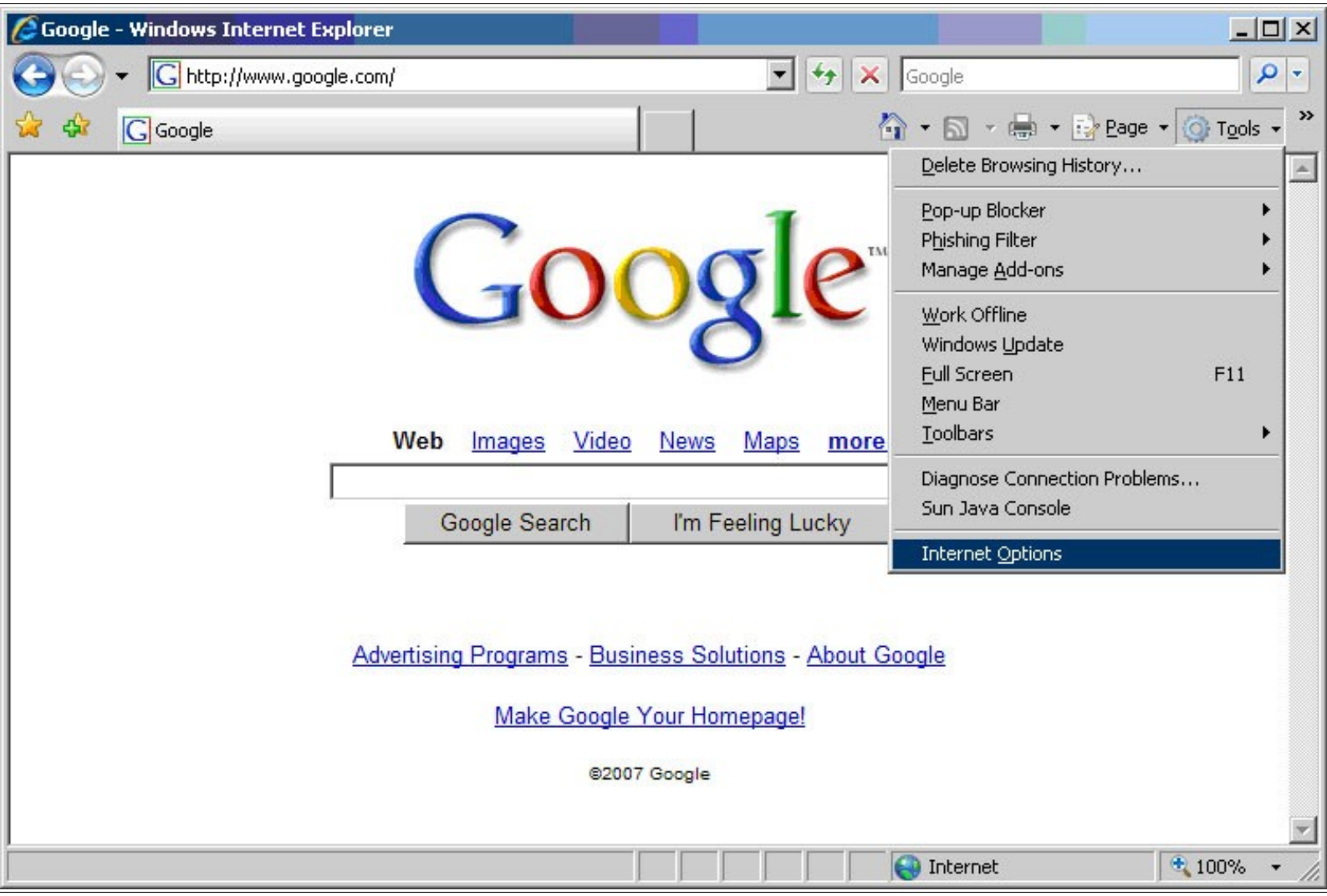


MBF 744

COMPUTER NETWORKS AND INTERNET

Course Code	MBF744
Course Title	Computer Network and Internet
Course Developer/Writer	Dr. G. A. Aderounmu Computer Department O.A.U. Ile-Ife.
Programme Leader	Dr. O. J. Onwe National Open University of Nigeria, Lagos
Course Coordinator	Abdullahi S. Araga National Open University of Nigeria, Lagos

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No 5 Dar es Salam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published by:
National Open University of Nigeria 2008

First Printed 2008

ISBN: 978-058-245-2

CONTENTS **PAGES**

MODULE 1

UNIT 1 INTRODUCTION TO COMPUTER NETWORK AND DATA COMMUNICATION

CONTENTS

1.0 INTRODUCTION

The world of computer networks and data communications is surprisingly vast and increasingly significantly field of study. One considered primarily the domain of network specialist and technicians, computer networks now involve business managers, computer programmers, system designers, office managers, home computer users, every citizens. It is virtually impossible for the average person on the street to spend 24 hours without directly or indirectly use some forms of computer network. Welcome to the amazing world of computer networks! Unless you have spent the last 24 hours in complete isolation, it is nearly impossible to not have used some form of computer networks and data communication. In this course, MCS 760, we shall be looking at a deeper level of computer network and the Internet. The first unit of this course treats the concepts behind computer networks.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- Define the basic terminology of computer networks.
- Recognize the individual component of computer networks.
- Outline the basic network connections.
- Cite the reasons for using a network architecture.
- List the he layers of the OSI model and describe the duties of each layer.
- Compare the OSI model and TCP/IP protocol suite and list their differences and similarities.

3.0 MAIN CONTENT

3.1 Language of Computer Networks

i. **Computer network** – an interconnection of computers and computing equipment using either wires or radio waves over small or large geographic areas.
ii. **Local area network –** networks that are small in geographic size spanning a room, floor, building, or campus.
iii. **Metropolitan area network** – networks that serve an area of 1 to 30 miles, approximately the size of a typical city.
iv. **Wide area network** – a large network that encompasses parts of states, multiple states, countries, and the world.
v. **Personal area network –** a network of a few meters, between wireless devices such as PDAs, laptops, and similar devices.
vi. **Voice network** – a network that transmits telephone signals
vii. **Data network** – a network that transmits computer data
viii. **Data communications** – the transfer of digital or analog data using digital or analog signals
ix. **Telecommunications** – the study of telephone and the systems that transmit telephone signals.
x. **Network management** – the design, installation, and support of a network, including its hardware and software.

3.2 Components of Computer Networks

If you could create one picture that tries to give an overview of a typical computer network, what might this picture include? Figure 1.1 shows such a picture, and it includes examples of local, personal, and wide area networks.

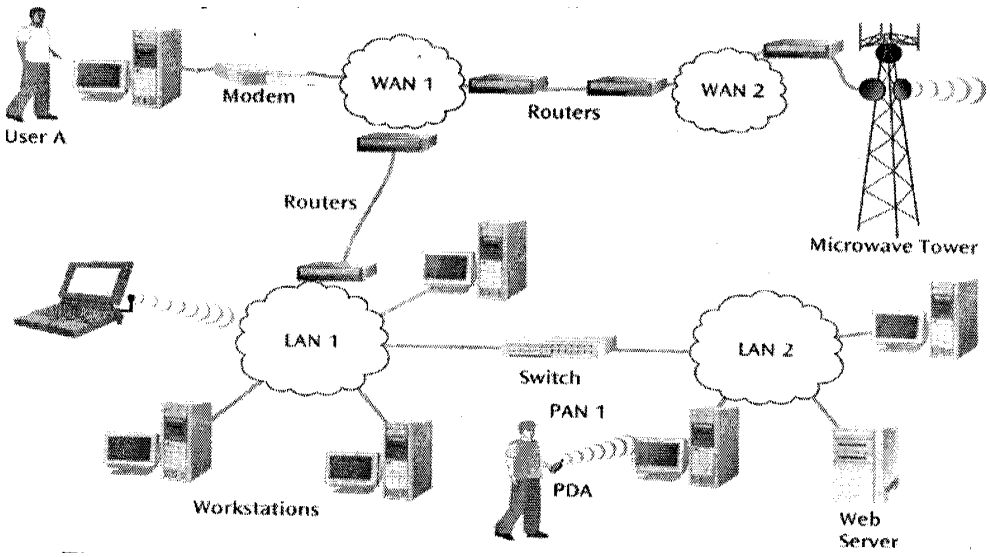


Figure 1.1: Overview of Interconnection between different types of Networks

It is important to note that most local area networks include the following hardware:

- **Workstations** – which are personal computers or microcomputers where users resides.
- **Servers** – which are computers that store network software and shared user files.
- **Network hubs** – which are the collection point for the wires that interconnects the workstations.
- **Switches** – which are more advanced devices that are replacing hubs and are capable of filtering out unnecessary traffic.
- **Routers** – which are the connecting devices between Local Area Network (LANs), and Wide Area Network (WANs).

Wide area networks also can be of many types. Though many different technologies are used to support wide area networks, all wide area networks include the following components:

- Nodes – which are the computing devices that allow workstations to connect to the network and that make the decisions about where to route a piece of data.
- Some type of high-speed transmission line, which runs from one node to another.
- A subnet which consists of the nodes and transmission line collected into a cohesive unit.

3.3 Basic Network Connection

In the last two sections, we examined the language and components of computer network. Let us now examine basic network systems and their connections to see how extensive the uses of data communications and computer networks are. The basic connections that we will examine in this section include:

- **Computer terminal/microcomputer to mainframe**

During the 1960s and 1970s, the computer terminal-to-mainframe connections was virtually in every office, manufacturing, and academic environment. These types of systems are still used in many types of businesses for data entry and data retrieval, such as you might find when applying for your car license. This scenario is shown in figure 1.2

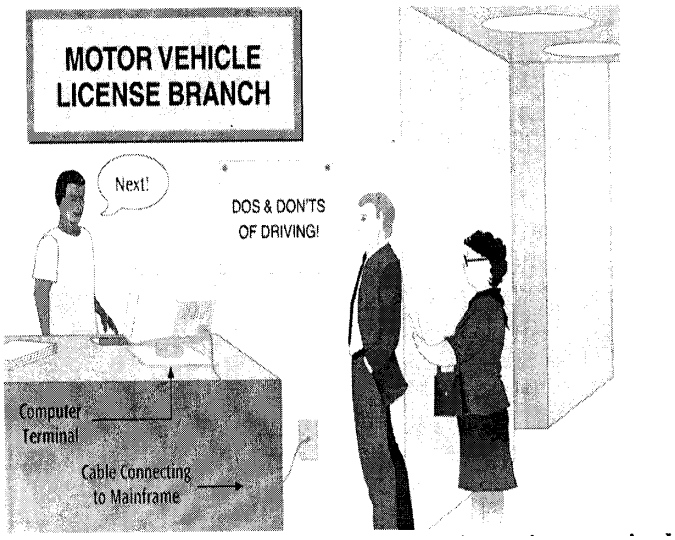


Figure 1.2: Text-based input transaction using terminal

The terminal-to-mainframe connections use "dumb" terminals because the end user is doing relatively simple data entry and retrieval operations and a workstation with a lot of computing power is not necessary. A computer terminal is a device that is essentially a keyboard and screen with no long-term storage capabilities, and little, if any processing power. The mainframe computer controls the sending and receiving of data to and from each terminal since the terminal does not posses a lot of computing power. The advent of microcomputers in the 1970s and early 1980s replaced the terminal.

- **Microcomputer to local area network**

The most common network connection today, is the microcomputer-to-local area network connection. It is found in virtually every business and academic environments, and now homes. Using microcomputers connected to a LAN, end users can request and download an application, then execute it on their computers. Figure 1.3 shows a diagram of this type of connection. Example of this type of connection in the business world is the client/server system where a user at a microcomputer issues a request for some form of data or service from the server.

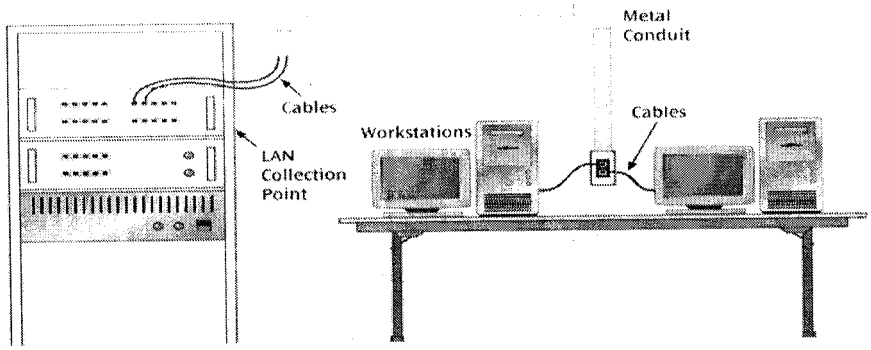


Figure 1.3: Microcomputer-to-LAN connection

- **Local area network to local area network**

Because the LAN is a standard in business and academic environments, it should come as no surprise that many organizations need the services of multiple LANs and that it may be necessary for these LANs to communicate with each other. For example, a company may want the LAN that supports its research department to share an expensive color laser printer with its marketing department's LAN. Fortunately, it is possible to connect two LANs so that they can share peripherals as well as software. The devices that usually connect two or more LANs are the switch, bridge, and router, already discussed in section 3.2. Figure 1.4 provides an example of two LANs connected by a switch.

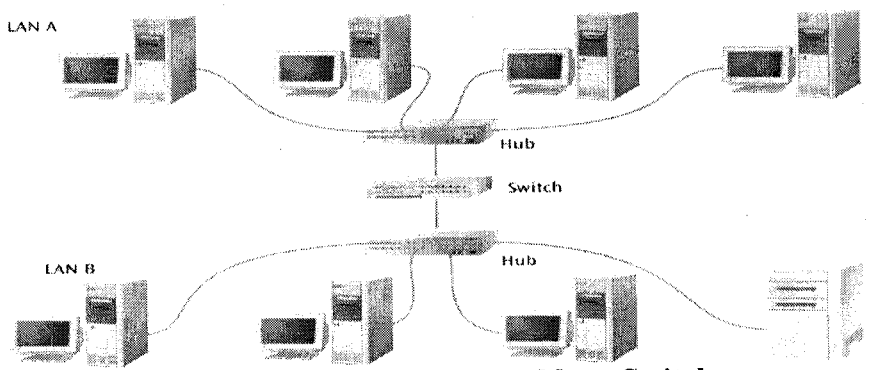


Figure 1.4: Two LANs connected by a Switch

- **Local area network to metropolitan area network**

This is a new form of network that interconnect businesses within a metropolitan area. Typically, this interconnection uses only fiber-optic links at extremely high speeds. A Metropolitan Area Network (MAN) is a high-speed network that interconnects multiple sites within a close geographic region, such as large urban area. MAN is cross between LANs and MANs. Figure 1.5 shows a simple scenario of LAN-to-MAN connections.

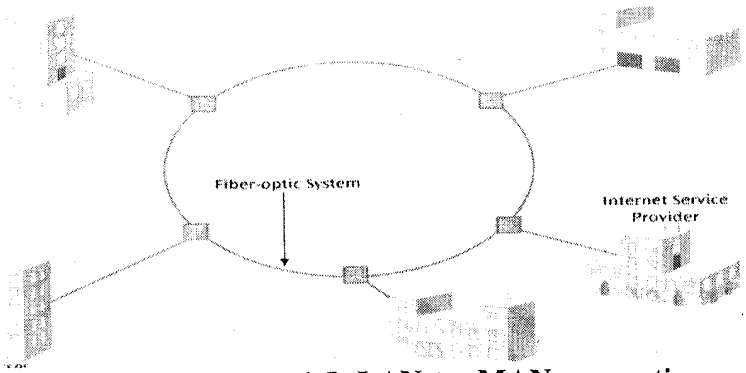


Figure 1.5: LAN-to- MAN connection

Figure 1.5: LAN-to-MAN connection

Other basic network connections include Local Area Network to Wide Area Network, sensor to Local Area Network, satellite and microwave, wireless telephone and wired telephone to network, etc.

SELF-ASSESSMENT EXERCISE

i. We have discussed to a good extent the language and components of computer network. Based on the knowledge you acquired in section 3.1 through 3.3, define the following terms:

- Computer network
- Data communications
- Computer terminal

ii. What are the advantages of using a Computer network?
iii. Mention two major components of a computer network.

3.4 Network Architecture

Now that you can identify different types of networks and connections, you need a framework to describe how all the various components of a network interoperate. When you use a computer network to perform an application, many pieces come together to assist in the operation. A

network architecture places the appropriate network pieces together the layers define a model for the functions or services that need to be performed. A network architecture is a reference model that describes the layers of hardware and software necessary to transmit data between two points or for multiple devices/applications to interoperate. Reference models are necessary to increase likelihood that different components from different manufacturers will converse. The most common architectures known today are in this section Open System Interconnection (OSI) model and Transmission Control Protocol/Internet Protocol suite. We will discuss these two architectures in the next subsection.

3.4.1 Open System Interconnection

The international organization for standardization (ISO) created the OSI reference model in an attempt to standardize the design of the communication systems and the interoperability between their components. As shown in figure 1.6, the model consist of seven layer: Application, presentation, session, transport, network, data link, and t he physical.

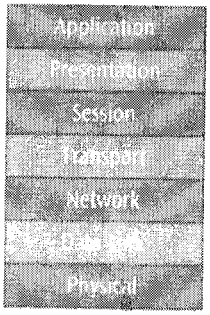


Figure 1.6: OSI reference model

- **Application layer –** This is the top layer in the OSI model where the application using the network resides. Common network applications include web browsing, e-mail, file transfers, and remote logins.
- **Presentation layer –** This layer performs a series of miscellaneous functions necessary for presenting the data package properly to the sender or receiver. For example the presentation layer might perform ASCII-to-non-ASCII character conversions, encryption and decryption of secure documents, and the compression of data into smaller units.
- **Session layer –** This is the least used layer of the OSI model. This layer is responsible for establishing sessions between users for token management (a service that controls which user's computer talks when during the current session by passing software token back and

forth). It also establishes synchronization points, which are backup points that are used in case of errors or failures.

- **Transport layer –** This layer ensures that the data packet that arrives the final destination is identical to the data packet that left the original station. The transport layer provides an end-to-end error-free network connection. It is used only at the two end-points of connection.
- **Network layer –** This layer is responsible for creating, maintaining and ending network connections. As this layer sends the package of data from node to node within a network and between multiple networks, it generates the network addressing necessary for the system to recognize the next intended receiver. To choose a path through the network, the network layer determines routing information and applies it to each packet or group of packets. It also performs congestions control, which ensures that network does not become saturated at any point.
- **Data link layer –**This layer is responsible for taking the data and transforming it into a *frame* with header, control and address information, and error diction code. The data link control layer is also responsible for flow and error control. It is to be noted t hat the data link operations are quite similar to the transport layer operations. The primary difference is that the transport layer performs its operations only at the end points, while the data link layer performs its operations at every stop (node) along the path.
- **Physical layer –** This is the bottom layer of the OSI model. It handles the transmission of its over a communications channel. To perform this transmission of bits, the physical layer handles voltage levels, connectors, media choice, and modulation techniques.

The first four layers described above are called end-to-end layers. They are responsible for the data that is transmitted between the endpoints of a network connection. In other words, these layers perform their operations only at the beginning point and ending point of the network connection. Where as the other three layers, network, data link, and physical layers are not end-to-end layers. They perform their operations at each node along the network path, not just at the end points.

A Typical Scenario

A communication application such as an e-mail system that accepts the message “Andy; how about lunch? Sharon” has many steps. To begin, the e-mail “application worker” prompts user to enter a message and specify an intended receiver. The application worker creates the appropriate data package with message contents and addresses and sends it to a “Presentation worker”. The presentation worker examines the data package to determine whether something like encryption or data

compression is required and, if so perform the necessary functions. The modified data package is then passed to a "Session worker", which is responsible for adding backup synchronization points in the case of network failures. Adding a backup synchronization point is similar to leaving a bookmark in a book while you read ahead. If the book should fall out of your hands and you lose the page, the bookmark will show you where you were a few pages ago and allow you to catch up to the current page. Next the updated data package goes to the "Transport worker" which is responsible for providing overall transport integrity. The transport worker may establish a connection with the intended receiver, monitor the flow between sender and receiver, and perform the necessary operations to recover lost data in case some data disappears or become unreliable.

The "Network worker" then takes the data package and may add routing information so that the data package can find its way through the network. Next to get the data package is the "Data link worker", which inserts error-checking information and prepares the data package for transmission. The final worker is the "Physical worker", which transmits the data package over some form of wire or through the air using radio waves. Each worker has his own job function as shown in figure 1.7.

SELF-ASSESSMENT EXERCISE 2

i. Using appropriate examples, explain the following as its relates to OSI model:

   a. Token management
   b. Synchronization point.

ii. What is the primary difference between the data link layer and the transport?

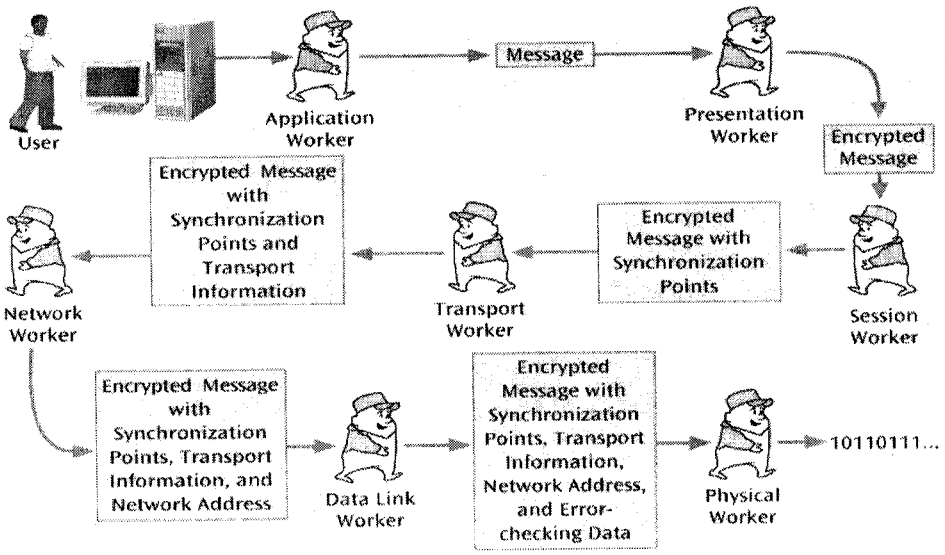


Figure 1.7: A layer function scenario

3.4.2 Transmission Control Protocol/Internet Protocol

As ISO was hammering out details of the OSI model, another network architecture that was already in place quietly continued to gain currency. The TCP/IP protocol suite, which was not created by standards-making organization but by a group of computer scientists, incorporates the TCP and IP protocols and has in fact always been more popular than the OSI model. Figure 1.8 shows the layers of the TCP/IP protocol suite (Internet suite) and hoe it compares to the OSI model. The TCP/IP protocol suite does not have rigidly defined layers as the OSI model does, some textbooks describe five TCP/IP layers, while other describe four.

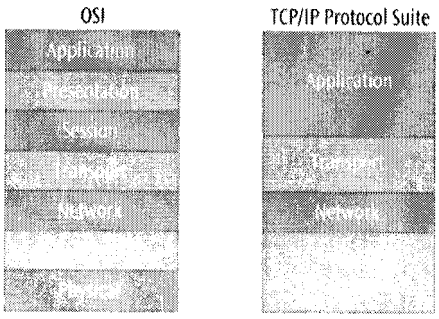


Figure 1.8: TCP/IP compared to OSI model

- **Application layer –** This layer is equivalent to OSI's application and presentation layers. Frequently used applications include File Transfer Protocol, Telnet, Simple mail Transfer protocol, Simple Network Management, and Hypertext Transfer Protocol. All these applications will be discussed in the later unit of this course.

- **Transport layer –** This layer is equivalent to OSI's transport layer. This layer commonly uses the TCP to maintain an error-free end-to-end connection. User datagram protocol is an alternative that is also used, though less frequently, in the TCP/IP protocol suite.
- **Network (Internet or internet work) layer –** This layer is equivalent to OSI's network layer. The protocol that is used at this layer to transfer data within and between networks is the Internet Protocol (IP).
- **Network access (data link/physical) layer –** This layer is roughly equivalent to OSI's data link and physical layers.

ANSWER TO SELF-ASSESSMENT EXERCISE

1(a)
- **Computer Network** – An interconnection of computers and computing equipment using either wires or radio waves over small or large geographic areas.
- **Data Communications** – This is the transfer of digital or analog data using digital or analog signal from one point to another.
- **Computer Terminal** – This is a device that is essentially a keyboard and screen with no long-term storage capabilities, and little if any, processing power.

1(b) Advantages of Computer Network
- Resource sharing
- Reliability

1(c) Major Components of Computer Network
- Server
- Router
- Workstation
- Hub
- Bridges, etc.

2(a) i. **Token management:** This is a service that controls which user's computer talks when during the current session y passing a software token back and forth.

ii. **Synchronization point:** These are backup points that are used in case of errors or failures. Use example of electronic book under the session layer in section 3.4.1 or any other related example.

(b) The primary difference is that the transport layer performs its operation only at the end points, while the data link layer performs its operations at every node along the path.

4.0 CONCLUSION

In this unit, which is our very first one, we have discussed generally the concept of computer network and data communications. We looked at the language of computer networks. We also identified the basic components of computer networks and network connections. We gave a brief overview of network architecture and described the two most common and important network architecture OSI model and TCP/IP protocol suite.

5.0 SUMMARY

The field of data communication and computer networks includes LAN, MAN, WAN, among others. The application area of computer networks and data communications can be understood in terms of general network connection: Terminal/microcomputer-to-mainframe computer, microcomputer-to-LAN, LAN-to-LAN, LAN-to-MAN, MAN-to-MAN etc. A network architecture, or communications model places network pieces in layers. The layers define a model for the functions or services that need to be performed. Each layer in the model defines the services that either hardware or software or both provide. To standardize the design of communication systems, the ISO created the OSI model which consist of seven layers. Another network architecture called the TCP/IP suite has surpassed the OSI model in popularity. This TCP/IP is also known as Internet model and is composed of four layers.

6.0 TUTOR-MARKED ASSIGNMENT

a. How do the layers of the OSI model compare with the layers of the TCP/IP protocol suite?

b. Distinguish between a Switch and a Router.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network,* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). *Data Communication Principles for Fixed and Wireless Networks*, Kluwer Academic Publisher, USA.

UNIT 2 FUNDAMENTALS OF DATA AND SIGNALS

CONTENTS

1.0 INTRODUCTION

In Unit 1 of this course, we discussed the concept of computer network and highlighted the major components of computer network. The components discussed in unit 1, section 3.2 are physical components. This unit will deal primarily with two ingredients that are more difficult to see physically: data and signals. The study of data and signals will explain why almost all forms of communication such as data, music, and video, are usually converted are slowly being converted from their original analog forms to the newer digital forms.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- Distinguish between logical and physical connections.
- Distinguish between data and signals, and cite the advantages of digital data and signals over analog data and signals.
- Identify three basic components of a signal.
- Discuss bandwidth of a signal and how it relates to data transfer speed.
- List and draw diagrams of the basic digital encoding schemes.

- Identify different shift keying techniques
- Identify the two most common digitization techniques and describe their advantages and disadvantages
- List the different data codes and how they are used in communication system.

3.0 MAIN CONTENT

3.1 Logical and Physical Connection

Before we proceed to fundamental of data signals, we will look at an important concept to understand with regards to layers of a communication model is the lines of communication between a sender and receiver. Let us consider figure 2.1, which shows sender and receiver using a network application that is designed on the OSI model.

Notice the dashed lines between the sender's and receiver's application layers, transport layers, network layers, and data link layers. No data flows over these dashed lines. Each dashed line indicate a logical connection. A logical connection is a non-physical connecting between sender and receiver that allows an exchange of command and responses. The sender's and receiver's transport layers for example share a set of commands that is used to perform transport-type functions, but the actual information or data has to be passed through the physical layers of the sender and receiver, as there is no direct connection between the two transport layers. The physical connection is the only connection between sender and receiver and is at the physical layer, where actual 1s and 0s – the digital content of the message are transmitted over wires or airwaves.

3.2 Introduction of Data and Signals

Data and signals are the two most basic building blocks of any computer network. It is important to understand that the terms "data" and "signals" do not mean the same thing, and that in order for a computer

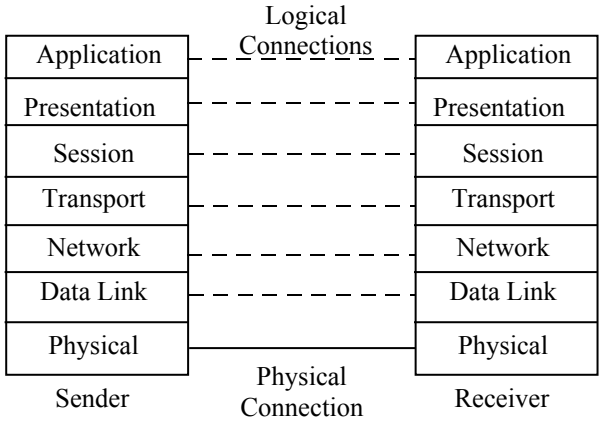


Figure 2.1: Sender and Receiver Communicating using OSI

network to transmit data, the data must first be converted into the appropriate signals. The one thing data and signals have in common is that both can be in either analog or digital form, which gives us four possible data-to-signal conversion combinations:

- Analog data-to-analog signal, which involves amplitude and frequency modulation techniques.
- Digital data-to-digital signal, which involves five encoding techniques.
- Digital data-to-analog signal, which involves three modulation techniques.
- Analog data-to-digital signal, which involves two digitization techniques.

Each of these four combinations occurs quite frequently in computer networks, and each has unique applications and properties, which are summarized in table 2.1.

Table 2.1: Four Combinations of Data and Signals

Data	Signal	Encoding or Conversion Technique	Common Devices	Common Systems
Analog	Analog	Amplitude modulation Frequency modulation	Radio tuner TV tuner	Telephone AM and FM radio Broadcast TV Cable TV
Digital	Digital	NRZ-L NRZI Manchester Differential Manchester Dipolar-AMI 4B/5B	Digital encoder	Local area networks Telephone systems
Digital	Analog	Amplitude shift keying Frequency shift keying Phase shift keying	Modern	Dial-up Internet access DSL Cable modems
Analog	Digital	Pulse code modulation Delta modulation	Codec	Telephone systems Music systems

3.2.1 Data and Signals (Concept)

Information that is stored within computer systems are transferred over a computer network can be divided into two categories: data and signals. Data is entities that convey meaning within a computer or computer system. Common example include:

- A computer file of names and addresses stored on a hard disk drive
- The bits or individual elements of a movie stored on a DVD.
- The binary ls and 0s of music stored on compact disk or inside an iPod.
- The digits 0 to 9, which might represent some kind of sales figures.

In each of the above examples, some kind of information has been electronically captured and stored on some type of storage device. Assuming you want to transfer this data from one point to another, either via physical wire or through radio waves, the data has to be converted into signal. Signals are the electric or electromagnetic impulses used to encode and transmit data. Common examples of signals include:

- A transmission of a telephone conversation over a telephone line.
- A live television news interview from Europe transmitted over a satellite system.
- A transmission of a term paper over the printer cable between a computer and a printer.

3.2.2 Analog Vs Digital

Although data and signals are two different entities that have little in common, one characteristic they do share is that they can exist in either analog or digital form. Analog data and analog signals are represented as continuous waveform, with examples such as (naturally occurring) music and voice-analog data and telephone system's electronic transmission of a voice conversation-analog signal. Figure 2.2 example of analog waveform.

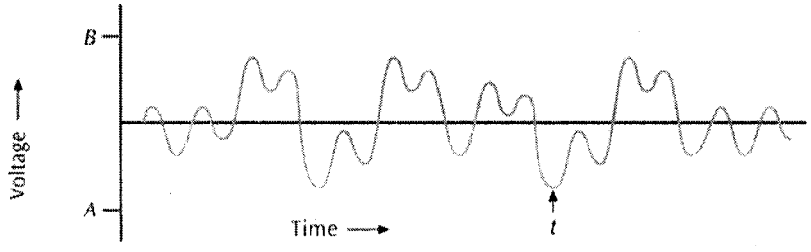


Figure 2.2: Example of Analog waveform

One primary shortcomings of analog data and analog signals is how difficult it is to separate noise from the original waveform. Noise is unwanted electrical or electromagnetic energy that degrades t he quality of signals and data.

Digital data and digital signals are discrete waveforms, rather than continuous waveforms. Digital waveform takes on only a finite number of values. Figure 2.3 is an example of a square wave where the digital waveform takes on only two different values.

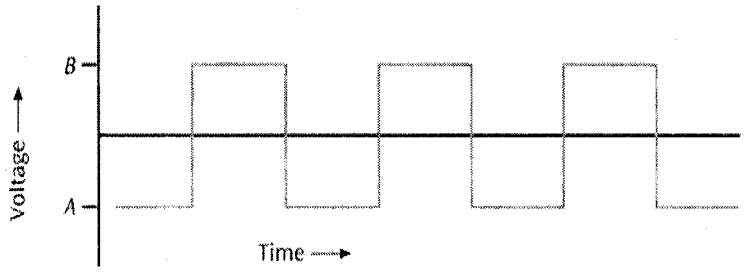


Figure 2.3: Example of Digital waveform

3.2.3 Fundamentals of Signals

There are three basic components of analog and digital signals: amplitude, frequency, and phase. As shown in figure 2.4, the amplitude of a signal is the height of the wave above (or below) a given reference point. This height often denotes the voltage level of the signal (measured in volts), but it also can denote the current level of the signal (measured in amps) or the power level of signal (measured in watts).

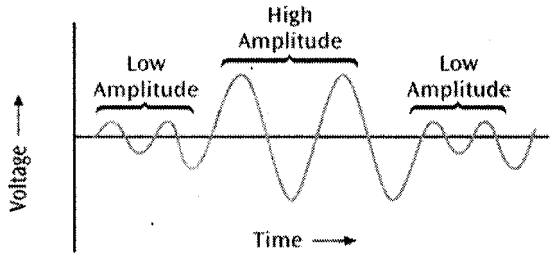


Figure 2.4: Signal with two Different Amplitude

The frequency of a signal is the number of times a signal makes a complete cycle within a given time frame. The length or time interval of one cycle is called its period. The period can be calculated by taking the reciprocal of the frequency. Figure 2.5 shows three different analog signals.

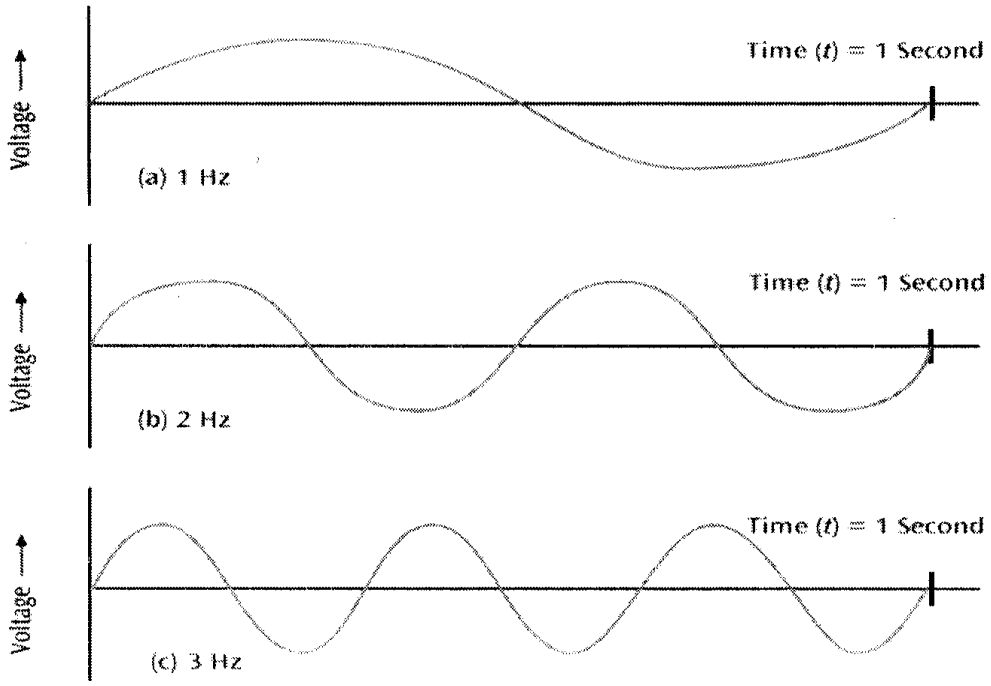


Figure 2.5: Three signals of 1Hz, 2Hz, and 3Hz

The frequency is measured in Hertz (Hz), or cycles per second. The range of frequencies that a signal spans from minimum to maximum is called the spectrum, while the bandwidth is the absolute value of the difference between the lowest and highest frequencies of a signal. For example, consider an average voice with a frequency range of roughly 300Hz to 3100Hz, the spectrum would be 300-3100Hz and the bandwidth would be 2800Hz.

The phase of a signal is the position of the waveform relative to a given moment of time or relative to time zero. A change in phase can be any number of angles between 0 and 360 degrees. Phase changes often occur on common angles, such as 45, 90, 135, etc.

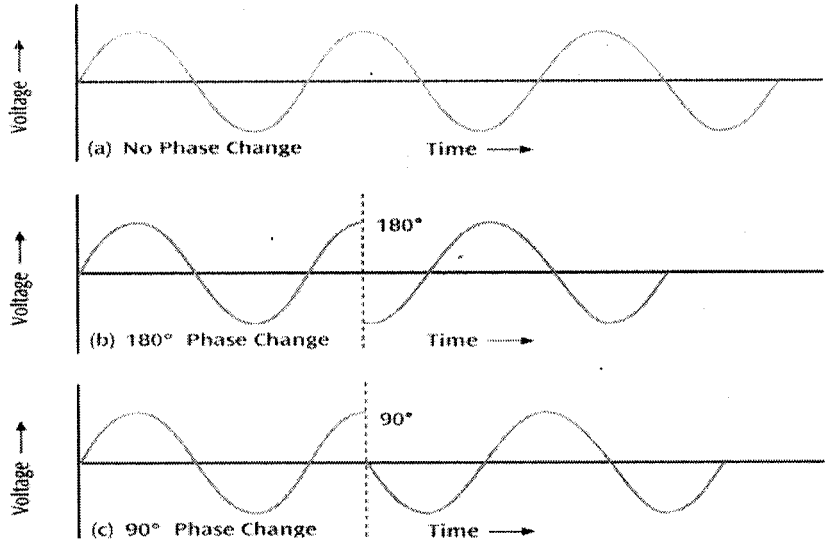


Figure 2.6: A sine wave showing (a) no phase change, (b) 180 degree phase change and (c) 90 degree phase change

Please note that all signals experience loss (attenuation). Attenuation is denoted as a decibel (Db) loss. Decibel losses (and gains) are additive. The dB is a relative measure of signal loss or gain and is expressed as dB = 10 log 10 (P2/P1) where P1 and P2 are the ending and beginning power levels, respectively of the signal expressed in watts. If a signal starts at a transmitter with 10 watts of power and arrives at a receiver with 5 watts of power, the signal loss in dB is calculated as follows:

dB = 10 log10 (5/10)
= 10 log10 (0.5)
= 10 (-0.3)
= -3

3.3 Converting Data into Signals

Like signals, data can be analog or digital. Typical analog signals convey analog data, digital signal convey digital data. However, you can use analog signals to convey digital data and digital signals to convey analog data. The decision about whether to use analog or digital signals often depends on the transmitting equipment and the environment in which the signals must travel. There are four main combinations of data and signals. Let us examine each of this in turn.

3.3.1 Analog Data Transmitted Using Analog Signal

In order to transmit analog data, you can modulate the data onto a set of analog signals. Modulation is the process of sending data over a signal by varying either its amplitude, frequency, or phase. Broadcast radio and television are two very common examples of this. Consider figure 2.7, which shows AM radio as an example. The audio data that is generated by radio station might appear like the first sine wave shown in the figure. To convey this analog data, the station uses a carrier wave signal, like that shown in figure 2.7(b). In modulation process, the original audio waveform and the carrier wave are essentially added together to produce the third waveform.

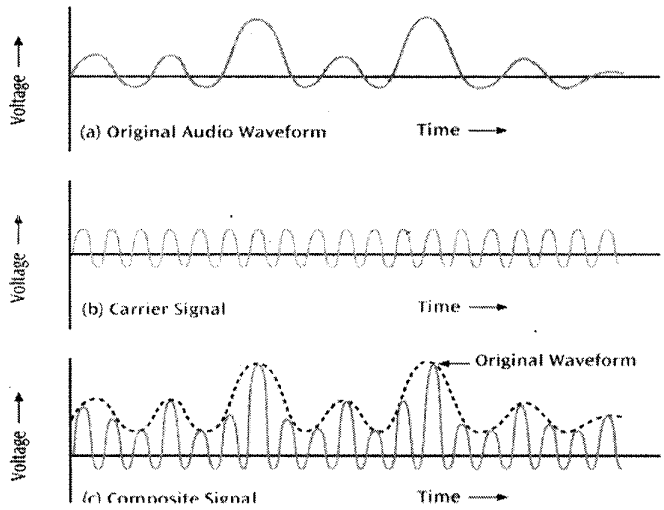


Figure 2.7: An audio waveform modulated into carrier frequency using Amplitude Modulation

3.3.2 Digital Data Transmitted Using Digital Signals

To transmit digital data using digital signals, the 1s and 0s of the digital data must be converted to the proper physical form that can be transmitted over a wire or airwave. Thus if you wish to transmit a data value 1, you could do this bit transmitting a positive voltage on the medium. If you wish to transmit a data value of 0, you could transmit a zero voltage. You could also use the opposite scheme: a data value of 0 is positive voltage, and a data value of 1 is a zero voltage. Digital encoding schemes like this are used to convert the 0s and 1s of digital data into appropriate transmission form. There are numerous techniques available to convert digital data into digital signals. Let's define five of these techniques:

- **Nonreturn-to-Zero-leve (NRZ-L)**
  0 = high level,
  1 = low level.
- **Nonreturn to Zero Inverted (NRZ-I)**
  0 = no transition at beginning of interval (one bit time)
  1 = transition at the beginning of interval
- **Manchester**
  0 = transition from high to low in middle of interval
  1 = transition from low to high in middle of interval
- **Differential Manchester**
  Always a transition in middle of interval
  0 = transition at the beginning of interval
  1 = no transition at the beginning of interval
- **Bipolar AMI**
  0 = no line signal

1 = positive or negative level, alternating for successive ones.

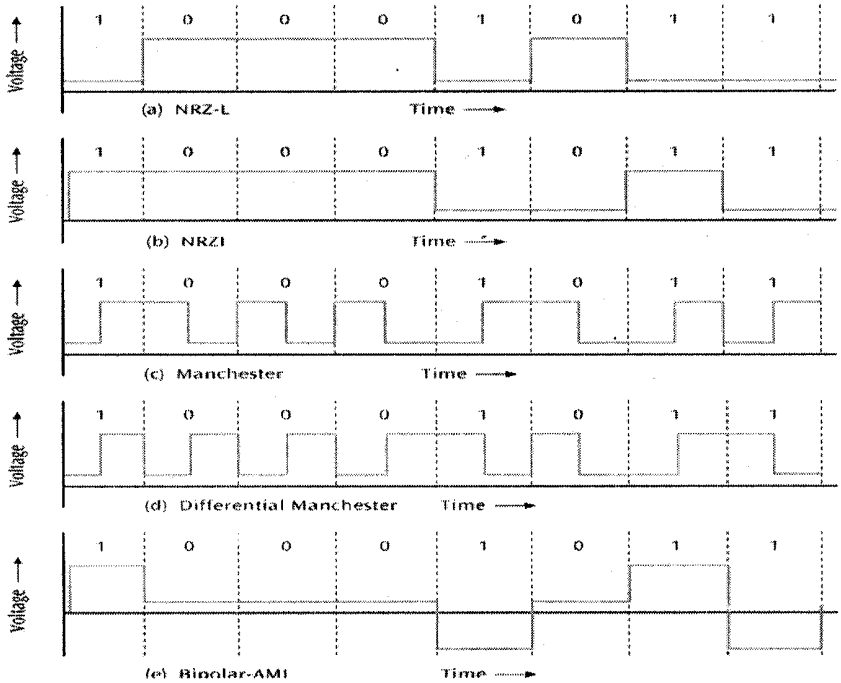


Figure 2.8: Five different encoding schemes

3.3.3 Digital Data Transmitted Analog Signals

This is also an example of modulation defined in section 3.3.1. Three currently modulated techniques for encoding digital data and transmitting it over analog signals are: Amplitude shift keying, frequency shift keying, and phase shift keying. Shift keying is a simpler form of modulation in which the binary 1s and 0s are represented by using different values of amplitude, frequency, or phase.

- **Amplitude shift keying**

In this technique one amplitude encodes a 0 while another amplitude encodes a 1 (a form of amplitude modulation). Some systems use multiple amplitudes. The question is why use multiple signal levels? The reason is not far fetched since we can represent two levels with a single bit, 0 or 1, we can represent four levels with two bits: 00, 01, 10, 11, we can represent eight levels with three bits: 000, 001, 010, 011, 100, 101, 110, 111, etc. It is to be noted that the number of levels is always a power of 2. Figure 2.9 shows amplitude shift keying.

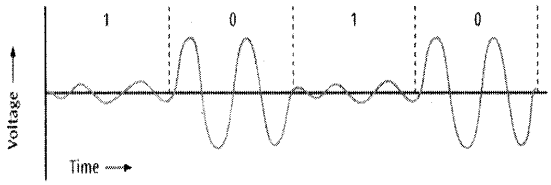


Figure 2.9(a): Example amplitude shift keying representing two amplitude levels

Amplitude shift keying has a weakness: it is susceptible to sudden noise impulses such as static charges created by a lighting storm. When transmitting data over standard telephone lines, amplitude shift keying does not exceed 1200 bps.

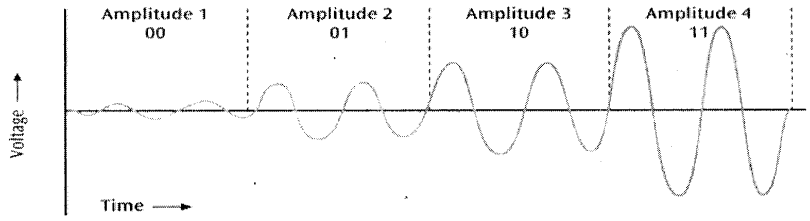


Figure 2.9(b): Example amplitude shift keying representing four amplitude levels

- **Frequency shifting keying**

Frequency shift keying uses two different frequency ranges to represent data values of 0 and 1. Lower frequency might represent a 1, while the higher frequency signal might represent 0. During each bit period, the frequency of the signal is constant. Unlike amplitude shift keying, frequency shift keying does not have a problem with sudden noise spike that can cause loss of data, which make it more robust encoding technique. Nonetheless, frequency shift keying is not perfect. It is subject to intermodulation distortion, and phenomenon that occurs when the frequencies of two or more signals mix together and create new frequencies. This technique is not used on systems that require a high data rate. Figure 2.10 show a typical example of frequency shift keying.

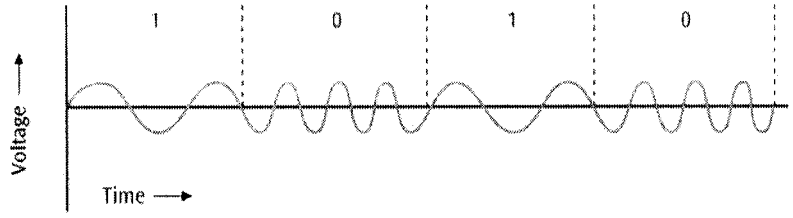


Figure 2.10: Frequency shift keying

- **Phase shift keying**

This technique represents 0s and 1s by different changes in the phase of a waveform. For example a 0 could be no phase change, while a 1 could be a phase change of 180 degrees. It is to be noted that phase changes are not affected amplitude changes, nor are they affected by intermodulation distortions. This techniques is less susceptible to noise and can be used at higher frequencies. Phase shift keying is so accurate that the signal transmitter can increase efficiency by introducing multiple phase-shift angles. For example Quadrature Phase Shift Keying incorporate four different phase angles 45 degrees, 135 degrees, 225 degrees, and 315 degrees, each of which represent 2 bits. Figure 2.11 (a) and (b) shows simple phase keying and Quadrature phase shift keying respectively.

Before we move on to last combination of data signals, let us look at data transfer rate. Data transfer rate is the number of bits transfer in one second measured in bps. One question is how do you send data faster? The best technique is to use a higher frequency signal (make sure the medium can handle the higher frequency), or use a higher number of signal levels. In both cases, noise can be a problem. The maximum data transfer rates can be calculated using **Shannon's equation:** S(f) x log2 (1 +S/N).

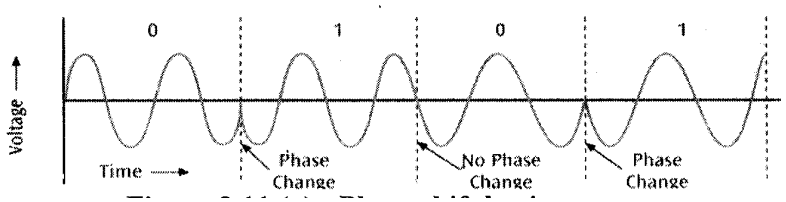


Figure 2.11 (a): Phase shift keying

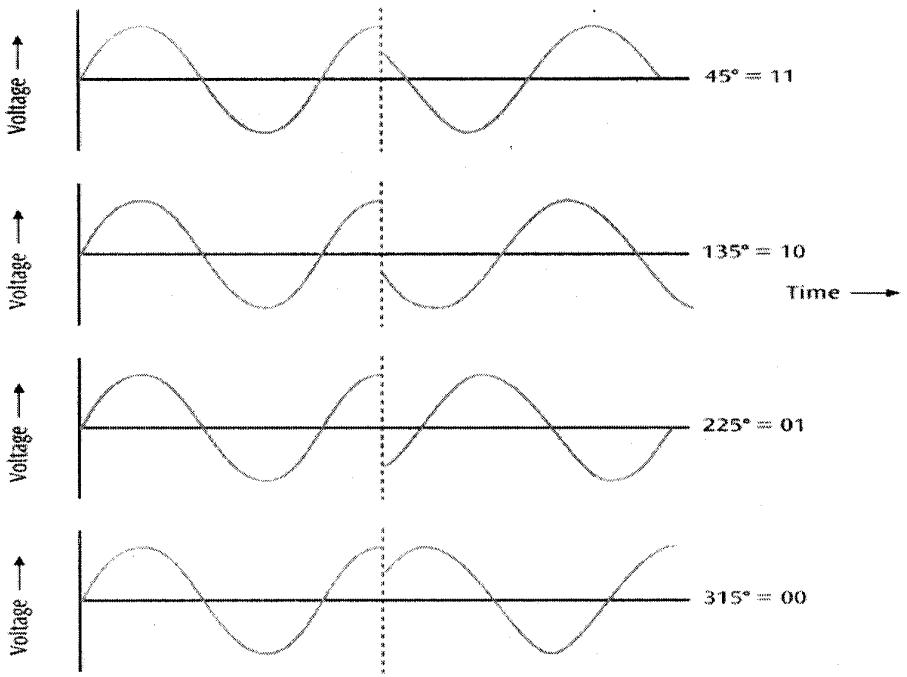


Figure 2.11 (b): Four phase angle in Quadrature phase shift keying

Where f = signal frequency (bandwidth), S is the signal power in watts, and N is the noise power in watts. For example, what is the data rate of a 3400Hz signal with 0.2 watts of power and 0.0002 watts of noise?

S(f) = 3400 x log2 (1 + 0.2/0.0002)
= 3400 x log2 (1001)
= 3400 x 9.97
= 33898 bps.

3.3.4 Analog Data Transmitted Using Digital Signals

It is often necessary to transmit data over a digital medium. For example, many scientific laboratories have testing equipment that generates the results as analog data. This analog data is converted to digital signals so that the original data can be transmitted through a computer system and eventually stored in memory or a magnetic disk. An artist performs a song that produces music, which is analog data. A device then converts this analog data to digital data so that the binary 1s and 0s of the digitized music can be stored, edited and eventually recorded on a compact disc. Two techniques for converting analog data to digital signals are pulse code modulation (PCM) Delta Modulation.

- **Pulse Code Modulation**

In this technique, the analog waveform is sampled at specific intervals and the "snapshots" are converted to binary values. When the binary values are later converted to an analog signal, a waveform similar to the original results. The more snapshots taken in the same amount of time, or the more quantization levels, the better the resolution Since telephone systems digitize human voice, and since the human voice has a fairly narrow bandwidth, telephone systems can digitize voice into either 128 or 256 levels. These are called quantization levels. If 128 levels, then each sample is 7 bits (2 /\ 7 = 128). If 256 levels, then each sample is 8 bits (2 /\ 8 = 256). Figure 2.12 shows an example of PCM. At time t (on the x-axis) a snapshot of the analog waveform is taken, resulting in the decimal value 14 on the y-axis. The 14 is converted to a 5-bit binary value(O1110) by codec and transmitted to the device for storage. In figure 2.12 the y- axis is divided into 32 gradations, or quantization levels. The encoding process continues this way. To reconstruct the original analog waveform from the store digital values, special hardware converts each binary value back to decimal and generates an electric pulse of appropriate magnitude.

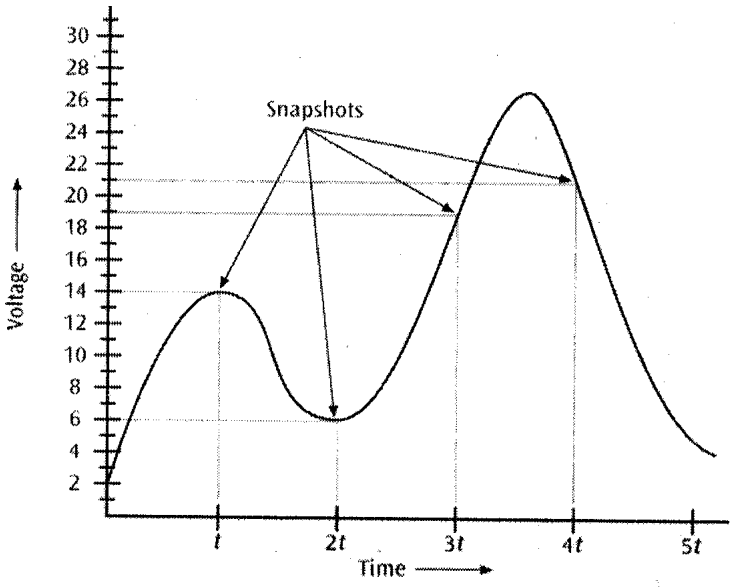


Figure 2.12: Example of taking Snapshots of an analog waveform

- **Delta Modulation**

In this technique an analog waveform is tracked, using a binary 1 to represent a rise in voltage, and a 0 to represent a drop. With this encoding technique, only 1 bit per sample is generated. See figure 2.13.

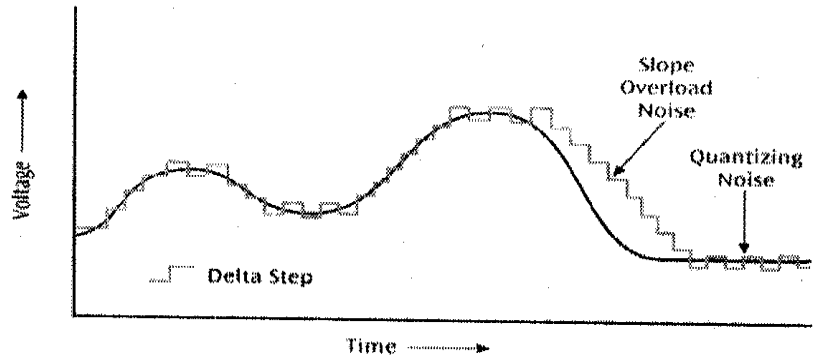


Figure 2.13: Delta Modulation

SELF-ASSESSMENT EXERCISE 1

i. What is the difference between Data and Signals?
ii. List the three basic components of all signals
iii. Define the spectrum of a signal
iv. Give two examples of (i) data (ii) signals
v. Consider a 300-Hz signal with a power level of 0.2 and a noise level of 0.0002 watts. Calculate the data rate.

3.4 Data Codes

One of the most common form of data transmitted between a transmitter and a receiver is textual data. For example banking institutions that wish to transfer money often transmit textual information, such as account numbers, name of account owners, bank names, addresses, and the amount of money to be transferred. This textual information is transmitted as a sequence of characters. To distinguish on character from another, each character is represented by unique binary pattern of 0s and Is. The set of all textual character s or symbols and their corresponding binary patterns is called EBCDIC, ASCII, and UNICODE.

3.4.1 Extended Binary Coded Decimal (EBCDIC)

The EBCDIC is an 8-bit code allowing 256 possible combinations of textual symbols. The 256 combinations of textual symbols include all uppercase and lowercase letters, digits 0 to 9, a large number of special symbols and punctuation marks, and a number of control characters. IBM mainframe computers are the major users of the EBCDIC character sets. Table 2.2 shows EBCDIC character sets.

Example: using EBCDIC show the binary equivalent of $1200.00

$ 01011011
1 11110001
2 11110010
0 11110000
0 11110000
. 01011100
0 11110000
0 11110000

3.4.2 American Standard Code for Information Interchange (ASCII)

The ASCII is a government standard in the United States and is one of the most widely used data codes in the world. ASCII character set exists

Table 2.2: EBCDIC character sets

Bits				4	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
				3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
				2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
				1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
8	7	6	5																	
0	0	0	0	NUL	SOH	STX	EXT	PF	HT	LC	DEL			SMM	VT	FF	CR	SO	SI
0	0	0	1	DLE	DC_1	DC_2	DC_3	RES	NL	BS	IL	CAN	EM	CC		IFS	IGS	IHS	IUS
0	0	1	0	DS	SOS	FS		BYP	LF	EOB	PRE			SM			ENQ	ACK	BEL
0	0	1	1			SYN		PN	RS	UC	EOT					DC_4	NAK		SUB
0	1	0	0	SP												<	(	+	|
0	1	0	1	&										!	$	.	)	:	¬
0	1	1	0	—												%	-	>	?
0	1	1	1													@		=	"
1	0	0	0		a	b	c	d	e	f	g	h	i						
1	0	0	1		j	k	l	m	n	o	p	q	r						
1	0	1	0			s	t	u	v	w	x	y	z						
1	0	1	1																
1	1	0	0		A	B	C	D	E	F	G	H	I						
1	1	0	1		J	K	L	M	N	O	P	Q	R						
1	1	1	0			S	T	U	V	W	X	Y	Z						
1	1	1	1	0	1	2	3	4	5	6	7	8	9						

To represent the value $1200.00 using ASCII the corresponding characters will be:

$ 0100100
1 0110001
2 0110010
0 0110000
0 0110000
. 0101110
0 0110000
0 0110000

3.4.3 UNICODE

One of the major problems with both EBCDIC and ASCII is that they cannot represent symbols other than those found in English language. Further they cannot even represent all the different types of symbols in English language, for example maI1Y of the technical symbols used in engineering and mathematics. And what if we want to represent the other languages around the world? For this what we need is a more powerful encoding technique-UNICODE. UNICODE is an encoding technique that provides a unique coding value for every character in every language, no matter what platform. Currently UNICODE supports more than 110 different code charts. Many of the computer companies such as Apple, HP, IBM, Microsoft, Oracle, Sun, and Unisys have accepted UNICODE(16 bit code). You can view the UNICODE web site at www.unicode.org.

SELF-ASSESSMENT EXERCISE 2

i. Using EBCDIC, ASCII, and UNICODE character code sets, what are the binary encodings of the message " Transfer".

ii. Explain the major advantages of UNICODE over other character code sets.

Table 2.3: ASCII character sets

Low-Order Bits (4, 3, 2, 1)	000	001	010	011	100	101	110	111
0000	NUL	DLE	SPACE	0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	'	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	—	o	DEL

ANSWER TO SELF-ASSESSMENT EXERCISE

i. (a) Data: Data are entities that convey meaning (computer file, music on CD, results from a blood gas analysis machine) while Signals are the electric or electromagnetic encoding of data (telephone conversation, we page download).

(b) Three basic components of all signals: Amplitude, Frequency, and Phase

(c) Spectrum of a signal is the range of frequencies that a signal spans from minimum to maximum

(d) $S(f) = 3000 * \log_2 (1 + 0.2/0.002)$
$= 3000 * 10g_2 (1001)$
$= 3000 * 9.97$
$= 29910$ bps

ii. (a)

	EBCDIC	ASCII	UNICODE
T	11100011	1010100	0000000001010100
r	10011001	1110010	0000000001110010
a	10000001	1100001	0000000001100001
n	10010101	1101110	0000000001101110
s	10100010	1110011	0000000001110011
f	10000110	1100110	0000000001100110
e	10000101	1100101	0000000001100101
r	10001001	1110010	0000000001110010

(b) UNICODE is a encoding technique that provides a unique coding value for every character in every language, no matter what platform. Currently UNICODE supports more than 110 different code charts.

4.0 CONCLUSION

In this unit, we discussed the two most basic building blocks of any computer network; data and signals. We identified three basic components of a signal. To convert data into signals we discussed various encoding and modulation techniques. Finally we identified three different data codes and how they are used to represent data.

5.0 SUMMARY

Data and signals are the two building blocks of computer networks. All signals consist of three basic components: amplitude, frequency, and phase. Noise and attenuation are the two important factors affecting the transfer of a signal over a medium. Because data and signals can be either digital or analog, four combinations of data and signals are possible: analog data converted to analog signal, digital data converted to digital signal, digital data converted to analog, and analog data converted to a digital signal. To transmit analog data overran analog signal, the analog waveform of the data is combined with another analog Wave form in a process known as modulation. Digital data carried by digital signals is represented by digital encoding formats, including the popular Manchester encoding schemes. For digital data to be transmitted

using analog signals, the digital data must first under go a process called shift keying or modulation. Three basic techniques of shift keying are amplitude shift keying, frequency shift keying, and phase shift keying. Two common techniques of converting analog data into digital signals are pulse code modulation and Delta modulation.

To transmit letters, numbers, symbols, and control characters found in text data, data codes are necessary. Three important data codes are EBCDIC, ASCII, and UNICODE.

6.0 TUTOR-MARKED ASSIGNMENT

1. What is the frequency in Hertz of a signal that repeats 80,000 times within 1 minute? What is its period?
2. Draw in chart form as shown in figure 2.8 the voltage representation of the bit pattern 11010010 for the digital encoding schemes NRZ-L, NRZI, Mnachester, Differential Manchester, and bipolar-AMI.
3. Using Shannnon' s theorem, calculate the data transfer rate given the following information:

signal frequency	=	10,000 Hz
signal power	=	5000 watts
noise power	=	230 watts.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). Data *Communications and Computer Network*, A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). *Data Communication Principles for Fixed and Wireless Networks,* Kluwer Academic Publisher, USA.

UNIT 3 TRANSMISSION MEDIA

CONTENTS

1.0 INTRODUCTION

The world of computer network will exists if there were no medium by which to transmit data. Recall that in Unit land 20f this course we described some major components of computer networks. The. transmission medium is another important component of computer networks. This unit will investigate each of these media.

2.0 OBJECTIVES

Up successful completion of this Unit, you should be able to:

- Outline the physical description, transmission characteristics, advantages, disadvantages, application areas of guided and unguided transmission media.
- Identify media selection criteria.

3.0 MAIN CONTENT

3.1 Design Factors

The transmission medium is the physical path between transmitter and receiver in a data transmission system. Transmission media can be classified as guided or unguided. In both cases, communication is in form of electromagnetic waves. With guided media, the waves are guided along the solid medium such as copper twisted, copper coaxial, and optical fiber. The atmosphere and outer space are examples of unguided media (wireless media) that provide a means of transmitting electromagnetic signals but do not guide them. The characteristics and quality of a data transmission are determined both by the characteristics of the medium and the characteristics of the signal. In considering the

design of data transmission systems, a key concern, generally is data rate and distance: the greater the data rate and distance the better. A number of design factors relating to the transmission medium and to the signal determine data rate and distance. Some of the design factors include:

- **Bandwidth**

  The greater the bandwidth of a signal, the higher the data rate that can be achieved.
- **Transmission impairments**
  Impairments, such as attenuation, limit the distance. For guided media, twisted pair generally suffer more impairment than coaxial cable.
- **Interference**
  Interference from competing signals in overlapping frequency bands can distort or wipe out a signal. This factor is of particular concern for the unguided media, but it is also a problem with guided media. For guided media, interference can be caused by emanations from nearby cables. For example, twisted pair are often bundled together, and conduits often carry multiple cables. Proper shielding of a guide medium can minimize this problem
- **Number of receiver**
  A guided medium can be used to construct a point-to-point link or a shared link with multiple attachments. In the later case, each attachment introduces some attenuation and distortion on the line, limiting distance and/or data rate.

3.2 Guided Transmission Media

For guided transmission media, the transmission capacity, in terms of either data rate or bandwidth, depends critically on the distance and on whether the medium is point-to-point or multipoint, such as in LAN. Table 3.1 indicates the type of performance typical for the common guided medium for long distance point-to-point applications. The three guided media commonly used for data transmission are twisted pair, coaxial cable, and optical fiber. We examine each of this in turn.

Table 3.1 : Point-to-point transmission characteristics of guided media

Transmission medium	Total data rate	Bandwidth	Repeater spacing
Twisted pair	4Mbps	3MHz	2 to 10km
Coaxial cable	500 Mbps	350 MHz	1 to 10km
Optical fiber	2 Gbps	2 GHz	10 to 100km

- **Twisted pair**

This is the least-expensive and most widely-used guided transmission medium. It consists of two insulated copper wires arranged in a spiral pattern. A wire pair acts as a single communication link. Typically, a number of these are bundled together into a cable by rapping them in tough protective sheath. Figure 3.1 shows the twisted pair.

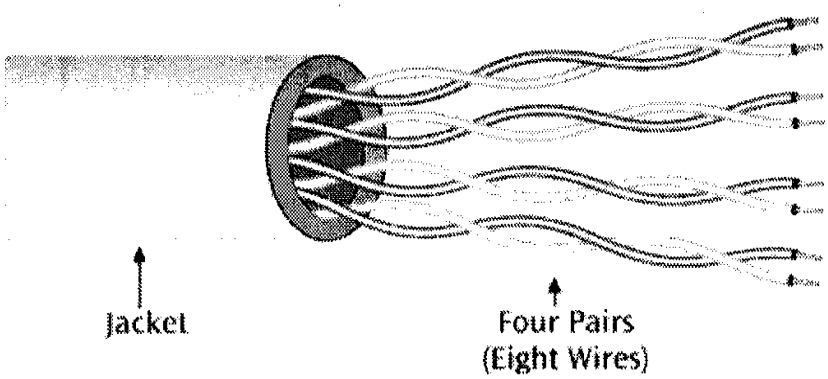


Figure 3.1 : Four pair twisted wire

Over longer distances, cables may contain hundreds of pairs. The twisted pairing tends to decrease the crosstalk interference adjacent pairs in a cable. This is as shown in figure 3.2.

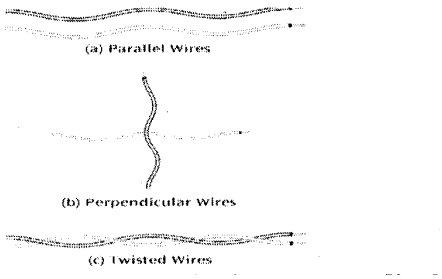


Figure 3.2 : Parallel wires, perpendicular wire, and twisted pair

The wires in a pair have thicknesses of from 0.016 to 0.036 inches. It is used for both analog and digital signals. For analog signals, amplifiers are required about every 5 to 6 km, while for digital signals, repeater are required every 2 or 3 km. It is the most commonly used medium in the telephone network as well as being the workhorse for communications within buildings for LANs supporting personal computers. Compared to other commonly used guided transmission media, twisted pair wire is limited in distance, bandwidth and data rate. The medium is quite susceptible to interference and noise because of its easy coupling with electromagnetic fields several measures are taken to reduce impairments. Shielding the wire with metallic braid or sheathing reduces

interference. Twisted pair wires in two varieties: unshielded and shielded. Unshielded twisted pair (UTP) is ordinary telephone wire. Office buildings by universal practice, are pre-wired with a great deal of excess UTP, more than is needed for simple telephone support. UTP is subject to external electromagnetic interference. A way to improve the characteristics of this medium is to shield the twisted pair with a metallic braid that reduces interference. Shielded Twisted Pair (STP) provides better performance at lower data rates. However it is more expensive and more difficult to work with than UTP. Figure 3.4 shows a typical STP.

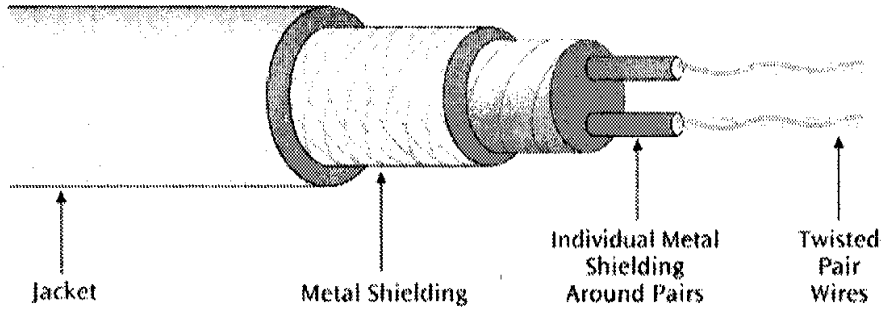


Figure 3.3: Example of Shielded Twisted Pair

Twisted pair wire is classified by category. Twisted pair is currently Category 1 through Category 7, although Categories 2 and 4 are nearly obsolete. Table 3.2 summarized the characteristics of twisted pair wire.

Table 3.2: A summary of the Characteristics of twisted pair wire

UTP Category	Typical Use	Maximum Data Transfer Rate	Maximum Transmission Range	Advantages	Disadvantages
Category 1	Telephone wire	<100 kbps	5-6 kilometers (3-4 miles)	Inexpensive, easy to install and interface	Security, noise
Category 2	T-1, ISDN	<2 Mbps	5-6 kilometers (3-4 miles)	Same as Category 1	Security, noise, obsolescence (?)
Category 3 Telephone circuits	LANs	10 Mbps	100m (328ft) with less noise	Same as Category 1	Security, noise
Category 4	LANs	20 Mbps	100m (328ft)	Same as Category 1, with less noise	Security, noise, obsolescence
Category 5	LANs	100 Mbps (100 MHz)	100m (328ft)	Same as Category 1, with less noise	Security, noise
Category 5e	LANs	250 Mbps per pair (125 MHz)	100m (328ft)	Same as Category 5. Also includes specifications for connectors, patch cords, and other components	Security, noise
Category 6	LANs	250 Mbps per pair (200 MHz)	100m (328ft)	Higher rates than Category 5, less noise	Security, noise, cost
Category 7	LANs	600 MHz	100m (328ft)	High data rates	Security, noise, cost

- **Coaxial cable**

Coaxial cable, like twisted pair consists of two conductors, but is constructed differently to permit it to operate over wider range of frequencies. It consists of a hollow outer cylindrical conductor that surrounds a single inner wire conductor. The inner conductor is held in place by either regularly spaced insulating rings or a solid dielectric material. The outer conductor is covered with a shield or jacket. A coaxial cable has a diameter of from 0.4 to about 1 inch. It much less susceptible to interference and crosstalk than twisted pair because of its shielded, concentric construction. The principal constraints on performance are attenuation, thermal noise, and intermodulation noise This cable is the most versatile transmission medium and is enjoining widespread use in a, wide variety of applications; the most important of these are: Television distribution, long-distance telephone transmission, short-run computer system links, LANs, etc.

This cable is used to transmit both analog and digital signals. Figure 3.4 and 3.5 shows example of coaxial cable and two types available (thick and thin) respectively.

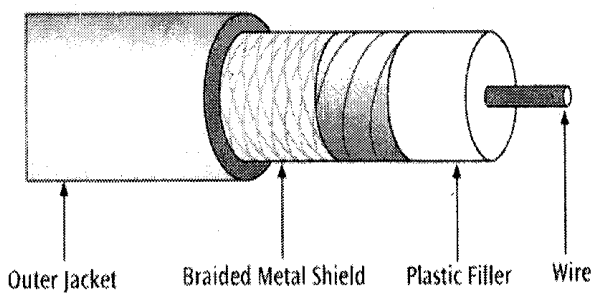


Figure 3.4: Example of coaxial cable

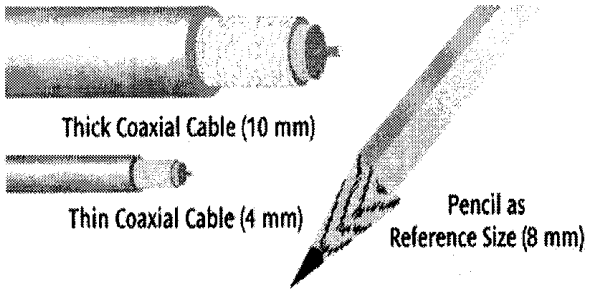


Figure 3.5 : Thick and Thin coaxial cable

- **Optical Fiber**

An optical fiber is a thin (2 to 125 micro meter), flexible medium capable of conducting an optical ray. Various and plastics can be used to make optical fibers. An optical fiber cable has a cylindrical shape and consists of three concentric sessions: the core, cladding, and the jacket. The core is the inner most section and consists of one or more very thin strands, or fibers, made of glass or plastic. Each fiber is surrounded by its own cladding, a glass or plastic coating that has optical properties different from those of the core. The outermost layer, surrounding one or a bundle of cladded fibers, is the jacket. The jacket is composed of plastic and other material layered to protect against moisture, abrasion, crushing and other environmental dangers.

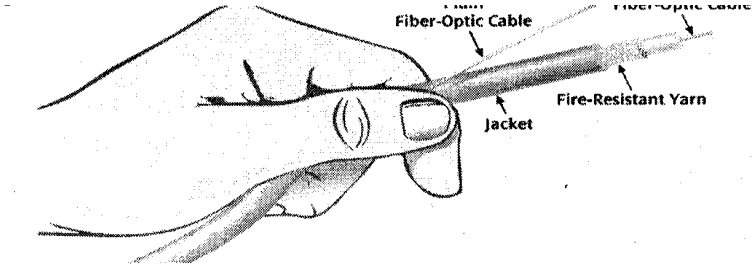


Figure 3.6: Fiber optic cable

The following characteristics distinguish optical fiber from twisted pair or coaxial cable:

Greater capacity: The potential bandwidth, and hence the data rate is of immense; data rate of 2 Gbps over tens of kilometers.

Smaller size and lighter weight: Considerably thinner than coaxial cable or bundled twisted pair.

Lower attenuation: Considerably lower and is constant over a range

Electromagnetic isolation: Not affected by external electromagnetic fields. The system is not vulnerable to interference, impulse noise, or cross talk Five basic categories of application have become important for optical fiber: Long-haul trunks, Metropolitan trunks, Rural-exchange trunks, subscriber loops, and local area networks. Figure 3.7 shows a typical application of fiber in a local area network.

Figure 3.7: Fiber linking two LANs

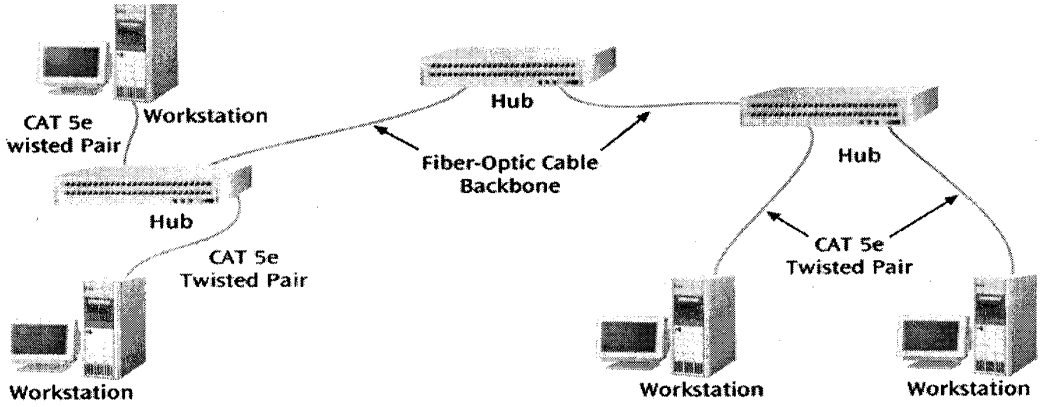


Figure 3.7: Fiber linking two LANs

It operate in the range of about 1014to 1015 Hz. The principle of optical transmission is as follows: Light from a source enters the cylindrical glass or plastic core. Rays at shallow angles are reflected and propagated along the fiber (multimode). When the fiber core radius is reduced, fewer angles will reflect. By reducing the radius of the core to the order of a wavelength, only a single angle or mode can pass. This single mode propagation provides superior performance for the following reason: with multi mode transmission, multiple propagation paths exist, each with a different path length and, hence, time to traverse the fiber; this causes signal elements to spread out in time, which limit the rate at which data can be accurately received. Because there is a single transmission path with single-mode transmission, such distortion cannot occur. Finally by varying the index of refraction of the core, a third type of transmission, known as multimode graded index, is possible. This type is intermediated between the other two in characteristics. Two different types of light source are used in fiber optic systems: the light emitting diode and the injection laser diode which are semiconductor devices which emit a beam of light when a voltage is applied Table 3.3 give a summary of the characteristics of conducted media.

Table 3.3: Summary of characteristics of conducted wire

Type of Conducted Medium	Typical Use	Maximum Data Rate	Maximum Transmission Range	Advantages	Disadvantages
Twisted pair Category 1, 3	Telephone systems	<2 Mbps	5-6 kilometers (3-4 miles)	Inexpensive, common	Noise, security
Twisted pair Category 5, 5e, 6, 7	LANs	100-1000 Mbps 10 Mbps	100m (328ft) 100m (328ft)	Inexpensive, versatile	Noise, security
Thin Coaxial Cable (basedband single channel)	LANs	10Mbps	100m (328ft)	Low noise	Security
Thick Coaxial Cable (broadband multichannel)	LANs, cable TV, long-distance telephone, short-run computer system links	10-100 Mbps	5-6 kilometers (3-4 miles) (at lower data rates)	Low noise, multiple channels	Security
LED Fiber-Optic	Data, video, audio, LANs,	Gbps	300 meters (approx. 1000ft)	Secure, high capacity, low noise	Interface expensive but decreasing in cost
Laser Fiber-Optic	Data, video, audio, LANs, WANs, MANs	100s Gbps	100 kilometers (approx. 60 miles)	Secure, high capacity, very low noise	Interface expensive

3.3 Wireless Transmission

In wireless transmission, various types of electromagnetic waves are used to transmit signals. Radio transmissions, satellite transmissions, visible light, infrared light, X-rays, and gamma rays are all examples of electromagnetic waves or electromagnetic radiation. In general, electromagnetic radiation is energy propagated through space and, indirectly, through solid objects in the form of advancing disturbance of electric and magnetic fields. The basic difference between various types of electromagnetic waves is their differing wavelengths, or frequencies, as shown in figure 3.8.

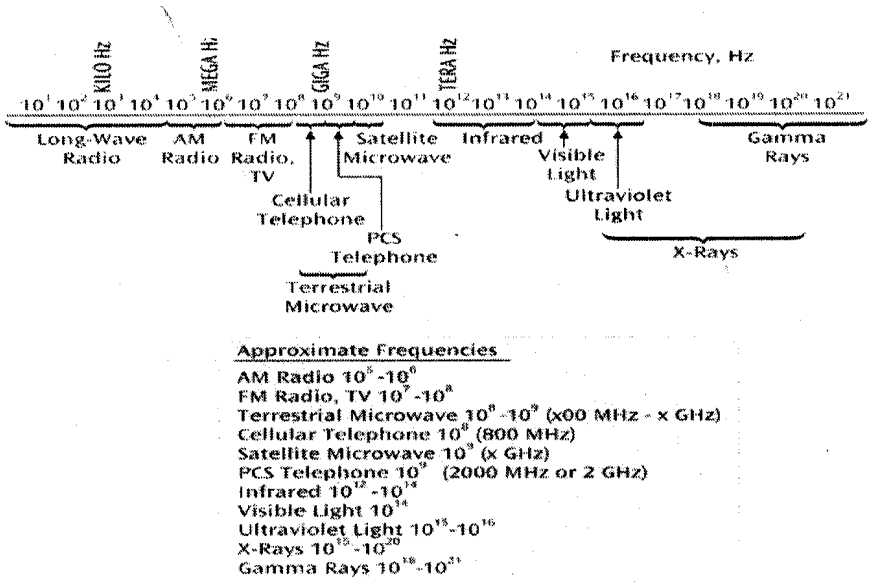


Figure 3.8: Electromagnetic wave frequencies

- **Terrestrial microwave**

Transmit tightly focused beams of radio signals from one-ground-based microwave transmission antenna to another. The two most common application areas are telephone communications and business intercommunication. The approximate distance between towers is between 20 -30 miles. The higher the tower, the farther the possible transmission distance. It is to be noted that signals will not pass through solid objects. Terrestrial microwave transmits data at hundred of millions of bits per second. Microwave antennas use line-of-site transmission, which means that to receive and transmit a signal, each antenna must be in sight of the next antenna as shown in figure 3.9. The disadvantages of terrestrial microwave can include loss of signal strength (attenuation) and interference from other signals (intermodulation), in addition to the costs of either leasing the service or installing and maintaining the antennas.

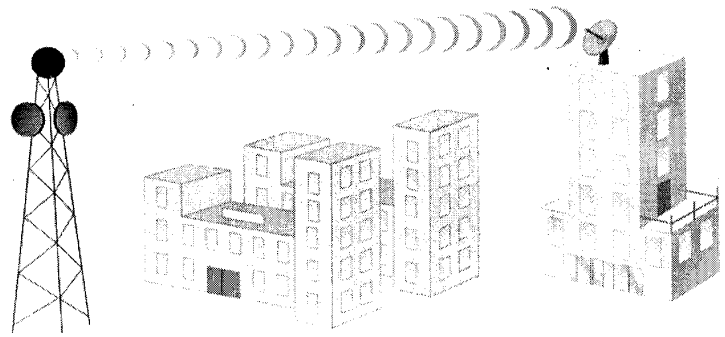


Figure 3.9: A microwave antenna

- **Satellite microwave transmission**

This transmission systems are similar to terrestrial microwave systems except that the signal travels from ground station on Earth to a satellite and back to another ground station on Earth, thus achieving much greater distances than Earth-bound line-of-sight transmissions. One way to categorizing systems is how far the satellite is from the Earth. The closer a satellite is to the Earth, the shorter the times required to send data to the satellite(uplink) and receive data from the satellite (downlink) .This transmission time from ground station to satellite and back to ground station is called propagation delay. Satellites can be classified by how far out into orbit each one is (LEO, MEO, GEO, and HEO). LEO –Low Earth Orbit --100 to 1000 miles out. Used for wireless e-mail, special mobile telephones, pagers, spying, videoconferencing. MEO -Middle Earth Orbit -1000 to 22,300 miles. Used for GPS (global positioning 'systems) and government. GEO -Geosynchronous Earth Orbit -22,300 miles. Always over the same position on earth (and always over the equator). Used for weather, television, government operations. HEO -Highly Elliptical Earth orbit

-satellite follows an elliptical orbit. HEO is used by the military for spying and by scientific organizations for photographing celestial bodies. Satellite microwave can also be classified by its configuration: Bulk carrier configuration, multiplexed configuration, and single-user earth station configuration (e.g. VSAT). Figure 3.10 shows Earth and four Earth orbits: LEO, MEO,GEO, HEO, figure 3.11 shows satellite configuration.

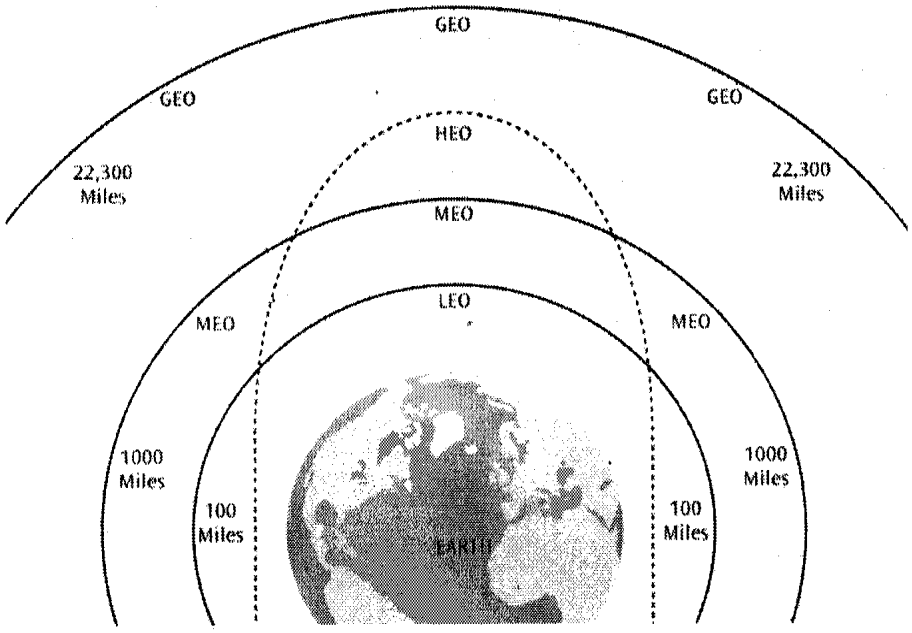


Figure 3.11: Satellite configuration

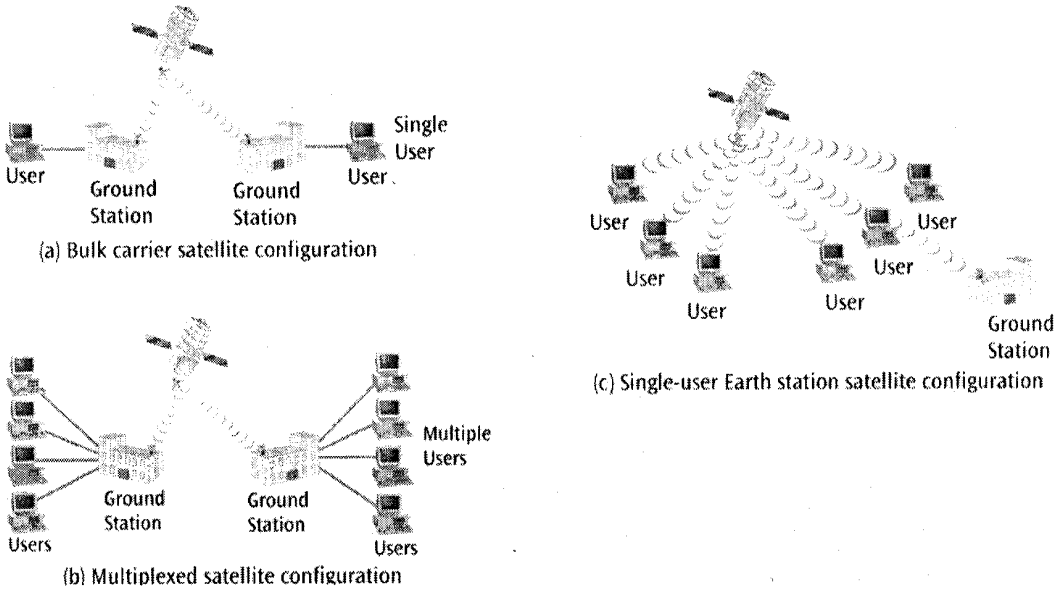


Figure 3.11: Satellite Configuration

- **Cellular Telephone**

This wireless technology uses radio waves. First generation include AMPS - Advanced Mobile Phone Service -first popular cell phone service; used analog signals and dynamically assigned channels, D-AMPS -Digital AMPS –applied digital multiplexing techniques on top of AMPS analog channels. The second generation include PCS-Personal Communication Systems -essentially all-digital cell phone service. PCS phones came in three technologies:

- TDMA -time division multiple access
- CDMA -code division multiple access
- GSM -global system for mobile communications

The third generation cellular phone include UMTS (Universal Mobile Telecommunications System) -also called Wideband CDMA. Other wireless transmission include: Infrared transmission, wireless application protocol, broadband wireless distribution services, short-range transmissions such as Bluetooth, and wireless area network systems. Table 3.4 summarized the wireless media.

SELF-ASSESSMENT EXERCISE 1

i. For what purposes are category 1 -7 twisted pair used for?
ii. Identify the characteristics that distinguishes optical fiber from twisted pair or coaxial
iii. What is the difference between Terrestrial microwave and Satellite microwave?

Table 3.4: Summary of wireless Media

Type of Wireless Medium	Typical Use	Maximum Data Transfer Rate	Maximum Transmission Range	Advantages	Disadvantages
Terrestrial Microwave	Long-haul telecommunications, building to building	100s-Mbps	20-30 miles	Reliable, high speed, high volume	Long haul, expensive to implement, line-of-sight
Satellite LEO	Communications such as e-mail, paging, worldwide mobile phone network, spying, remote sensing, video conferencing	100s-Mbps	Depends on number of satellites	High-speed transfers, very wide distance, some applications inexpensive	Some application expensive, interference
Satellite MEO	GPS-style surface navigation systems	100s-Mbps	Depends on number of satellites	High-speed transfers, wide distance	Expensive to lease, some interference
Satellite GEO	Signal relays for cable and direct television	100s-Mbps	One-third the Earth's circumstance (8000 miles)	Very long distance, high speed, and high volume	Expensive to lease, some interference
Satellite HEO	Global surveillance, scientific applications	100s-Mbps	Variable	Variability of distance	Expensive
Cellular (AMPS and	Cellular telephones	19.2 kbps	Each cell: 0.5-50 mile radius, but	Widespread, inexpensive	Noise

D-AMPS)			nationwide coverage	applications	
PCS	Cellular telephones	9.6 kbps	Each cell: 0.5-25 mile radius	Digital, low noise	Slow data rates
GPRS, 1XRTT	Cellular telephones	30-75 kbps	Each cell: 0.5-25 mile radius	Digital, low noise	Slow data rates
UMTS	Cellular telephones	320 kbps		Digital, low noise	
EV-DO	Cellular telephones	500 kbps	Each cell: 0.5-25 mile radius	Digital, low noise	
Infrared	Short-distance data transfer	16 Mbps	1.5 miles	Fast, inexpensive, secure	Short distances, line of sight

3.4 Media Selection Criteria

When designing or updating a computer network, the selection of one type of medium over another is an important issue, Computer-based network projects have perform poorly and possibly even failed as a result of a poor decision about the appropriate type of medium. The principal factors to be considered include cost, speed, expandability and distance, environment.

ANSWER TO SELF-ASSESSMENT EXERCISE

1(a) Cat 1: Telephone wire
Cat 2: Telephone circuits Tl, ISDN
Cat 3-7: LANs

(b) The characteristics that distinguishes optical fiber from twisted pair or coaxial.
Greater capacity, smaller size and light weight, lower attenuation, and electromagnetic isolation.

(c) Satellite microwave transmission systems are similar to terrestrial microwave systems except that the signal travels from ground station on Earth to a satellite and back to another ground station on Earth, thus achieving much greater distances than Earth-bound line-of-sight transmissions.

4.0 CONCLUSION

In this unit, we highlighted a number of design factors (bandwidth, transmission impairments, interference, and number of receiver) relating to the transmission medium and to the signal determine data rate and distance of a data transmission systems. We discussed two categories of transmission media (wired and wireless) in terms of their physical characteristics, applications and transmission characteristics. We rounded up this unit with some of the criteria for media selection.

5.0 SUMMARY

All data communication media can be divided into two basic categories (1) conducted media such as wires and (2) radiated or wireless such as satellite systems. The three types of conducted media include twisted pair, coaxial cable, and fiber optic cable. Twisted pair and coaxial cable are both metal wires and are subject to electromagnetic interference.

Fiber optic cable is impervious to electromagnetic interference; therefore, it experiences a low noise level than coaxial and twisted pair. Fiber optic has the best transmission speeds and long-distance performance of all conducted media. Several basic groups of wireless media exists; terrestrial microwave transmissions, satellite transmissions, cellular telephone systems, infrared transmissions, WiMAX, Bluetooth, etc. Each of the wireless technologies is deigned for specific applications. When trying to select a particular medium for an application, it helps to compare the different media using five these five criteria: cost, speed, expandability, and distance, right-of-way, environment, and security.

6.0 TUTOR-MARKED ASSIGNMENT

(a) A telephone line with a bandwidth 100 kHz is known to have a loss of 20dB. The input signal power is measured as 0.5 watt, and the output signal noise level is measured as 2.5microwatt. Using this information, calculate the output signal-to-noise ratio.

(b) Explain a few common applications for terrestrial microwave.

7.0 REFERENCES/FURTHER READINGS

Stallings, W (1997). *Data and Computer Communications*, Prentice-Hall International, Inc. ,USA

Curt, M W (2007). *Data Communications and Computer Network*. A Business User's Approach Fourth Edition. Bob Woo bury, Canada.

UNIT 4 DATA COMMUNICATION INTERFACES

CONTENTS

1.0 INTRODUCTION

Recall that we identified microcomputer (server or workstation) as one of the major components of computer networks. A computer will be of no use if we could not connect it to anything. Many people feel that their computers will not be worth much if there were no way to connect the computer to printer, or if they could not connect it to modem to surf the Internet. Many people in the corporate world depend almost exclusively on an interconnection between their computer and a company's local area network. To better understand the interconnection between a computer and a peripheral device, you must be come familiar with the concept of interfacing which is the focus of this unit.

2.0 OBJECTIVES

After reading this Unit, you should be able to:

- List the four components of all interface standard.
- Recognize the difference between half-duplex and full-duplex connection.
- Outline the characteristics of asynchronous, synchronous, and isochronous data link interfaces.

3.0 MAIN CONTENT

3.1 Interfacing Standards

You will agree that a computer is fantastic tool. This tool is capable of performing a myriad of operations primarily because of two important facts: a computer is programmable, and it can connect to a wide range of

input/output devices, or peripherals. The connection to a peripheral is often called the interface, and the process of providing all the proper interconnections between a computer and a peripheral device is called interfacing. You can't discuss interfacing without discussing standards.

There are essentially two types of standards:

- Official standards, created by standards making organizations such as ITU (International Telecommunications Union), IEEE (Institute for Electrical and Electronics Engineers), EIA (Electronics Industries Association), ISO (International Organization for. Standards), and ANSI (American National Standards Institute).
- De-facto standards, protocols created by other groups that are not official standards but because of their widespread use, become "almost" standards.

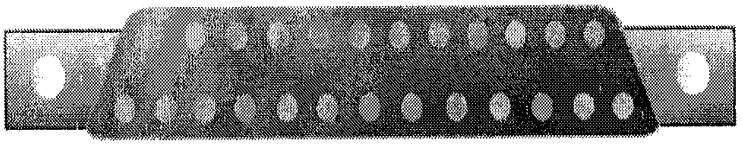
There are four possible components to an interface standard:

- The **electrical component:** deals with voltages, line capacitance, and other electrical characteristics.
- The **mechanical component** deals with items such as the connector or plug description.
- The **functional component** describes the function of each pin or circuit that is used in a particular interface.
- The **procedural component** describes how the particular circuits are used to perform an operation.

In order to better understand the four components of an interface, let's examine two popular interface standards: EIA-232F -an older standard originally designed to connect a modem to a computer, USB (Universal Serial Bus) -a newer standard that is much more powerful than EIA-232F.

- **RS-232 and EIA-232F**

Originally named RS-232 but has gone through many revisions. The electrical component is defined by another standard: V.28 while the mechanical component is often defined by ISO 2110, the DB-25 connector. The DB-9 connector is now more common than the DB-25. Figure 4.1 shows a typical DB 25 and DB 9 connection. Can you find these two connectors at the back of your computer?



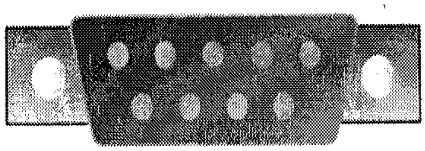


Figure 4.1: DB 25 and DB 9 connection

The functional and procedural components are defined by the V.24 standard. For example, V.24 defines the function of each of the pins on the DB-9 connector, as shown table 4.1. Now let us see an example of the procedural dialog that can be used to create a connection between two endpoints in table 4.2.

As you go through table 4.2 please note the level of complexity needed to establish a full-duplex connection. The full-duplex connection will be discussed later in this section.

Table 4.1: EIA-232F Interchange Circuits for a DB 9 Connector

Pin Number	Circuit Originates From	Circuit Names
1	DCE	Data Carrier Detect (Received Line Signal Detector)
2	DCE	Received Data
3	DTE	Transmitted Data
4	DTE	Data Terminal Ready (DTE Ready)
5	---	Ground
6	DCE	Data Set Ready (DCE Ready)
7	DTE	Request to Send
8	DCE	Clear to Send
9	DCE	Ring Indicator

Table 4.2: Sequence of command between local DTE and DCE and between remote DCE and DTE to establish a connection.

1. Local DTE turns on Data Terminal Ready (pin 4) to te11 its DCE (modem) it wants to place a call.
2. Local modem sends a DCE Ready signal (pin 6) to its DTE.
3. Local DTE sends phone number to its modem over Transmitted Data (pin 3). Modem dials number.
4. Remote DCE (modem) alerts its DTE to incoming call via Ring Indicator (pin 9).
5. Remote DTE turns on its DTE Ready (pin 4).

6. Remote modem sends a carrier signal back to local modem.
7. Remote modem sends a DCE Ready signal (pin 6) to its DTE.
8. Local modem detects carrier and alerts its DTE via Received Line Signal Detector (pin 1).
9. Local modem sends a carrier signal to remote modem.
10. Remote modem detects carrier and alerts its DTE via Received Line Signal Detector (pin 1).
11. Local DTE wishes to send data, so asserts Request to send (pin 7).
12. Local modem responds with Clear to Send (pin 8).
13. Local DTE sends digital data over Transmitted Data (pin 3) to its modem, which converts it to analog and sends over carrier to remote modem.
14. Signal arrives at remote DCE, which converts the analog signal to digital and sends it to its DTE via Received Data (pin 2).

The USB interface is a modern standard for interconnecting a wide range of peripheral devices to computers. It Supports plug and play. It can also daisy-chain multiple devices. USB 2.0 can support 480 Mbps (USB 1.0 is only 12 Mbps). The USB interface defines all four components. The electrical component defines two wires VBUS and Ground to carry a 5 volt signal, while the D+ and D- wires carry the data and signaling information. The mechanical component precisely defines the size of four different connectors and uses only four wires (the metal shell counts as one more connector).

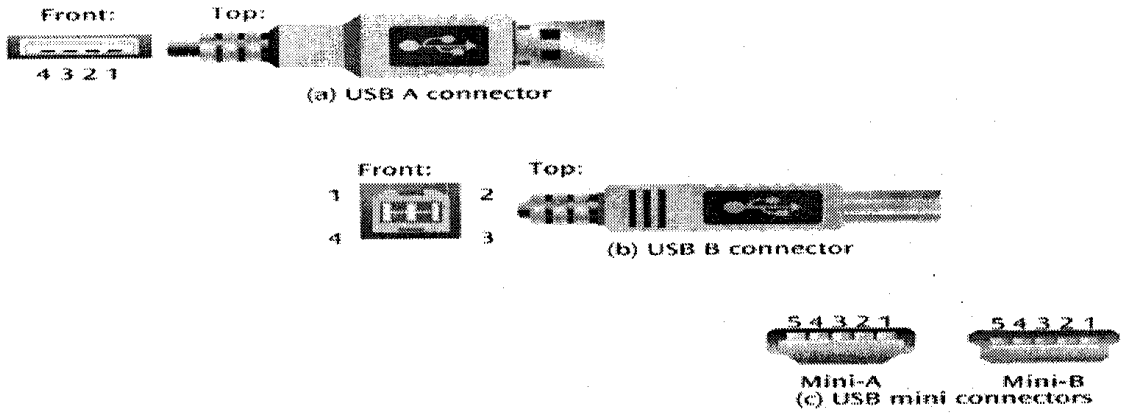


Figure 4.2: Four types of USB connector

The functional and procedural components are fairly complex but are based on the polled bus, that is the computer takes turns asking each peripheral if it has anything to send.

Other interface standards

In addition to USB, other interface standards have been created over the years to provide high-speed connections to various types of peripheral devices. Example include Fire Wire, SCSI (Small Computer System Interface),-iSCSI, InfiniteBand, and Fiber channel. Please read more on these other interfaces.

3.2 Data link connections

As you can see from section 3.1 interface standards such as EIA-232F, USB, ETC, consist of four components: electrical, mechanical, functional, and procedural. Because these four components define the physical connection between a computer and a peripheral, they reside at the physical layer of the OSI model. In order to transmit data successfully between two points on a network, such as between a network sender and a network receiver, there is the need to define the data link connections. While examining the data link connection, recall the duties of the data link layer discussed in unit 1 of this course- two of which are to create a frame of data for transmission between the sender and receiver and to provide some way of checking for errors during transmission. Keep these duties in mind as we examine three different types of data link connections.

- **Asynchronous connections**

This is one of the simplest examples of a data link protocol and is found primarily in microcomputer-to-modem and terminal-to-modem connections. To transmit data from sender to receiver, an asynchronous connection creates a one-character package called a frame, added to the front of the frame is a start bit, while a stop bit is added to the end of the frame. An optional parity bit can be added which can be used to detect errors.

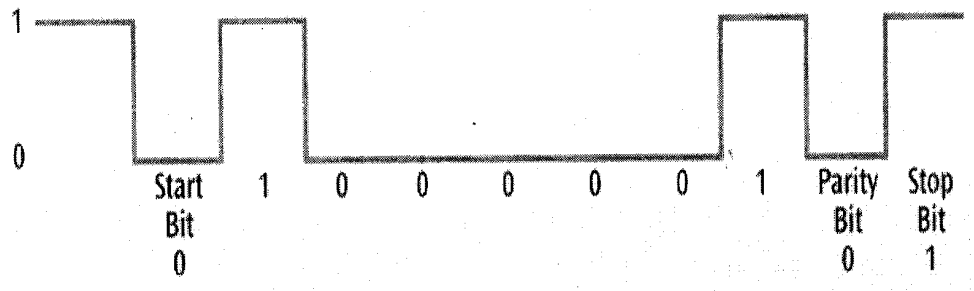


Figure 4.3: Example of character A with one start bit, one stop bit and even parity

Because each character has its own start, stop, and parity bits, the transmission of multiple characters, such as HELLO, is possible. Figure

4.4 demonstrates the transmission of HELLO. An asynchronous connection has advantages and disadvantages. On the positive side, generation of start, stop, and parity bit is simple and requires little hardware or software. On the negative side, an asynchronous connection has one disadvantage in particular that cannot be overlooked. Given that seven data bits (ASCII character code set) are often combined with one start bit, one stop bit, and one parity bit, the resulting transmitted character contains three check bits and seven data bits, for a 3:7 ratio.

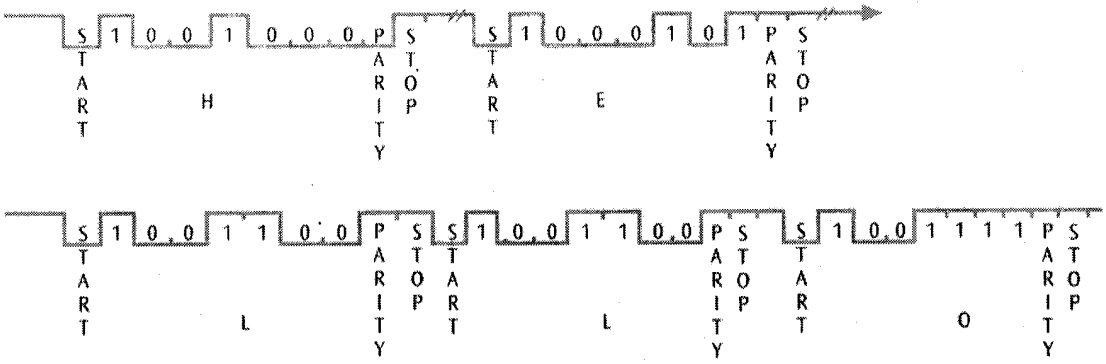


Figure 4.4: Example of Asynchronous connections

In this scenario of 10 total check and data bits, 3 out of 10 or 30% of the bits are used as check bits. This ratio of check bits to data bits is not very efficient for high amounts of data transfer and therefore, results in slow data transfers. The term asynchronous is misleading here because you must always maintain synchronization between the incoming data stream and the receiver. Asynchronous connections maintain synchronization by using small frames with a leading start bit.

Synchronous connections

The second type of data link connection(with less misleading name) is the synchronous connection. With a synchronous connection, the unit of transmission is a sequence of characters. This sequence of characters may be thousands of characters in size. A synchronous connection creates a large frame that consists of header and trailer flags, control information, optional address information, error detection code, and data. Figure 4.5 shows block diagram of synchronous connection.

01111110	…010110…	…110010…	…101110110…	…0110011…	01111110
Flag	Control	Address	Data	Checksum	Flag

Figure 4.5: Block diagram of the ports of a generic synchronous connection

The starting and ending sequences of the synchronous connection are called flags are each typically 8 bits in length. Following the start

sequence flag is usually one or more bytes of control information which provides information about the enclosed data or provides status information pertaining to the sender or receiver or both. The address field indicates the destination of the frame. Following the data is almost always some form of error-checking sequence, such as cyclic check sum which is an error-checking technique than parity checking and is used in many modern implementations of computer networks. Detail of error-checking will be discussed in later Unit of this course.

Three ways are used to maintain synchronization in synchronous connections:

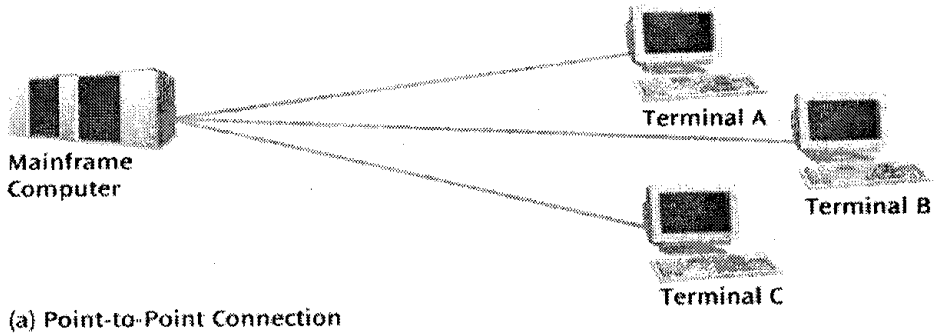
(i) send a synchronizing clock over a separate line that runs parallel to the data stream. As the data arrives on one line, a clock signal arrives on the second line. The receiver can use this clock signal to stay synchronized with the incoming data.
(ii) If transmitting a digital signal, use a Manchester code discussed in Unit 2 which is an example of a self-clocking signal.
(iii) If transmitting an analog signal, use the properties of the analog signal itself for self-clocking.

- **Isochronous connections**

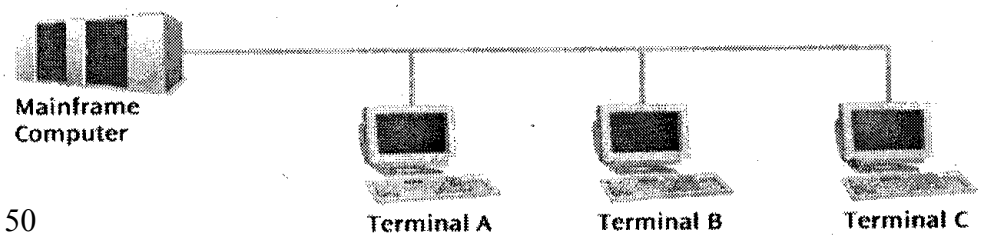
This is a special kind of data link connection used to support various types of real-time applications. Examples of real-time applications usually include: streaming voice, and video.

3.3 Terminal-to-Mainframe connection

One type of connection that is based upon synchronous and asynchronous connections is the terminal-to-mainframe computer connection. Recall that the terminal-to-mainframe connection was introduced in Unit 1 of this course. A direct connection between a



(a) Point-to-Point Connection



(b) Multipoint Connection

Figure 4.6: Terminal-to-mainframe connection

It is to be noted that the mainframe is the primary and the terminals are the secondary. To allow a terminal to transmit data to a mainframe, the mainframe must poll the terminal. There two basic forms of polling: roll-call polling and hub polling. In roll-call polling, the mainframe polls each terminal in a round-robin fashion while in hub polling, the mainframe polls the first terminal, and this terminal passes the poll onto the next terminal. Figure 4.7 shows a typical diagram of polling.

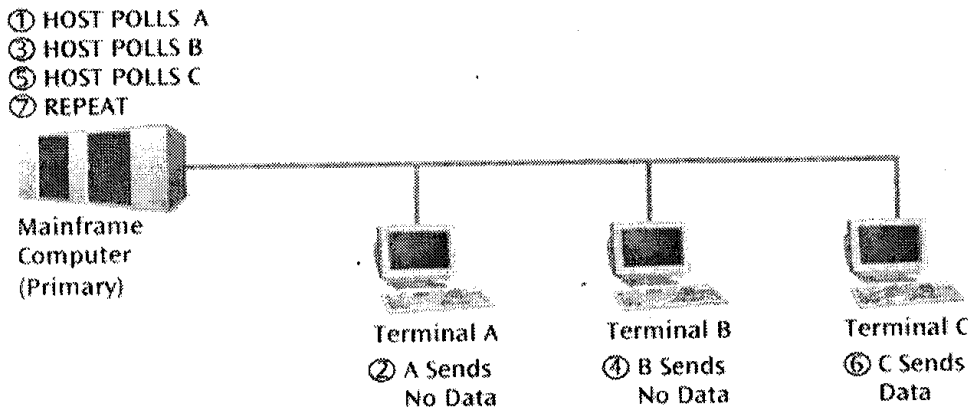


Figure 4.7: Concept of polling

SELF-ASSESSMENT EXERCISE

i. What are the four components of an interface?
ii. What are the features of asynchronous connections that help them maintain synchronization?
iii. Describe how a synchronous connection keep the sender and receiver synchronized.

ANSWER TO SELF-ASSESSMENT EXERCISE

1a. The **Electrical component**: deals with voltages, line capacitance, and other electrical characteristics.
The **mechanical component** deals with items such as the connector or plug description.
The **functional component** describes the function of each pin or circuit that is used in a particular interface
The **procedural component** describes how the particular circuits are used to

b. Asynchronous connections maintain synchronization by using small frames with a leading start bit.

c. Three ways are used to maintain synchronization in synchronous connections:

- send a synchronizing clock over a separate line that runs parallel to the data stream. As the data arrives on one line, a clock signal arrives on the second line. The receiver can use this clock signal to stay synchronized with the incoming data.
- If transmitting a digital signal, use a Manchester code which is an example of a self-clocking signal.
- If transmitting an analog signal, use the properties of the analog signal itself for self-clocking.

3.4 Transmission Modes

There are three mode of data transfer in computer network: a half duplex connection transmits data in both directions but in only one direction at a time, while a full duplex connection transmits data in both directions and at the same time. The third technique, simplex connection transmit data in only one direction. Based on your experience, identify some examples of each of the transmission modes above.

4.0 CONCLUSION

In this unit we described the basics of connecting a computer to other devices in terms of standards and interfacing. We moved on to identify three popular data link layer configurations: asynchronous connections, synchronous connections, and isochronous connections. Finally, we examined the connection between a terminal and mainframe computer.

5.0 SUMMARY

The connection between a computer and a peripheral is often called the interface, and the process of providing all the proper interconnections between a computer and a peripheral is called interfacing. Over the years, a number of interface standards have been developed. Two that are worthy of are EIA-232F and the USB. Transmission systems that are half-duplex can transmit data in both directions, but in only one direction at a time. Full-duplex systems can transmit data in both directions. Asynchronous connection use single-character frames and start and stop bits to establish the beginning and ending points of the frame. Synchronous connections use multiple-character frames, sometimes consisting of thousands of characters. Isochronous connections provide real-time connections between computers and peripherals. A connection between a computer terminal and a mainframe computer that is dedicated to one terminal is called a point-to-point connection, while a shared connection between more than one computer terminal and a mainframe computer is called a multipoint connection.

6.0 TUTOR-MARKED ASSIGNMENT

1. Terminals A, B, and C connected to a mainframe computer. Only terminal C has data to transmit. Show the sequence of messages sent between the mainframe and three terminals using roll-call polling.
2. What types of devices are best served with an isochronous connection?

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network*, A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). *Data Communication Principles for Fixed and Wireless Networks*, Kluwer Academic Publisher, USA.

UNIT 5 MULTIPLEXING AND COMPRESSION

CONTENTS

1.0 INTRODUCTION

Under the simplest conditions, a medium can carry only one signal at any moment in time. For example, the twisted pair cable that connects a keyboard to a microcomputer carries a single digital signal. Likewise category 5e twisted pair wire (discussed in Unit 3) that connects a microcomputer to a LAN carries only one digital signal at a time. The technique of transmitting multiple signals over a single medium is called Multiplexing. Another way to make a connection between two devices more efficient is to compress the data that transfer over the connection. This unit address these two techniques.

2.0 OBJECTIVES

At the end of this Unit, you should be able to:

- Describe the different multiplexing techniques, list their applications, advantages and disadvantages
- Describe the different between lossy and lossless compression.
- Describe the basic operation of run-length compression.

3.0 MAIN CONTENT

3.1 Frequency Division Multiplexing(FDM)

FDM is the assignment of non-overlapping frequency ranges to each "user". or signal on a medium. Thus, all signals are transmitted at the same time, each using different frequencies. To allow multiple users share a single medium, FDM assigns each user a separate channel. A channel is an assigned set of frequencies that is used to transmit the user's signal. In FDM, this signal is analog. Cellular telephone systems

is a common example of FDM. In general, device that accepts input from one or more users is called multiplexor, while the device that is attached to the receiving end of the medium and splits off each signal to deliver it to the appropriate receiver is called demultiplexor. In all FDM systems the multiplexor accepts input from the user, convert the data streams to analog signals using either fixed or dynamically assigned frequencies, and transmits the combined analog signals over a medium that has a wide enough bandwidth to support the total range of all assigned frequencies. rhe demultiplexor then accepts combined analog signals, and deliver this to the appropriate user(s). Figure 5.1 shows a simplified diagram of FDM.

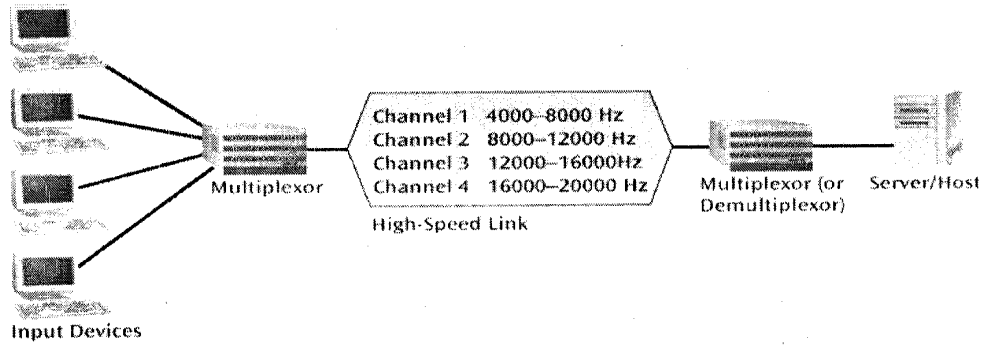


Figure 5.1: Example of FDM

3.2 Time Division Multiplexing(TDM)

Recall that in section 3.1, FDM takes the available bandwidth on a medium and divides the frequencies among multiple channels, or users. Essentially, this division enables multiple users to transmit at the same time. In contrast, TDM allows only one user at a time to transmit, and the sharing of the medium is accomplished by dividing available transmission time among users. In this technique, a user uses the entire bandwidth of the channel, but only for a brief moment. A DM calls on one input device after another, giving each device a turn at transmitting its data over a high-speed line. For example, suppose two users, A and B, wish to transmit data over a shared medium to a distant computer. We can create 'a rather simple TDM scheme by allowing user A to transmit
during the first second, then user B during the following second, followed again by user A during the third second, and so on. We have two types: Synchronous TDM and Statistical TDM.

Synchronous TDM

Sync. TDM gives each incoming source signal a turn to be transmitted, proceeding through the sources in round-robin fashion. Given n inputs, a synchronous TDM accepts one piece of data, such as byte, from the first device, transmits it over a high-speed link, accepts one byte from the second device, transmits it over the high-speed link and continues this process until a byte is accepted from the nth device. After the nth device's first byte is transmitted, the multiplexor returns to the first device and continues in round-robin fashion.(see figure 5.2)

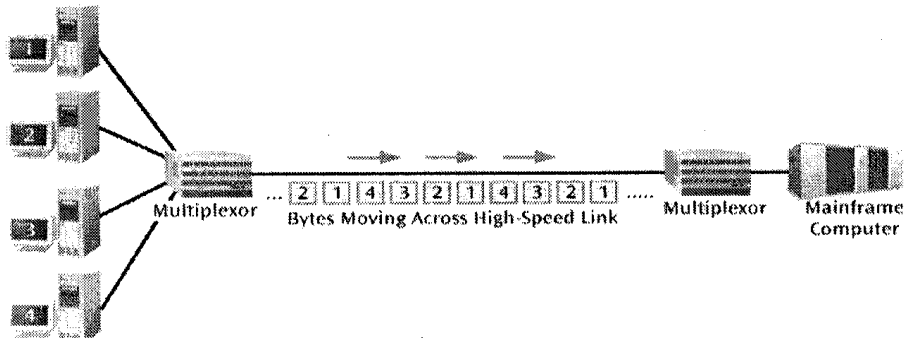


Figure 5.2: Sample Synchronous TDM

Please note that if one device generates data at a faster rate than other devices, then the multiplexor must either sample the incoming data stream from that device more often than it samples the other devices, or buffer the faster incoming stream. If a device has nothing to transmit, the multiplexor must still insert something into the multiplexed stream. Maintaining synchronization across a multiplexed link is important. To maintain synchronization between sending multiplexor and receiving demultiplexor, the data from the input sources is often packed into a simple frame, and synchronization bits are added somewhere within the frame. Three types of synchronous TDM that are popular today are T -1, ISDN, and SONET /SDH.

Statistical TDM

As you've seen in the preceding discussions, both frequency division multiplexing and synchronous time division multiplexing can waste unused transmission space. One solution to this problem is Statistical TDM (sometimes called asynchronous TDM). Statistical TDM transmits only from active users and does not transmit empty slots. To transmit data only from active users, the multiplexor creates a more complex frame that contains data only from those input sources that have something to send.

For example, consider the following simplified scenario. If four station, A,B,C and D, are connected to a statistical multiplexor, but only stations A and C are currently transmitting, the statistical multiplexor transmits only the data from A and C as shown in figure 5.3.

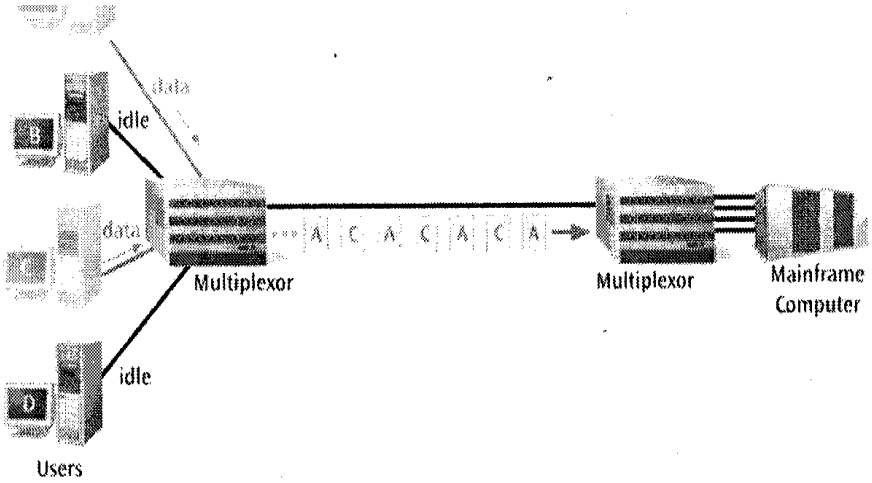


Figure 5.3: Statistical Multiplexing

Note that at any moment, the number of station transmitting can change from two to zero, one, three or four. If that happens, the statistical multiplexor must create a new frame containing data from the currently transmitting stations. Because only two out of four is transmitting, how does the demultiplexor on the receiving end recognize the correct recipients of the data? Some type of address must be included with each byte of data, to identify who sent the data and for whom it is intended.

Wavelength Division Multiplexing (WDM)

WDM multiplexes multiple data streams onto a single fiber-optic line. It is in essence, a frequency division multiplexing technique that assign input sources to separate sets of frequencies. WDM uses different wavelength(frequency) lasers to transmit multiple signals at the same time over a single medium. The wavelength of each differently colored laser is called the lambda. Thus WDM supports multiple lambdas. The technique assigns a uniquely colored laser to each input source and combines the multiple optical signals of the input sources so that they can be amplified as a group and transported over a single fiber. It is interesting to note that each to note that because of the properties of the signals and glass fiber, plus the nature of light itself, each signal carried on the fiber can be transmitted at a different rate from other signals. This means that a single fiber-optic line can support simultaneous transmission speeds such as 51.84 Mbps, 1555.52 Mbps 622.08 Mbps, and 2.488 Gps

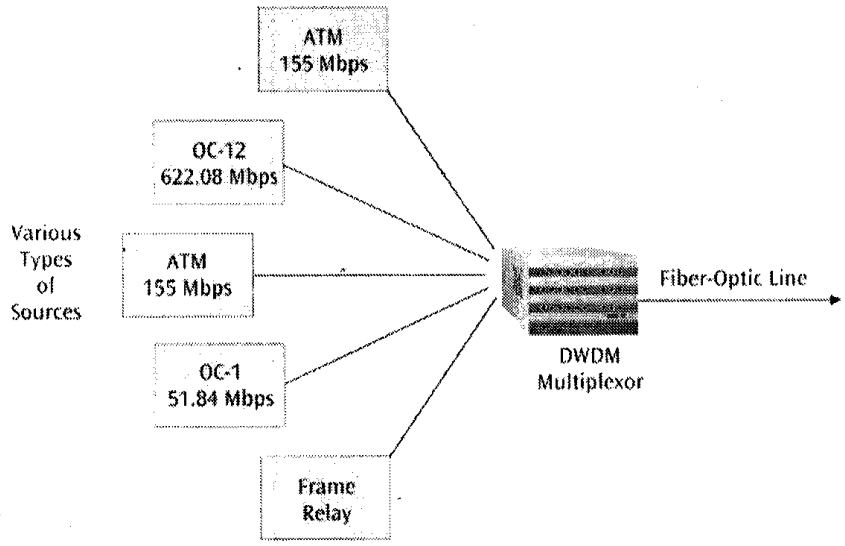


Figure 5.4: Wavelength Division Multiplexing

WDM is also scalable. As the demands on the systems and its application grow, it is possible to add additional wavelengths, or lambdas onto the fiber, thus further multiplying the overall capacity of the original fiber-optic system. While most systems support fewer than 100 lambdas, some of the ultra high-priced systems can handle more than 100 lambdas. When WDM can support a large number of lambdas, it is often called dense WDM. Dense WDM is an expensive way to transmit signals from multiple devices due to high number of differently colored lasers required in one unit. One less expensive variation on Dense WDM is coarse wavelength division multiplexing CWDM. It is less expensive technology because it is designed for short-distance connections and has only few lambdas, with a greater space between lambdas.

Discrete Multitone

This is a multiplexing technique commonly found in digital subscriber line(DSL) systems. DSL is a technology that allows a high-speed data signal to traverse a standard copper-based telephone line. It is to be noted that the highest transmission speed we can achieve with a standard dial-up telephone line is 56kbps. DSL however, is capable of achieving speeds into the millions of bits per second using DMT multiplexing technique. For example one form of DMT supports 256 subchannels, each of which is capable of a 60-kbps yielding 15.36 millions-bps system as shown in figure 5.5

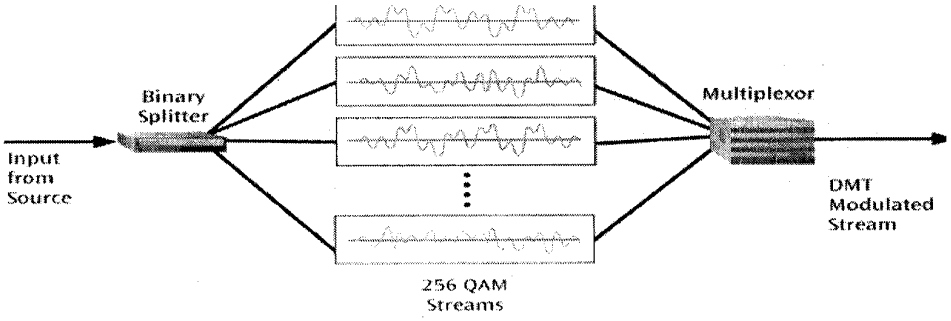


Figure 5.5: DMT technique

Code Division Multiplexing (CDM)

COM is a relatively new technology that has been used extensively by both the military and cellular telephone companies. Whereas other multiplexing techniques differentiate one user from another by either assigning frequency ranges or interleaving bit sequences in time, CDM allows multiple users to share a common set of frequencies by assigning a unique digital code to each user. CDM is based on upon class of Modulation techniques known as spread spectrum which is a technology used in the communications industry for modulating a signal into a new signal that is more secure and thus more resistant to wire-tapping. This technology falls into two categories- frequency hopping and direct sequence. CDM uses direct sequence spread spectrum technology, a technique that spreads the transmission of a signal over a wide range of frequencies, using mathematical values. As the original data is input into a direct sequence modulator, each binary 1 and 0 is replaced with a larger, unique bit sequence. Fro example each, each device in a cell phone market that uses CDM to transmit its signal is assigned its own bit sequence. When the bit sequences arrive at the destination station, the CDM is capable of telling one mobile device's bit sequence from another's. Let's create an example using three mobile users: A,B, and C. Suppose Mobile A has been the binary code 10111001, mobile user B ith code 01101110, and mobile user C with code 11001101 which are called chip spreading codes. In the real word, these codes are 64 bits in length. If a mobile user A wishes to send a binary 1, a mobile device transmits the unique code. If mobile user A wishes to send a binary 0, mobile devices transmits the inverse of the code. In this example assuming mobile user A transmits a binary 1, mobile user B transmits a binary 0, and mobile user C transmits a binary 1, the following is actually transmitted. The following is actually transmitted:

Mobile A sends a 1, or 10 111 001, or +-+++--+
Mobile B sends a 0, or 10010001, or +--+---+
Mobile C sends a 1, or 11001101, or ++--++-+

For simplicity, assume 8-bit code, signal code: 1- chip = +N volt; 0-chip = - N volt.

The receiver receives all three signals at the same time and adds the voltages as shown below:

+	-	+	+	+	-	-	+
+	-	-	+	-	-	-	+
+	+	-	-	+	+	-	+

Summed signal received by base station: +3, -1, -1, +1, +1, -1, -3, +3.
Then, to determine what each mobile user transmitted, the receiver multiplies the sum by the original code of each mobile users, expressed as + and -values, then take the sum of those products.

Base station decode for Mobile A:

- Signal received: +3, -1, -1, + 1, + 1, -1, -3, +3
- Mobile A's code: +1, -1,+1,+1,+1,-1, -1,+1
- Product result: +3,+1, -1,+1,+1,+1,+3,+3
- Sum of Products: + 12
- Because the sum of product is greater than or equal 8 in this 8-bit example the value transmitted must have been a binary 1.

Base station decode for Mobile B:

- Signal received: +3, -1, -1, +1, +1, -1, -3, +3
- Mobile A's code: -1, +1, +1, -1, +1, +1, +1,-1
- Product result: -3, -1, -1, -1, +1, -1, -3,-3
- Sum of Products: -12
- Because the sum of product is less than 8 in this 8-bit example the value
- transmitted must have been a binary O.

Using a 64 -bit code, it is theoretically possible to support 264 (18,446,744,073,709,551,616) cellular telephones in the same metropolitan area at the same time. Techniques such as this will allow data communications systems to grow in response to an ever increasing demand for communication services. Table 5.1 shows the advantages and disadvantages of multiplexing techniques.

Table 5.1: Advantages and disadvantages of Multiplexing techniques

Multiplexing Technique	Advantages	Disadvantages
Frequency Division Multiplexing	Simple Popular with radio, Cable TV All the receivers, such as cellular telephones, do not need to be at the same location	Noise problems due to analog signals Wastes bandwidth Limited by frequency ranges
Synchronous Time Division Multiplexing	Digital signals Relatively simple Commonly used with T-1, ISDN	Wastes bandwidth
Statistical Time Division Multiplexing	More efficient use of bandwidth Frame can contain control and error information Packets can be of varying size	More complex than synchronous time division multiplexing
Wavelength Division Multiplexing	Very high capacities over fiber Signals can have varying speeds Scalable	Cost Complexity
Discrete Multitone	Capable of high transmission speeds	Complexity, noise problems
Code Division Multiplexing	Large capacities Scalable	Complexity Primarily a wireless technology

SELF-ASSESSMENT EXERCISE

Mobile user A is using code division multiplexing and has been assigned a binary code of 10010101. Mobile user B, also using code division multiplexing, has been assigned a binary code of 00011100. Mobile user C, also using code division multiplexing, has been assigned 00110011. Mobile user A transmits a1, while mobile user B transmits a 0, and mobile user C transmits O. Show the sum products that results and your calculation.

3.3 Compression

This is another technique used to squeeze more data over a communications line. If you can compress a data file down to Y2 of its original size, the file will obviously transfer in less time. We have two basic groups of compression:

- Lossless - when data is uncompressed, original data returns (compress financial file).
- Lossy - when data is uncompressed, you do not have the original data(Compress a video image, movie, or audio file).

Examples of lossless compression include Huffman codes, run-length compression, and Lempel-Ziv compression, while examples of lossy compression include MPEG, JPEG, MP3.

3.3.1 Lossless Compression

One of the more common and simpler examples of lossless compression is run-length encoding. This technique replaces any repetitions of the same bit or byte that occur in a sequence of data with single occurrence of the bit/byte and a run count. Let's take a look at the example below:

0000000000000010000000011000000000000000000001000...00110000000

(30 0s)

Runs: 14 9 0 20 30 0 11

A compression technique based on run-length encoding would compress the 0s by first counting the "runs" of 0s- that is, it would start by counting the 0s until a binary 1 is encountered. If there are no Os between a pair of 1 s the pair would be considered a run that contains Os. Performing .this gave the above runs. The next step is to convert each of the decimal value (14, 9, etc) into 4-bit binary values, or nibbles. Note: If you need to code a value larger than 15, you need to use two code two consecutive 4-bit nibbles. The first is decimal 15, or binary 1111, and the second nibble is the remainder. For example, if the decimal value is 20, you would code 1111 0101 which is equivalent to 15 + 5.

If you want to code the value 15, you still need two nibbles: 1111 0000. The rule is that if you ever have a nibble of 1111, you must follow it with another nibble. Thus converting the above runs of 14, 9, 0, 20, 30, 0, and 11 would produce the following nibbles: 1110100100001111010111111110000 1011

In this example, note that the original bit string, which consisted of 91 bits, is compressed to 36 bits- a reduction of 60% and that no data is lost (hence the name lossless). One disadvantage of this technique is that it is worthwhile only if the original data consists predominantly of binary 0s.

3.3.2 Lossy Compression

The compression technique described above is an example of lossless compression. Lossless compression is necessary when the nature of the data is such that it is important that no data be lost during the compression and decompression stages. Program, text, and image files, video, images and high-quality audio files can also be compressed using, lossless compression but the percentage of reduction is usually not as significant. This is due to the nature of data of the data in video and audio files-'there is not, one' symbol or set of symbols that occur frequently enough to produce a reasonable level of compression. For example, if you take some music and digitize it, you will produce a long stream of binary 1s and 0s. To compress this stream, you can chose to perform a lossless run-length encoding on either the -1 s or the 0s. Unfortunately, however, because this type of data is dynamic, there will probably not enough repeating runs of either bit to 'produce a reasonable compression. Thus you need some other compression technique.

MP3, which is an abbreviation for MPEG (Moving Picture Experts Group) Audio layer 3, is a common form of audio compression. JPEG, which stands for Joint Photographic Experts Group, is a technique that is very commonly used to compress video images. The process of converting an image to JPEG format involves three phases: discrete cosine transformation, quantization, and run length encoding.

SELF-ASSESSMENT EXERCISE

Given the following bit string, show the run-length encoding that would result:

0000010000000001100000000000000010000110000000000000000000
01000000

ANSWER TO SELF-ASSESSMENT EXERCISE

1. Mobile A sends a1, or 10010101, or +--+-+-+
   Mobile B sends a0, or 00011100, or ---+++--
   Mobile C sends a1, or 00110011, or --++--++

For simplicity, assume 8-bit code, signal code: 1-chip = + N volt; 0-chip = - N volt.

The receiver receives all signals at the same time and adds the voltages as shown below:

```
+    -    -    +    -    +    -    +
-    -    -    +    +    +    -    -
-    -    +    +    -    -    +    +
```

Summed signal received by base station: -1, -3, -1, +3, -1, +1, -1, + 1

Then, to determine what each mobile user transmitted, the receiver multiplies the sum by the original code of each mobile users, expressed as + and -values, then take the sum of those products.

Base station decode for Mobile A:

- Signal received: -1, -3, -1, +3, -1, +1, -1, +1
- Mobile A's code: +1, -1, -1, +1, -1, +1, -1, +1
- Product result: -1, +3, +1 +3 +1 +1 +1 +1
- Sum of Products: + 10
- Because the sum of product is greater than or equal 8 in this 8-bit example the value transmitted must have been a binary 1.

Base station decode for Mobile B:

Signal received: -1, -3, -1, +3, -1, +1, -1,+1
Mobile B' s code: -1, -1, -1, + 1, + 1, + 1, -1, -1
Product result: +1,+3,+2, +3, - 1 +1 +1 -1
Sum of Products: 9

Using the concept for Mobile user Sum product is -8.

2. The run-length for:

```
00000100000000011000000000000000100001100000000000000000000001000000
  5         9      0         15            4  0         20                  6
```

4.0 CONCLUSION

In this unit, we looked at three basic techniques for dividing a medium into multiple channels which are: a division of frequencies, a division of time, a division of transmission codes. We finally discussed two basic forms of compression: lossy and lossless used to compacts data into a small package.

5.0 SUMMARY

For multiple signals to share a single medium, the medium must be divided into multiple channels. The three basic techniques for dividing a medium into multiple channels are: a division of frequency, a division of time, a division of transmission codes. Frequency division multiplexing involves assigning" nonoverlapping frequency ranges to different signals. Frequency division multiplexing uses digital signals. Time multiplexing of a medium involv~s dividing the available transmission time on a medium among users.

Time division multiplexing has two basic forms: synchronous time division multiplexing and statistical time division multiplexing. Synchronous time division multiplexing accepts input from fixed number of devices and transmits their data in an unending repetitious pattern. T -1, ISDN, and SONET/SDH telephone systems are common examples of systems that use synchronous tie division multiplexing. Statistical time division multiplexing accepts input from a set of devices that have data to transmit, creates a frame with data and control information, and transmits that frame.

Wavelength division multiplexing involves fiber-optic systems and transfer of multiple streams of data over a single fiber using multiple, colored laser transmitters. Discrete multi tone is a technology used in DSL systems. Code division multiplexing allows multiple users to share the same set of frequencies by assigning a unique digital code to each user.

Compression is the process that compacts data into a small package. When stored, compressed, data saves space; when transmitted, it results in shortest transmission times. Two basic forms of compression exist: lossless, in which no data is lost during the compression and decompression stages; and lossy, in which some of the original data is lost. Two popular form of loss less compression include run-length encoding and the Lempel-Ziv compression technique. Lossy compression is the basis of a number of compression techniques, including MP3 for audio, JPEG for still images, and MPEG for moving video.

6.0 TUTOR-MARKED ASSIGNMENT

XYZ Corporation has two buildings separated by a distance of 300 meters. A 3-inch diameter tunnel extends underground between the two buildings. Building A has a mainframe computer and Building B has 66 terminals. List some efficient techniques to link the two buildings.

7.0 REFERENCES/FURTHER READINGS

Curt, W (2007). *Data Communications and Computer Network,*: A Business User's Approach Fourth Edition. Bob Woo bury, Canada.

Aftab, A (2003). *Data Communication Principles for Fixed and WirelessNetworks,* Kluwer Academic Publisher, USA.

MODULE 2

UNIT 1 ERROR DETECTION

CONTENTS

1.0 INTRODUCTION

The process of transmitting data over a medium often works according to Murphy's Law: If something can go wrong, it probably will. Even if all possible error reducing measures are applied before and during data transmission, something will invariably alter the form of the original data. If this alteration is serious enough, the original data becomes corrupt, and the receiver does not receive the data that was originally transmitted. Even with the highest-quality fiber-optic cable, noise eventually creeps in and begins to disrupt data transmission. Thus, despite one's best efforts to control noise, some noise is inevitable.

When the ratio of noise power to signal power becomes such that the noise overwhelms the signal, errors occur. It is at this point that

error-detection techniques become valuable tools. Given that noise is inevitable and errors happen, something needs to be done to detect error conditions. This unit examines some of the more common error detection methods and compares them in terms of efficiency and efficacy.

2.0 OBJECTIVES

At the end of this Unit, you should be able to:

- Identify the different types of noise commonly found in computer networks.
- Specify the different error-prevention techniques, and be able to apply an error-prevention technique to a noise.
- Compare the different error-detection techniques in terms of efficiency and efficacy.
- Performs simple parity and longitudinal parity calculations, and enumerate their strengths and weaknesses.
- Cite the advantages of cyclic redundancy checksums, and specify what types of errors cyclic redundancy checksums will detect.
- Differentiate between the basic forms of error control, and describe the circumstances under which each may be used.

3.0 AIN CONTENT

3.1 Noise and Errors

As you might expect, a number of errors can occur during data transmission. From a simple blip to a massive outage, transmitted data-both analog and digital-is susceptible to many types of noise and errors. Copper-based media have traditionally been plagued by many types of interference and noise. Satellite, microwave, and radio networks are also prone to interference and crosstalk. Even the near-perfect fiber-optic cables can introduce errors into a transmission system, although the probability of this happening is less than with the other types of media. Let's examine several of the major types of noise that occur in transmission systems.

3.1.1 White Noise

White noise, which is also called thermal noise or Gaussian noise, is a relatively continuous type of noise and is much like the static you hear when a radio is being tuned between two stations. It is always present to some degree in transmission media and electronic devices and is dependent on the temperature of the medium. As the temperature increases, the level of noise increases due to the increased activity of the

electrons in the medium. Because white noise is relatively continuous, it can be reduced significantly but never completely eliminated. White noise is the type of interference that makes an analog or digital signal become fuzzy (Figure 6.1). Removing white noise from a digital signal is relatively straightforward if the signal is passed through a signal regenerator before the noise completely overwhelms the original signal. Removing white noise from an analog signal is also possible and involves passing the noisy analog signal through an appropriate set of filters, which (one hopes) leaves nothing but the original signal.

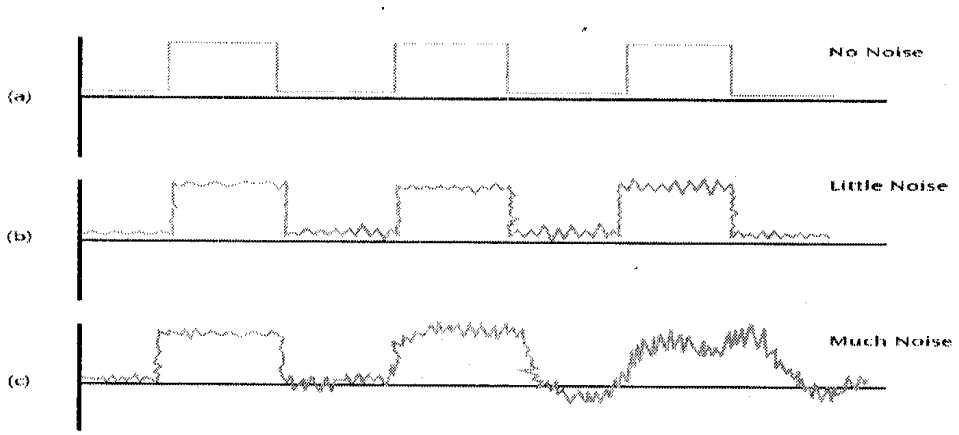


Figure 6.1: White noise as it interfere with Digital signal

3.1.2 Impulse Noise

Impulse noise, or noise spike, is a noncontinuous noise and one of the most difficult errors to detect, because it can occur randomly. The difficulty comes in separating the noise from the signal. Typically, the noise is an analog burst of energy. If the impulse spike interferes with an analog signal, removing it without affecting the original signal can be extremely difficult. For an example, consider what happens when you listen to AM radio during a thunderstorm. The lightning strikes in the area cause severe static on the radio- so severe that you cannot hear the normal radio transmissions.

3.1.3 Crosstalk

Crosstalk is an unwaI1ted coupling between two different signal paths. This unwanted coupling can be electrical, as might occur between two sets of twisted pair wire (as in a phone line), or it can be electromagnetic (as when unwanted signals are picked up by microwave antennas). Telephone signal crosstalk was a more common problem 20 to 30 years ago, before telephone companies used fiber-optic cables and other well-shielded wires. When crosstalk occurs during a phone conversation, you can hear another telephone conversation in the

background (Figure 6.2). High humidity and wet weather can cause an increase in electrical crosstalk over a telephone system. Even though crosstalk is relatively continuous, it can be reduced with proper precautions and hardware, as you will see shortly.

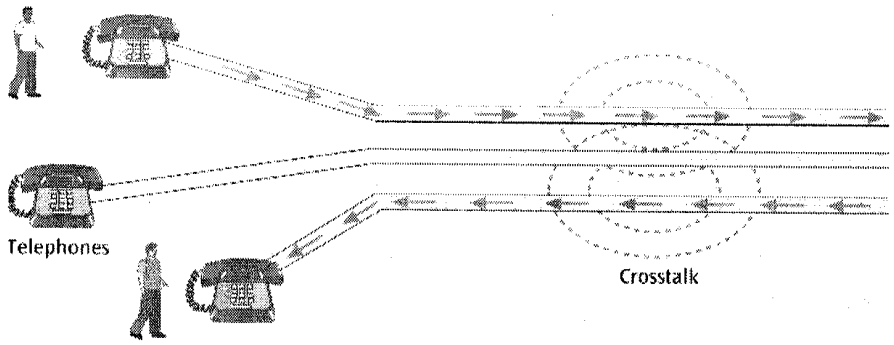


Figure 6.2: Three telephone circuit experiencing crosstalk

3.1.4 Echo

Echo is the reflective feedback of a transmitted signal as the signal moves through a medium. Much like the way a voice will echo in an empty room, a signal can hit the end of a cable, bounce back through the wire, and interfere with the original signal. This error occurs most often at junctions where wires are connected or at the open end of a coaxial cable. Figure 6.3 demonstrates a signal bouncing back from the end of a cable and creating an echo. To minimize the effect of echo, a device called an echo suppressor can be attached to a line. An echo suppressor is essentially a filter that allows the signal to pass in one direction only. For local area networks that use coaxial cable, a small filter is usually placed on the open end of each wire to absorb any incoming signals.

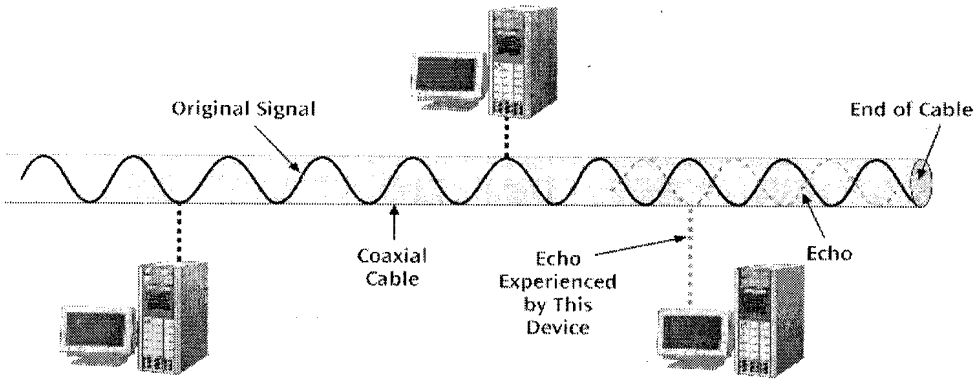


Figure 6.3: An example of echo noise

3.1.5 Jitter

Jitter is the result of small timing irregularities that become magnified during the transmission of digital signals as the signals are passed from one device to another. To put it another way, when a digital signal is being transmitted, the rises and falls of the signal can start to shift, or become blurry, and thus produce jitter. If unchecked, jitter can cause video devices to flicker, audio transmissions to click and break up, and transmitted computer data to arrive with errors. If jitter becomes too great, correcting it can require the transmitting devices to slow down their transmission rates, which in turn limits overall system performance. Figure 6-4 shows a simplified example of a digital signal experiencing jitter.

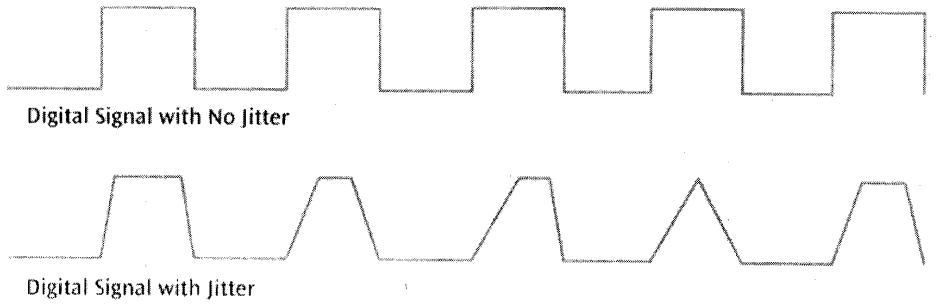


Figure 6.4: Original digital signal & signal with Jitter

Causes of jitter can include electromagnetic interference, crosstalk, passing the signal through too many repeaters, and the use of lower-quality equipment. Possible solutions to the jitter problem involve installing proper shielding, which can reduce or eliminate electromagnetic interference and crosstalk, and limiting the number of times a signal is repeated.

3.1.6 Attenuation

Attenuation is the continuous loss of a signal's strength as it travels though a medium. It is not necessarily a form of error, but can indirectly lead to an increase in errors affecting the transmitted signal.

3.2 Error Prevention

Because there are so many forms of noise and errors, and because the presence of one form of noise or another in a system is virtually a given, every data transmission system must take precautions to reduce noise and the possibility of errors. An unfortunate side effect of noise during a transmission is that the transmitting station has to slow down its transmission rate. For this reason, when a modem first makes a connection with another modem, the two modems participate in fallback

negotiation. This means that if the transmitting modem sends data and the data arrives garbled, the receiving modem may ask the transmitting modem to fall back to a slower transmission speed. This slowdown creates a signal in which the bit duration of each 0 and 1 is longer, thus giving the receiver a better chance of distinguishing one value from the next, even in the presence of noise. If you can reduce the possibility of noise before it happens, however, the transmitting station may not have to slow down its transmission stream. You can prevent the occurrence of many types of transmission errors by applying proper error-prevention techniques, including those listed below:

- Install wiring with the proper shielding to reduce electromagnetic interference and crosstalk.
- Use telephone line conditioning or equalization (provided by the telephone company), in which filters are used to help reduce signal irregularities. For an additional charge, the telephone company will provide various levels of conditioning to leased lines. This conditioning provides a quieter line, which minimizes data transmission errors.
- Replace older equipment with more modern, digital equipment; although initially expensive, this technique is often the most cost-effective way to minimize transmission errors in the long run.
- Use the proper number of digital repeaters and analog amplifiers to increase signal strength, thus decreasing the probability of errors.

Table 6.1 list the different types of error that can arise and includes one or more possible error-prevention techniques for each.

3.3 Error Detection

Despite one's best attempts at prevention, errors still occur. Since most data transferred over a communication line is important, it is usually necessary to apply an error-detection technique to the received data to ensure that no errors were introduced into the data during transmission. If an error is detected, a typical response is to perform some type of request for transmission. Error detection can be performed in several places within a communications model. One of the most common places is the data link layer. When a device creates a frame of data at the data link layer, it inserts some type of error-detection code. When the frame arrives at the next device in the transmission sequence, the receiver extracts the error-detection code and applies it to the data frame. Then the data frame is reconstructed and sent to .the next device in the transmission sequence. Some protocols perform an error detection

routine at the final destination. As we saw in Unit one, TCP performs error detection at the end points of the connection.

Table 6.1: Summary of errors and error-prevention mechanism

Type of Error	Error-Prevention Technique
White noise	Install special filters for analog signals; implement digital signal regeneration for digital signals
Impulse noise	Install special filters for analog signals; implement digital signal processing for digital signals
Crosstalk	Install proper shielding on cables
Echo	Install proper termination of cables
Jitter	Use better-quality electronic circuitry, use fewer repeaters, slow the transmission speed
Attenuation*	Install device that amplifies analog signals; implement digital signal regeneration of digital signals.

* Not a type of error, but indirectly affects error.

Regardless of where the error detection is applied, all systems still recognize the importance of checking for transmission errors. The error-detection techniques themselves can be relatively simple or relatively elaborate. As you might expect, simple techniques do not provide the same degree of error checking as the more elaborate schemes. Let's examine some error-detection techniques and evaluate the strengths and weaknesses of each.

3.3.1 Parity Checks

The most basic error-detection techniques are parity checks, which are used with asynchronous connections. Although there are various forms of single-character parity checking, one fact remains constant: parity checks let too many errors slip through undetected. For this reason alone, parity checks are rarely, if ever, used in any kind of serious data transmissions. Despite this, two forms of parity checks- simple parity and longitudinal parity-do still exist and are worth examining. Simple Parity (occasionally known as vertical redundancy check) is the easiest error-detection method to incorporate into a transmission system; it comes in two basic forms: even parity and odd parity.

The basic concept of parity checking is that a bit is added to a string of bits to create either even parity or odd parity. With even parity, the 0 or 1 added to the string produces an even number of binary 1s. With odd parity, the 0 or 1 added to the string produces an odd number of binary 1s. If the 7 -bit ASCII character set is used, a parity bit is added as the

eighth bit. Suppose, for example, that the character "k"-which is 1101011 in binary-is transmitted and even parity is being applied. In this case, a parity bit of 1 would be added to the end of the bit stream, as follows: 11010111. There is now an even number (six) of 1s. (If odd parity were used, a 0 would be added at the end, resulting in 11010110.). Now, if a transmission error causes one of the bits to be flipped (the value is erroneously interpreted as a 0 instead of a1, or vice versa), the error can be e detected if the receiver understands that it needs to check for even parity. Returning to the example of the character "k" sent with even parity, if you send 11010111 but 01010111 is received, the receiver will count the 1s, see that there is an odd number, and know there is an error.

What happens if 11010111 with even parity is sent and two bits are corrupted? For example, 00010111 is received. Will an error be detected? The answer is no, an error will not be detected, because the number of 1 s is still even. Simple parity can detect only an odd number of erroneous bits per character. Is it possible for more than one bit in a character to be altered as a result of transmission error? Yes, isolated single-bit errors occur 50 to 60 percent of the time. Error bursts, in which two erroneous bits are separated by fewer than 10 uncorrupted bits, occur 10 to 20 per- cent of the time. Note that when the 7 -bit ASCII character set is used, a. parity bit is added for every 7 bits of data, resulting in a 1:7 ratio of parity bits to data bits. Thus, simple parity produces relatively high ratios of check bits to data bits, while achieving only mediocre 50% error-detection results.

3.3.2 Longitudinal Parity

Longitudinal parity, sometimes called longitudinal redundancy check or horizontal parity, tries to solve the main weakness of simple parity-that all even numbers of errors are not detected. To provide this extra level of protection, longitudinal parity needs to use additional parity check bits, as you will see shortly. The first step of this parity scheme involves grouping individual characters together in a block, as shown in Table 6-2. Each character (also called a row) in the block has its own parity bit. In addition, after a certain number of characters are sent, a row of parity bits, or a block character check, is also sent.

Each parity bit in this last row is a parity check for all the bits in the column above it. If one bit is altered in Row 1, the parity bit at the end of Row 1 signals an error. In addition, the parity bit for the corresponding column also signals an error. If two bits in Row 1 are flipped, the Row 1 parity check will not signal an error, but two column parity, checks will signal errors. This is how longitudinal parity is able to detect more errors than simple parity. Note, however, that if two bits

are flipped in Row 1 and two bits are flipped in Row 2, and the errors occur in the same column, no errors will be detected. This scenario, which is shown in Table 6-3, is a limitation of longitudinal parity.

Table 6.2: Sample example of longitudinal parity

			Data					Parity
Row l	1	1	0	1	0	1	1	1
Row 2	1	1	1	1	1	1	1	1
Row 3	0	1	0	1	0	1	0	1
Row 4	0	0	1	1	0	0	1	1
Parity Row	0	1	0	0	1	1	1	0

Table 6.3: Errors and longitudinal parity

			Data					Parity
Row 1	1	40	0 1	1	0	1	1	1
Row 2	1	40	40	1	1	1	1	1
Row 3	0	1	0	1	0	1	0	1
Row 4	0	0	1	1	0	0	1	1
Parity Row	0	1	0	0	1	1	1	0

Although longitudinal parity provides an extra level of protection by using a double parity check, this method, like simple parity; also introduces a high number of check bits relative to data bits, with only slightly better than 50% error-detection results. If n characters in a block are transmitted, the ratio of check bits to data bits is n+8:7n. In other words, to transmit a 20-character block of data, for example, a simple parity bit needs to be added to each of the 20 characters, plus a full 8-bit block character check is added at the end, producing a ratio of check bits to data bits that is 28:140, or 1:5.

3.3.3 Cyclic Redundancy Checksum

Unlike the simple parity and longitudinal parity techniques of error detection, which produce high ratios of check bits to data bits with only 50% error-detection results, the cyclic redundancy checksum (CRC), or cyclic checksum, method typically adds 8 to 32 check bits to potentially large data packets and yields an error detection capability approaching 100 percent. The CRC error-detection method treats the packet of data to be transmitted (the message) as a large polynomial. The rightmost bit of the data becomes the X^0 term, the next data bit to the left is the X I term, and so on. When a bit in the message is 1, the corresponding polynomial term is included. Thus, the data 101001101 would be equivalent to the polynomial:

X^8		$+ x^6$			$+ X^3$	$+ x^2 +$		X^0
1	0	1	0	0	1	1	0	1

(Because any value raised to the 0th power is 1, the X^0 term is always written as a 1.) The transmitter takes this message polynomial and, using polynomial arithmetic, divides it by a given generating polynomial, and produces a quotient and a remainder. The quotient is discarded, but the remainder (in bit form) is appended to the end of the original message polynomial., and this combined unit is transmitted over the medium. When the data plus remainder arrive at the destination, the same generating polynomial is used to detect an error. A generating polynomial is an industry-approved bit string that is used to create the cyclic checksum remainder. Some common generating polynomials that are in widespread use include:

- CRC-12: $X^{l2} + X^{11} + x^3 + x^2 + x + 1$
- CRC-16: $X^{l6} + X^{15} + x^2 + 1$
- CRC-CCITT: $X^{l6} + X^{l5} + x^5 + 1$
- CRC-32: $X^{32} + X^{26} + X^{23} + X^{22} + X^{16} + X^{12} + x^{11} + X^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$
- Asynchronous Transfer Mode CRC: $x^8 + x^2 + x + 1$

The receiver divides the incoming data (the original message polynomial plus the remainder) by the exact same generating polynomial that was used by the transmitter. If no errors were introduced during data transmission, the division should produce a remainder of zero. If an error was introduced during transmission, the arriving original message polynomial plus the remainder will not divide evenly by the generating polynomial and will produce a nonzero remainder, signaling an error condition. In real life, the transmitter and receiver do not perform polynomial division with software. Instead, hardware designed into an integrated circuit can perform the calculation much mote quickly.

The CRC method is almost foolproof. Table 6,.4 summarizes the performance of the CRC technique. In cases where the size of the error burst is less than r + 1, where

Table 6.4: Error-detection performance of cyclic redundancy checksum

Type of Error	Error Detection Performance
Single-bit errors	100 percent
Double-bit errors	100 percent, as long as the generating polynomial has atleast three 1s (they all do)
Odd number of bits in error	100 percent, as long as the generating polynomial contains a factor x + 1 (they all do)
An error burst of length < r + 1	100 percent
An error burst of length = r + 1	probability = $1 - (1/2)^{(r-1)}$
An error burst of length > r + 1	probability = $1 - (1/2)^{r}$

r is the degree of the generating polynomial, error detection is 100 percent. For example, suppose the CRC-CCITT is used, and so the degree, or highest power, of the polynomial is 16. In this case, if the error burst is less than r + 1 or 17 bits in length, CRC will detect it. Only in cases where the error burst is greater than or equal to r + 1 bits in length is there a chance that CRC may not detect the error. The chance or probability of an error burst of size r + 1 being detected is $1 -(1/2)^{(r-1)}$. Assuming again that r = 16, $1- (1/2)^{(16-1)}$ equals 1 -0.0000305, which equals 0.999969. Thus, the probability of a large error being detected is very close to 1.0 (100 percent).

The cyclic redundancy checksum is one of the few cases in the field of computer science in which you get more than you paid for. In contrast to parity checking, cyclic redundancy checksum detects almost 100 percent of all errors. Recall that parity checking, depending on whether it is simple parity or longitudinal parity, can detect only between 50 percent and approximately 80 percent of errors. You can perform manual parity calculations quite quickly, but the hardware methods of cyclic redundancy checksum are also fairly quick. As you saw earlier, parity schemes require a high number of check bits per data bits. In contrast, cyclic redundancy checksum requires that a remainder-sized number of check bits (either 8, 16, or 32- as you can see from the list of generating polynomials) be added to a message. The message itself can be hundreds to thousands of bits in length. Therefore, the number of check bits per data bits in cyclic redundancy can be relatively low. Cyclic redundancy checksum is a very powerful error-detection technique and should be seriously considered for all data transmission systems. Indeed, all local area networks use CRC techniques (CRC-32 is found in Ethernet LANs), the Inter- net uses a 16-bit CRC, and most other wide area

network protocols incorporate a cyclic checksum. Now that we understand the basic error-detection techniques, let's look at what happens once an error is detected.

SELF-ASSESSMENT EXERCISE

i. Describe white noise and how does it affect a signal.
ii. Briefly enumerate the error-prevention techniques to prevent occurrence of transmission error.
iii. Assuming ODD parity concept, determine the parity bit for the following group of characters: 0001111, 1110011, and 1011111
iv. What is the main difference between longitudinal parity and simple parity checks?

ANSWER TO SELF-ASSESSMENT EXERCISE

- White noise is also called thermal noise or Gaussian noise, and is a relatively continuous type of noise and is much like the static you hear when a radio is being tuned between two stations. It is always present to some degree in transmission media and electronic devices and is dependent on the temperature of the medium. As the temperature increases, the level of noise increases due to the increased activity of the electrons in the. medium. Because white noise is relatively continuous, it can be reduced significantly but never completely eliminated. White noise is the type of interference that makes an analog or digital signal become fuzzy.

Error prevention techniques

- Install wiring with the proper shielding to reduce electromagnetic interference and crosstalk.
- Use telephone line conditioning or equalization (provided by the telephone company), in which filters are used to help reduce signal irregularities. For an additional charge, the telephone company will provide various levels of conditioning to leased lines. This conditioning provides a quieter line, which minimizes data transmission errors.
- Replace older equipment with more modern, digital equipment; although initially expensive, this technique is often the most cost-effective way to minimize transmission errors in the long run.
- Use the proper number of digital repeaters and analog amplifiers to increase signal strength, thus decreasing the probability of errors.
- **Using ODD parity concept**
  0001111, parity bit is 1

1110011, parity bit is 0
1011111, parity bit is 1

4.0 CONCLUSION

In this unit, we looked at the different types of noise including white noise, impulse noise, crosstalk, echo, jitter, and attenuation which may be present in computer networks. We also identified other reasonable options for reducing the possibility of errors in computer networks. Attempt was made to described two forms error detection techniques: parity checks and cyclic redundancy checks.

5.0 SUMMARY

Noise is always present in computer networks, and if the noise level is too high, errors will be introduced during the transmission of data. The types of noise include white noise, impulse noise, crosstalk, echo, jitter, and attenuation. Only impulse noise is considered a noncontinuous noise, while the other forms of noise are continuous. Among the techniques for reducing noise are proper shielding of cables, telephone line conditioning or equalization, using modern digital equipment, using digital repeaters and analog amplifiers, and observing the stated capacities of media. Other reasonable options for reducing the possibility of errors include reducing the number of devices in a transmission stream, decreasing the length of cable runs, and reducing data transmission speed. Two basic forms of error detection are parity and cyclic redundancy checksum. Simple parity adds one additional bit to every character and is a very simple error-detection scheme that suffers from low error-detection rates and a relatively high ratio of check bits to data bits. Longitudinal parity adds an entire character of check bits to a block of data and improves error detection, but still suffers from inadequate error-detection rates and a relatively high ratio of check bits to data bits. Cyclic redundancy checksum is a superior error-detection scheme with almost 100 percent capability of recognizing corrupted data packets. Calculation of the checksum remainder is fairly quick when performed by hardware, and it adds relatively few check bits to potentially large data packets.

6.0 TUTOR-MARKED ASSIGNMENT

1. Which type of cable is most susceptible echo and why?
2. Given the character 1010010, what bit will be added to support odd parity?
3. Generate the parity bits and longitudinal parity bits for even parity for the characters: 0101010,0011010,0011110, and 0000110.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). Data Communications and Computer Network, A Business User's Approach Fourth Edition: Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks, Kluwer Academic Publisher, USA

UNIT 2 ERROR CONTROL

CONTENTS

1.0 INTRODUCTION

In Unit 6 we focused on error prevention and detection in when data is transferred from on computer to another in a network. Once an error in the received data transmission stream is detected, what is, the receiver going to do about it? The action that the receiver takes is called error control, which essentially involves taking one of three actions:

- Do nothing.
- Return a message to the transmitter asking it to resend the data packet that was ill error.
- Correct the error without retransmission.

In this Unit, we set out to examine each of these options in more detail.

2.0 OBJECTIVES

After reading through this unit, you should be able to:

- Differentiate between the basic forms of error control.
- Describes the circumstances under which each maybe used.

3.0 MAIN CONTENT

3.1 Error control options

As earlier mentioned under the introduction in this unit there are three possible error control options. The subsection describe these three options.

3.1.1 Do nothing

The first error-control option-doing nothing-'-hardly seems like an option at all. Yet doing nothing for error control is becoming a mode of operation for some newer wide area network transmission techniques. For example, frame relay, which has only been in existence since 1994 and is offered by telephone companies to transfer data over wide areas, supports the "do nothing" appr°a.ch to error control. If a data frame arrives at a frame relay switch and an error is detected after the cyclic check- sum is performed, the frame is simply discarded. The rationale behind this action is twofold. Frame relay networks are created primarily of fiber-optic cable. Because fiber-optic cable is the medium least prone to generating errors, it is assumed that the rate of errors is low and that error control is unnecessary. If a frame is in error and is discarded, frame relay assumes that either the transport layer or the higher-layer application that is using frame relay to transmit data will keep track of the frames and will notice that a frame has been discarded. It would then be the responsibility of the application to request that the dropped frame be retransmitted. Consider the example in which a company has a database application that sends database records across the country between two corporate locations. The database application (at the application layer) is using frame relay at the data link layer to transfer the actual records. If a record or part of a record is dropped by frame relay because of a transmission error, frame relay does not inform the application. Instead, the database application has to keep track of all records sent and received, and if one record does not arrive at the destination, the database application has to ask for a retransmission.

3.1.2 Return a Message

The second option-sending a message back to the transmitter-is probably the most common form of error control. Returning a message was also one of the first error-control techniques developed and is closely associated with a particular flow control technique. Recall from Unit One that flow control is a process that keeps a transmitter from sending too much data to a receiver, thus overflowing the receiver's buffer. Over the years, two basic versions of return-a-message error control have emerged: Stop-and-wait and sliding window. Let's look at the Stop-and- wait error control first.

Stop-and-wait error control is a technique usually associated with the Stop-and-wait flow control protocol. This protocol and its error-control technique are the oldest, simplest, and thus most restrictive. A workstation (Station A) transmits one packet of data to another workstation (Station B), then stops and waits for a reply from Station B.

Four things can happen at this point. First, if the packet of data arrives without the error, Station B responds with a positive acknowledgment, such as ACK. When Station A receives an ACK, it transmits the next data packet. Second, if the data arrives with an error, Station B responds with a negative acknowledgment, such as NAK or REJ (for reject). If Station A receives a NAK, it resends the previous data packet. Figure 7.1 shows an example of these transactions.

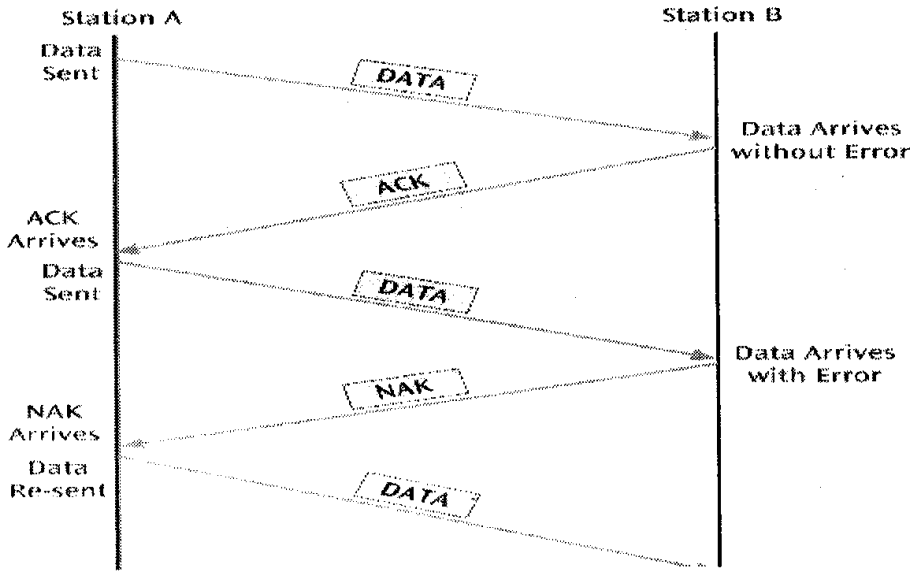


Figure 7.1: Some dialogue using stop-and-wait ARQ

Third, a packet arrives at Station B uncorrupted, Station B transmits an ACK, S but the ACK is lost or corrupted. Because Station A must wait for some form of acknowledgment, it will not be able to transmit any more packets. After a certain amount of time (called a timeout), Station A resends the last packet. But now if this packet arrives uncorrupted at Station B, Station B will not know that it is the same packet as the last one received. To avoid this confusion, the packets are numbered 0, 1,0, 1, and so on. If Station A sends packet 0, and the ACK for packet 0 is lost, Station A win resend packet 0. Station B will notice two packet 0s in a row (the original and a duplicate) and deduce that the ACK from the first packet 0 was lost. Fourth, Station A sends a packet, but the packet is lost. Since the packet did not arrive at Station B, Station 13 will not return an ACK. Because Station A does not receive an ACK, it will time out and resend the previous packet. For example, Station A sends packet 1, times out, and resends packet 1. If this packet 1 arrives at Station B, Station B responds with at 1 ACK. How does Station A know whether the ACK is acknowledging the first packet or the second? To avoid confusion, the ACKs are numbered, just like the packets. In contrast to packets (which are numbered 0, I, 0, I, and &0 on), however, ACKs are numbered 1, 0, 1, 0, and so on. If Station B receives packet 0, it

responds with ACK 2. The ACK 1 tells Station A that the next packet expected is packet 1. Because packet 0 just arrived, packet 1 is expected next. One of the most serious drawbacks to the simple Stop-and-wait error control is its high degree of inefficiency. Stop-and-wait error control is a half-duplex protocol, which means that only one station can transmit at one time. The time the transmitting station wastes waiting for an acknowledgment could be better spent transmitting additional packets. More efficient techniques than Stop-and-wait are available. One of these protocols is the sliding window technique. Sliding Window Error Control is based on the sliding window protocol, which is a flow control scheme that allows a station to transmit a number of data packets at one time before receiving some form of acknowledgment. Sliding window protocols have been around since the 1970s, a time when computer networks had two important limitations.

First, line speeds and processing power were much lower than they are today. For this reason, it was important that a station transmitting data did not send data too quickly and overwhelm the receiving station. Second, memory was more expensive, and so network devices had limited buffer space in which to store incoming and outgoing data packets. Because of these limitations, standard sliding window protocols set their maximum window size to seven packets. A station that had a maximum window size of 7 (as some of the early systems did) could transmit only seven data packets at one time before it had to stop and wait for an acknowledgment. Because a window size of 7 was small, extended sliding window protocols were soon created that could support 127 packets. Today, the TCP protocol used on the Internet can dynamically adjust its window size into the thousands for optimum performance. For simplicity, the following examples will consider the standard protocol with a maximum window size of 7. To follow the flow of data in a sliding window protocol with window size 7, packets are assigned numbers 0, 1, 2, 3, 4, 5, 6, and 7. Once packet number 7 is transmitted, the number sequencing starts back over at 0. Even though the packets are numbered 0 through 7, which corresponds to eight different packets, only seven data packets can be outstanding (unacknowledged) at one time. (The reason for this will become apparent shortly.). Because a maximum of seven data packets can be outstanding at one time, two data packets of the same number (for example, both numbered 4) can never be transmitted at the same time. If a sender has a maximum window size of 7 and transmits four packets, it can still transmit three more packets before it has to wait for an acknowledgment. If the receiver acknowledges all four packets before the sender transmits any more data, the sender's window size once again returns to 7.

Consider a scenario in which the sender transmit five packets and stops. The receiver receives the five packets and acknowledges them. Before the acknowledgment is received, the sender can send two more packets, because the window size is 7. On receipt of the acknowledgment, the sender can send an additional five packets before it has to stop again. The acknowledgment that a receiver transmits to the sender is also numbered. In the sliding window protocol, acknowledgments always contain a value equal to the number of the next expected packet. For example, if the sender, as shown in Figure 7.2, transmits three packets numbered 0, I, and 2, and the receiver wishes to acknowledge all of them, the receiver will return an acknowledgment (ACK) with value 3, l

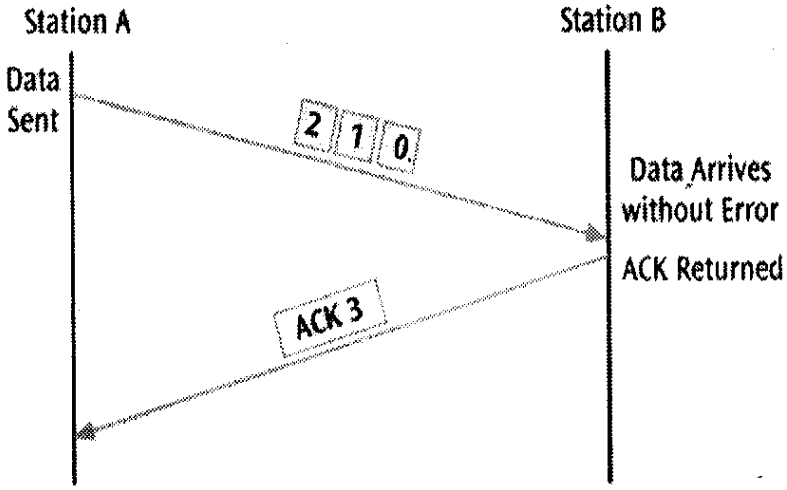


Figure 7.2: Example of sliding window

Let's go back now and consider what would happen if the protocol allowed eight packets to be sent at one time. Assume the sender sends packets numbered 0 through 7. The receiver receives all of them and acknowledges all by sending an acknowledgement numbered 0 (the next packet expected). But what if none of the packets arrived at the receiver? The receiver would not respond with a positive acknowledgment, and the sender would hear nothing. If after a short waiting period (a timeout), the sender asked the receiver for the number of the next packet expected, the receiver would answer with 0. The sender would never know if that meant all the packets were received or none of the packets was received. This potential confusion could lead a sender to resend all the packets when it's not necessary, or to resend none of the packets when the receiver is trying to indicate an error. Now let's add error control to the sliding window protocol.

Essentially four things can happen to a packet when it is transmitted: the packet arrives without error, the packet is lost (never arrives), the packet is corrupted (arrives but has a cyclic checksum error), the packet is delayed (if the packet is delayed long enough, a duplicate packet may be transmitted, resulting in two copies of the same packet). A sliding

window protocol with error control must be able to account for each of these four possibilities. As you read on, it might help to remember an important t distinction. A sliding window protocol's function is simply to inform the transmitter what piece of data is expected next. The function of a sliding window protocol with error control is to further specify what will occur if something goes wrong during a sliding window operation. While we examine the four possible things that can go wrong during transmission, it is also important to note a difference in the way different sliding window protocols number data. Older sliding window protocols, such as the High-level Data Link Control protocol, number the transmitted packets, each of which may contain hundreds of bytes of data. Thus, if a transmitter sends four packets, they might be numbered 0, 1, 2, and 3, respectively. If something goes wrong with a packet, the receiver will request that packet n be transmitted again. In contrast, newer sliding window protocols such as the TCP protocol of TCP/IP numbers the individual bytes. In this case, if a transmitter sends a packet with 400 bytes of data, the bytes may, for example, be numbered 8001 to 8400 if something goes wrong with the packet, the receiver will indicate that it needs bytes 8001 to 84000 to be transmitted again. Let's look at some examples that illustrate the four basic error control scenarios possible with sliding window protocols. For the first scenario, you'll see a packet-numbering example, but in general, we will concentrate more on examples of the byte-numbering scheme of the TCP protocol, as it is the more popular of the protocols. In the first scenario (shown in Figure 7.3), one or more packets, numbered individually, are transmitted, and all arrive without error. More specifically, Station: A transmits four packets numbered 2,3,4, and 5, and Station B receives them and! sends an ACK 6 acknowledging all four. Notice that Station B is also telling Station A what packet it expects next (packet 6). Station A responds by sending five more packets numbered 6, 7, 0, 1, and 20. Station B acknowledges all the packets by returning an ACK 3.

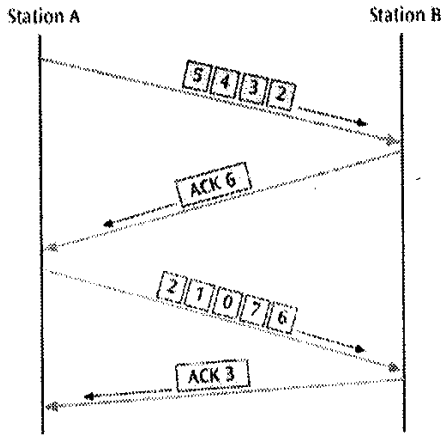


Figure 7.3: Normal transfer of data between two stations with numbering of the packets

If the sliding window protocol numbers bytes instead of packets, we might have an example such as that shown in Figure 7.4. Station A transmits one packet with bytes 0-400, followed by a second packet with bytes 401-800. Station B receives both packets and acknowledges all the bytes. Note once again that the ACK tells Station A the next byte that Station B expects (801).

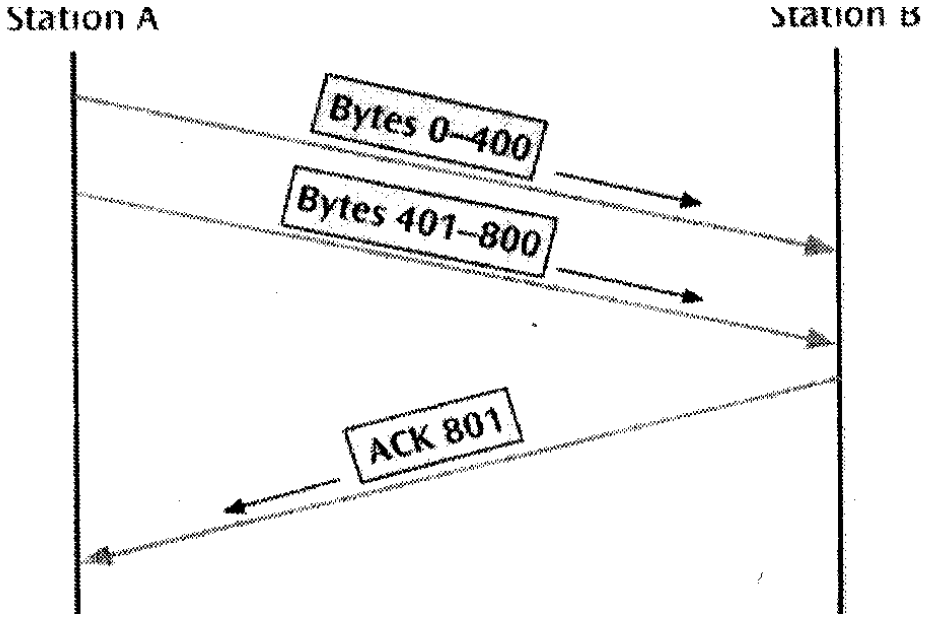


Figure 7.4: Normal transfer of data between two stations with numbering of the packets

An interesting question arises: Does a receiver have to acknowledge the data every time something is received? Or can a receiver wait a while to see if something or more is coming in before it sends an acknowledgment? In the world of TCP/IP, receiving stations follow a handful of rules for resolving this question. The first rule (shown in Figure 7.5) is that if a receiver just received some data and wishes to resend data back to the sender, then the receiver should include an ACK with the data it's about to send. This is called piggybacking, and it saves the receiver from sending a separate ACK message. The second rule is that if the receiver does not have any data to return to the sender, and the receiver has just acknowledged the receipt of a previously sent packet of data, then the receiver must wait 500ms to see if another packet arrives. If a second packet arrives before the 500 ms expire, however, then the receiver must immediately send an ACK.

Lastly, the third rule states that if the receiver is waiting for a second packet to arrive, and the 500 milliseconds expire, then the receiver doesn't wait for a second packet and instead issues an ACK immediately.

What happens when a packet is lost? Figure 7.6 illustrates the situation in which Station A transmits a sequence of packets and the second one is lost in the network. When the receiver, Station B, sees the third packet out of sequence, it returns an ACK with the sequence number of what it was expecting (byte 2401).Station A sees that something is wrong and retransmits the second packet. A similar result would occur if the second packet arrived, but with a CRC error. In both cases the packet is considered "lost." What happens if a packet is delayed, or a duplicate packet arrives at the destination? If the packet is delayed enough that it comes in out of order, the receiver also treats this as a lost packet and sends an ACK with the appropriate value. When the delayed packet or a duplicate packet arrives, the destination will see a packet with a sequence number less than the previously acknowledged bytes and simply discard the duplicate.

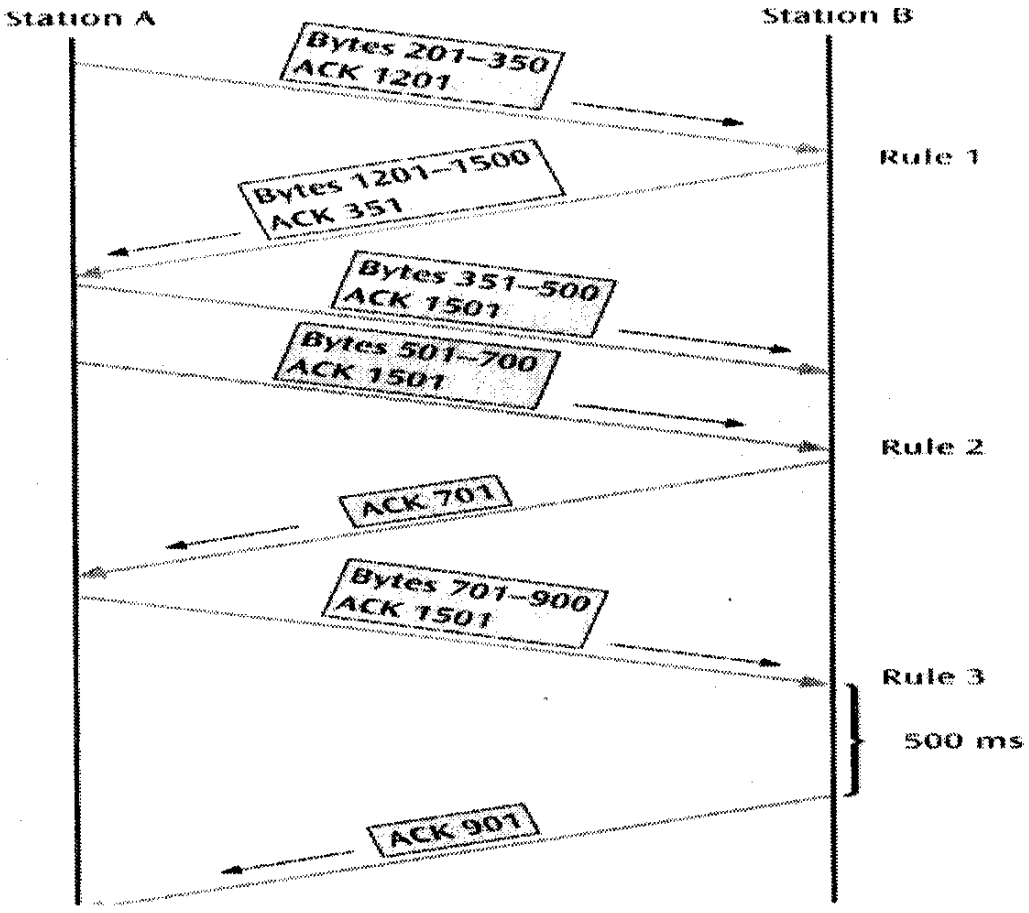


Figure 7.5: Three examples of returning an acknowledgement

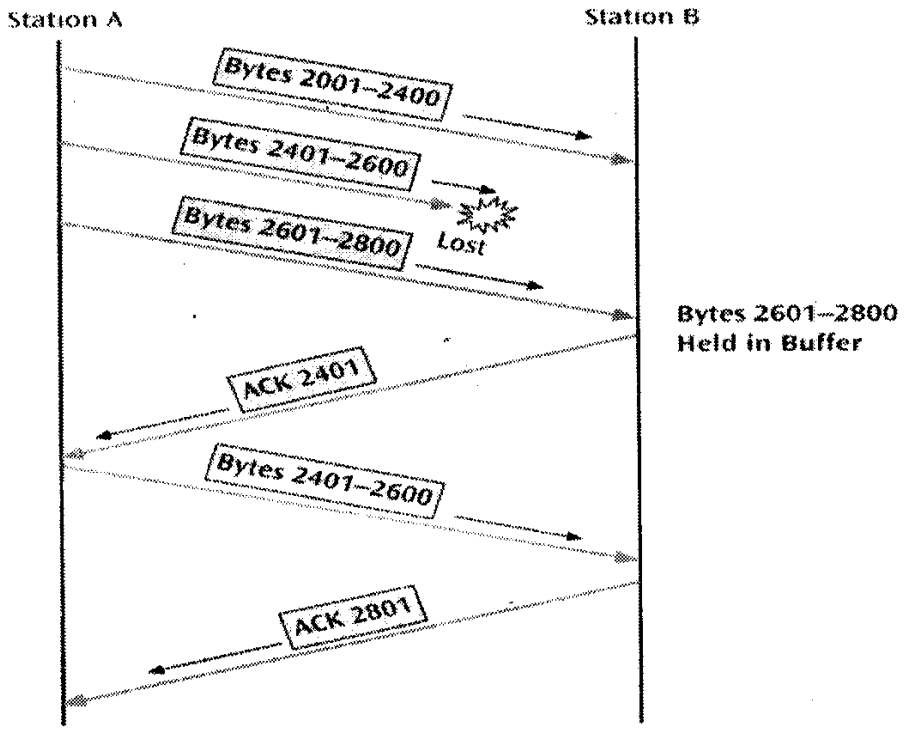


Figure: 7.6: A lost packet and station B's response

Finally, what happens if an acknowledgment is lost? Two possible scenarios exist. If a lost acknowledgment command is followed shortly by another acknowledgment command that does not get lost, no problem should occur, because the acknowledgments are cumulative (the second acknowledgment will have an equal or later packet number). If an acknowledgment command is lost and is not followed by any subsequent acknowledgment commands, the transmitting station will eventually time out and treat the previous packet as a lost packet and retransmit it (Figure 7.7).

3.2 Error Correction

The beginning of this section on error control listed three actions a receiver can take if an error packet is deemed corrupted: do nothing, return an error message, or correct the error. Correcting the error seems like a reasonable solution. The data packet has been received, and error-detection logic has determined that an error has occurred. Why not simply correct the error and continue processing? Unfortunately, correcting an error is not that simple. For a receiver to be able to fix an error-in a process called forward error correction-redundant information must be present so that the receiver knows which bit or bits are in error and what their original values were. For example, if you were given the data 0110110 and informed that a parity check had detected an error,

could you determine which bit or bits were corrupted? No, you would not have enough information. To see the full extent of the problem, consider what would happen if you transmitted three identical copies of each bit. For example, you transmitted 0110110 as 000 111 111 000 111 111 000. Now, let's corrupt one bit:000 111 111 001 111 111 000.

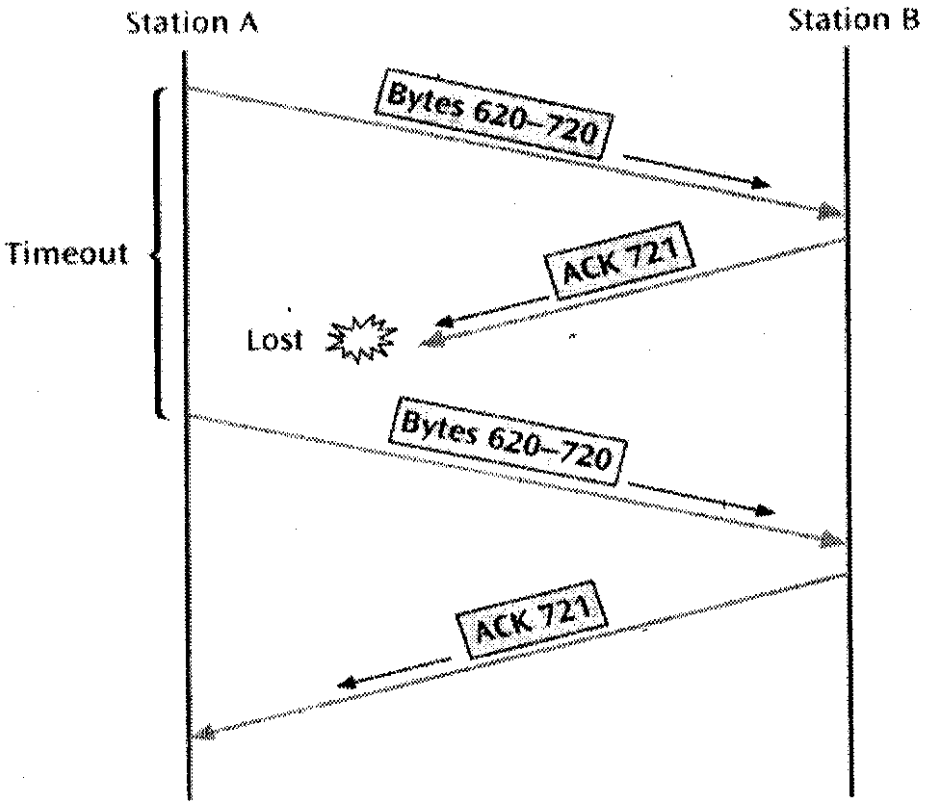


Figure 7.7: A lost acknowledgement and the retransmission of a packet

Can you determine which bit was corrupted? If you assume that only one bit has been corrupted, you can apply what is known as the majority rules principle and determine that the error bit is the final I in the fourth group, 001. Note, however, that even in this simple example, forward error correction entailed transmitting three times the original amount of data, and it provided only a small level of error correction. This level of overhead limits the application of forward error correction. A more useful type of forward error correction is a Hamming code. A Hamming code is a specially designed code in which special check bits have been added to data bits such that, if an error occurs during transmission, the receiver may be able to correct the error using the included check and data bits. For example, let's say we want to transmit an 8-bit character, for example, the character 01010101 seen in Figure 7.8. Let's number the bits of this character b12, b11, bl0, b9, b7, b6, b5, and b3. (We will number

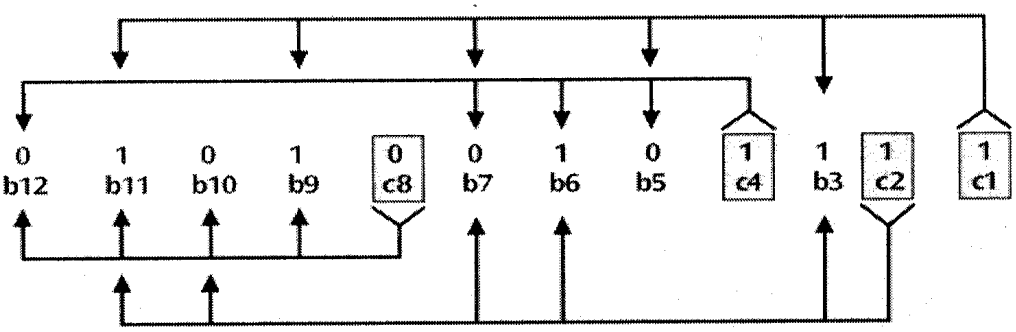


Figure 7.8: Example of Hamming code

the bits from right to left, leaving spaces for the soon-to-be-added check bits.) No add to these data bits the following check bits: c8, c4, c2, and cl, where c8 generates a simple even parity for bits b 12, b11, b 10, and b9. The check bit c4 will generate a simple even parity for bits b 12, b7, b6, and b5. Check bit c2 will generate a simple even parity for bits b11, b10, b7, b6, and b3. Finally, cl will generate a simple even parity for bits b11, b9, b7, b5, and b3. Note that each check bit here is checking different sequences of data bits.

Let's take a closer look at how each of the Hamming code check bits in Figure 6-11 works. Note that c8 "covers" bits b12, b11, bl0, and b9, which are 0101. If we generate an even parity bit based on those four bits, we would generate a 0 (there are an even number of 1s). Thus, c8 equals 0. c4 covers b12, b7, b6, and b5, which are 0010, so c4 equals 1. c2 covers b11, bl0, b7, b6, and b3, which are 10011, so c2 -equals 1. cl covers b11, b9, b7, b5, and b3, which are 11001, so cl equals 1. Consequently, if we have the data 01010101, we would generate the check bits 0111, as shown in Figure 6-11. This l2-bit character is now transmitted to the receiver. The receiver accepts the bits and performs the four parity checks on the check bits c8, c4, c2, and c 1. If nothing happened to the 12-bit character during transmission, all four parity checks should result in no error. But what would happen if one of the bits is corrupted and somehow ends up the opposite value? For example, what if bit b9 is corrupted? With the corrupted b9, we would now have the string 010000101111. The receiver would perform the four parity checks, but this time there would be parity errors. More precisely, because c8 checks b12, b11, bl0, b9, and c8 (01000), there would be a parity error as you can see, there are an odd number of in the data string, but the check bit is returning a 0 c4 checks b12, b7, b6, b5, and c4 (00101), and thus would produce no parity error, c2 checks b11, bl0, b7, b6, b3, and c2 (100111), and would produce no parity error. cl checks bits b11, b9, b7, b5, b3, and cl (100011), which would result in a parity error.

Notice that if we examine just the check bits and denote a 1 if there is a parity error and a 0 if there is no parity error, we would get 1001 (c8 error, c4 no error, c2 no error, cl error). 1001 is binary for 9, telling us that the bit in error is in the ninth position. Despite the additional costs of using forward error correction, are there applications that would benefit from the use of this technology? Two major groups of applications can in fact reap a benefit:

- applications that send sensitive data and
- applications that send data over very long distances.

Any organization that sends highly sensitive information such as financial data, health records, or government data does not want to transmit the same record multiple times. Despite the fact that anyone sending sensitive information is likely to use encryption software, data is vulnerable to interception every time it is sent. Thus, if a record is transmitted and arrives in error but can be corrected without retransmission, the likelihood of interception decreases. Likewise, if data has to be sent over a long distance, it is costly time-wise and money-wise to retransmit a packet that arrived in error. For example, the time required for a branch in Lagos to send a message to another branch in Abuja is several minutes. If the data arrives garbled, it will be another several minutes before the negative acknowledgment is received, and another several minutes before the data can be retransmitted. If a large number of data packets arrive garbled, transmitting data to Abuja could be a very long, tedious process.

SELF-ASSESSMENT EXERCISE

In sliding window error control system in which each packet is numbered, station A sends packet 4,5,6,and 7. Station B receives them and wants to acknowledge all of them. What does station B send back to station A?

4.0 CONCLUSION

Our discussion so far in this Unit is the identification of mechanisms to detect and correct errors. This mechanism is called error control. We identified Stop-and-wait protocol and Sliding window error control. Correcting an error is not that simple. We also described the use of Hamming code as a technique to fix error in computer networks.

5.0 SUMMARY

Once an error has been detected, there are three possible options: do nothing, return an error message, and correct the error. The "doing nothing" option is used by some of the newer transmission technologies, such as frame relay. Frame relay assumes that fiber-optic lines will be used, which significantly reduces the chance for errors. If an error does occur, a higher-layer protocol will note the frame error and will perform some type of error control. The option of returning an error message to the transmitter is the most common response to an error and involves using Stop-and- wait protocols and sliding window protocols. A Stop-and-wait protocol allows only one packet to be sent at a time. Before another packet can be sent, the sender has to receive a positive acknowledgment. A sliding window protocol allows multiple packets to be sent at one time. A receiver can acknowledge multiple packets with a single acknowledgment. Error correction is a possibility if the transmitted data contains enough redundant information so that the receiver can properly correct the error without asking the transmitter for additional information. This form of error control requires a high amount of overhead and is used only in special applications in which the retransmission of data is not desirable.

6.0 TUTOR-MARKED ASSIGNMENT

1. In a stop-and-wait error control system, station A sends packet 0, it arrives without error, and an ACK is returned, but the ACK is lost. What happens next?
2. In a sliding window error control system, Station A sends packet with bytes 501-700, followed immediately by a packet with bytes 701-900. Create a diagram of this error control scenario, and show the response(s) that Station B will send if there are no errors

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network*, A Business User's Approach Fourth Edition: Bob Woo bury, Canada.

Aftab, A (2003). *Data Communication Principles for Fixed and Wireless Networks*, Kluwer Academic Publisher, USA.

UNIT 3 LOCAL AREA NETWORKS: THE BASICS

CONTENTS

1.0 INTRODUCTION

A local area network (LAN) is a communications network that interconnects a variety of data communications devices within a small geographic area and broadcasts data at high data transfer rates. Recall that in Unit 1, we briefly looked at language of computer network which include: local area network, and data communication definition. The phrase "data communications devices" covers computers such as personal computers, computer workstations, and mainframe computers, as well as peripheral devices such as disk drives, printers; and modems. Data communications devices could also include items such as motion, smoke, and heat sensors; fire alarms; ventilation systems; and motor speed controls. These latter devices are often found in businesses and manufacturing environments where assembly lines and robots are commonly used. ! The next piece of the definition, "within a small geographic area" usually implies that a local area network can be as small as one room, or can extend over multiple rooms, over multiple floors within a building, and even over multiple buildings within a single campus-area. The most common geographic areas, however, are a room or multiple rooms within a single building. This unit will discuss into detail the concept of local area network.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- State the definition of a local area network
- List the primary functions, activities, and application areas of local area networks
- Identify physical and logical topologies of local area networks.

3.0 MAIN CONTENT

3.1 Primary Function of Local Area Networks

To better understand the capabilities of local area networks, let's examine their primary function and some typical activities and application areas. The majority of users expect a local area network to provide access to hardware and software resources that will allow them to perform one or more of the following activities in an office, academic, or manufacturing environment: file serving, database and application serving, print serving, Internet accessing, e-mailing, video and music transfers, process control and monitoring, and distributed processing. A local area network performs file serving when it's connected to a workstation with a large storage disk drive that acts as a central storage repository, or file server. -For example, when the local area network offers access to a high-level application such as a commercial project management application, the network stores the project management software (or a portion of it) on the file server and transfers a copy of it to the appropriate workstation on demand. By keeping all of the application on the server or more likely, part of it on the server and part of it on the client work station, the network can control access to the software and can reduce the amount of disk storage required on each user's workstation for this application. For a second example, suppose two or more users wish to share a data set. In this case, the data set, like the application software, would be stored on the file server, while the network provided access to those users who had the appropriate permissions. A local area network can also provide access to one or more high-quality printers. The local aria network software called a print server provides workstations with the authorization to access a particular printer, accepts and queues prints jobs, prints cover sheets, and allows users access to the job queue for routine administrative functions.

Most local area networks provide the service of sending and receiving e-mail. This e-mail service can operate both within the local area network and between the local area network and other networks, such as the Internet (to be discussed later in this course). Stored somewhere on the

network is a database of e-mail messages, both old and new. When users log in to access their e-mail, their messages are stored and retrieved from the e-mail server. A local area network can interface with other local area networks, wide area networks (such as the Internet), and mainframe computers. Thus, a local area network is often the glue that holds together many different types of computer systems and networks. A company can use a local area network's interfacing ability to enable its employees to interact with people external to the company, such as customers and suppliers. For example, if employees wished to send purchase orders to vendors, they could enter transactions on their workstations. These transactions would travel across the company's local area network, which would be connected to a wide area network. The suppliers would eventually receive the orders by being connected to this wide area network through their own local area network.

Figure 8.1 shows typical interconnections between a local area network and other entities. It is common to interconnect one local area network to another local area network via a device such as a switch. Equally common is the interconnection of a local area network to a wide area network via a router. A local area network can also be connected to a mainframe computer to enable the two entities to share each other's resources.

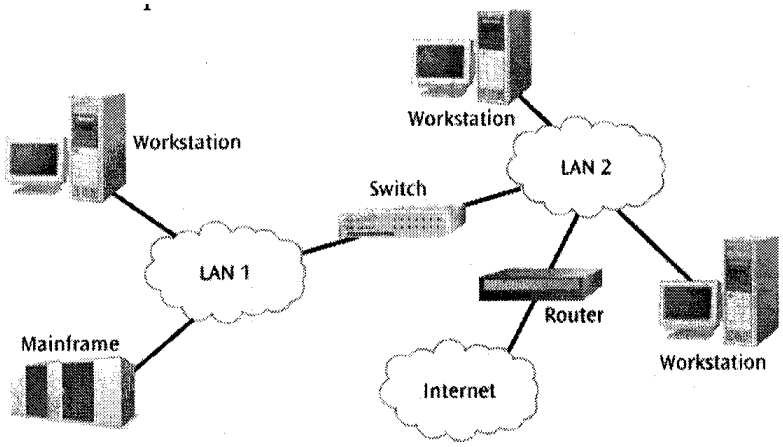


Figure 8.1: A LAN interconnecting another LAN the Internet, and mainframe computer

Many higher-speed local area networks provide the capabilities of transferring video images and video streams. For example, a local area network could allow a user to transfer high-resolution graphic images, transfer video streams, and perform teleconferencing between two or more users. In manufacturing and industrial environments, local area networks are often used to monitor manufacturing events and report and control their occurrence. The local area network provides process control and monitoring. An automobile assembly line that uses sensors

to monitor partially completed automobiles and control robots for assembly is an excellent example of a local area network per- forming process control functions.

Depending on the type of network and the choice of network operating system, a local area network may support distributed processing, in which a task is subdivided and sent to remote workstations on the network for execution. Often- times, these remote workstations are idle; thus, the distributed processing task amounts to the "stealing" of CPU time from other machines (and is often called grid computing). The results of these remote executions are then returned to the originating workstation for dissemination or further processing. By delegating tasks .to those computers that are most capable of handling specific chores, the distribution of tasks or parts of tasks can lead to an increase in execution speed. In addition to performing these common activities, a local area network can be an effective tool in many application areas. One of the most common application areas is an office environment. A local area network in an office can provide word processing, spreadsheet operations, database functions, electronic mail (e-mail) access, Internet access, electronic appointment scheduling, and graphic image creation capabilities over a wide variety of platforms and to a large number of workstations.

Completed documents can be routed to high-quality printers to produce letterheads, graphically designed newsletters, and formal documents. A second common application area for a local area network is an academic environment. In a laboratory setting, for example, a local area network can provide students with access to the tools necessary to complete homework assignments, send e-mail, and interact with the Internet. In a classroom setting, a local area network can enable professors to deliver tutorials and lessons with high-quality graphics and sound to students. Multiple workstations can be used to provide students with instruction at their own pace, while the instructor monitors and records each student's progress at every workstation. A third common application area for a local area network is manufacturing. In fact, modern assembly lines operate exclusively under the control of local area networks. As products move down the assembly line, sensors control position, robots perform mundane, exacting, or dangerous operations; and product subassemblies are inventoried and ordered. The modern automobile assembly line is a technological tour de force, incorporating numerous local area networks and mainframe computers. Now that we're familiar with the more common activities and applications of local area networks, let's examine some advantages and disadvantages.

3.2 Advantages and Disadvantages of Local Area Networks

One of the biggest advantages of local area networks is their ability to share resources in an economical and efficient manner. Shared hardware resources can include high-quality printers, tape-backup systems, plotters, CD-jukeboxes, mass storage systems, and other hardware devices. On the software end, local area networks allow the sharing of commercial applications, in-house applications, and data sets with all user workstations.

Also, with respect to communications, each workstation in a local area network can send and receive messages to and from other workstations and networks. This intercommunication allows users to send e-mail, access Web pages, send print jobs, and retrieve database records. (An interesting side effect of local area networks is that an individual workstation can survive a network failure if the workstation does not rely on software or hardware found on other workstations or on the server.) An additional advantage is that component evolution can be independent of system evolution, and vice versa. For example, if new workstations are desired, it is possible to replace older workstations with newer ones with few, if any, changes to the network itself. Likewise, if one or more network components become obsolete, it is possible to upgrade the network component without replacing or radically altering individual workstations. Under some conditions, local area networks allow equipment from different manufacturers to be mixed on the same network. For example, it is possible to create a local area network that incorporates IBM-type personal computers with Sun workstations and Apple microcomputers. Two other advantages' of local area networks are high transfer rates and low error rates. Local area networks typically have data transfer rates from 10 million bits per second to 10 billion bits per second. Because of these rates, documents can be transferred across a local area network quickly and with confidence.

Finally, because local-area networks can be purchased outright, the entire network and all workstations and devices can be privately owned and maintained. Thus, a company can offer its desired services using the hardware and software it deems best for employees. Interestingly, however, some companies are beginning to view equipment purchases as a disadvantage. Supporting an entire corporation with the proper computing resources is expensive. It does not help that as a computer reaches its first birthday, there is a newer, faster, and less expensive computer waiting to be purchased. Thus, some companies lease local area network equipment or hire a third party to support their networks. In addition, local area networks have a number of disadvantages. For one, local area network hardware, the operating systems, and the software that runs on the network can be expensive. The components of

LANs that require significant funding include the network server, the network operating system, the network cabling system including hubs and switches, the network-based applications, network security, and support and maintenance. Despite the fact that a local area network can support many types of hardware and software, the different types of hardware and software may not be able to interoperate. For example, even if a local area network supports two different types of database systems, users may not be able to share data between the two database systems. Another disadvantage is the potential for purchasing software with the incorrect user license. For example, it is almost always illegal to purchase a single-user copy of software and then install it on a local area network for multiple use. To avoid using software illegally, companies must be aware of the special licensing agreements associated with local area networks. An important disadvantage that has often been overlooked in the past is that the management and control of the local area network requires many hours of dedication and service.

A manager, or network administrator, of a local area network should be properly trained and should not assume that the network can support itself with only a few hours of attention per week. Therefore, a local area network requires specialized staff and knowledge, and the right diagnostic hardware and software. Unfortunately, many hours of this support time are very often spent fighting virus; and other network security issues. Finally, a local area network is only as strong as its weakest link. For example, a network may suffer terribly if the file server cannot adequately serve all the requests from users of the network. Upon upgrading a server, a company may discover that the cabling is no longer capable of supporting the higher traffic. Upon upgrading the cabling, it may become apparent that the network operating system is no longer capable of performing the necessary functions. Upgrades to part of a network can cause ripple effects throughout the network, and the cycle of upgrades usually continues until it is once again time to upgrade the server.

Considering all the advantages and disadvantages associated with local area networks, it should not be surprising that the decision to incorporate a local area network into an existing environment requires much planning, training, support, and money. Let's now look more closely at how the workstations in a local area network are interconnected to serve the activities and applications discussed so far.

3.3 Basic Local Area Network Topologies

The workstations within a local area network today are interconnected in two basic configurations, or topologies: the star-wired bus, and wireless. But this has not always been the case. Originally there were four

configurations-: bus/tree, star-wired bus, star-wired ring, and wireless. (While recent market demand has favored and given prominence to two of these four configurations, understanding how all four topologies operate is useful because the topologies build on each other in important ways.) The choice of local area network topology is occasionally dictated by the physical environment in which the local area network is to be placed. More likely, how- ever, the choice of a topology is determined by other factors such as a preferred access method, desired data transfer speeds, availability of applications and hardware, and brand loyalty. Let's examine each of the four topologies in the chronological order of their emergence, paying special attention to their advantages and disadvantages.

3.3.1 Bus/Tree Topology

The bus/tree topology, often simply called the bus topology, was the first topology used when local area networks became commercially available in the late 1970s, and it essentially consists of a simple cable, or bus, to which all devices attach. As shown in Figure 8.2, the bus is simply a linear coaxial cable into which multiple devices or workstations tap.

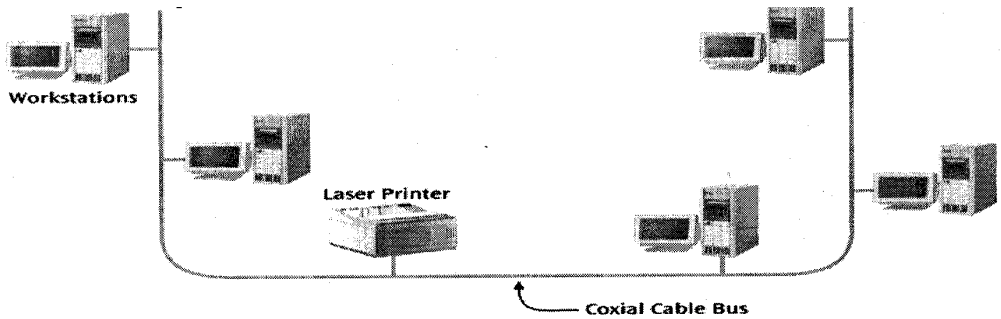


Figure 8.2: Simple diagram of a local area network bus topology

When a device transmits on the bus, all other attached devices receive the trans- mission. Connecting to the cable requires a simple device called a tap (Figure 8-3). This tap is a passive device, as it does not alter the signal and does not require electricity to operate. On the workstation end of the cable is a network interface card. The network interface card (NIC) is an electronic device, typically in the form of a computer circuit board, that performs the necessary signal conversions and protocol operations that allow the workstation to send and receive data on the network.

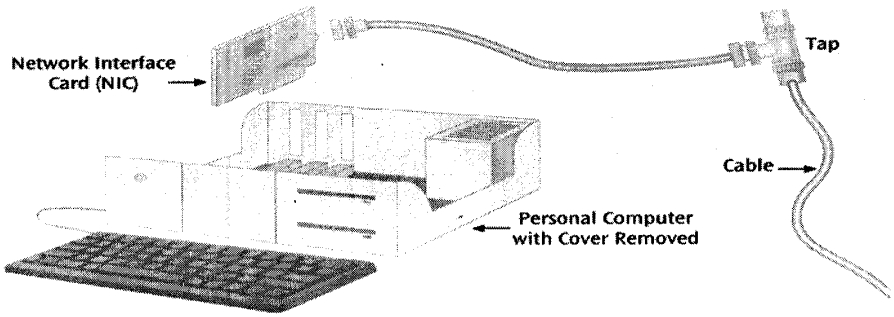


Figure 8.3: Tap used to interconnect a workstation and a LAN cable

Two different signaling technologies can be used with a bus network: baseband signaling and broadband signaling. Baseband signaling typically uses a single digital signal (with Manchester encoding) to transmit data over the bus. This single digital signal uses the entire spectrum of the cable; therefore, only one signal at a time can be transmitted on the cable. All workstations must be aware that another workstation is transmitting, so they do not attempt to transmit and thereby inadvertently destroy the signal of the first transmitter. Allowing only one workstation access to the medium at one time is the responsibility of the medium access control protocol, which will be discussed in detail in Unit 9. Another characteristic of baseband technology that's worth noting is that baseband transmission is bidirectional, which means that when the signal is transmitted from a given workstation, the signal propagates away from the source in both directions on the cable (Figure 8-4).

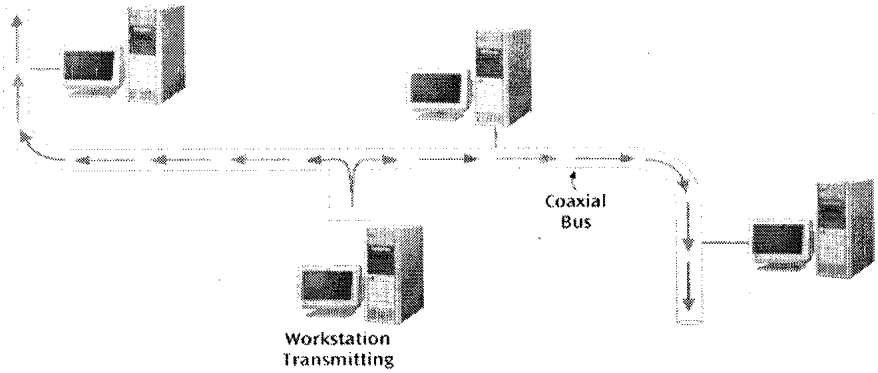


Figure 8.4: Bidirectional propagation of a baseband signal

The second type of signaling technology used on the bus local area network is broadband technology. Broadband technology very often uses analog signaling in the form of frequency division multiplexing to divide the available medium into multiple channels. Each channel is capable of carrying a single communication between two workstations. Because the medium can be divided into multiple channels, broadband signaling allows multiple concurrent communications. It is also possible to split and join broadband cables and signals to create con- figurations

more complex than a single linear bus. These more complex bus topologies consisting of multiple interconnected cable segments are termed trees. Figure 8.5 shows an example of a tree network.

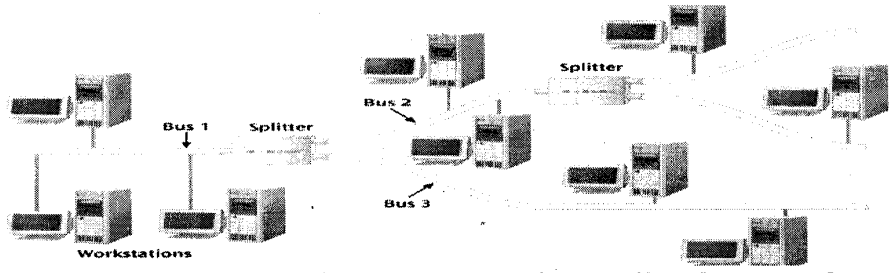


Figure 8.5: Simple example of a broadband tree topolofy

All bus networks whether broadband or baseband share a major disadvantage: In general, it is difficult to add a new workstation if no tap currently exists. Because there is no tap, the cable has to be cut, and a tap has to be inserted. Cutting the cable and inserting a tap disrupts the traffic on the network and is a somewhat messy job. The best way to avoid this is to anticipate where workstations will be and have the installation team install all the necessary taps in advance. As you might expect, however, predicting the exact number and location of taps is virtually impossible. With the introduction of other topologies, bus-based local area networks have lost popularity to the point that relatively few bus-based local area networks are installed today. The only bus that is still regularly used is the bus that delivers the video and data signals of cable television. One reason for this is that, if you recall, coaxial cable is a good medium for transmitting the high-frequency signals of cable television.

Let's examine the topology that replaced the local area network bus: the star- wired bus.

3.3.2 Star-wired Bus Topology

The most popular configuration for a local area network is the star-wired topology, or simply, star topology. This topology should not be confused with an older topology also called the star topology. The older star topology supported a local area network called the Star LAN, in which one computer at the center of the star controlled the transmissions of all the other workstations. Today's modem star-wired bus topology acts like a bus but looks like a star. To be a little more precise, the topology logically acts as a bus, but it physically looks like a star. The logical design of a network determines how the data moves around the network from, workstation to workstation. The physical design refers to the pattern formed by the locations of the elements of the network, as it

would appear if drawn on a sheet of paper. Let's explore the details of this important distinction further. In a star-wired bus topology, all workstations connect to a central device such as the hub as you can see in Figure 8.6.

The hub is a nonintelligent device that simply and immediately retransmits the data it receives from any workstation out to all other workstations (or devices) connected to the hub. All workstations hear the transmitted data, because there is only a single transmission channel, and all workstations are using this one channel to send and receive. Sending data to all workstations and devices generates a lot of traffic but keeps the operation very simple, because there is no routing to any particular workstation. Thus, with regard to its logical design, the star-wired bus is acting as a bus: when a workstation transmits, all workstations (or devices) immediately receive the data. The network's physical design, however, is a star, because all the devices are connected to the hub and radiate outward in a starlike (as opposed to linear) pattern.

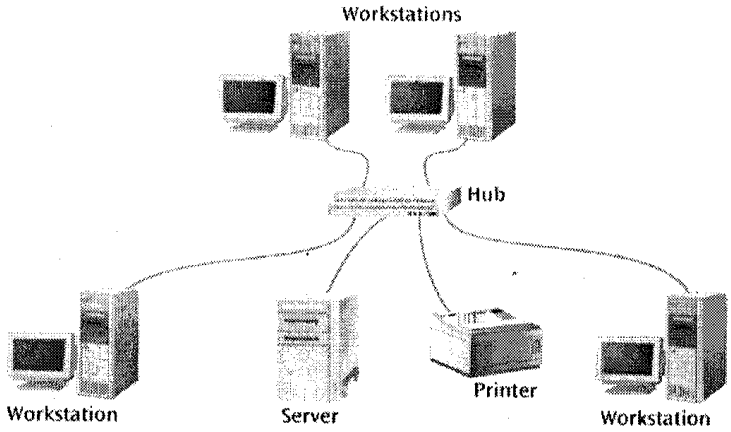


Figure 8.6: Simple example of a star-wired bus

Many hubs support multiple types of media. Twisted pair cabling has become the preferred medium for star-wired bus topologies, while fiber-optic cable is typically used as a connector between multiple hubs. The connectors on the ends of the twisted pair cables are simple-to-use modular RJ-45 connectors. The RJ-45 connector is very similar to (but a little bit wider than) the modular connector that connects a telephone to the wall jack (an RJ-II connector). Twisted pair cable and modular connectors have made it much simpler to add workstations to a star-wired bus than to a coaxial-cabled bus. The many advantages of a star-wired bus topology include simple installation and maintenance, low-cost components (such as hubs and twisted pair wiring), and high volume of compatible products due to major market share. Perhaps the only disadvantage of a star design is the amount of traffic its hub(s) must handle.

When two or more hubs are interconnected and a workstation transmits data, all the workstations connected to all the hubs receive the data. This is an example of a shared network. As has been noted, the hub is a relative nonintelligent device. It does not filter out any data frames, and it does not perform any routing. Later, you will see that the hub can be replaced with a more advanced device called a switch, which can reduce the amount of traffic on the network.

3.3.3 Star-wired Ring Topology

With respect to logical design, the star-wired ring topology is a circular connection of workstations, as Figure 8.7 depicts. The star-wired ring is essentially a marriage of the star-wired topology and topology that was used (albeit sparingly) in the 1980s and is known simply as ring topology. Because star-wired ring topologies support baseband signals, the star-wired ring is capable of supporting only one channel of information. This channel of information flows in one direction around the ring, moving from workstation to workstation. Because the star-wired ring is a closed loop of wire, it is important for some devices to remove a circling piece of data from the ring; otherwise, the piece of data will keep circling. The device that removes the data is the workstation that originally transmitted the data. Although the logical organization of the workstations in a star-wired ring topology is circular, the physical organization of a star-wired ring is not circular. Physically, a star-wired ring looks much like a star-wired bus design, with all its workstations connected to a central device. This central device is not a hub, however, but rather a multi station access unit. A Multistation Access Unit (MAU) accepts data from a workstation and 'transmits this data to the next workstation downstream in the ring (see Figure 8.8).

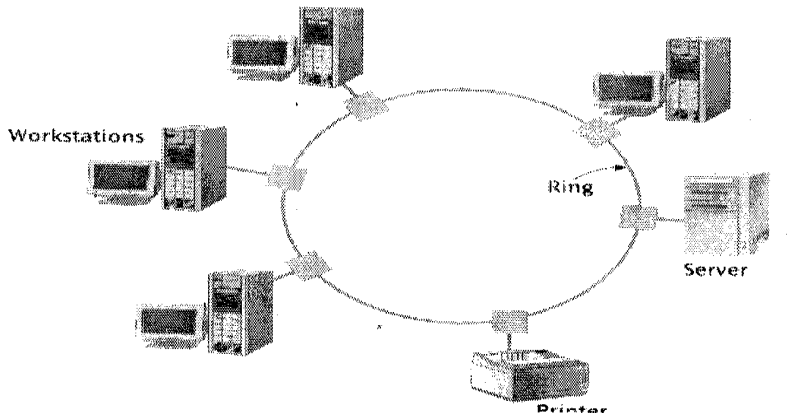


Figure 8.7: Star-wired ring topology

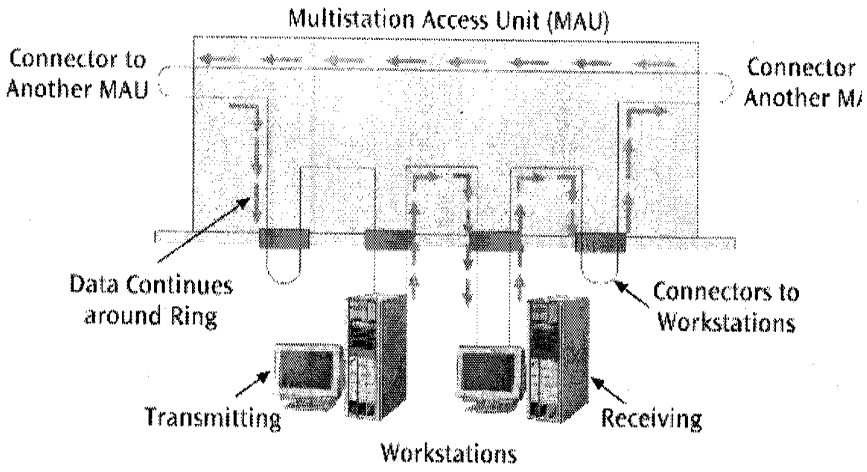


Figure 8.8: Multistation Acess Unit on a star-wired ring topology

A Multistation Acess Unit is quite a bit different from a hub in that it does not send a copy of the incoming data immediately out to every connection unlike a hub that broadcasts all incoming signals onto all connected links immediately. Like hubs, MAUs . The star-wired ring topology has many of the same advantages as the star-wired us topology. The star-wired ring topology is based on twisted pair wiring, which makes installing new workstations easy, and this makes the topology easy to maintain. Some of the disadvantages of star-wired rings include: slower transmission speeds, higher costs, and more complex software.

3.3.4 Wireless Topology

This is not really a specific topology since a workstation in a wireless LAN can be anywhere as long as it is within transmitting distance to an access point. Several versions of IEEE 802.11 standard defines various forms of wireless LAN connections. Workstations reside within a basic service set, while multiple basic service sets create an extended service set. There are two basic components necessary: the client radio, usually a PC card with an integrated antenna installed in a laptop or workstation, and the access point (AP), which is an Ethernet port plus a transceiver. The AP acts as a bridge between the wired and wireless networks and can perform basic routing functions. Workstations with client radio cards reside within a basic service set, while multiple basic service sets create an extended service set. Figure 8.9 and 8.10 shows single and multiple cell wireless configuration respectively.

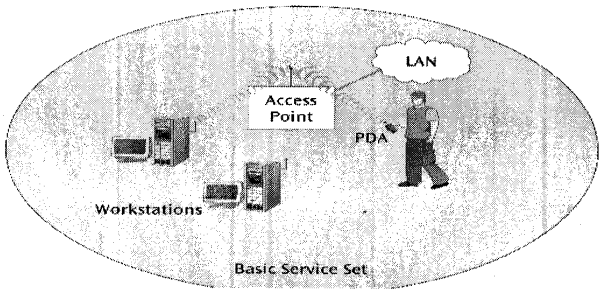


Figure 8.9: Single-cell wireless LAN configuration

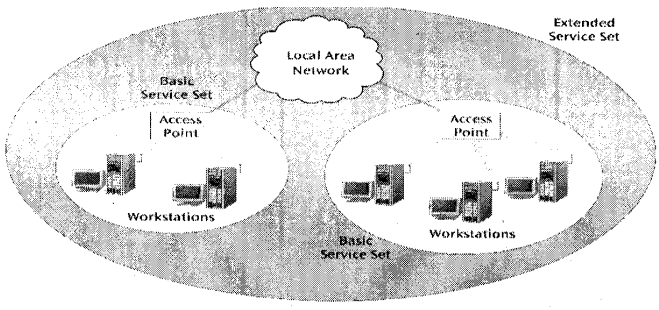


Figure 8.10: Multiple-cell wireless LAN configuration

The original wireless standard, capable of transmitting data at 2 Mbps is IEEE 802.11, while the second wireless standard, capable of transmitting data at 11 Mbps is IEEE 802.11 b. With directional antennae designed for point-to-point transmission (rare), 802.11 b can transmit for more than 10 miles, while an omni -directional antenna on a typical AP, range may drop to as little as 100 feet. IEEE 802.11 a is one of the more recent standards, capable of transmitting data at 54 Mbps (theoretical) using the 5 GHz frequency range and IEEE 802.11 g -The other recent standard, also capable of transmitting data at 54 Mbps (theoretical) but using the same frequencies as 802.11 b (2.4 GHz) and is backwards compatible with 802.11 b. To provide security, most systems use either Wired Equivalent Privacy (WEP), which provides either 40- or 128-bit key protection. Wireless LANs may also be configured without an access point. These configurations are called "ad-hoc". Table 7.1 compared the bus, star-wired bus, star-wired ring, and wireless topologies.

SELF-ASSESSMENT EXERCISE

i. What are the primary functions of a Local Area Network?
ii. Identify the advantages and disadvantages of a Local Area Network
iii. List three application areas of a Local Area Network

Table 7.1: Comparison of the bus, star-wired bus, star-wired ring & wireless topologies

	Baseband Bus	**Broadband Bus/Tree**	**Star-Wired Bus**	**Star-Wired Ring**	**Wireless**
Signaling technique	Digital	Analog	Digital	Digital	Analog
Physical layout	Linear	Linear	Star-like	Star-like	Star-like
Usual media type	Coaxial cable	Coaxial cable	Twisted pair	Twisted pair	Airwaves
Installation ease	Moderate	Moderate	Easy	Easy	Easy
New workstation installation	Hard (if no tap available)	Hard (if no tap available)	Easy (if port available)	Easy (if port available	Very easy
Concurrent channels	No	Yes	No	No	Yes
New LAN installations	No	No	Yes	No	Yes

ANSWER TO SELF-ASSESSMENT

Functions of LANs

The primary functions of a LAN are to enable the sharing of data, software, and peripherals and to provide common services such as file serving, print serving, support for electronic mail, and process control and monitoring in office, academic, and manufacturing environments.

Advantages and Disadvantages of LANs Local area networks have numerous advantages, including resource sharing, separate component and network evolution, high data transfer rates, and low error rates. Local area networks also have numerous disadvantages, including relatively high costs, a high degree of maintenance, and the constant need for upgrades.

Application Areas LAN

Access to remote databases
Value added network
Access to remote program

4.0 CONCLUSION

This unit begins by discussing the primary function of a local area network as well as its advantages and disadvantages, and application areas Next, the basic physical (hardware) layouts or topologies of the most commonly found local area networks are discussed. We conclude the unit comparing different existing topologies.

5.0 SUMMARY

A local area network is a communications network that interconnects a variety of data communications devices within a small area and transfers data at high transfer rates with very low error rates. The primary functions of a LAN are to enable the sharing of data, software, and peripherals and to provide common services such as file serving, print serving, support for electronic mail, and process control and monitoring in office, academic, and manufacturing environments. Local area networks have numerous advantages, including resource sharing, separate component and network evolution, high data transfer rates, and low error rates. Local area networks also have numerous disadvantages, including relatively high costs, a high degree of maintenance, and the constant need for upgrades. A local area network can be configured as a bus/tree topology, a star-wired bus topology, a star-wired ring topology, or a wireless network. A baseband bus topology local area, network uses digital signaling and supports one channel. A broadband bus topology local area network uses analog signaling and can support hundreds of simultaneous channels. With both baseband and broadband buses, expansion in the form of adding a new workstation is difficult when a tap is not available. The star-wired bus topology is a variation on the bus topology, but because it includes a hub, it has the advantages of easier installation and maintenance. The star-wired bus has essentially replaced the baseband bus and broadband bus LANs. The star-wired ring topology is a circular connection of workstations in which data is passed from workstation to workstation around the ring. Star-wired buses have also replaced star-wired rings. The wireless topology allows a highly flexible placement of workstations and requires no wiring to transmit and receive data.

6.0 TUTOR-MARKED ASSIGNMENT

1. Describe an example of a broadband bus system
2. Physically, a hub looks the same as a MAU. Logically they are different. Explain how they are different.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network*, A Business User's Approach Fourth Edition: Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks: Kluwer Academic Publisher, USA.

UNIT 4 MEDIUM ACCESS CONTROL

CONTENTS

1.0 INTRODUCTION

In Unit 8 we discussed the concept of local area networks. For a workstation to place data onto a local area network, the network must have a medium access control protocol. A medium access control protocol is the software that allows a workstation to place data onto a local area network. Depending on the network's topology, several types of medium access control protocols may be applicable. In a broadcast network, it is imperative that only one workstation at a time wave be allowed to transmit its data onto the network. In this present unit, however, we will concentrate on one workstation transmitting at a time, while unit 10 will introduce switches and full-duplex connections in more detail.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- Specify the different medium access control protocols
- Recognize the different IEEE 802 frame formats
- Describe the common local area network systems

3.0 MAIN CONTENT

3.1 Categories of Medium Access Control

There are two basic forms of medium access control protocols for local area network. These are contention-based (such as Carrier Sense Multiple Access/Collision Detection(CSMAlCD), which is found on star-wired bus local area networks, and Carrier Sense Multiple Access/Collision A voidance(CSMAlCA), which is found on wireless local area networks),and round-robin (such as token passing or token ring). Let's examine both of these protocols.

3.1.1 Contention-based Protocols

A contention-based protocol is basically a first-come, first-served protocol- the first station to recognize that no other station is transmitting data and place its data onto the medium is the first station to transmit. The most popular contention-based protocol is carrier sense multiple access with collision detection(CSMAlCD). The CSMAlCD medium access control protocol is found almost exclusively on star-wired bus and bus local area networks and is therefore the most widely used medium access control protocol. The name of this protocol is so long that it almost explains itself. With the CSMA/CD protocol, only one workstation at a time can transmit, and because of this, the CSMAlCD protocol is basically a half-duplex protocol. A workstation listens to the medium-that is, senses for a carrier on the medium-to learn whether any other workstation is transmitting. If another workstation is transmitting, the workstation wanting to transmit will wait and try again to transmit. The amount of time the workstation waits depends on the particular type of CSMAlCD protocol. If no other workstation is currently transmitting, the workstation transmits its data onto the medium. The CSMA/CD access protocol is analogous to human beings carrying on a conversation at the dinner table in the dark. If no one is talking, someone can speak. If someone is talking, everyone else hears this and waits. If two human beings start talking at the same time, they both stop immediately (or at least polite people do) and wait a certain amount of time before trying again. In most situations, the data being sent by a workstation is intended for one other workstation, but all the workstations on the CSMAlCD protocol-based network receive the data. (Once again, we will see that this is no longer true with the switched local area networks introduced in the next unit). Only the intended workstation (the workstation with the intended address) will do something with the data. All the other workstations will discard the frame of data.

As the data is being transmitted, the sending workstation continues to listen to the medium, listening to its own transmission. Under normal conditions, the workstation should just hear its own data being transmitted. If the workstation hears garbage, however, it assumes a collision has occurred. A collision occurs when two or more workstations listen to the medium at the same moment, hear nothing: and then transmit their data at the same moment. Actually, the two workstations do not need to begin transmission at exactly the same moment for a collision to occur. Consider a situation in which two workstations are at opposite ends of a bus. A signal propagates from one end of the bus to the other end in time n. A workstation will not hear a collision until its data has, on average, travelled halfway down the bus, collided with the other workstation's signal, and then propagated back down the bus to the first station (Figure 9.1). This interval, during which the signals propagate down the bus and back, is the collision window. During this collision window, a workstation might not hear a transmission, falsely assume that no one is transmitting, and then transmit its data.

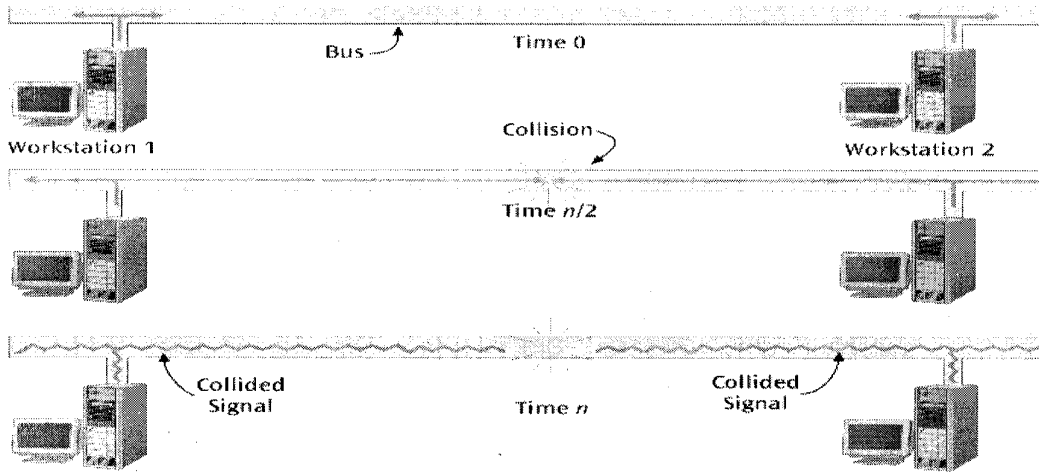


Figure 9.1: Two workstation experiencing a collision

If the network is experiencing a small amount of traffic, the chances for collision are small. The chance for a collision increases dramatically when the network is under a heavy load and many workstations are trying to access it simultaneously. Studies have shown that as the traffic on a CSMA/CD network increases, the rate of the collisions increases, which further degrades the service of the network. If a workstation detects a collision, it will immediately stop its transmission, wait some random amount of time, and try again. If another collision occurs, the workstation will wait once more. Because of these collisions, busy CSMA/CD networks rarely exceed 40 percent throughput. In other words, busy CSMA/CD networks waste 60 percent of their time dealing with collisions and other overhead. Many users find this rate

unacceptable and consider modifications to standard CSMAlCD, such as using switches instead of hubs.

Because the number of times a workstation will have to wait is unknown, it is not possible to determine exactly when a workstation will be allowed to transmit its data without collision. Thus, CSMA/CD is a nondeterministic protocol. A nondeterministic protocol is one in which you cannot calculate the time at which a workstation will transmit. If your application must have its workstations transmit data at known times, you might want to consider a medium, access consider a medium access protocol other than CSMA/CD.

3.1.2 Wireless CSMA/CA

The contention-based medium access control protocol that supports wireless local area networks has two interesting differences, from the CSMA/CD protocol found on wired LANs. First, there is no collision detection. In other words, the transmitter does not listen during its transmission to hear if there was a collision with another signal somewhere on the network. The reasoning is that if two workstations are so far apart that they cannot hear each other's transmission signal, then they won't hear a collision. Instead, the algorithm of the protocol supporting wireless LANs limits when a workstation can transmit, in an attempt to reduce the number of collisions. The type of algorithm that tries to avoid collisions is called carrier sense multiple access with collision avoidance (CSMAlCA). How does the algorithm limit when a workstation can transmit? That answer is tied to the second interesting difference-priority levels. In an attempt to provide a certain level of priority to the order of transmission, the CSMA/CA algorithm has been modified to function according to the following rule: If a user device wishes to transmit and the medium is idle, the device is not allowed to transmit immediately. Instead, the device is made to wait for a small period of time called the interframe space (IFS). If the medium is still idle after this interframe space, the device is then allowed to transmit.

How does the interframe space provide a priority system? There are up to three different interframe space times. The first IFS time-short IFS-is used by devices that require an immediate response, such as an acknowledgment, a clear to send, or a response to a poll. The second IFS time-midlength IFS-is used by the access device when it's issuing polls to the user devices. The third IFS time-long IFS-is used as a minimum delay for ordinary user devices when they are contending for access to the network. Thus, before a standard user device can transmit, it must wait and give higher priority devices a chance to transmit first. If the medium is initially busy, the device simply continues to listen to the medium. When the medium becomes idle, the user device delays for the

interframe space. If the medium is still idle after the interframe space, the user device selects a random backoff factor. When the backoff counter reaches zero, the device transmits the packet. This procedure helps prevent a number of users from hearing an idle medium, transmitting at the same moment, and causing a collision.

3.1.3 Round-robin Protocols

A round-robin protocol is a protocol in which each workstation takes a turn at transmission, and the turns are uniformly distributed over all workstations. In contrast to the first-come, first- served contention-based protocols, round-robin protocols specify that if multiple workstations are waiting to transmit, each workstation will have to wait until its turn comes around. The advantage is that each workstation will eventually get a turn and cannot be forced out by another workstation that seizes the communications channel first. The most popular example of the round-robin protocol is the token-passing protocol. Before a workstation can transmit, it must possess the token. Once the transmission is complete, the workstation releases the token to the next workstation, in round-robin order. Eventually, the token is passed around to all workstations and returns to the first workstation, to begin another cycle. Two types of token-passing protocols exist: token ring and token bus. Although they both use a token-passing algorithm, they have different underlying topologies. Let's introduce the more popular of the two algorithms-token ring.

Token Ring

The token ring local area network uses the star-wired ring topology for the hardware and a round-robin protocol for the software. It operates on the principle that to transmits data onto the ring, your workstation must be currently in possession of a software token. There is typically only one token in the entire network, so only one workstation may transmit at a time. When a workstation has completed its transmission, it passes the token on to the downstream neighboring workstation. Only the workstation holding the token can transmit, so there is no need for any workstation to listen for a collision while transmitting, because collisions cannot occur. As has been mentioned, collisions are one of the main problems of CSMA/CD networks. As the number of concurrent users rises, the number of collisions rises. As the collisions rise, more workstations are forced to retransmit their messages, and overall throughput declines. Because the token ring does not experience any collisions, overall throughput remains high even under heavy loads. This capability of token ring to give every workstation a turn is attractive and is valuable for applications that require uniform response times. Because the order of transmission by each workstation is known, the wait time to

transmit can be determined (as opposed to being unpredictable); thus, token ring is a deterministic protocol. Let's take a quick look at how the token ring protocol works. Consider Figure 9.2 in which Station A has just released the token. Because Station B is the next down-stream neighbor from Station A, and Station B has data to transmit on the ring, Station B seizes the token. After seizing the token, Station B transmits its data, which is destined for Station M. As Station M copies in the data frame, the data continues around the ring until it returns to Station B, which removes the data from the ring. After Station B has removed its data from the ring, it passes the token to Station C.

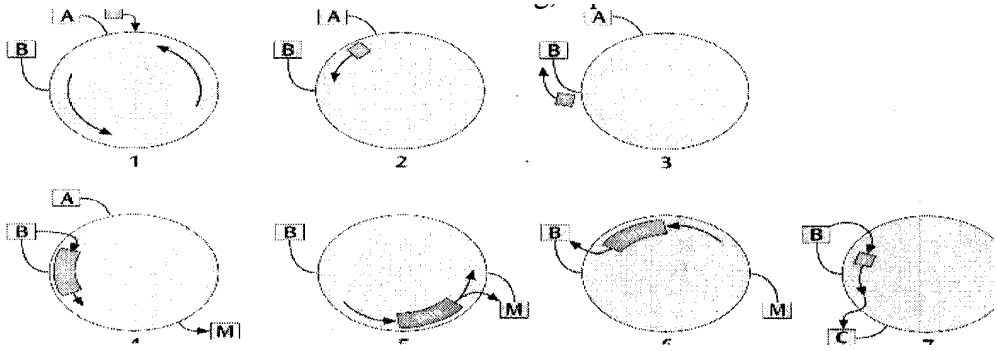


Figure 9.2: Data transmission on a token ring local area network

A major disadvantage of the token ring access protocol is the complexity of the software needed to maintain the token. This software has to address important questions such as: What happens if the token disappears? (A workstation does not forward it.). If the token disappears, who generates a new token? Is it possible for two stations to generate a token, thus resulting in two tokens on the ring? Although token ring has the definite advantage of being a deterministic protocol and performing quite well under heavy loads, it has had a difficult time competing with CSMA/CD networks. In fact, an overwhelming majority of local area networks use CSMA/CD as the medium access control protocol of choice. Some reasons that CSMA/CD is more popular than token ring are:

- CSMA/CD was the first local area network medium access control method, and thus got a good jump on installations and support.
- Token ring local area networks have almost always lagged behind CSMA/CD networks with regard to transmission speed. When CSMA/CD first became popular, the typical transmission speed was 10 Mbps. Token' ring, when it first appeared, had a transmission speed of only 4 Mbps. For a while, token ring jumped ahead with a 16-Mbps version, but CSMA/CD caught 'up with a 100-Mbps version, and then a 1000-Mbps version. Token

ring finally announced a 100-Mbps version, but this was too late to save the protocol in the marketplace. Many people feel it will just be a matter of time before token ring fades into the history books.

- CSMA/CD is less expensive to implement, due in part to its widespread marketing and acceptance.
- CSMA/CD.is a simpler protocol.

3.2 IEEE 802

When ISO created the OSI model in the 1970s, local area networks were just beginning to appear. In order to better support the unique nature of local area networks and to create a set of industry-wide standards, the IEEE produced a series of protocols under the name 802 (some of which you've already encountered in your reading). One of the first things the IEEE 802 protocols did was to split the data link layer into two sub layers: the medium access control sub-layer and the logical link control sub-layer (Figure 9.3).

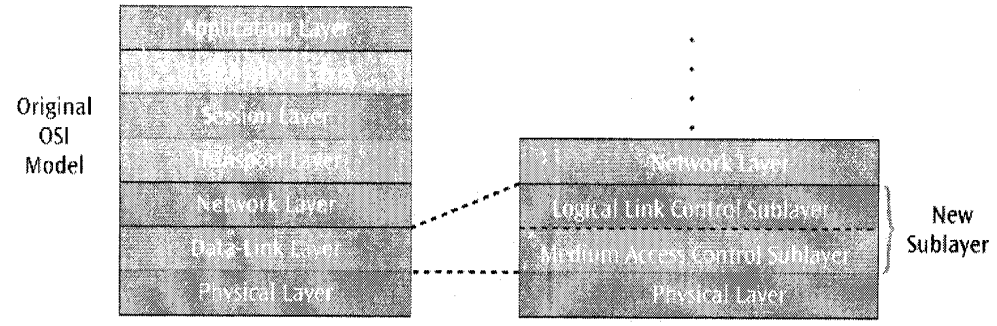


Figure 9.3: Data link layer split into two sublayers

The medium access control (MAC) sub layer works more closely with the physical layer and contains a header, computer (physical) addresses, error-detection codes, and control information. Because of this closeness with the physical layer, there is not a strictly defined division between the MAC sub layer and the physical layer. The logical link control (LLC) sub layer is primarily responsible for logical addressing and providing error control and flow control information. The medium access control sublayer defines the layout or format of the data frame, simply called the frame. As you will see in the next section, there are a number of different frame formats, depending on the type of local area network. For example, CSMA/CD LANs have one frame format, while token ring LANs have another format. Within this frame format are the fields for error detection, workstation addressing, and various types of control information. Thus, the MAC sub-layer is a very important layer when it comes to describing a local area network. Let's examine the

most common MAC sub layer frame format: IEEE 802.3 format for CSMA/CD networks.

3.2.1 IEEE 802.3 Frame Format

The IEEE 802.3 standard for CSMA/CD uses the frame format shown in Figure 9.4. The preamble and start of frame byte fields combine to form an 8-byte flag that the receiver locks onto for proper synchronization. The destination address and source address are the 2- or 6-byte addresses (of these, the 2-byte is less common) of the receiving computer and sending computer. More precisely, each network interface card in the world has a unique 6-byte (48-bit) address. When CSMA/CD sends data to a particular computer, it creates a frame with the appropriate NIC address of the intended computer. The data length is simply the length in bytes of the data field, which is the following entry. The PAD field adds characters "to the frame (pads the frame). The minimum size frame that any station can transmit is 64 bytes long.

Preamble	Start of Frame Byte	Destination Address	Source Address	Data Length	Data	PAD	Checksum
7 bytes of 10101010	10101011	2 or 6 bytes	2 or 6 bytes	2 bytes	0-1500 bytes	0-46 bytes	4 bytes

Frames shorter than 64 bytes are considered runts, or frame fragments, that resulted from a collision, and these are automatically discarded. Thus, if a workstation attempts to transmit a frame in which the data field is very short, PAD characters are added to ensure that the overall frame length equals at least 64 bytes. Finally, the checksum field is a 4-byte cyclic redundancy checksum.

Our discussion of local area network technology started with an examination of the main types of network topologies: bus, star-wired bus, star-wired ring, and wireless in unit 8, while unit 9 continues with examination of the two major categories of medium access control protocols that operate on these different topologies: contention-based (CSMAlCD) and round-robin (token passing). Let's now turn our attention to the actual products or local area network systems that are found in a typical computer environment.

SELF-ASSESSMENT EXERCISE

i. Define Medium Access Control
ii. List the two categories of Medium Access Control and briefly explain the two categories.

3.3 Local Area Network Systems

Four of the most popular local area network systems are Ethernet, IBM Token Ring, Fiber Distributed Data Interface, and Wireless Ethernet. Let's examine each of these in the order in which they were introduced.

3.3.1 Wired Ethernet

Ethernet was the first commercially available local area network system and remains, without a doubt, the most popular local area network system today. The wired version of Ethernet is based primarily on the star-wired bus topology and uses the CSMAlCD medium access protocol. Because Ethernet is so popular and has been around the longest, it has evolved into a number of different forms. To avoid mass mayhem, the IEEE created a set of individual standards specifically for Ethernet or CSMAlCD local area networks, all under the category of 802.3. For your reference, the 802.3 standards to be discussed are summarized in Table 9-1.

Table 9.1: Summary of Ethernet Standards

Ethernet Standard	Maximum Transmission Speed	Signal Type	Cable Type	Maximum Segment Length
10Base5	10 Mbps	Baseband	Coaxial	500 meters
10Base2	10 Mbps	Baseband	Coaxial	200 meters
1Base5	1 Mbps	Baseband	Unshielded twisted pair	500 meters
10BaseT	10 Mbps	Baseband	Unshielded twisted pair	100 meters
10Broad36	10 Mbps	Broadband	Coaxial	3600 meters
100BaseTX	100 Mbps	Baseband	2-pair Category 5 or higher unshielded twisted pair	100 meters
100BaseT4	100 Mbps	Baseband	4-pair Category 3 or higher unshielded twisted pair	100 meters
100BaseFX	100 Mbps	Baseband	Fiber optic	1000 meters
1000BaseSX	1000 Mbps	Baseband	Fiber optic	100 meters
1000BaseLX	1000 Mbps	Baseband	Fiber optic	100 meters
1000BaseCX	1000 Mbps	Baseband	Specialized balanced copper	25 meters
1000BaseT	1000 Mbps	Baseband	Twisted pair-four pairs	100 meters
10GBase-fiber	10 Gbps	Baseband	Fiber optic	
10GBase-T	10 Gbps	Baseband	Cat 5e/6	100 meters
10GBase-CX	10 Gbps	Baseband	Twin axial	30 meters

Note the following: **MBasen, M** represent the speed, **Base** an abbreviation for baseband and **n**. For example: 100Base2: 100Mbps baseband signal with maximum distance of 200m.

3.3.2 IBM Token Ring

Token Ring was created and popularized by IBM, which was one of the few manufacturers of token ring products. Thus, token ring and IBM Token Ring are usually two names for the same thing. As these names imply, IBM Token Ring uses the star-wired ring topology and token ring access method described in unit 8. Because IBM is not a standards-making organization, the standard for this technology was created by IEEE. More precisely, the IEEE 802.5 standard defined a token ring specification for three data transmission rates: 4 Mbps, 16 Mbps, and 100 Mbps. The 100-Mbps token ring (IEEE 802.5t) is designed for workstation-to-MAU connections using either Category 5 twisted pair wire or fiber-optic cable.

3.3.3 Fiber Distributed Data Interface (FDDI)

Although many people liked the token ring local area network for its deterministic protocol and its high throughput under heavy loads, many other people were not happy with its slow 4-Mbps and 16-Mbps transmission speeds. In an attempt to marry a deterministic access protocol with a high transmission rate, the Fiber Distributed Data Interface (FDDI) ring was created. The FDDI protocol resembles a token ring network that has been lifting weights. With a data transmission speed of 100 Mbps, network distances up to 200 km, and a possible interconnection of 500 stations, an FDDI network is a vastly updated token ring network. To achieve these impressive figures, several
modifications were made to the original token ring design. These modifications included using fiber-optic cable to connect each workstation to the central MAU, creating a dual-ring topology (of a ring within a ring) in order to provide fault tolerance, using the more efficient 48/58 encoding instead of differential Manchester encoding to increase the bit per baud ratio, and making various adjustments to the token-passing protocol itself.

Despite the advantages offered by FDDI, it appears that the faster versions of Ethernet are providing fierce competition and are driving out currently existing installations of FDDI systems. Industry experts indicate that FDDI may not be around much longer. Once again, the popularity of Ethernet and all its various forms may have forced another communications protocol into near oblivion, Wireless Ethernet

3.3.4 Wireless Ethernet

This is actually another term for the three formats that we learned about earlier in unit 8: IEEE 802.llb, IEEE 802.lla, and IEEE 802.llg. As you know, 802.llb transmits data at a theoretical speed of 11 Mbps, while both 802.lla and 802.llg transmit data at a theoretical speed of 54 Mbps. It is quite common nowadays to purchase a single NIC or access point that is capable of sup-porting all three technologies. Many businesses and educational institutions now offer some form of wireless Ethernet connectivity in their offices and campuses. Advantages include a reduction in wiring and a much more flexible approach to connecting to the network. Disadvantages include implementing new technologies and, wireless security issues.

ANSWER TO SELF-ASSESSMENT EXERCISE

- A medium access control protocol is the software that allows workstations to "take turns" at transmitting data.
- Two basic categories:
  Contention-based protocols
  Round robin protocols

Contention-Based Protocols -CSMA/CD

This is essentially first come first served. Most common example is carrier sense multiple access with collision detection (CSMA/CD). In CSMA/CD ff no one is transmitting, a workstation can transmit. If two workstations transmit at the same time, a collision occurs. When the two workstations hear the collision, they stop transmitting immediately. Each workstation backs off a random amount of time and tries again. Hopefully, both workstations do not try again at the exact same time. CSMAlCD is an example of a non-deterministic protocol. Another example of contention-based protocol is CSMA/Collision Avoidance. This protocol does not listen and detect collisions. Instead, it tries to avoid collisions before they happen. All devices, before they transmit, must wait an amount of time called an intra-frame space (IFS). Some applications have a short IFS, while others have a long IFS. If two applications want to transmit at the same time, the application with the shorter IFS will go first.

Round Robin Protocols

Each workstation takes a turn transmitting and the turn is passed around the network from workstation to workstation. Most common example is

token ring LAN in which a software token is passed from workstation to workstation. Token ring is an example of a deterministic protocol.

4.0 CONCLUSION

This unit begins by defining medium access control and we identified two categories of medium access control: contention based and round-robin. In the last part of this unit, we described LAN system. Example of LAN system include: Ethernet, IBM Token Ring, FDDI, and wireless network.

5.0 SUMMARY

For a workstation to place data onto a local area network, the network must have a medium access control protocol. The two basic forms of medium access control protocols are contention-based (such as CSMAlCD, which is found on star-wired bus local area networks, and CSMA/CA, which is found on wireless local area networks), and round-robin (such as token passing or token ring). CSMAlCD works on a first-come, first-served basis, supports half-duplex and full-duplex connections, and is clearly the most popular access protocol, but it suffers from collisions of data frames during high-usage periods (when hubs are employed). Round-robin protocols are good under heavy loads but require more software support. To standardize the medium access control protocols, IEEE created the 802 series of network standards. The most popular types of local area network systems are Ethernet (CSMA/ CD) and, wireless Ethernet. Ethernet LANs have the most variations of products and continue to dominate the local area network market.

6.0 TUTOR-MARKED ASSIGNMENT

(a) Explain the difference between 1000BaseSX and 1000BaseLX
(b) What are the disadvantages unique to Ethernet?

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network*, A Business User's Approach Fourth Edition: Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks: Kluwer Academic Publisher, USA.

UNIT 5 LOCAL AREA NETWORKS: INTERNET WORKING

CONTENTS

1.0 INTRODUCTION

Unit 8 introduced local area networks by discussing their primary functions, typical activities ad applications, advantages and disadvantages, basic topologies, while unit 9 explained medium access control techniques and the most common local area networks systems. This unit introduces an exciting and often confusing topic: internetworking of local area networks to other local area networks, and of local area networks to wide area networks. This unit will examine four devices: the hub, the bridge, the switch, and the router.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- List the reasons for interconnecting multiple LANs segments and interconnecting LANs and WANs.
- Identify the functions and purposes of the various interconnection devices that have been used over time.

3.0 MAIN CONTENTS

3.1 Reasons for Interconnections of LANs

Suppose a company has a sufficiently large LAN that is extremely busy, with high demands on its file server and printer. The response time for retrieving files from the server is very slow, and employees have to wait long periods of time for a print job. The company would like to divide the network into two LAN segments, thus dividing the traffic and workload between two different networks. Once this division is made, the response times should improve, and wait times for print jobs should drop to an acceptable level. As the preceding example demonstrates, a single internal local area network is often not sufficient to support the needs of its users. The network may not be capable of supporting a large number of users, or it may not be able to provide users access to the resources they need, because those resources are located on a different internal local area network or on an external wide area network. In these situations, having multiple local area networks or a local area network that has access to a wide area network might provide better service. If such access to multiple local area networks or to a wide area network is desired, steps must be taken to interconnect the multiple networks. Interconnecting multiple networks or multiple segments of networks is called internetworking, and the large diversity of networks in existence makes this a complex process. Breaking a large network into smaller networks is called segmentation. Given all this complexity, why would anyone want to connect two or more networks? There are, in fact, several valid reasons for internetworking. Consider, for example, a company that has two local area networks, one for the research department and one for the design department. If the company could interconnect the two networks, it would be possible for the employees in research to share date and resources with the employees in design, and vice versa. In addition, multiple repositories of data could be created, to which both research and design have access, to which only research has access, and to which only design has access. Another company might want to interconnect a local area network and a wide area network, such as the Internet. With this interconnection, users of the local area network could gain access to the World Wide Web, external electronic mail, and remote logins, and perform file uploads and downloads to any location in the world.

Consider yet another company in which the computer workstations are spread across a number of adjacent buildings. To, enable management to better manage the workload demand that is distributed across multiple buildings, each building could house at least one local area network, and all the buildings could be interconnected. This internetworking of small LANs could be implemented on floors within a building or even within divisions on a floor within a building. Suppose a company maintains a mainframe computer that processes a 'number of crucial computer programs or legacy applications (older programs that have been used by the company for many years). If the company internet-worked a local

area network to its mainframe, it could provide all workstation users with access to the applications on the mainframe. These programs could then be transmitted from the mainframe to the workstation (downloaded) and executed on the user workstation. This downloading relieves the amount of work on the mainframe computer. Internetworking isn't just for large corporations. Suppose you have two or more personal computers in your home for family use. Suppose also that you have installed a Small Office/Home Office network so that the computers can share data, software, and peripherals such as a high-quality printer. The next step would be to provide the home network with an interface to the Internet so that all the home computers could access Internet resources. Many devices are available for performing the interconnection of two or more networks. We will classify these devices into four basic categories: hubs, bridges, switches, and routers. Let's discuss each of these devices in turn and examine examples that demonstrate their use.

3.2 Hubs

Creating a local area network does not simply involve running a cable from one workstation to another. In almost all cases, you need at least one device that acts as a collection point of the cables. Some people think that any device in a local area network' - that is a collection point for the cables coming from the workstations is a hub. Thus, bridges, switches, MAUs (Multi station Access Units), and routers might all be considered hubs. But not all professionals in the field agree with this definition. To avoid confusion, this definition will not be used here. Instead, let's define a hub as a device that interconnects two or more workstations in a star-wired bus local area network and broadcasts incoming data onto all outgoing connections (in other words, no routing or selective forwarding is performed). It is also important to note that a hub is a device that works at the physical layer. The hub simply accepts a frame and immediately forwards it. It does not examine or modify the contents of the frame.

Although hubs come in many configurations with many different types of options, most hubs can be sorted into one of two categories: managed hubs and unmanaged hubs. A managed hub possesses enough processing power to be managed from a remote location. The management operations performed on a hub include inventory management (knowing what devices are where on the network), traffic and environmental monitoring, and power management An unmanaged hub contains little or no intelligence at all and cannot be controlled from a remote location. Unmanaged hubs are less expensive than managed hubs, but they cannot participate in any kind of network management operations. Their sole function is to allow the interconnection of two or

more workstations in a local area network. One characteristic that all hubs seem to continue to share is their low cost.

3.3 Bridges

A hub is a simple device that requires virtually no overhead to operate. But it is also inefficient. When a network is experiencing a high level of traffic, a hub compounds the problem by taking any incoming frame and retransmitting it out to all connections. In contrast, the next interconnection device we'll be looking at, the bridge, can use processing power to direct a frame out a particular port, thus reducing the amount of traffic on the network. Originally, a bridge was a device that interconnected two local area networks that either used the same protocol or had two different protocols. Now, equipment catalogs show bridges as specialized devices that can interconnect two dissimilar local area networks. Let's define a bridge as a device that interconnects two local area networks that both have a medium access control sub layer. Although the bridge, in most applications, has been replaced by the switch, the bridge is a simple form of the switch and thus serves as a good starting point for learning about switches. When interconnecting two local area networks that have medium access control sub layers, the basic bridge has one primary function-if a data frame originates on one network and is destined to arrive at a workstation on the same network, the bridge needs to realize this destination information and not forward the data frame to a second network. Thus, the bridge acts as a filter. A filter examines the destination address of a frame and either forwards or does not forward the frame, depending on some address information stored within the bridge. The bridge will reduce the amount of traffic on the interconnected networks by dropping frames that do not have to be forwarded. Let's examine this filtering function a little more closely (Figure 10-1). As a frame of data moves across the first CSMAlCD network and enters the bridge, the bridge examines the medium access control sub layer source and destination addresses. A MAC sub layer address is the address assigned to the network interface card (NIC) when the NIC is manufactured. (All companies that produce NICs have agreed to use a formula that ensures that every NIC in the world has a unique MAC address.)

Figure 8-1
A bridge interconnecting two CSMA/CD local area networks

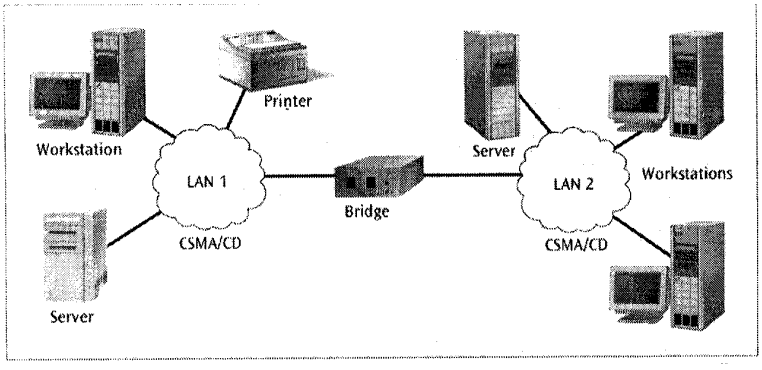


Figure 10.1: A Bridge interconnecting two CSMA/CD LANs

The bridge, using some form of internal logic, determines if a data frame's destination address belongs to a workstation on the current network. If it does, the bridge does nothing more with the frame, because it is already on the appropriate network. If the destination address is not an address on the current network, the bridge passes the frame on to the next CSMA/CD network, assuming that the frame is intended for a station on that network. Additionally, the bridge can check for transmission errors in the data by performing a cyclic checksum computation. The question is how does the bridge know what addresses are on which networks? Did a technician sit down and type the address of every NIC on each interconnected network? Not likely. Most bridges (and switches) are transparent, which means they earn by themselves. Let's examine the transparent bridge more closely.

3.3.1 Transparent Bridge

The transparent bridge is designed for CSMA/CD LANs. The transparent bridge observes network traffic flow and uses this information to make future decisions regarding frame forwarding. Upon installation, the bridge begins observing the addresses of the frames in transmission on the current network and "'creates an internal port table to be used for making future routing decisions. The bridge creates the internal port table by using a form of backward learning that is, by observing the location from which a frame has come. If a frame is on the current network, the bridge assumes that the frame originated from somewhere on that network. The bridge takes the source address from the frame and places it into an internal table. After watching traffic for a while, the bridge has a table of workstation addresses for that network. If a frame arrives at the bridge with a destination address that does not match any address in the table, the bridge assumes the frame is intended for a workstation on some other network and passes the frame on to the next network. For an example of how the transparent bridge learns, examine Figure 10-2 and the following scenario. The bridge has two ports, one for CSMA/CD LAN A and the second for CSMA/CD LAN B. When the bridge is first activated, its internal port tables, one for Port A and one for Port B, are empty. Figure 10.3(a) shows the two

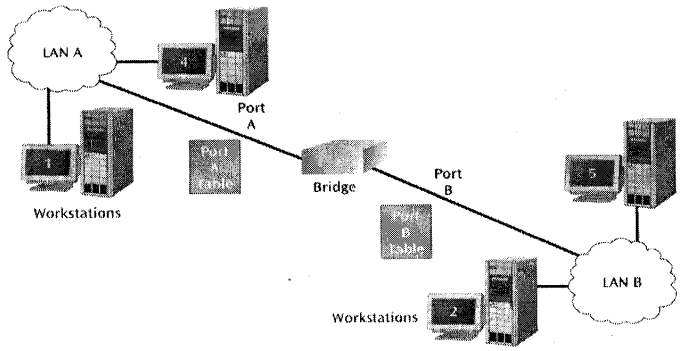


Figure 10.2: An example of a Transparent Bridge

tables as being initially empty. Now suppose a workstation 1 transmits a frame intended for Workstation 4. Because CSMA/CD is a broadcast network, the frame goes to all devices on the network, including the bridge. The bridge extracts workstation's address and puts it in Port A's table. It has just learned that Workstation 1 is on LAN A, as shown in Figure 10-3(b). The bridge still doesn't know the address of workstation 4, however. Even though Workstation 4 is also on LAN A, the bridge does not know this fact, because there is no entry from an incoming frame in the Port A table for Workstation 4. Consequently, the bridge unnecessarily forwards the frame out Port B onto LAN B.

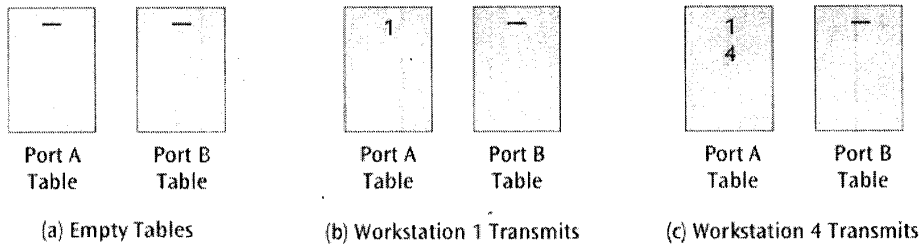


Figure 10.3: The two internal port tables and their new entries

Now suppose Workstation 4 returns a frame to Workstation 1. The bridge extracts the address of Workstation 4 and places it in the Port A table, as shown in Figure 1 0.3(c). The frame is destined for Workstation 1, and the bridge sees that there is an entry for workstation 1 in Port A's table. Now, the bridge knows Workstation 1 is on LAN A and does not forward the frame on to LAN B. In addition, if Workstation 1 sends another frame to workstation 4, the bridge will see that workstation 4 is on LAN A (because of the entry in the Port A table) and will not forward the frame on to LAN B. If workstation 1 sends a frame to workstation 5, the bridge will not recognize the address of Workstation 5, because there is no entry in Port A's table! and it will forward the frame on to LAN B. The bridge will perform the same learning function for LAN B and update Port B's table accordingly. Thus, the bridge learns where workstations are and then uses that information for future forwarding decisions. As we will see shortly, the switch learns in much the same manner as the transparent bridge. As traffic enters on each port, a table is updated to reflect the source address of the received frame. Later, when a frame is to be transmitted to another workstation, this table of forwarding addresses is consulted and the frame is sent out the optimal port.

3.3.2 Remote Bridge

Another form of bridge is the remote bridge. A remote bridge is capable of passing a data frame from one local area network to another when the two local area networks are separated by a long distance and there is a wide area network connecting them. For a remote bridge to pass a data frame over a wide area network, the frame must be converted into the proper form required for traversal over the intervening wide area network. In Figure 10.4, a frame (LAN X Packet) that leaves workstation 1 destined for the server on LAN Y travels across LAN X and arrives at Bridge A. Remote Bridge A realizes that the frame is not intended for a workstation on LAN X and prepares to forward the frame. But to reach the server, the frame must traverse a frame relay wide area network. The remote bridge adds frame relay header information in front of the existing LAN frame. The frame relay header information is used to get the frame across the frame relay wide area network. Once the frame reaches remote Bridge B, the frame relay header is removed, leaving the original LAN frame. The original LAN frame then travels across LAN Y to the destination server.

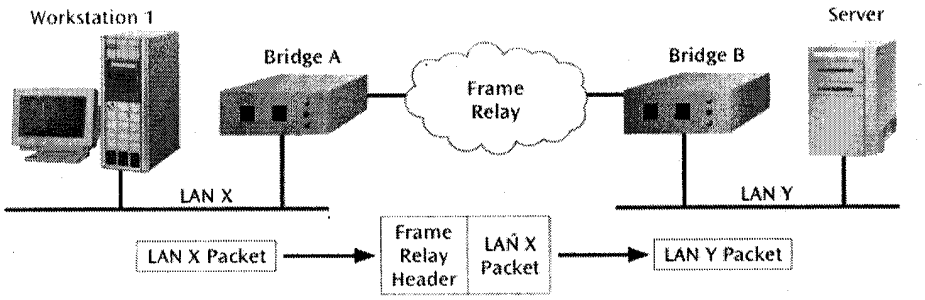


Figure 10.4: Two LANs with intervening frame relay network

Because of the capability of a remote bridge to transfer data from local area networks across a wide area network, the remote bridge is a more complicated and thus more expensive device than a transparent bridge. Selecting and installing a pair of remote bridges is even more complicated and difficult, because there are many different types of local area networks (10BaseT, 100BaseFX, 1000BaseSX, and 10Gbps, to name a few) and many different types of wide area networks (such as frame relay and Asynchronous Transfer Mode). In addition, a unique remote bridge must be used for each local area network-wide area network combination. For example, if you want to connect two 100BaseFX local area networks with a frame relay wide area network, you will need to obtain the appropriate 100BaseFX/frame relay remote bridge. We now have an understanding of the basic device that can be used to interconnect local area networks: the bridge. But the bridge has, to a high degree been replaced with a more sophisticated device: the switch. Let's examine the switch next and see how it shares many characteristics of both the bridge and the hub.

3.4 Switches

A switch is a combination of a hub and a bridge. It can interconnect multiple workstations (like a hub) but can also filter out frames, thereby providing a segmentation of the network (like abridge). Although a switch is very similar to a bridge functionally in that it acts as a filter, the switch operates in place of a hub. Stated another way, the switch is a multi-port bridge that can have as many ports (connections) as a hub. Switches can significantly decrease interconnection traffic and increase the throughput of interconnected networks or segments, without requiring additional cabling or rearranging of the network devices. As you may recall, CSMAlCD networks experience collisions. By reducing the number of unnecessarily transmitted packets, a switch can cause the number of collisions to decline. As the number of collisions declines, the overall throughput of the network should increase. Despite all the similarities between a switch and a bridge, there are a number of significant differences. First, a switch is a more powerful device than a bridge (and a hub). A single switch can be designed to support multiple I-,AN technologies. A switch can also be designed to provide redundant circuits to avoid hardware failures and has the capability of automatically switching to a backup data circuit controller, or management module. These controller and management modules support the operations and control of the switch.

Another important advantage of a switch is that it is designed to perform much faster than a bridge, especially switches that use cut-through architecture. In a cut-through architecture, the data frame begins to exit the switch almost as soon as it begins to enter the switch. In other words, a cut-through switch does not store a data frame and then forward it. In contrast, a store-and-forward device holds the entire frame for a small amount of time while various fields of the frame are examined, a procedure that diminishes the overall network throughput. The cut-through capability allows a switch to pass data frames very quickly, thus improving the overall network throughput. The major disadvantage of cut- through architecture is the potential for the device to forward faulty frames. For example, if a frame has been corrupted, a store-and-forward device will input the frame, perform a cyclic checksum, detect the error, and perform some form of error control. A cut-through device, however, is so fast that it begins forwarding the frame before the cyclic checksum field can be calculated. If there is a cyclic check- sum error, it is too late to do anything about it. The frame has already been transmitted. If too many corrupted frames are passed around the network, network integrity suffers.

Finally, a switch can have several ports, unlike a bridge, which has only two ports. To support each of these ports efficiently, the main hardware of the switch- called the backplane-has to be fast enough to support the aggregate or total bandwidth of all the ports. For example, if a switch has eight 100-Mbps ports, the backplane has to support a total of 800 Mbps. This backplane is similar to a bus inside a microcomputer. It allows you to plug in one or more printed circuit cards. Each circuit card supports one port, or connection, to a workstation or other device. If the circuit cards are hot swappable, it is possible to insert and remove cards while the power to the unit is still on. This capability allows for quick and easy maintenance of the switch. The switch has one physical similarity to the hub. If you decide to install a switch into a local area network, in many cases this is as simple as unplugging a hub and plugging a switch in the hub's place. Logically, however, the switch is not like the hub. The switch can examine frame addresses and, based on the contents of an address, direct the frame out the one appropriate path. Whereas the hub simply blasts a copy of the frame out all connections, the switch (like the bridge) uses intelligence to determine the one best connection for outgoing transmission. Depending on user requirements, a switch can interconnect two different types of CSMAlCD network segments: shared segments and dedicated segments. In shared segment networks, as shown in Figure 10.5, a switch may be connected to a hub (or several hubs), which then connects multiple workstations. Because the workstations are first connected to a hub, they all share the one channel, or band-width, of the hub, which limits the transfer speeds of individual stations. In dedicated segment networks, a switch may be directly connected to a workstation, and the switch connects to the hub. Each workstation then has a private or dedicated connection. This dedicated connection increases the bandwidth for each workstation over what the bandwidth would be if the workstation were connected to the hub. Dedicated segments are useful for more powerful workstations with high communication demands.

One of the more interesting applications for a dedicated segment network and a switch is creating a virtual LAN. A virtual LAN, or VLAN, is a logical subgroup within a local area network that is created via switches and software rather than by manually moving wiring from one network device to another. For example, if a company wishes to create a workgroup of employees to work on a new project, network support personnel can create a VLAN for that workgroup.

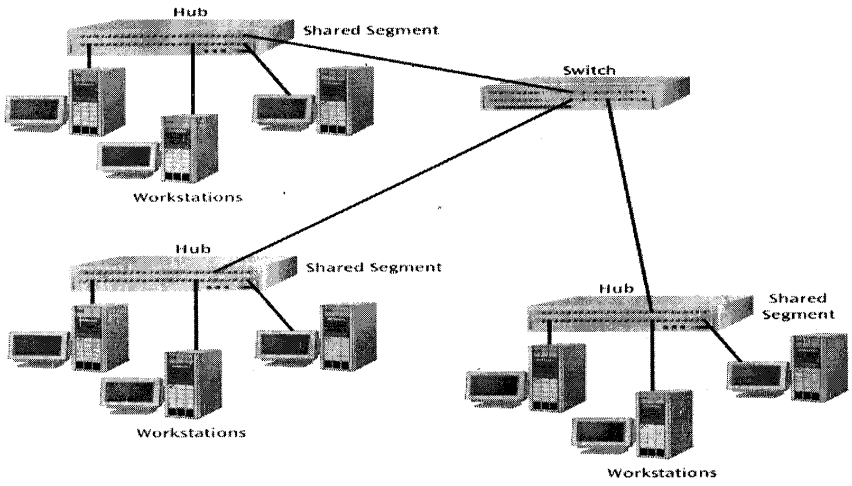


Figure 10.5: Workstations, connected to a shored segment local area network

Whether shared or dedicated segments are involved, the primary goal of a switch is to isolate a particular pattern of traffic from other patterns of traffic or from the remainder of the network. Consider a situation in which two servers, along with a number of workstations, are connected to a switch (Figure 10-6). If Workstation A wishes to transmit to Server 1, the switch forwards the data packet/frame directly

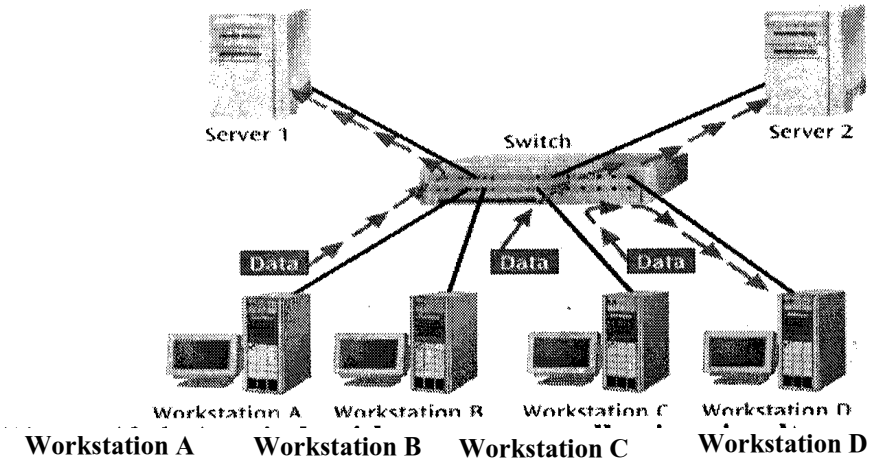


Figure 10.6: A switch with two servers allowing simultaneous access to each server

to Server 1 and to nowhere else on the network. Workstations B, C, and D do not receive a data frame from Workstation A. Furthermore, Workstation A can transmit to Server 1 at the same time that Workstation B transmits to Server 2. Finally, the switch can accommodate a high degree of intercommunication between the two

servers without sending the data to any workstations on the network. Because many local area networks have a high degree of inter-server communication, this use of a switch can effectively reduce overall network traffic.

All of the network devices introduced thus far-hubs, bridges, and switches are useful and sometimes vital to the proper support of a local area network. One device that has been mentioned numerous times and is essential if we want to interconnect our local area network with a wide area network is the router.

SELF-ASSESSMENT EXERCISE

i. State the reasons for interconnecting.
ii. Explain the concept of cut-through architecture and mention the disadvantage(s) of using this architecture to interconnect devices in network environment.

3.5 Routers

The most common function of a router is- to transmit data frames between two networks, one of which has a medium access control sublayer, while the second network does not. A common use of a router is to interconnect a local area network, such as a CSMA/CD LAN, with the Internet. The Internet, as you will see in Unit 14, does not use the 48-bit MAC address to find workstations or devices, but is based on the TCP/IP protocol suite and IP addressing instead. In other words, to perform the routing, the router cannot just look at the MAC-layer addresses. Instead, the router must dig further into the data frames for the IP address. More specifically, it examines the network-layer addresses and uses those to perform the routing. The byproduct of digging deeper into the data frames for network routing information is increased processing time, which in part makes the router a slower and more elaborate device than a bridge. What sequence of events occurs when a CSMA/CD data frame is converted by a router to an Internet data frame? Figure 10.7 shows a data frame passing from a CSMA/CD LAN to the Internet. A data packet originates in an upper-layer application, such as an e-mail program, as can be seen in Figure 10.7(a). In Figure 10.7(b), the data packet is passed to the transport and network layers of the CSMA/CD LAN, and the appropriate transport and network headers are inserted.

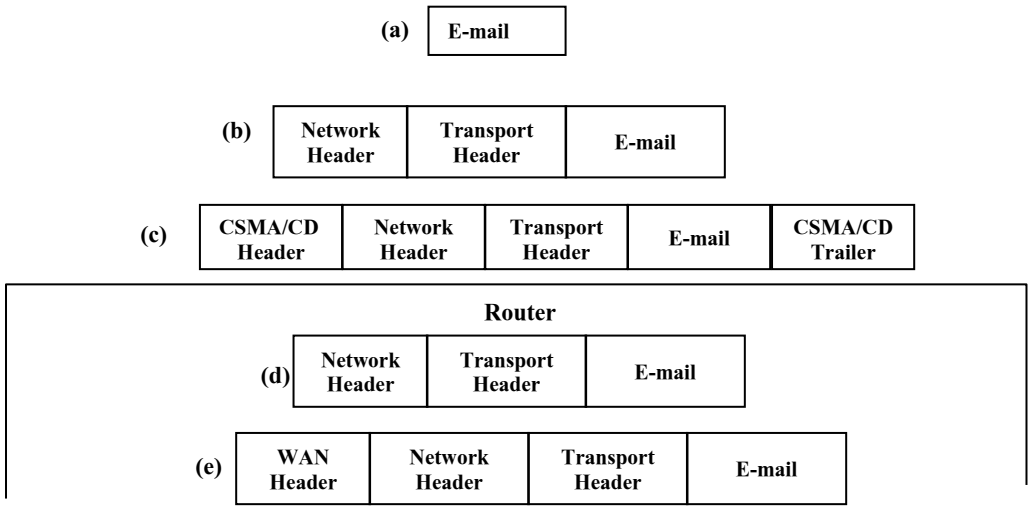


Figure 10.7: A data frame as it passes from CSMA/CD LAN to the Internet

After the network information has been inserted, the packet is given to the medium access control sub layer, where CSMA/CD header and trailer information is added (including the NIC address of the router), as shown in Figure 10. 7(c). The packet is then placed onto the medium of the CSMA/CD network, where it travels to the router. Upon receiving the packet, the router does not need the MAC header and trailer, so they are removed, as shown in Figure 10. 7(d). Now, the router will extract the destination network address from the network header and determine that the packet has to go out on the Internet. The router then adds the appropriate wide area network header (if any is needed) to the packet, as shown in Figure 10. 7(e), and sends the packet out of the router and onto the Internet. From there, the packet will make its way through the Internet to its final destination. As you can see from Figure 8-14, the router has to discard the MAC header and trailer information to get at the network header information, which is used for routing across a wide area network. In contrast, a bridge or a switch extracts only the MAC layer information and does not have to dig any deeper than that layer. Routers are similar to switches in that both are getting more advanced every day. A common feature found in many routers is some form of firewall protection. A firewall is a system or combination of systems that supports an access control policy between two networks. In the case of a router, a firewall acts as a protection system between the local area network and the wide area network. Some of the newer routers also contain switches, thus combining the services of a router with the advantages of a switch. Modem routers can also accept data in one format and convert the data to another format. For example, when a data packet comes in from the Internet in Internet form (TCP/IP), the

router can convert the packet to an appropriate form for traversal over the local area network. The more advanced routers can also perform management functions such as monitoring network traffic, providing accounting information, and incorporating quality of service functions.

Quality of service (QoS) is the concept that data transmission rates, error rates, and other network traffic characteristics can be measured, improved, and (one hopes) guaranteed in advance. Thus, if a user wishes to use an application that incorporates a high bandwidth transmission, such as real- time video, a network connection that contains the appropriate QoS parameters is requested. The network examines the request and, based on current network demands and network limitations, either guarantees or does not guarantee the connection.

ANSWER TO SELF-ASSESSMENT EXERCISE

Reasons for Interconnecting

- To separate / connect one corporate division with another.
- To connect two LANs with different protocols.
- To connect a LAN to the Internet.
- To break a LAN into segments to relieve traffic congestion.
- To provide a security wall between two different types of users.

Cut-through architecture

The cut-through capability allows a switch to pass data frames very quickly, thus improving the overall network throughput. The major disadvantage of cut- through architecture is the potential for the device to forward faulty frames.

A virtual LAN

A virtual LAN, or VLAN, is a logical subgroup within a local area network that is created via switches and software rather than by manually moving wiring from one network device to another.

4.0 CONCLUSION

This unit examined into detail four devices: the hub, the bridge, the switch, and the router used to interconnect multiple LANs segments and interconnecting LANs and WANs. The unit also highlighted the functions and purposes of the various interconnection devices that have been used over time.

5.0 SUMMARY

As local area networks grow in size, and as the need to connect a local area network to other local area networks and to wide area networks arises, it is necessary to understand how to interconnect similar and dissimilar networks. The reasons for interconnecting networks within a business include the need to interconnect two local area networks from different departments or different geographic locations, to provide local area network users with access to the Internet or a legacy system, to segment a network that is growing too quickly, to provide a level of security between two application groups, and to provide a means of sharing software and peripherals. The reasons for interconnecting networks within a home include the need to share software, peripherals, and Internet connections.

Four devices can provide varying levels of interconnection: hubs, bridges, switches, and routers. A hub is a device that interconnects multiple workstations within a local area network and, like most interconnection devices, can be either managed or unmanaged.

Bridges are relatively simple devices that could interconnect two (typically identical) networks at the medium access control sublayer. A bridge acted as a filter as it examined each frame and decided whether the frame should be forwarded on to the next network. Some bridges could also convert one medium access control format to another medium access control format. A transparent bridge builds its own routing tables by observing the flow of traffic on the networks, a process that is referred to as backward learning. A remote bridge connects two local area networks that are separated by large distances and a wide area network. As the data frame leaves one local area network, the remote bridge adds the necessary control information for the frame to traverse the wide area network.

Although most bridges and many hubs are being replaced with switches, the operation of a switch is similar to the basic operation of a bridge. A switch can provide a significant decrease in interconnection traffic and increase the throughput of inter- connected networks, while still using conventional cabling and adapters. A switch replaces a hub and isolates the traffic flow between segments of the network by examining the address of the transmitted frame and directing the frame to the appropriate port. A switch that employs a cut-through architecture is the opposite of a store-and-for- ward device, in that the data frame is leaving the switch almost as soon as it begins to enter the switch. Switches can create shared segments in which all workstations hear all the traffic or dedicated segments in which other workstations do not hear the local traffic. Routers interconnect local area networks with wide area

networks. A router's most common function is to route data packets between two networks, one of which uses the addresses in the medium access control sublayer, while the other uses the addresses in a layer other than the medium access control sublayer. Routers provides all users on a local area network with access to outside networks. Routers perform more slowly than bridges and require more processing, because they have to dig deeper into the data frame for control information.

6.0 TUTOR-MARKED ASSIGNMENT

1. What does it mean when a switch or device is cut-through? What is the main disadvantage of a cut-through switch? Is there a way to solve the disadvantage of a cut-through switch without losing its advantages? Defend your answer.

2. Distinguish between a shared segment network and a dedicated segment network.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.

MODULE 3

UNIT 1 INTRODUCTION TO MANs AND WANs

CONTENTS

1.0 INTRODUCTION

A local area network, as you may recall, is typically confined to a single building or set of buildings that are in close proximity (as in a campus). What happens when a network expands into a metropolitan area, across & state, or across the entire country? A network that expands into a metropolitan area and exhibits high data rates, high reliability, and low data loss is called a metropolitan area network (MAN). What happens when a network is larger than a metropolitan area? A network that expands beyond a metropolitan area is a wide area network. Wide area networks share a few characteristics with local area networks: they interconnect computers, use some form of medium for the interconnection, and support network applications.

In this unit, we will examine metropolitan area networks and wide area networks and see how they compare and contrast to other forms of networks.

2.0 OBJECTIVES

At the end of this Unit, you should be able to:

- Distinguish LAN s, MANs, and WAN s
- Identify the characteristics of MANs, and explain how they compare and contrast with W ANs and LANs.
- Describe how circuit-switched, datagram packet-switched and virtual circuit packet-switched networks work.
- Identify the differences between a connection-oriented network and a
- Connectionless, and give an example of each.

3.0 MAIN CONTENT

3.1 Metropolitan Area Network (MAN) Basics

Many of the same technologies and communications protocols found in local area networks (and wide area networks) are used to create metropolitan area networks. Yet MANs are often unique with respect to topology and operating characteristics. MANs can be used to support high-speed disaster recovery systems and real-time transaction backup systems. They can also provide interconnections between corporate data centers and Internet service providers, and support high-speed connections among government, business, medical, and educational facilities. MANs are almost exclusively fiber-optic networks, and thus capable of supporting data rates into the tens of millions and hundreds of millions of bits per second. For the same reason, they are advertised as networks with very low error rates and extremely high throughput. Although these characteristics are not that different from those of many local area networks, a few characteristics distinguish MANs from LANs. The first characteristic is that MANs cover much greater distances than LANs do. As the name implies, metropolitan area networks are quite capable of supporting entire metropolitan areas, such as Lagos, Ibadan, and lie-Ife. Local area networks rarely extend beyond the walls of a single building; and thus are smaller than MANs. A second characteristic that distinguishes MANs from LANs (but not necessarily from W ANs) is that most MANs can recover very quickly from a link or switch/router failure. MANs are designed to have highly redundant circuits so that in the event of a component failure, the network can quickly reroute traffic away from the failed component. This ability to reroute in the event of a failure is called failover, and the speed at which a failover is performed is the failover time. While not all MANs have low failover times, achieving them is certainly the goal of any company that offers a MAN service. A third characteristic that distinguishes many MANs from both LANs and WANs is that some

MAN topologies are based on a ring. Last, a feature that is beginning to appear in MANs but that neither LANs nor WANs currently have is the ability of a user to dynamically allocate more bandwidth on demand. Now that we have examined some of the basic characteristics that distinguish MANs from LANs and WANs, let's turn our attention to the technologies that support MANs.

3.1.1 SON ET vs. Ethernet

Almost all MANs are based on one of two basic forms of supporting technology: SONET or Ethernet. SONET, as we saw in the previous unit, is a synchronous time division multiplexing technique that is capable of sending data at hundreds of mil lions of bits per second. The network topology is a ring, but this ring is actually composed of multiple rings that enable the network to provide backup in the event of a segment failure (Figure 11.1). This is one of the characteristic of SONET rings that allows them to have a very low failover me. At the present time, most MANs are supported by SONET ring technology.

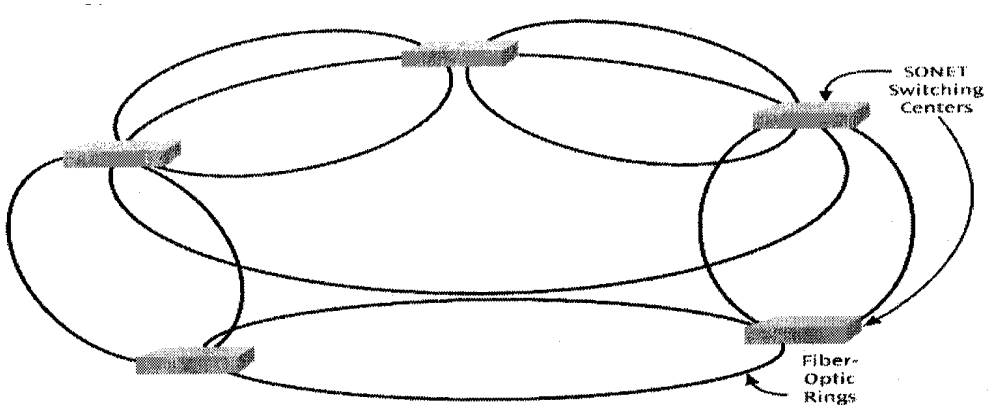


Figure 11.1: SONET systems

Unfortunately, SONET has a number of disadvantages. It is a complex, fairly expensive technology that cannot be provisioned dynamically. Furthermore, SONET was designed to support multiple streams of voice channels (such as multiple T -Is, which transmit at 1.544 Mbps). Ethernet MANs are less expensive than SONET systems, well understood, easily scalable from 10 Mbps to 100 Mbps to 1000 Mbps to 10 Gbps, and the best technology for carrying IP traffic (the type of traffic that runs over the Internet).

One disadvantage of Ethernet in the MAN is higher failover time. Ethernet MANs do not recover as quickly as SONET rings and can potentially leave customers without service for seconds. Although SONET rings typically have a failover time of 50 milliseconds, Ethernet

failover times can be multiple seconds. Nonetheless, Ethernet MANs have a number of attractive characteristics and are growing in popularity.

Ethernet MANs have given rise to a newer service whose popularity has grown in the last few years: Metro Ethernet. Metro Ethernet is a data transfer service that can connect your business to another business (or businesses) using a standard Ethernet connection. With Metro Ethernet, you may connect your company directly to another company using a point-to-point connection, or, for example, to two other companies using two point-to-point connections, as is shown in Figure 11.2 (a). Alternatively, you may connect your company to multiple 'Companies. as though they were all part of a large local area network, as shown in Figure 11-2(b).

Now that we have examined MAN basics, let's look at the basics of wide area networks.

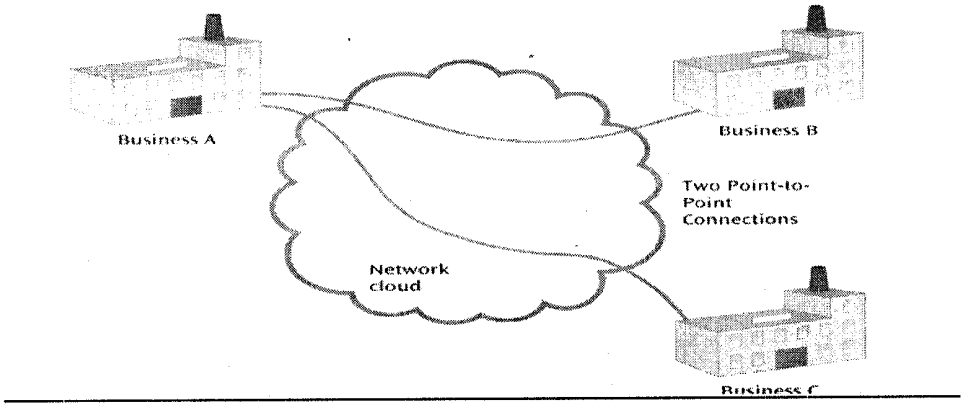


Figure 11.2(a): Two point-to-point connections

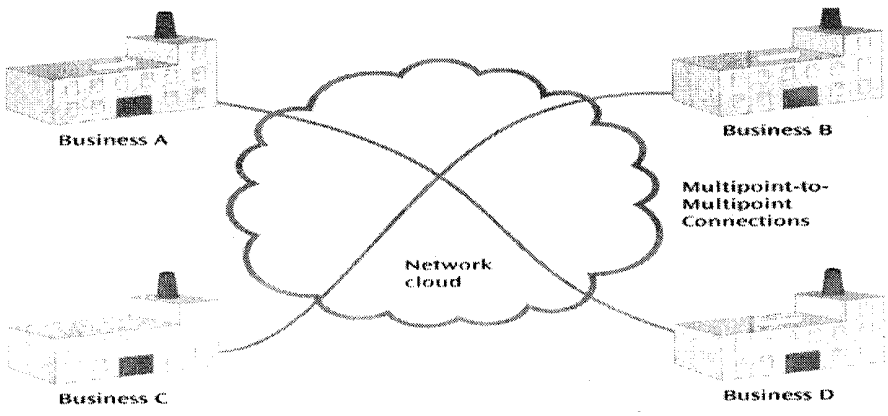


Figure 11.2(b): Multipoint-to-Multipoint connections

3.2 Wide Area Network Basics

A wide area network (WAN) is a collection of computers and computer-related equipment interconnected to perform a given function or functions and typically using local and long-distance telecommunications systems. The types of computers used within a wide area network range from microcomputers to mainframes. The telecommunications lines can be as simple as a standard telephone line or as advanced as a satellite system. Wide area networks are typically used to transfer bulk data between two endpoints and provide users with electronic mail services, access to database systems, and access to the Internet. Wide area networks can also assist with specialized operations in many fields, such as manufacturing, medicine, navigation, education, entertainment, and telecommunications. As you may recall, a local area network works as a bus-based network in that clusters of workstations are connected to a central point (hub or switch) through which workstations can transmit messages to one another. Because there are so many workstations in a wide area network and they are spread over large (possibly very large) distances, this type of interconnection is not feasible. Likewise, a network in which each workstation is connected to every other network workstation is also impractical, as there would be so many connections into each workstation that the technology would be totally unmanageable. Instead, a wide area network connects its workstations through the use of a mesh design and requires routing to transfer data across the network. A network that is connected in a mesh is one in which neighbours are connected only to neighbours (Figure 11-3). Thus, to be transmitted across a mesh network, the data has to be passed along a route from workstation to workstation.

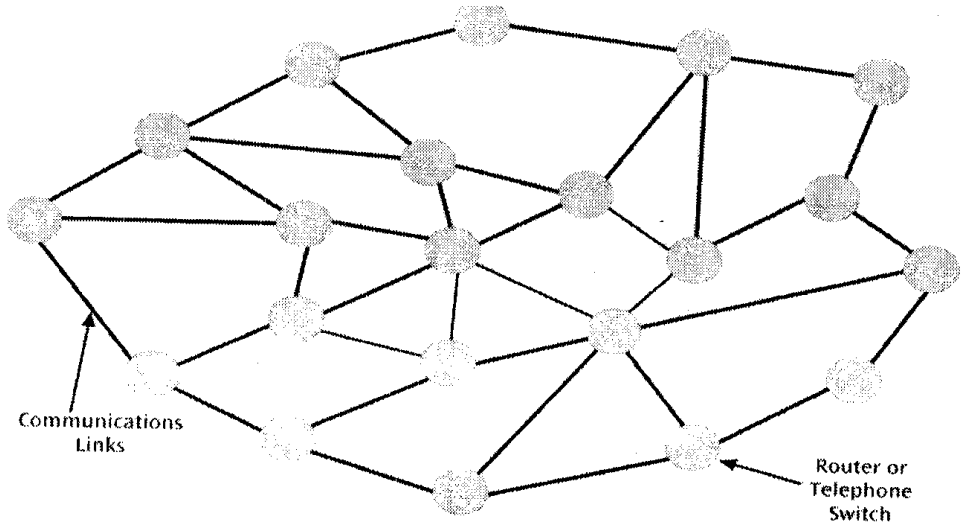


Figure 11.3: A simple mesh network

All wide area networks are collections of at least two basic types of equipment: a station and a node. A station is a device that interfaces a

user to a network, while a node is a device that allows one or more stations to access the physical network and is a transfer point for passing information through a network. A node is often a computer, a router, or a telephone switch. The support structure of a wide area network is the sub-network. A sub-network is a collection of nodes and interconnecting telecommunication links, as shown in Figure 11.4. The sub-network is responsible for getting the data to the proper destination node, which then delivers it to the appropriate destination station. Clearly, a network would not exist without a sub-network, The sub-network is simply the vehicle for getting the data from sender to destination.

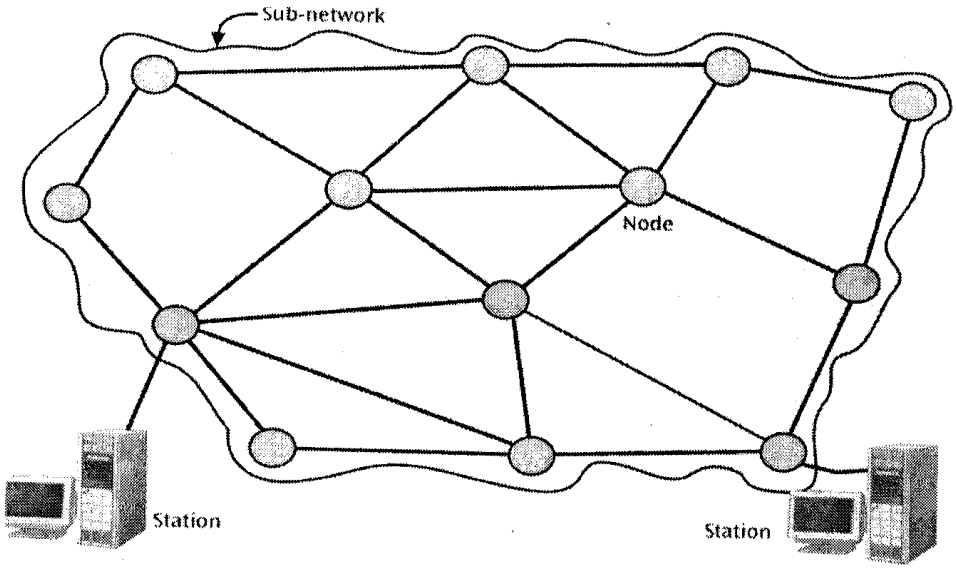


Figure 11.4: Sub-network, nodes, and two end stations

The topics introduced in the rest of this unit and unit 12 -sub-networks, routing, and congestion-are handled by the network layer of a network architecture model, regardless of whether the model is OSI or TCP/IP. Let's first examine the topic of sub-networks in detail, and then explore routing 'and congestion in Unit 12.

3.2.1 Types of Sub-networks

A wide area network's sub-network may be categorized by the way it transfers information from one end of the sub-network to the other. The three basic types of sub-networks are circuit-switched, packet-switched, and broadcast. For simplicity, we will omit the term "sub" and simply refer to these as circuit-switched networks, packet-switched networks, and broadcast networks.

Circuit-Switched Network

A circuit-switched network is a sub-network in which a dedicated circuit is established between sender and receiver, and all data passes over this

circuit. One of the best examples of a circuit-switched network is the dial-up telephone system. When someone places a call on a dial-up telephone network, a circuit, or path, is established between the person placing the call and the recipient of the call. This physical circuit is unique, or dedicated, to this one call and exists for the duration of the call. The information (the telephone conversation) follows this dedicated path from node to node within the network, as shown in Figure 11.5. A wide area network in which information follows a dedicated path from node to node within the network is a circuit-switched wide area network.

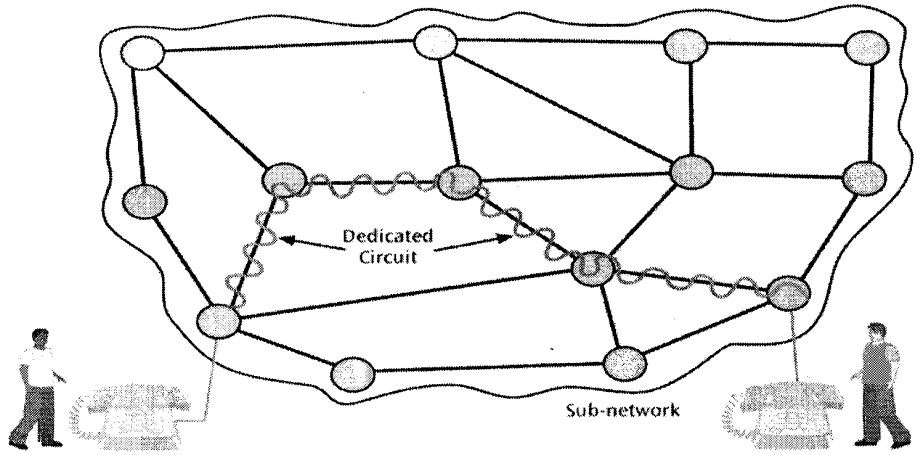


Figure 11.4: An example Circuit-switched network

When a telephone call is placed over a circuit-switched network, the network, needs time to establish the circuit and to tear down the circuit. But once the circuit is established, all subsequent data travels quickly from node to node. A circuit switched network has two key disadvantages. First, each circuit is dedicated to one connection. Second, when the circuit is used to transfer data (as oppose voice), it is probably not being utilized fully, because computer data transfer is often sporadic.

Packet-Switched Network

The packet-switched network is most often found in networks designed to transfer computer data (such as the Internet). In a packet-switched network, all data messages are transmitted using fixed-sized packages, called packets, and no unique dedicated physical path is established to transmit the data packets across the network. (To distinguish between a piece of data processed at the data link layer and a piece of data processed at the network layer, the term "frame" is used at data link layer, and the term "packet" is used at the network layer.) .If the message to be transferred is large, it is broken into multiple packets. When the multiple packets arrive at the destination, they are reassembled into the original message. The two types of packet-switched networks are the datagram and the virtual-circuit. In a

datagram packet-switched network, each data packet can follow its own, possibly unique, course through the sub-network. As each packet arrives node, a decision is made as to which path the packet will follow next. This dynamic decision making allows for great flexibility should the network experience congestion or failure. For example, if a group of data packets is currently being routed through node A on its way to node E and node A experiences router problems, the network can reroute the packets through B instead.

The problem with datagram networks is that when a large group of packets is addressed to the same destination, the network's nodes have to examine each packet individually and determine each packet's next path. This can lead to inefficiency, or simply to wasted time. To solve this problem, the virtual circuit packet-switched network was created. In a virtual circuit packet-switched network, all packets that belong to a logical connection can follow the same path through the network. For example, one station may want to transfer a large amount of data, such as the entire electronic contents of a book, across the network to another station. To accomplish this, a virtual circuit breaks the large amount of data into n packets and determines an optimal temporary path through the network. Each router along the path is then informed that it will be participating in a particular virtual circuit. When the data arrives with the address of that particular virtual circuit, the router simply sends the data out the router connection that is associated with that virtual circuit. When the data transfer is complete, the temporary path is dissolved (that is, each router tosses that virtual circuit information). This type of packet-switched network is called a virtual circuit because the path followed by the packets acts like a circuit but is not an actual, physical circuit like a telephone circuit. Although this type of sub-network sounds similar to a circuit-switched network in that all data follow a fixed path, it is substantially different.

The path in a virtual circuit packet-switched network exists only in the software and only when the network creates the necessary routing tables at the appropriate nodes. These routing tables are similar to the port tables used by switches to deter mine a path through a local area network. Another difference between a virtual circuit and a circuit-switched network is that the path in a virtual circuit network may be shared by other traffic. To appreciate the significance of this, recall that when you place a call on a circuit-switched network, you have a dedicated circuit from one end of the connection to the other. Thus, if you could see data as it travels through the wire in your circuit, you would see that the only data in this circuit is your data. In contrast, each user in a virtual circuit network is sharing one or more circuits with other users. If you could see into a wire in this circuit, you might see data from several users, or data from one user, then another, traveling

across the network. Thus, the various wires that constitute a circuit in a virtual circuit network are carrying data streams from multiple users, and it is the software that keeps each data stream separate from the others. To summarize, packet-switched networks break long computer messages into fixed-sized packets and are thus designed to support computer data transmissions. In a datagram packet-switched network, each packet is an entity by itself. When a large message has been broken into n packets, each packet in a datagram network enters a node, where a unique routing decision is made. Because there is no fixed circuit to follow, the network spends no time in creating a circuit. Time is spent, however, in determining each packet's route. Nevertheless, datagram networks are quite flexible, as they can react quickly to network changes. The other type of packet-switched network, the virtual circuit packet-switched network, is a marriage of the packet-switched network and the circuit-switched network. When a large message in this network is broken into multiple packets, all these packets follow the same path through the sub-network. This path is determined before the first packet is transmitted-an activity that requires circuit setup time-and a virtual circuit, with its internal routing tables, helps route the packets from source to destination.

The use of packet-switched networks in the transmission of telephone conversations involves digitizing voice data and converting it into packets. Once voice data is in packet form, it can be transferred over a packet-switched network along with data and video. This type of conversion will lead to more economical communications because all the forms of data being transmitted will be sharing one form of network.

Broadcast Network

As in the case of a node in the broadcast design of most local area networks, when a node on a wide area network broadcast network transmits its data, the data is received by all the other nodes. This form of wide area network sub-network is, at the moment, relatively rare. Some systems, however, are in existence. These systems use radio frequencies to broadcast data to all workstations and typically operate as radio broadcast networks in rural areas or in areas where there are many islands surrounded by large bodies of water. Some of the new wireless Internet access services such as Wi-Max are also based on a broadcast network, but they are more often considered to be metropolitan area networks than wide area networks. Because broadcast networks are not as common as circuit-switched and packet-switched networks, the remaining discussions in this unit will not include them. In summary, the physical design of a wide area network, or its sub-network, has three basic forms: circuit-switched, packet-switched (datagram and virtual circuit), and broadcast. The characteristics of these forms are

summarized in Table 11.1. Note how circuit-switched and virtual circuit networks require path setup time and cannot dynamically reroute packets should a network problem occur. It is also worth noting from Table 11.1 that the circuit-switched network was designed primarily for voice signals and is the only network that offers a dedicated path.

Table 11.1: Summary of sub-network characteristics

Characteristic	Circuit-Switched	Datagram Packet-Switched	Virtual Circuit Packet-Switched	Broadcast
Path setup time?	Yes	No	Yes	No
Routing decision for each packet?	No	Yes	No	Typically no routing
Dedicated path?	Yes	No	No	No
Can dynamically reroute if problems occur?	No	Yes	No	Typically no routing
Designed originally for data or voice?	Voice	Data	Data	Data
Connection dedicated to your transfer only?	Yes	No	No	No

Having examined the physical design of wide area networks, let's now turn our attention to their logical design.

3.2.2 Connection-oriented Vs. Connectionless Network Applications

The sub-network of a wide area network is the physical infrastructure and thus consists of nodes (routers or telephone switches) and various types of interconnecting media. What about the logical entity that operates over this physical infrastructure? This logical entity often takes the form of a software application. For example, if you are using an e-mail application to send a message to a friend across the country, the e-mail application is the logical entity that uses the network's physical infrastructure or sub-network to deliver the message. Many different types of applications are found on wide area networks, including e-mail, Web browsing, and other commercial applications. Let's categorize all the network applications, or logical entities, into two basic categories: connection-oriented applications and connectionless applications.

A **connection-oriented** network application, such as the one that performs a file transfer using FTP, provides some guarantee that information traveling through the network will not be lost and that the information packets will be delivered to the intended receiver in the same order in which they were transmitted. Thus, a connection-oriented

network application provides what is called a reliable service. To provide a reliable service, the network requires that a logical connection be established between the two endpoints. If necessary, connection negotiation is performed to help establish this connection. For example, consider the following scenario: A bank wants to transfer a large sum of money electronically to a second bank. The first bank creates a connection with the second bank. As part of establishing this connection, the two banks agree to transfer the funds using data encryption. After the first bank sends the transfer request, the second bank checks the request for accuracy and returns an acknowledgment to the first bank. The first bank will wait until the acknowledgment arrives before doing anything else. All packets transferred during this period are part of this connection and are acknowledged for accuracy. If this is the only electronic transfer, the first bank will say goodbye, and the second bank will acknowledge the goodbye. Note that the type of sub-network used in this transfer process is not an immediately relevant issue. The sub-network could have been circuit-switched or packet-switched. All the application required was that a reliable connection be used to transfer the funds. You can perform online banking requests (that is, use a connection-oriented network application) over a local area network (a packet-switched network) at work or school, just as you can perform them over a dial-up telephone connection (a circuit-switched network) from home.

A **connectionless** network application does not require a logical connection to be made before the transfer of data. Thus, a connectionless application does not guarantee the delivery of any information or data. Data may be lost, delayed, or even duplicated. No connection establishing or terminating procedures are followed, because there is no need to create a connection. Each packet is sent as a single entity and not as part of a network connection. A common example of a connectionless network application is Domain Name System (DNS), a program that converts a URL, such as www.oauife.edu.ng, into an IP address. When you request a Web page using its URL, no connection is created between you and the DNS system. You simply click the browser button, and DNS converts the URL of the request into an IP address and sends the request along. Connectionless applications do not negotiate a connection, and the transfer of data is rarely, if ever, acknowledged. Additionally, if you send a second URL request, it has no relationship (network-wise) to the first. As in the case of a connection-oriented network application, the underlying sub-network of a connectionless application is, again, not really an issue. It can be either a circuit-switched network or a packet-switched network. You can send a URL request from work or school over a local area network (packet-switched network), just the same as you can from home over a dial-up connection

(circuit-switched network). Another good example that illustrates the difference between connection-oriented and connectionless networks is the relationship between the telephone system and the postal service. When you call someone on the telephone, that person will answer the telephone if he or she is available. Once the telephone has been answered, a connection is established. The conversation follows, and when one person or the other has finished with the conversation, some sort of ending statement is issued, both parties hang up, and the connection is terminated (Figure 11.5). Thus, when you use the telephone system, you are using a connection-oriented network that provides a reliable service.

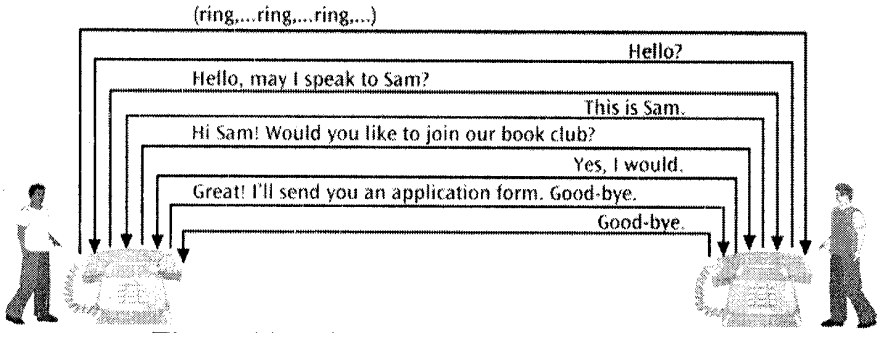


Figure 11.5: Connection-oriented telephone

In contrast, if you send a standard letter to someone through the Nigeria Postal Service, it will most likely be delivered, but there is no guarantee. There is also no guarantee of when it will arrive. If the letter is lost, you will not know until some time passes and you begin to think, "I haven't heard from Femi yet; did she receive my letter? The postal service, then, is similar to a connection less network that provides (no offense intended) an unreliable service. The unreliable service offered by the connectionless postal "network" requires that you take additional; actions if you want to ensure that the letter is delivered as intended.

SELF-ASSESSMENT EXERCISE

i. Differentiate between a network node and a network station.
ii. State the characteristics, advantages, and disadvantages of the following types of Network: Circuit-switched network and Packet-switched network
iii. How does a connectionless application differs from a connection-oriented application.

ANSWER TO SELF-ASSESSMENT EXERCISE

A station is a device that interfaces a user to a network, while a node is a device that allows one or more stations to access the physical network and is a transfer point for passing information through a network.

Main characteristics, advantages, and disadvantages of Circuit-switched network

A circuit-switched network is a sub-network in which a dedicated circuit is established between sender and receiver, and all data passes over this circuit. But once the circuit is established, all subsequent data travels quickly from node to node. A circuit switched network has two key disadvantages. First, each circuit is dedicated to one connection. Second, when the circuit is used to transfer data (as oppose voice), it is probably not being utilized fully, because computer data transfer is often sporadic.

Packet-Switched Network

Packet-switched networks break long computer messages into fixed-sized packets and are thus designed to support computer data transmissions. In a datagram packet-switched network, each packet is an entity by itself. When a large message has been broken into **n** packets, each packet in a datagram network enters a node, where a unique routing
decision is made ,Because there is no fixed circuit to follow, the network spends no time in creating a circuit. Time is spent, however, in determining each packet's route.

Nevertheless, datagram networks are quite flexible, as they can react quickly to network changes. The other type of packet-switched network, the virtual circuit packet-switched network, is a marriage of the packet-switched network and the circuit-switched network. When a large message in this network is broken into multiple packets, all these packets follow the same path through the sub-network. This path is determined before the first packet is transmitted-an activity that requires circuit setup time-and a virtual circuit, with its internal routing tables, helps route the packets from source to destination.

Connection-Oriented V s Connectionless Applications

A connection-oriented network application provides some guarantee that information traveling through the network will not be lost and that the information packets will be delivered to the intended receiver in the same order in which they were transmitted. A connectionless network application does not require a logical connection to be made before the

transfer of data. Thus, a connectionless application does not guarantee the delivery of any information or data.

4.0 CONCLUSION

In this unit, we looked at some of the technologies and communication protocols use to create MANs and W ANs. We discussed SONET and ETHERNET used to create MANs Also different physical sub-networks for W ANs such as circuit-switched, packet-switched, and broadcast were identified and discussed. We concluded this unit with the identification of two categories of network applications or logical entities: connection oriented applications and connectionless applications.

5.0 SUMMARY

A network that expands into a metropolitan area and exhibits high data rates, high reliability, and low data loss is called a metropolitan area network (MAN). Many of the same technologies and communications protocols found in local area networks (and wide area networks) are used to create metropolitan area networks. Metropolitan area networks are based upon either SONET or Ethernet backbones. SONET backbones consist of fiber-optic rings, while Ethernet backbones are mesh networks. A Metro Ethernet service provides an Ethernet interface to a business and can transfer data at high rates over metropolitan areas. Wide area networks cover larger geographic areas than both local area networks and metropolitan area networks, and they are based on potentially different physical sub-networks: circuit-switched, packet-switched, and broadcast. A circuit-switched network creates a dedicated circuit between sender and receiver, and all data passes over this circuit. A packet-switched network transmits fixed-sized packages of data called packets. Packet-switched networks fall into two subcategories: datagram networks and virtual circuit networks. The datagram packet-switched network transmits each packet independently of every other packet. Each packet is considered a single entity and is not part of a larger grouping of packets. The virtual circuit packet-switched network creates a virtual circuit using routing tables and transmits all packets belonging to a particular connection over this virtual circuit. A broadcast network transmits its data to all workstations at the same time. Broadcast networks are more often used in local area networks than in wide area networks. The network application that runs over a sub-network can be either connection oriented or connectionless. A connection-oriented network application provides some guarantee that the information traveling through the network will not be lost and that the information packets will be delivered to the intended receiver in the same order in which they were transmitted. To provide this service, a logical

connection is established before any data transfer takes place. A connectionless network application does not require a logical connection to be made before data is transferred.

6.0 TUTOR-MARKED ASSIGNMENT

1. State three advantages of a SONET-based metropolitan area network over and Ethernet-based metropolitan area network.
2. Which type of network application requires more elaborate software: connection oriented or connectionless? Explain.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.

UNIT 2 ROUTING AND NETWORK CONGESTIONS

CONTENTS

1.0 INTRODUCTION

Recall that a wide area network's underlying sub-network consists of multiple nodes, each with multiple possible connections to other nodes within the sub-network as discussed in Unit 11. Each' node is a router that accepts an input packet, examines the destination address of the packet and forwards the packet onto a particular communications line. In the case of multiple-linked nodes, there may be one or more paths into a node as well as one or more paths out of a node. If most of the nodes in the sub-network have multiple inputs and outputs, numerous routes from a source node to a destination node may exist. How is routing through a wide area network accomplished?

This unit attempt to answer this question by examining different routing and network congestion schemes.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- Describe the differences between centralized routing and distributed routing, and cite the advantages and disadvantages of each.
- Describe the differences between static routing and adaptive routing

- Document the main characteristics of flooding
- Discuss the basic concepts of network congestion, including quality of service

3.0 MAIN CONTENT

3.1 Routing

How is routing through a wide area network accomplished? Let's consider the Internet (to be discuss fully in Unit 14) as an example: It is a massive collection of networks, routers, and communications lines (various types of telephone lines). When a data packet enters a router, the router examines the IP address encapsulated in the network layer of the packet and determines where the packet should go next. When there are multiple routes through a network such as the Internet, how is anyone particular route selected? Although routing on the Internet is fairly complex, it is possible to examine the, basic routing techniques that all types of wide area networks employ. But keep in mind that a wide area network does not use only one form of routing. The routing algorithms used within the Internet, for example, are actually combinations of several types of basic routing techniques. To begin to understand the often complex issue of routing, it is helpful to think of the sub-network as a graph consisting of nodes (computers, routers, or telephone switches) and edges (the communications links between the nodes), as shown in Figure 12-1. In this network graph, the edge between each pair of nodes can be assigned a weight or associated cost, as has been done in Figure 12-1, to form a structure called a weighted network graph.

You can assign many meanings to the weights in a weighted network graph.

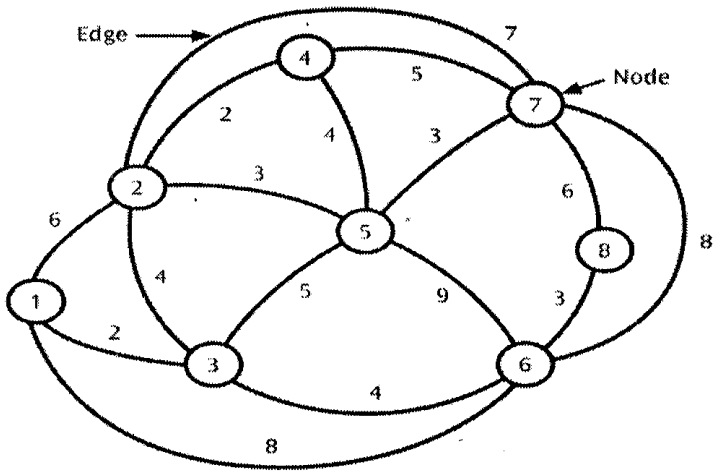


Figure 12.1: Example of a Network Graph

For example, a weight can correspond to the naira cost of using the communications link between two nodes. A weight could also represent the time-delay cost associated with transmitting data on that link between the source and destination: nodes. Another factor that is commonly represented as a weight is the size of the queue that has backed up while waiting for a packet to be transmitted onto a link. Each of these weights may be useful for determining a route through a network.

Once you consider the sub-network as a graph and assign weights to the paths between nodes, you can develop an algorithm for traversing the network. There are, in fact, many algorithms for selecting a route through a network. Often, algorithms strive for an optimal route through a network, but there are different ways to define "optimal." For example, one algorithm might define the optimal route as one that generates the least naira cost. Another algorithm might consider the path with the least time delay to be the optimal route. A third algorithm might define the optimal route as the one having the smallest queue lengths at the nodes along the path. Some algorithms use criteria other than optimality. For example, they might try to balance the network load over a number of different paths. Another kind of algorithm might favor one type of traffic over another, for example, real-time traffic over non-real-time traffic. A third type m ay try to remain robust, responding to changing network demands as nodes and communications links fail or become congested. Yet a fourth kind may try to remain static and not switch between possible paths. As you can see, routing is a complicated topic. To get a feel for routing in wide area networks, let's examine several of the most commonly used routing algorithms (Dijkstra's least-cost algorithm and flooding) and several techniques for managing routing information (centralized vs. distributed routing; adaptive vs. fixed routing). Most wide area networks use a combination of these routing techniques to achieve a routing algorithm that is fair, efficient, and robust, but at the same time stable.

3.1.1 Dijkstra's Least-Cost Algorithm

One possible method for selecting a route through a network is to choose a route that minimizes the sum of the costs of all the communications paths along that route. A classic algorithm that calculates a least-cost path through a network is '" Dijkstra's least-cost algorithm. This algorithm is executed by each node, and the results are stored at the node and sometimes shared with other nodes. Because this calculation is time-consuming, it is done only on a periodic basis or when something in the network changes-for example, when there is a connection or node failure. Let's look at an example. Figure 12.2 depicts

sub-network with cost associated with each link. Path A-B-G has a cost of 9 (2 + 7), A-D-G; has a cost of 10 (5 + 5), and , path A-B-E-G has a cost of 8 (2 + 4 + 2). To ensure that you find the least-cost route, you need to use a procedure that calculates the cost of every possible route, starting from a given node. Although the human eye can quickly pick out a path through a network graph, there are a number of drawbacks to "eyeballing" the data to find a solution:

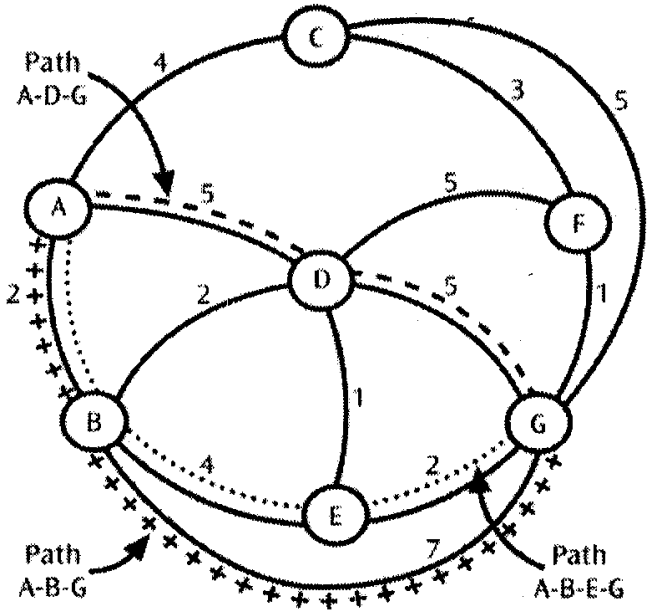


Figure 12.2: Network with costs associated with each link

- You can easily miss one or more paths.
- You may not find the least-cost path.
- Wide area networks are never as simple as the network in Figure 12.2; thus, eyeballing the data could never be sustained as a reliable, long-term procedure.

Most wide area networks use some form of Dijkstra's algorithm to determine a least-cost route through a network, whether that cost is a measure of time or money.

3.1.2 Flooding

Compared to Dijkstra's least-cost routing algorithm, the flooding technique seems simple. Flooding states that each node takes the incoming packet and retransmits it onto every outgoing link. For example, assume a packet originates at Node A, as shown in Figure 12.3. Node A simply transmits a copy of the packet on everyone of its outgoing links. Thus, a copy of the packet (the first copy, or copy 1) is sent to Nodes B, C, and D. When the packet arrives at Node B, B simply transmits a copy of the packet to each of its outgoing nodes (D, E, and

G). Likewise, Node C will transmit a copy of the packet to each of its outgoing nodes (F and G). Node D will also transmit a copy of the packet to each of its outgoing nodes (B, E, F, and G). Figure 12.4 shows the second copies of the packets leaving Nodes B, C, and D. It should not take you long to realize that the network will very quickly be flooded with copies of the original data packet.

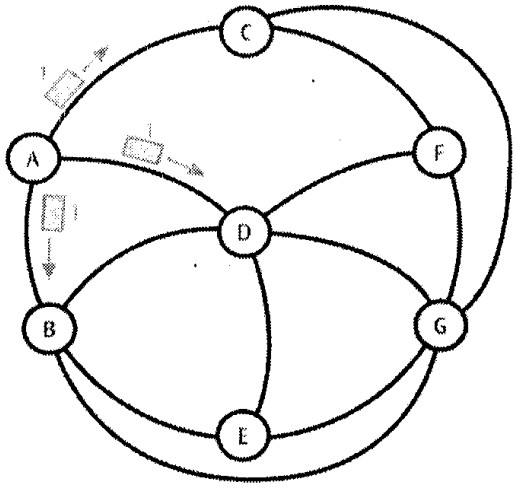


Figure 12.3 Network with flooding starting from Node A

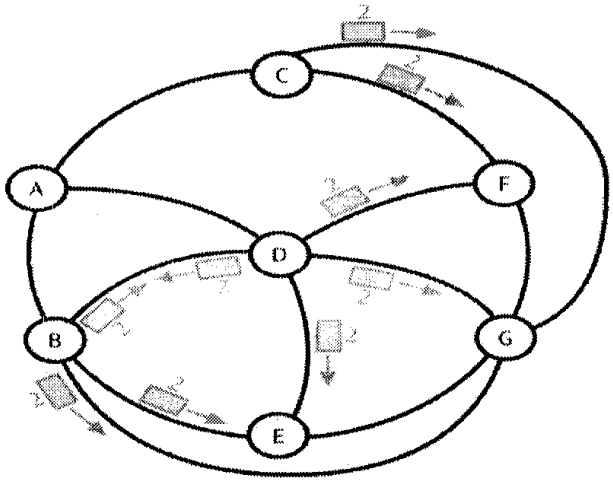


Figure 12.4: Flooding has continued to Nodes B, C, ad D

To prevent the quantity of copied packets from becoming overwhelming, two common sense rules can be established. First, a node need not send a copy of the packet back to the link from which the packet just arrived. Thus, when Node A sends a copy of the packet to Node C, C does not need to send a copy immediately back to A. Second, a network limit, called the hop limit, can be placed on how many times any packet is copied. Each time a packet is copied, a counter associated

with the packet increases by one. This counter is called the hop count. When the hop count equals the network hop limit, this particular packet will not be copied anymore. For example, suppose the network has a hop limit of 3. When Node A first sends copies to B, C, and D, each of the three copies has a hop count of 1. When the packet arrives at Node C, copies with hop counts of 2 will be sent to F and G (Figure 12.4). When the copy arrives at Node F, two copies with hop counts of 3 will be transmitted to D and G, and the packets that arrive at D and G with hop counts of 3 will go no farther. Although flooding may seem like a strange way to route a packet through a network, the procedure does have its merits. If a copy of a packet must be sent to a particular node, flooding will get it there, assuming, of course, there is at least one active link to the receiving node, and the network hop limit is not set at too small a value. Flooding is also advantageous when a copy of a packet needs to get to all nodes-for example, when emergency information or network initialization information is sent. The major disadvantage of flooding is the large number of copied packets distributed throughout the network.

3.1.3 Centralized Vs. Distributed Routing

Centralized and distributed routing are not so much algorithms for routing data packets through a network as they are techniques for providing routing information. Centralized routing involves storing all the routing information at one central location (Table 12.1). Whenever any router in a network needs routing information, this central location is queried and the routing results are returned. For example, if a packet arrives at Node A and is destined for Node G, Table 12.1 conveys that it should be sent to Node B next. These days, centralized routing is rarely used in wide area networks. Instead, most wide area networks now employ distributed routing.

Table 12.1: Routing table kept at a centralized network site

		Destination Node						
		A	B	C	D	E	F	G
Origination Node	A	-	B	C	B	B	C	B
	B	A	-	A	D	D	D	D
	C	A	A	-	A	F	F	F
	D	B	B	F	-	E	E	E
	E	D	D	G	D	-	G	G
	F	C	G	C	G	G	-	G
	G	E	E	F	E	E	F	-

Distributed routing is a technique that uses a routing algorithm, such as a least-cost algorithm, to generate routing information and dictates that this information be stored (in the form of routing tables) at distributed locations-typically, routers within the network. When a data packet enters the network at node x, that node consults its own routing table to determine the next node that should receive the packet. In this scheme, each node needs routing information only for its own locale. For example, the routing table for Node C would be Table 12.2. From this table, you can see that if a data packet arrives at Node C and is destined for Node G, the packet should be sent to Node F next. (Once the packet arrives at Node F, Node F's routing table will have to be examined for the next hop in the path.)

Table 12.2: Local routing table for node C

		Destination Node						
		A	B	C	D	E	F	G
Origination Node	C	A	A	-	A	F	F	F

One of the primary advantages of distributed routing is the fact that no single node (or central router) is responsible for maintaining all routing information. This situation confers a number of benefits. First, if any node crashes, it will probably not disable the entire network. Second, a node will not need to send a request to a central router, because each node has its own table. One disadvantage of distributed routing is related to the problems that arise if the routing tables need to be updated. When all the routing information is in one place (that is, in a single table), it is simple to make updates. When routing information is scattered throughout a network, getting the appropriate routing information to each node is a complex problem. Another consequence of storing routing information at multiple locations is that at any given point in time there may be one or more routing tables that contain old or incorrect information.

3.1.4 Adaptive Vs. Fixed Routing

Centralized and distributed routing are methods for sending routing information. They are typically used in conjunction with some form of least-cost routing algorithm. Regardless of whether routing information is centralized or distributed, when networks change, routing information needs to change too. When routing tables adapt to network changes, the routing system is called adaptive. Adaptive routing is a dynamic technique in which routing tables react to network fluctuations, such as congestion and node/link failure.When a problem occurs in a network with adaptive routing, the appropriate information is
transmitted to the routing tables, and new routes that avoid the problem areas are created. Adaptive routing raises some questions and issues:

How often should information be shared, and how often should routing tables be updated? How much additional traffic is generated by messages transmitting routing information? Unfortunately, adaptive routing can add to network congestion. Each time the network experiences a change in congestion, information about this change is transmitted to one or more nodes. The transmission of this information adds to the congestion, possibly making it worse. In addition, if the network reacts too quickly to a congestion problem and reroutes all traffic onto a different path, it can create congestion problems in a different area. The network might then detect the congestion in this different area and possibly reroute all traffic back toward the first problem area. This back and forth rerouting produces a yo-yo effect, affecting network stability and decreasing efficiency.

The opposite of adaptive routing is fixed routing. With fixed routing, routing tables are created once, typically when the network is installed, and then never updated again. While this method is simple and eliminates the need for routers to talk to one another (thus avoiding additional traffic), it can also yield networks with out-of,-date information and thus inefficient or slow routing. It is debatable if fixed routing exists anymore on wide area networks (or for that matter, on any other type of network).

3.1.5 Routing Examples

The Internet is covered in detail in Unit 14, but let's take a few minutes to examine two of the routing algorithms that have been used within the Internet over the years. By examining these algorithms, you will see how a real-life routing protocol is actually composed of many of the algorithms and techniques introduced in the preceding sections. The first routing algorithm used within the Internet (when it was still called ARPANET) was called a distance vector routing algorithm. The distance vector routing algorithm was an adaptive algorithm in which each node maintained a routing table called a vector. Because each node maintained its own routing table, the routing algorithm was also a distributed algorithm. Every 30 seconds, each node exchanged its vector with its neighbors. When all its neighbors' vectors came in, a node would update its own vector with the least-cost values of all the neighbors. This adaptive, distributed algorithm had the formal name Routing Information Protocol (RIP). This protocol had two unfortunate side effects. The first was that good news (routing information that indicated a shorter path) moved relatively slowly through the network, one router at a time. The second side effect was that bad news, such as ,a router or link failure, very often moved even more slowly through the network.

The next routing protocol that was implemented on the Internet (in approximately 1979) was called a link state routing algorithm. Link state routing essentially involves four steps. The first step is to measure the delay or cost to each neighboring router. For example, each router can send out a special echo packet that gets bounced back almost immediately. If a timestamp were placed on the packet as it left and again as it returned, the router would know the transfer time to and from a neighbouring router. The second step is to construct a link state packet containing all this timing information. The third step is to distribute the link state packets via flooding. In addition to using flooding, the link state routing algorithm is a distributed algorithm. The fourth and final step is to compute new routes based on the updated information. Once a router collects a full set of link state packets from its neighbours, it creates its routing table, usually using Dijkstra's least-cost algorithm. Open Shortest Path First (OSPF) protocol is a link state algorithm and is still used today by many Internet routers.

Now that we know how data is routed in a network, we need to look at a common side effect of routing too much data at a time-congestion.

SELF-ASSESSMENT EXERCISE

i. Describe the concept of flooding as it relates to routing in computer network.
ii. Explain the difference between Centralized and Distributed routing

3.2 Network Congestion

When a network or a part of a network becomes so saturated with data packets that packet transfer is noticeably impeded, network congestion occurs. Congestion may be a result of a short-term problem, such as a temporary link or node failure, or it may be a result of a longer-term problem, such as inadequate planning for future traffic needs or poorly created routing tables and routing algorithms. As with so many things in life, the network is only as strong as its weakest link. If network designers could properly plan for the future, network congestion might exist only in rare instances. But the computer industry, like most other industries, is filled with examples of failures to plan adequately for the future. Computer networks are going to experience congestion, and no amount of planning can avoid this situation. Thus, it is important to consider effective congestion-avoidance and congestion-handling techniques.

3.2.1 Problem Associated with Network Congestion

Networks experience congestion for many reasons. A network failure-either failure on a communications link between nodes, or the failure of the node itself-may lead to network congestion. If the network cannot quickly detect the point of failure and dynamically route around this point, it may experience a wide range of congestion problems, from a small slowdown on an individual link to total network collapse. Even if the network were to begin the rerouting process, it might still experience congestion because one less network path would be available. But communications link and node failures are not the only causes of congestion. Insufficient buffer space at a node in a sub-network can also cause network congestion. It is not uncommon to have hundreds or even millions of packets arriving at a network node each second. If the node cannot process the packets quickly enough, incoming packets will begin to accumulate in a buffer space. When packets sit in a buffer for an appreciable amount of time, network throughput begins to suffer. If adaptive routing is employed, this congestion can be recognized, and updated routing tables can be sent to the appropriate nodes (or to a central routing facility). But changing routing tables to deflect congestion might provide only a temporary fix, if any fix at all. What is needed is a more permanent solution. Two possible more permanent solutions would be increasing the speed of the node processor responsible for processing the incoming data packets and increasing the amount of buffer space in the node. Unfortunately, both of these solutions may take a large amount of time and money to implement. Perhaps less costly alternatives are possible.

What happens if the buffer space is completely full and a node cannot accept any additional packets? In many systems, packets that arrive after the buffer space is full are discarded. Although this is not a very elegant solution, it momentarily solves the problem of too many packets. Unfortunately, this is like bad medicine- it treats the symptoms, but not the disease. What is needed is a solution that reacts quickly to network congestion and addresses the real problem-too many packets. Let's examine several possible solutions that have been suggested or tried in the past several years.

3.2.2 Possible Solutions to Congestion

Solutions to network congestion generally fall into two different categories. The first category contains solutions that are implemented after congestion has occurred. The second category of solutions contains techniques that attempt to avoid congestion before it happens. Let's assume that congestion has already occurred and examine the first category of solutions. Most, if not all, of the solutions in the first

category involve telling the transmitting station to slow down or stop its transmission of packets into the network. For example, if an application on the network suddenly realizes that its packets are being discarded, the application may inform the transmitting station to slow down until further notice. Because the application is simply observing its own throughput and not relying on any special types of signals coming from the network, this is called implicit congestion control. Often, however, the network itself sends one or more signals to a transmitting station, informing the station to slow down or stop its insertion of packets into the network. When the network signals the transmitting station to slow down, this is called explicit congestion control. For an example, consider frame relay which is designed to use two explicit congestion control techniques: forward explicit congestion notification and backward explicit congestion notification. With forward explicit congestion notification (FECN), when a frame relay router experiences congestion, it sends a congestion signal (inside the data frames) forward to the destination station, which in turn tells the originating station to slow down the transfer of data. With backward explicit congestion notification (BECN), the frame relay router experiencing congestion sends a signal back to the originating station, which then slows down its transmission. Other congestion control methods are based on simpler techniques such as the flow control methods introduced in Unit 7. Flow control at the data link layer allows two adjacent nodes to control the amount of traffic passing between them. When the buffer space of a node becomes full, the receiving node informs the sending node to slow or stop transmission until further notice.

Now let's consider the second category of congestion control techniques those that try to prevent congestion before it even happens. The old saying "An ounce of prevention is worth a pound of cure" certainly applies to computer net-works. Avoiding congestion not only leads to fewer lost and delayed packets but also to many happier customers. One possible solution to controlling the flow of packets between two nodes is buffer pre-allocation. In buffer pre-allocation, before one node sends a series of n packets to another node, the sending node inquires in advance whether the receiving node has enough buffer space for the **n** packets. If the receiving node has enough buffer space, it sets aside the **n** buffers and informs the sending node to begin transmission. Although this scheme generally works, it does introduce extra message passing, additional delays and possible wasted buffer space if all n packets are not sent. But the alternative-discarding packets due to insufficient buffer space are worse.

More recent network technologies such as Asynchronous Transfer Mode (ATM) approach network congestion in a very serious fashion. Because ATM networks transfer data at very high speeds, congestion can occur

rather quickly and be devastating. Thus, it is extremely important to keep congestion from occurring in the first place. ATM uses a congestion avoidance technique that appears to work quite well. This technique-connection admission control-avoids congestion by requiring users to negotiate with the network regarding how much traffic they will be sending, or what resources the network must provide to satisfy the user's needs before the user sends any data. If the network cannot satisfy the user's demands, the user connection is denied. In negotiating the viability of a connection, users and networks must resolve questions such as the following:

- What is the average (or constant) bit rate at which a user will transmit?
- What is an average peak bit rate at which a user might transmit?
- At what rate might a network start discarding packets in the event of congestion?
- What is the average bit rate that the network can provide?
- What is the average peak bit rate that the network can provide?

Many networks relate these issues to quality of service (QoS), a concept in which a network user and the sub-network agree on a particular level of service (acceptable guidelines for the proper transfer of data). For example, a user who requires a very fast, real-time connection to support live action video will negotiate with the network for a particular quality of service. If the network can provide this level of quality, a contract is agreed on, the user is charged accordingly, and a connection is established. If a second user requires a slower connection for e-mail, a different level of quality is agreed on, and this connection is established. Very often these agreements between service provider and a service user are formalized in a service level agreement, a legally binding, written document, that can include service parameters offered in the service, various types of service/support options, incentives if the service levels are exceeded,) and penalties if service levels are not met. Unfortunately, only one network technology successfully supports connection admission control and I quality of service, and that is Asynchronous Transfer Mode. Most A TM systems can provide a range of services from high-speed constant bit rate down to slower-speed bit rate on demand. CSMA/CD, the most popular LAN protocol, does not provide different levels of service, nor does the Internet, with its TCP and IP protocols. (As we will see in the next Unit 14, however, the newer Internet protocol, IPv6, does include some form of labeling to support customer-specified connections.)

ANSWER TO SELF-ASSESSMENT EXERCISE

Concept of Flooding

When a packet arrives at a node, the node sends a copy of the packet out to every link except the link the packet arrived on. Traffic grows very quickly when every node floods the packet. To limit uncontrolled growth, each packet has a hop count. Every time a packet hops, its hop count is incremented. When a packet's hop count equals a global hop limit, the packet is discarded.

Centralized Routing

One routing table is kept at a "central" node. Whenever a node needs a routing decision, the central node is consulted. To survive central node failure, the routing table should be kept at a backup location. The central node should be designed to support a high amount of traffic consisting of routing requests.

Distributed Routing

Each node maintains its own routing table. No central site holds a global table. Somehow each node has to share information with other nodes so that the individual routing tables can be created. Possible problem with individual routing tables holding inaccurate information.

4.0 CONCLUSION

This Unit discussed into detail routing and congestion control mechanisms. There various techniques for selecting the optical route in a network such as Dijkstra's least-cost algorithm, arid flooding were discussed. We also described the various methods for providing routing information such as centralized, distributed, adaptive and static routing. Finally, we discussed the concept of network congestion, the problems associated with network congestion and possible solutions to congestion.

5.0 SUMMARY

Selecting the optimal route for the transfer of a data packet through a network is a common service of many networks. This optimal route is obtained by combining two or more of the many routing algorithms and techniques available today. One possible way to select an optimal route through a network is to choose a path whose total path costs have the smallest value. This technique is based on Dijkstra's least-cost algorithm and is a common method for determining the optimal route. Flooding is

a routing technique that requires each node to take the incoming packet and retransmit it onto every outgoing link. If a copy of a packet must be sent to a particular node, flooding will get it there, but it will also create a very large number of copied packets that are distributed throughout the network. Centralized routing is a technique for providing routing information that dictates that the routing information, which is generated by a method such as the least-cost algorithm, be stored at a central location within the network. Another technique for providing routing information, distributed routing allows each node to maintain its own routing table. Adaptive routing allows a network to establish routing tables that can change frequently, as network conditions change.

When a network or a part of a network becomes so saturated with data packets that packet transfer is noticeably impeded, network congestion has occurred. Congestion may be the result of too much traffic, network node failure, network link failure, or insufficient nodal buffer space. Remedies for network congestion include implicit congestion control, explicit congestion control, flow control, pre-allocation of nodal buffers, and connection admission control parameters. Quality of service (QoS) parameters can be used by network users and the service providers to establish acceptable guidelines for the proper transfer of data. These guidelines can cover transmission speed, level of errors, and overall network throughout

6.0 TUTOR-MARKED ASSIGNMENT

1. List the steps involved in creating, using and terminating a virtual circuit.
2. Using the concept of flooding and graph shown in figure 12.5, how many packets will be created if a packet originates at Node A and there is a network hop limit of three?

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.

UNIT 3 NETWORK SECURITY

CONTENTS

1.0 INTRODUCTION

Computer network security has reached a point at which it can best be characterized by two seemingly conflicting 'statements: Never has a network security been better than it is today; and never have computer networks been more vulnerable than they are today. Today, the Internet(to be discussed fully in Unit 14 and 15) allows anyone in the world to access or attempt to access any computer system that it is connected to the Internet. This interconnectivity between computer systems and networks is both a boon and a bane. It allows us to download and order toys for kids, but it also exposes all Internet-attached systems to invasion.

In this unit, we will look first at the individual elements of security, especially encryption and authentication. We will then see how these elements are combined into complete integrated security systems. We will finish with discussions of Internet firewalls and other security issues.

2.0 OBJECTIVES

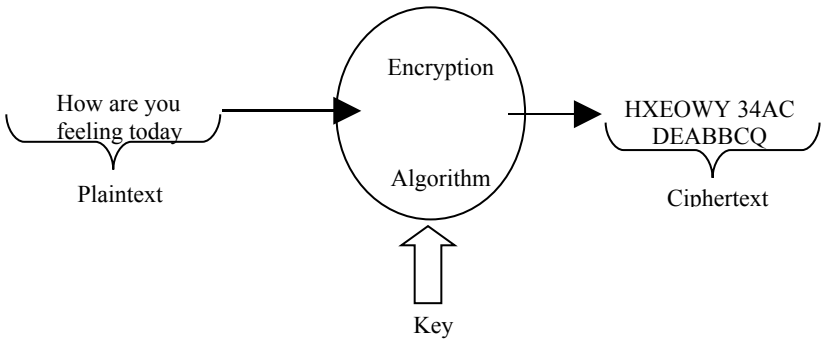
After studying this unit, you should be able to describe:

- Symmetric key encryption and public key encryption
- Authentication
- Integrated security system
- Other security matters, including multilayer security, firewalls, intrusion detection, etc.

3.0 MAIN CONTENTS

3.1 Encryption

The most basic building block of security is encryption, which scrambles a message before transmission, so that an interceptor cannot read the message as it flows over the network. However, the receiver knows how to decrypt (descramble) the encrypted message making it readable again. Encryption provides privacy, which is also called confidentiality. If you are transmitting your credit card number over a network, or a bank is transmitting an electronic transaction, privacy is critical. Figure 1. 3.1 look more closely at encryption procedure. Before a message is encrypted, it is called plaintext. After a plaintext message is encrypted, we called the stream of bits that the encryption generates the ciphertext. At the final destination the ciphertext is now converted back



to the original message. There are two elements in encryption: an encryption method and a key. The encryption method specifies the mathematical process that will be used in the encryption. Each encryption method uses a special string of bits called a **key**.

There are many encryption methods. These method falls into two categories: symmetric key encryption and public key encryption. Examples of symmetric key encryption included: data encryption standard **(DES), RCS, Blofwish, 3DES,IDEA, etc.** Under public key encryption,' common algorithm include, RSA, elliptical curve cryptosystem(ECC), and EI Gamal.

We will look first at symmetric key encryption, which is the simplest form of encryption. Figure 13.1 specifically illustrates symmetric key encryption. When party A sends to party B, party A encrypts with the single symmetric key and party B decrypts with same key. When party B sends to party A, party B encrypts with the single symmetric key and party A decrypts with same key. This is called symmetric key encryption because the encryption process is symmetrical: the same key both encrypts and decrypts. One problem with symmetric key encryption is that both parties must keep the symmetric key secret. Key distribution is a special problem in symmetric key encryption because each pair of partners needs different keys. For example, if there are N users, each of which need to communicate with every other user, N*(N-1)/2 symmetric keys would be needed. For only 100 users, this would mean 4,950 keys! The most widely used symmetric key encryption algorithm is the data encryption standard(DES). DES breaks the plaintext into chunks of 64 bits. It then encrypts each chunk of plaintext using 64-bit key. However, it is possible to say that DES uses a 56-bit key, because 8 of the key bits are redundant, to check for incorrect keys. At the other end, each 64-bit chunk is decrypted back to plaintext. With a key length of only 56 bits, DES is no longer considered adequate for applications such as electronic fund transfer. Many security systems no use a stronger variant 0 DES called Triple DES. As the name suggests, 3DES involves applying the DES algorithm three times instead of once. However, this has to be done with more than one key to be successful.

The second is public key encryption. Recall that in symmetric key encryption, both parties use a single key to encrypt and decrypt. In contrast, in public key encryption, when one party sends to another, there are two keys: the receiver's public key and the receiver's private key. Note that both of these keys are those of receiver, not of the sender. In this method, for N communicating partners, there are only N public keys. For N users, only N private key-public keys are needed, not N*(N-I)/2. One major problem with public key encryption and decryption is that they are about a hundred times slower than symmetric key encryption and decryption is that they are about a hundred times slower than symmetric key encryption and decryption. It cannot encrypt and decrypt long messages fast enough for most purposes. It is only used to encrypt brief messages.

3.2 Authentication

One purpose of encryption is to prevent anyone who intercepts a message form being able to read the message. Encryption brings confidentiality, which is also called privacy. In contrast, authentication

has a different purpose, namely to prove the sender's identity. There are many forms of authentication. We will look briefly at four.
The first form of authentication is **password**, which offer only a weak form of authentication. Many people select passwords that are common words. If users are given long and meaningless passwords, in turn, they either write them down or have their computer remember them(allowing anyone using their computer to pose as them). Of course, having a good password is, worthless if the user has his or her computer remember important passwords so that anyone walking up to their computer can impersonate them. The second form of authentication is authentication cards. Automatic teller machines use authentication cards that contain coded information. Users swipe them through a card reader slot and usually type a password as well.

Biometric authentication' is another form of authentication that measures body dimensions. Finger print analyzers are inexpensive and fairly good. At the other extreme, iris analyzers, which look at the iris in one of your eyes, are much more precise but also more expensive. Currently, Biometric products suffer from lack of vendor interoperability because there are no effective standards. The fourth to be discussed is the public key authentication. Recall that private keys must be kept secret. Therefore, if a person or other entity can demonstrate possession of their private key, the person or the entity will be authenticated. One basic way to implement public key authentication is by the use of public key challenge-response authentication. In this form of authentication there is a verifier, which wishes to authenticate the identity of its communication partner, the applicant. First the verifier creates a short plaintext message called the challenge message. The challenge message has to be short because it will be encrypted with public key encryption, which is good only for short messages. The verifier sends this plaintext to the applicant. Second, the applicant encrypts the challenge message with the applicant's own private key. This create the response message. The applicant sends this response message back to the verifier. Third, the verifier attempt to decrypt the response message using the applicant's known public key. If the public key successfully decrypts the response message back to the original challenge message, then the applicant holds the private key of the person or entity the applicant claims to be. The applicant is authenticated. Digital signatures allow us to authenticate each message that an applicant sends

3.3 Digital Certificates and Public Key Infrastructures (PKI)

Unfortunately, public key encryption can be used deceptively. For instance, an impostor claims to be a certain true person in an attempt to deceive a verifier. First, the impostor sends its own public key to the verifier. In the delivery message, the impostor says, in effect, "Hi. I am

the true person. Here is my public key". The verifier now incorrectly believes that it has the true person's public key, although it really holds the impostor's public key. This is a critical deception. Everything is easy. When the impostor send messages with digital signatures using its own private key, the public key that the verifier holds will wrongly "authenticate" the digital signature as being that of the true person.

Thanks on the above illustration; public key encryption is useless unless you can independently verify the true party's public key through a trusted third party. Digital certificates make this possible. Note that digital certificates by themselves do not authenticate an applicant. They merely provide the true public key for a certain person, program, or hardware device.

3.4 Integrated Security System

So far, we have only looked at the element of secure communication. However, if two processes on different computers are to communicate smoothly, we need to implement several forms of security, and we need a process for doing this systematically automatically, Integrated security system(ISSs) implement this broad spectrum of activities automatically.

3.4.1 Secure Socket Layer (SSL)

SSL is an example of ISS which secures most electronic commerce transactions. Every time your browser issues' an https://command, it begins an SSL session. Figure 13.2 shows a summary of communication between customer and the merchant host.

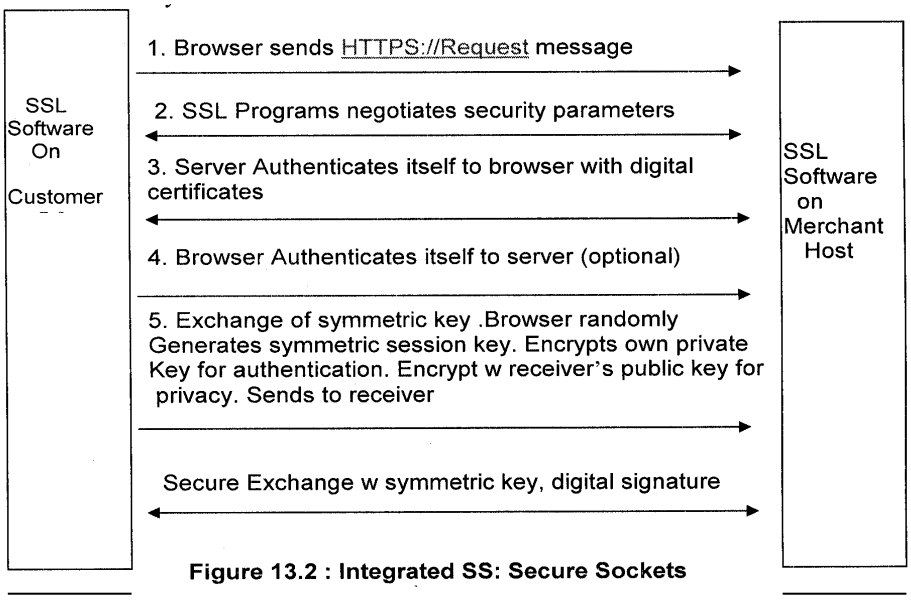


Figure 13.2: Integrated SS: Secure Sockets

3.5 Other Security Issues

We have discussed only some of the most important security issues. There are other important security issues to consider.

3.5.1 Multilayer Security

Security can be applied at any layer, Often, as figure 13,2 illustrates, ISSs are implemented at several layers, Old and established security algorithms have a nasty record of having hackers discover security problems after years of effective use. If security is employed at multiple layers, a single breakdown in an algorithm will not compromise security. On the negative side, each layer of security produces delays and increase costs.

3.5.2 Firewalls

A firewall is a system or combination of systems that supports an access control policy between two networks. A firewall can limit the types of transactions that enter a system, as well as the types of transactions that leave a system. Firewalls can be programmed to stop certain types or ranges of IP addresses, as well as certain types of TCP port numbers (applications), A packet filter firewall is essentially a router that has been programmed to filter out or allow to pass certain IP addresses or TCP port numbers. A proxy server is a more advanced firewall that acts as a doorman into a corporate network. Any external transaction that requests something from the corporate network must enter through the proxy server.

Proxy servers are more advanced but make external accesses slower. Figure 13.3 shows a firewall as its stops certain internal and external transactions, while figure 13.4 shows a proxy server sitting outside the protection of the corporate network.

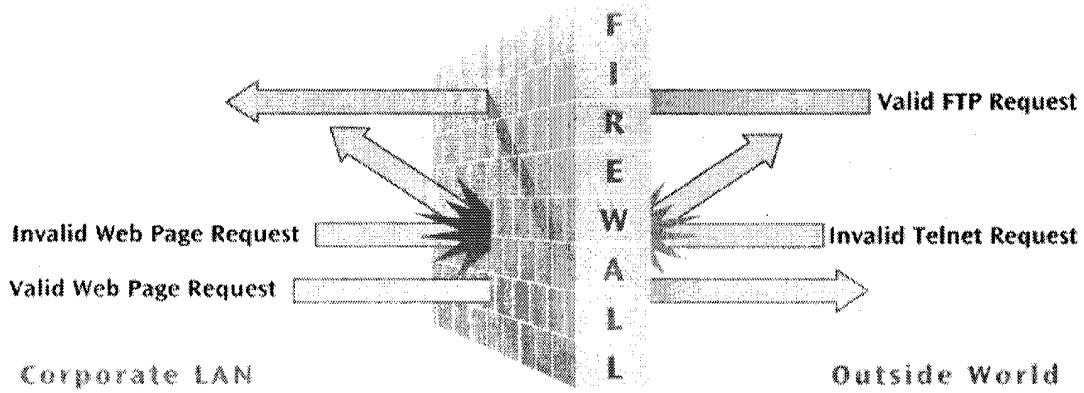


Figure 13.3: An example firewall

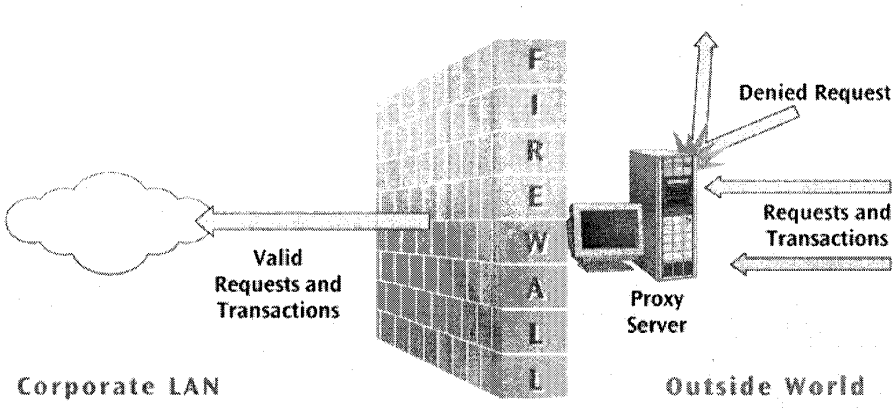


Figure 13.4: An example Proxy server used as firewall

3.5.3 Wireless Security

How do you make a wireless LAN secure?
WEP (Wired Equivalency Protocol) was the first security protocol used with wireless LANs. It had weak 40-bit static keys and was too easy to break.WPA (Wi-Fi Protected Access) replaced WEP. Major improvement including dynamic key encryption and mutual authentication for wireless clients.

3.5.4 Intrusion Detection

When adversaries attack, we need intrusion detection systems that detects when an attack is under way and notify the administrator. The administrator may the then take steps to thwart the attack. In fact, the intrusion detection system may take steps to thwart the attack itself. It is important to detect intrusions. Otherwise, hackers will simply keep trying different attacks until they find on that succeeds. Intrusion detection allows a company to realize that an attack is under way and to limit the time that the hacker has to attack the system. Intrusion detection is also important in helping a company assess the security dangers facing it. Many companies that install intrusion detection systems are amazed by the number of times their systems are attacked. They often find that outsiders are already breaking into their systems and are reading sensitive files. Intrusion detection systems keep an audit log giving the details of an attack. It is crucial to maintain this audit log if the company wishes to take legal measures against attackers.

SELF-ASSESSMENT EXERCISE

i. What are the elements of encryption.

ii. Highlight the different forms of authentication you know
iii. What are the main steps in a conversation using an integrated security system(ISS)?
iv. Define the role of a firewall in data communication network.

ANSWER TO SELF-ASSESSMENT EXERCISE

Element of Encryption: Encryption method, and a key.
Forms of authentication: password, biometrics, public key authentication.

ISS steps

- Browser sends request message
- SSL programs negotiates security parameters
- Server authenticates itself to browser with digital certificates
- Browser authenticates itself to server (optional)
- Exchange of symmetric key
- Secure exchange

Roles of Firewall in data communication network

supports an access control policy between two networks.
limit the types of transactions that enter a system, as well as the types of transactions that leave a system.

4.0 CONCLUSION

This topic is wide, hence, in this unit we described the concept of encryption and authentication. We identified two major categories of encryption: symmetric key encryption and public key encryption. We discussed the different form of authentication viz-a-viz: password, authentication cards, biometrics, and public key authentication. We also highlighted the steps involve when using secure socket layers to secure most electronic commerce transactions. Finally, we conclude this unit by looking at other network security issues such as multilayer security, firewalls" and intrusion detection.

5.0 SUMMARY

Encryption uses a method and a key to convert plaintext to ciphertext and ciphertext back to plaintext. The key must be secret, and to be secure, keys must be quite long. The two general categories of encryption are symmetric key encryption and public key encryption. Both categories are used widely and often in conjunction with one

another. In authentication, the sender is required to prove his or her identity. In public key authentication, the sender proves that it knows its private key, which only the sender should know. This can be done through challenge-response processes and digital signatures, in conjunction with digital certificates. Digital certificates by themselves do not authenticate a person or process. They merely verify that a particular public key belongs to a particular user.

Encryption and authentication are merely elements in integrated security systems, which provide fully secure communication. Usually, ISSs begin with a negotiation phase, implement authentication, distribute session keys, and communicate with symmetric key encryption and perhaps digital signatures. Other security matters include multilayer security, packet filter and application (proxy) firewalls, and intrusion detection.

6.0 TUTOR-MARKED ASSIGNMENT

1. Distinguish between packet filter firewalls and application firewalls in terms of what they examine. Which type of firewall offers more sophisticated protection? What are the weaknesses of application firewalls?
2. What is multilayer security? Highlight the advantages of this form of security.

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.

UNIT 4 THE INTERNET

CONTENTS

1.0 INTRODUCTION

The previous unit described various type of network ranging from LANs, MAN, and WANs. This collection of networks became known as the Internet. Today's present Internet is a vast collection of thousands of networks and their attached devices. The Internet began as the Arpanet during the 1960s. One high-speed backbone connected several university, government, and research sites. The backbone was capable of supporting 56 kbps transmission speeds and eventually became financed by the National Science Foundation (NSF). To support the Internet and all its services, many protocols are necessary. Some of the protocols that we will look at:

- Internet Protocol (IP)
- Transmission Control Protocol (TCP)
- Address Resolution Protocol (ARP)
- Dynamic Host Configuration Protocol (DHCP)
- Network Address Translation (NA T).

Let's begin our study of the fascinating world of the Internet by examining several of the more important Internet protocols.

2.0 OBJECTIVES

At the end of this Unit, you should be able to:

- Discuss the responsibilities of the Internet Protocol and Transmission Control Protocol and how IP can be used to create a connection between networks.
- Identify the relationships between TCP/IP and the protocols ICMP, UDP, ARP, DHCP, NA T, and Tunneling protocols.

3.0 MAIN CONTENT

3.1 Internet Protocols

From simple e-mail to the complexities of the Web, many services are available on the Internet (to be treated in Unit 15). What enables these varied services to work? What enables the Internet itself to work? The answer to these questions is Internet protocols. Recall that the Internet with all its protocols follows the TCP/IP protocol suite (Internet model). See figure 14. 1 (Recall the layers of the TCP/IP protocol suite that was introduced in Unit one). An application, such as e-mail, resides at the highest layer. A transport protocol, such as TCP, resides at the transport layer. The Internet Protocol (IP) resides at the Internet or network layer. A particular media and its framing resides at the network access (or data link) layer.

Other Internet protocol include: Internet Control Message Protocol (ICMP), and User Datagram Protocol (UDP). We now examine these protocols into detail.

3.2 Internet Protocol (IP)

The Internet Protocol (IP) provides a connectionless data transfer service over heterogeneous networks by passing and routing IP datagrams. IP datagram is essentially another name for a data packet. To be passed and routed on the Internet, all IP datagrams or packets that are passed down from the transport layer to the network layer (the connectionless packet delivery layer) are encapsulated with an IP header (see Figure 14.2

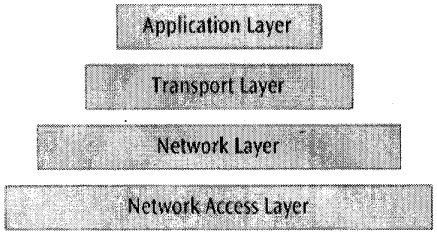


Figure 14.1: Hierarchy of layers as created by the Department of Defence

that contains the information necessary to transmit the packet from one network to another. The format of this header will be explained in the next few paragraphs. Consider once again the example of a workstation performing a network operation such as sending an e-mail message to a distant workstation, a process that is depicted in Figure 14.2. Suppose both workstations are on local area networks, and the two local area networks are connected via a wide area network. As the local workstation sends the e-mail packet down through the layers of the first internal network, the IP header is encapsulated over the transport layer packet, creating the IP datagram. The appropriate MAC layer headers are encapsulated over the IP datagram, creating a frame, and this frame is sent through LAN 1 to the first router. Because the router interfaces LAN 1 to a wide area network, the MAC layer information is stripped off, leaving the IP datagram. At this time, the router may use any or all of the IP information to perform the necessary internetworking functions. The necessary wide area network level information is applied, and the packet is sent over the WAN to Router 2. When the packet arrives at the second router, the wide area network information is stripped off, once again leaving the IP datagram. The appropriate MAC layer information is then applied for transfer of the frame over LAN 2, and the frame is transmitted. Upon arrival at the remote workstation, all header information is removed, leaving the original data.

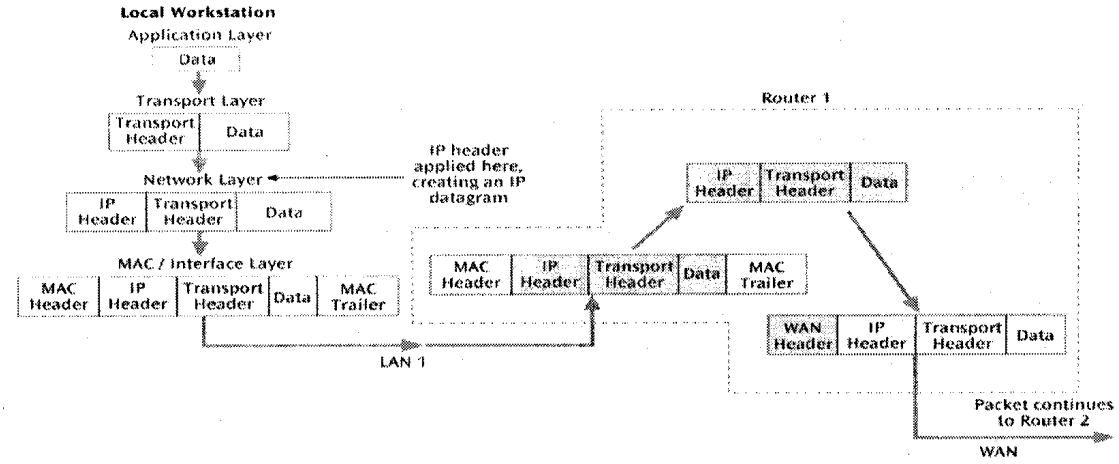


Figure 14.2: Movement of a packet from one network to another

When a router has the IP datagram, it may make several decisions affecting the datagram's future. In particular, the router must perform the following functions:

- Make routing decisions based on the address portion of the IP datagram.

- Fragment the datagram into smaller datagrams if the next network to be traversed has a smaller maximum packet size than the current size of the packet.
- Decide that the current datagram has been hopping around the network for too long and delete the datagram.

To perform these functions, the router needs address information, datagram size, and the time the datagram was created. This information is found in the IP header, which was the information applied to the transport packet at the network layer. Each of these router functions, along with the IP header fields that support these functions, will be examined in more detail in the upcoming section.

3.2.1 IP Datagram Format

Figure 14.3 shows that the network layer of the communications software added an IP header to the transport layer packet-thereby creating an IP datagram-before passing the packet on to the next layer of software. The information included in this IP header and the way the header is packaged allow the local area networks and wide area network to share data and create internetwork connections. Exactly what is in this IP header that allows this internetworking to happen? Figure 14.3 shows the individual fields of the IP header in more detail. Even though all 14 fields are important, let's examine the field that specifies the version of the Internet Protocol and those fields that affect three of the primary functions of IP: fragmentation, datagram discard, and addressing. By examining these fields, you will begin to discover how IP works and why it is capable of interconnecting so many different types of networks.

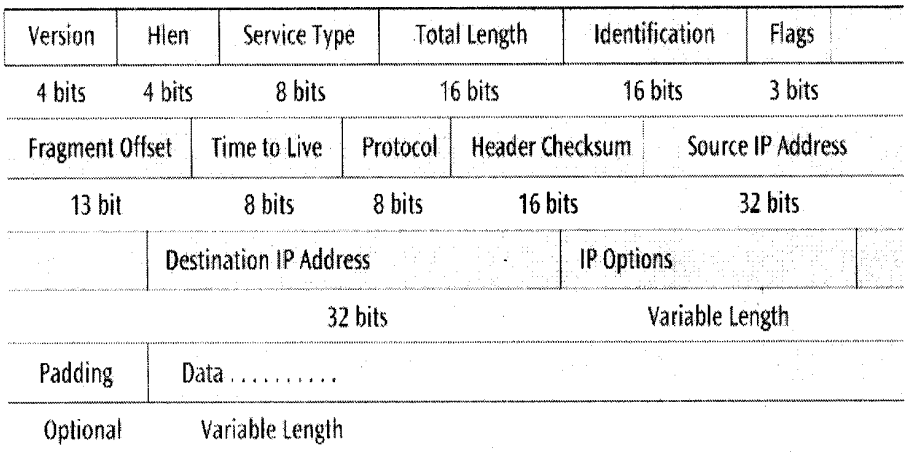


Figure 14.3: Format of the IP datagram

The first field of interest is the Version field. The Version field contains the version number of IP being used, just in case a new version becomes available. Currently, almost all networks involved in the Internet use IP version 4. IP version 6, which was created during the late 1990s and is discussed in detail in a later section of this unit, should eventually replace version 4. The Version field is important because two users using two different versions of the IP format at the same time would find that their packets could not talk to each other and would experience a nonfunctional system.

The next three fields of interest-Identification, Flags, and Fragment Offset-are used to fragment a datagram into smaller parts. Why would we want to fragment a datagram? When the Internet Protocol was created, the maximum packet size of some older networks that were still in existence was small. This maximum size was limited by network hardware, software, or other factors. Because IP was designed to work over practically any type of network, it had to be able to transfer datagrams of varying sizes. Rather than limit IP to the smallest maximum packet size in existence (who even knows what that is?), the Internet Protocol allows a router to break or fragment a large datagram into smaller fragments so it will fit onto the next network. Fortunately, hardly any modem networks have a maximum packet size that is small enough to be a concern. Thus, someday fragmenting a datagram into smaller packets will no longer be an important issue. In fact, as you will learn shortly, IP version 6 does not even have a field in the header to perform fragmentation. The next field we will examine is the **Time to Live** field, which enables the network to discard a datagram that has been traveling the Internet for too long. The **Time to Live** field indicates how long a particular datagram is allowed to live- bounce from router to router-within the system. When a packet is first created, this field is set to its maximum value: 255. Each router along the route from source to destination decreases the number in the Time to Live field by 1. When the value of the Time to Live field reaches zero, a router deletes the datagram. The Time to Live field is analogous to the hop count and hop limit introduced in Chapter Ten. The final two fields of interest, Source IP Address and Destination IP Address, contain, respectively, the 32-bit IP initial source and final destination addresses of the datagram. A 32-bit address uniquely defines a connection to the Internet-usually a workstation or device, although one workstation or device may support multiple Inter- net connections. As an IP datagram moves through the Internet, the Destination IP Address field is examined by a router. The router, using some routing algorithm, forwards the datagram onto the next appropriate communications link. The details of IP addresses will be covered in the section titled "Locating a document on the Internet" in unit 15. The Internet Protocol is definitely one of the most important communications protocols. Because of its simple design, it is relatively

easy to implement in a wide variety of devices. And because of its power, it is capable of interconnecting net- works of virtually any type. Even though the Internet Protocol is powerful, its primary objective is getting data through one or more networks. It is not responsible for creating an error-free, end-to-end connection. To accomplish this, the Internet Protocol relies on the Transmission Control Protocol, or TCP.

3.3 The Transmission Control Protocol

Perhaps one of the most common examples of a transport layer protocol is the other half of the popular TCP/IP protocol. The primary function of Transmission Control Protocol (TCP) is to turn an unreliable network (such as the one created by IP) into a reliable network that is free from lost and duplicate packets. Thus, TCP essentially fills in the holes created by IP. But how can a transport layer protocol make an unreliable network reliable? To make a network more reliable, TCP (as well as most transport layer protocols) performs the following six functions:

- **Create a connection:** The TCP header includes a port address that indicates a particular application on a machine. Used in conjunction, the port address and the IP address identify a particular application of a particular machine. When TCP creates a connection between a sender and a receiver, the two ends of the connection use a port number to identify the particular application's connection. This port number is found within the TCP datagram and is passed back and forth between sender and receiver.

- **Release a connection:** The TCP software can also dissolve a connection after all the data has been sent and received.

- **Implement flow control:** To make sure the sending station does not overwhelm the receiving station with too much data, the TCP header includes a field, called the Window value that allows the receiver to tell the sender to slow down. This Window value is similar in operation to the sliding window used at the data link layer. The difference between the two window operations is that the data link layer's sliding window operates between two nodes or between a workstation and a node, while the TCP window operates between the two endpoints (sender and receiver) of a network connection.

- **Establish multiplexing:** Because the TCP header includes a port number instead of an IP address, it is possible to multiplex multiple connections over a single IP connection. This

multiplexing can be done by creating a different connection that has a port number different from a previous connection.

- **Perform error recovery**: TCP numbers each packet for transmission with a sequence number. As the packets arrive at the destination site, the receiving TCP software checks these sequence numbers for continuity. If there is a loss of continuity, the receiving TCP software uses an acknowledgment number to inform the sending TCP software of a possible error condition.

- **Establish priority:** If the sender has to transmit data of a higher priority, such as an error condition, TCP can set a value in a field (the Urgent Pointer) that indicates that all or a portion of the enclosed data is of an urgent nature. To perform these six functions, TCP places a header at the front of every data packet that travels from sender to receiver, or from one end of the connection to the other. As we did with the IP header, let's examine the more important fields in the TCP header.

3.3.1 TCP Datagram Format

A user is sitting at a workstation and running a network application-for example, an e-mail program. When the user wants to send an e-mail message, the e-mail program takes the e-mail message and passes it to the transport layer of the software. If the e-mail is heading out onto the Internet, the transport layer adds a TCP header to the front of the e-mail message. The information in this header is used by the TCP layer at the receiving workstation to perform one or more of the six transport functions. The TCP header contains the fields shown in Figure 14-4. Let's examine only those fields that assist TCP in performing the six functions listed earlier. The first two TCP header fields, Source Port and Destination Port, contain the addresses of the application programs at the two ends of the transport connection. These port addresses are used in creating and terminating connections. The port number can also be used to multiplex multiple transport connections over a single IP connection. It is important to note the difference between an IP address and a port number. The IP address identifies a device connected to the Internet, while the port number identifies an application on that device.

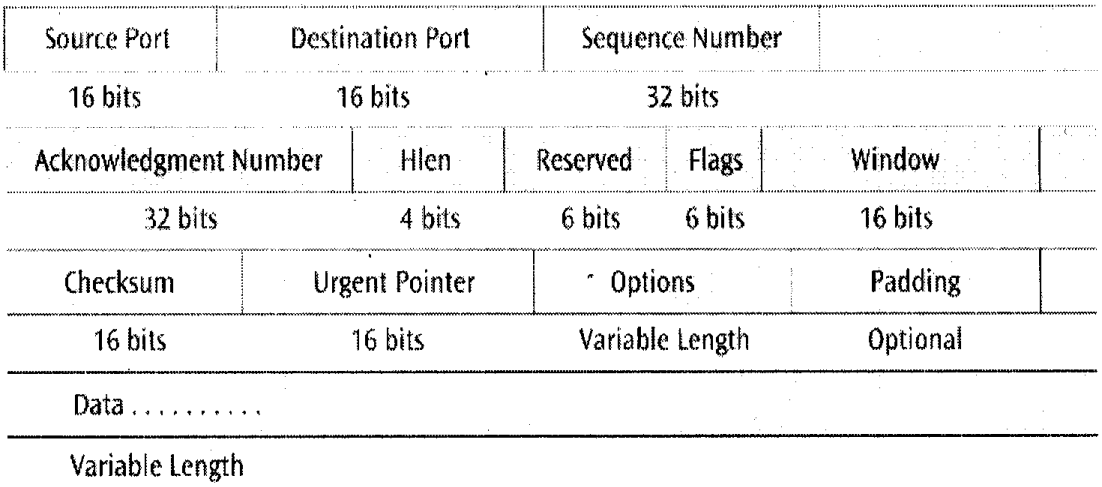


Figure 14.4: Te fields of the TCP header

Working together, the two create what is called a socket-a precise identification of a particular application on a particular device. What if your company has one server that handles both e-mail and FTP connections? The server would have one IP address but two different port numbers- one for the e-mail application and one for the FTP application. Now let's add the fact that this server is more than likely on a local area network, and thus has a network interface card with a unique 48-bit NIC address. Now we have three addresses. The NIC address is used only on the local area network to find a particular device. The IP address is used to move the data packet through the Internet. The port number is used to identify the particular application on a device. The Sequence Number field contains a 32-bit value that counts bytes and indicates a packet's data position within the connection. For example, if you are in the middle of a long connection in which thousands of bytes are being transferred, the Sequence Number tells you the exact position of this packet within that sequence. This field can be used to reassemble the pieces at the receiving workstation and determine if any packets of data are missing. The Window field contains a sliding window value that provides flow control between the two endpoints. If one end of the connection wants the other end of the connection to stop sending data, the Window field can be set to zero. The Checksum field is the next field and provides for a cyclic checksum of the data field that follows the header. The Urgent Pointer is used to inform the receiving workstation that this packet of data contains urgent data. Like its counterpart IP, TCP is a fairly streamlined protocol. Its primary goal is to create an error-free, end-to-end connection across one or more networks. TCP and IP do have their shortcomings, however, and these have typically led to problems such as security weaknesses, quality of service issues, and congestion control.

3.4 Internet Control Message Protocol (ICMP)

As an IP datagram moves through a network, a number of things can go wrong. As a datagram nears its intended destination, a router may determine that the destination host is unreachable (the IP address is wrong, or the host does not exist), the destination port is unknown (there is no application that matches the TCP port number), or the destination network is unknown (again, the IP address is wrong). If a datagram has been on the network too long and its Time to Live value expires, the datagram will be discarded. Also, there could be something wrong with the entire IP header of the datagram. In each of these cases, it would be nice if a router or some other device would send an error message back to the source workstation, informing the user or the application software of a problem. The Internet Protocol was not designed to return error messages, so something else is going to have to perform these operations. The Internet Control Message Protocol (ICMP), which is used by routers and nodes, performs this error reporting for the Internet Protocol. All ICMP messages contain at least three fields: a type, a code, and the first eight bytes of the IP datagram that caused the ICMP message to be generated. A type is simply a number from 0 to n that uniquely identifies the kind of ICMP message, such as invalid port number or invalid IP address. A code is a value that provides further information about the message type. Together, ICMP and IP provide a relatively stable network operation that can report some of the basic forms of network errors.

3.5 User Datagram Protocol (UDP)

TCP is the protocol used by most networks and network applications to create an error-free, end-to-end network connection. TCP is connection-oriented in that a connection via a port number must be established before any data can be transferred between sender and receiver. What if you don't want to establish a connection with the receiver but simply want to send a packet of data? In this case, User Datagram Protocol is the protocol to use. User Datagram Protocol (UDP) is a no-frills transport protocol that does not establish connections, does not attempt to keep data packets in sequence, and does not watch for datagrams that have existed for too long. Its header contains only four fields-Source Port, Destination Port, Length, and Checksum-and it is used by a small number of network services, such as DNS that do not need to establish a connection before sending data.

3.6 Address Resolution Protocol (ARP)

The Address Resolution Protocol is another small but important protocol that is used to support TCP/IP networks. Address Resolution Protocol

(ARP) takes an IP address in an IP datagram and translates it into the appropriate medium access control layer address for delivery on a local area network. As mentioned earlier, every workstation that has a connection to the Internet is assigned an IP address. This IP address is what a packet uses to find its way to its intended destination. There is one problem, however, when a workstation is on, for example, an Ethernet or CSMAICD local area network. Recall that the piece of data that traverses a CSMNCD LAN is called a frame. This frame consists of a number of fields of information, none of which is an IP address. If the frame is supposed to go to a particular workstation with a unique IP address, but the frame does not contain an IP address, how does the frame know where to go? ARP provides the answer to this question. After an IP datagram enters a CSMAICD LAN through a router and before its IP header is stripped off to leave only the CSMNCD frame, ARP broadcasts a message on the LAN asking which workstation belongs to this IP address. The workstation that recognizes its IP address sends a message back saying, "Yes, that is my IP address, and here is my 48-bit CSMNCD (NIC) address. Please forward that IP packet to me via my address." (The 48-bit NIC address is stored in a buffer just in case it will be needed again in the near future.)

3.7 Dynamic Host Configuration Protocol (DHCP)

When a company installs a number of computer workstations and intends to give them all Internet access, it must assign each of them an IP address. This IP address, as we have learned, allows a workstation to send and receive information over the Internet. Two basic methods are used to assign an IP address to a workstation: static assignment and dynamic assignment. With static assignment, somebody sits down at each machine and, using the network operating system, installs an IP address. The person installing the address must then record that IP address on paper somewhere so that the IP address is not accidentally assigned to another machine. What happens if a workstation with an IP address that was statically assigned is then removed from service? Someone has to make sure the IP address assigned to that machine is removed for use in another machine. What if a mistake is made, and the same IP address is assigned to two machines? In this case, there would be an IP address conflict, and a network administrator would have to locate the two machines with the same IP address and rectify the situation. Another issue to consider is related to IP address procurement. If a company has 1000 workstations and each workstation has access to the Internet, then the company has to acquire (typically lease) 1000 IP addresses. This is not cost-effective, particularly if only one-half of the users are on the Internet (using their IP address) at a time. Static assignment of IP addresses can lead to a waste of resources. Dynamic assignment of IP addresses solves these three problems. The most

popular protocol that handles dynamic assignment is Dynamic Host Configuration Protocol (DHCP). When a workstation running the DHCP client software needs to connect to the Internet; the protocol issues an IP request, which prompts the DHCP server to look in a static table of IP addresses. If this particular workstation has an entry, then that IP address is assigned to that workstation. But if there is no entry in the static table, the DHCP server selects an IP address from an available pool of addresses and assigns it to the workstation. The IP address assignment is
temporary, with the default time limit being one hour. DHCP clients may negotiate for a renewal of the assignment if the workstation is still accessing the Internet when the temporary assignment is nearing expiration. Thus, with DHCP, the three problems introduced with static assignment are solved. No individual has to assign IP addresses to workstations, two workstations never get assigned the same IP address, and if only 200 workstations out of 1000 are ever using the Internet at the same time, the company can probably get by with acquiring only 200 IP addresses. Users who dial in to the Internet from home often use DHCP without knowing it. Rather than assigning an IP address to every potential dial-in user, an Internet service provider's DHCP server assigns an IP address during the session creation period. The computer uses this temporary IP address until the user logs off, at which time the address is placed back into the pool, ready for the next dial-in user.

3.8 Network Address Translation (NA T)

Another protocol that is used to assign IP addresses is Network Address Translation (NAT). More precisely, NAT lets a router represent an entire local area network to the Internet as a single IP address. When a user workstation on a company local area network sends a packet out to the Internet, NA T replaces the IP address of the user workstation with a corporate global IP address. In fact, all packets that leave the corporate network contain this global IP address. Thus, the only IP address that anyone sees outside of the corporate network is the one global IP address. If all packets from all workstations leave the corporate network with the same IP address, how do the responses that come back from the Internet get directed to the proper machine? The NAT software maintains a cache listing of all IP packets that were sent out and who sent each packet. When a response comes back, NAT checks the cache to see who originally sent the request. When NAT finds the match, it removes the global IP address, reinserts the user workstation's IP address, and places the packet on the corporate network.

SELF-ASSESSMENT EXERCISE

Explain the following Internet protocol

- NAT
- UDP
- ARP.

An interesting feature of using NAT is that, because the outside world never sees any of the IP addresses used within the corporate network, a level of security has been added. Additionally, the company does not need to use purchased IP addresses on the corporate network. To support this feature, a number of IP addresses have been designated as "phony" IP addresses. When a workstation with a phony IP address issues an Internet request, the NA T software replaces the phony IP address with the corporate global IP address. The use of NAT and phony IP addresses is another way to save money on leasing IP addresses. Home and small business local area networks often use NAT to conserve IP addresses. By keeping track of each request in its internal cache, NAT allows multiple workstations to access the Internet with only one IP address. If the computers are also using DHCP, a home user can dial in to the Internet, have an IP address dynamically assigned, and use that IP address' for the NAT operation. NAT is so useful that many routers now incorporate it as a standard security (firewall) feature.

3.9 Tunneling Protocols and Virtual Private Networks (VPNs)

One of the more serious problems with the Internet is its lack of security. Whenever a transmission is performed, it is susceptible to interception. Retailers have solved part of the problem by using encryption techniques to secure transactions dealing with credit card numbers and other private information. Businesses that want their employees to access the corporate computing system from a remote site have found a similar solution-virtual private networks. A virtual private network (VPN) is a data network connection that makes use of the public telecommunications infrastructure but maintains privacy through the use of a tunneling protocol and security procedures. A tunneling protocol, such as the Point-to- Point Tunneling Protocol (PPTP), is the command set that allows an organization to create secure connections using public resources such as the Internet.

Proposed by CISCO Systems, PPTP is a standard sponsored by Microsoft and other companies. It is an extension of the Internet's Point-to-Point Protocol (PPP), which is used for communication between two computers using a serial connection. The most common example of a

serial connection is a dial-up modem connection between a user's workstation and an Internet service provider: (PPP is usually the standard preferred over an earlier protocol, Serial Line Internet Protocol, or SLIP.) Employees who are located outside a company building can use PPTP to create a tunnel through the Internet into the corporate computing resources. Because this connection runs over the Internet and may be used to transmit confidential corporate data, it must be secure. The security of the connection is often supported by IPsec. IPsec, an abbreviation for IP Security, is a set of protocols developed by the Internet Engineering Task Force to support the secure exchange of data packets at the IP layer. In order for IPsec to work, both sender and receiver must exchange public encryption keys.

Besides being a secure connection, a tunnel is also a relatively inexpensive connection, because it uses the Internet as its primary form of communication. Alternatives to creating a tunnel are acquiring a dial-up telephone line or leasing a telephone line both of which can be more expensive.

ANSWER TO SELF-ASSESSMENT EXERCISE

NAT: NAT lets a router represent an entire local area network to the Internet as a single IP address. Thus it appears all traffic leaving this LAN appears as originating from a global IP address.

UDP: A transport layer protocol used in place of TCP. Where TCP supports a connection-oriented application, UDP is used with connectionless applications. UDP also encapsulates a header onto an application packet but the header is much simpler than TCP.

ARP: ARP translates an IP address into a MAC layer address so a frame can be delivered to the proper workstation.

4.0 CONCLUSION

In this unit we defined the Internet and described protocols (IP, TCP, ICMP, UDP, ARP,DHCP, and NA T) ,that support the Internet. We also identified the relationship between TCP/IP and the above mentioned protocols.

5.0 SUMMARY

To support the Internet, many protocols, such as IP, TCP, ICMP, UDP, ARP, DHCP, and NAT, are necessary. The Internet Protocol (IP) provides a connectionless transfer of data over a wide variety of network

types. Transmission Control Protocol (TCP) resides at the transport layer of a communications model and provides an error-free, end-to-end connection. Internet Control Message Protocol (ICMP) performs error reporting for the Internet Protocol. User Datagram Protocol (UDP) provides a connectionless transport layer protocol, in place of TCP. Address Resolution Protocol (ARP) translates an address into a CSMA/ CD MAC address for data delivery on a local area network. Dynamic Host Configuration Protocol (DHCP) allows a network to dynamically assign IP addresses to workstations as they are needed. Network Address Translation (NAT) allows a network to replace local IP addresses with one, global-type IP address. Tunneling protocols allow a company to create virtual private network connections into a corporate computing system.

6.0 TUTOR-MARKED ASSIGNMENT

1. Why is ARP necessary if every workstation connected to the Internet has a unique IP address?
2. If I dial in to the Internet from home, is it likely that my workstation is using DHCP? Explain

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.

UNIT 5 THE WORLD WIDE WEB

CONTENTS

1.0 INTRODUCTION

Now that we have an understanding of the protocols that enable the Internet function, in this unit we closely look at the World Wide Web, the area of the Internet most familiar to all of us and services available on the Internet.

2.0 OBJECTIVES

At the end of this unit, you should be able:

- Describe the major Internet applications and services
- Discuss the business advantages of the world wide web
- Recognize that the Internet is constantly evolving and that IPv6 and Internet 2 demonstrate that evolution.

3.0 MAIN CONTENT

3.1 The World Wide Web

The World Wide Web (WWW) is a vast collection of electronic documents that are located on many different Web servers and contain text and images that can be accessed by simply clicking links within a browser's Web page. Using a Web browser, you can download and view Web pages on a personal computer. Of all the Internet services, the World Wide Web is probably the one that has had the most profound impact on business. Internet retail sales and service support have

exploded with the use of personal computers and Web browsers. Virtually any and every imaginable product and service is now sold on the Web. In a single day, you can purchase clothing and groceries, buy an airline ticket, select and purchase an automobile, plan a funeral, submit an auction bid on a toy you had as a child, find a new job, rent a videotape, and order pizza for dinner-=-all online. To do this, all you need is a personal computer with a connection to the Internet and a Web browser. The Web pages you download can consist of text, graphics, links to other Web pages, sometimes music and video, and even executable programs. Web pages are created using Hypertext Markup Language (HTML), which can be generated manually with a text-based editor such as Notepad, or through using a Web page authoring tool. A Web page authoring tool is similar to a word processor, except that instead of creating text documents, you create HTML-based Web pages. The Web page authoring tool has a graphical user interface that allows you to enter text and insert graphics and other Web page elements and also arrange these elements on the page using drag-and-drop techniques. Once you've created your Web page, you store it on a computer that contains Web server software and has a connection to the Internet. The Web server software accepts Hypertext Transfer Protocol (HTTP) requests from Web browsers connected to the Internet, retrieves a requested Web page from storage, and returns that Web page to the requesting computer via the Internet. The Hypertext Transfer Protocol, or HTTP, is an application layer protocol. Now that we have an understanding of how documents are passed across the Internet, the question remains: How are Internet documents addressed and found? To answer this question, we need to examine URLs, the Domain Name System, and IP addresses.

3.1.1 Locating a Document on the Internet

When a user is running a browser on a workstation and clicks on a link, the browser attempts to locate the object of the link and bring it across the Internet to the user's workstation. This object can be a document, a Web page, an image, an FTP file, or any of a number of different types of data objects. How does the Internet locate each object? Stated simply, every object on the Internet has a unique English-based address called its Uniform Resource Locator (URL). The Internet, however, does not recognize URLs directly. In order for the Internet to find a document or object, part of the object's URL has to be translated into the IP address that identifies the Web server where the document or object is stored. This translation from URL to IP address is performed by the Domain Name System (DNS). The 1P address itself is not as simple as it appears. The assignment of IP addresses is complex and requires an understanding of the different classes of addresses and of a concept called subnet masking. Let's examine each of these concepts- URLs,

DNS, IP addresses, and subnet masking-further to gain a better understanding of how the Internet finds one document from among billions of documents.

1. Uniform Resource Locator (URL)

In order for users to be able to find something on the Internet, every object located on the Internet must have a unique "address." This address, or Uniform Resource locator (URL), uniquely identifies files, Web pages, images, or any other types of electronic documents that reside on the Internet. When you are using a Web browser and click a link to a Web page, you are actually sending a command out to the Internet to fetch that particular Web page from a specific location that is based on the Web page's URL. All Uniform Resource Locators consist of four parts, as shown in Figure 15.1.

http://cs.depaul.edu/public/utilities/ada/example.htm
1 2 3 4
(a)

ftp://lgatekeeper.dec.com/pub/games/starwars.exe
1 2 3 4
(b)

Figure 15.1: Parts of Uniform Resource Locator

The first part, indicated by a 1 in the figure, is the service type. The service type identifies the protocol that is used to transport the requested document. For example, if you request a Web page, then Hypertext Transfer Protocol (http://) is the service type used to retrieve the Web page, as shown in Figure 15.1(a). If you are requesting an FTP document, then the FTP protocol (ftp://) is used, as in Figure 15.1(b). Other service types include telnet:/ / to perform a remote login, news:// to access a Usenet group, and mailto:// to send an electronic mail message. The second part of the URL, indicated by a 2 in Figure 15.1 is the domain name. This portion of the URL specifies a particular server at a particular site that contains the requested item. In the example in Figure 15.1, cs.depaul.edu is one of the servers supporting-the computer science program at DePaul University. Starting on the right, edu is the top-level domain and indicates that the Web site is an educational site. Other top-level domains are com (commercial), gov (government), mil (military), org (non-profit organization), net (network-based), biz, name, info, pro, museum, aero, and coop. Each country also has its own top-level domain name. For example, Canada is ca and the United Kingdom is uk. The domain name at the next level-called the mid-level domain name-is usually the name of the organization (often a company or

school) or host, such as depaul. Any other lower- level domains are further subdivisions of the host and are usually created by the host. For example, suppose a company named FiberLock applies to the agency handling domain name registration to ask for the mid-level domain name fiberlock. Because FiberLock is a commercial business, its top-level domain name will be .com. If no one else is using fiberlock, the company will be grantedfiberlock.com as its domain name. The company may then add more domain levels, such as www or email, to create entities such as www.fiberlock.com or email.fiberlock.com, which could correspond to the company's Web page server and e-mail server, respectively. The third part. of a URL, labeled 3 in Figure 15-1, is the directory or subdirectory information. For example, the URL http://cs.depaul.edu/public/utilities/ada specifies that the requested item is located in the subdirectory *ada,* under the subdirectory utilities, under the subdirectory public. The final part of the URL, specified by the number 4 in Figure 15.1, is the filename of the requested object. In this case, it is a document titled example. htm. If no filename were specified in the URL, then a default file, such as default.htm or index.html, would be retrieved.

2. Domain Name System (DNS)

When referencing an Internet site, we often refer to its domain name. Computers, however, do not use domain names. They use 32-bit binary addresses called IP addresses. For example, a valid IP address has the following form: 10000000100111000000111000000111. To make IP addresses a little easier for human beings to understand, these 32-bit binary addresses are represented by dotted decimal notation. This dotted decimal notation is created by converting each 8-bit string in the 32-bit IP address into its decimal equivalent. Thus the IP address above becomes 128.156.14.7, as shown here:

```
10000000  10011100  00001110 00000111
   128       156        14        7
```

But even the decimal equivalent to the IP address is not convenient for us humans. Because computers use 32-bit binary addresses and almost all human beings use the domain name form, the Internet converts the binary forms into English-based domain names, and vice versa. To do this, it uses the Domain Name System (DNS), which is a large, distributed database of Internet addresses and domain names. This distributed database consists of a network of local DNS servers, mid-level DNS servers, and higher-level DNS servers. To keep the system manageable, the DNS database is distributed according to the top-level domains: edu, gov, com, mil, and so on.

Converting a domain name into a binary IP address can be simple or complicated. The level of complexity depends on whether or not a local network server on the originating local area network recognizes the domain name. If a network server cannot resolve an address locally, it will call upon a higher authority. A local DNS server will send a DNS message to the next higher DNS server until the address is found, or it is determined that the address does not exist. If the address does not exist, an appropriate message is returned. Consider, first, what happens when a local server recognizes a domain name. When the new domain name appears, an application program, such as a Web browser, calls a library procedure named the resolver. The resolver sends a DNS message to a local DNS server, which looks up the name and returns the IP address to the resolver. The resolver then returns the IP address to the application program. But what happens if the local DNS server does not have the requested information? It may query other local DNS servers, if there are any. The information concerning the existence of other local servers and remote server locations would be kept in a file on the local computer network. If the answer again is no, the local DN,S server tries the next level up-perhaps a mid-level server. If the mid-level server does not recognize the domain name, or there is no mid-level server, the top-level name server for the domain is queried. If the top-level name server does not recognize the domain name, it will either return a "URL Not Found"
message t or go down a level and query a local DNS server. To understand this better, let's look at an example.

Consider a scenario in which a user at cs.waynestate.edu in Wayne State University wants to retrieve a Web page from www.trinity.edu at Trinity University. The message originates from cs. waynestate.edu and goes to the waynestate.edu name server. The waynestate.edu name server does not recognize www.trinity.edu, because the domain name www.trinity.edu is not in a list of recently referenced Web sites. Therefore, the waynestate.edu name server sends a DNS request to edu-server.net. Although edu-server. net does not recognize www.trinity.edu, it does recognize trinity.edu, so it sends a query to trinity.edu. The trinity.edu name server recognizes its own www.trinity.edu and returns the result to edu-server.net, which returns the result to the user's computer, which then inserts the result-that is, the appropriate 32-bit binary IP address- into the browser request.

3. IP Addresses

Now let's talk a little more about the IP address. When IP and IP addresses were created in the 1960s, an IP address belonged to a particular class. This type of addressing was called classful addressing, and it was based, as we shall see shortly, on five different classes. In

approximately 1996, a new form of addressing became available: classless addressing. Although classless addressing is more common today, it is still important to understand how both of these systems work. To start things off, we will examine classful addressing.

As you have already learned, IP addresses are currently 32 bits long. In classful addressing, the addresses are not, however, simple 32-bit integers. Instead, they can consist of three specific pieces of information. The size and value of these three pieces of information depend on the basic form of the address. There are five basic forms of an IP address: Class A, B, C, D, and E (Table 15.1).

As mentioned earlier, each IP address can consist of three parts:

- A 1-,2-,3-, or 4-bit identifier field (also known as a beginning bit pattern).
- A net 10, which indicates a particular network.
- A host ID, which indicates a particular host, or computer, on that network.

As Table 15.1 shows, given a beginning bit pattern of 0, there are 128 Class A addresses, or networks, in existence. Each Class A address can have 16,777,216 hosts, or computers. Clearly, 128 is not very many networks; in fact, all the 128 Class A addresses were assigned a long time ago. Another impractical feature of the Class A address type is the allocation of 16,777,216 computers per network. Attaching 16,777,216 computers to one network is beyond imagination. As we have seen, many local area networks rarely have more than a few hundred computers attached to them. Unfortunately, for this reason, many Class A addresses go unused. Class B addresses allow for 16,384 net IDs, or networks, each supporting 65,536 host IDs- meaning that each of the 16,384 networks can have 65,536 host computers attached to it. Class C addresses allow for 2,097,152 net IDs, or networks, and 256 host IDs. In the case of the Class C address type, the number of host computers allowed is too small to accommodate any networks but the smallest. Class 0 addresses are available for networks that allow multicasting of messages.

Table 15.1: Five basic forms of a 32-bit IP Address

Address Type	**Beginning Bit Pattern**	**Network Address (net ID)**	Host Address (host ID)
Class A	0	128 addresses (7 bits)	**16,777,216 addresses (24 bits)**
Class B	10	16,384 addresses (14 bits)	**65,536 addresses (16 bits)**
Class C	110	2,097,152 addresses (21 bits)	**256 addresses (8 bits)**
Class D	1110	Multicast address	
Class E	**1111**	**Reserved addresses**	

IP multicasting is the capability of a network server to transmit a data stream to more than one host at a time. Consider a scenario in which a company wants to download a streaming video of a training exercise to 20 users sitting at separate workstations. If a server has to transmit 20 individual copies to the 20 workstations (unicast), a very high bandwidth signal will be necessary. If the server could instead multicast one copy of the video stream to the 20 workstations, a much smaller bandwidth would be necessary. To multicast, the server could insert a Class 0 address into the IP datagram, and each of the 20 workstations could tell its IP software to accept any datagrams with this Class 0 address. Although IP multicasting' has some very promising advantages, it suffers from lack of security. Because it is relatively easy for a workstation to tell its IP software to accept a particular Class 0 address, any workstation-even one that doesn't have permission-could potentially receive the multicast.

3.1.2 Creating Web Pages

To transmit a Web page, a Web browser, Web server, and the Internet use the Hypertext Transfer Protocol. HTTP is not, however, used to display the Web page once it reaches the intended destination. To control how a Web page is displayed, another specification is used-Hypertext Markup Language (HTML). Although HTML was the original and is still the most commonly used method for controlling the display of web pages two new forms of HTML have emerged that offer more power and flexibility in Web-page creation-dynamic HTML and extensible Markup Language. Let's examine each of these three specifications and discuss their importance to businesses and consumers. Before we begin, however, the question of why you should know the basic methods of constructing a Web page is important and should be addressed. The Web page is the fundamental element of the World Wide Web. If you understand what is involved in creating a Web page, you can communicate better with individuals who create Web pages, and you can create Web pages yourself if you are ever called upon to do so.

1. Markup Languages

In order to create and display Web pages, some type of markup language is necessary. While there are many types of markup languages, we will briefly introduce three common types here: Hypertext Markup Language (HTML), dynamic Hyper- text Markup Language (dynamic HTML), and eXtensible Markup Language (XML). Hypertext Markup Language (HTML) is a set of codes inserted into a document that is intended for display on a Web browser. The codes, or markup symbols, instruct the browser on how to display a Web page's text, images, and

other elements. The individual markup codes are often referred to as tags and are surrounded by brackets « ». Most HTML tags consist of an opening tag, followed by one or more attributes, and a closing tag. Closing tags are preceded by a forward slash (/). Attributes are parameters that specify various qualities that an HTML tag can take on. For example, a common attribute is HREF, which specifies the URL of a file in an anchor tag «A».

Figure 15.2 shows an example of a small HTML file on the left, and a list of descriptive comments on the right, figure 15.3

<HTML>	1. Begins every HTML document
<HEAD>	2. Begins the head section
<TITLE>DePaulUniversity Home Page </TITLE>	3. Title that appears on browser title bar
</HEAD>	4. Closes the head section
<BODY>	5. Begins the body section
This is the first line 	6. A text line followed by a line break
<P>Start a new paragraph</P>	7. Begins a paragraph
<H1>First Heading</H1>	8. Level 1 head
<H2>A second-level Heading</H2>	9. Level 2 head
<HR>	10. Inserts a horizontal rule
<B>Bold this line</B> 	11. Bold text
<I>Italicize this line</I> 	12. Italicized text
<IMG SRC="\images\banner.gif">	13. Inserts an image
<A HREF="http://www.cs.depaul.edu> DePaul CS Page</A>	14. Link to another Web page
</BODY>	15. Closes the body section
</HTML>	16. Ends every HTML document

Figure 15.2: Example of an HTML file

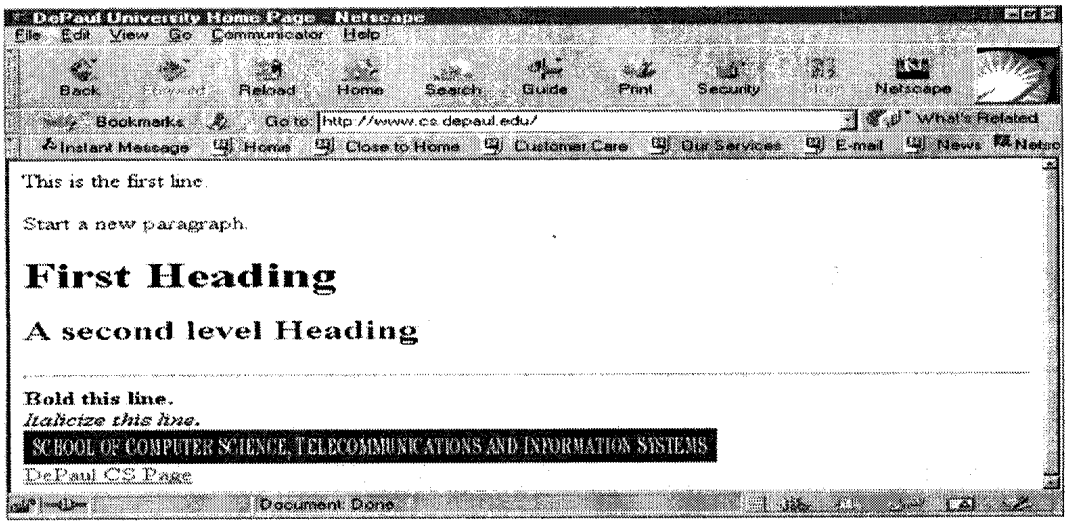


Figure 15.3: Web page generated by example figure 15.2

Each line in the file includes an HTML tag, which tells the browser how to display text or images. Although HTML is relatively simple to use, it suffers from a number of short-comings. One of the most serious

shortcomings is HTML's inability to allow a user to place text or an image at a precise position on the Web page. With HTML, it is also difficult for a user to specify and switch between sets of font styles and colors easily and quickly. To better support these types of functionality, dynamic HTML was created. Rather than functioning as a single specification, dynamic HTML. (DHTML) is a collection of newer markup tags and techniques that can be used to create more flexible and more powerful Web pages. HTML pages are simple, static text documents that browsers read, interpret, and display on the screen. In contrast, dynamic HTML pages have additional functionality that allows them to be, among other things, interactive.

Stated simply, dynamic HTML can grab any element on a Web page and change its appearance, content, or location on the page dynamically. Features of dynamic HTML include:

- active pop-ups (when a user moves his or her mouse pointer over an area on a page, additional text can be made to appear on the page)
- the live positioning of elements or layers (in other words, the placement of an object on a Web page can be specified by using x,y coordinates)
- cascading style sheets (CSS), which allow a Web page author to incorporate multiple styles (fonts, styles, colors, and so on) in an individual HTML page.

Another member of this family of markup languages is eXtensible Markup Language (XML). Whereas HTML determines only how the content of a document is to be displayed by a Web browser, XML also defines the content of the document. When an XML document is passed between two entities, the document contains the data and a detailed description of the data. This dual construction eliminates the need for sending additional earlier documents that describe the format of the data. The syntax of XML is fairly similar to that of HTML; however, there are a number of very important differences. First, XML is extensible, which means a user can define his or her own tags. You can create tags that define entire data structures. For example, the entries in an auto parts catalog might require tags such as <PARTNAME>, <MAKE>, <MODEL>, <YEAR>, <DESCRIPTION>, and <P ARTCOST>.

Second, XML is much less forgiving than HTML. XML documents have many more precise rules for the creation of tags and the elements within a document. For example, all tags must be properly nested, all attribute values must have quotation marks around them, and all tags with empty content must end in "...!>". Unlike an HTML document, a

document created in XML will not be displayed if the coding contains a mistake. A mistake in HTML coding is often ignored by the browser, and the rest of the document is displayed.

3.2 Internet Services

When the Internet came into existence as the ARPANET, most people used it for e-mail, file transfers, and remote logins. In addition to studying those services, let's examine several of the more popular services that the Internet provides today.

1. **Electronic mail (e-mail):**

Electronic mail, or e-mail, is the computerized version of writing a letter and mailing it at the local post office. Many people are so committed to using e-mail that if it were taken away tomorrow, some serious social and economic repercussions would be felt throughout the Nigeria and the rest of the world. Many commercial e-mail programs are inexistence, as well as a number of free ones that can be downloaded from the Internet. Although each e-mail program has its own unique feel and options, most offer the following services:

- Creating an e-mail message
- Sending an e-mail message to one recipient, multiple recipients, or a mailing list
- Receiving, storing, replying to, and forwarding e-mail messages
- Attaching a file, such as a word-processing document, a spreadsheet, an image, or a program to an outgoing e-mail message

Most e-mail systems consist of two parts: (1) the user agent, which is the portion of the e-mail program that allows a user to create, edit, store, and forward e-mail messages and (2) the message transfer agent, which is the portion of the e-mail program that prepares and transfers the e-mail message. Each transmitted e-mail also consists of two basic components: an envelope, which contains information describing the e-mail message, and the message, which is the contents of the envelope. Most messages consist of plain text and are written in simple ASCII characters. What if you want to send (or attach) a nontext-based item, such as a spreadsheet, a database, or an image? The e-mail program then creates a Multipurpose Internet Mail Extensions (MIME) document and attaches it to the e-mail message. Once the e-mail and optional attachment have been created, it is time to transmit the message. The Simple Mail Transfer Protocol (SMTP) is an Internet protocol for sending and receiving e-mail and is used to perform the transfer. To send the e-mail message, the source computer establishes a TCP

connection to port number 25 (typically) on the destination computer. The destination computer has an e-mail daemon - a program that is always running in the background and waiting to perform its function- that supports the SMTP protocol. The e-mail daemon watches port 25, accepts incoming connections, and copies messages to the appropriate mailbox. The European and Canadian equivalent to SMTP is the XAOO protocol. How many times do you receive an e-mail message even though your machine is not turned on? When you turn your computer on and run your e-mail program, a message informs you that you have n new e-mail messages waiting. What software performs this operation? Post Office Protocol version 3 (POP3) is the soft- ware that allows the user to save e-mail messages in a server mailbox and download them when desired from the server. POP3 is useful if you do not have a permanent connection to a network and must dial in using a temporary Internet connection. POP3 will hold your e-mail messages until the next time you dial in and access your mailbox. Thus, POP3 software is commonly found on mobile laptop computers or home computers without permanent network connections, but also on computer systems that have permanent connections. An alternative to POP3 is the more sophisticated Internet Message Access Protocol (IMAP). IMAP (the latest version is IMAP4) is a client/server protocol in which e-mail is received and held for you at your Internet server. You can view just the heading of the e-mail or view the sender of the message and then decide if you want to download the mail. You can also create and manipulate folders or mailboxes on the server, delete old e-mail messages, or search or certain parts of an e-mail message.

2. File Transfer Protocol (FTP)

The File Transfer Protocol, or FTP, was one of the first services offered on the Internet. Its primary functions are to allow a user to download a file from a remote site to the user's computer and to upload a file from the user's computer to a remote site. These files could contain data, such as numbers, text, or images, or executable computer programs. Although the World Wide Web has become the major vehicle for retrieving text and image based documents, many organizations still find it useful to create an FTP repository of data and program files. Using a Web browser or specialized FTP software, you can easily access an FTP site. If you desire privacy and wish to restrict access to an FTP site, the site can be designed to require a user ill and password for entry. To access an FTP site via a Web browser and download a file, you need at least three pieces of information. First, you must know the name of the FTP site. Second, you must know the name of the directory or “subdirectory in which to look for the file. Third, you must know the name of the file that you want to download, Thus, downloading FTP files is not a "browsing" activity. You must have a good idea of what you are looking

for and where it is located. As an example, let's say you are reading an article in a computer magazine, and the article describes a free utility program that organizes your time, keeps you on schedule, and helps you make new friends. You must have this program, but it is on some FTP site in the middle of nowhere. No problem. All you need is the utility program's URL and a Web browser. As explained in an earlier section, every document on the Internet has a unique URL.

3. Remote login (Tel net)

Remote login, or Telnet, is a terminal emulation program for TCP/IP networks, such as the Internet, that allows users to log in to a remote computer. The Telnet program runs on your computer and connects your workstation to a remote server on the Internet. Once you are connected to the remote server or host, you can enter commands through the Telnet program, and those commands will be executed as if you were entering them directly at the terminal of the remote computer. There are three reasons for using a Telnet program. First, Telnet allows a user to log in to a personal account and execute programs no matter where he or she is. For example, you might have a computer account at two different companies or two different schools. Although you may be physically located at one site, you can use Telnet to log in to the other computer site. Through this login, you can check your e-mail or run an application. Second, remote login allows a user to access a public service on a remote computer site. Third, Telnet enables a network administrator to control a network server and communicate with other servers on the network. This control can be performed from a remote distance, such as another city or the network administrator's home.

4. Voice over IP

One of the newer services that is attracting the interest of companies and home users alike is the sending of voice signals over an IP-based network such as the Internet. The practice of making telephone calls over the Internet has had a number of different names, including packet voice, voice over packet, voice over the Internet, Internet telephony, and Voice over IP (VoIP). But it appears the industry has settled on the term "Voice over IP," in reference to the Internet Protocol (IP), which controls the transfer of data over the Internet. Whatever its title, Voice over IP has emerged as one of the hottest Internet services and has certainly drawn the attention of many companies. Because this technology is relatively new, the implementation of Voice over IP varies greatly, depending on your level of involvement. Due to the allure of a potentially large market, man' companies are offering complete packages that can be installed on a local area network system. These packages involve large amounts of equipment, such as VoIP servers,

high-speed switches, special IP-enabled telephones, and routers that can direct telephone calls. At the other end of the spectrum, some telecommunications companies offer a service only for individual/small business users and home users. These services use traditional telephone lines and telephones but require special 'adapters (that convert traditional telephone signals to IP packets and back) to be inserted between the line and the phone. When a reliable and high-quality telephone system is already available, why explore new technology that provides the same service? One of the earlier advantages of Voice over IP was simply related to the fact that long-distance calls, especially overseas calls, cost money, while sending data-or voice-over the Internet is essentially free. As it turns out, this advantage has become less important for many corporate users. One reason for this is that long-distance telephone rates have dropped significantly over the years; another is that the call quality of long-distance VoIP is usually much worse than conventional long-distance. But many companies are now finding other, more important advantages in being able to treat voice data like other forms of data. For one, if both voice and data can travel 'over the same network, companies would realize savings in both equipment and infrastructure. This would contribute to yet another significant advantage: the need for separate telephone management personnel, and local area network management personnel would be reduced or even eliminated.

Voice over IP has a number of disadvantages as well. The statement that sending data over the Internet is essentially free is misleading. Nothing, of course, is free. All Internet users must pay an Internet service provider for access, the interconnecting phone line, and any necessary hardware and software. A second, and more important disadvantage, is that transmitting voice over a corporate network can be demanding on the network's resources. If the current corporate network system is straining to deliver data, adding voice to this system can cause severe service problems. These service problems can be compounded because voice systems require networks that can pass the voice data through in a relatively small amount of time. A network that delays voice data by more than 20 milliseconds from end to end will introduce a noticeable echo into the transmission. If the delay becomes longer than 250 milliseconds (that's only a quarter of a second), the system will be basically unusable.

5. Listservs

A listserv is a popular software program used to create and manage Internet mailing lists. Listserv software maintains a table of e-mail addresses that reflects the cur- rent members of the listserv. When an individual sends an e-mail to the listserv address, the listserv sends a

copy of this e-mail message to every e-mail address stored in the listserv table. Thus, every member of the listserv receives the e-mail message. Other names for listserv or other types of listserv software include *mailserv, majordomo, and almanac*. To subscribe to a listserv, you send a specially formatted message to a special listserv address. This address is different from the address for sending an e-mail to all the listserv members.

6. Streaming audio and video

Streaming audio and video involves the continuous download of a compressed audio or video file, which can then be heard or viewed on the user's workstation. Typical examples of streaming audio are popular and classical music, live radio broadcasts, and historical or archived lectures, music, and radio broadcasts. Typical examples of streaming video include pre-recorded television shows and other video productions, lectures, and live video productions. Businesses can use streaming audio and video to provide training videos, product samples, and live feeds from corporate offices, to name a few examples. To transmit and receive streaming audio and video, the network server requires the space necessary to store the data and the software to deliver the stream, and the user's browser requires a streaming product such as RealPlayer from Real Networks to accept and display the stream. All audio and video files must be compressed, because an uncompressed data stream would occupy too much bandwidth and would not travel in real time. Real-Time Protocol (RTP) and Real-Time Streaming Protocol (RTSP) are two common application layer protocols that servers and the Internet use to deliver streaming audio and video data to a user's browser.

7. Instant messaging

More formally titled Instant Messaging and Presence, instant messaging (1M) allows a user to see if people are currently logged in on the network and, if they are, to send them short messages in real time. Many users, especially those in the corporate environment, are turning away from e-mail and using instant messaging as a means of communicating. The advantages of 1M include real-time conversations, server storage savings (because you are not storing and forwarding instant messages, as you would e-mails), and the capability to carryon a "silent" conversation between multiple parties. Service providers such as AOL, Microsoft's MSN, and Yahoo!, as well as a number of other software companies, incorporate instant messaging into their products.

3.3 The Internet and Business

Throughout this unit, numerous references have been made to how the Internet affects a business. The term that has come to represent a business's commercial dealings over the Internet is e-commerce. E-commerce can also be defined as the buying and selling of goods and services via the Internet, and in particular via the World Wide Web. To understand the important issues and trends associated with this intersection between technology and business, let's sub-divide e-commerce into the following four areas:

- E-retailing-E-retailing is the electronic selling and buying of merchandise using the Web. Virtually every kind of product imaginable is available for purchase over the Web. Sophisticated Web merchandisers can track their customers' purchasing habits, provide online ordering, allow credit card purchases, and offer a wide selection of products and prices that might not normally be offered in a brick-and-mortar store.
- Electronic data interchange (ED I)-Electronic data interchange (ED I) is the electronic commercial transaction between two or more companies. For example, a company wishing to purchase a large number of cell phones may send an electronic request to a number of cell phone manufacturers. The manufacturers may bid electronically, and the company may accept a bid and place an order electronically. Bank funds will also be transferred electronically between companies and their banks.
- Micro-marketing-Micro-marketing is the gathering and use of the browsing habits of potential and current customers, which is important data for many companies. When a company knows and understands its customers' habits, it can target particular products to particular individuals.
- Internet security- The security systems that support all Internet transactions are also considered an important part of e-commerce.

3.4 The Future of Internet

The Internet is not a static entity. It continues to grow by adding new networks and new users every day. People are constantly working on updating and revising the Internet's myriad components. The driving force behind all these changes, as well as all Internet protocols, is the self-regulating government of the Internet Based on a committee structure, this government consists of many committees and groups, including:

- The Internet Society (ISO C)-A volunteer organization that decides the future direction of the Internet

- The Internet Architecture Board (IAB)-A group of invited volunteers that approves standards
- The Internet Engineering Task Force (IETF)-A volunteer group that discusses operational and technical problems
- The Internet Research Task Force (IRTF)-The group that coordinates research activities
- The World Wide Web Consortium (W3C)-A Web industry consortium that develops common protocols and works to ensure interoperability among those protocols.
- Many more steering groups, research groups, and working groups

One of the biggest changes to affect the Internet will be the adoption of a new version of the Internet Protocol, version IPv6. Currently the Internet is using IPv4, which was the version presented earlier in unit 14. (In case you are wondering, IPv5was not a version open to the public but one used for testing new concepts.) Let's take a closer look at the details of version 6 and see how they compare to the current version, IPv4.

3.4.1 IPv6

The next version of the Internet Protocol. Main features include:

- Simpler header
- 128-bit IP addresses
- Priority levels and quality of service parameters
- No fragmentation.

Figure 15.4 shows the fields in the IPv6 header.

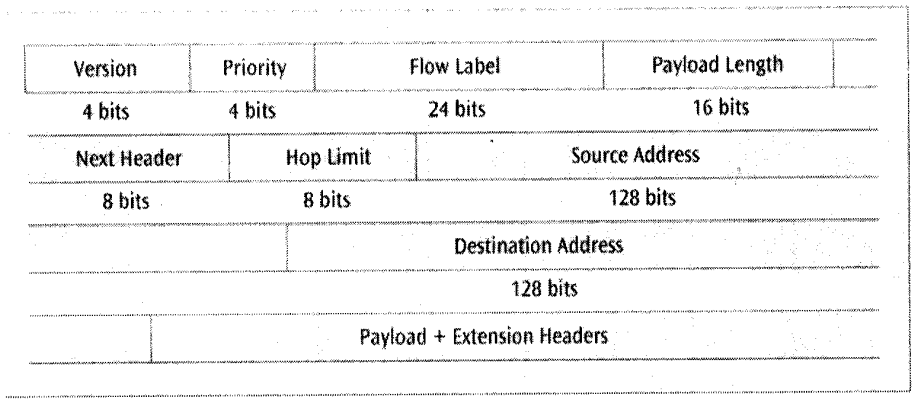


Figure 15.4: IPv6 header

SELF-ASSESSMENT EXERCISE

i. What is an instant messaging?
ii. E-commerce consists of four major area. Identify and explain fully the four major areas.

ANSWER TO SELF-ASSESSMENT EXERCISE

Instant Messaging

Allows a user to see if people are currently logged in on the network and then send short messages in real time. Consumes less resources than e-mail, and faster. Numerous Internet service providers .such as America Online, Yahoo!, and Microsoft MSN offer instant messaging.

Four major areas of e-commerce

- e-retailing
- Electronic Data Interchange (EDI)
- Micro-marketing
- Electronic security.

4.0 CONCLUSION

This unit described WWW tools to transmit and receive Web pages. These tools include; HTTP, HTML, XML etc. We also demonstrated how a document on the Internet using URL. The unit also highlighted the protocols the support the operations of e-mail such as POP3, IMAP ,SMTP, etc. We concluded this unit with the future of the Internet including IPv6.

5.0 SUMMARY

The World Wide Web is a vast collection of electronic documents containing text, and images that can be accessed by simply clicking a link within a browser's Web page. The browser uses HTTP to transmit and receive Web pages and HTML to display those Web pages. HTML and dynamic HTML are markup languages used to define how the contents of a Web page will be displayed by a Web browser. XML is a set of rules for defining your own markup language. By creating your own markup language, you can create a document that contains both the data and a definition of that data. To locate a document on the Internet, you usually refer to its Uniform Resource Locator (URL). A URL uniquely identifies each document on every server on the Internet. One component of the URL is the site address of the requested document. Each site address consists of a network ID and a host or device ID. The

site address is converted to a 32-bit IP address by the Domain Name System (DNS). The Internet consists of many commonly used network applications. Electronic mail, or e-mail, is a standard requirement for most business operations and can transfer standard text messages and include MIME-encoded attachments. Protocols such as Simple Mail Transfer Protocol (SMTP), Post Office Protocol version 3 (POP3), and Internet Message Access Protocol (!MAP) support the operations of e-mail. File Transfer Protocol (FTP) is useful for uploading or downloading files across the Internet. Remote login using Telnet allows an individual to log in to a remote computer site and perform operations as if the user were physically located at the remote site. VoIP (Internet telephony) offers an inexpensive alternative to long-distance calling, but with questionable quality. The Internet was not designed to transfer real- time data, which is a capability that is necessary to support interactive voice. Nevertheless, many businesses are embracing VoIP internally as a way to deliver combined voice and data applications. A listserv is a popular software program used to create and manage Streaming audio and video are the continuous downloading of a compressed audio or video file, which is then heard or displayed on the user's workstation. Streaming audio and video require support protocols such as Real-Time Protocol (RTP) and Real-Time Streaming Protocol (RTSP). Instant messaging is growing in popularity as a way to maintain real-time communications between multiple users. E-commerce, a rapidly growing area of the Internet, is the buying and selling of goods and services electronically. Many companies are investing heavily in e-commerce in hopes that it will increase their market share and decrease their costs.

6.0 TUTOR-MARKED ASSIGNMENT

Some of the new protocols, such as IPv6, are not including any kind of error-detection scheme on the data portion of the packet. What is the significant of this trend?

7.0 REFERENCES/FURTHER READINGS

Curt, M W (2007). *Data Communications and Computer Network.* A Business User's Approach Fourth Edition. Bob Woobury, Canada.

Aftab, A (2003). Data Communication Principles for Fixed and Wireless Networks. Kluwer Academic Publisher, USA.