EIILM UNIVERSITY
SIKKIM

COMPUTER NETWORKS AND DATA COMMUNICATION

Subject: COMPUTER NETWORKS AND DATA COMMUNICATION **Credits: 4**

SYLLABUS

Data Communications Fundamentals

Communication model, computer communications architecture, OSI model, TCP/IP, standards, signals, analog and digital transmission, transmission media, line configuration, topologies, data communications codes, error detection and correction methods.

Data Modem & Modulation

Data encoding methods, analog to digital, digital to analog etc., data modulation methods: ASK, FSK, PSK, QAM, M-ary systems. Data modems: modulation interface, operation on 2 wire, 4 wire and dial up lines.

Data Communications Methods

Data communication interface, line control unit, UART, USRT, Serial interface, terminal types.

Data Link Control & Multiplexing

Flow control, error detection & control, IBM Bisync protocol, SDLC, HDLC, HDLC line procedures, time division multiplexing (TDM), FDM, line encoding, carriers.

Data Communication Networking

Switched networks, circuit switching, packet switching, broadcast networks, LAN, WAN topologies, ATM & Frame relay, cell relay.

Suggested Readings:

1. Data Communications and Networking, Fourth Edition by Behrouza A. Forouzan,TMH
2. Computer Networks,A.S.Tanenbaum,4th edition,Pearson education
3. Introduction to Data communications and Networking,W.Tomasi,Pearson education
4. Data and Computer Communications,G.S.Hura and M.Singhal,CRC Press,Taylor and Francis Group
5. An Engineering Approach to Computer Networks-S.Keshav,2nd Edition,Pearson Education
6. Understanding communications and Networks,3rd Edition, W.A.Shay,Cengage Learning

COURSE OVERVIEW

This course presents an overview of data communications and computer networks, including network hardware and network software, as well as reference models, example networks, example data communication services and network standardization. The ISO and the Internet (TCP/IP) computer network models are discussed in details. The course covers each network layer in details, starting from the Physical layer to the Application layer. It also includes an overview of network security topics. The physical network layer covers a guided and wireless transmission media, the telephone system, and Internet access networks, such as DSL, ISDN and Cable modem. It also presents and overview of ATM networks. The Data Link Layer covers design issues, error detection and correction, elementary data link protocols, sliding window protocols and example data link protocols. The Medium Access Sub-layer covers the following subjects: the channel allocation problem, multiple access protocols, IEEE standard 802 for LAN's and MAN's, bridges, high-speed and wireless LAN's. However, with phenomenal advances in fields like Local Area Networks (LANs), client/server architecture and the advent of high speed technologies such as ADSL, Frame Relay and ATM, the network has become inseparable from the computers it links together.

The functions of the Network Layer are explained in context of design issues, routing algorithms, congestion control and internetworking. The IPv4 and IPv6 protocols are covered in detail. The Transport Layer includes the transport service, elements of transport protocols, and a detailed presentation of TCP and UDP internet transport protocols. The Application Layer issues cover major Internet applications, such as, DNS (Domain Name System), Electronic Mail, FTP and the World Wide Web. It also includes an introduction to the main network security principals.

It is now feasible to connect the best of desktops, minis and mainframes, and to bring them in a common orbit with the objective of enhancing organizational productivity. The dramatic improvement in microprocessor power per dollar and the pressure towards globalization are also accelerating the trend towards distributed and desktop oriented computing. The ongoing construction of the digital Information Superhighway, the escalating importance of the Internet, the rush to build secure and reliable Internet applications for electronic commerce, and the emergence of new computing paradigms such as network computing have created a new vista of opportunities and challenges for Information Technology (IT) managers.

COMPUTER NETWORKS AND DATA COMMUNICATION

CONTENT

COMPUTER NETWORKS AND DATA COMMUNICATION

CONTENT

LESSON 1

Today's Topics

- Communication and data communication
- A transmission system
- The basic data communication model

Objectives

Upon completion of this Lesson, you should be able to:

- Explain how data communication happens
- Explain the basic data communication model

1.1 Communication

What is communication?

You are right! It is just exchange of information between two humans.

Data communication is similarly exchange of information between two computers.

We will first see a basic data communication model before we proceed any further.

1.2 Basic Communication Model

Communication defined more technically can be described as the conveyance of a **message** from one entity, called the **source or transmitter**, to another, called the **destination or receiver**, via a **channel** of some sort.

A very basic example of such a communication system is conversation; people commonly exchange verbal messages, with the channel consisting of waves of compressed air molecules at frequencies, which are audible to the human ear. This is depicted in Figure 1.1.

The conveyance of a message could be followed by a reciprocal **response** message from the original destination (now a source) to the original source (now a destination) to complete one cycle in a dialogue between corresponding entities. Depending on the application or need for the information exchange, either atomic one-way transactions or a two-way dialogue could be appropriate.

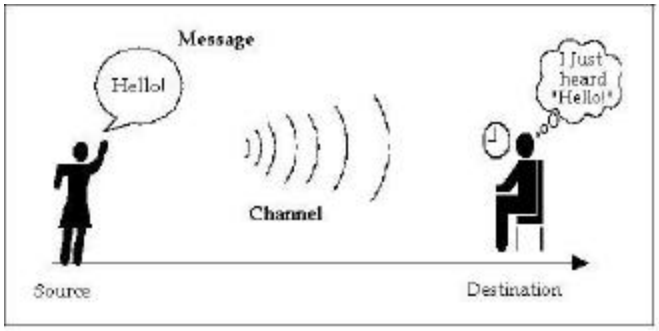


Figure 1.1: Basic Communication Model

Now how do you make out that the message has correctly reached the destination?

In our example of conversation between two people- the receiver would usually acknowledges the receipt of the message either by nodding his head or with some reply like 'I understand' . This acknowledged form of dialogue is the basis of reliable communication - somehow the source must get feedback that the destination correctly received the message.

The basic model explained above wherein originator and destination are human beings can be generalized as in Figure 1.2. To suit the data communication between two systems exchanging information between them.

1.3 Data Communication

So in data communication who is the sender and the receiver? Let us try to understand this along with a few definitions.

1.3.1 Source

Device that generates the data to be passed on to the **Destination** device. It could be a user computer trying to make a query to a server computer.

1.3.2 Transmitter

If the data generated by the Source device has to be transmitted through **Transmission Channel** or **Transmission System** then it has to be presented in a form that is acceptable to the Transmission system. This job is done by the **Transmitter**. For example, a modem takes a digital bit stream from the attached computer and transforms that stream of bits into an analog signals which can be handled by the telephone network.

Now that you know that at the sending end the data is converted to a form suitable for transmission, naturally what should be done at the receiving end?

1.3.3 Receiver

This receives the signal from the transmission system and converts it into a form that is suitable to the destination device. For example, a modem accepts analog signal from a transmission channel and transforms it into digital bit stream.

1.3.4 Destination

Device to which the source device sends data.

1.3.5 Transmission System

What is the role of the transmission system?

This can be a single transmission line connecting the two systems communicating or a complex network to which numerous communicating systems are connected.

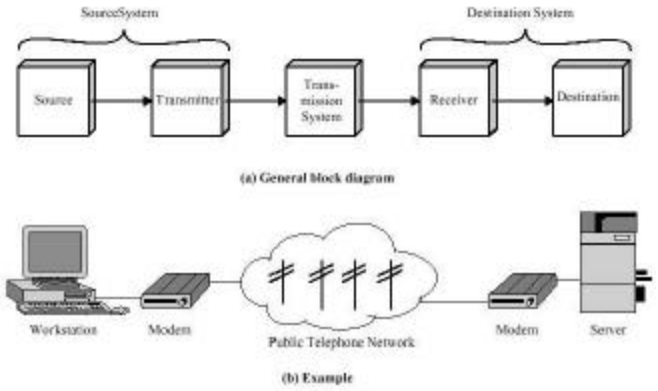


Figure 1.2 Simple Data Communication model

1.4 Data Communication Model

Figure 1.3 provides a new perspective on the data communication model of Figure1.2. Let us trace through the details of this figure using electronic mail as an example.

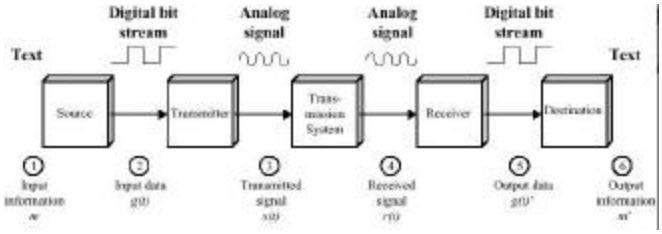


Figure 1.3 Data Communication model

Consider that the source input device and transmitter are components of a personal computer. The user of the PC wishes to send a message to another user—for example, “Have a nice day” (m). The user activates the electronic mail package on the PC and enters the message via the keyboard. The character string is briefly buffered in main memory. We can view it as a sequence of bits (g) in memory. The personal computer is connected to some transmission medium, such as a local network or a telephone line by an I/O device (transmitter), such as a local network transceiver or a modem. The input data are transferred to the transmitter as a sequence of voltage shifts g(t) representing bits on some communications bus or cable. The transmitter is connected directly to the medium and converts the incoming stream [g(t)] into a signal [s(t)] suitable for transmission. The transmitted signal s(t) presented to the medium is subject to a number of impairments, before it reaches the receiver. Thus, the received signal r(t) may differ to some degree from s(t). The receiver will attempt to estimate the original s(t), based on r(t) and its knowledge of the medium, producing a sequence of bits g’(t). These bits are sent to the output personal computer, where they are briefly buffered in memory as a block of bits (g’). In many cases, the destination system will attempt to determine if an error has occurred and, if so, will cooperate with the source system to eventually obtain a complete, error-free block of data. These data are then presented to the user via an output device, such as a printer or a screen. The message (m’), as viewed by the user, will usually be an exact copy of the original message (m).

Now consider a telephone conversation. In this case, the input to the telephone is a message (m) in the form of sound waves. The sound waves are converted by the telephone into electrical signals of the same frequency. These signals are transmitted without modification over the telephone line. Hence, the input signal g(t) and the transmitted signal s(t) are identical. The signal s(t) will suffer some distortion over the medium, so that r(t) will not be identical to s(t). Nevertheless, the signal r(t) is converted back into a sound wave with no attempt at correction or improvement of signal quality. Thus **m’** is not an exact replica of **m**. However, the received sound message is generally comprehensible to the listener.

1.5 Data Communication- its Importance

This is the era of global economy. Data communications has become an extremely important aspect of this modern era. Business documents and information has to be exchanged over geographical boundaries as transactions happen. For enabling data communications a combination of software and hardware is essential. The rapid growth in these two fields has made rapid advances in data communication possible.

1.6 Characteristics Essential to a Data Communication System

There are in short three characteristics essential to any data communication system

1. Correct Delivery: When a sender transmits data for an intended recipient, the data must reach only the intended recipient and not another.
2. Accurate delivery: The data sent must be received in the same form as the one in which it was sent. There must not be ant sort of alterations to it in transit.
3. Timely delivery: The data must travel from the sender to the receiver in a finite amount of time.

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- Data and Network CommunicationMiller, Vikas publishing house
- New Horizons in Data Acquisition and Computer Interfaces, Omega Press LLC
- www.ctr.kcl.ac.uk/lectures

LESSON 2

Today's Topics

- Protocols
- Protocol architecture
- OSI Model
- TCP/IP model

Objectives

Upon completion of this Lesson, you should be able to:

- Define the term protocol
- Define OSI model and its layers
- Define TCP/IP model

So now that you have the basic concept of how data is communicated over a network you are ready to proceed further with the network architecture:

2.1 Network Architecture

is a fancy term for the way that networking products are constructed. Networking hardware and software is implemented on systems via a mechanism called **network or communication architecture**. **Communication architecture** is the layering of software based upon the functionality of each layer.

In discussing computer communications architecture the concept of protocols is paramount.

2.2 What is a Protocol? Try Answering This Before you go on!

A **network protocol** is a set of rules for communicating between computers. Protocols govern format, timing, sequencing, and error control. Without these rules, the computer cannot make sense of the stream of incoming bits.

A **protocol** is used for communication between entities in different systems. For two entities to communicate successfully, they must speak the same language. What is communicated, how it is communicated, and when it is communicated must conform to some mutually acceptable convention or protocol between the entities involved. The key elements of a protocol are

- **Syntax** Includes such things as data format and signal levels.
- **Semantics** Includes control information for coordination and error handling.
- **Timing** Includes speed matching and sequencing.

Having introduced the concept of a protocol, we can now introduce the concept of protocol architecture.

Two protocol architectures have served as the basis for the development of interoperable communications standards: the **TCP/IP protocol suite and the OSI reference model.** TCP/IP is the most widely used interoperable architecture, and OSI has become the standard model for classifying communication functions.

Therefore let us go briefly into the description of both

2.3 The OSI (Open System Interconnection) Protocol Architecture

The **Open System Interconnection (OSI)** model includes a set of protocols that attempt to define and standardize the data communications process. The OSI protocols were defined by the International Organization for Standardization (ISO).

The OSI model is not a single definition of how data communications actually takes place in the real world. Numerous protocols may exist at each layer. The OSI model states how the process should be divided and what protocols should be used at each layer. If a network vendor implements one of the protocols at each layer, its network components should work with other vendors' offerings. The OSI model has seven layers.

Application
Presentation
Session
Transport
Network
Data Link
Physical

1. The **Physical layer** provides the electrical and mechanical interface to the network medium (the cable). This layer gives the data-link layer (layer 2) its ability to transport a stream of serial data bits between two communicating systems; it conveys the bits that move along the cable. It is responsible for making sure that the raw bits get from one place to another, no matter what shape they are in, and deals with the mechanical and electrical characteristics of the cable.
2. The **Data-Link layer** handles the physical transfer, framing (the assembly of data into a single unit or block), flow control and error-control functions over a single transmission link; it is responsible for getting the data packaged for the Physical layer. The data link layer provides the network layer (layer 3) reliable information-transfer capabilities. The data-link layer is often subdivided into two parts-Logical Link Control (LLC) and Medium Access Control (MAC)-depending on the implementation.
3. The **Network layer** provides for the transfer of data in the form of packets across the communication networks. It establishes, maintains, and terminates logical and physical connections across multiple interconnected networks. A key aspect of this transfer is the routing of packets from the source to the destination machine typically traversing a

number of transmission links and network nodes where routing is carried out. Routing is the process by which a path is selected out of many available paths to the destination so that data packet reaches the destination fast, efficiently, reliably as required. This function makes the network most complex layer in the reference model. Also network layer is responsible for translating logical addresses, or names, into physical (or data-link) addresses. It provides flow-control functions across the computer-network interface.

4. The **Transport layer** ensures data is successfully sent and received between two end nodes. If data is sent incorrectly, this layer has the responsibility to ask for retransmission of the data. Also it ensures data are passed onto the upper layers in the same order in which they were sent. Specifically, it provides a reliable, network-independent message-interchange service to the top three application-oriented layers. This layer acts as an interface between the bottom and top three layers. By providing the session layer (layer 5) with a reliable message transfer service, it hides the detailed operation of the underlying network from the session layer.
5. The **Session layer** decides when to turn communication on and off between two computers. It provides the mechanisms that control the data-exchange process and coordinates the interaction between them. It sets up and clears communication channels between two communicating components. Unlike the network layer (layer 3), it deals with the programs running in each machine to establish conversations between them. Some of the most commonly encountered protocol stacks, including TCP/IP, don't implement a session layer.
6. The **Presentation layer** performs code conversion and data reformatting (syntax translation). It is the translator of the network, making sure the data is in the correct form for the receiving application. Of course, both the sending and receiving applications must be able to use data subscribing to one of the available abstract data syntax forms. Most commonly, applications handle these sorts of data translations themselves rather than handing them off to a Presentation layer.
7. The **Application layer** provides the interface between the software running in a computer and the network. It provides functions to the user's software, including file transfer access and management (FTAM) and electronic mail service.

2.4 TCP/IP Protocol Suite

TCP/IP is a result of protocol research and development conducted on the experimental packet-switched network, ARPANET, funded by the Defense Advanced Research Projects Agency (DARPA) in the U.S, and is generally referred to as the TC/IP protocol suite. This protocol suite consists of a large collection of protocols that have been issued as Internet standards by the Internet Activities Board (IAB).

There is no official TCP/IP protocol model as there is in the case of OSI. However, based on the protocol standards that have been developed, we can organize the communication task for TCP/IP into five relatively independent layers;

- Application layer
- Transport layer (TCP)
- Internet layer (IP)
- Network access layer
- Physical layer

The **physical layer** covers the physical interface between a data transmission device (e.g., workstation, computer) and a transmission medium or network. This layer is concerned with specifying the characteristics of the transmission medium, the nature of the signals, the data rate, and related matters.

The **network access layer** is concerned with the exchange of data between an end system and the network to which it is attached. The sending computer must provide the network with the address of the destination computer, so that the network may route the data to the appropriate destination. The specific software used at this layer depends on the type of network to be used; different standards have been developed for circuit switching, packet switching (e.g., X.25), local area networks (e.g., Ethernet), and others. Thus, it makes sense to separate those functions having to do with network access into a separate layer. By doing this, the remainder of the communications software, above the network access layer, need not be concerned about the specifics of the network to be used. The same higher-layer software should function properly regardless of the particular network to which the computer is attached.

The network access layer is concerned with access to and routing data across a network for two end systems attached to the same network. In those cases where two devices are attached to different networks, procedures are needed to allow data to traverse multiple interconnected networks. This is the function of the **Internet layer**. The **Internet Protocol (IP)** is used at this layer to provide the routing function across multiple networks. This protocol is implemented not only in the end systems but also in routers. A router is a processor that connects two or more networks and whose primary function is to relay data from one network to the other on its route from the source to the destination end system.

Regardless of the nature of the applications that are exchanging data, there is usually a requirement that data be exchanged reliably. That is, we would like to be assured that all of the data arrive at the destination application and that the data arrive in the same order in which they were sent. As we shall see, the mechanisms for providing reliability are essentially independent of the nature of the applications. Thus, it makes sense to collect those mechanisms in a common layer shared by all applications; this is referred to as the **transport layer**. The **Transmission Control Protocol (TCP)** is the most commonly used protocol to provide this functionality. This protocol model TCP/IP derived its name from the above two protocols.

Finally, the **application layer** contains the logic needed to support the various user applications. For each different type of application, such as file transfer, separate module is needed that is peculiar to that application.

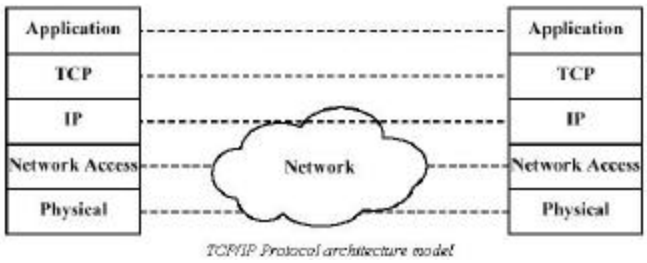


Figure 2.1 TCP/IP Protocol Architecture Model

Figure 2.1 shows how the TCP/IP protocols are implemented in end systems. Note that the physical and network access layers provide interaction between the end system and the network, whereas the transport and application layers are what is known as end-to-end protocols; they support interaction between two end systems. The Internet layer has the flavor of both. At this layer, the end system communicates routing information to the network but also must provide some common functions between the two end systems.

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- Data and Network CommunicationMiller, Vikas publishing house
- Data CommunicationsPrakash C Gupta, Eastern Economy edition
- www.ctr.kcl.ac.uk/lectures

Notes

LESSON 3

Today's Topics

- Standards and organizations to set up standards
- Signals
- Analog Signal
- Digital Signal
- Sampling rate and Nyquist Criterion

Objectives

Upon completion of this Lesson, you should be able to:

- Discuss on Organisations which set standards
- Define an analog and digital signal
- Discuss the limitations of analog signals
- Explain sampling rate and nyquist criterion

3.1 Standards: You Can Answer That Definitely!

Since standards are necessary in regulating data communication there are a no: of organizations serving as standards creation committees. A few of them in this area are

3.1.1 International standards Organization (ISO)

is a well-known multinational standards body. Most members of ISO are their respective government representatives. Members from over 80 developed nations actively reperesent the ISO.

The American National standards Institute: is a private non profit organization that does not have any direct ties with the US federal Govt. Professional groups, industry representatives, consumer groups etc represent ANSI.

3.1.2 The Institute of Electrical and Electronic Engineers:

is the biggest professional engineering body in the world. It oversees the development of international computer and communication standards.

3.2 Signals

At the core of data communications is the concept of signal propagation. Any signal can be classified into one of the two types - **analog or digital**

3.2.1 Analog Signals

An analog signal is a continuously varying signal, similar to a sinusoidal waveform. For e.g.: if we measure the room temperature continuously and plot its graph with time on the X axis and temperature on the Y-axis, we would get a continues waveform. This is an example of an analog signal.
Therefore an analog signal is smooth and continuous whereas a digital signal represents a sudden jump from one value to another.

The sine wave is the fundamental continuous signal. A general sine wave can be represented by three parameters: amplitude (A), frequency (f), and phase (ô). The amplitude is the peak value of strength of the signal over time; typically, this value is measured in volts or watts. The frequency is the rate (in cycles per second, or Hertz (Hz)) at which the signal repeats. An equivalent parameter is the period (T) of a signal, which is the amount of time it takes for one repetition; therefore, T = 1/f. Phase is a measure of the relative position in time within a single period of a signal, as illustrated below. The general sine wave can be written as

$$S(t) = A\sin(2\pi ft + \phi)$$

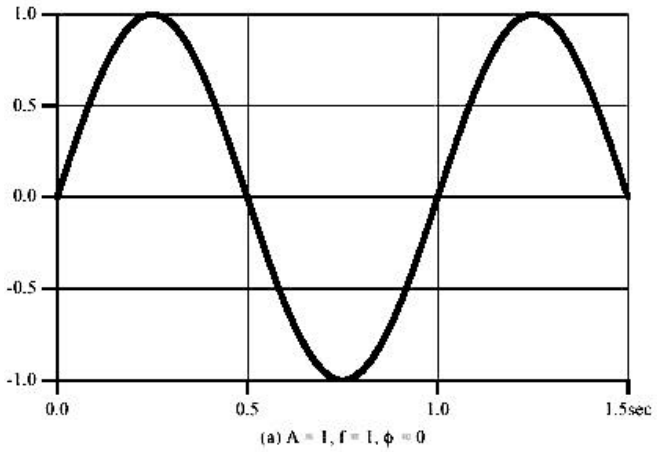


Fig 3.1

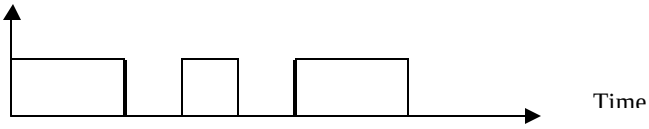
Digital signals

In contrast to analog O/P instruments that sense continuous variables such as pressure and temperature, many instruments provide an output that has one of two states: high or low, open or closed. A pressure might be too high or a temperature too low, triggering closure of a switch. Outputs, too, are not strictly analog-solenoid valves typically are opened or closed, many pumps and heaters are simply turned on or off. Pulse signals are another form of digital I/O, with one rotation of a turbine flow meter or tachometer corresponding to a single, countable event. Digital I/O also can be used for parallel communications among plug-in expansion cards, and to generate clock and other timing signals.

Already in the binary language of computers, these types of digital, or discrete, inputs and outputs (I/O) are much easier for microprocessor-based data acquisition systems to deal with than analog signals. Similar to analog-to-digital converters used for analog I/O, digital I/O is designed to deal directly with transistor-to-transistor logic (TTL) level voltage changes. TTL typically sets the low voltage level between 0 and 0.8 V and the high voltage level between 2.0 and 5.0 V. Voltage levels between 0.8 and 2.0 V are not allowed. A voltage change, then, from the

high range to the low range (or vice versa) represents a digital change of state from high to low, on to off, etc.

And because acquiring an analog signal is more complex than acquiring a digital one, analog I/O channels also are more expensive. Hence, if digital I/O is adequate, do not bother with analog.



Digital Signal

Fig 3.2

3.4 Why Digital Communication

A transmission system makes use of a physical **transmission media** or **channel** that allows the propagation of electromagnetic energy in the form of pulses or variations in voltage, current, or light intensity. In analog communication the objective is to transmit a signal waveform, which is a function that varies continuously with time, as shown in Figure 3.3a. For example, the electrical signal coming out of a microphone corresponds to the variation in air pressure corresponding to sound. This function of time must be reproduced exactly at the receiver output of the analog communication system. In practice, communications channels do not satisfy this condition, so some degree of distortion is unavoidable.

In digital transmission the objective is to transmit a given symbol that is selected from some finite set of possibilities. For example, in binary digital transmission the objective is to transmit either a 0 or a 1. This can be done, for instance, by transmitting positive voltage for a certain period of time to convey a 1 or a negative voltage to convey a 0, as shown in Figure 3.3b. The task of the receiver is to determine the input symbol with high probability. The positive or negative pulses that were transmitted for the given symbols can undergo a great degree of distortion. Where signaling uses positive or negative voltages, the system will operate correctly as long as the receiver can determine whether the original voltage was positive or negative.

The cost advantages of digital transmission over analog transmission become apparent when transmitting over a long distance. Consider, for example, a system that involves transmission over a pair of copper wires. As the length of the pair of wires increases, the signal at the output is attenuated and the original shape of the signal is increasingly distorted.

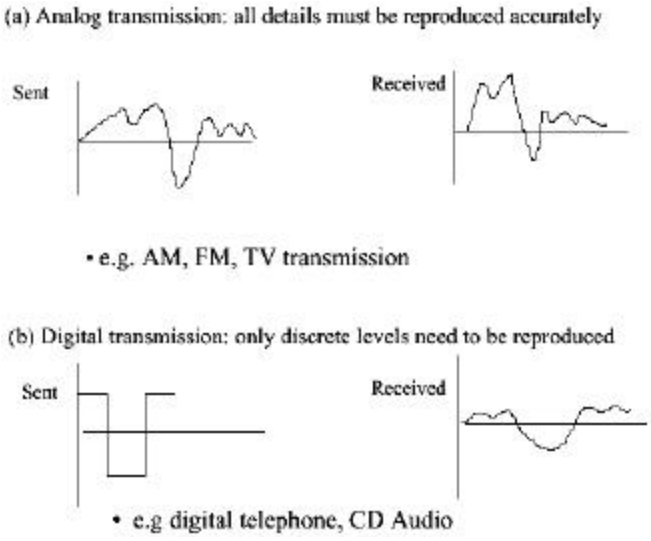


Figure 3.3 Analog versus Digital signal transmission

In addition, interference from extraneous sources, such as radiation from car ignitions and power lines, as well as noise inherent in electronic systems result in the addition of random noise to the transmitted signal. To transmit over long distances, it is necessary to introduce **repeaters** periodically to regenerate the signal, as shown in Figure 3.4. Such signal regeneration is fundamentally different for analog and digital transmissions.

Figure 3.4 Typical long-distance link

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- The Data Acquisition Systems Handbook, Omega Press LLC, 1997.
- New Horizons in Data Acquisition and Computer Interfaces, Omega Press LLC, 199
- users.telenet.be/educypedia

LESSON 4

Today's Topics

- Repeaters
- Bandwidth
- Sampling rate
- Nyquist Criterion

Objectives

Upon completion of this Lesson, you should be able to:

- Discuss on repeaters
- Define bandwidth
- Explain sampling rate and nyquist criterion

4.1 Repeater

In an analog communication system, the task of the repeater is to regenerate a signal that resembles as closely as possible the signal at the input of the repeater segment. Figure 4.1 shows the basic functions carried out by the analog repeater.

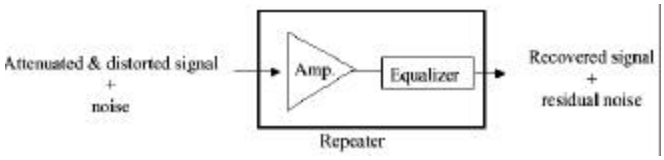


Figure 4.1 analog repeaters

The input to the repeater is an attenuated and distorted version of the original transmitted signal plus the random noise added in the segment. First the repeater deals with the attenuation by amplifying the received signal. To do so the repeater multiplies the signal by a factor that is the reciprocal of the attenuation '**a'.** The resulting signal is still distorted by the channel.

The repeater next uses a device called an **equalizer** in an attempt to eliminate the distortion. The source of the distortion in the signal shape has two primary causes. The first cause is that different frequency components of the signal are attenuated differently. In general, high frequency components are attenuated more than low-frequency components. The equalizer compensates for this situation by amplifying different frequency components by different amounts. The second cause is that different frequency components of a signal are delayed by different amounts as they propagate through the channel. The equalizer attempts to provide differential delays to realign the frequency components. In practice it is very difficult to carry out the two functions of the equalizer. For the sake of argument, suppose that the equalizer is perfect. The output of the repeater then consists of the original signal plus the noise.

In the case of analog signals the repeater is limited in what it can do to deal with noise. If it is known that the original signal does not have components outside a certain frequency band, then the repeater can remove noise components that are outside the signal band. However, the noise within the signal band cannot be reduced and consequently the signal that is finally recovered by repeater will contain some noise. The repeater then proceeds to send the recovered signal over the next transmission segment. By the time signal reaches the destination after going through many repeaters as in the case of long distance transmission its quality degrades considerably as the noises accumulates at each segment.

Next consider the same copper wire transmission system for digital communications. Suppose that a string of 0s and 1s is conveyed by a sequence of positive and negative voltages. As the length of the pair of wires increases, the pulses are increasingly distorted and more noise is added. A digital repeater is required as shown in Figure 4.2. The sole objective of the repeater is to determine with high probability the original binary stream. The repeater also uses an equalizer to compensate for the distortion introduced by the channel. However, the repeater does not need to completely regenerate the original shape of the transmitted signal. It only needs to determine whether the original pulse was positive or negative. To do so, the repeater is organized in the manner shown in Figure 4.2.

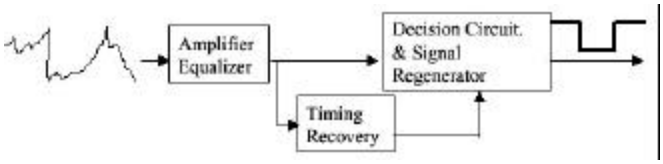


Figure 4.2 A Digital repeater

A timing recovery circuit keeps track of the intervals that define each pulse. The decision circuit then samples the signal at the midpoint of each interval to determine the polarity of the pulse. In a properly designed system, in the absence of noise, the original symbol would be recovered every time, and consequently the binary stream would be regenerated exactly over any number of repeaters and hence over arbitrarily long distances. However, noise is unavoidable, which implies that errors will occur from time to time. An error occurs when the noise signal is sufficiently large to change the polarity of the original signal at the sampling point. Digital transmission systems are designed for very low bit error rates, for example, 10^{-7},10^{-9}, or even 10^{-12}, which corresponds to one error in every trillion bits!

The impact on signal quality in multiple digital repeaters is similar to the digital recording of music where the signal is stored as a file of binary information. We can copy the file digitally any number of times with extremely small probabilities of errors being introduced in the process. In effect, the quality of the sound is unaffected by the number of times the file is copied.

The preceding discussion shows that digital transmission has superior performance over analog transmission. Digital repeaters eliminate the **accumulation of noise** that takes place in analog systems and provide for long-distance transmission

that is nearly independent of distance. Digital transmission systems can operate with lower signal levels or with greater distances between repeaters than analog systems can. This factor translates into lower overall system cost and was the original motivation for the introduction of digital transmission.

Over time, other benefits of digital transmission have become more prominent. Networks based on digital transmission are capable of handling any type of information that can be represented in digital form. Thus digital networks are suitable for handling many types of services. Digital transmission also allows networks to exploit the advances in digital computer technology to increase not only the volume of information that can be transmitted but also the types of processing that can be carried out within the network, that is, error correction, data encryption, and the various types of network protocol processing that are the subject of this book.

4.2 Bandwidth, Data Rate and Channel Capacity

We have seen that there are a variety of impairments that distort or corrupt a signal. For digital data, the question that then arises is to what extent these impairments limit the data rate that can be achieved. The rate at which data can be transmitted over a given communication path, or channel, under given conditions, is referred to as the **channel capacity.**

There are four concepts here that we are trying to relate to one another:

4.2.1 Data Rate.

This is the rate, in bits per second (bps), at which data can be communicated.

4.2.2 Bandwidth

This is the maximum bandwidth of the transmitted signal as constrained by the nature of the transmission medium or transmission channel, expressed in cycles per second, or hertz (Hz).

4.2.3 Noise

The average level of noise over the communications path.

4.2.4 Error Rate

The rate at which errors occur, where an error is the reception of a 1 when a 0 was transmitted, or the reception of a 0 when a 1 was transmitted.

The problem we are addressing is this: communications facilities are expensive, and, in general, the greater the bandwidth of the transmission facility, the greater the cost. Furthermore, all transmission channels of any practical interest are of limited bandwidth. The limitations arise from the physical properties of the transmission medium or from deliberate limitations at the transmitter on the bandwidth to prevent interference from other sources. Accordingly, we would like to make as efficient use as possible of a given bandwidth. For digital data, this means that we would like to get as high a data rate as possible at a particular limit of error rate for a given bandwidth. The main constraint on achieving this efficiency is noise.

4.2.5 Nyquist Sampling Rate

To begin, let us consider the case of a channel that is noise-free. In this environment, the limitation on data rate is simply the bandwidth of the signal. A formulation of this limitation, due to **Nyquist**, states that if the signal with frequencies components not greater than **W** Hz (cycles per second) is given, then it is possible to represent completely that signal by sampling it at **2W** samples per second.

Extending this to the transmission media, given a transmission medium of bandwidth W Hz, the highest signal rate that can be carried by it is 2W samples per second.

Note that in the last paragraph, we referred to samples. If the signals to be transmitted are binary (two voltage levels), then each sample would represent a binary bit, and the data rate that can be supported by W Hz is 2W bps. As an example, consider a voice channel being used, via modem, to transmit digital data. Assume a bandwidth of 3100 Hz. Then the capacity, C, of the channel is 2W = 6200 bps. However, signals or samples with more than two levels can be used; that is, each signal element can represent more than one bit. For example, if four possible voltage levels are used as sampled signals, then each signal element can represent two bits. With multilevel signaling, the Nyquist formulation becomes

$$C = 2W \log_2 M$$

Where M is the number of discrete signal or voltage levels. Thus, for M = 8, a value used with some modems, channel capacity C becomes 18,600 bps. So, for a given bandwidth, increasing the number of different signals can increase the data rate. However, this places an increased burden on the receiver: Instead of distinguishing one of two possible signals during each signal time, it must distinguish one of M possible signals. Noise and other impairments on the transmission line will limit the practical value of M.

Thus, all other things being equal, doubling the bandwidth doubles the data rate. Now consider the relationship between data rate, noise, and error rate. The presence of noise can corrupt one or more bits. If the data rate is increased, then the duration of bit becomes shorter so that more bits are affected by a given pattern of noise. Thus, at a given noise level, the higher the data rate, the higher the error rate.

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- The Data Acquisition Systems Handbook, Omega Press LLC, 1997.
- New Horizons in Data Acquisition and Computer Interfaces, Omega Press LLC, 199
- users.telenet.be/educypediadc.qut.edu.au/sedc/

LESSON 5

Today's Topics

- Transmission media
- UTP cables
- STP cables
- Coaxial cables
- Optic Fiber cable**s**

Objectives

Upon completion of this Lesson, you should be able to:

- Explain the construction of different types of cables
- Identify the application of each
- Discuss the Advantages and Limitations of the different types of cables

5.1 Transmission Media

When you speak to your friend over the telephone how does your voice get carried forward? Definitely there is some kind of media, which carries your voice.

The first layer (physical layer) of the OSI Seven layer model is dedicated to the transmission media.

Due to the variety of transmission media and network wiring methods, selecting the most appropriate media can be confusing - what is the optimal cost-effective solution???

When choosing the transmission media, what are the factors to be considered?

- Transmission Rate
- Distances
- Cost and Ease of Installation
- Resistance to Environmental Conditions

5.1.1 Unshielded Twisted Pair (UTP)

UTP is the copper media, inherited from telephony, which is being used for increasingly higher data rates, and is rapidly becoming the de facto standard for horizontal wiring, the connection between, and including, the outlet and the termination in the communication closet.

A **Twisted Pair** is a pair of copper wires, with diameters of 0.4-0.8 mm, twisted together and wrapped with a plastic coating. The twisting increases the electrical noise immunity, and reduces the bit error rate (BER) of the data transmission. A UTP cable contains from 2 to 4200 twisted pairs.

UTP is a very flexible, low cost media, and can be used for either voice or data communications. Its greatest disadvantage is the limited bandwidth, which restricts long distance transmission with low error rates.

5.1.2 Shielded Twisted Pair

STP is heavier and more difficult to manufacture, but it can greatly improve the signaling rate in a given transmission scheme Twisting provides cancellation of magnetically induced fields and currents on a pair of conductors. Magnetic fields arise around other heavy current-carrying conductors and around large electric motors. Various grades of copper cables are available, with Grade 5 being the best and most expensive. Grade 5 copper, appropriate for use in 100-Mbps applications, has more twists per inch than lower grades. More twists per inch means more linear feet of copper wire used to make up a cable run, and more copper means more money.
Shielding provides a means to reflect or absorb electric fields that are present around cables. Shielding comes in a variety of forms from copper braiding or copper meshes to aluminized Mylar tape wrapped around each conductor and again around the twisted pair.

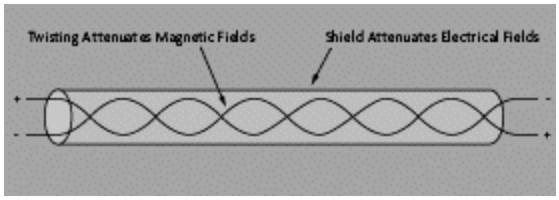
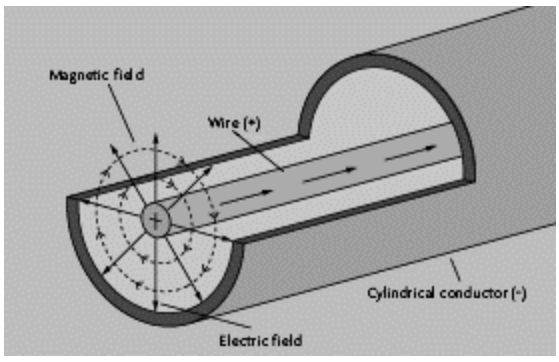


Fig 5.1 Shielded Twisted Pair Cable

5.1.3 Coaxial Cable

Coaxial cable is a two-conductor cable in which one conductor forms an electromagnetic shield around the other. The two conductors are separated by insulation. It is a constant impedance transmission cable. This media is used in base band and broadband transmission.

Coaxial cables do not produce external electric and magnetic fields and are not affected by them. This makes them ideally suited, although more expensive, for transmitting signals.



Coaxial Cable Construction

Fig 5.2

5.1.4 Fiber Optic Cable

The optical fiber construction:

The Fiber Optic Cable contains one or many fibers, each of them wrapped with a plastic tube and an external coating. Its construction includes

- CORE - made of glass.
- CLADDING - made of glass.
- Plastic coating

Fiber optics is being used more often as user applications demand higher and higher bandwidths. The term “bandwidth” technically means the difference between the highest and lowest frequencies of a transmission channel, in hertz (Hz). More commonly, it means the capacity or amount of data that can be sent through a given circuit.

A bandwidth of 100 Mbps is standard using fiber optic cables. When first introduced, fiber was considered only for special applications because it was expensive and difficult to work with. In recent years, the quest for greater bandwidth combined with easier-to-use fiber has made it more common. Tools and training for installing and troubleshooting fiber are readily available.

There are three basic fiber optic cables available: multimode step index, multimode graded index, and single mode. Multimode fibers usually are driven by LEDs at each end of the cable, while single mode lasers usually drive fibers. Single mode fibers can achieve much higher bandwidths than multimode fibers, but are thinner (10 microns) and physically weaker than multimode. Equipment costs for transmitting and receiving single mode fiber signals are much higher (at least four times) than for multimode signals.

One distinct advantage of fiber optic cables is noise immunity. Fiber optic cables can be routed indiscriminately through high noise areas with impunity, although fire ratings should be observed. Cables that pass through multiple spaces in a plant should be rated for heating/ ventilation/air conditioning (HVAC) plenums where they can withstand fires per National Fire Protection Association (NFPA) requirements

5.1.5 Media Comparison Chart

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- Data and Network CommunicationMiller, Vikas publishing house
- Data CommunicationsPrakash C Gupta, Eastern Economy edition
- www.cs.ucl.ac.uk/staff/S.Bhatti/D51-notes

MEDIA	ADVANTAGES	DISADVANTAGES
Twisted Pair Cable	Inexpensive Well understood Easy to add nodes	Sensitive to noise Short distances Limited bandwidth Security - easily tapped
Coaxial Cable	High bandwidth Long distances Noise immunity	Physic dimensions Security - easily tapped
Optical Fiber Cable	Very high bandwidth Noise immunity Long distances High security Small size	Connections T splitters

LESSON 6
LINE CONFIGURATION, TOPOLOGIES

Today's Topics

- Point to point Communication
- Multipoint Communication
- Bus topology
- Ring topology

Objectives

Upon completion of this Lesson, you should be able to:

- Distinguish between point to point and multipoint communication
- Differentiate between the different network topologies
- List out the advantages/ limitations of each

6.1 How Can Lines Generally be Configured?

Under the aspect of networking data communication can be done in two different ways: by

Point-to-Point Communication or by Multi-Point Communication

6.1.1 Point-to-Point Communication

is present when each terminal is connected to its main computer via a dedicated link. This kind of connection is fast and easy to install/maintain, but it needs a lot of cable and a lot of I/O-slots at the server's side. So it is most often used for very small networks.

Point-to-Point Communication can exist permanently or for the duration of a call (eg: telephone) only.

6.1.2 Multi-Point Communication

is present when several devices share the same transmission medium. Multi-Point Communication has several advantages. It can cut cabling costs and it solves the problem of too few available I/O-slots but it asks for additional software and to distinguish all terminals. For avoiding collisions on the channel software can be used (allowing only one device to send at a time), but most often a hardware solution is taken that prevents the signals from colliding by combining several signals into one (Multiplexing).

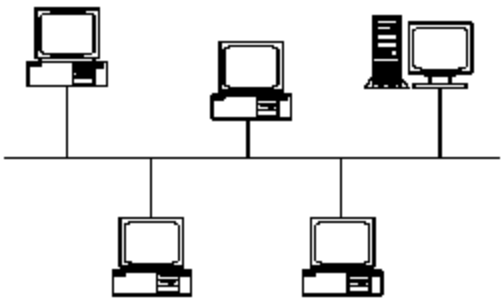
We will be studying multiplexing in detail at a later stage.

6.2 Network Topologies

The term topology refers to the method used to connect components on a network. The most common topologies are ring, bus, and star topologies, but they can take on the appearances of each other and still maintain their characteristics. For example, a token ring network segment can be wired in a star configuration, where components are cabled back to a hub where the ring is "inside" the hub. This allows a common wiring closet for a given building or area with home-run cable pulls for each component. A hub offers the advantages of centralized maintenance and configuration control.

6.2.1 Bus Topology

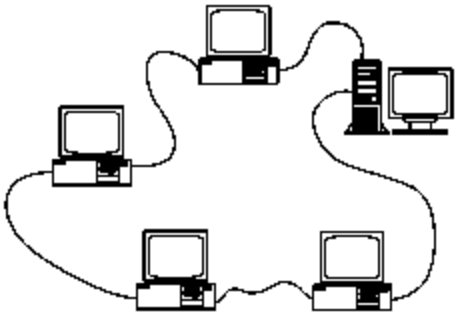
Bus topology uses multipoint philosophy. In this case, a long cable called bus forms the backbone to all the nodes. A node wanting to send some data to some other node pushes the data on the bus, which carries it to the other node. Looks similar to your bus transport? Now you know where the name came from.



Bus Network Topology

6.2.2 Ring Topology

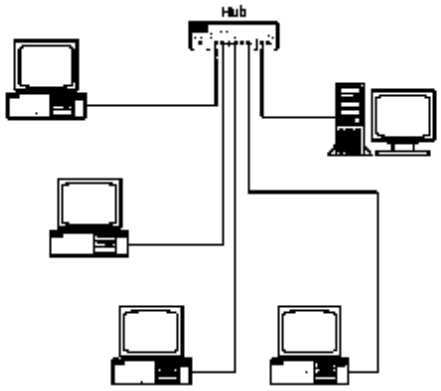
In ring topology each node is directly connected to only its two adjacent neighbours. If a node wants to send something to a distant node on a ring, it has to go through many intermediate nodes, which acts like repeaters, reproducing the same incoming bit stream with full signals on the outgoing line. A ring is easy to configure and install. If a node in a simple ring fails, the whole ring cannot function. Anothere demerit is that the traffic is only in one direction.



Ring Network Topology

6.2.3 Star Topology

Here there is a central node, often called a hub. If a node wants to send some data to another node, it sends it to this hub. The hub in turn sends it to the appropriate node.



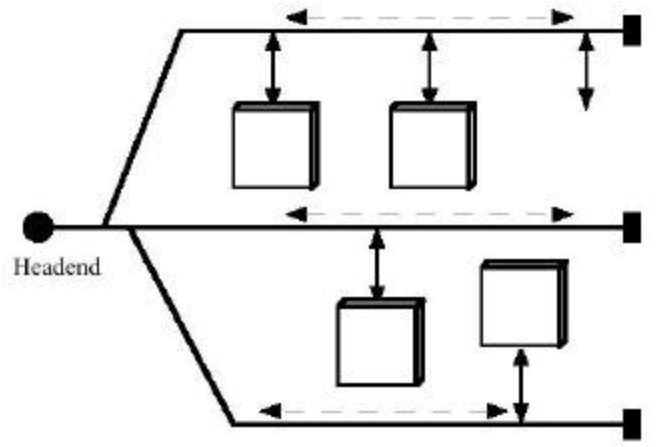
Star Network Topology

	Bus topology	Ring topology	Star topology
Application	Small networks and networks with low traffic	Small networks where speed is a criterion	Best way to integrate voice and data services and used extensively in digital PBXs
Complexity	Not complex	Relatively complex	Quite complex
Vulnerability	Failure of one workstation does not affect the network	Failure in a single workstation or in the channel can cause system failure due to the interdependence of workstations	If server fails all activity on the network ceases. Failuere of an individual workstation does not affect the system.
Expandability	Expansion is easy	Moderately easy to add or delete workstations	Is severely restricted because most servers can support a limited no: of network interfaces

Other common Topologies

6.2.4 Tree topology

In a tree network several devices or computers are linked in a hierarchical fashion. This type is commonly used in the organization where headquarters communicate with regional offices and regional offices communicate with district offices and so on.



(b) Tree

6.2.5 Hybrid Topology

Hybrid topology is one that uses two or more of the topologies mentioned above together. For example a token ring network segment can be wired in a star configuration, where components are cabled back to a hub where the ring is "inside" the hub. This allows a common wiring closet for a given building or area with home-run cable pulls for each component.

References and Further Reading

- Data and Network CommunicationMiller, Vikas publishing house
- Computer NetworksTannenbaum
- www.atis.org/tg2k/_tree_topology.html

Notes

LESSON 7

Today's Topics

- The different data encoding methods
- NRZ Encoding
- Bipolar Encoding

Objectives

Upon completion of this Lesson, you should be able to:

- Explain the different types of coding
- Compare the relative advantages/ limitations

7.1 Line Coding

Line coding is the method used for converting a binary information sequence into a digital signal in a digital communications system. The selection of a line coding technique involves several considerations. Maximizing bit rate is the main concern in digital transmission when bandwidth is at a premium. However, in other situations, such as in LANs, other concerns are also of interest. For example, another important design consideration is the ease with which the bit timing information can be recovered from the digital signal. Also, some line coding methods have built-in error detecting capabilities, and other methods have better immunity to noise and interference. Finally, the complexity and the cost of the line code implementations are always factors in the selection for a given application.

7.1.1 Non Return to Zero (NRZ) Encoding

This is the most common and easiest encoding scheme**.** A '0' is transmitted by raising the voltage level to high and a 1 is transmitted using a low voltage. Thus any sequence of 0 s and 1 s is transmitted by alternating between high and low. The name NRZ refers to the fact that the voltage level remains constant, i.e. does not return to zero during the time a bit is transmitted .If binary 1 is transmitted by sending a +A voltage level, and a 0 is transmitted by sending a 0 voltage and if binary 0s and 1s both occur with probability 1/2, then the average transmitted power for this line code is

$$(1/2)A^2 + (1/2)0^2 = A^2/2$$

frequency components because they produce essentially the same variations in a signal as a function of time. Strings of consecutive 0s and consecutive 1s lead to periods where the signal remains constant for long time producing low frequency components. These strings of 0s and 1s occur frequently enough to produce a spectrum that has its components concentrated at the lower frequencies. This situation presents a problem when the communications channel does not pass low frequencies. For example, most telephone transmission systems do not pass the frequencies below about 200 Hz.

7.1.2 Bipolar Encoding Method

Here is another encoding scheme, which may be more suitable for low frequencies.

The **bipolar encoding** method was developed to produce a spectrum that is more amenable to channels that do not pass low frequencies. In this method binary 0s are mapped into 0 voltage, thus making no contribution to the digital signals; consecutive 1s are alternately mapped into +A/2 and -A/2. Thus a string of consecutive 1s will produce a square wave with the frequency (1 / 2T) Hz. As a result, the spectrum for the bipolar code has its frequency content centered around the frequency (1/2T) Hz and has small content at low frequencies. Timing recovery is an important consideration in the selection of a line code. The timing-recovery circuit in the receiver, monitors the transitions at the edge of the bit intervals to determine the boundary between bits. Long strings of 0s and 1s in the binary and the polar binary encoding can cause the timing circuit to lose synchronization because of the absence of transitions. In the bipolar encoding long strings of 1s result in a square wave that has strong timing content; however, long strings of 0s still pose a problem. To address this problem, the bipolar line codes used in telephone transmission systems place a limit on the minimum number of 0s that may be encoded into the digital signal. Whenever a string of N consecutive 0s occurs, the string is encoded into a special binary sequence that contains 0s and 1s. To alert the receiver that a substitution has been made, the sequence is encoded so that the mapping in the bipolar line code is violated; that is, two consecutive 1s do not alternate in polarity.

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Center Ltd, Manchester

LESSON 8

Today's Topics

- Return to Zero Encoding
- Manchester Encoding
- Differential Manchester encoding

Objectives

Upon completion of this Lesson, you should be able to:

- Explain the different types of coding
- Compare the relative advantages/ limitations

8.1 Return to Zero Encoding

This usually ensures signal patterns for any bit pattern and thus overcome the limitation of NRZ codes. Two of these techniques, which are very widely used, are Manchester Encoding and Differential Manchester Encoding.

8.2 Manchester Encoding

In the Manchester encoding, there is a transition at the middle of each bit period. The mid bit transition serves as a clocking mechanism and also a data. a low to high transition represents a 1, and a high to low transition represents a zero. The Manchester encoding can be viewed as the transmission of two pulses for each binary bit. A binary 1 is mapped into the binary pair of 10, and the corresponding polar encoding for these two bits is transmitted: A binary 0 is mapped into 01.

8.3 Differential Manchester Encoding

In differential Manchester, the mid bit transition is used only to provide clocking. A '0' causes the signal to change at the start of the interval. A '1' causes the signal to remain where it was at the end of the previous interval. Thus a '0' may go from low to high or high to low depending on the initial value of the signal.

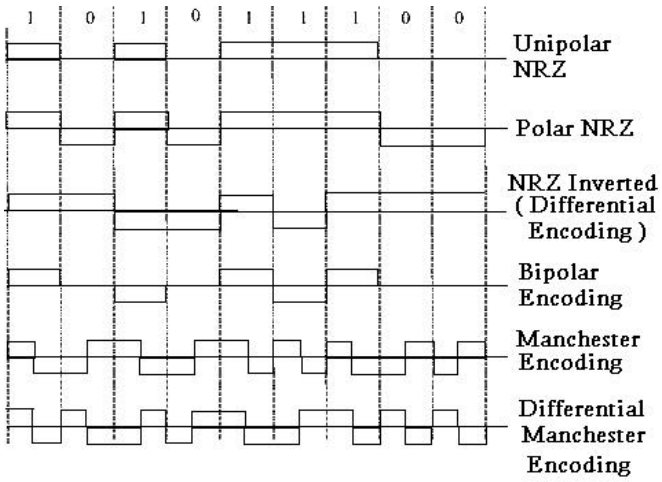


Figure 8.1 Line coding methods

For you to learn and know better here are a few references. You might be able to find many more.

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Center Ltd, Manchester
- www.rhyshaden.com/encoding.htm

Notes

LESSON 9

Now that you have learnt how to encode your data, you are ready to see how your data travels.

Today's Topics

- Modulation
- Amplitude shift Keying
- Frequency shift keying
- Phase shift Keying

Objectives

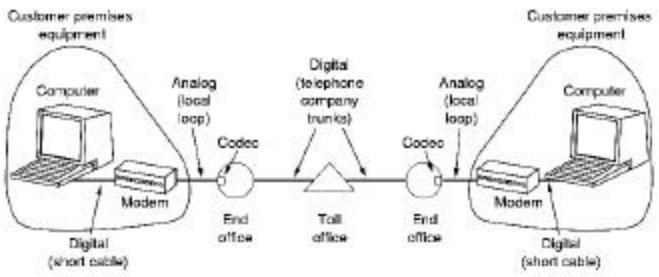
Upon completion of this Lesson, you should be able to:

- Define Modulation
- Distinguish between different types of Modulation
- Explain the advantages and disadvantages if any of each type of modulation

9.1 Modulation

Modulation is the process, or result of the process, of varying a characteristic of a carrier in accordance with an information-bearing signal. **For the past 100 years, analog transmission has dominated all the communication**. In particular, the telephone system was originally based entirely on analog signaling. With the advance in technology, the long-distance trunks between telephone exchanges are converted to digital, but the local loops between a telephone exchange and the telephone at the user are still analog. Consequently, when a computer wishes to send digital data it produces, over the telephone line, the data must be first converted to analog form by a device for transmission over telephone line and at the receiver end this received analog signal must be converted back to digital data.

A device that accepts a serial stream of digital bits as input and produces modulated analog carrier signal as output (or vice versa) is called a **modem** (for **mo**dulator- **dem**odulator). The modem is inserted between the digital computer and the analog telephone system. A continuous tone in the 1000 to 2000 Hz range, called a **sine wave carrier** is modulated according to the input digital signal at the transmitting end and at the receiver end the received modulated signal is converted back to digital stream of bits by the process of demodulation.



There are three basic types of modulation methods for transmission of digital signals. These methods are based on the three attributes of a sinusoidal signal, amplitude, frequency and phase. The corresponding modulation methods are called Amplitude Shift Keying (ASK), Frequency shift keying (FSK), and Phase Shift Keying (PSK). In addition a combination of ASK and PSK is employed at high bit rates. This method is called **Quadrature Amplitude Modulation(QAM).**

We will now see these methods in detail.

9.1.1 Amplitude Phase Shift Keying

This you can say is the simplest mode of digital modulation. In ASK the carrier amplitude is multiplied by the binary 1 or 0.The amplitude modulated carrier signal can be written as

$$V(t) = d \sin (2PI\ fct)$$

Where fc is the carrier frequency and d is the data bit variable that can take values '1' or '0'. depending on the state of the digital signal.

9.1.2 Frequency Shift Keying

In FSK the frequency of the carrier is shifted between two discrete values, one representing binary '1' and the other representing binary '0'. The carrier amplitude does not change. FSK is relatively simple to implement. It is used extensively in low speed modems having bit rates below 1200 bps. The instantaneous value of the FSK signal is given by

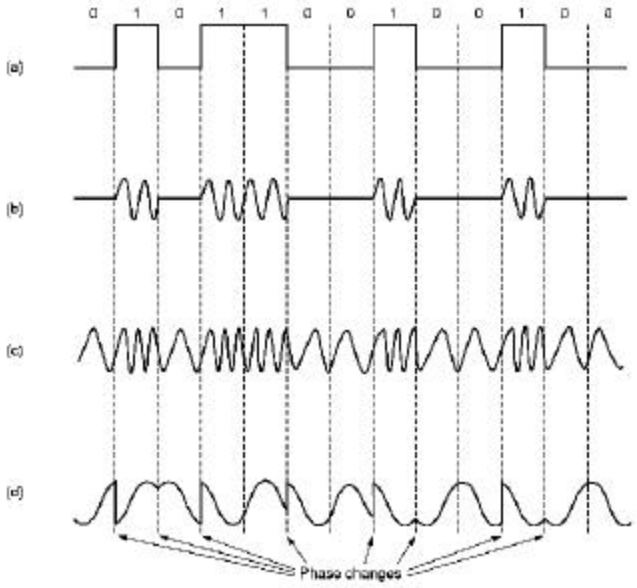
$$V(t) = d \sin(2Pift(t) + d \sin(2Pifot)$$

Where f1 and f0 are the frequencies corresponding to binary '1' and '0' respectively and d is the data signal variable.

9.3 Phase Shift Keying

This is the most efficient of the three modulation methods and is used for high bit rates. Inn PSK, phase of the carrier is modulated to represent the binary values. The simplest for m of PSK is called as BPSK. . In the simplest form of **phase modulation**, the carrier wave is systematically shifted 45, 135, 225, or 315 degrees at uniformly spaced intervals. Each phase shift transmits 2 bits of information.

The figure below will illustrate to you better the three forms of modulation.



a) Binary signal. (b) Amplitude modulation (c) Frequency Modulation (d) Phase Modulation

Fig 9.1

To go to higher and higher speeds, it is not possible to just keep increasing the sampling rate. The Nyquist theorem says that even with a perfect 3000-Hz line (which a dial-up telephone is not), there is no point in sampling the signal faster than 6000 Hz as all frequency components higher than 3000Hz in the input signal are going to be filtered out by the 3000Hz telephone line. Thus all research on faster modem is focused on getting more bits per sample (i.e. per baud).

Most modems use a combination of modulation techniques to transmit multiple bits per-baud. In Figure 9.2a we see dots at 0,90, 180, and 270 degrees, with two amplitude levels per phase shift. The distance from the origin indicates amplitude. In Figure 9.2b we see a different modulation scheme, in which 16 different combinations of amplitude and phase shift are used.

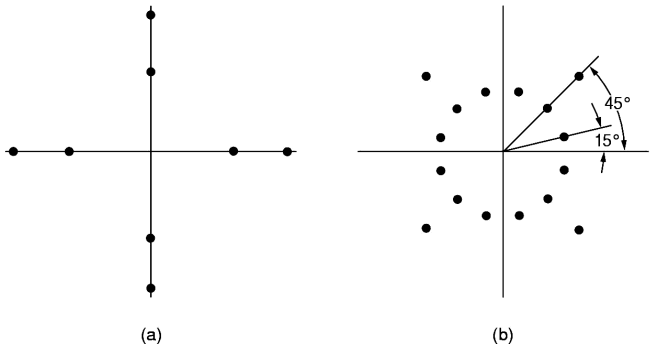


Fig 9.2

Number of combination used will be a power of 2 and each sample or baud represents number of bits that is equal to the power. For example with $8 = 2^3$ different combination each baud represents 3 bits and with $16 = 2^{4\ \text{different}}$ combination each baud represents 4 bits. The scheme of Figure 9.2 b when used to transmit 9600 bps over a 2400 baud line is called **QAM** (**Quadrature Amplitude Modulation).**

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Center Ltd, Manchester
- modems.rosenet.net/

Notes

LESSON 10

Today's Topics

- Data modems
- Operation on two wire and four wire
- Dial up lines

Objectives

Upon completion of this Lesson, you should be able to:

- Be able to define what is a modem
- Explain the different types of modem

Now that you know the different modulation schemes it is apt to have a look at data modems

Many of you might be having one at home and must be already familiar with its use.

Try defining what a modem is before you continue. The answer is there in the term itself.

10.1 Modem

The term modem is derived from the words, modulator and demodulator. The digital modulation and demodulation schemes discussed previously are implemented in the modems. Most of the modems are designed for utilizing the analog voice band service offered by the telecommunication network. Therefore the modulated carrier wave generated by a modem fits into the 300- 3400 Hz bandwidth of the speech channel.

10.2 Types of Modems

Modems can be of several types and they can be categorized in a no: of ways.

- Directional Capability – Half Duplex and Full Duplex Modems
- Connection to the line- 2 wire modem and 4 wire modem
- Transmission mode- Asynchronous modem and synchronous modem

Half Duplex and full Duplex modems- A half duplex modem permits transmission in one direction at a time. If a carrier is detected on the line by the modem, it gives an indication of the incoming carrier to the DTE (Data terminal equipment) through a control signal of its digital interface. So long as the carrier is being received the modem does not give clearance to the DTE to transmit.

A full duplex modem allows transmission in both directions. Thus there are two carriers on the line, one outgoing and the other incoming.

10.2.1 2W – 4W Modems

The line interface of the modem can have a 2 wire or a 4 wire connection to the transmission medium. In a 4 wire connection, one pair of wires is used for the outgoing carrier and the other is used for the incoming carrier. Full duplex and half duplex modes of data transmission are possible on a 4 wire connection. As the physical transmission path for each direction is separate, the same carrier frequency can be used for both the directions.

A leased two wire connection is cheaper than a 4 wire connection because only one pair of wires is extended to the subscriber's premises. The data connection established through telephone exchanges is also a two-wire connection. For the 2 wire connections, modems with a 2 wire line interface are required. Such modems use the same pair of wires for outgoing and incoming carriers. The same frequency is used for half duplex modems but for full duplex mode of operation on a two wire connection it is necessary to have two transmission channels, one for transmit direction and the other for the receive direction. This is achieved by frequency division multiplexing of two different carrier frequencies.

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Centre Ltd, Manchester
- modems.rosenet.net/

Notes

Today's Topics

- Synchronous data transmission
- Asynchronous data transmission
- UART
- USRT

Objectives

Upon completion of this Lesson, you should be able to:

- Explain the difference between synchronous and asynchronous data transmission
- Explain the working of UART
- Differentiate between UART and USRT

11.1 Serial Transmission

There are two primary forms of serial transmission: Synchronous and Asynchronous. Both forms are described below.

11.1.1 Synchronous Serial Transmission

requires that the sender and receiver share a clock with one another, or that the sender provide a strobe or other timing signal so that the receiver knows when to "read" the next bit of the data. In most forms of serial Synchronous communication, if there is no data available at a given instant to transmit, a fill character must be sent instead so that data is always being transmitted. Synchronous communication is usually more efficient because only data bits are transmitted between sender and receiver, and synchronous communication can be more costly if extra wiring and circuits are required to share a clock signal between the sender and receiver.

11.1.2 Asynchronous Transmission

allows data to be transmitted without the sender having to send a clock signal to the receiver. Instead, the sender and receiver must agree on timing parameters in advance and special bits are added to each word, which are used to synchronize the sending and receiving units.

11.2 UART(Universal Asynchronous Receiver/Transmitter)

The Universal Asynchronous Receiver/Transmitter (UART) controller is the key component of the serial communications subsystem of a computer. The UART takes bytes of data and transmits the individual bits in a sequential fashion. At the destination, a second UART re-assembles the bits into complete bytes. Serial transmission is commonly used with modems and for non-networked communication between computers, terminals and other devices.

When a word is given to the UART for Asynchronous transmissions, a bit called the "Start Bit" is added to the beginning of each word that is to be transmitted. The Start Bit is used to alert the receiver that a word of data is about to be sent, and to force the clock in the receiver into synchronization with the clock in the transmitter. These two clocks must be accurate enough to not have the frequency drift by more than 10% during the transmission of the remaining bits in the word. (This requirement was set in the days of mechanical teleprinters and is easily met by modern electronic equipment.)

After the Start Bit, the individual bits of the word of data are sent, with the Least Significant Bit (LSB) being sent first. Each bit in the transmission is transmitted for exactly the same amount of time as all of the other bits, and the receiver "looks" at the wire at approximately halfway through the period assigned to each bit to determine if the bit is a 1 or a 0. For example, if it takes two seconds to send each bit, the receiver will examine the signal to determine if it is a 1 or a 0 after one second has passed, then it will wait two seconds and then examine the value of the next bit, and so on.

The sender does not know when the receiver has "looked" at the value of the bit. The sender only knows when the clock says to begin transmitting the next bit of the word.

When the entire data word has been sent, the transmitter may add a Parity Bit that the transmitter generates. The Parity Bit may be used by the receiver to perform simple error checking. Then at least one Stop Bit is sent by the transmitter.

When the receiver has received all of the bits in the data word, it may check for the Parity Bits (both sender and receiver must agree on whether a Parity Bit is to be used), and then the receiver looks for a Stop Bit. If the Stop Bit does not appear when it is supposed to, the UART considers the entire word to be garbled and will report a Framing Error to the host processor when the data word is read. The usual cause of a Framing Error is that the sender and receiver clocks were not running at the same speed, or that the signal was interrupted.

Regardless of whether the data was received correctly or not, the UART automatically discards the Start, Parity and Stop bits. If the sender and receiver are configured identically, these bits are not passed to the host.

If another word is ready for transmission, the Start Bit for the new word can be sent as soon as the Stop Bit for the previous word has been sent.

Because asynchronous data is "self synchronizing", if there is no data to transmit, the transmission line can be idle.

11.3 USRT

(Universal Synchronous Receiver-Transmitter) is a circuit capable of receiving and sending data without requiring a start and/or stop code.

References and Further Reading

- www.techweb.com/encyclopedia
- Handbook of data communication, National Computing Centre Ltd, Manchester
- isp.webopedia.com/TERM/U/USRT.html

Today's Topics

- Data interface standards
- RS 232
- EIA 232E

There is usually a special purpose device that handles the interface between the computer and the transmission line. This may take the form of a card that plugs into the motherboard (internal) or as an external device connected to the computer by a cable. If it is internal, it communicates with the computer using the computer's bus communication protocols; if external, it likely communicates through a serial communications port on the back of the computer. Either way there is a standardized interface that the device (modem, network card, etc.) presents to the computer that allows it to be configured and controlled.

A number of standards exist. These usually cover 4 areas relating to the interface:

- Mechanical specifies connectors (e.g. plugs and jacks), wire (circuit) connections between connectors, wire type and gauge, etc.
- Electrical specifies voltage levels, current, etc. on the various circuits.
- Functional specifies functions of the circuits (e.g. data, control, timing, ground, etc.). One trend to note is that increasingly, control and timing are embedded in the data circuit. This simplifies the interface at the expense of increasing the logic in the transmission device. For example, RS-232, a common standard for serial communications (e.g. modems, etc.), has 16 control lines, 4 data lines, 3 timing lines and a ground, whereas ISDN has only 4 data lines and 2 power lines.
- Procedural specifies the use of the functions to accomplish transmission and reception.

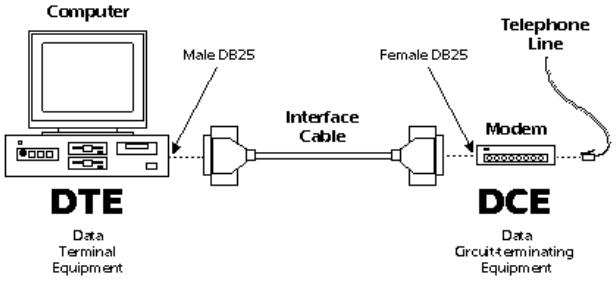
RS 232 Standard

In the early 1960s, a standards committee, today known as the Electronic Industries Association, developed a common interface standard for data communications equipment. At that time, data communications was thought to mean digital data exchange between a centrally located mainframe computer and a remote computer terminal, or possibly between two terminals without a computer involved. These devices were linked by telephone voice lines, and consequently required a modem at each end for signal translation. While simple in concept, the many opportunities for data error that occur when transmitting data through an analog channel require a relatively complex design. It was thought that a standard was needed first to ensure reliable communication, and second to enable the interconnection of equipment produced by different manufacturers, thereby fostering the benefits of mass production and competition. **From these ideas, the RS232 standard was born** It specified signal voltages, signal timing, signal function, a protocol for information exchange, and mechanical connectors.

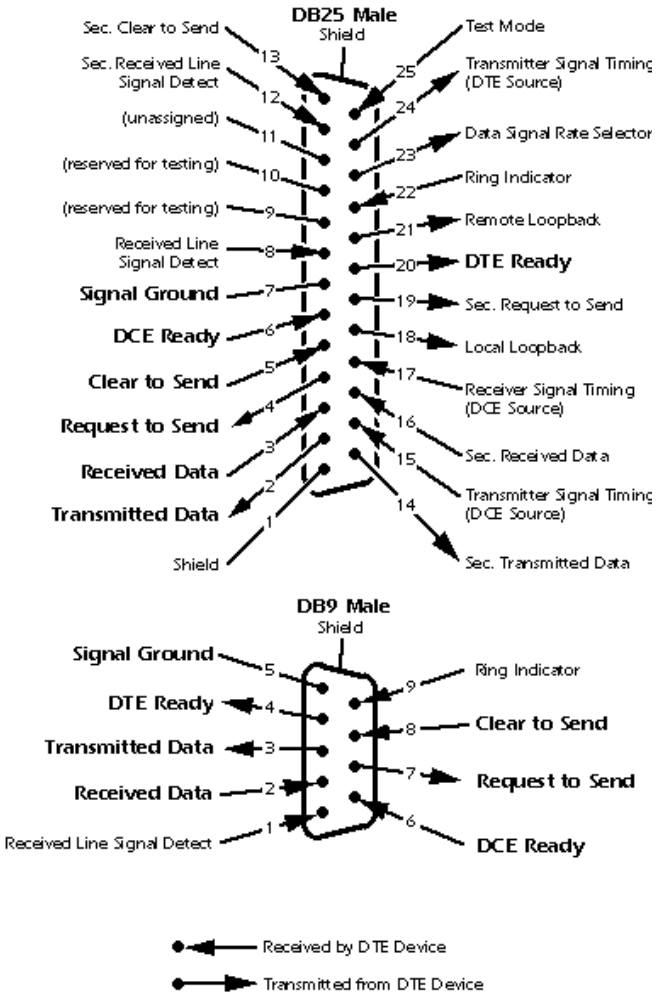
12.1 EIA 232E Standard

Over the 40+ years since this standard was developed, the Electronic Industries Association published three modifications, and the most the recent is EIA232E standard introduced in 1991. Besides changing the name from RS232 to EIA232, some signal lines were renamed and various new ones were defined, including a shield conductor.

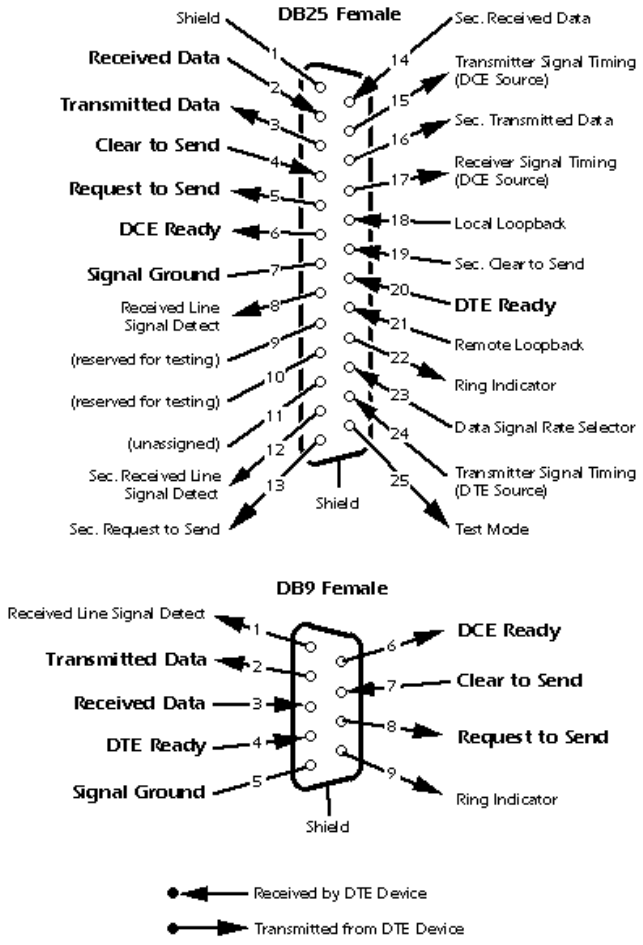
If the full EIA232 standard is implemented, the equipment at the far end of the connection named the DTE device (Data Terminal Equipment, usually a computer or terminal), has a male DB25 connector, and utilizes 22 of the 25 available pins for signals or ground. Equipment at the near end of the connection (the telephone line interface) named the DCE device (Data Circuit-terminating Equipment, usually a modem), has a female DB25 connector, and utilizes the same 22 available pins for signals and ground. The cable linking DTE and DCE devices is a parallel straight through cable with no crossover or self-connects in the connector hoods. If all devices exactly followed this standard, all cables would be identical, and there would be no chance that an incorrectly wired cable could be used. This drawing shows the orientation and connector types for DTE and DCE devices:

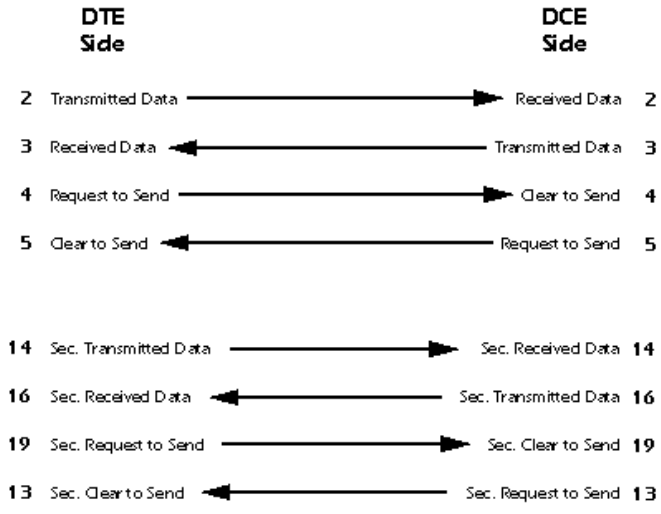


Looking Into the DTE Device Connector



Looking Into the DCE Device Connector





Limitations of EIA 232 D

- Although EIA 232 D is the most popular physical layer interface its use in computer networking is limited to low data rates and short distance data transmission applications. The distance between a DTE and DCE is limited to 15 mtrs.
- The absence or misconnection of flow control (handshaking) signals, resulting in buffer overflow or communications lock-up.
- Incorrect communications function (DTE versus DCE) for the cable in use, resulting in the reversal of the Transmit and Receive data lines as well as one or more handshaking lines.
- Incorrect connector gender or pin configuration, preventing cable connectors from mating properly

Notes

LESSON 13

Today's Topics

- Error detection
- Single Parity Check

Here the emphasis is sending data over a data communications link.

To achieve the necessary control, a layer of logic is added above the physical transmission layer; this logic is referred to as data link control or a data link control protocol. When a data link control protocol is used, the transmission medium between systems is referred to as a data link.

To see the need for data link control, we list some of the requirements and objectives for effective data communication between two directly connected transmitting-receiving stations:

- **Frame synchronization.** Data are sent in blocks called frames. The beginning and end of each frame must be recognizable.
- **Flow control.** The sending station must not send frames at a rate faster than the receiving station can absorb them.
- **Error control.** Any bit errors introduced by the transmission system must be corrected.
- **Addressing.** On a multipoint line, such as a local area network (LAN), the identity of the two stations involved in a transmission must be specified.
- **Control and data on same link.** It is usually not desirable to have a physically separate communications path for control information. Accordingly, the receiver must be able to distinguish control information from the data being transmitted.
- **Link management.** The initiation, maintenance, and termination of a sustained data exchange requires a fair amount of coordination and cooperation among stations. Procedures for the management of this exchange are required.

A data link protocol that satisfies these requirements is a rather complex affair.

First we look into the various techniques used to implement the error control and flow control mechanism. Then we look at the most important example of a data link control protocol: HDLC (High-level Data Link Control).

This protocol is important for two reasons: First, it is a widely used standardized data link control protocol. And secondly, HDLC serves as a baseline from which virtually all other important data link control protocols are derived.

13.1 Error Detection

In this section we discuss the idea of error detection in general terms, using the single parity check code as an example throughout the discussion. Also we briefly discuss Cyclic Redundancy Check (CRC), a most commonly used error detection mechanism. The basic idea in performing error detection is simple.

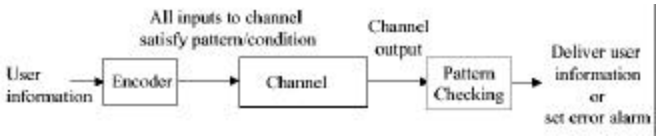


Figure 13.1 General Error detection system

As illustrated in Figure13.1, the information produced by an application programme is encoded so that the stream that is input into the communication channel satisfies a specific pattern or condition. The receiver checks the stream coming out of the communication channel to see whether the pattern is satisfied. If it is not, the receiver can be certain that an error has occurred and therefore sets an alarm to alert the user. This certainty stems from the fact that, no such pattern would have been transmitted by the encoder.

13.2 Single Parity Check

The Simplest code is the **single parity check code** that takes k information bits and appends a single **check bit** to form a **codeword**, which will be transmitted over the channel**.** The parity check at the receiver ensures that the total number of 1s in the received codeword is even; that is, the codeword has even parity. The check bit in this case is called a parity bit. Here the received codeword is valid if it has even number of 1s, otherwise it is invalid and there is some error in the received codeword. This error-detection code is used in ASCII where characters are represented by seven bits and the eighth bit consists of a parity bit. This code is an example of the so-called linear codes because the parity bit is calculated as the modulo 2 sum of the information bits:

$$b_{k+1} = (b_1 + b_2 + b_3 + \ldots. + b_k) \text{ modulo } 2$$

where b_1, b_2,b_3 , …. ,b_k are the information bits. And b_{k+1} is the parity bit generated. The modulo 2 arithmetic is given below

0	0	1	1
+0	+1	+0	+1
0	1	1	0

Thus, if the information bits contain an even number of 1s, then the parity bit will be 0; and if they contain an odd number of 1s, then the parity bit will be 1. Consequently, the above rule will assign the parity bit a value that will produce a codeword that always contains an even number of Is. This pattern defines the single parity check code.

If a codeword undergoes a single error during transmission, then the corresponding binary block at the output of the channel will contain an odd number of 1s and the error will be detected. More generally, if the codeword undergoes an odd number of errors, the corresponding output block will also

contain an odd number of 1s. Therefore, the single parity bit allows us to detect all error patterns that introduce an odd number of errors. On the other hand, the single parity bit will fail to detect any error patterns that introduce an even number of errors, since the resulting codeword will have even parity which is a valid codeword. Nonetheless, the single parity bit provides a remarkable amount of error-detection capability, since the addition of a single check bit results in making half of all possible error patterns detectable, regardless of the value of k.

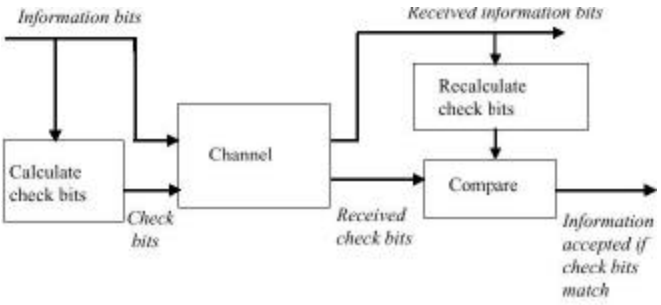


Figure 13.2 Error detection system using check bits

Figure 13.2 shows an alternative way of looking at the operation of this example. At the transmitter a checksum is calculated from the information bits and transmitted along with the information. At the receiver, the checksum is recalculated, based on the received information. The received and recalculated checksums are compared, and the error alarm is set if they disagree.

This simple example can be used to present two fundamental observations about error detection. The first observation is that error detection requires **redundancy** in that the amount of information that is transmitted is over and above the required minimum. For a single parity check code of length k+ 1, k bits are information bits, and one bit is the parity bit. Therefore, the fraction l/(k+1) of the transmitted bits is redundant.

The second fundamental observation is that **every error-detection technique will fail to detect some errors**. In particular, an error-detection technique will always fail to detect transmission errors that convert a valid codeword into another valid codeword. For the single parity check code, an even number of transmission errors will always convert a valid codeword to another valid codeword.

The objective in selecting an error-detection code is to select the code words that reduce the likelihood of the transmission channel converting one valid codeword into another. To visualize how this is done, suppose we depict the set of all possible binary blocks as the space shown in Figure 13.3, with codeword shown by **x**s in the space and noncode word by **o** s.

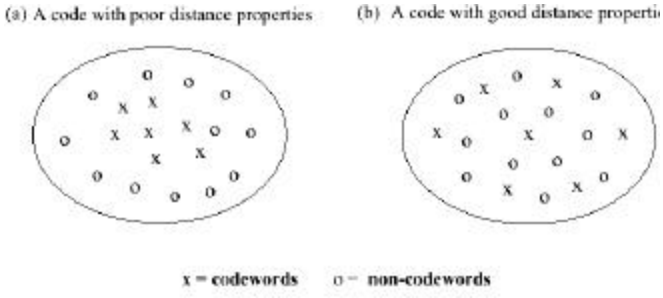


Figure 13.3 Distance properties of codes.

To minimize the probability of error-detection failure, we want the code words to be selected so that they are spaced as far away from each other as possible. Thus the code in Figure 13.3a is a poor code because the code words are close to each other or in other words distance between two valid code words are low. On the other hand, the code in Figure 13.3b is good because the distance between code words is maximized.

The effectiveness of a code clearly depends on the types of errors that are introduced by the channel.

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Centre Ltd, Manchester
- modems.rosenet.net/

Notes

LESSON 14

Today's Topics

- Cyclic Redundancy Check
- Introduction of error control mechanisms

14.1 Cyclic Redundancy Check (CRC)

One of the most common, and most powerful, error-detecting codes is the **Cyclic Redundancy Check (CRC**), which can be described as follows. Given a k-bit block of bits, or message, the transmitter generates an n-bit sequence, known as a **Frame Check Sequence (FCS**), so that the resulting frame, consisting of k +n bits, is exactly divisible by some predetermined number called CRC polynomial. The receiver then divides the incoming frame of k+n bits by the same CRC polynomial number and, if there is no remainder, assumes there was no error. The CRC polynomial number by which the information frame bits are divided are selected such that distance between two valid codeword is high and with proper selection of CRC polynomial it is possible to detect errors with very high probability i.e. more than 99.9% of the errors can be detected. Even it is possible to find out which bits are in error when only few bits are corrupted in the information bit stream. So with CRC it is also possible to correct the error though it is not possible for all possible errors. Even interesting thing about CRC is that the generation of the FCS can be implemented with very simple electronic circuitry. The generation of FCS does not take any extra time – at the end of transmission of k information bits, n bits FCS will be ready. Theory behind the CRC would make very interesting reading.

14.2 Error Control Mechanisms

Error control refers to mechanisms to detect errors that occur in the transmission of frames and take corrective steps to make sure frames are received correctly. In the model used, which covers the typical case, data are sent as a sequence of frames; frames arrive in the same order in which they are sent; and each transmitted frames suffer an arbitrary and variable amount of delay before reception. Two types of error are possible:

- **Lost frame.** A frame fails to arrive at the receiver. For example, a noise burst may damage a frame to the extent that the receiver is not aware that a frame as been transmitted.
- **Damaged Frame.** A recognizable frame arrives but part of its content is in error.

  The most common techniques for error control are based on some or all of the following steps.
- **Error detection.** This is done as discussed in previous section.
- **Positive acknowledgement.** The destination returns a positive acknowledgement to successfully received, error-free frames.
- **Retransmission after timeout.** The source retransmits a frame that has not been acknowledged after a predetermined amount of time.
- **Negative acknowledgement and retransmission.** The receiver returns a negative acknowledgement to frames in which an error is detected. The source retransmits those frames again.

Collectively these mechanisms are referred to as **Automatic repeat ReQuest(ARQ)**. In effect ARQ provides reliability over an unreliable data link. **ARQ** is a technique used to ensure that a data stream is delivered accurately to the user despite errors that occur during transmission. And we refer to the set of rules that govern the operation of the transmitter and receiver based on ARQ as the **ARQ protocol.** ARQ forms the basis for Data Link Control protocols that provide for the reliable transfer of information. In this section we discuss the three basic types of ARQ protocols, starting with the simplest and up to the most complex.

Notes

LESSON 15

Today's Topics

- Stop-and-wait-ARQ
- Go-back-N ARQ
- Selective-reject ARQ

There are three version of ARQ, which are standardized.

- Stop-and-wait-ARQ
- Go-back-N ARQ
- Selective-reject ARQ

In the data transmission model, which covers the typical case it is assumed that a user generates a sequence of information blocks for transmission. The ARQ mechanism requires the block to contain a header with control information that is essential to proper operation, as shown in Figure 15.1

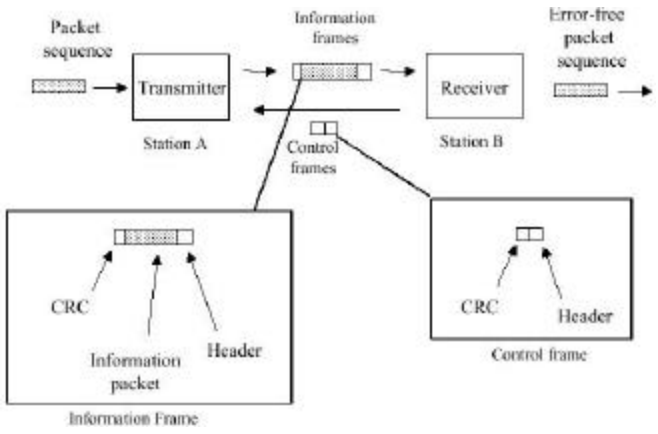


Figure 15.1 Basic elements of ARQ

The transmitter will also append CRC check bits that cover the header and the information bits to enable the receiver to determine whether errors have occurred during transmission.

We assume that the design of the CRC ensures that transmission errors can be detected with very high probability

In addition to the error-detection code, the other basic elements of ARQ protocols consist of **information frames (I-frames)** that transfer the user packets, control frames, and time-out mechanisms, as shown in Figure15.1. **Control frames** are short binary blocks that consist of a header that provides the control information followed by the CRC. The control frames include ACKs bits, which acknowledge the correct receipt of a given frame or group of frames; NAKs bits, which indicate that a frame has been received in error and that the receiver is taking certain action; and an **enquiry frame** ENQ, which commands the receiver to report its status. The time-out mechanisms are required to prompt certain actions to maintain the flow of frames. We can visualize the transmitter and receiver as working jointly on ensuring the correct and orderly delivery of the sequence of packets provided by the sender.

15.2 Stop-and-Wait ARQ

The first protocol considered is **Stop-and-Wait ARQ** where the transmitter and receiver work on the delivery of one frame at a time through an alternation of actions. Figure 15.2a shows how ACKs and time-outs can be used to provide recovery from transmission errors, in this case a lost frame. At the initial point in the figure, stations A and B are working on the transmission of frame 0. Note that each time station A sends an I-frame, it starts an **I-frame timer** that will expire after some time-out period. The time-out period is selected so that it is greater than the time required to receive the corresponding ACK frame. Figure 15.2 a shows the following sequence of events:

1. Station A transmits frame 0 and then waits for an ACK frame from the receiver.
2. Frame 0 is transmitted without error, so station B transmits an ACK frame.
3. The ACK from station B is also received without error, so station A knows the frame 0 has been received correctly.
4. Station A now proceeds to transmit frame 1 and then resets the timer.
5. Frame 1 undergoes errors in transmission. It is possible that station B receives frame 1 and detects the errors through the CRC check; it is also possible that frame 1 was so badly garbled that station B is unaware of the transmission. In either case station B does not take any action.
6. The time-out period expires, and frame 1 is retransmitted.

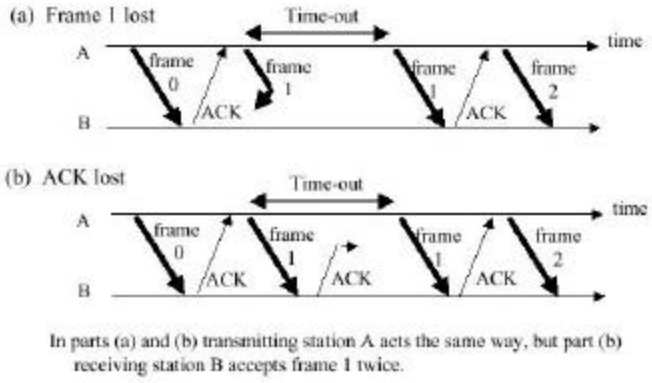


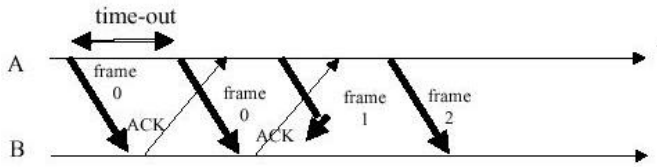
Figure 15.2 Possible ambiguities when frames are unnumbered

The protocol continues in this manner until frame 1 is received and acknowledged. The protocol then proceeds to frame 2, and so on.

Transmission errors in the reverse channel lead to ambiguities in the Stop-and-Wait protocol that need to be corrected. Figure

15.2 b shows the situation that begins as in Figure 15.2 a, but where frame 1 is received correctly, and its acknowledgment undergoes errors. After receiving frame 1 station B delivers its contents to the destination. Station A does not receive the acknowledgment for frame 1, so the time-out period expires. Note that at this point station A cannot distinguish between the sequence of events in parts (a) and (b) of Figure 15.2 . Station A proceeds to retransmit the frame. If the frame is received correctly by station B, as shown in the figure, then station B will accept frame 1 as a new frame and redeliver it to the user. Thus we see that the loss of an ACK can result in the delivery of a duplicate packet. The ambiguity can be eliminated by including a sequence number in the header of each I-frame. Station B would then recognize that the second transmission of frame 1 was a duplicate, discard the frame, and resend the ACK for frame 1.

A second type of ambiguity arises if the ACKs do not contain a sequence number. In Figure 15.3 frame 0 is transmitted, but the time-out expires prematurely. Frame 0 is received correctly, and the (unnumbered) ACK is returned. In the meantime station A has resent frame 0.



Transmitting station A misinterprets duplicate ACKs

Figure 15.3 Possible ambiguities when ACKs are unnumbered

Shortly thereafter, station A receives an ACK and assumes it is for the last frame. Station A then proceeds to send frame 1, which incurs transmission errors. In the meantime the second transmission of frame 0 has been received and acknowledged by station B. When station A receives the second ACK, the station assumes the ACK is for frame 1 and proceeds to transmit frame 2. The mechanism fails because frame 1 is not delivered. This example shows that premature time-outs (or delayed ACKs) combined with loss of I-frames can result in gaps in the delivered packet sequence. This ambiguity is resolved by providing a sequence number in the acknowledgment frames that enables the transmitter to determine which frames have been received.

Stop-and-Wait ARQ becomes inefficient when the propagation delay is much greater than the time to transmit a frame. For example, suppose that we are transmitting frames that are 1000 bits long over a channel that has a speed of 1.5 megabits/ second and suppose that the time that elapses from the beginning of the frame transmission to the receipt of its acknowledgment is 40 ms. The number of bits that can be transmitted over this channel in 40 ms is $40 \times 10^{-3} \times 1.5 \times 10^{6} =$ 60,000 bits. However, Stop-and-Wait ARQ can transmit only 1000 bits in this period time. This severe inefficiency is due to the requirement that the transmitter wait for the acknowledgment of a frame before proceeding with other transmissions. The situation becomes much worse in the presence of transmission errors that trigger retransmissions.

The **delay-bandwidth product** is the product of the bit rate and the delay that elapses before an action can take place. In the preceding example the delay-bandwidth product is 60,000 bits. In Stop-and-Wait ARQ the delay-bandwidth product can be viewed as a measure of lost opportunity in terms of transmitted bits. This factor arises as a fundamental limitation in many network problems.

Notes

LESSON 16

- Go-Back-N ARQ
- Selective Repeat ARQ

In this section we show that the inefficiency of Stop-and-Wait ARQ can be overcome by allowing the transmitter to continue sending enough frames so that the channel is kept busy while the transmitter waits for acknowledgments.

Suppose for now that frames are numbered 0,1,2,3,... The transmitter has a limit on the number of frames $\mathbf{W_s}$ that can be outstanding. W_s is chosen larger than the delay-band-width product to ensure that the channel can be kept busy.

The idea of the basic **Go-Back-N ARQ** is as follows: Consider the transfer of a reference frame, say, frame 0. After frame 0 is sent, the transmitter sends (Ws – 1) additional frames into the channel, optimistic that frame 0 will be received correctly and not require retransmission. If things turn out as expected, an ACK for frame 0 will arrive in due course while the transmitter is still busy sending frames into the channel, as shown in Figure16.1 . The system is now done with frame 0. Note, however, that the handling of frame 1 and subsequent frames is already well underway. A procedure where the processing of a new task is begun before the completion of the previous task is said to be **pipelined.** In effect Go-Back-N ARQ pipelines the processing of frames to keep the channel busy.

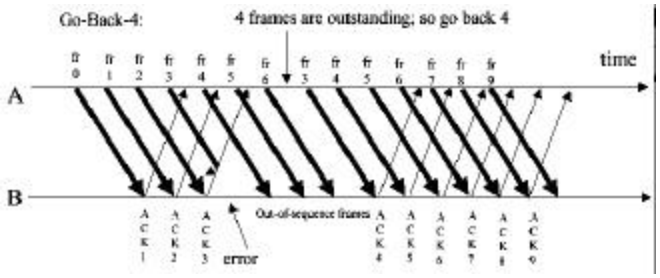


Figure 16.1 Basic Go-Back-N ARQ

Go-Back-N ARQ gets its name from the action that is taken when an error occurs. As shown in Figure16.1, after frame 3 undergoes transmission errors, the receiver ignores frame 3 and all subsequent frames. Eventually the transmitter reaches the maximum number of outstanding frames. It is then forced to go back N frames, where N = Ws, and begin retransmitting all packets from 3 onwards.

The Go-Back-N ARQ as stated above depends on the transmitter exhausting its maximum number of outstanding frames to trigger the retransmission of a frame. Thus this protocol works correctly as long as the transmitter has an unlimited supply of packets that need to be transmitted. In situations where packets arrive sporadically, there may not be (Ws – 1) subsequent transmissions. In this case retransmissions are not triggered, since the window is not exhausted. This problem is easily resolved by modifying Go-Back-N ARQ such that a timer is associated with each transmitted frame.

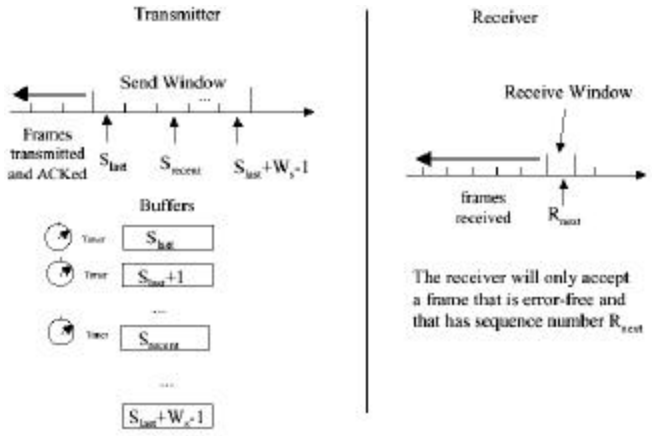


Figure 16.2 Go-Back-N ARQ

Figure 16.2 shows how the resulting Go-Back-N **ARQ** protocol operates. The transmitter must now maintain a list of the frames it is processing, where Slast is the number of the last transmitted frame that remains unacknowledged and Srecent is the number of the most recently transmitted frame. The transmitter must also maintain a timer for each transmitted frame and must also buffer all frames that have been transmitted but have not yet been acknowledged. At any point in time the transmitter has a **send window** of available sequence numbers. The lower end of the window is given by Slast, and the upper limit of the transmitter window is (Slast + Ws –1). If Srecent reaches the upper limit of the window, the transmitter is not allowed to transmit further new frames until the send window slides forward with the receipt of a new acknowledgment.

16.2 Selective Repeat ARQ

In channels that have high error rates, the Go-Back-N ARQ protocol is inefficient because of the need to retransmit the frame in error and all the subsequent frames. A more efficient ARQ protocol can be obtained by adding two new features: first, the receive window is made larger than one frame so that the receiver can accept frames that are out of order but error free; second, the retransmission mechanism is modified so that only individual frames are retransmitted. This protocol is referred to as **Selective Repeat ARQ.**

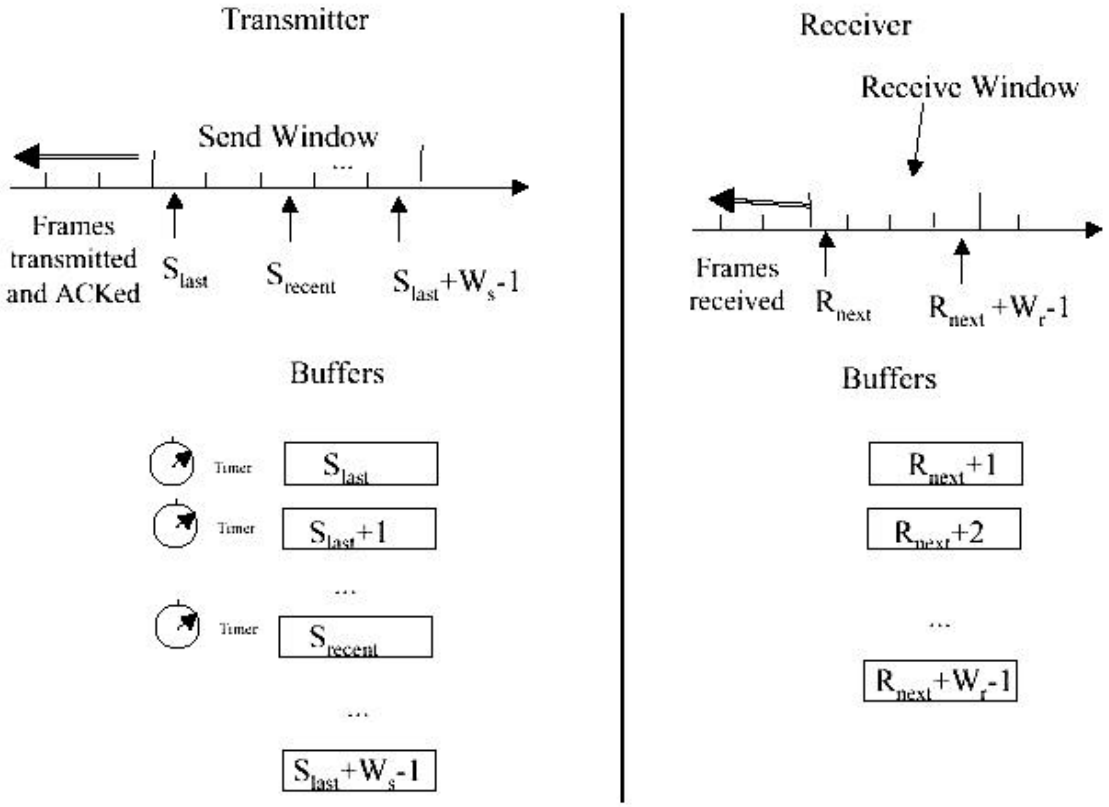


Figure 16.3 Selective Repeat ARQ

We continue to work under the constraint that the ARQ protocol must deliver an error-free and ordered sequence of packets to the destination. Figure 16.3 shows that the send window at the transmitter is unchanged but that the receive window now consists of a range of frame numbers spanning from Rnext to (Rnext + Wr –1), where Wr is the maximum number of frames that the receiver is willing to accept at a given time. As before, the basic objective of the protocol is to advance the values of Rnext and Slast through the delivery of the oldest outstanding frame. Thus ACK frames carry Rnext the oldest frame that has not yet been received. The receive window is advanced with the receipt of an error-free frame with sequence number Rnext.

Unlike the case of Go-Back-N ARQ, the receive window may be advanced by several frames. This step occurs when one or more frames that follow Rnext have already been received correctly and are buffered in the receiver. Rnext and the following consecutive packets are delivered to the destination at this point. Now consider the retransmission mechanism in Selective Repeat ARQ. The handling of timers at the transmitter is done as follows. When the timer expires, only the corresponding frame is retransmitted.

Notes

LESSON 17

Today's Topics

- Flow Control
- Sliding Window Flow Control
- HDLC

17.1 Flow Control

Flow control is a technique used for assuring that a receiving entity is not overwhelmed by the data from the transmitting entity. At the receiving entity the data received from the transmitter are buffered till they are processed and passed on to the higher layer software and it allocates some finite data memory for this buffering purpose. In the absence of flow control, the receiver's buffer may fill up and overflow while it is still processing old data.

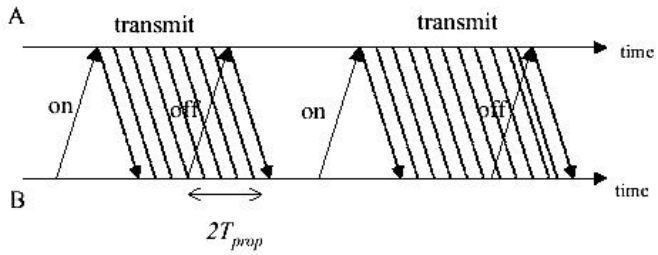


Figure 17.1 On-OFF Flow Control

The simplest procedure for exercising flow control is to use signals that direct the sender to stop transmitting the data. Suppose entity A is transmitting to entity B at a rate R bps. If entity B detects that its buffer are filling up, it issues a stop signal to entity A. After approximately one propagation delay Tprop entity A stops transmitting as shown in Figure 17.1

From the instant that B sent its signal, it receives an additional 2Tprop R bits, which is equal to the delay-bandwidth product of the link. Thus entity B must send the off signal when its buffer contents exceed a threshold value. This type of flow control is used in the X-ON / X-OFF protocol that is used between a terminal and a computer. This is also used in various Data link controls.

In the previous section we discussed ARQ sliding-window protocols. The objective there was to provide a reliable transfer of a sequence of data over an unreliable communication channel. Here we show how some of the elements of ARQ protocols can in fact provide flow control functionality also.

17.2 Sliding-Window Flow Control

The sliding-window protocols that were used in ARQ mechanism can also be used for flow control. In the simplest case the size of the send window Ws is made equal to the number of buffers that are available at the receiver. Because Ws is the maximum number of outstanding frames from the transmitter, buffer overflow cannot occur at the receiver.

Figure 17.2 shows an example where receiver sends an acknowledgement after the last frame in a window has been received. In this figure tcycle is the basic delay that elapses from the time the first frame is transmitted to the receipt of its acknowledgement.

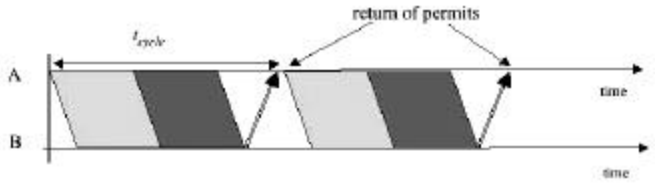


Figure 17.2 Sliding window Flow Control

The delay in sending the acknowledgements has the effect of pacing or controlling the rate at which the transmitter sends frames to the receiver.

17.3 High-Level Data Link Control (HDLC)

The most important data link control protocol is HDLC (ISO 33009, ISO 4335). Not only is HDLC widely used, but also it is the basis for many other important data link control protocols, which use the same or similar formats and the same mechanisms as employed in HDLC. Accordingly, in this section we provide a detailed discussion of HDLC.

17.3.1 Basic Characteristics of HDLC

To satisfy a variety of applications, HDLC defines three types of stations, two link configurations, and three data-transfer modes of operation.

The three station types are

- **Primary station.** Has the responsibility for controlling the operation of the link. Frames issued by the primary are called commands.
- **Secondary station.** Operates under the control of the primary station. Frames issued by a secondary are called responses. The primary maintains a separate logical link with each secondary station on the line.
- **Combined station.** Combines the Features of primary and secondary. A combined station may issue both commands and responses.

The two link configurations are

- **Unbalanced configuration.** Consists of one primary and one or more secondary stations and supports both full-duplex and half-duplex transmission.
- **Balanced configuration.** Consists of two combined stations and supports both full-duplex and half-duplex transmission.

The three data transfer modes are

- **Normal Response Mode** (NRM). Used with an unbalanced configuration. The primary may initiate data transfer to a secondary, but a secondary may only transmit data in response to a command from the primary.
- **Asynchronous balanced mode** (ABM). Used with a balanced configuration. Either combined station may initiate transmission without receiving permission from the other combined station.
- **Asynchronous response mode** (ARM). Used with an unbalanced configuration. The secondary may initiate transmission without explicit permission of the primary. The primary still retains responsibility for the line, including initialization, error recovery, and logical disconnection.

NRM is used on mulitdrop lines, in which a number of terminals are connected to a host computer. The computer polls each terminal for input. NRM is also sometimes used on point-to-point links, particularly if the link connects a terminal or other peripheral to a computer. ABM is the most widely used of the three modes; it makes more efficient use of a full-duplex point-to-point link as there is no polling overhead. ARM is rarely used; it is applicable to some special situations in which a secondary may need to initiate transmission.

Notes

LESSON 18

Today's Topics

- HDLC Frame Structure
- Bit Stuffing

18.1 HDLC Frame Structure

HDLC uses synchronous transmission. All transmissions are in the form of frames, and a single frame format suffices for all types of data and control exchanges. Figure 18.1a depicts the structure of the HDLC frame. The flag, address, and control fields that precede the information field are known as a header. The FCS and flag fields following the data or information field are referred to as a trailer.

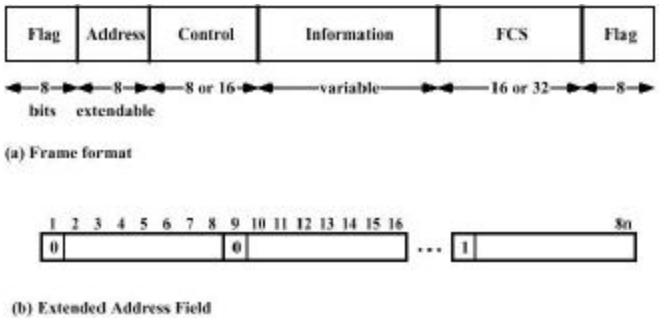


Figure 18.1 (a), (b) HDLC frame structure

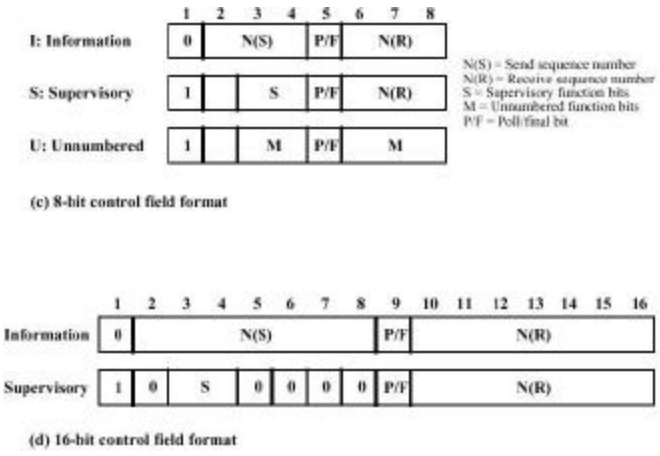


Figure 18.1 (c), (d) HDLC frame structure

18.1.1 Flag Fields

Flag fields delimit the frame at both ends with the unique pattern 01111110. A single flag may be used as the closing flag for one frame and the opening flag for the next. On both sides of the user-network interface, receivers are continuously hunting for the flag sequence to synchronize on the start of a frame. While receiving a frame, a station continues to hunt for that sequence to determine the end of the frame. However, it is possible that the pattern 01111110 will appear somewhere inside the frame, thus destroying frame-level synchronization. To avoid this, a procedure known as bit stuffing is used.

Between the transmission of the starting and ending flags, the transmitter will always insert an extra 0 bit after each occurrence of five 1s in the frame. After detecting a starting flag, the receiver monitors the bit stream. When a pattern of five 1s appears, the sixth bit is examined. If this bit is 0, it is deleted. If the sixth bit is a 1 and the seventh bit is a 0, the combination is accepted as a flag. If the sixth and seventh bits are both 1, the sender is indicating an abort condition.

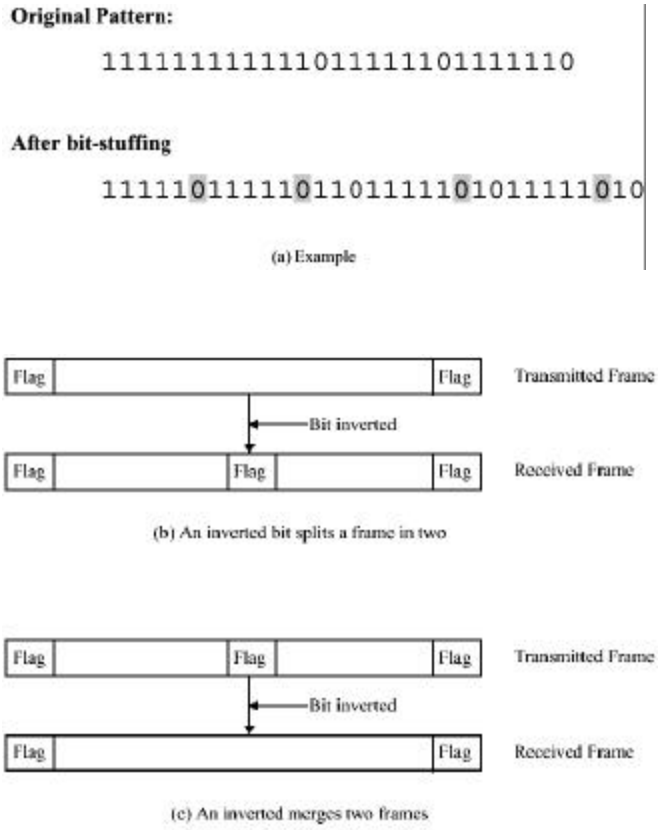


Figure 18.2 Bit stuffing

With the use of bit stuffing, arbitrary bit patterns can be inserted into the data field of the frame. This property is known as data transparency.

Figure 18.2 shows an example of bit stuffing. Note that in the first two cases, the extra 0 is not strictly necessary for avoiding a flag pattern, but is necessary for the operation of the algorithm. The pitfalls of bit stuffing are also illustrated in this figure. When a flag is used as both an ending and a starting flag, a 1-bit error merges two frames into one; conversely, a 1-bit error inside

the frame could split it in two. However these errors will be detected by the CRC checksum - FCS field and there will be retransmission of those frames.

18.1.2 Address Field

The address field identifies the secondary station that transmitted or is to receive the frame. This field is not needed for point-to-point links, but is always included for the sake of uniformity. The address field is usually eight bits long but, by prior agreement, an extended format may be used in which the actual address length is a multiple of seven bits (Figure18.1 b). The least significant bit of each octet is 1 or 0 depending on whether it is or is not the last octet of the address field. The remaining seven bits of each octet form part of the address. The single-octet address of 11111111 is interpreted as the all-stations address in both basic and extended formats. It is used to allow the primary to broadcast a frame for reception by all secondaries.

18.1.3 Control Field

HDLC defines three types of frames, each with a different control field format. Information frames (I-frames) carry the data to be transmitted for the user (the logic above HDLC that is using HDLC). Additionally, flow and error-control data using the ARQ mechanism, are piggybacked on an information frame. Supervisory frames (S-frames) provide the ARQ mechanism when piggybacking is not used. Unnumbered frames (U-frames) provide supplemental link control functions. The first one or two bits of the control field serves to identify the frame type. The remaining bit positions are organized into subfields as indicated in Figure 18.1 c and d. Their use is explained below in the discussion of HDLC operation.

Note that the basic control field for S and I-frames uses 3-bit sequence numbers. With the appropriate set-mode command, an extended control field can be used for S- and I-frames that employs 7-bit sequence numbers. U-frames always contain an 8-bit control field.

18.1.4 Information Field

The information field is present only in I-frames and some U-frames. The field can contain any sequence of bits but must consist of an integral number of octets. The length of the information field is variable up to some system-defined maximum.

18.1.5 Frame Check Sequence Field

The frame check sequence (FCS) is an error-detecting code calculated from the remaining bits of the frame, exclusive of flags. The normal code is the 16-bit CRC. An optional 32-bit FCS, using CRC-32, may be employed if the frame length or the line reliability dictates this choice.

Notes

LESSON 19

Today's Topics

- HDLC Operation
- Bisync protocol

Objectives

- Describe the operation of high level data link control
- Describe the bisync protocol

19.1 HDLC Operation

HDLC operation consists of the exchange of I-frames, S-frames, and U-frames between two stations. In describing HDLC operation, we will discuss these three types of frames.

The operation of HDLC involves three phases. First, one side or another initializes the data link so that frames may be exchanged in an orderly fashion. During this phase, the options that are to be used are agreed upon. After initialization, the two sides exchange user data and the control information to exercise flow and error control. Finally, one of the two sides signals the termination of the operation.

19.1.1 Initialization

Initialization may be requested by either side by issuing one of the six set-mode commands. This command serves three purposes:

1. It signals the other side that initialization is requested.
2. It specifies which of the three modes (NRM, ABM, ARM) is requested.
3. It specifies whether 3- or 7-bit sequence numbers are to be used.

If the other side accepts this request, then the HDLC module on that end transmits an unnumbered acknowledged (UA) frame back to the initiating side. If the request is rejected, then a disconnected mode (DM) frame is sent.

19.1.2 Data Transfer

When the initialization has been requested and accepted, then a logical connection is established. Both sides may begin to send user data in I-frames, starting with sequence number 0. The N(S) and N(R) fields of the I-frame are sequence numbers that support flow control and error control. An HDLC module sending a sequence of I-frames will number them sequentially, modulo 8 or 128, depending on whether 3- or 7-bit sequence numbers are used, and place the sequence number in N(S). N(R) is the acknowledgment for I-frames received; it enables the HDLC module to indicate which number I-frame it expects to receive next. S-frames are also used for flow control and error control.

19.1.3 Disconnect

Either HDLC module can initiate a disconnect, either on its own initiative if there is some sort of fault, or at the request of its higher-layer user. HDLC issues a disconnect by sending a disconnect (DISC) frame. The other side must accept the disconnect by replying with an acknowledgement.

19.2 Bisync Protocol

Bisync is an abbreviation shortened from "binary synchronous". Sometimes you may also see the acronym BSC. Bisync is a block-oriented, error-correcting, synchronous data communications protocol introduced by IBM back in 1964 with the introduction of a product called the 270X Transmission Control Unit.

To be precise, 2780 and 3780 were model numbers of IBM remote job entry (RJE) data terminals — namely the. IBM 2780 Data Communications Terminal and theIBM 3780 Data Communications Terminal .These terminals used punch cards and consisted of a card reader, a card punch, and a line printer. They used the bisync protocol to transmit and receive data with an IBM mainframe computer. Usually dial-up or leased telephone lines and 2400 bps Bell 201C modems and then later 4800 bps Bell 208B modems were used to connect the terminal to the mainframe.RJE was how programs, often referred to as jobs, were submitted to be run on mainframe computers back in the 1960 and 1970s. That was the era of keypunch machines and punched cards. The statements for a computer program, usually COBOL or FORTRAN, and the input data for the programs were punched onto cards using a keypunch machine. The resulting card deck was carried, often wheeled over on carts, to an RJE terminal, placed into a card reader hopper, a button was pushed, and the card images were transmitted to the mainframe.

An immediate response may have come back to printer, exchange, or card punch devices of the terminal. This output might have been the result of program just submitted, or from some previous run. Very frequently, the program and data were held at the mainframe for execution at a later time. This is known as batch processing.

The 3780 terminal was a later model than the 2780 terminal and used a more robust version of the bisync protocol — hence the terms "3780 bisync" vs. "2780 bisync". Virtually all bisync in use today conforms with the 3780 version.

While it is true that "real" IBM 3780 and 2780 terminals are not in use today, the underlying bisync protocol became the defacto standard file transfer protocol for a wide array of devices in the days before the PC revolutionized computing. Much like Zmodem and FTP today, if you needed to get a file from one machine to another during that time, very often bisync was protocol used.

And bisync wasn't solely used in "true" computers. The bisync protocol ended up in ATM machines, check sorting machines, radar systems, cash registers, radio dispatching systems, telephone switches, and countless other devices.

This massive array of hardware is not about to disappear overnight. There is still a huge installed base of bisync-equipped machinery in North America and to a lesser extent in the rest of the world.If somehow all of the bisync interconnected machines in the world were to disappear all at once, the results would be catastrophic. Many banks would cease to function. Some air traffic control systems would collapse. Many of the point-of-sale systems in retail stores would fail. Many credit and debit cards would become useless. EDI (electronic data interchange) networks that manage much of the business-to-business commerce would crash. There is no doubt that bisync is still a vital link in the chain of the world's computer infra-structure

Notes

LESSON 20

Today's Topics

- Multiplexing
- TDM
- FDM

Objectives

Upon completion of this Lesson, you should be able to:

- Define Modulation
- Distinguish between different types of Modulation
- Exlpain the advantages and disadvantages if any of each type of modulation

20.1 Multiplexing

involves the sharing of expensive network resources by several connections or information flows. The network resource of primary importance is bandwidth of the communication channel, which is measured in Hertz for analog transmission system and bits/second for digital transmission system. Here we consider multiplexing techniques that are used to share a set of transmission lines among a community of users.

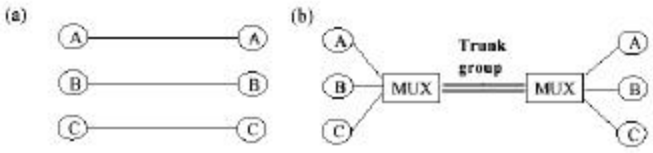


Figure 20.1 Multiplexing

Figure 20.1 a shows an example where three pairs of users communicate by using three separate sets of wires. This method becomes inefficient as the number of users increases. A better approach is to dynamically share the network resources among a community of users. Figure 20.1 b show a multiplexer which allows this sharing. When a user wants to communicate with another user at the other end the multiplexer dynamically assigns a communication line for the duration of the call. When the call is completed, the transmission line is returned to the pool that is available to meet new connection requests. Note that signaling is required between two multiplexer to set up and terminate each call.

These multiplexing schemes can be divided into two basic categories: FDM **(Frequency Division Multiplexing),** and TDM (**Time Division Multiplexing).** In FDM the frequency spectrum is divided among the logical channels, with each user having exclusive possession of some frequency band. In TDM the users take turns (in a round robin), each one periodically getting the entire bandwidth for a little burst of time.

20.1.1 Frequency-Division Multiplexing (FDM)

Suppose that the transmission system line has a bandwidth that is much greater than the required by a single connection. For example in Figure 20.2 each user has a signal of W Hz and the channel that is available is greater than 3W Hz. In such a case available bandwidth can be shared by the individual users and each user will have the required bandwidth at his disposal for the complete duration of allotment.

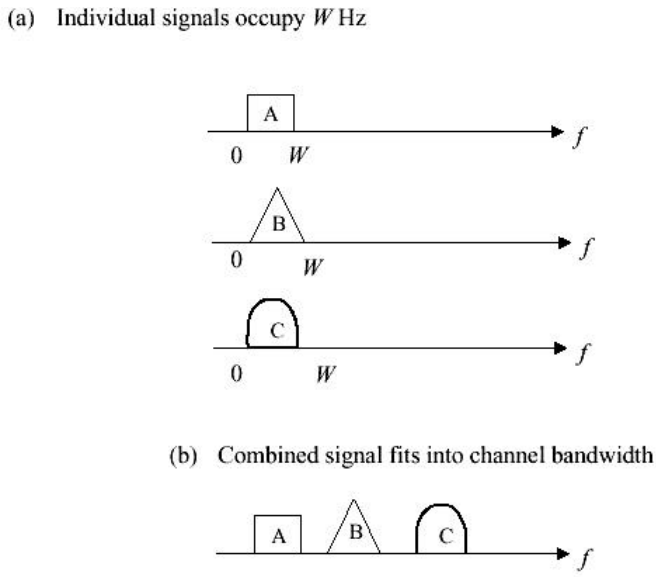


Figure 20.2 Frequency Division Multiplexing, FDM

In **Frequency-division multiplexing (FDM),** the bandwidth is divided into a number of frequency slots, each of which can accommodate the signal of an individual connection. The multiplexer assigns a frequency slot to each connection and uses modulation with appropriate carrier frequencies to place the signal of the different connection in corresponding frequency slot.

This process results in an overall combined signal that carries all the connections as shown in Figure 20.2 b. The combined signal is transmitted, and the demultiplexer recovers the signals corresponding to each connection. Reducing the number of wires that need to be handled reduces the overall cost of the system.

FDM was introduced in the telephone network in the 1930s. The basic analog multiplexer combines 12 voice channels in one line. Each voice signal occupies 4 kHz of bandwidth. The multiplexer modulates each voice signal so that it occupies a 4 kHz slot in the band between 60 and 108 kHz. The combined signal is called a group. A hierarchy of analog multiplexers has been defined. For example, a supergroup (that carries 60 voice signals) is formed by multiplexing five groups, each of bandwidth 48 kHz, into the frequency band from 312 to 552 kHz. Note that for the purposes of multiplexing, each group is

treated as an individual signal. Ten supergroups can then be multiplexed to form a mastergroup of 600 voice signals that occupies the band 564 to 3084 kHz. Various combinations of mastergroups have also been defined.

Familiar examples of FDM are broadcast radio and broadcast cable television, where each station has an assigned frequency band. Stations in AM, FM, and television are assigned frequency bands of 10 kHz, 200 kHz, and 6 MHz, respectively. FDM is also used in cellular telephony where a pool of frequency slots, typically of 25 to 30 kHz each, are shared by the users within a geographic cell. Each user is assigned a frequency slot for each direction. Note that in FDM the user information can be in analog or digital form and that the information from all the users flows simultaneously.

20.1.2 Time Division Multiplexing TDM

In time-division multiplexing (TDM), the transmission between the multiplexers is provided by a single high-speed digital transmission line. Each connection produces a digital information flow that is then inserted into the high-speed line. For example in Figure 20.3 a each connection generates a signal that produces one unit of information every 3T seconds. This unit of information could be a bit, a byte, or a fixed-size block of bits. Typically, the transmission line is organized into frames that in turn are divided into equal-sized slots. For example, in Figure 20.3 b the transmission line can send one unit of information every T seconds, and the combined signal has a frame structure that consists of three slots, one for each user. During connection setup each connection is assigned a slot that can accommodate the information produced by the connection.

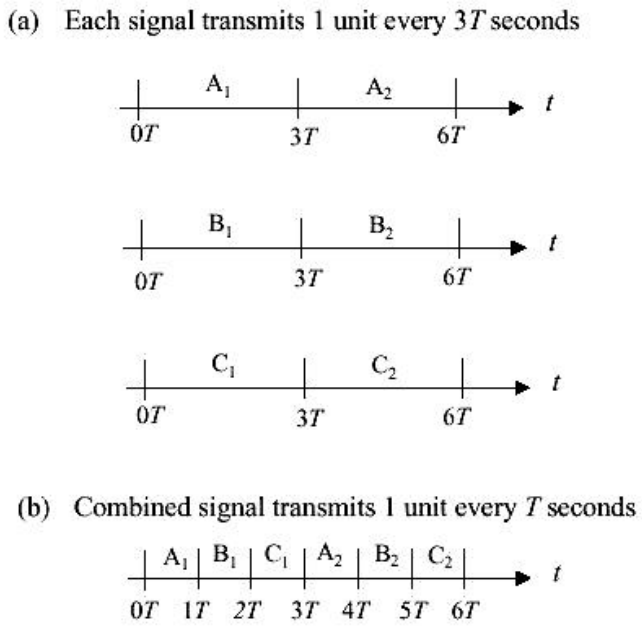


Figure 20.3 Time Division Multiplexing

TDM was introduced in the telephone network in the early 1960s. The T1 carrier system that carries 24 digital telephone connections is shown in Figure 20.4

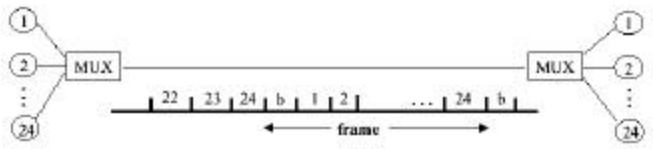


Figure 20.4 Time Division Multiplexing

Notes

Today's Topics

- Switching Basics

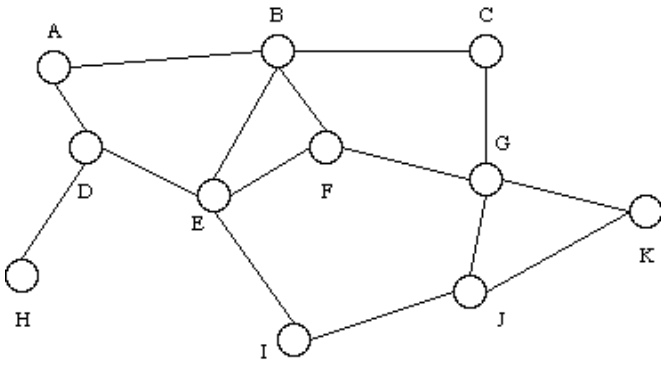
21.1 Switching Basics

Often multiple devices have to be connected over long distances so that they can communicate with each other.. At such times bus and ring topologies cannot be used because of the distance and large number of nodes.At such times bus and ring topologies cannot be used because of the distance and large no: of nodes.

As the number of connected users increased, it has become infeasible to provide a circuit which connects every user to every other user, and some sharing of the transmission circuits (known as "switching") has become necessary. To accomplish this goal, the data communications network has evolved. A network is a set of nodes that are interconnected to permit the exchange of information. Three switching techniques have been proposed for building networks:

- Circuit Switching
- Packet Switching
- Message Switching

Each allows sharing communication facilities among multiple users, and each uses equipment located at the nodes to replace the patch-panels used in a point-to-point connection. Packet switching is most often used for data communication. Most networks consist of many links (see the figure below) which allow more than one path through the network between nodes. A data communications network must be able to select an appropriate path for each required connection.

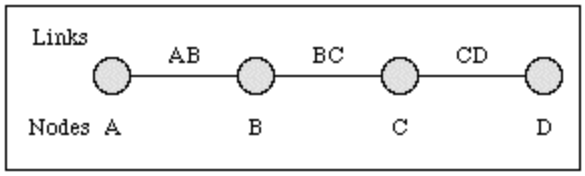


Any of the three approaches could yield minimum delay in a particular situation, though situations where message switching yields minimum delay are rare. The relative performance of circuit switching and packet switching depends strongly on the speed and "cost" of establishing a connection.

21.2 Circuit Switching

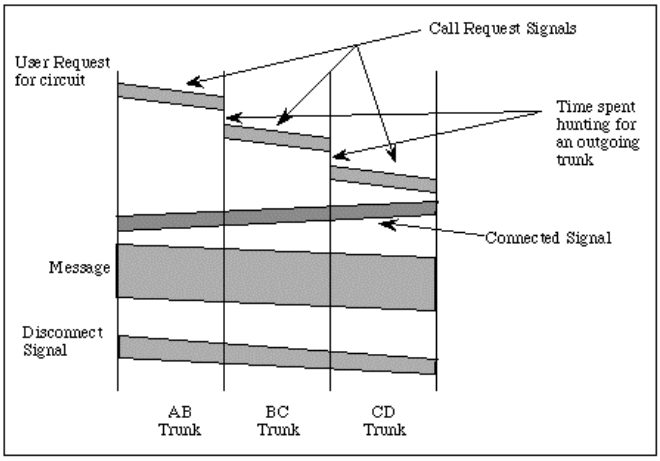
In a circuit-switched network, a dedicated communication path is established between two stations through the nodes of the network. That path is a connected sequence of physical links between nodes. On each link, a logical channel is dedicated to the connection. Data generated by the source station are transmitted along the dedicated path as rapidly as possible. At each node, incoming data are routed or switched to the appropriate outgoing channel without delay. The most common example of circuit switching is the telephone network.

Circuit switching is the most familiar technique used to build a communications network. It allows communications equipment and circuits, to be shared among users. Each user has sole access to a circuit (functionally equivalent to a pair of copper wires) during network use. Consider communication between two points A and D in a network. The connection between A and D is provided using (shared) links between two other pieces of equipment, B and C.



A connection between two systems A & D formed from 3 links

Network use is initiated by a connection phase, during which a circuit is set up between source and destination, and terminated by a disconnect phase. These phases, with associated timings, are illustrated in the figure below.



(Information flows in two directions. Information sent from the calling end is shown in pink and information returned from the remote end is shown in blue)

After a user requests a circuit, the desired destination address must be communicated to the local switching node (B). In a telephony network, this is achieved by dialing the number.

Node B receives the connection request and identifies a path to the destination (D) via an intermediate node (C). This is followed by a circuit connection phase handled by the switching nodes and initiated by allocating a free circuit to C (link BC), followed by transmission of a call request signal from node B to node C. In turn, node C allocates a link (CD) and the request is then passed to node D after a similar delay.

The circuit is then established and may be used. While it is available for use, resources (i.e. in the intermediate equipment at B and C) and capacity on the links between the equipment are dedicated to the use of the circuit.

After completion of the connection, a signal confirming circuit establishment (a connect signal in the diagram) is returned; this flows directly back to node A with no search delays since the circuit has been established. Transfer of the data in the message then begins. After data transfer, the circuit is disconnected; a simple disconnect phase is included after the end of the data transmission.

Notes

LESSON 22

Today's Topics

- Packet Switching
- Message Switching

Objectives

Upon completion of this Lesson, you should be able to:

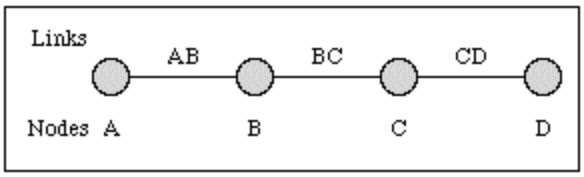
- Explain what is packet switching
- Explain what is message switching
- Compare packet and message switching

22.1 Packet Switching

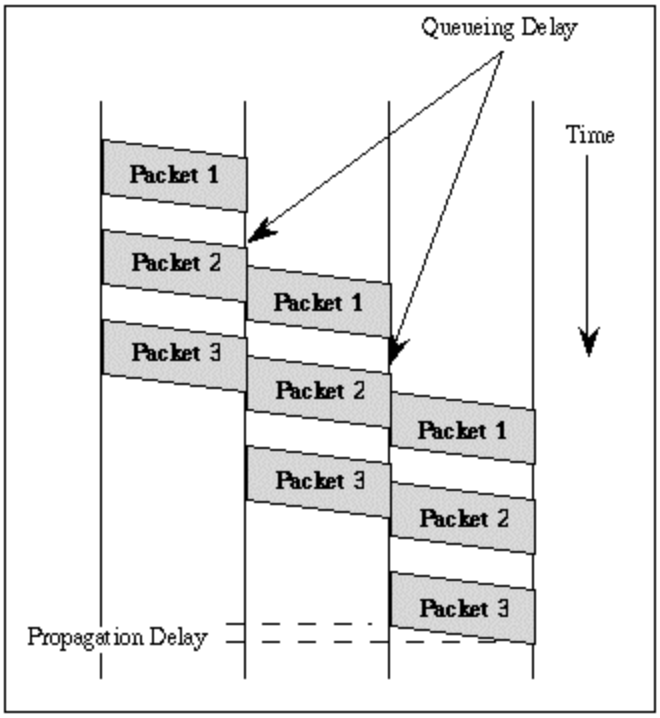
A quite different approach is used in a packet-switched network. In this case, it is not necessary to dedicate transmission capacity along a path through the network. Rather, data are sent out in a sequence of small chunks, called packets. Each packet is passed through the network from node to node along some path leading from source to destination. At each node, the entire packet is received, stored briefly, and then transmitted to the next node. Packet-switched networks are commonly used for terminal-to-computer and computer-to-computer communications.

Packet switching is similar to message switching using short messages. Any message exceeding a network-defined maximum length is broken up into shorter units, known as packets, for transmission; the packets, each with an associated header, are then transmitted individually through the network. The fundamental difference in packet communication is that the data is formed into packets with a pre-defined header format, and well-known "idle" patterns which are used to occupy the link when there is no data to be communicated.

A packet network equipment discards the "idle" patterns between packets and processes the entire packet as one piece of data. The equipment examines the packet header information (PCI) and then either removes the header (in an end system) or forwards the packet to another system. If the out-going link is not available, then the packet is placed in a queue until the link becomes free. A packet network is formed by links which connect packet network equipment.



Communication between A and D using circuits which are shared using packet switching.



Packet-switched communication between systems A and D

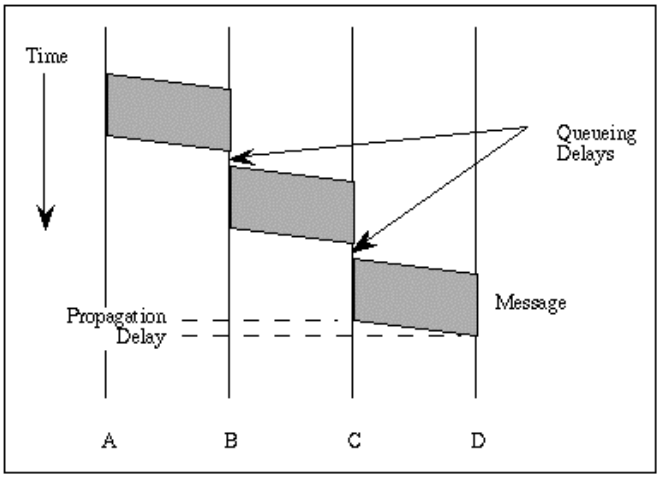
(The message in this case has been broken into three parts labeled 1-3)

22.1.1 There are two Important Benefits from Packet Switching

1. The first and most important benefit is that since packets are short, the communication links between the nodes are only allocated to transferring a single message for a short period of time while transmitting each packet. Longer messages require a series of packets to be sent, but do not require the link to be dedicated between the transmission of each packet. The implication is that packets belonging to other messages may be sent between the packets of the message being sent from A to D. This provides a much fairer sharing of the resources of each of the links.
2. Another benefit of packet switching is known as "pipelining". Pipelining is visible in the figure above. At the time packet 1 is sent from B to C, packet 2 is sent from A to B; packet 1 is sent from C to D while packet 2 is sent from B to C, and packet 3 is sent from A to B, and so forth. This simultaneous use of communications links represents a gain in efficiency, the total delay for transmission across a packet network may be considerably less than for message

switching, despite the inclusion of a header in each packet rather than in each message.

22.2 Message Switching

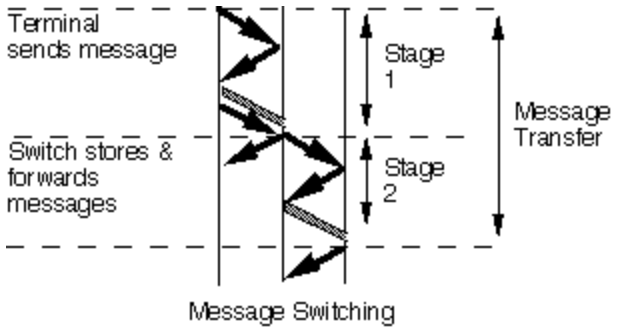


The use of message switching to communicate between A and D

The figure illustrates message switching; transmission of only one message is illustrated for simplicity. As the figure indicates, a complete message is sent from node A to node B when the link interconnecting them becomes available. Since the message may be competing with other messages for access to facilities, a queuing delay may be incurred while waiting for the link to become available. The message is stored at B until the next link becomes available, with another queuing delay before it can be forwarded. It repeats this process until it reaches its destination.

Circuit setup delays are replaced by queuing delays. Considerable extra delay may result from storage at individual nodes. A delay for putting the message on the communications link (message length in bits divided by link speed in bps) is also incurred at each node en route. Message lengths are slightly longer than they are in circuit switching, after establishment of the circuit, since header information must be included with each message; the header includes information identifying the destination as well as other types of information.

Most message switched networks do not use dedicated point-to-point links and therefore a call must be set-up using a circuit switched network. The figure below illustrates the use of message switching over a circuit switched network, in this case using one intermediate message switch.



Message switching using circuit switched connections between message switches.

Although message switching is still used for electronic mail and telex transmission, it has largely been replaced by packet switching.(in fact, most electronic mail is carried using message switching with the links between message switches provided by packet or circuit switched networks).

Notes

LESSON 23

Today's Topics

- Frame Relay
- Frame Format
- Data Integrity
- Virtual Circuits

Objectives

Upon completion of this Lesson, you should be able to:Upon completion of this Lesson, you should be able to:

- Explain what is frame relay
- State the advantages/disadvantages of frame relay

23.1 Frame Relay

Frame Relay is a simplified form of Packet Switching,in which synchronous frames of data are routed to different destinations depending on header information. Frame Relay switches packets end to end much faster, but **there is no guarantee of data integrity at all.**

Because Frame Relay does not 'care' whether the frame it is switching is error free or not, a Frame Relay node can start switching traffic out onto a new line as soon as it has read the first two bytes of addressing information at the beginning of the frame. Thus a frame of data can travel end to end, passing through several switches, and still arrive at its destination with only a few bytes delay. These delays are small enough that network latency under Frame Relay is not noticeably different from direct leased line connections. So the performance of a Frame Relay network is virtually identical to that of a leased line, but because most of the network is shared, costs are lower.

23.1.1 Frame Format

Frame Relay uses the synchronous HDLC frame format up to 4kbytes in length. Each frame starts and ends with a Flag character (7E Hex). The first 2 bytes of each frame following the flag contain the information required for multiplexing across the link. The last 2 bytes of the frame are always generated by a Cyclic Redundancy Check (CRC) of the rest of the bytes between the flags. The rest of the frame contains the user data.

23.1.2 Data Integrity

There is none. The network delivers frames, whether the CRC check matches or not. It does not even necessarily deliver all frames, discarding frames whenever there is network congestion. Thus it is usual to run an upper layer protocol above Frame Relay that is capable of recovering from errors, such asTCP/IP, X.25 or IPX.

In practice, however, the network delivers data quite reliably. Unlike the analog communication lines that were used in the past, modern digital lines have very low error rates. Very few frames are discarded by the network, particularly at this time when the networks are operating at well below design capacity.

23.1.3 Virtual Circuits

Packets are routed through one or more Virtual Circuits known as Data Link Connection Identifiers (**DLCI**s). Most Virtual Circuits are Permanent Virtual Circuits or PVCs, which means that the network provider sets up all DLCI connections at subscription time. Switched Virtual Circuits (**SVC**s) are also part of the Frame Relay specification. They provide a link that only lasts only as long as the session.

23.1.4 Flow Control and Information Rates

There is no true flow control on Frame Relay. The network simply discards frames it cannot deliver. However, the protocol does include features designed to control and minimize frame loss at the user level.When you subscribe, you will specify the line speed (e.g., 56kbps or T1) and also, typically, you will be asked to specify a Committed Information Rate (CIR) for each DLCI. This value specifies the maximum average data rate that the network undertakes to deliver under "normal conditions". If you send faster than the CIR on a given DLCI, the network will flag some frames with a Discard Eligibility (DE) bit. The network will do its best to deliver all packets but will discard any DE packets first if there is congestion. For example, your Frame Relay access may be a full T1 (1.54Mbps), but you may have subscribed to a CIR of only 512kbps. What happens is that the Access Node measures your average throughput over a time period, usually one second. If the average throughput is over 512kbps, then the 'extra' frames are marked with a DE bit, and will be discarded first.

Notes

LESSON 24

Today's Topics

- Asynchronous Transfer Mode
- ATM Devices and the Network Environment
- ATM Cell Basic Format
- ATM Devices
- ATM Network Interfaces

Objectives

Upon completion of this Lesson, you should be able to:Upon completion of this Lesson, you should be able to:

- Explain what is ATM switching
- Explain ATM Switches, ATM endpoints UNI,NNI

24.1 Asynchronous Transfer Mode Switching

Asynchronous Transfer Mode (ATM) is an International Telecommunication Union-Telecommunications Standards Section (ITU-T) standard for cell relay wherein information for multiple service types, such as voice, video, or data, is conveyed in small, fixed-size cells. ATM networks are connection-oriented. Figure 24.1 illustrates a private ATM network and a public ATM network carrying voice, video, and data traffic.

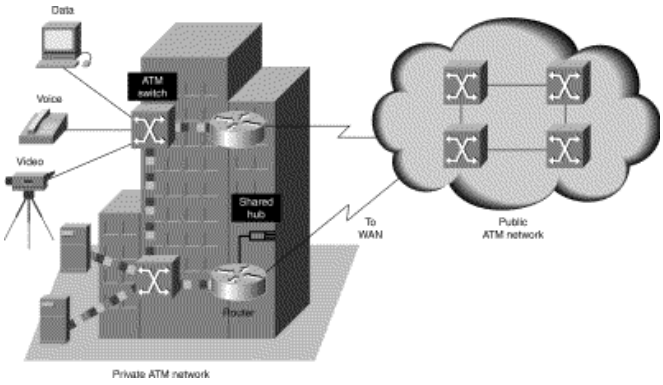


Figure 24.1: A Private ATM Network and a Public ATM Network Both Can Carry Voice, Video, and Data Traffic

24.2 ATM Devices and the Network Environment

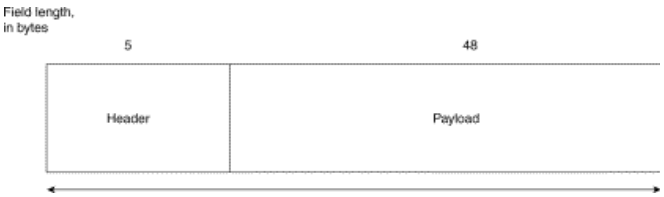
ATM is a cell-switching and multiplexing technology that combines the benefits of circuit switching (guaranteed capacity and constant transmission delay) with those of packet switching (flexibility and efficiency for intermittent traffic). It provides scalable bandwidth from a few megabits per second (Mbps) to many gigabits per second (Gbps). Because of its asynchronous nature, ATM is more efficient than synchronous technologies, such as time-division multiplexing (TDM).

With TDM, each user is assigned to a time slot, and no other station can send in that time slot. If a station has much data to send, it can send only when its time slot comes up, even if all other time slots are empty. However, if a station has nothing to transmit when its time slot comes up, the time slot is sent empty and is wasted. Because ATM is asynchronous, time slots are available on demand with information identifying the source of the transmission contained in the header of each ATM cell.

24.3 ATM Cell Basic Format

ATM transfers information in fixed-size units called cells. Each cell consists of 53 octets, or bytes. The first 5 bytes contain cell-header information, and the remaining 48 contain the payload (user information). Small, fixed-length cells are well suited to transferring voice and video traffic because such traffic is intolerant of delays that result from having to wait for a large data packet to download, among other things. Figure 23-2 illustrates the basic format of an ATM cell.

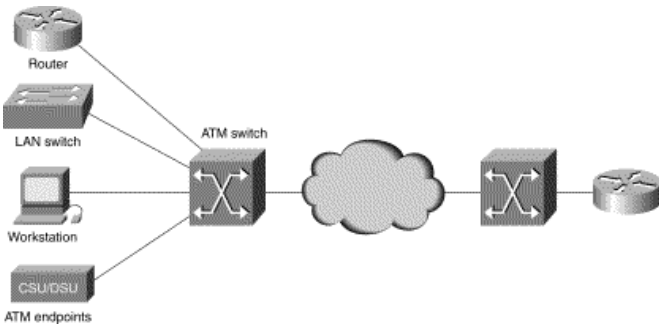
Figure 24-2: An ATM Cell Consists of a Header and Payload Data



24.4 ATM Devices

An ATM network is made up of an ATM switch and ATM endpoints. An ATM switch is responsible for cell transit through an ATM network. The job of an ATM switch is well defined: It accepts the incoming cell from an ATM endpoint or another ATM switch. It then reads and updates the cell header information and quickly switches the cell to an output interface toward its destination. An ATM endpoint (or end system) contains an ATM network interface adapter. Examples of ATM endpoints are workstations, routers, digital service units (DSUs), LAN switches, and video coder-decoders (CODECs). Figure 23-3 illustrates an ATM network made up of ATM switches and ATM endpoints.

Figure 24-3: An ATM Network Comprises ATM Switches and Endpoints



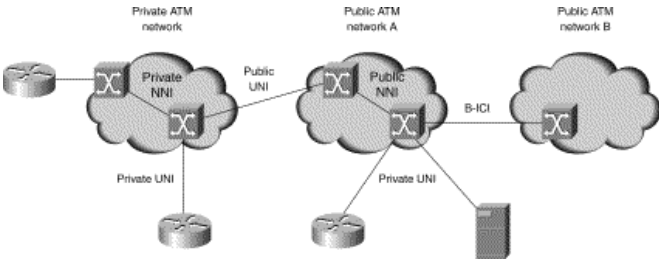
24.5 ATM Network Interfaces

An ATM network consists of a set of ATM switches interconnected by point-to-point ATM links or interfaces. ATM switches support two primary types of interfaces: UNI and NNI. The UNI connects ATM end systems (such as hosts and routers) to an ATM switch. The NNI connects two ATM switches.

Depending on whether the switch is owned and located at the customer's premises or is publicly owned and operated by the telephone company, UNI and NNI can be further subdivided into public and private UNIs and NNIs. A private UNI connects an ATM endpoint and a private ATM switch. Its public counterpart connects an ATM endpoint or private switch to a public switch. A private NNI connects two ATM switches within the same private organization. A public one connects two ATM switches within the same public organization.

An additional specification, the broadband intercarrier interface (B-ICI), connects two public switches from different service providers. Figure 23-4 illustrates the ATM interface specifications for private and public networks.

Figure 24-4: ATM Interface Specifications Differ for Private and Public Networks



Notes

LESSON 25

Today's Topics

- ATM Cell Header Format
- ATM Cell Header Fields
- ATM Reference Model

Objectives

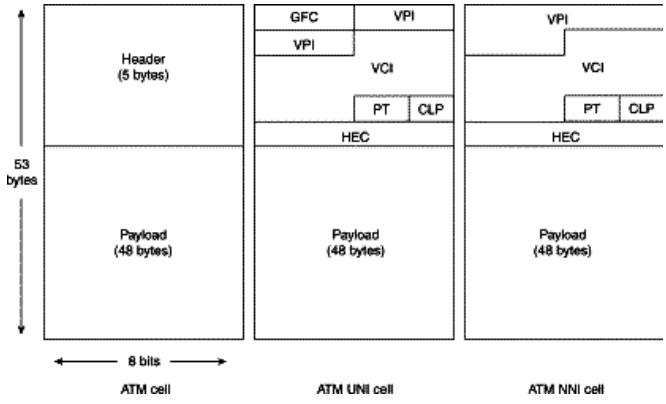
Upon completion of this Lesson, you should be able to:Upon completion of this Lesson, you should be able to:

- Explain the different fields in ATM cell header
- Explain the ATM Reference model

25.1 ATM Cell Header Format

An ATM cell header can be one of two formats: UNI or NNI. The UNI header is used for communication between ATM endpoints and ATM switches in private ATM networks. The NNI header is used for communication between ATM switches. Figure 25.1 depicts the basic ATM cell format, the ATM UNI cell header format, and the ATM NNI cell header format.

Figure 25.1: An ATM Cell, ATM UNI Cell, and ATM NNI Cell Header Each Contain 48 Bytes of Payload

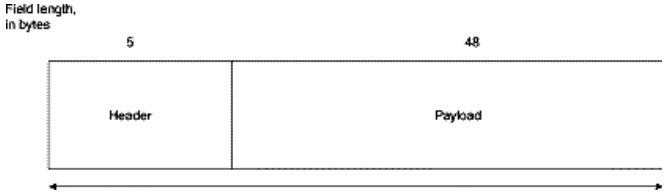


Unlike the UNI, the NNI header does not include the Generic Flow Control (GFC) field. Additionally, the NNI header has a Virtual Path Identifier (VPI) field that occupies the first 12 bits, allowing for larger trunks between public ATM switches.

25.2 ATM Cell Header Fields

In addition to GFC and VPI header fields, several others are used in ATM cell header fields. The following descriptions summarize the ATM cell header fields illustrated in Figure 25.1:

- **·Generic Flow Control (GFC)**—Provides local functions, such as identifying multiple stations that share a single ATM interface. This field is typically not used and is set to its default value of 0 (binary 0000).
- **·Virtual Path Identifier (VPI)**—In conjunction with the VCI, identifies the next destination of a cell as it passes through a series of ATM switches on the way to its destination.
- **Virtual Channel Identifier (VCI)**—In conjunction with the VPI, identifies the next destination of a cell as it passes through a series of ATM switches on the way to its destination.
- **Payload Type (PT)**—Indicates in the first bit whether the cell contains user data or control data. If the cell contains user data, the bit is set to 0. If it contains control data, it is set to 1. The second bit indicates congestion (0 = no congestion, 1 = congestion), and the third bit indicates whether the cell is the last in a series of cells that represent a single AAL5 frame (1 = last cell for the frame).
- **Cell Loss Priority (CLP)**—Indicates whether the cell should be discarded if it encounters extreme congestion as it moves through the network. If the CLP bit equals 1, the cell should be discarded in preference to cells with the CLP bit equal to 0.
- **Header Error Control (HEC)**—Calculates checksum only on the first 4 bytes of the header. HEC can correct a single bit error in these bytes, thereby preserving the cell rather than discarding it.



25.3 ATM Reference Model

The ATM architecture uses a logical model to describe the functionality that it supports. ATM functionality corresponds to the physical layer and part of the data link layer of the OSI reference model.

The ATM reference model is composed of the following planes, which span all layers:

- **Control**—This plane is responsible for generating and managing signaling requests.
- **User**—This plane is responsible for managing the transfer of data.
- **Management**—This plane contains two components:
  - Layer management manages layer-specific functions, such as the detection of failures and protocol problems.
  - Plane management manages and coordinates functions related to the complete system.

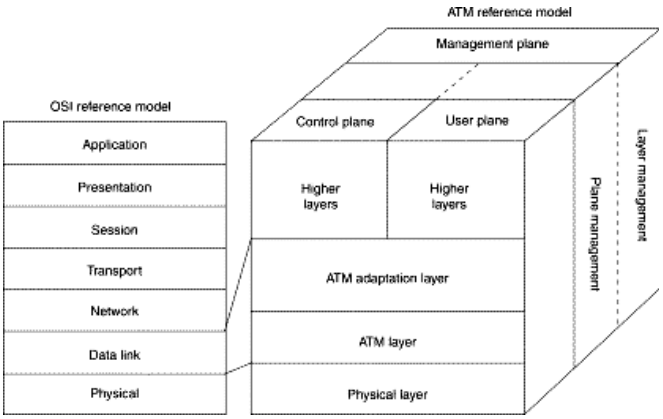
The ATM reference model is composed of the following ATM layers:

- **Physical layer**—Analogous to the physical layer of the OSI reference model, the ATM physical layer manages the medium-dependent transmission.

- **ATM layer**—Combined with the ATM adaptation layer, the ATM layer is roughly analogous to the data link layer of the OSI reference model. The ATM layer is responsible for the simultaneous sharing of virtual circuits over a physical link (cell multiplexing) and passing cells through the ATM network (cell relay).
- **ATM adaptation layer (AAL)**—Combined with the ATM layer, the AAL is roughly analogous to the data link layer of the OSI model. The AAL is responsible for isolating higher-layer protocols from the details of the ATM processes. The adaptation layer prepares user data for conversion into cells and segments the data into 48-byte cell payloads.

Finally, the higher layers residing above the AAL accept user data, arrange it into packets, and hand it to the AAL. Figure 25.2 illustrates the ATM reference model.

Figure 25.2 : The ATM Reference Model Relates to the Lowest Two Layers of the OSI Reference Model



Notes

LESSON 26

Today's Topics

- Broadcast Networks
- Point to point Networks
- LAN
- Ethernet
- Token Ring
- ARCNET
- FDDI
- WAN

Objectives

Upon completion of this Lesson, you should be able to::

- Explain the difference between point to point and broadcast networks
- Explain and Compare the main LAN toplogies
- Describe a WAN

Broadly speaking there are two types of transmission technology.

- Broadcast Networks
- Point to point networks

26.1 Broadcast Networks

have a single communication channel that is share d by all the machines on the network.Short messages called packets in certain contexts sent by any machine are received Broadcast networks generally allow the possibility of addressing by using a special code in the address field.a packet to all destinations by all the otheres.An address field within the packet specifies for whom it is intended.Upon receiving a packet, a machine checks for itself the address field. If it is meant for itself it processes the packet , otherwise it is ignored.

26.2 Point to Point

consisits of manyu connections between individual apir of machines.To go from the source to the destination a packet on this **type of** network may have to first visit one or more intermediate machines.

An alternative solution for classifying networks is their scale. In this classification comes the true networks , which can be further divided as LANs , WAN s and MANs.

26.3 LAN

A local area network (LAN) is a group of computers and associated devices that share a common communications line or wireless link and typically share the resources of a single processor or server within a small geographic area (for example, within an office building). Usually, the server has applications and data storage that are shared in common by multiple computer users. A local area network may serve as few as two or three users (for example, in a home network) or as many as thousands of users (for example, in an FDDI network).

The main local area network technologies are:

- Ethernet
- Token Ring
- Arcnet
- FDDI(Fiber Distributed Data Interface)

Typically, a suite of application programs can be kept on the LAN server. Users who need an application frequently can download it once and then run it from their local hard disk. Users can order printing and other services as needed through applications run on the LAN server. A user can share files with others at the LAN server; read and write access is maintained by a LAN administrator. A LAN server may also be used as a Web server if safeguards are taken to secure internal applications and data from outside access.

26.3.1 Ethernet

is the most widely-installed local area network (LAN) technology. Specified in a standard, IEEE802.3 , Ethernet was originally developed by Xerox and then developed further by Xerox, DEC, and Intel. An Ethernet LAN typically uses coaxial cable or special grades of twisted pair wires. Ethernet is also used in wireless LANs. The most commonly installed Ethernet systems are called 10BASE-Tand provide transmission speeds up to 10mbps.. Devices are connected to the cable and compete for access using a Carrier Sense Multiple Access with Collision Detection(CSMA/CD) protocol.

26.3.2 A Token Ring

network is a local area network (LAN) in which all computers are connected in a ring or star topology and a bit- or token-passing scheme is used in order to prevent the collision of data between two computers that want to send messages at the same time. The Token Ring protocol is the second most widely-used protocol on local area networks after Ethernet.The IBM Token Ring protocol led to a standard version, specified as IEEE 802.5. Both protocols are used and are very similar. The IEEE 802.5 Token Ring technology provides for data transfer rates of either 4 or 16 megabits per second. Very briefly, here is how it works:

1. Empty information frames are continuously circulated on the ring.
2. When a computer has a message to send, it inserts a token in an empty frame (this may consist of simply changing a 0 to a 1 in the token bit part of the frame) and inserts a message and a destination identifier in the frame.
3. The frame is then examined by each successive workstation. If the workstation sees that it is the destination for the message, it copies the message from the frame and changes the token back to 0.

4. When the frame gets back to the originator, it sees that the token has been changed to 0 and that the message has been copied and received. It removes the message from the frame.

The frame continues to circulate as an “empty” frame, ready to be taken by a workstation when it has a message to send.

26.3.3 Arcnet

is a widely-installed local area network (LAN) technology that uses a token-bus scheme for managing line sharing among the workstations and other devices connected on the LAN. The LAN server continuously circulates empty message frames on a bus(a line in which every message goes through every device on the line and a device uses only those with its address). When a device wants to send a message, it inserts a “token” (this can be as simple as setting a token bit to 1) in an empty frame in which it also inserts the message. When the destination device or LAN server reads the message, it resets the token to 0 so that the frame can be reused by any other device. The scheme is very efficient when traffic increases since all devices are afforded the same opportunity to use the shared network.

ARCNET can use coaxial or fibreoptic lines.

26.3.4 FDDI

(Fiber Distributed Data Interface) is a set of ANSI and ISO standards for data transmission on fibre optic lines in a local area network(LAN) that can extend in range up to 200 km (124 miles). The FDDI protocol is based on the token ring protocol. In addition to being large geographically, an FDDI local area network can support thousands of users. FDDI is frequently used on thebackbone for a wide area network(WAN).

An FDDI network contains two token rings, one for possible backup in case the primary ring fails. The primary ring offers up to 100 mbpscapacity. If the secondary ring is not needed for backup, it can also carry data, extending capacity to 200 Mbps. The single ring can extend the maximum distance; a dual ring can extend 100 km .

26.4 WAN

A wide area network spans a large geographical area often a country or continent. It contains a collection of machines intended for a country or continent. It contains a collection of machines intended for running user (application) programs called hosts. The hosts are connected by a communication subnet . The job of the subnet is to carry messages from host to host , just as the telephone system carries words from speaker to listener. By separating the pure communication aspects of the network (subnet) from the application aspects (hosts) the complete network design is greatly simplified.

In most WANs the subnet consists of two distinct components, transmission lines and switching elements. Transmission lines (also called circuits, channels, or trunks) move bits between machines. The switching elements are specialized computers used to connect two or more transmission lines. When data arrive on an incoming line the switching element must choose an outgoing line to forward them on.

In most WANs the network contains numerous cables or telephone lines, each one connecting a pair of routers. If two routers that do not share a cable wish to communicate they must do this indirectly via other routers. When a packet is sent from one router to another via one or more intermediate routers the packet is received at each intermediate router in its entirety, stored there until the required output line is free and then forwarded. A subnet using this principle is called a point to point network In contrast to local area networks, which have a symmetric topology, WANS have irregular topologies.

A second possibility for a WAN is a satellite or ground radio system. Each router has an antenna through which it can send and receive. All routers can hear the output from the satellite and in some cases they can also hear the upward transmissions of their fellow routers to the satellite as well. Sometimes the routers are connected to a substantial point-to-point subnet, with only some of them having a satellite antenna. Satellite networks are inherently broadcast and are more useful when the broadcast property is important.

For more information on topics in chapters 15 –26 please refer to

References and Further Reading

- Data Communications, Prakash C Gupta, Eastern Economy Edition
- Handbook of data communication, National Computing Center Ltd, Manchester
- Comput er Networks- Tannenbaum

Notes

EIILM UNIVERSITY
SIKKIM

Jorethang, District Namchi, Sikkim- 737121, India
www.eiilmuniversity.ac.in