mark the company's perimeter.
- Fire extinguishers for fire-sensitive areas like server rooms and security rooms.
- Security guards can help to maximize security.

School

- An adjustable firewall and proxy to allow authorized users access from the outside and inside.
- Strong Antivirus software and Internet Security Software packages.
- Wireless connections that lead to firewalls.
- Children's Internet Protection Act compliance.
- Supervision of network to guarantee updates and changes based on popular site usage.
- Constant supervision by teachers, librarians, and administrators to guarantee protection against attacks by both internet and sneakernet sources.

Large Government

- A strong firewall and proxy to keep unwanted people out.
- Strong Antivirus software and Internet Security Software suites.
- Strong encryption.
- White list authorized wireless connection, block all else.
- All network hardware is in secure zones.
- All host should be on a private network that is invisible from the outside.
- Put Web servers in a DMZ, or a firewall from the outside and from the inside.
- Security fencing to mark perimeter and set wireless range to this.

5.9.1 Key Management

The management of keys is the chief problem area for all encryption systems. The keys are the most valuable information. If anyone can get a key, anyone can decrypt everything that has been encrypted by that key. In some cases, one may also be able to get succeeding keys. The management of keys is not just about securing them while in use. It is also about creating strong keys, securely distributing the keys to correct users, and revoking the keys, if they have been compromised or expired.

NOTES

Key Creation

Obviously, all keys must be created with care. Certain keys have poor security performance with certain algorithms. For example, when creating keys for use with RSA, care must be used to choose a and b from the set of prime numbers. Likewise, a key of all 0's when used with DES does not provide strong security. Most encryption systems have some method for generating keys. In certain cases, people are allowed to choose the key by creating a password. In this case, it may be a good idea to instruct the users on how to choose strong passwords, ideally, which should include numbers and special characters. Otherwise, the total key space is significantly reduced (this allows quicker brute-force key searches).

Some keys are chosen from random numbers. Unfortunately, there are very few truly random number generators. Most are pseudo-random. If the generator is not truly random, it may be possible to predict the next number. If anyone bases the keys on the output of the random number generator and you can predict the output, you may be able to predict the key.

The length of the key may also need to be chosen. Some algorithms use fixed key lengths (such as RSA with 512-bit key). For example, a 1,024-bit RSA key is stronger than a 512-bit RSA key. You cannot, however, compare the strength of the DES key to a RSA key in a similar manner. Table 5.1 shows the relative strengths of keys for different types of algorithms.

Table 5.1 *Relative Key Strengths for Different Types of Algorithms*

Private Key Encryption (DES, RC5)	Public Key Encryption (RSA, Diffie-Hellman)	Elliptic Curve Encryption
40 bits	–	–
56 bits	400 bits	–
64 bits	512 bits	–
80 bits	768 bits	–
90 bits	1,024 bits	160 bits
120 bits	2,048 bits	210 bits
128 bits	2,304 bits	256 bits

Key Distribution

Once the keys have been generated, they must get to various locations and to be equipment to be used. If the key is not secure during transit, it may be copied or stolen and the entire encryption system is now insecure. Therefore, the distribution channel must itself be secure. Keys could be moved out-of-band. In other words, the keys could be transported by administrators by hand. This may work successfully if the remote sites are short distances apart. But what if the remote sites are continents apart? The problem becomes much harder.

There is a partial solution to this problem, however. It may be possible to use the Diffie–Hellman Key Exchange in order to create and distribute many session keys (short-term keys used for a single session or a small amount of traffic). This may reduce the need to travel to remote locations. Longer-term keys (RSA keys, for example) require more care. It is not appropriate to use the Diffie-Hellman Key Exchange algorithm to distribute the RSA key pairs. In the case of RSA key pairs, one key must be kept secret and one can be published. The key that is published must be published in such a way as to preclude being tampered with. If the pairs are to be generated by a central authority, the private key must be securely transmitted to the pair owner. If the owner will generate the key pair, the public key may need to be transmitted to the central authority in a secure manner.

Key Certification

If keys are transmitted to a remote destination by some means, they must be checked once they arrive to be sure that they have not been tampered with during transit. This can be a manual process or it can be done via some type of digital signature. Public keys are intended to be published or given out to other users and must also be certified as belonging to the owner of the key pair. This can be done through a central authority (normally called a certificate authority, or CA). In this case, the CA provides a digital signature on the public key and this certifies that the CA believes the public key belongs to the owner of the key pair (Refer Figure 5.12). Without proper certification, an attacker could introduce her own keys into the system and thus compromise the security of all information transmitted or authenticated.

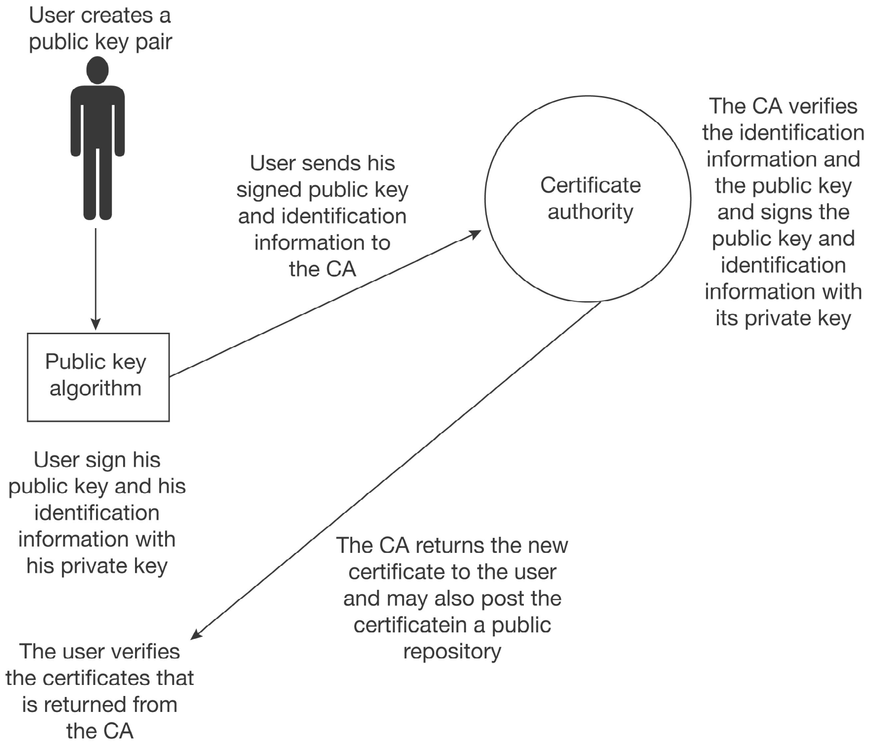

User creates a public key pair
User sends his signed public key and identification information to the CA
Certificate authority
The CA verifies the identification information and the public key and signs the public key and identification information with its private key
Public key algorithm
User sign his public key and his identification information with his private key
The CA returns the new certificate to the user and may also post the certificatein a public repository
The user verifies the certificates that is returned from the CA

Fig. 5.12 Key Protection

Key Protection

The public keys of a key pair do not require confidentiality protection. They only require the integrity protection which is provided by their certification. The private key of a public key pair must be protected at all times. If an attacker were to gain a copy of the private key, he could read all confidential traffic addressed to the key pair owner and also digitally sign information as if he was the key owner. The protection of the private key includes all copies of it. Therefore, the file that holds the key must be protected just like any backup tape that may include the file. Most systems protect the private key with a password. This will protect the key from casual snooping but not from a concerted attack.

The password used to protect the key must be well chosen to resist brute-force attacks. However, the best way to protect the key is to prevent an attacker from gaining access to the file in the first place. All keys to a private key system must be protected. If the key is kept in a file, this file must be protected wherever it may reside. If the key will reside in memory, care must be taken to protect the memory space from examination by a user or process. Likewise, in the case of a core dump, the core file must be protected since it may include the key.

Key Revocation

Keys do not have infinite lives. Session keys only exist for a given session. There is no need to revoke the key as it becomes invalid at the end of the session. Some keys may be certified for a given period of time. Generally speaking, public keys pairs are certified for one or two years. The certified public key will recognize if the date has expired. Systems that read the certificate will not consider it valid after that date so it becomes unnecessary to revoke an expired certificate. However, keys can also be lost or compromised. If this happens, the owner of the key must inform other users of the fact that the key is no longer valid and thus it should not be used. In the case of a private key encryption system, if a key is compromised (and if the users of the system know it) they can communicate this information to each other and begin using a fresh key. The case of public key encryption systems is a little different.

If a key pair is compromised and revoked, there is no obvious way to inform all of the potential users about the public key which is no longer valid. In some cases, public keys are published to key servers. Someone wishing to communicate with the owner of the key may go to the server once to retrieve the certified public key. If the key is compromised and revoked, how does another person find out? The solution is that one must periodically visit the key server to see if there is a revocation of the key and the owner of the key must post the revocation to all of the potential key servers. The key servers must also hold this revocation information at least until the original certificate expires.

5.10 ELECTRONIC MAIL AND INTERNET SERVICES

Of the various applications of TCP/IP the most important one is the internetworking equivalent of the real-world postal delivery system, commonly called *electronic mail* or *e-mail*. The history of e-mail goes back to the very earliest days of TCP/IP's development. Today millions of people every day send both simple and complex messages around the world through e-mail. TCP/IP e-mail is not any one application. It is implemented as a complete system comprising several protocols, software elements and components. All these elements perform one or the other part of the complete communication process of e-mail. These include a standard message format, a specific syntax for recipient addressing, and protocols to both deliver mail and allow access to mailboxes from intermittently connected TCP/IP clients.

Mail Communication Process Steps

NOTES

The modern TCP/IP e-mail communication process consists of the following five basic steps.

1. **Mail Composition**

E-mail journey begins with the creation of a message, that is, electronic mail message. There are two parts of a message: the *header* and the *body*. the header contains data that describes the message and controls how it is delivered and processed, the body of the message is the actual information that is to be communicated. The message must be created as per the standard message format for the e-mail system so that it can be processed. It must also specify the e-mail addresses of the intended recipients for the message.

By way of analogy to real mail, the body of the message is like a letter, and the header is like the envelope into which the letter is placed.

2. **Mail Submission**

There are various other internetworking applications besides e-mail. But, electronic mail is different from many other internetworking applications in that the sender and receiver of a message do not necessarily need to be connected to the network simultaneously, nor even continuously, to use it. The system is so designed that after composing the message, the user decides when to submit the message to the electronic mail system so it can be delivered. The mail is transferred by using the Simple Mail Transfer Protocol (SMTP).

This is analogous to dropping off an envelope at the post office, or to a postal worker picking up an envelope from a mailbox and carrying it to the local post office to insert into the mail delivery stream.

3. **Mail Delivery**

Once the user has submitted the electronic mail message, it is accepted by the sender's local SMTP system for delivery through the mail system to the destination user. Today, this is accomplished by performing a Domain Name System (DNS) lookup of the intended recipient's host system and establishing an SMTP connection with that system. SMTP also supports the ability to specify a sequence of SMTP servers through which a message must be passed to reach the desired destination. One way or the other, eventually the message arrives at the recipient's local SMTP system.

This is like the transportation of the envelope through the postal system's internal 'internetwork' of trucks, airplanes and other equipment to the intended recipient's local post office.

4. **Mail Receipt and Processing**

Now, the local SMTP server accepts the e-mail message for further processing. It places the mail into the intended recipient's mail box, where it waits for the user to retrieve it.

In our physical analogy, this is the step where the recipient's local post office sorts mail coming in from the postal delivery system and puts the mail into individual post office boxes or bins for delivery.

NOTES

5. **Mail Access and Retrieval**

The intended recipient periodically checks with its local SMTP server to see if there is any mail for him/her. If so, the recipient can retrieve the mail, open it and read its content. This is done by a special mail access protocol or method and not by SMTP. The access protocol and client e-mail software may allow the user to scan the headers of received mail (such as the subject and sender's identity) to decide which mail messages to download. This saves quite a lot of time as user need not actually open up every mail.

This is the step where mail is physically picked up at the post office or delivered to the home.

Electronic Mail Message Communication Model, Devices and Protocol Roles

One of the critical requirements of an electronic mail system is that the sender and receiver of a message need not be online at the time when mail is sent. TCP/IP therefore uses a communication model with several devices that allow the sender and recipient to be *decoupled*. The sender's client device spools mail and moves it to the sender's local SMTP server when it is ready for transmission. The e-mail is then transmitted to the receiver's SMTP server using SMTP where it remains for an indefinite period of time. When the recipient is ready to read it, he or she retrieves it using one or more of a set of mail access protocols and methods, the two most popular of which are POP and IMAP (Refer Figure 5.13).

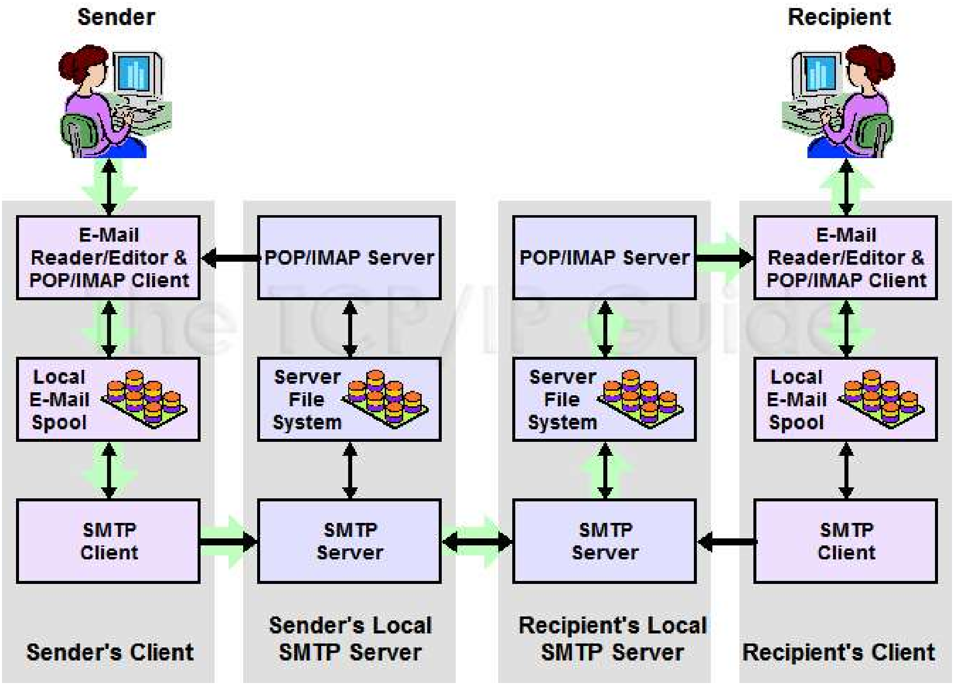


Fig. 5.13 Electronic Mail Communication Model

- **Sender's Client Host:** The sender of the mail composes an electronic mail message, generally using a mail client program on his or her local machine. The mail, once composed, is not immediately sent out over the Internet; it is held in a buffer area called a *spool*. This allows the user to be "detached" for the entire time when a number of outgoing messages are created. When the user is done, all of the messages can be sent at once.
- **Sender's Local SMTP Server:** When the user's mail is ready to be sent, he or she connects to the internetwork. The messages are then communicated to the user's designated local SMTP server, normally run by the user's Internet Service Provider (ISP). The mail is sent from the client machine to the local SMTP server using SMTP. (It is possible in some cases for the sender to be working directly on a device with a local SMTP server, in which case sending is simplified.)
- **Recipient's Local SMTP Server:** The sender's SMTP server sends the e-mail using SMTP to the recipient's local SMTP server over the Internetwork. There, the e-mail is placed into the recipient's incoming mailbox (*inbox*). This is comparable to the outgoing spool that existed on the sender's client machine. It allows the recipient to accumulate mail from many sources over a period of time, and retrieve them when it is convenient.
- **Recipient's Client Host:** In certain cases the recipient may access his or her mailboxes directly on the local SMTP server. More often, however, a mail access and retrieval protocol, such as POP3 or IMAP, is used to read the mail from the SMTP server and display it on the recipient's local machine. There, it is displayed using an e-mail client program, similar to the one the sender used to compose the message in the first place.

Some form of addressing is required for all network communications. Since electronic mail is *user-oriented*, e-mail addresses are based on users as well. In modern TCP/IP e-mail, standard addresses consist of a *user name*, which specifies who the recipient is, and a *domain name*, which specifies the DNS domain where the user is located. A special DNS *mail exchange* (MX) record is set up for each domain that accepts e-mail, so a sending SMTP server can determine what SMTP server it should use to send mail to a particular recipient.

User Agent

The user agent (UA) makes the sending and receiving any message easier and provides service to the user. Some examples of command-driven user agents are mail, pine, and elm. Some examples of GUI-based user agents are Eudora, Outlook, and Netscape (Refer Figure 5.14).

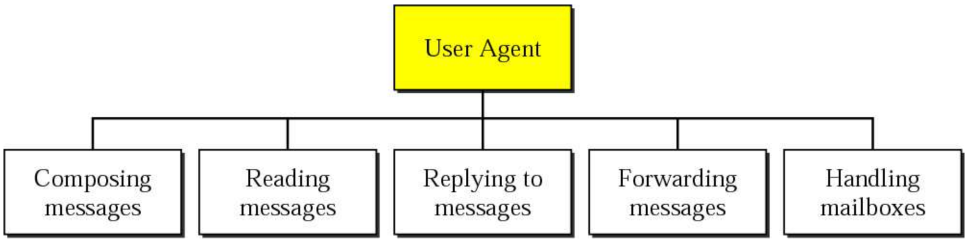


Fig. 5.14 *Functions of User Agent*

NOTES

Simple Mail Transfer Protocol

The most important component of the TCP/IP electronic mail system is the *Simple Mail Transfer Protocol* (SMTP). Derived from Mail Transfer Protocol (MTP), SMTP is the mechanism used for the delivery of mail between TCP/IP systems and users. The only part of the e-mail system for which SMTP is not used is the final retrieval step by an e-mail recipient.

In the early days of SMTP, mail was delivered using the relatively inefficient process of relaying from server to server across the internetwork. Today, when an SMTP server has mail to deliver to a user, it determines the server that handles the user's mail using the Domain Name System (DNS) and sends the mail to that server directly.

SMTP servers both send and receive e-mail; the device sending mail acts as a client for that transaction; the one receiving it acts as a server (Refer Figure 5.15). To avoid confusion, it is easier to refer to the device sending e-mail as the *SMTP sender* and the one receiving as the *SMTP receiver*; these were the terms used when SMTP was originally created.

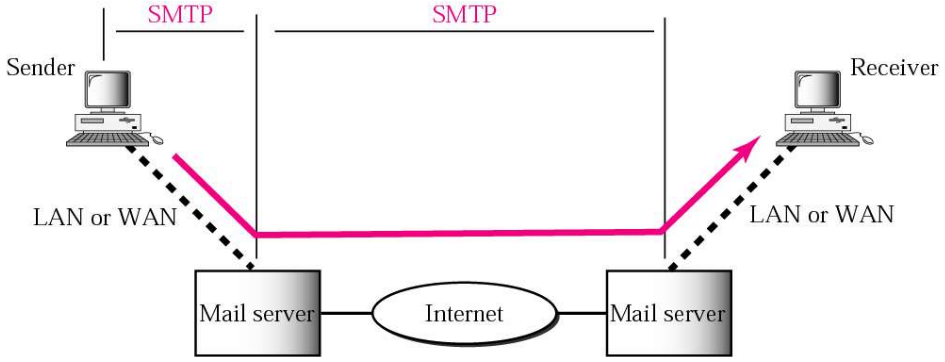


Fig. 5.15 *Simple Mail Transfer Protocol (SMTP)*

SMTP Connection and Session Establishment and Termination

The delivery of electronic mail using the Simple Mail Transfer Protocol involves the regular exchange of e-mail messages between SMTP servers. SMTP servers are responsible for sending e-mail that users submit for delivery. They also receive e-mail either intended for local recipients, or in some cases for forwarding or relaying to other servers.

An SMTP session consists of three basic phases (Refer Figure 5.16):

1. First, the session is **established** through the creation of a TCP connection and the exchange of identity information between the SMTP sender and receiver using the *HELO* command.
2. Once the session is established, **mail transactions** can be performed.
3. Finally, when the SMTP sender is done with the session, it **terminates** it using the *QUIT* command.

If SMTP extensions are supported, the SMTP sender uses the *EHLO* (extended hello) command instead of *HELO*, and the SMTP receiver replies with a list of extensions it will allow the SMTP sender to use.

SMTP mail transaction process

The delivery of e-mail message begins with the establishment of an SMTP session between the devices sending and receiving the message. The SMTP sender initiates a TCP connection to the SMTP receiver, and then sends a *HELO* or *EHLO* command, to which the receiver responds. Assuming there are no problems, the session is then established and ready for actual e-mail message transactions.

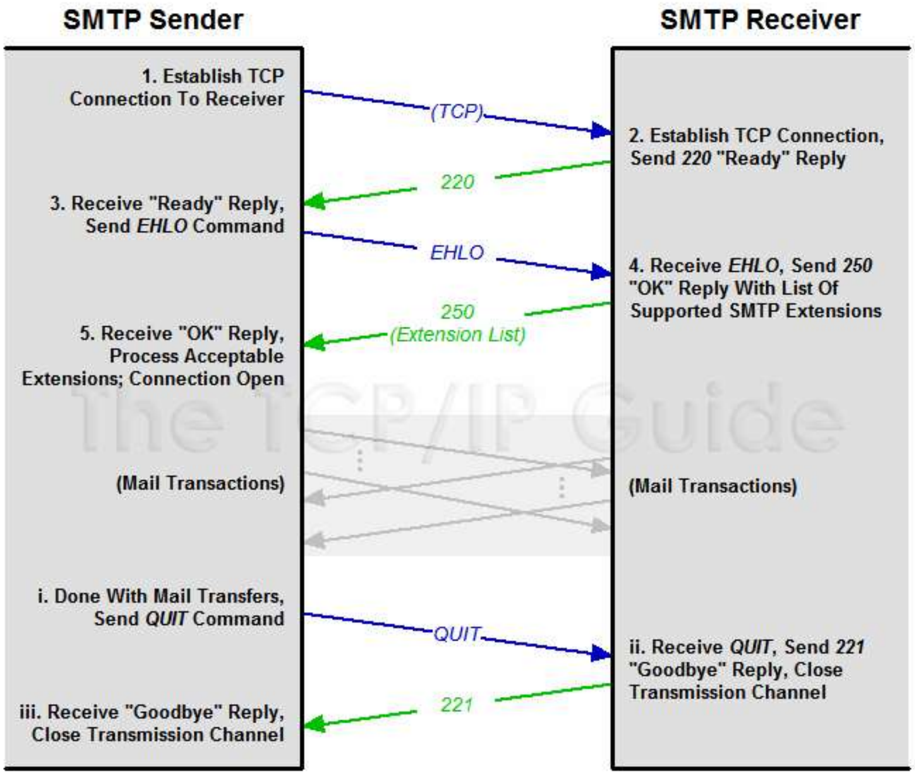

SMTP Sender
SMTP Receiver
1. Establish TCP Connection To Receiver
(TCP)
2. Establish TCP Connection, Send 220 "Ready" Reply
220
3. Receive "Ready" Reply, Send EHLO Command
EHLO
4. Receive EHLO, Send 250 "OK" Reply With List Of Supported SMTP Extensions
250 (Extension List)
5. Receive "OK" Reply, Process Acceptable Extensions; Connection Open
The TCP/IP Guide
(Mail Transactions)
(Mail Transactions)
i. Done With Mail Transfers, Send QUIT Command
QUIT
ii. Receive QUIT, Send 221 "Goodbye" Reply, Close Transmission Channel
221
iii. Receive "Goodbye" Reply, Close Transmission Channel

Fig. 5.16 *SMTP Session*

SMTP Mail Transaction Overview

The SMTP mail transaction process consists of three steps:

1. **Transaction initiation and sender identification:** The SMTP sender tells the SMTP receiver that it wants to start sending a message, and gives the receiver the e-mail address of the message's originator.
2. **Recipient identification:** The sender tells the receiver the e-mail address(es) of the intended recipients of the message.
3. **Mail transfer:** The sender transfers the e-mail message to the receiver. This is a complete e-mail message meeting the RFC 822 specification (which may be in MIME format as well) (Refer Figure 5.16).

SMTP Security Issues

The base protocol does not include any security mechanism as Internet security was not an issue in the times when SMTP was designed. But with the change in current scenario, e-mail is so often abused today, most modern SMTP servers incorporate one or more security features to avoid any problem.

SMTP Commands

The SMTP sender performs operations using a set of *SMTP commands*. Each command is identified using a four-letter code. Since SMTP only supports a limited number of functions, it has a small command set (Refer Figure 5.17).

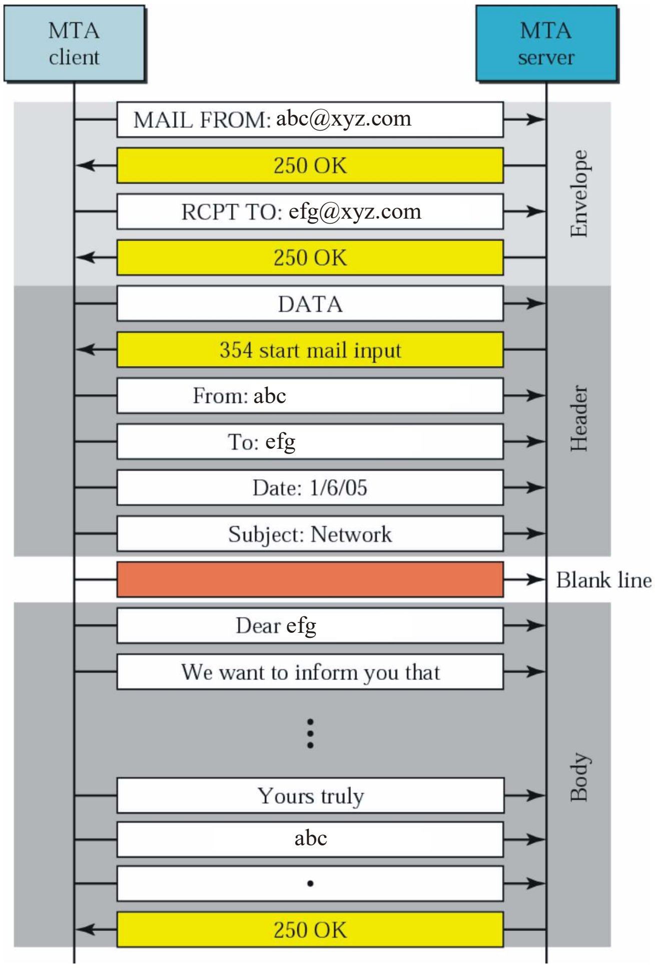


Fig. 5.17 *SMTP Mail Transaction*

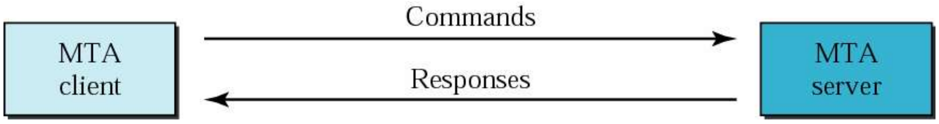


Fig. 5.18 *SMTP Commands and Responses*

- Command format

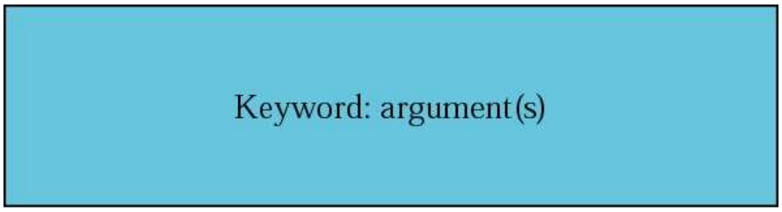


Fig. 5.19 *Command Format*

- Commands

Keyword	*Argument(s)*
HELO	Sender's host name
MAIL FROM	Sender of the message
RCPT TO	Intended recipient of the message
DATA	Body of the mail
QUIT	
RSET	
VRFY	Name of recipient to be verified
NOOP	
TURN	
EXPN	Mailing list to be expanded
HELP	Command name
SEND FROM	Intended recipient of the message
SMOL FROM	Intended recipient of the message
SMAL FROM	Intended recipient of the message

Fig. 5.20 *Commands*

SMTP responses

Each time the SMTP sender issues a command, it receives a *reply* from the SMTP receiver (Refer Figure 5.21). These replies are similar to FTP replies, and uses both a three-digit reply code and a descriptive text line. A special *enhanced status codes* SMTP extension is also defined; when enabled, this causes the SMTP receiver to return more detailed result information after processing a command.

Code	*Description*
Positive Completion Reply	
211	System status or help reply
214	Help message
220	Service ready
221	Service closing transmission channel
250	Request command completed
251	User not local; the message will be forwarded
Positive Intermediate Reply	
354	Start mail input
Transient Negative Completion Reply	
421	Service not available
450	Mailbox not available
451	Command aborted: local error
452	Command aborted; insufficient storage

Permanent Negative Completion Reply	
500	Syntax error; unrecognized command
501	Syntax error in parameters or arguments
502	Command not implemented
503	Bad sequence of commands
504	Command temporarily not implemented
550	Command is not executed; mailbox unavailable
551	User not local
552	Requested action aborted; exceeded storage location
553	Requested action not taken; mailbox name not allowed
554	Transaction failed

Fig. 5.21 *SMTP Responses*

Mail Access Protocols

For flexibility, TCP/IP uses a variety of mailbox access and retrieval protocols and methods to allow users to read e-mail. Three different models describe how these different methods work:

- The *online* model, in which e-mail is accessed and read on the server.
- The *offline* model, in which mail is transferred to the client device and used there.
- The *disconnected* model, where mail is retrieved and read offline but remains on the server with changes synchronized for consistency.

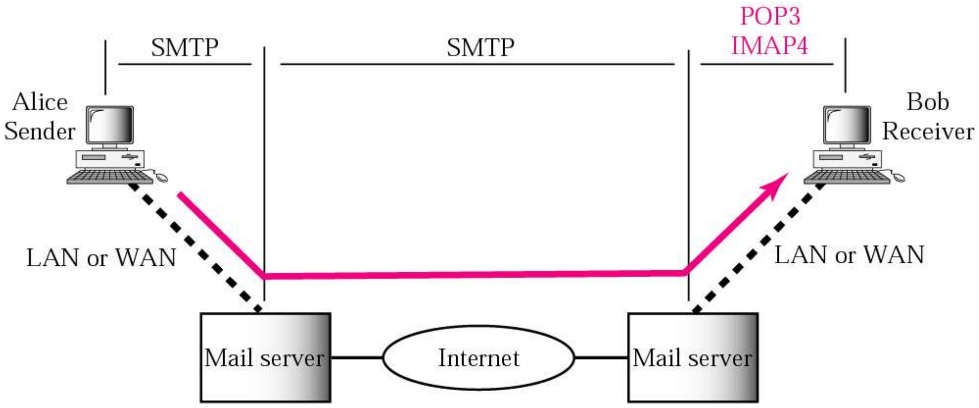


Fig. 5.22 *POP and IMAP*

TCP/IP Post Office Protocol (POP/POP3)

The *Post Office Protocol (POP)* is currently the most popular TCP/IP e-mail access and retrieval protocol. It implements the offline access model, allowing users to retrieve mail from their SMTP server and use it on their local client computers. It is specifically designed to be a very simple protocol and has only small number of commands. The current revision of POP is version 3, and the protocol is usually abbreviated *POP3* for that reason (Refer Figure 5.22).

NOTES

- POP3 General Operation, Client/Server Communication and Session States

POP3 is a regular TCP/IP client/server protocol. To provide access to mailboxes, POP3 server software must be installed and continuously running on the server where the mailboxes are located. POP3 uses the Transmission Control Protocol (TCP) for communication, to ensure the reliable transfer of commands, responses and message data. POP3 servers 'listen' on well-known port number 110 for incoming connection requests from POP3 clients. After a TCP connection is established, the POP3 session is activated. The client sends commands to the server, which replies with responses and/or e-mail message contents (Figure 5.23).

POP3 is a client/server protocol that is described using a simple linear sequence of states.

1. ***Authorization*** **State:** The server provides a greeting to the client to indicate that it is ready for commands. The client then provides authentication information to allow access to the user's mailbox. By default, POP3 uses only a simple user name/password authentication method.
2. ***Transaction*** **State:** The client is allowed to perform various operations on the mailbox. These include listing and retrieving messages, and marking retrieved messages for deletion. The client normally begins by first retrieving statistics about the mailbox from the server, and obtaining a list of the messages in the mailbox. The client then retrieves each message one at a time, marks each retrieved message for deletion on the server.
3. ***Update*** **State:** When the client is done with all of its tasks and issues the *QUIT* command, the session enters this state automatically, where the server actually deletes the messages marked for deletion in the *Transaction* state. This concludes the session and the TCP connection between the two is terminated.

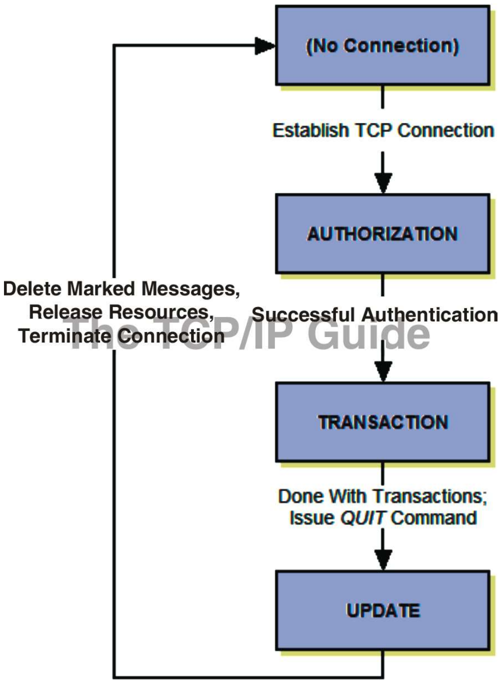

(No Connection)
Establish TCP Connection
AUTHORIZATION
Successful Authentication
TRANSACTION
Done With Transactions; Issue QUIT Command
UPDATE
Delete Marked Messages, Release Resources, Terminate Connection
The TCP/IP Guide

Fig. 5.23 *POP General Operation*

NOTES

TCP/IP Internet Mail Access Protocol (IMAP/IMAP4)

The Post Office Protocol is popular because of its simplicity and long history, but POP has few features and normally only supports the rather limited *offline* mail access method. To provide more flexibility for users in how they access, retrieve and work with e-mail messages, the *Internet Message Access Protocol* (*IMAP*) was developed. IMAP is primarily used in the online and disconnected access models; it allows users to access mail from many different devices, manage multiple mailboxes, select only certain messages for downloading, and much more. Due to its many capabilities, it is growing in popularity.

IMAP general operation, client/server communication and session states

IMAP is a client/server application, and an IMAP session begins with the client making a TCP connection to the server (Figure 5.23).

The session between an IMAP4 client and server is described in the IMAP standards.

The following are the IMAP states, in the usual sequence in which they occur for a session:

1. ***Not Authenticated* State:** The session normally begins in this state after a TCP connection is established; unless the special IMAP *preauthentication* feature has been used (we will get to this feature shortly). The client at this point cannot really do much aside from providing authentication information so it can move to the next state.
2. ***Authenticated* State:** The client has completed authentication, either through an authentication process in the prior state or through preauthentication. The client is now allowed to perform operations on whole mailboxes. The client must select a mailbox before individual message operations are permitted.
3. ***Selected* State:** After a mailbox has been chosen, the client is allowed to access and manipulate individual messages within the mailbox. When the client is done with the current mailbox it can close it and return to the *Authenticated* state to select a new one to work with, or can log out to end the session.
4. ***Logout* State:** The client may issue a *Logout* command from any of the other states to request that the IMAP session be ended. The session may also enter this state if the session inactivity timer expires. The server sends a response and the connection is terminated.

5.10.1 Sending E-Mails via Internet

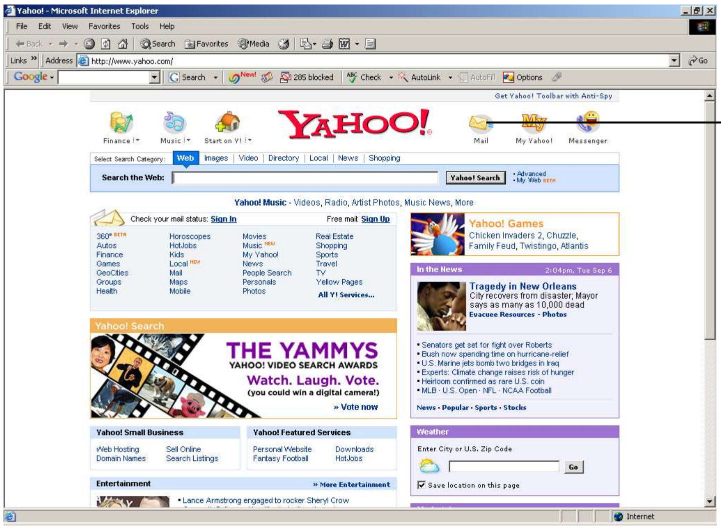
E-mail, short for electronic mail, enables you to send your correspondence instantaneously anywhere in the world via the Internet. E-mail has made the world a ‘smaller place’.

The popularity of e-mailing is because of its capability to send and receive messages anytime, anywhere without any cost. An e-mail allows you to send and receive a variety of file types such as text, image, video, sound and graphics to a single recipient or multiple recipients using broadcasting. To use the e-mail feature, you just need to create an e-mail account for yourself using a website that offers such services. Various sites provide the e-mail facility. Some of them such as Yahoo.com, Rediff.com, hotmail.com and lycos.com provide it free of cost while others charge for it.

Since by now you would be quite keen to use the e-mail facility, let us run through the process of creating and using an e-mail account on Yahoo.com.

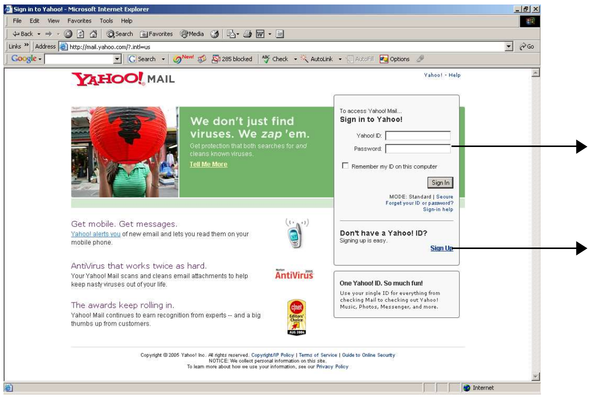
Creating a User ID

Type the URL '**http://www.yahoo.com**' in the address bar of a Web browser such as Internet Explorer to visit the Yahoo homepage.


Yahoo! - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.yahoo.com/
YAHOO!
Mail
My Yahoo!
Messenger
Search the Web:
Yahoo! Search
THE YAMMYS
Internet

Mail

The page that is now displayed is the 'Sign in' page. If you are an existing user, you need to type in your user id and password to log on to your account. If you are a first time user, you need to first create an account for yourself.

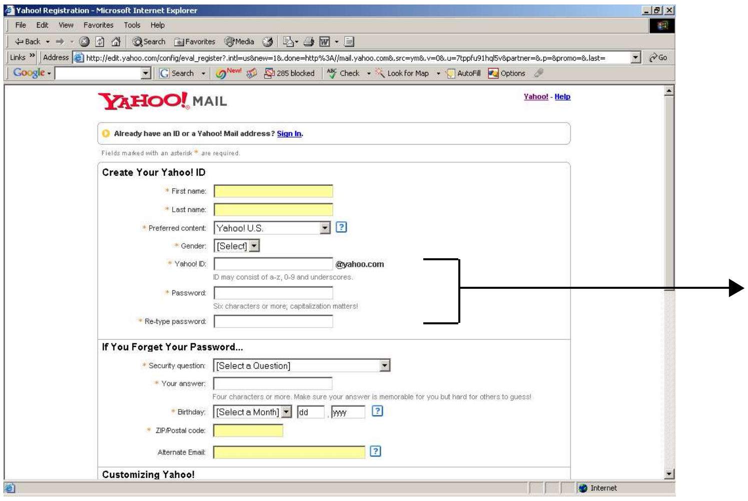

Sign in to Yahoo! - Microsoft Internet Explorer
Address http://mail.yahoo.com/?.intl=us
YAHOO! MAIL
We don't just find viruses. We zap 'em.
Sign in to Yahoo!
Yahoo! ID:
Password:
Remember my ID on this computer
Sign In
Don't have a Yahoo! ID?
Sign Up
Internet

Existing
User

New User

Click 'Sign Up' to create a new user ID. The page that is displayed is a registration form that requires you to fill in your details along with the user ID and password for your new e-mail account.

NOTES


Yahoo! Registration - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Links Address http://edit.yahoo.com/config/eval_register?.intl=us&new=1&.done=http%3A//mail.yahoo.com&.src=ym&.v=0&.u=7tppfu91hql5v&partner=&.p=&promo=&.last= Go
Google Search New! 285 blocked Check Look for Map AutoFill Options
YAHOO! MAIL
Yahoo! - Help
Internet

Already have an ID or a Yahoo! Mail address? Sign In.

Fields marked with an asterisk * are required.

Create Your Yahoo! ID

* First name:
* Last name:
* Preferred content: Yahoo! U.S.
* Gender: [Select]
* Yahoo! ID: @yahoo.com
ID may consist of a-z, 0-9 and underscores.
* Password:
Six characters or more; capitalization matters!
* Re-type password:

If You Forget Your Password...

* Security question: [Select a Question]
* Your answer:
Four characters or more. Make sure your answer is memorable for you but hard for others to guess!
* Birthday: [Select a Month] dd , yyyy
* ZIP/Postal code:
Alternate Email:

Customizing Yahoo!

User Id and Password

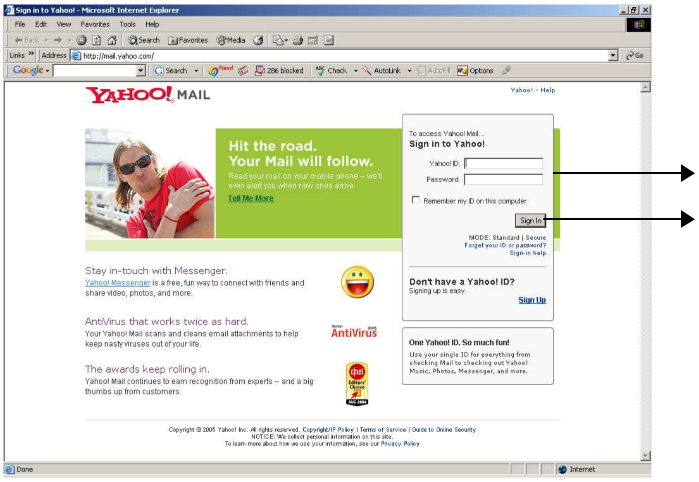
Once you have registered yourself on a website, you become a member and can simply log on to your mail account to start sending and receiving e-mails. For all future access, you would have to remember your user ID and password because that is the key to your login.

Checking Your E-Mail

You can access your e-mail anytime by logging on to your mail. To do so, carry out the following steps:

– Type the URL 'http://www.yahoo.com' in the address bar of a Web browser.

– Enter your user ID and password.


Sign in to Yahoo! - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Links Address http://mail.yahoo.com/ Go
Google Search New! 286 blocked Check AutoLink AutoFill Options
YAHOO! MAIL
Yahoo! - Help
AntiVirus
cnet Editors' Choice AUG 2004
Done
Internet

Hit the road.
Your Mail will follow.
Read your mail on your mobile phone -- we'll even alert you when new ones arrive.
Tell Me More

To access Yahoo! Mail...
Sign in to Yahoo!
Yahoo! ID:
Password:
□ Remember my ID on this computer
Sign In
MODE: Standard | Secure
Forget your ID or password?
Sign-in help

Don't have a Yahoo! ID?
Signing up is easy.
Sign Up

One Yahoo! ID. So much fun!
Use your single ID for everything from checking Mail to checking out Yahoo! Music, Photos, Messenger, and more.

Stay in-touch with Messenger.
Yahoo! Messenger is a free, fun way to connect with friends and share video, photos, and more.

AntiVirus that works twice as hard.
Your Yahoo! Mail scans and cleans email attachments to help keep nasty viruses out of your life.

The awards keep rolling in.
Yahoo! Mail continues to earn recognition from experts -- and a big thumbs up from customers.

Copyright © 2005 Yahoo! Inc. All rights reserved. Copyright/IP Policy | Terms of Service | Guide to Online Security
NOTICE: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy

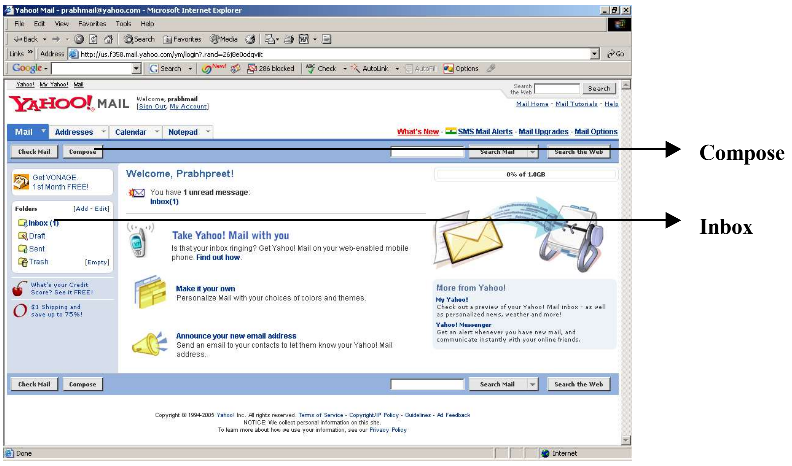
User Id and Password

Sign In

Once you have signed in successfully, you can access your e-mail account. You can access your '**Inbox**' to view any incoming mail, or write a new mail through the **Compose Mail** option.

Clicking the Inbox button displays all the received messages or mails.

Clicking the e-mail subject displays the contents of the e-mail that can be read to take necessary action.


Yahoo! Mail - prabhmail@yahoo.com - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Links Address http://us.f358.mail.yahoo.com/ym/login?.rand=26j8e0odqviit Go
Google Search New! 286 blocked Check AutoLink AutoFill Options
Yahoo! My Yahoo! Mail
Search the Web Search
YAHOO! MAIL
Welcome, prabhmail [Sign Out, My Account]
Mail Home - Mail Tutorials - Help
Mail Addresses Calendar Notepad
What's New - SMS Mail Alerts - Mail Upgrades - Mail Options
Check Mail Compose Search Mail Search the Web
Get VONAGE. 1st Month FREE!
Folders [Add - Edit]
Inbox (1)
Draft
Sent
Trash [Empty]
What's your Credit Score? See it FREE!
$1 Shipping and save up to 75%!
0% of 1.0GB
Done Internet
Compose
Inbox

Welcome, Prabhpreet!

You have **1 unread message**: Inbox(1)

Take Yahoo! Mail with you
Is that your inbox ringing? Get Yahoo! Mail on your web-enabled mobile phone. **Find out how**.

Make it your own
Personalize Mail with your choices of colors and themes.

Announce your new email address
Send an email to your contacts to let them know your Yahoo! Mail address.

More from Yahoo!
My Yahoo!
Check out a preview of your Yahoo! Mail inbox - as well as personalized news, weather and more!
Yahoo! Messenger
Get an alert whenever you have new mail, and communicate instantly with your online friends.

Copyright © 1994-2005 Yahoo! Inc. All rights reserved. Terms of Service - Copyright/IP Policy - Guidelines - Ad Feedback
NOTICE: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy

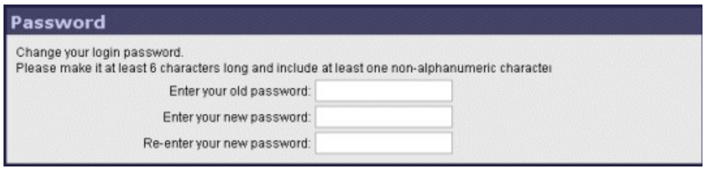
Changing the Password

An e-mail password is used for security reasons. A password is the personal code of a user and should not be disclosed to anyone. Hence, never give your password to anyone and do not write it down where someone else may find it. You can change your password in the following manner:

Click on **Options** on the top menu bar.

Click on **Password** under **Your Information**.

The **Change Mail Password** window should appear:



Password

Change your login password.
Please make it at least 6 characters long and include at least one non-alphanumeric characte

Enter your old password:
Enter your new password:
Re-enter your new password:

In the **Old Password** box, type in your current password.

In both the **New Password** and **Re-enter New Password** boxes, enter what you want your new password to be.

Click on the **Save Options** button.

A password should be at least eight characters in length.

This will only change your e-mail account password, not the password you use to log on to the computer.

Composing and Sending E-mails

The Compose option on the left corner of your screen allows you to write an e-mail message. You can also attach documents to your mail. When you select the Compose option, the following screen appears.

You can use the following option while composing or writing an e-mail message.

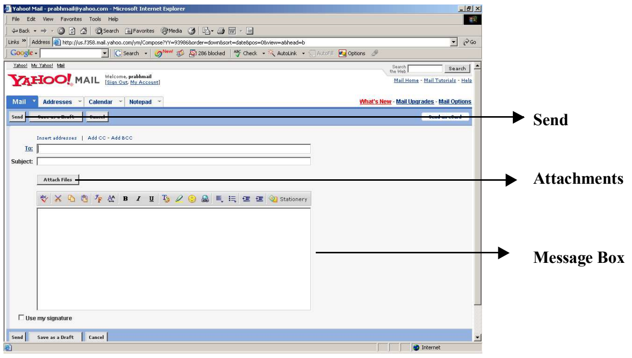
- **To:** Specifies the e-mail address of a recipient such as recipient@domain.com and user@abcdomain.com**.** This should be a valid email id for the delivery of your message. You can specify multiple recipients' addresses separated by commas.
- **Cc:** Specifies the address of the recipient to whom you want to send the carbon copy (cc) of your message. You can specify multiple

NOTES

recipients' addresses separated by commas.

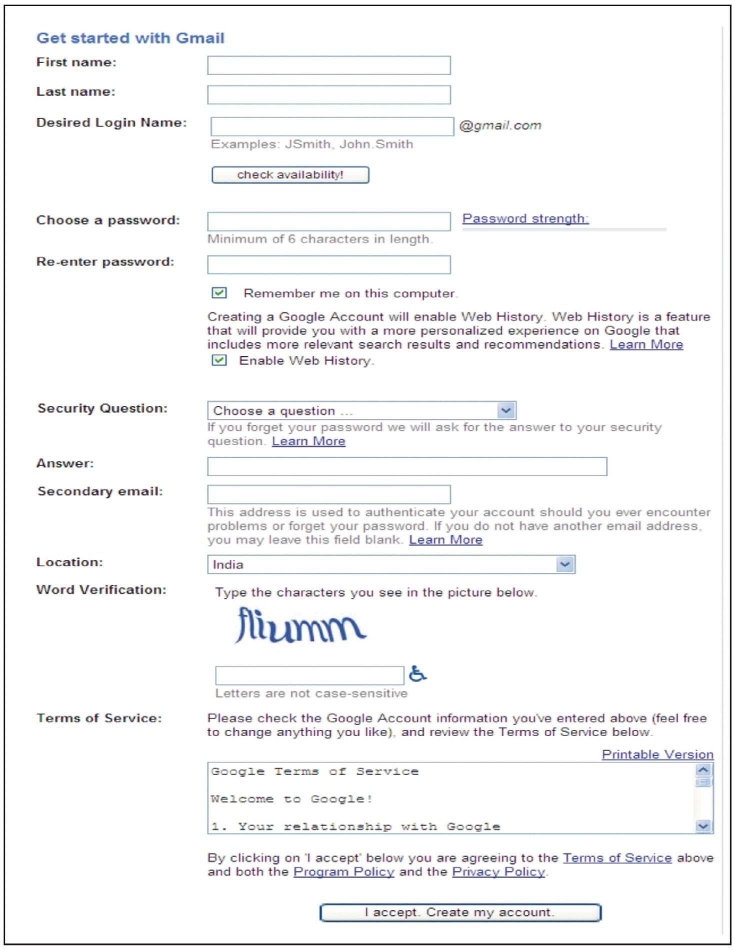
– **Subject:** Refers to the subject of the e-mail message. It provides a fair idea to the recipient about what the mail contains.

– **Message Box:** Provides a text area for composing e-mail content.



E-Mailing with Google

- Open the web browser and type the name of the e-mail site on which you would like to create an account. For example, gmail.com.
- Type in the following URL: http://mail.google.com/mail/signup in the address bar.
- Follow the steps mentioned in the web page in order to successfully create your e-mail account. The following page will be displayed.



Get started with Gmail

First name:

Last name:

Desired Login Name: @gmail.com
Examples: JSmith, John.Smith

check availability!

Choose a password: Password strength:
Minimum of 6 characters in length.

Re-enter password:

☑ Remember me on this computer.

Creating a Google Account will enable Web History. Web History is a feature that will provide you with a more personalized experience on Google that includes more relevant search results and recommendations. Learn More
☑ Enable Web History.

Security Question: Choose a question ...
If you forget your password we will ask for the answer to your security question. Learn More

Answer:

Secondary email:
This address is used to authenticate your account should you ever encounter problems or forget your password. If you do not have another email address, you may leave this field blank. Learn More

Location: India

Word Verification: Type the characters you see in the picture below.
Letters are not case-sensitive

Terms of Service: Please check the Google Account information you've entered above (feel free to change anything you like), and review the Terms of Service below.

Printable Version

Google Terms of Service

Welcome to Google!

1. Your relationship with Google

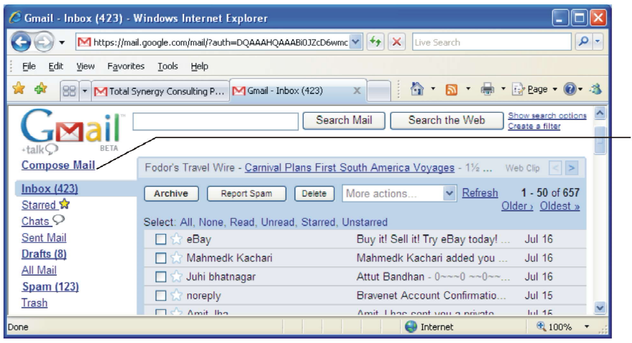
By clicking on 'I accept' below you are agreeing to the Terms of Service above and both the Program Policy and the Privacy Policy.

I accept. Create my account.

NOTES

Sending and Receiving E-Mails

① Click on **Compose Mail** to create a new message.



Gmail - Inbox (423) - Windows Internet Explorer

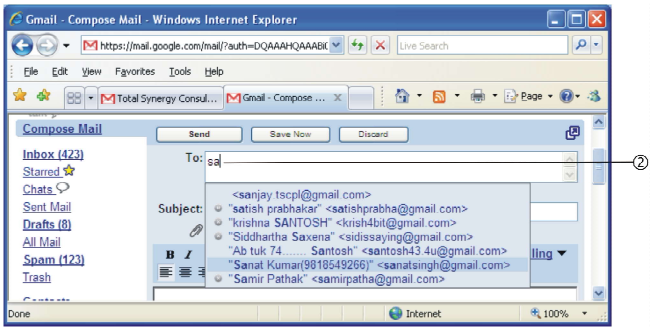
https://mail.google.com/mail/?auth=DQAAAHQAAABi0JZcD6wmc

Fodor's Travel Wire - Carnival Plans First South America Voyages - 1½ ...

1 - 50 of 657

eBay	Buy it! Sell it! Try eBay today! ...	Jul 16
Mahmedk Kachari	Mahmedk Kachari added you ...	Jul 16
Juhi bhatnagar	Attut Bandhan - 0~~~0 ~~0~~...	Jul 16
noreply	Bravenet Account Confirmatio...	Jul 15

② You can select addresses from your Contacts list or type the address in the **To**:, **Cc**:, or **Bcc**: fields. When you begin to type an address in these fields, a complete address will be suggested from your **Contacts** list.



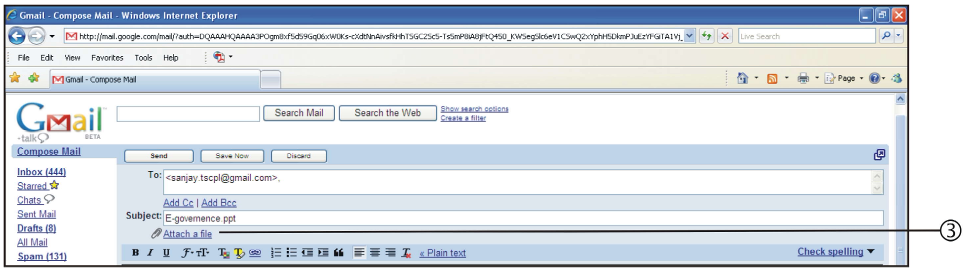
Gmail - Compose Mail - Windows Internet Explorer

https://mail.google.com/mail/?auth=DQAAAHQAAABi(

To: sa

<sanjay.tscpl@gmail.com>
"satish prabhakar" <satishprabha@gmail.com>
"krishna SANTOSH" <krish4bit@gmail.com>
"Siddhartha Saxena" <sidissaying@gmail.com>
"Ab tuk 74....... Santosh" <santosh43.4u@gmail.com>
"Sanat Kumar(9818549266)" <sanatsingh@gmail.com>
"Samir Pathak" <samirpatha@gmail.com>

③ Select the **Attach a file** link in order to attach any file with the e-mail message. (The figure shows Attach a file. Attach another file is displayed only after the first file has been attached).



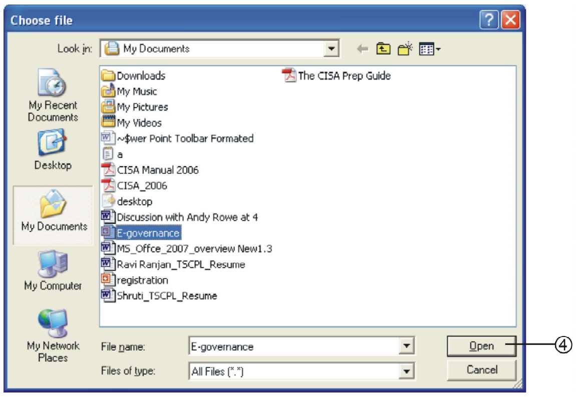
Gmail - Compose Mail - Windows Internet Explorer

http://mail.google.com/mail/?auth=DQAAAHQAAAA3POgm8xf5d59Gq06xW0Ks-cXdtNnAivsfkHhTSGC2Sc5-Ts5mP8iA8jFtQ450_KWSegSlc6eV1CSwQ2xYphH5DkmPJuEzYFGiTA1Vj_

To: <sanjay.tscpl@gmail.com>,

Subject: E-governence.ppt

④ Select the file you want to attach. Then click on **Open**.



Choose file

Look in: My Documents

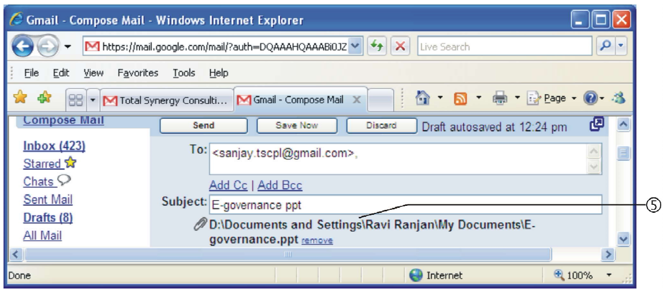
Downloads
My Music
My Pictures
My Videos
~$wer Point Toolbar Formated
a
CISA Manual 2006
CISA_2006
desktop
Discussion with Andy Rowe at 4
E-governance
MS_Offce_2007_overview New1.3
Ravi Ranjan_TSCPL_Resume
registration
Shruti_TSCPL_Resume
The CISA Prep Guide

File name: E-governance

Files of type: All Files (*.*)

NOTES

⑤ Your file will now be attached to your e-mail message.


Gmail - Compose Mail - Windows Internet Explorer
https://mail.google.com/mail/?auth=DQAAAHQAAABi0JZ
Live Search
File Edit View Favorites Tools Help
Total Synergy Consulti...
Gmail - Compose Mail
Page
Compose Mail
Send
Save Now
Discard
Inbox (423)
Starred
Chats
Sent Mail
Drafts (8)
All Mail
Add Cc | Add Bcc
remove
Done
Internet
100%
⑤

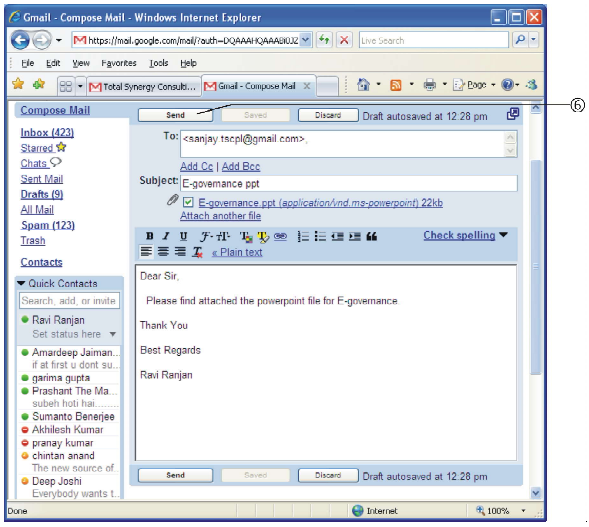
Draft autosaved at 12:24 pm

To: <sanjay.tscpl@gmail.com>,

Subject: E-governance ppt

D:\Documents and Settings\Ravi Ranjan\My Documents\E-governance.ppt

⑥ Now click on the **Send** button to send the e-mail.



Draft autosaved at 12:28 pm

To: <sanjay.tscpl@gmail.com>,

Subject: E-governance ppt

☑ E-governance.ppt (application/vnd.ms-powerpoint) 22kb

Dear Sir,

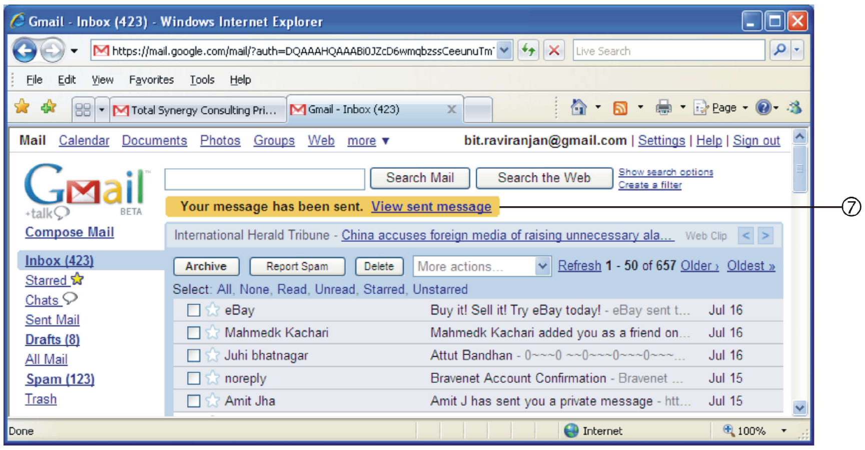
Please find attached the powerpoint file for E-governance.

Thank You

Best Regards

Ravi Ranjan

⑦ You can now see that the message has been sent.


Gmail - Inbox (423) - Windows Internet Explorer
https://mail.google.com/mail/?auth=DQAAAHQAAABi0JZcD6wmqbzssCeeunuTm
Live Search
File Edit View Favorites Tools Help
Total Synergy Consulting Pri...
Gmail - Inbox (423)
Page
Mail Calendar Documents Photos Groups Web more ▾
Settings | Help | Sign out
Gmail +talk BETA
Search Mail
Search the Web
Show search options
Create a filter
Compose Mail
Web Clip
Inbox (423)
Starred
Chats
Sent Mail
Drafts (8)
All Mail
Spam (123)
Trash
Archive
Report Spam
Delete
More actions...
Refresh
Older ›
Oldest »
Select: All, None, Read, Unread, Starred, Unstarred
Done
Internet
100%
⑦

bit.raviranjan@gmail.com

Your message has been sent. View sent message

International Herald Tribune - China accuses foreign media of raising unnecessary ala...

1 - 50 of 657

□ eBay	Buy it! Sell it! Try eBay today! - eBay sent t...	Jul 16
□ Mahmedk Kachari	Mahmedk Kachari added you as a friend on...	Jul 16
□ Juhi bhatnagar	Attut Bandhan - 0~~~0 ~~0~~~0~~~0~~~...	Jul 16
□ noreply	Bravenet Account Confirmation - Bravenet ...	Jul 15
□ Amit Jha	Amit J has sent you a private message - htt...	Jul 15

Receiving E-Mails

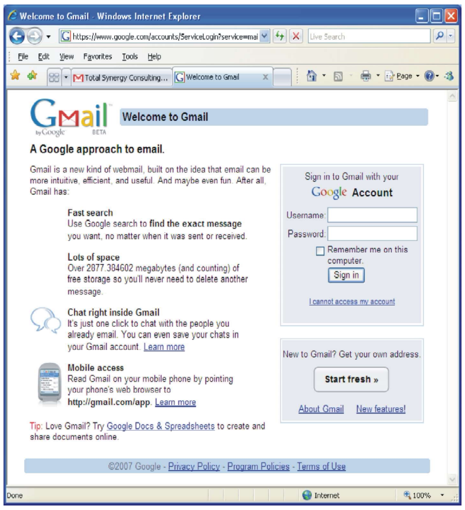
You can check the received mails by clicking on the **Inbox** tab.

Some Popular Websites

Given below are the names of some popular free e-mail websites.

1. www.gmail.com
2. www.mail.yahoo.com
3. www.hotmail.com

4. www.rediffmail.com
5. www.indiatimes.com


Welcome to Gmail - Windows Internet Explorer
https://www.google.com/accounts/ServiceLogin?service=mai
Live Search
File Edit View Favorites Tools Help
Total Synergy Consulting...
Welcome to Gmail
Page
Gmail by Google BETA
Username:
Password:
Sign in
Start fresh »
Done
Internet
100%

Welcome to Gmail

A Google approach to email.

Gmail is a new kind of webmail, built on the idea that email can be more intuitive, efficient, and useful. And maybe even fun. After all, Gmail has:

Fast search
Use Google search to **find the exact message** you want, no matter when it was sent or received.

Lots of space
Over 2877.384602 megabytes (and counting) of free storage so you'll never need to delete another message.

Chat right inside Gmail
It's just one click to chat with the people you already email. You can even save your chats in your Gmail account. Learn more

Mobile access
Read Gmail on your mobile phone by pointing your phone's web browser to **http://gmail.com/app**. Learn more

Tip: Love Gmail? Try Google Docs & Spreadsheets to create and share documents online.

Sign in to Gmail with your Google Account

□ Remember me on this computer.

I cannot access my account

New to Gmail? Get your own address.

About Gmail New features!

©2007 Google - Privacy Policy - Program Policies - Terms of Use

Popular Networking Sites

Some of the networking and community websites include:

1. www.orkut.com
2. www.fropper.com

5.10.2 Some Important Features of E-Mail Services Available on The Internet

Following are some of the significant features of e-mail services that are available on the Internet for the Web e-mail users:

- Automatic reply to messages
- Auto-forward and redirection of messages
- Facility to send copies of a message to many people
- Automatic filing and retrieval of messages
- Addresses can be stored in an address book and retrieved instantly
- Notification if a message cannot be delivered
- E-mails are automatically date and time stamped
- Signatures can be attached
- Files, graphics or sound can be sent as attachments, often in compressed formats
- Webmail and mobile email can be used to receive and send messages while on the move

Check Your Progress

10. Write the main feature of data encryption.

11. What is authentication?

13. Define internet privacy.

13. What are the two parts of a message?

14. Why an E-mail password is used?

NOTES

5.11 ELECTRONIC COMMERCE

These days, e-commerce uses electronic technology for its high growth; thus there is a high demand for the latter. It is well-known that computer increases our capacity to store, search and retrieve information. With the tremendous growth in computer usage for communication and other purposes, people from various fields are forming virtual society on the Internet. The concept is quite simple; if one has access to a personal computer (PC) and can connect it to the Internet with a browser he/she can do an online business. You have to just get on the Web, open an online store, and watch your business grow. This wired world of business, where technology, human talent, and a new method of doing business, make up today's growing worldwide economy. The backbone of this electronic commerce is the Internet. E-commerce is not only about technology; it is also about information, decision-making and communication. Use of e-commerce refers to purchase, or sale, advertising and servicing of goods or services over the Internet. Currently though not big enough as compared to traditional peer markets, E-commerce is expected to grow in the near future.

According to a survey, the e-commerce industry in India is expected to grow very quickly. The total number of Internet users are rising very rapidly. Worldwide, the growth of e-commerce has gained popularity due to online shopping but this has not happened in the case of the Indian market. Here, it is mainly concentrated on online travel and the banking sector.

However, growth of the industry is expected to go up very high in the near future that will include both the Internet and mobile banking users.

5.11.1 Internet: A Tool for Electronic Commerce

E-business deals with the buying and selling of information, products and services through the computer network. E-business is also defined as a business activity which uses an electronic medium. It also refers to the buying or selling of goods and services without visiting a store. E-business involves activities, such as the delivery of information, products, services and payment through the electronic medium. In addition, e-business refers to paperers business activities, such as supply chain management, enterprise resource planning, customer relationship management, and knowledge management.

In the 1950s, computers were used by organizations to process and store records of internal transactions. However, information between various businesses continued to be exchanged on paper, like purchase orders, invoices, cheques, remittance devices and other standard forms, which were used to document transactions. IBM was the first company which used the term 'e-business' internationally. In 1972, IBM used the term 'e-business' and the first successful transaction was executed between the US and the European Union in 1993, with the invention of personal computers.

History of the Internet

1969	: The US Department of Defense started the first network among major research centres in the US.
1971	: Major connections or nodes were established. E-mail was introduced.

1973 : Defense Department started developing various forms of file transfer.

1984 : Domain Name Service (DNS) was introduced.

1986 : The US National Service Foundation created Internet-based telephone lines.

1987 : The number of hosts (computers on the Internet) reached 10,000.

1988 : The number of hosts on the Internet crossed over 60,000.

1989 : Over 100,000 hosts on the Internet were registered.

1991 : The World Wide Web (WWW) was created by CERN in Switzerland.
(Conseil European pour la Recherché Nuclearire)

1992 : One million hosts were found on the Internet.

1995 : There were a total of 6.6 million hosts or computers on the Internet.

July 1997 : 1.3 million domain names were registered.

Dec. 1997 : 22 million servers, 40 million users on the WWW.

2000 : 110 million users and 72 million domain names.

2003 : 802.2 million users and 233 hosts.

***Table 5.2** Growth of the Internet in India*[3]

Years	Internet Subscribers	Internet Users
1997	25	45000
1998	250	200000
1999	359	1000000
2000	650	2000000
2001	1130	6668000
2002	1763	10684000
2003	3661	29000000
2004	4403	31723000
2005	6674	52875000

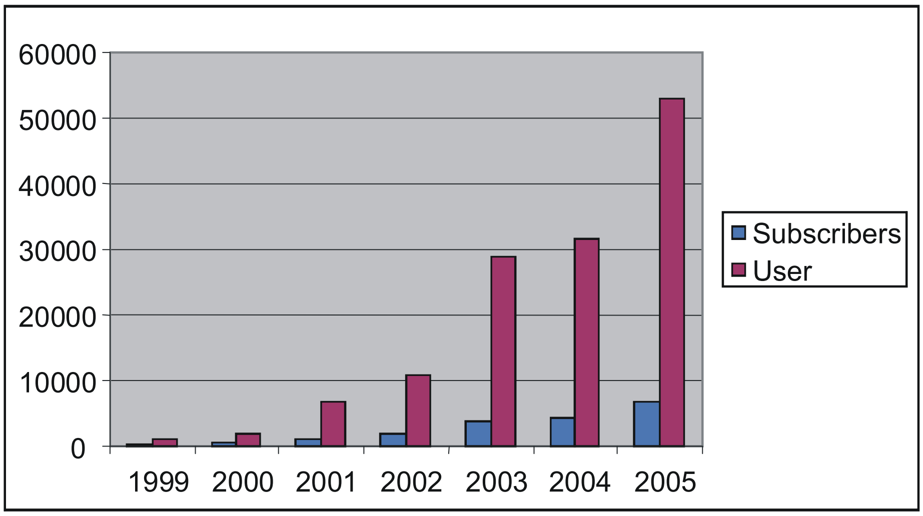


***Fig. 5.24** Internet Usage in India*[1]

Table 5.3 Growth in the Number of Hosts Over the Years[2]

Year	Number of hosts advertised in the DNS
1993	₹ 1,313,000 crore
1994	₹ 2,217,000 crore
1995	₹ 4,852,000 crore
1996	₹ 9,472,000 crore
1997	₹ 1,6146,000 crore
1998	₹ 2,967,0000 crore
1999	₹ 4,323,0000 crore
2000	₹ 72,398,092 crore
2001	₹ 109,574,429 crore
2002	₹ 147,344,723 crore
2003	₹ 171,638,297 crore
2004	₹ 233,101,481 crore

E-business opportunities for businesses

Many businesses need e-business software services to help take advantage of e-business areas.

1. **Tourism and travel sector:** This sector has updated its system with E-business services. Consumers can make online reservation of hotels, motels, air tickets, railway tickets, etc.
2. **Banking sector:** Most banks have changed their working style by making their services available online through their respective websites.
3. **Health care sector:** This sector is a large one and uses a major part of government expenses. So, most of the health care companies communicate or exchange their services with each other.
4. **Stock sector:** In the stock exchange sector, e-business services provide Demat Account facilities for customers who can conduct an overall analysis of the status of the stock areas and carry out their respective transactions.
5. **Financial sector:** In India, this sector has adopted E-business services and the users make full use of the same.

Working of e-business

To understand the operation of e-business, consider a customer who wants to make an online purchase. He is moved to the online transaction server where entire the information is encrypted. Once he has placed his order, the information moves through a private gateway to a processing network, where the issuing and acquiring banks complete or deny the transaction. This process takes only few seconds.

NOTES

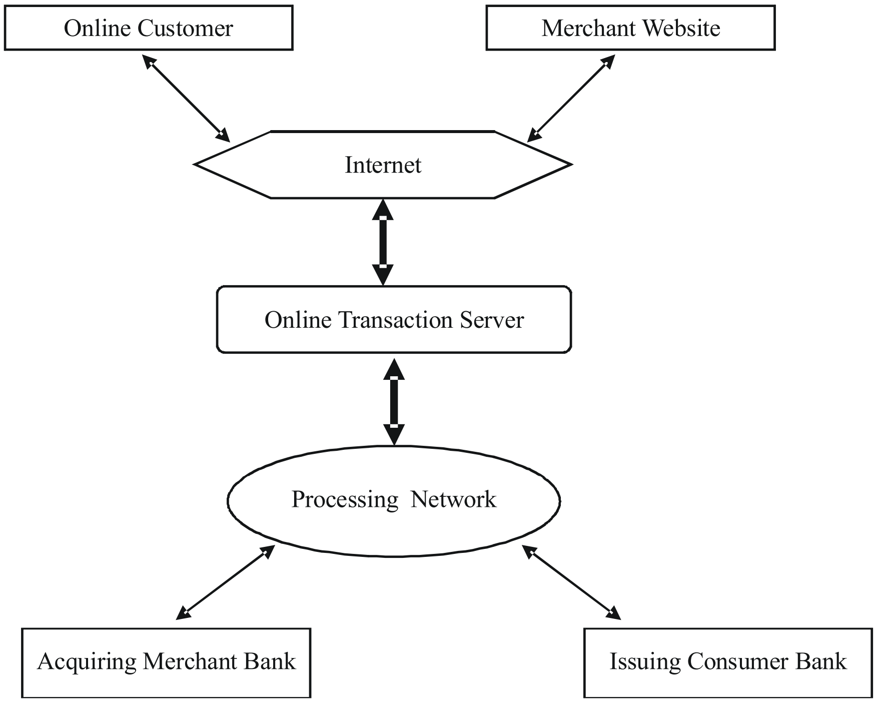


Fig. 5.25 *An Online Transaction*

Difference between e-business and traditional business mechanisms

S. No.	Basis	E-business	Traditional
1	Reduction of data error	It does not involve data at multipoints. With e-business, data goes directly from one computer to another without involving a human being.	The buyer and the seller create purchase orders on their systems, print them or e-mail them to the receiver. The receiver then re-enters the same information on the computer. This creates the error.
2	Reduction of cost	Initial cost of e-business is very high as compared to the paper process. However, over a period of time, it is very effective.	As time is money, time is directly linked to saving money. There is a repetition of the same work at every level. So it involves a lot of time and if there is an error, it may lead to wastage of money.
3	Reduction of paper work	E-business data in the electronic form is easy to share across the organization.	It requires re-entry of data at every level and also lot of time. So previous time is wasted in re-entering and printing reports.
4	Reduction of process cycle time	E-business reduces the processing cycle time of complete cycles as the data is entered into the system. It is a simulating process.	In the traditional system, when the buyer orders in a paper format, the data is re-entered into the seller's computer and only then processing takes place. This is time consuming and requires full commitment.

Advantages of e-business

1. **All-time processing:** Customers can use the marketplace at all times with the use of E-business services.
2. **Better service:** Customers are fully satisfied and receive better service.
3. **Removing mediators:** Customers can directly contact the suppliers and remove all mediators.
4. **Data on consumer performance:** Using the e-business services, one can understand consumer behaviour, for example, websites, products, schemes and modes of payment which are preferred by the customer.
5. **Time saving:** Customers can save time because they can purchase anything through the merchant websites.
6. **Improved customer services:** These days, consumers want better services. Therefore, E-business services offer a means of communication between the consumer and the company. The consumer can even make online complaints to a company.
7. **Origin of new business opportunity:** The biggest network between consumers and companies can lead to the origin of new business opportunities, like infinite possibilities for businesses to develop and increase their consumer base.
8. **Enhanced speed and accuracy of a product:** The usage of e-business services reduces human errors and other problems like duplication of proceedings. This perfection in speed and accuracy, plus easy access to documents and information affect the increase in production.
9. **Product cost saving:** Despite the fact that you can reduce the cost of a product by the use of e-business services, it also reduces the errors and the cost of sending the information to partners.

Other advantages

- It reduces the cost of the product.
- It reduces paper work as the entire work is done electronically.
- The product is directly supplied to the customer because all orders and enquiries are processed online. This eliminates the need for wholesellers and retailers and brings down the cost.
- Improved customer relationship is achieved by fast dissipation of information.
- E-business minimizes the time taken from order to delivery.
- Provides better, faster and effective linkage with clients.
- Enhances the organization’s product and also does a market analysis, as the organization gets feedback from the customer.
- E-business helps to create knowledge markets. Small groups inside big firms can be funded with seed money to develop new ideas.
- E-business helps people work together.
- E-business is a 24 × 7 operation and has a global reach.

Disadvantages of e-business

1. **Lack of customer awareness:** Mostly people have no knowledge about electronic communication like the Internet, computers, etc. Therefore, they are not able to transact electronically.
2. **Not for small businesses:** Small businessmen do not want to take any extra burden because they have no knowledge of e-business functions.
3. **Does not support all types of businesses:** Some types of businesses are not fit for e-business services.
4. **Legal formalities:** If you want to use e-business services in your business, you have to complete certain legal formalities like authorization and authentication.

Other disadvantages

- High risk for Internet startup organization
- E-business is not free
- Security problems
- Customer relation problems
- Data integrity problems
- Customer satisfaction problems

Goals of E-business

The main goals of e-business are to understand how the:

1. Needs of a consumer, merchant and organization can be met
2. Quality and quantity of goods can be improved
3. Speed of services can be increased

Prerequisites for E-business procedure

In order to conduct e-business, you will require:

1. A commercial website, for example, www.futurebazaar.com.
2. The product or services you want to sell through the respective websites.
3. Shopping carts or purchase order forms.
4. Current credit card account that will be accepted on e-payment.
5. An online payment gateway, if you plan to process credit cards in real time, over the Internet.
6. A secure socket layer (SSL) which will secure the gateway.

Functions of E-business

E-business applications enable various business functions and transactions to be conducted electronically. Some of the functions are discussed as follows:

E-Advertising: Advertising of information is currently the largest commercial activity on the Web. For example:

(a) A company's website contains its profile and all the information on its products and services.

(b) It displays banners that can be clicked.

(c) E-business portals like www.yahoo.com, are used for advertising.

(d) Newsgroups also provide publicity.

E-Catalogue: Web pages that the information on products or services that a company offers are available on an e-catalogue. An e-catalogue provides information on:

(a) Packaging

(b) Product attributes and characteristics

(c) Availability

(d) Payment modes

(e) Cost, etc.

E-Publishing: This sector was among the first to spend on this novel technology, especially on the Internet. E-publishing has led to several successful e-commerce endeavours, such as an independent publication through the Internet and electronic newspapers.

Online publications offer services, such as:

(a) Online reading/browsing

(b) Online search

(c) Customized information services

E-Banking: This facility offers remote banking electronically. Electronic banking is also referred to as online banking, cyber banking, home banking or virtual banking. It enables Web users to make online purchases and pay for the same, using an online-banking facility. It is cost effective, simple and available round the clock. The customers have access to several services, such as:

(a) Bill payment

(b) Electronic cheque writing

(c) Record keeping

(d) Tracking of bank account, credit cards

5.12 ELECTRONIC DATA INTERCHAGE (EDI)

Electronic Data Interchange (EDI) related with exchange of documents which are used in business electronically. In other words exchange of business document from compute to computer is known as EDI. Documents are in standard form and used to share among business partner. EDI brings many advantages like high processing speed, low cost, less errors, and improve relationship among business partners. EDI is faster than emails, fax and postal mails. EDI documents reach to receiver's computer and start processing immediately. EDI is programmed documents and handled by computer and thus the processing speed will increase. If people involved then the process is getting slow and changes of errors are also increased.

Business entities conducting business electronically are called trading partners. All the online shopping sites uses EDI to place order and inform business partner.

5.9.1 EDI Implementation

EDI Scope includes some typical steps toward a successful EDI implementation. Following are the steps required:

NOTES

- **Define the need that EDI is going to fulfil**: Any stakeholder who is interested in starting EDI exchanges of a given type in the process of identifying the best solution to fulfil the need expressed by the users. Business experts provide up-to-date input regarding the requirements to be considered while developing a certain exchange (For example, which data elements need to be implemented to successfully provide Returns Services, to comply with airline security regulations, or provide despatch accounting information).
- **Provide information about specific industry group requirements**: Throughout the implementation process, business experts provide additional documentation and support to ensure compliance with the business requirements of specific industry groups, which can exceed the technical requirements in the EDI messaging standards.
- **Plan the different steps required to implement EDI**: Assisting posts, airlines and other industry stakeholders who are planning to implement EDI for many years, providing standard and customised deployment plans with the key milestones and implementation scheduling based on the requirements and limitations of the implementing party. Describe the technical requirements and the necessary configurations for the involved systems: a Help Desk and Systems Technology team provide detailed documentation regarding the setup of an EDI exchange system, information on available system providers, and the key settings to be considered when configuring an EDI system.
- **Prepare the technical infrastructure to support the exchanges**: The Help Desk is the central point that puts in place the required technical infrastructure for the exchanges, mainly by the setup and management of EDI mailboxes and the exchanges between them.
- **Test and validate the EDI messages during the preparation phase**: Prior to release in production, users can provide test EDI messages to the Help Desk, which uses developed tools to test them and validate their compliance with the defined requirements.
- **Communicate with the receiving parties to confirm readiness and activation of EDI message exchanges**: Once the EDI exchanges are ready to be triggered, the Help Desk communicates through its extensive global technical contact network to coordinate and announce the start-up dates, and provides the necessary information for a successful exchange.
- **Monitor and support the new exchanges**: To ensure success during the initial period, the new exchanges are closely monitored by the Help Desk, assisting as needed.
- **Provide continuous technical advice and the right reporting tools to help the users monitor their own EDI message quality**: The applicable tools and reports to support permanent and proactive monitoring of the

quality, availability, and timeliness of EDI messages, supporting seamless day-to-day operations and facilitating the generation of value-adding, business relevant data from EDI.

NOTES

E-Commerce provides the following needs:

- **Non-Cash Payment**: E-Commerce enables the use of credit cards, debit cards, smart cards, electronic fund transfer via bank's website, and other modes of electronics payment.
- **24x7 Service Availability:** E-Commerce automates the business of enterprises and the way they provide services to their customers. It is available anytime, anywhere.
- **Advertising / Marketing**: E-Commerce increases the reach of advertising of products and services of businesses. It helps in better marketing management of products/services.
- **Improved Sales**: Using E-Commerce, orders for the products can be generated anytime, anywhere without any human intervention. It gives a big boost to existing sales volumes.
- **Support**: E-Commerce provides various ways to provide pre-sales and post-sales assistance to provide better services to customers.
- **Inventory Management:** E-Commerce automates inventory management. Reports get generated instantly when required. Product inventory management becomes very efficient and easy to maintain.
- **Communication Improvement**: E-Commerce provides ways for faster, efficient, reliable communication with customers and partners.

Advantages of an Electronic Data Interchange (EDI) System

Following are the advantages of having an EDI system:

- **Shorter Processing Life Cycle**: Orders can be processed as soon as they are entered into the system. It reduces the processing time of the transfer documents. EDI saves our time of processing and generate automatic software generated purchase order, invoice and order detail immediately.
- **Electronic Form of Data**: It is quite easy to transfer or share the data, as it is present in electronic format.
- **Reduction in Data Entry Errors**: Chances of errors are much less while using a computer for data entry. Data repetition and document repetition is reduced to some extent. All the documents are centrally located and easily accessible by business partner.
- **Maintain detail Customer Information in Database:** Customer information is stored in database and provide improved client service. Help in maintaining detail information of customer in database. Many large manufacturers and retailers are ordering their suppliers to institute an EDI program for maintaining and generating EDI.
- **Improve Client Service:** The fast transfer of enterprise documents and assessed decline in mistakes allow you to do business faster and more efficient.

- **Reduction in Paperwork**: As a lot of paper documents are replaced with electronic documents, there is a huge reduction in paperwork.
- **Cost Effective**: As time is saved and orders are processed very effectively, EDI proves to be highly cost effective. The cost of paper processing is very high as compared to EDI. It saves our money and expenses.
- **Standard Means of Communication**: EDI enforces standards on the content of data and its format which leads to clearer communication.

Disadvantages of Electronic Data Interchange (EDI)

1. **Too Many Measures:** There are too numerous measures bodies developing standard documents formats for EDI. Developing EDI is difficult for programmers.
2. **Changing Standards:** Each year, most measures bodies publish modifications to the measures. This impersonates a difficulty to EDI users.
3. **Limit your Selling Partners:** Some large companies are inclined to halt doing enterprise with enterprises who don't comply with EDI.
4. **Intranet:** An Intranet is a type of private Internet. Intranet uses IP protocol to sharing information. It is used to share information, computation services and operational system within an organisation. Intranet is also an example of WAN. It is a type of private network is not open for public. The information is limited within an organisation. It can be organisation internal website, which is spread over the multiple LAN on different locations. Intranet may be private website, internal communication tool used to communicate internally in an organisation for important information exchange and focus on important decisions.

Overview of the Technology Involved in EDI

Electronic Data Interchange (EDI) can be transmitted using any methodology agreed to by the sender and recipient, but as more trading partners began using the Internet for transmission, standardized protocols have emerged.

This includes a variety of technologies, including:

- Modem (Asynchronous and Synchronous)
- FTP(File Transfer Protocol), SFTP(Secure File Transfer Protocol) and FTPS(File Transfer Protocol Secure)
- E-mail
- HTTP(Hyper Text Transfer Protocol)
- AS1
- AS2
- AS4
- Odette File Transfer Protocol [OFTP (and OFTP2)]
- Mobile EDI
- And more technologies

NOTES

When some people compared the synchronous protocol 2400 bit/s modems, CLEO devices, and value-added networks used to transmit EDI documents to transmitting via the Internet, they equated the non-Internet technologies with EDI and predicted erroneously that EDI itself would be replaced along with the non-Internet technologies. In most cases, these non-internet transmission methods are simply being replaced by Internet protocols, such as FTP, HTTP, telnet, and e-mail, but the EDI documents themselves still remain.

In 2002, the IETF published RFC 3335, offering a standardized, secure method of transferring EDI data via e-mail. On July 12, 2005, an IETF working group ratified RFC4130 for MIME-based HTTP EDIINT (AS2) transfers, and the IETF has prepared a similar RFC for FTP transfers (AS3). EDI via web services (AS4) has also been standardised by the OASIS standards body. While some EDI transmission has moved to these newer protocols, the providers of value-added networks remain active.

5.13 ELECTRONIC COMMERCE USER CHARACTERISTIC AND ISSUES

E-commerce is a form of commerce or business through which consumers are able to buy or sell products or merchandise electronically over the Internet. E-commerce takes place between organizations and between organizations and their customers. It includes transaction of goods and other materials, and includes accessing information, trading goods and electronic materials.

E-commerce Definitions (from various perspectives)

1. **From an interface perspective:** E-commerce includes various information and business exchanges between a consumer and an organization.
2. **From communications perspective:** E-commerce is a way by which a user can supply items, information or transactions via networks.
3. **From an online perspective:** E-commerce provides an electronic environment that makes it possible for the purchasing and selling of items on the Internet, such as furniture, books and electronic items.
4. **As a market:** E-commerce is a global set of connections.

In a nutshell, e-Commerce is a form of commerce or business through which users are able to buy or sell items electronically over the Internet.

E-commerce, E-business and E-transaction

E-commerce

E-commerce can be:

- Business-to-business selling and purchasing
- The security of business transactions
- E-retailing with online catalogues
- The assembly and use of demographic data through the Web

- Business-to-business exchange of data through electronic data interchange (EDI)
- E-mail and fax (e.g., with the help of newsletters)

NOTES

E-business

E-business refers to business with customers, vendors and suppliers—via the Internet. E-business provides an environment to enhance businesses and also provides an interface between businesses and customers. E-business conducts business on the Internet, not only by selling and purchasing, but also by providing services to customers and collaborating with business partners.

E-transaction

E-transaction means commercial transactions with anyone, anywhere and anytime. It provides new business opportunities that result in greater efficiency and effective transactions between customers and business partners.

Scale of E-Commerce

In E-commerce, the scale of work consists of communication and information exchange as follows:

- Exchange of secure documents, contents and values
- Platforms for e-commerce communications
- Navigation, advertising and exchange of catalogue
- Negotiation and contract making protocols in interactions among consumers, businesses and public administration
- Mobile technology-based applications
- Devices and protocols which support mobility

Drivers of E-Commerce

The drivers of e-commerce are:

1. **Anytime, anywhere, anyone**

   Today, any user can access information any time. E-commerce binds organization, business and other sectors with the help of video, multimedia, text and other technologies.

2. **Digital revolution**

   With the help of digital revolution it is possible for digital devices to communicate with each another.

3. **Increase in access**

   Due to tremendous increase in the number of computers worldwide it has greatly increased the demand for information and communication for business as well as pleasure.

4. **Organizational changes**

   E-commerce makes it possible to change the approach of any organization. There is a tendency of owners and managers within the departments to develop a chain of relationships within the organization.

Basics of E-Commerce

The basics features of e-commerce are:

- Business process that helps buying and selling items on the Internet
  - Supplier, inventory, distribution, payment management
  - Financial management, purchasing products and information
- Customer purchasing on the Internet
- Transactions conducted between businesses on the Internet

Myths about e-Commerce

The following are some of the commonly noticed myths:

1. **E-commerce is innovative**

   Unfortunately, many Internet retailers spend a disproportionate amount on the innovative tasks of website construction and marketing and concentrate little on customer support and fulfilment of their requirements.

2. **Creation of website is easy**

   This is true to some extent; however, ensuring availability and performance of the site is not an easy task. There is technology and networking infrastructure to consider for effective use of a website.

4. **Customers can be lured**

   All companies know that customers can be lured with price promotions and giveaways. There are rarely loyal customers. The moment a competitor lowers the price, they click over to the site. The best customer can be lured only with quality service once an item has been purchased.

5. **Everyone is doing it**

   It is true, but a Web presence is not commerce.

Features of E-Commerce

The features of e-commerce are:

- The facility to retrieve orders from the Internet.
- The capacity to permit users to accesses accounting data securely over the Internet.
- The web page catalogue in several cases is actually associated directly to the software data based on accounting. The main advantage is that the buyer observes real-time information related to cost, quality and measure.
- The ability to send computerized information and data to users/groups of users.
- To get printouts of all reports in web page (HTML) formats.
- Web-enabled accounting software's help menu is connected directly to pages on the Internet through the WWW.

E-Commerce Framework

An e-commerce framework supposes that e-commerce applications will be built on the existing technology infrastructure—group of computers, communication networks and communication software to develop the information superhighway.

E-commerce architectural framework

(*i*) ***Main platforms***

The risk to the Internet is through digital disorder, closed markets that cannot use each other's services, incompatible applications and frameworks that interoperate or build upon each other, and an array of security and payment options that confuse the consumers.

One solution to these problems is an object-oriented architectural framework for Internet commerce. Several vendors of e-commerce solutions have declared descriptions of such a framework. The most important platforms are:

- IBM commerce point
- Microsoft Internet commerce framework
- Netscape ONE (Open Network Environment)
- Oracle NCA (Network Computing Architecture)
- Sun/Javasoft JECF (Java e-commerce Framework)

(*ii*) ***General model***

Recently, four of these companies have settled to hold a common distributed object model based on Common Object Request Broker Architecture Internet Inter-ORB Protocol (CORBA IIOP). For the commerce on the Internet to be successful, such systems must also interoperate at a business application level. A consumer or business using one framework is supposed to be able to shop for, buy and make payments for products and services offered on dissimilar frameworks. This is not possible at present.

(*iii*) ***CommerceNet***

CommerceNet is a non-profit society that has been formed to help businesses and customers to utilize the Internet for buying and selling. It is a cross-industry effort to build a framework of frameworks, involving both e-commerce merchants and clients.

The victory of this development certainly depends on market leaders in each area who participate vigorously in their respective task forces**.** All users should use similar software because no single company can control what platform its customers will use.

Mechanics of E-Commerce

1. The business aspect of E-commerce

There are two bases and interactive business dimensions to e-commerce, and these are:

(*i*) ***The customer aspect.*** This refers to placing refined goods with the final clients.

(*ii*) ***The enterprise aspect.*** This is primarily an intercorporate or inter-organizational supply chain management, etc.

2. **The technological aspect of E-commerce**

   It can be classified according to the three basic functions of any market environment.

   (*i*) ***Access environment.*** It makes use of private and public network technologies, such as the Internet, LAN and WAN.

   (*ii*) ***Transaction aspects.*** These are EDI, point of scale device, credit, debit and smart card, automated teller machine (ATM) and electronic fund transfer (EFT).

   (*iii*) ***Support aspects.*** These are support services, such as card validation technologies, bar coding device, among others.

3. **The configuration of E-commerce**

   E-commerce to become operational requires three things to happen.

   (*i*) ***The organizational configuration.*** Integrating business process electronically.

   (*ii*) ***The network configuration.*** Providing a backbone for e-commerce.

   (*iii*) ***The media configuration.*** Getting access to the electronic marketplace.

E-commerce Applications

By using online business one can place goods or products online. A well-made application in e-commerce provides all the information to satisfy the customers' needs. This provides a sensible amount of product with the purchase ability to the customers. It is important to note that a website must be product specific and it must also supports the transaction process when business is being done. Some of these consist of:

A. 1. **Search capability for the product.** It provides a way through which a consumer can search products of their interest and switch directly to the interested product over the Internet.

2. **Data sheets can be downloaded.** Consumers can download products and other supporting information and make their purchase decision.

3. **Support for customers online.** It allows staff to focus more on customer services issued online.

4. **FAQ based on products.** Once the customer buys the product then they expect that their problems be sorted out directly without having to communicate through the use of quality sites.

5. **Message board to support customers.** Message board provides customers access to information any time they need. New customers can benefit from the questions and solutions provided by the message board.

6. **Product newsletters.** These allow customers to be up to date with product information. Users can easily subscribe mailing lists for product information in which they are interested.

7. **Support sales process.** E-commerce sites support the sales process through purchase and also provides the necessary information to the customer.

B. E-commerce communication mechanism

Nowadays, the Internet is the finest means of communication between businessmen and clients. Due to various advance technology-oriented concepts, purchasing and selling of goods through websites has become popular. Online business is growing speedily through a variety of software that helps consumers to learn the tricks of buying and selling. Online business works by the following methods:

- Shopping cart software
- Online e-telephony

Shopping cart software is the means of online presentation of goods for sale. It provides the idea of goods to choose from, online payment facility, joint selection of goods in the form of list, etc. By puting all the chosen goods in the cart and paying for all the selected items, shopping cart software has become the simplest way of shopping online. There are many features provided by this software, such as:

- Credit card adequacy
- Simple navigation system for the consumers
- Consumer account ability
- Order management ability
- Web-based administration ability
- Flexible shipping and tax options
- Built-in site optimization tools
- Inventory management ability

One of the finest ways to communicate regarding business is online telephony. It is the technology used to convert voice signals into data packets which are then are transported to a data network runs on the Internet Protocol (IP). It allows the consumer to call through the same phone line which he uses for the Internet connection. It is cheaper than making calls on the basic telephone line. This online communication technology is known in the web world as voice over IP.

C. Online E-telephony benefits

- A user is able to distinguish calls as business calls, personal calls or consumer service calls even as they are on same line.
- A user can direct the calls to a particular department and take automated orders.
- A user can screen the callers without any information to caller.
- A user can get forwarded calls from all over the world.
- There will be no busy line problems.
- Voice mails can be received on the computer.

In addition to these, there are many other facilities which can be availed by using e-telephony. Thus, communication on the Internet provides numerous facilities to ease business complexities and raise profits.

NOTES

5.13.1 Advantages and Disadvantages of E-commerce

Advantages of E-Commerce to Organizations

The various advantages of E-commerce to business organizations are as follows:

- Users and firms can do their business online through wired or wireless devices and will be able to increase their sale by using e-commerce.
- Companies will be able to offer their products or services at lower prices.
- It increases the business both at the local and global level markets.
- The cost of manufacturing products, processing items, distributing goods, storing data or information and accessing information can be reduced. E-commerce brings the universal access of the Internet to the core business processes of buying and selling goods and services. It helps to generate demand for products and services and improves order management, payment, and other support functions. The overall goal is to cut expenses by reducing transaction costs and streamlining all kinds of processes. E-commerce helps the process manufacturing companies by replacing paper catalogs, phone sales, and faxes. It also helps in the reduction in the cost of obtaining commodity products and the ability to sometimes get better prices for the products. E-commerce uses Internet technologies to enable better and faster collaboration between buyers and sellers. It is the practice of buying and selling varied good and services on the World Wide Web (Internet) over wired communication lines connected throughout the globe where the World Wide Web serves as the central medium for all trading transactions. It also enables sell and purchase of commodities and services right from your home thus reducing the related cost expenses. Because, to purchase the virtual products and services online, one has to simply order it and the products and services will be sent once your payment is acknowledged. Contribution to digital goods and services help the manufacturers to reduce operating cost and increase profit.

  Thus, e-commerce is very cost efficient and economical. General costs of running a business otherwise are far higher than that operated with the help of technology and e-commerce. Staffing, middlemen, overhead costs, etc. can be reduced drastically. Most of the transaction procedures can be automated without any human intervention.
- The seller's website gives greater accessibility on the products available. Reviews on the products bought on the website are convenient and useful for other prospective buyers. Costs such as rent, employment, marketing and other similar expenses are little. Also, the cost of advertising on the Internet is minimal and the reach is much wider than traditional business.
- The business organization will be able to reduce paperwork.
- Dropdown processing permits customization of products and services which provides competitive advantage to its implementers.
- It decreases the time between the cost of funds; and between the products and services.
- It supports the efforts for business process re-engineering **(**BPR**).**

- It decreases the product cost over the Internet, which is much more cheaper than value-added networks (VANs).

  It enables to build more collaborative and stronger relationships with suppliers. This includes streamlining and automating the underlying business processes, enabling areas such as:

  - o Direct marketing selling
  - o Customer services (call centres)
  - o Fulfilment
  - o Procurement
  - o Replenishment and information management

NOTES

Advantages of E-Commerce to Consumers

The following are the advantages of E-commerce to consumers:

- It allows customers to shop or perform any transaction at any time from any location in the world.
- It provides customers with better selection of products and services.
- Consumers may make quick comparisons among expensive products.
- Consumers can interact with other customers, share their ideas, views and experiences.
- Significant discounts on different products or items are available due to high competition.
- It allows fast delivery of products and services.
- Consumers can get information in this seconds.
- It is possible to participate in virtual auctions.
- Consumers can get additional information about the goods, and make a more informed decision. This helps in the following ways:
  - o Better information opens the way to more assurances and to make a better choice.
  - o Additional information also leads to improved consumer fulfilment because consumers have a better idea of using the goods.

Advantages of E-Commerce to Society

The following are the advantages of E-commerce to the society:

- It permits persons to work from home, due to which there is less traffic on roads, and this in turn reduces air pollution.
- It helps products to be sold at competitive prices.
- It allows people in remote areas to connect through the Internet and enjoy products, goods and others services which are generally not easily available to them.
- Delivery of services at reduced cost.
- It improves the quality of products.

NOTES

- More people can work offsite.
  - o This decreases HR costs for companies, because they can have smaller office buildings, less parking spaces, fewer IT services, etc.
- It facilitates the delivery of with the help of postal services
- Higher standard of living: Some goods can be sold at low prices, allowing less affluent people to buy more and increase their standard of living.

Disadvantages of E-Commerce

Some disadvantages of E-commerce are as follows:

- Customers will not be satisfied until they see the products or goods physically.
- Security problems might arise when doing business online.
- Lack of security measures
- Some legal issues

5.13.2 limitations of E-commerce

Technical Limitations of E-Commerce

There are various technical limitations of E-commerce, such as the following:

- There is lack of security, consistency, standards and other protocols.
- The bandwidth is insufficient for telecommunication.
- Development tools for software are changing speedily.
- There are some technical difficulties to integrate the Internet and E-commerce software.
- There is a requirement for web servers and other infrastructure instead of network servers which causes additional cost.
- There are some problems related to software that does not fit with some hardware, operating systems or other components.

Non-Technical Limitations of E-Commerce

Some of the non-technical limitations of E-commerce are as follows:

- Cost and justification
  - o The cost of developing an e-commerce application at home can be very high.
  - o There may be security and privacy problems.
- Online business provides lack of touch and feel to customers
- Lack of bargaining, trust and user conflict
- Control conflict

Check Your Progress

15. What is E- commerce?
16. What are the goals of e-business?
17. Define EDI.
18. What do you understand by E-Commerce?

NOTES

5.14 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Broadband LANs are multichannel, analog LANs. They are typically based on coaxial cable as the transmission medium, although fibre optic cable is also used. Individual channels offer bandwidth of 1 to 5 Mbps, with 20 to 30 channels typically supported.
2. The 'Line Coding', also called digital baseband modulation or digital baseband transmission, is a process carried out by a transmitter that converts data, in the form of binary digits, into a baseband digital signal that will represent the data on a transmission line.
3. ADSL is widely used to connect most of the homes and small business subscribers to the Internet. ADSL divides the available frequencies in a telephone line by assumimg that most of the Internet subscribers are more inclined to download information from the Internet than upload.
4. High bit-rate Digital Subscriber Line (HDSL) is defined as a telecommunications protocol which was standardized in 1994. It was the first Digital Subscriber Line (DSL) technology which used a higher frequency spectrum over copper and twisted pair cables.
5. Wireless Local Loop (WLL) is the use of a wireless communications link as the 'Last Mile / First Mile' connection for delivering Plain Old Telephone Service (POTS) or Internet access (marketed under the term 'Broadband') to telecommunications users.
6. Data security is concerned with the protection of data contained in a file or many files in a computer either as a standalone or on a network, from unauthorized interception.
7. The following are the three types of security measures:
   (i) Invalid access/Possibility of eavesdropping
   (ii) Firewall security
   (iii) Encryption (VPN Function)
8. A firewall is a combination of software and hardware components that controls the traffic between a secure network (usually an office LAN) and an insecure network (usually the Internet), using rules defined by the system administrator.
9. The basic aim of firewall is to provide only one entrance and exit to the network. There are two firewalls. One blocks the undesirable traffic, while the other allows traffic.

10. The main features of data encryption are:
    (i) Prevents unwanted access to documents and e-mail messages.
    (ii) Even the strongest levels of encryption are very difficult to break.
11. Authentication is any process by which one verifies that someone is who they claim they are. Basically, it involves a username and a password. It can also include any other method of demonstrating identity, such as a smart card, retina scan, voice recognition, or fingerprints.
12. The *Internet privacy* is a broad term referring to the various concerns, technologies, and strategies for protecting information, communications, and choices that are meant to be private. In general, using the Internet often means giving up some measure of privacy.
13. There are two parts of a message: the *header* and the *body*. the header contains data that describes the message and controls how it is delivered and processed, the body of the message is the actual information that is to be communicated.
14. An e-mail password is used for security reasons. A password is the personal code of a user and should not be disclosed to anyone. Hence, never give your password to anyone and do not write it down where someone else may find it.
15. E-commerce refers to purchase, or sale, advertising and servicing of goods or services over the Internet. Currently though not big enough as compared to traditional peer markets, E-commerce is expected to grow in the near future.
16. The main goals of e-business are to understand how the:
    (i) Needs of a consumer, merchant and organization can be met
    (ii) Quality and quantity of goods can be improved
    (iii) Speed of services can be increased
17. Electronic Data Interchange (EDI) related with exchange of documents which are used in business electronically. In other words exchange of business document from compute to computer is known as EDI.
18. E-commerce is a form of commerce or business through which consumers are able to buy or sell products or merchandise electronically over the Internet.

5.15 SUMMARY

- Broadband LANs are multichannel, analog LANs and typically based on coaxial cable as the transmission medium, although fibre optic cable is also used.
- Local Loop (LL) is referred as an electronic circuit line from a subscriber's phone to the local exchange office termed as Local Central Office (LCO).
- A Local Loop (LL) is a physical connection from the end user site to a providers Point of Presence (POP).

- In ADSL technology, there has been a new progress which intends to use two copper loops at a data rate of 1.544Mbps.
- ADSL offers download speed in the range of 1-2 Mbps and upload speed in the range of 64-640 Kbps.
- Line coding is the process of converting digital data to digital signals. With the help of this technique a sequence of bits can be converted to a digital signal.
- TCP/IP protocols are used globally irrespective of the nature of the organizations, whether they are general category organizations or security-specific sensitive organizations.
- Networks are designed to share information. Therefore, the network must be clearly configured to identify the shareable information and non-shareable information.
- Data security is concerned with the protection of data contained in a file or many files in a computer either as a standalone or on a network, from unauthorized interception.
- The intranet is a TCP/IP network that is modelled after the Internet that only works within the organization.
- In public key (asymmetric) encryption, two mathematically-related keys are used, one to encrypt the message and the other to decrypt it.
- Private Key encryption (symmetric) is also known as conventional or single-key encryption.
- A user remotely located must be first authenticated before accessing the network or Intranet of an organization.
- Managing Windows security is required to manage the complete system for running the applications, downloading the update features for Windows, runtime programs, etc.
- The private key encryption contains a secret key that is taken as code.
- Windows has a built-in database and control system to keep track of all of the software and critical information that lives on your PC.
- Viruses are frequently transmitted through e-mail attachments, peer to peer downloads, phishing and instant messages.
- The Internet privacy is a broad term referring to the various concerns, technologies, and strategies for protecting information, communications, and choices that are meant to be private.
- An interruption can be defined as a state where the asset of a system gets destroyed or becomes unavailable.
- A security policy can be defined as the framework within which an organization establishes needed levels of information security to achieve the desired confidentiality goals.
- Fabrication occurs when an attacker inserts forged objects into the system without the sender's knowledge or involvement.

- The management of keys is the chief problem area for all encryption systems. The keys are the most valuable information. If anyone can get a key, anyone can decrypt everything that has been encrypted by that key.
- The public keys of a key pair do not require confidentiality protection. They only require the integrity protection which is provided by their certification.
- One of the critical requirements of an electronic mail system is that the sender and receiver of a message need not be online at the time when mail is sent.
- E-business deals with the buying and selling of information, products and services through the computer network**.**
- E-business applications enable various business functions and transactions to be conducted electronically.
- Electronic Data Interchange (EDI) is related with exchange of documents which are used in business electronically.
- Electronic Data Interchange (EDI) can be transmitted using any methodology agreed to by the sender and recipient, but as more trading partners began using the Internet for transmission, standardized protocols have emerged.
- E-commerce is a form of commerce or business through which consumers are able to buy or sell products or merchandise electronically over the Internet.

5.16 KEY TERMS

- **Local Loop (LL):** It refers to an electronic circuit line from a subscriber's phone to the local exchange office termed as Local Central Office (LCO)
- **Line Coding:** It is the process of converting digital data to digital signals.
- **Intranet:** It is a TCP/IP network that is modelled after the Internet that only works within the organization.
- **Ciphertext:** This is the encrypted or scrambled message produced by applying the algorithm to the plaintext message using keys.
- **Authentication**: It is any process by which one verifies that someone is who they claim they are. Basically, it involves a username and a password.
- **Private Key Encryption**: The private key encryption contains a secret key that is taken as code. This mechanism encrypts a packet of information if it passed across network to the other computer.
- **Security Attack**: It refers to any action that compromises the security of information owned by an organization.
- **Interruption:** It can be defined as a state where the asset of a system gets destroyed or becomes unavailable.
- **Electronic Data Interchange (EDI**): It is the Exchange of business document from computer to computer over internet.

- **E-commerce:** It is a form of commerce or business through which consumers are able to buy or sell products or merchandise electronically over the Internet.
- **E-business**: E-business refers to business with customers, vendors and suppliers—via the Internet. E-business provides an environment to enhance businesses and also provides an interface between businesses and customers.

NOTES

5.17 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the advantage and disadvantages of broadband network?
2. What are line coding techniques?
3. State the basic techniques of data security.
4. Write the procedures to validate remote login.
5. What is key certification?
6. Write the steps of mail communication process.
7. What do you understand by SMTP?
8. Write some E-business opportunities for expanding business.
9. What are differences between E-business and traditional business mechanisms?
10. What are the advantages of E-commerce?
11. How the E-advertising is proven very useful in present days?
12. What are the disadvantages of EDI?
13. Write the features of e-commerce.
14. What are the myths related to e-commerce?
15. Discuss the advantages and disadvantages of E-Commerce.

Long-Answer Questions

1. Elaborate on local loop technologies.
2. Briefly describe the basic requirements of network security.
3. Explain the network architecture of a firewall in detail.
4. What are the various types of security attacks? Explain.
5. Discuss key management (SMTP) in detail.
6. Explain simple mail transfer protocol in detail.
7. Elaborate on the working of E-business.
8. Briefly describe the various steps involved in EDI implementation.

5.18 FURTHER READING

NOTES

Forouzan, Behrouz A. *Data Communications and Networking*. New Delhi: Tata McGraw-Hill, 2004.

Stallings, William and Richard Van Slyke. *Business Data Communications*. New Jersey: Prentice Hall, 1998.

Black, Uyless. *Computer Networks*. New Jersey: Prentice Hall, 1993.

Stallings, William. *Data and Computer Communications*. New Jersey: Prentice Hall, 1996.

Tanenbaum, Andrew S. *Computer Networks*. New Jersey: Prentice Hall PTR, 2002.

Stallings, William. *Data and Computer Communications*. NJ: Prentice-Hal, 1996.