SYLLABUS


3. **Data Link Control**:- Introduction, Error Detection, Single Parity Check, Cyclic Redundancy Check (CRC), Error Control mechanisms, Stop-and-Wait ARQ, Go-Back-N ARQ, Selective Repeat ARQ, Flow Control

   High-Level Data Link Control (HDLC), Basic Characteristics of HDLC, HDLC Frame Structure, HDLC Operation

4. **Local Area Networks and MAC Protocols**:- Introduction, LAN Architecture, LAN Topologies, Bus and Tree Topologies, Ring Topology, Star Topology, Medium Access Control, Round Robin Reservation, Contention, MAC Frame Format, Logical Link Control, LAN Systems, Ethernet and Fast Ethernet (CSMA/CD), Token Ring, IEEE 802.5 Medium Access Control, FDDI, Interconnection of LANs - Bridges

**Reference:**
- Data communication and Networking – Behrouz A. Forouzan
- Data and Computer Communication- William Stalling
- Computer Network – Tannebaum
Data Communication Model, Data Communications System Tasks, Communication Network and Services

Data Communications Networking, Wide-Area Networks, Local Area Networks, Protocol and protocol Architectures, The OSI model, The TCP/IP model


Transmission Media, Guided Transmission Media, Wireless (Radio) Transmission, Multiplexing, Frequency-Division Multiplexing (FDM), Time Division Multiplexing TDM, Wavelength Division Multiplexing, Space-Division Switches, Time-Division Switches

Introduction, Error Detection, Single Parity Check, Cyclic Redundancy Check (CRC), Error Control mechanisms, Stop-and-Wait ARQ, Go-Back-N ARQ, Selective Repeat ARQ, Flow Control
High-Level Data Link Control (HDLC), Basic Characteristics of HDLC, HDLC Frame Structure, HDLC Operation

Introduction, LAN Architecture, LAN Topologies, Bus and Tree Topologies, Ring Topology, Star Topology, Medium Access Control, Round Robin Reservation, Contention, MAC Frame Format, Logical Link Control, LAN Systems, Ethernet and Fast Ethernet (CSMA/CD), Token Ring, IEEE 802.5 Medium Access Control, FDDI, Interconnection of LANs - Bridges
# TABLE OF CONTENTS

## UNIT1

**Fundamentals of Data Communication**

1.0 Introduction  
1.1 Data Communication Model  
1.2 Data Communications System Tasks  
1.3 Communication Network and Services  
1.4 Data Communications Networking  
  1.4.1 Wide-Area Networks  
  1.4.2 Local Area Networks  
1.5 Protocols and Protocol Architectures  
  1.5.1 The OSI model  
  1.5.2 The TCP/IP model

## UNIT2

**Data Transmission**

2.1 Concepts and Terminology  
  2.1.1 Time-Domain Concepts  
  2.1.2 Frequency Domain Concepts  
2.2 Why digital communication?  
2.3 Bandwidth, Data rate and Channel Capacity  
  2.3.1 Nyquist sampling rate  
  2.3.2 Shannon Channel Capacity  
2.4 Line Coding  
2.5 Modems and Digital Modulation  
2.6 Transmission Media  
  2.6.1 Guided Transmission Media  
  2.6.2 Wireless (Radio) Transmission  
2.7 Multiplexing  
  2.7.1 Frequency-Division Multiplexing (FDM)  
  2.7.2 Time Division Multiplexing TDM  
  2.7.3 Wavelength Division Multiplexing  
2.8 Circuit Switching  
  2.8.1 Space-Division Switches  
  2.8.2 Time-Division Switches
UNIT 3
Data Link Control
3.0 Introduction
3.1 Error Detection
   3.1.1 Single Parity Check
   3.1.2 Cyclic Redundancy Check (CRC)
3.2 Error Control mechanisms
   3.2.1 Stop-and-Wait ARQ
   3.2.2 Go-Back-N ARQ
   3.2.3 Selective Repeat ARQ
3.3 Flow Control
3.4 High-Level Data Link Control (HDLC)
   3.4.1 Basic Characteristics of HDLC
   3.4.2 HDLC Frame Structure
   3.4.3 HDLC Operation

UNIT 4
Local Area Networks and MAC Protocols
4.0 Introduction
4.1 LAN Architecture
4.2 LAN Topologies
   4.2.1 Bus and Tree Topologies
   4.2.2 Ring Topology
   4.2.3 Star Topology
4.3 Medium Access Control
   4.3.1 Round Robin
   4.3.2 Reservation
   4.3.3 Contention
4.4 MAC Frame Format
4.5 Logical Link Control
4.6 LAN Systems
   4.6.1 Ethernet and Fast Ethernet (CSMA/CD)
   4.6.2 Token Ring
      4.6.2.1 IEEE 802.5 Medium Access Control
      4.6.2.2 FDDI
4.7 Interconnection of LANs - Bridges
1.0 INTRODUCTION
1.1 DATA COMMUNICATION MODEL
1.2 DATA COMMUNICATIONS SYSTEM TASKS
1.3 COMMUNICATION NETWORK AND SERVICES
1.4 DATA COMMUNICATIONS NETWORKING
   1.4.1 WIDE-AREA NETWORKS
   1.4.2 LOCAL AREA NETWORKS
1.5 PROTOCOLS AND PROTOCOL ARCHITECTURE
   1.5.1 THE OSI MODEL
   1.5.2 THE TCP/IP MODEL

This chapter provides the big picture of the data communication networks. The tasks involved in the communication networks, different communication services required, the kind of networks available, protocol architectures and OSI, TCP/IP protocol models are discussed in brief. In essence fundamental concepts of the data communication networks are discussed briefly, providing the basis for detailed discussion in later chapters. At the end of the chapter the student will have a context in which to place various topics as they progress through the later chapters.

1.0 Introduction

Basic Communication Model

Communication is 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. To give 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.
The only way that a message source can be certain that the destination properly received the message is by some kind of acknowledgment response from the destination. Conversing people might say "I understand" or nod their head in response to a statement made by their peer. This acknowledged form of dialogue is the basis of reliable communications - 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 suite the data communication between two systems exchanging information between them.

**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.

**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.

**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.

**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.

**Destination:** Device to which the source device sends data.
1.1 Data Communication Model

Figure 1.3 provides a new perspective on the data communication model of Figure 1.2. Let us trace through the details of this figure using electronic mail as an example.

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, “Good luck for your exams” (m). The user activates the electronic mail package on the PC and enters the message via the keyboard (input device). 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
This transmitted signal $s(t)$ 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.

### 1.2 Data Communications System Tasks

Some of the Key tasks to be performed by a Data Communications System are listed below.

- Signal Generation
- Interfacing
- Synchronization
- Exchange Management
- Transmission System Utilisation
- Error Detection and Correction
- Flow Control
- Addressing
- Routing
- Message Formatting

**Signals:** All the data that are transmitted over the transmitting system propagate as Electromagnetic signals. Hence the communicating device must be able to generate and receive these signals. **Signal generation** should be such that the resultant signal is capable of being propagated through the transmission medium and interpretable as data at the receiver.

**Interface:** A device must **interface** with the transmission system in order to communicate.
SOLUTION:

**SYNCHRONIZATION** :- Unless the receiver and transmitter are in Synchronization the receiver will not be able to make sense out of received signals. Receiver should know when the transmission of data starts, when it ends.

**EXCHANGE MANAGEMENT** :- For meaningful data transaction there should be some kind management of data being exchanged. Both the transmitter and receiver should adhere to some common convention about the format of data, amount of data that can be sent at a time and so on. This requires a prior definition of message formatting.

**TRANSMISSION SYSTEM UTILISATION** :- It refers to the need to make efficient use of Transmission Channel, which are generally shared by many communicating devices. Various techniques (Multiplexing) are available to allocate the total capacity of a transmission channel among connected devices. Care should be taken to avoid probable Congestion in some kind of multiplexing.

**ERROR DETECTION AND CORRECTION** :- In any communication system transmitted data is prone to error. Either it is because of transmitted signal getting distorted in the transmission medium leading to misinterpretation of signal or errors introduced by the intermediate devices. Error detection and Correction is required in cases where there is no scope for error in the data transaction. We can think of file transfer between two computers where there is a need for this. But in some cases it may not be very important as in the case of telephonic conversation.

**FLOW CONTROL** :- There is a possibility of transmitter generating data faster than the receiver device capable of handling. To handle this there should be some kind of flow control mechanism agreed upon between the two communicating devices.

**ADDRESSING** :- When more than two devices share a transmitting facility, a source system must somehow indicate the identity (or address) of the destination.

**ROUTING** :- The transmission system must ensure that the data being sent are routed only to the destination system.

The list of tasks explained above is not exhaustive one. But sure it indicates the complexities involved in a communication system. The data communication system must be designed to handle the above tasks as required so that meaningful communication can take place between two devices efficiently.

1.3 Communication Network and Services

A communication network, in its simplest form, is a set of equipment and facilities that provides a communication service: the transfer of information between users located at various geographical points. Here we focus on networks that use electronic or optical technologies. Examples of such networks include telephone networks, computer networks, television broadcast networks, cellular telephone networks, and the Internet. The ability of communication network to transfer information at extremely high speeds allows users to gather information in large volumes, nearly instantaneously and, with the aid of computers, to almost immediately exercise action at a distance. These two unique capabilities form the basis for many existing services and an unlimited number of future network-based services. We will now discuss several services that are
supported by current networks. The services are examined from the point of view of user requirements, that is, quality of service, features, and capabilities. The viewpoint here is that networks should ultimately be designed to meet the requirements of the user applications.

**Radio and television broadcasting** are probably the most common communication services. Various stations transmit an ensemble of signals simultaneously over radio or cable distribution networks. Aside from selecting the station of interest, the role of the user in these services is passive. Relatively high audio and video quality is expected, but a significant amount of delay (in the order of seconds or more) can be tolerated even in live broadcasts.

**Telephone service** is the most common real-time service provided by a network. Two persons are able to communicate by transmitting their voices across the network. The service is *connection-oriented* in the sense that the users must first interact with the network to set up a connection. The telephone service has the real-time requirement in that users cannot interact as in face-to-face conversation if the delays are greater than a fraction of second (approximately 250 milliseconds). The service must also be reliable in the sense that once the connection is established it must not be interrupted because of failures in the network. At the minimum the delivered voice signal must be intelligible, but in most situations the users expect a much higher quality that enables the listener not only to recognize what the speaker says but also to discern subtleties in intonation, mood, and so on. A high degree of availability is another requirement: Telephone users expect the network to be capable of completing the desired connection almost all the time. Security and privacy of the conversation are consideration in some situations.

The telephone service can be enhanced in a number of ways. For example, the toll-free service is provided wherein the caller will not be billed but costs of the call are automatically billed to the subscriber of the service. Similarly, in credit-card or calling-card services, the cost of a call is automatically billed to the holder of the card. Clearly, security and fraud are issues here.

Telephone networks provide a broad class of call management services that use the originating number or the destination number to determine the handling of a call. **Caller ID** allows the originating number, and sometimes name, of the originating call to be displayed to the destination user when the receiving device is display capable. Voice mail allows a destination user to have calls forwarded to a message-receiving device when the destination user is not available.

**Cellular telephone service** extends the normal telephone service to mobile users who are free to move within a regional area covered by an interconnected array of smaller geographical areas called cells. Each cell has a radio transmission system that allows it to communicate with users in its area. The use of radio transmission implies design compromises that may result in lower voice quality, lower availability, and greater exposure to eavesdropping. In addition, the cellular system must handle the *handing off* of users as they move from one cell to another so that an ongoing conversation is not terminated abruptly. Some cellular providers also support a roaming service where a subscriber is able to place calls while visiting regional areas other than the subscriber's
home base. Note that the mobility aspect to the roaming service is not limited to cellular (or wireless) communications. Indeed, the need for mobility arises whenever a subscriber wishes to access a service from anywhere in the world.

**Electronic mail** (e-mail) is another common network service. The user typically provides a text message and a name and/or address to a mail application. The application interacts with a local mail server, which in turn transmits the message to a destination server across a computer network. The destination user retrieves the message by using a mail application, such as *Outlook Express* software package of Microsoft. E-mail is not a real-time service in that fairly large delays can be tolerated. It is also not necessarily connection-oriented in that a network connection does not need to be set up expressly for each individual message. The service requires reliability in terms of the likelihood of delivering the message without errors and to the correct destination. In some instances the user may be able to request delivery confirmation. Again security and privacy may be a concern.

Many applications that involve an interaction between processes running in two computers may be characterized by *client/server* interaction. For example, a user likes to view a document file of a particular company and he initiates a process - *client* - to access a given file on the *server* which hosts that particular document. The **World Wide Web** (**WWW**) application typifies this interaction. The **WWW** consists of a framework for accessing documents that are located in computers connected to the **Internet**. These documents consist of text, graphics, and other media and are interconnected by links that appear within the documents. The WWW is accessed through a browser program that displays the documents and allows the user to access other documents by clicking on one of these links. Each link provides the browser with a **Uniform resource locator** (**URL**) that specifies the name of the system where the document is located as well as the name of the file that contains the requested document. For example the URL

http://www.abce.org/prog.html

specifies the document in the **Internet**

The first term, **http**, specifies the retrieval mechanism to be used, in this case, the **Hyper Text Transfer Protocol** (**HTTP**). Next the URL specifies the name of the host machine, namely, **www.abce.org**. The remaining data gives the path to the document within the host machine, that is, the URL identifies the document file on the server hosting the desired document. By entering the above URL in the address field of a browser such as **Internet Explorer** or **Netscape Navigator**, user can view the document.

In addition to text the files the WWW may contain audio and images that involve large amounts of information. While the user does not require real-time response, excessive delay in retrieving files reduces the degree of interactivity of the overall application where the user seeks information, reads it, and again seeks additional information by clicking on other items or on links to other Web sites. The overall delay is determined by the delays in accessing the servers as well as the time required to transmit the files through the network.

There are many other network services like **Video on demand**, **Streamed Audiovisual Services**, **Audio Conferencing**, **Audio-Visual Conferencing**. The basic information in these services is of audio and visual in nature and they require huge data to convey the
information. Apart from having huge data to be transported between communicating systems, since these services are of real-time in nature the bandwidth of the underlying communication network required for these services is also huge. Looking at the technological development happening in the data communication field, days are not for off when these services are provided to common user over the Internet at affordable cost.

1.4 Data Communications Networking

In its simplest form, data communication takes place between two devices that are directly connected by some form of point-to-point transmission medium. Often, however, it is impractical for two devices to be directly, point-to-point connected. This is so for one (or both) of the following contingencies:

- The devices are very far apart. It would be inordinately expensive, for example, to string a dedicated link between two devices thousands of miles apart.
- There is a set of devices, each of which may require a link to many of the others at various times. Examples are all of the telephones in the world and all of the terminals and computers owned by a single organization. Except for the case of a very few devices, it is impractical to provide a dedicated wire between each pair of devices.

The solution to this problem is to attach each device to a communications network. Figure 1.4 relates this area to the communications model of Figure 1.2a and also suggests the two major categories into which communications networks are traditionally classified: Wide-Area Networks (WANs) and Local-Area Networks (LANs). The distinction between the two, both in terms of technology and application, has become somewhat blurred in recent years, but it remains a useful way of organizing the discussion.
1.4.1 Wide-Area Networks

Wide-area networks have traditionally been considered to be those that cover a large geographical area, require the crossing of public right-of-ways, and rely at least in part on circuits provided by a common telephone carrier. Typically, a WAN consists of a number of interconnected switching nodes. Transmission from any one device is routed through these internal nodes to the specified destination device. These nodes (including the boundary nodes to which the devices are connected) are not concerned with the content of the data; rather, their purpose is to provide a switching facility that will move the data from node to node until they reach their destination. Traditionally, WANs have been implemented using one of two technologies: circuit switching and packet switching. More recently, frame relay and ATM networks have assumed major roles.

**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.

**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.

**Frame Relay**

Packet switching was developed at a time when digital long-distance transmission facilities exhibited a relatively high error rate compared to today's facilities. As a result, there is a considerable amount of overhead built into packet-switched schemes to compensate for errors. The overhead includes additional bits added to each packet to introduce redundancy and additional processing at the end stations and the intermediate switching nodes to detect and recover from errors.

With modern high-speed telecommunication systems, this overhead is unnecessary and counterproductive. It is unnecessary because the rate of errors has been dramatically lowered and any remaining errors can easily be caught in the end systems by logic that operates above the level of the packet-switching logic; it is counterproductive because the overhead involved soaks up a significant fraction of the high capacity provided by the network.

Frame relay was developed to take advantage of these high data rates and low error rates. Whereas the original packet-switching networks were designed with a data rate to the end user of about 64 kbps (kilo bits per second, 1 kilo=10³), frame relay networks are designed to operate efficiently at user data rates of up to 2 Mbps (Mega bits per second, 1 Mega=10⁶). The key to achieving these high data rates is to strip out most of the overhead involved with error control.

**ATM**

Asynchronous Transfer Mode (ATM), sometimes referred to as cell relay, is a culmination of all of the developments in circuit switching and packet switching over the past 25 years.

ATM can be viewed as an evolution from frame relay. The most obvious difference between frame relay and ATM is that frame relay uses variable-length packets, called frames, and ATM uses fixed-length packets, called cells. As with frame relay, ATM provides little overhead for error control, depending on the inherent reliability of the transmission system and on higher layers of logic in the end systems to catch and correct errors. By using a fixed-packet length, the processing overhead is reduced even
further for ATM compared to frame relay. The result is that ATM is designed to work in the range of 10s and 100s of Mbps, compared to the 2-Mbps target of frame relay. ATM can also be viewed as an evolution from circuit switching. With circuit switching, only fixed-data-rate circuits are available to the end system. ATM allows the definition of multiple virtual channels with data rates that are dynamically defined at the time the virtual channel is created. By using full, fixed-size cells, ATM is so efficient that it can offer a constant-data-rate channel even though it is using a packet-switching technique. Thus, ATM extends circuit switching to allow multiple channels with the data rate on each channel dynamically set on demand.

**ISDN and Broadband ISDN**

Merging and evolving communications and computing technologies, coupled with increasing demands for efficient and timely collection, processing, and dissemination of information are leading to the development of integrated systems that transmit and process all types of data – plain text data, audio, video. A significant outgrowth of these trends is the Integrated Services Digital Network (ISDN). Through this user with a single access point to ISDN network can avail of different kinds of communication – his computer can access the Internet, he can use the network for his telephone usage and also probably video communication.

The ISDN is intended to be a worldwide public telecommunications network to replace existing public telecommunications networks and deliver a wide variety of services. The ISDN is defined by the standardization of user interfaces and implemented as a set of digital switches and paths supporting a broad range of traffic types and providing value-added processing services. In practice, there are multiple networks, implemented within national boundaries, but, from the user's point of view, there is intended to be a single, uniformly accessible, worldwide network.

Despite the fact that ISDN has yet to achieve the universal deployment hoped for, it is already in its second generation. The first generation, sometimes referred to as narrowband ISDN, is based on the use of a 64-kbps channel as the basic unit of switching and has a circuit-switching orientation. The major technical contribution of the narrowband ISDN effort has been frame relay. The second generation, referred to as broadband ISDN, supports very high data rates (100s of Mbps) and has a packet-switching orientation. The major technical contribution of the broadband ISDN effort has been asynchronous transfer mode (ATM), also known as cell relay.

**1.4.2 Local Area Networks (LAN)**

As with Wide-Area Networks, a Local-Area Network or LAN is a communications network that interconnects a variety of devices and provides a means for information exchange among those devices. There are several key distinctions between LANs and WANs;

- The scope of the LAN is small, typically a single building or a cluster of buildings. This difference in geographic scope leads to different technical solutions.
• It is usually the case that the LAN is owned by the same organization that owns the attached devices. For WANs, this is less often the case, or at least a significant fraction of the network assets are not owned. This has two implications. First, care must be taken in the choice of LAN, as there may be a substantial capital investment (compared to dial-up or leased charges for wide area networks) for both purchase and maintenance. Second, the network management responsibility for a local network falls solely on the user.

• The internal data transfer rates of LANs are typically much greater than those of wide-area networks.

Traditionally, LANs make use of a broadcast network approach rather than a switching approach. With a broadcast communication network, there are no intermediate switching nodes. At each station, there is a transmitter / receiver that communicates over a medium shared by other stations. A transmission from any one station is broadcast to and received by all other stations. We will be concerned with networks used to link computers, workstations, and other digital devices. Also in the case of LANs, data are usually transmitted in packets. Because the medium is shared, only one station at a time can transmit a packet.

More recently, examples of switched LANs have appeared, the two most prominent examples are ATM LANs, which simply use an ATM network in a local area, and Fiber Channel. Various topologies are possible for broadcast LANs. Figure 1.5 shows two of them.

![Figure 1.5 Two broadcast networks. (a) Bus, (b) Ring](image)

**Ethernet**

In a bus (i.e. a linear cable) network, at any instant one machine is the master and is allowed to transmit. An arbitration mechanism is needed to resolve conflicts when two or more machines want to transmit simultaneously. The arbitration mechanism may be centralized or distributed. IEEE 802.3 CSMA/CD popularly known as Ethernet, for example, is a bus based broadcast network with decentralized control operating at 10 or 100 Mbps. Computers on an Ethernet can transmit whenever they want to; if two or more packets collide, each computer just waits a random time and tries again later.

**Token Ring**
A second type of broadcast system is the token ring network. A ring really consists of a collection of ring interfaces connected by point-to-point lines. Each bit arriving at an interface is copied back onto the ring at the other interface. In a token ring, a special pattern of bits called *token*, circulates around the ring whenever all stations are idle. When a station wants to transmit a frame, it is required to seize the token and remove it from the ring before transmitting. Because there is only one token, only one station can transmit at a given instant, thus solving the channel access problem.

### 1.5 Protocols and Protocol Architecture

When computers, terminals, and/or other data processing devices exchange data, the scope of concern is much broader than the concerns we have discussed so far. Consider, for example, the transfer of a file between two computers. There must be a data path between the two computers, either directly or via a communication network. But more is needed. Typical tasks to be performed are:

1. The source system must either activate the direct data communication path or inform the communication network of the identity of the desired destination system.
2. The source system must ascertain that the destination system is prepared to receive data.
3. The file transfer application on the source system must ascertain that the file management program on the destination system is prepared to accept and store the file for this particular user.
4. If the file formats used on the two systems are incompatible, one or the other system must perform a format translation function.

It is clear that there must be a high degree of *cooperation between the two computer systems*. The exchange of information between computers for the purpose of cooperative action is generally referred to as *computer communications*. Similarly, when two or more computers are interconnected via a communication network, the set of computer stations is referred to as a *computer network*.

In discussing computer communications and computer networks, two concepts are paramount:

- Protocols
- Protocol architecture or Computer Communications architecture

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.
What is a protocol, really? It is software that resides either in a computer's memory or in the memory of a transmission device, like a network interface card. When data is ready for transmission, this software is executed. The software prepares data for transmission and sets the transmission in motion. At the receiving end, the software takes the data off the wire and prepares it for the computer by taking off all the information added by the transmitting end.

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

It is clear that there must be a high degree of cooperation between the two computers. Instead of implementing the logic for this as a single module, the task is broken up into subtasks, each of which is implemented separately. As an example, Figure 1.6 suggests the way in which a file transfer facility could be implemented. Three modules are used. Tasks 3 and 4 in the preceding list could be performed by a file transfer module. The two modules on the two systems exchange files and commands. However, rather than requiring the file transfer module to handle the details of actually transferring data and commands, the file transfer modules each rely on a communications service module. This module is responsible for making sure that the file transfer commands and data are reliably exchanged between systems. Among other things, this module would perform task 2. Now, the nature of the exchange between systems should be independent of the nature of the network that interconnects them. Therefore, rather than building details of the network interface into the communications service module, it makes sense to have a third module, a network access module that performs task 1 by interacting with the network.

The file transfer module contains all of the logic that is unique to the file transfer application, such as transmitting passwords, file commands, and file records. There is a need to transmit these files and commands reliably. However, the same sorts of reliability requirements are relevant to a variety of applications (e.g., electronic mail, document transfer). Therefore, these requirements are met by a separate communications service module that can be used by a variety of applications. The communications service module is concerned with assuring that the two computer systems are active and ready for data transfer and for keeping track of the data that are being exchanged to assure delivery. However, these tasks are independent of the type of network that is being used. Therefore, the logic for actually dealing with the network is
separated out into a separate network access module. That way, if the network to be used is changed, only the network access module is affected. Thus, instead of a single module for performing communications, there is a structured set of modules that implements the communications function. That structure is referred to as protocol architecture. Each module in a layer is defined by a protocol and a set of protocols which work together are termed as protocol stack. The terms protocol architecture and protocol stack are used interchangeably.

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. Hence a brief introduction to both of them is given below.

1.5.1 The OSI 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.

<table>
<thead>
<tr>
<th>Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
</tr>
<tr>
<td>Presentation</td>
</tr>
<tr>
<td>Session</td>
</tr>
<tr>
<td>Transport</td>
</tr>
<tr>
<td>Network</td>
</tr>
<tr>
<td>Data Link</td>
</tr>
<tr>
<td>Physical</td>
</tr>
</tbody>
</table>

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.

Unfortunately, protocols in the real world do not conform precisely to these neat definitions. Some network products and architectures combine layers. Others leave
layers out. Still others break the layers apart. But no matter how they do it, all working network products achieve the same result - getting data from here to there.

### 1.5.2 The TCP/IP Protocol Architecture

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.

![TCP/IP Protocol Architecture Model](image)

Figure 1.7 TCP/IP Protocol Architecture Model

Figure 1.7 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.

**Exercise**

1. Discuss the different entities involved in a simple communication network model.
2. List the various tasks involved in data communication system and briefly explain them.
3. List the categories into which communication networks are classified. List different kind of networks under each category and briefly discuss one network each under a category.
4. What is protocol? What are the elements of a protocol?
5. Explain the concept of protocol architecture using File transfer application as an example.
6. Briefly discuss the OSI and TCP/IP protocol models.
UNIT 2
DATA TRANSMISSION

2.1 Concepts and Terminology

2.1.1 Time-Domain Concepts
2.1.2 Frequency Domain Concepts

2.2 Why digital communication?

2.3 Bandwidth, Data rate and Channel Capacity

2.3.1 Nyquist sampling rate
2.3.2 Shannon Channel Capacity

2.4 Line Coding

2.5 Modems and Digital Modulation

2.6 Transmission Media

2.6.1 Guided Transmission Media
2.6.2 Wireless (Radio) Transmission

2.7 Multiplexing

2.7.1 Frequency-Division Multiplexing (FDM)
2.7.2 Time Division Multiplexing TDM
2.7.3 Wavelength Division Multiplexing

2.8 Circuit Switching

2.8.1 Space-Division Switches
2.8.2 Time-Division Switches

Successful transmission of data depends principally on two factors: the quality of the signal being transmitted and the characteristics of the transmission medium. The objective of this chapter is to provide the student with an intuitive feeling for the nature of these two factors.

2.1 Concepts and Terminology

In this section we introduce some concepts and terms that will be referred to throughout the rest of the chapter.

Data transmission occurs between transmitter and receiver over some transmission medium. Transmission media may be classified as guided or unguided. In both cases, communication is in the form of electromagnetic waves. With guided media, the waves are guided along a physical path; examples of guided media are twisted pair, coaxial cable, and optical fiber. Unguided media provide a means for transmitting electromagnetic waves but do not guide them: examples are propagation through air, vacuum and seawater.

The term direct link is used to refer to the transmission path between two devices in which signals propagate directly from transmitter to receiver with no intermediate
devices, other than amplifiers or repeaters used to increase signal strength. A guided
transmission medium is point-to-point if, first, it provides a direct link between two
devices and, second, those are the only two devices sharing the medium (Figure 2.1a).

In a multipoint configuration, more than two devices share the same medium (Figure
2.1b).

A transmission may be simplex, half-duplex, or full-duplex. In simplex transmission,
signals are transmitted in only one direction; one station is the transmitter and the
other is the receiver. In half-duplex operation, both stations may transmit, but only one
at a time. In full-duplex operation, both stations may transmit simultaneously. In the
latter case, the medium is carrying signals in both directions at the same time.

**Frequency, Spectrum and Bandwidth**

Here, we are concerned with electromagnetic signals, used as a means to transmit data.
The electromagnetic signal is generated by the transmitter and transmitted over a
medium. The signal is a function of time, but it can also be expressed as a function of
frequency; that is, the signal consists of components of different frequencies. It turns
out that the frequency-domain view of a signal is far more important to the
understanding of data transmission than the time-domain view. Both views are
introduced here.
2.1.1 Time-Domain Concepts

Viewed as a function of time, an electromagnetic signal can be either continuous or discrete. A continuous signal is one in which the signal intensity or signal strength varies in a smooth fashion over time. In other words there are no breaks or discontinuities in the signal. A discrete signal is one in which the signal intensity maintains a constant level for some period of time and then changes to another constant level. Figure 2.2 shows both kinds of signals. The continuous signal might represent speech, and the discrete signal might represent binary 1s and 0s.

![Continuous Signal](image1)

**Figure 2.2a Continuous Signal**

![Discrete Signal](image2)

**Figure 2.2b Discrete Signal**

Periodic signal

The simplest sort of signal is a periodic signal, in which the same signal pattern repeats over time. Figure 2.3 shows an example of a periodic analog signal (sine wave) and a periodic digital signal (square wave). Mathematically, a signal $s(t)$ is defined to be periodic if and only if $s(t + T) = s(t)$ for $-00 < t < +00$

where the constant $T$ is the period of the signal. ($T$ is the smallest value that satisfies the equation.) Otherwise, the signal is aperiodic.
The sine wave is the fundamental continuous signal. A general sine wave can be represented by three parameters: amplitude ($A$), frequency ($f$), and phase ($\phi$). 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

$$s(t) = A \sin (2 \pi f t + \Phi)$$
Paper Name: Data Communication

(a) $A = 1, f = 1, \phi = 0$

(b) $A = 0.5, f = 1, \phi = 0$

(c) $A = 1, f = 2, \phi = 0$
Figure 2.4 shows the effect of varying each of the three parameters. In part (a) of the figure, the frequency is 1 Hz; thus, the period is $T = 1$ second. Part (b) has the same frequency and phase but an amplitude of $1/2$. In part (c), we have $f = 2$, which is equivalent to $T = 1/2$ seconds. Finally, part (d) shows the effect of a phase shift of $\pi / 4$ radians, which is 45 degrees ($2 \pi$ radians = $360^\circ = 1$ period). In Figure 2.4, the horizontal axis is time; the graphs display the value of the signal strength at a given point in space as a function of time.

### 2.1.2 Frequency Domain Concepts

In practice, an electromagnetic signal will be made up of many frequencies. For example, the signal $s(t) = \sin(2 \pi f_1 t) + 1/3 \sin(2 \pi (3f_1)t)$ is shown in Figure 2.5. The components of this signal are just sine waves of frequencies $f_1$ and $3f_1$; parts (a) and (b) of the figure show these individual components. There are several interesting points that can be made about this figure: the second frequency is an integer multiple of the first frequency. When all of the frequency components of a signal are integer multiples of one frequency, the latter frequency is referred to as the fundamental frequency.
Figure 2.5 Addition of frequency components

The period of the total signal is equal to the period of the fundamental frequency. The period of the fundamental signal $\sin(2 \pi f_1 t)$ is $T = 1/f_1$, and the period of $s(t)$ is also $T$, as can be seen from Figure 2.5c. It can be shown, using a discipline known as Fourier Analysis that any signal is made up of components at various frequencies, in which case each component is sinusoid. This result is of tremendous importance, because the effects of various transmission media on a signal can be expressed in terms of frequencies and the signal can be analysed to great extent.
The **spectrum** of a signal is the range of frequencies that it contains. For the signal in Figure 2.5c, the spectrum extends from $f_1$ to $3f_1$. The **absolute bandwidth** of the signal is the width of spectrum; in the above case it would be $3f_1 - f_1 = 2f_1$ Hz. However most of the energy in an actual signal will be contained in a relatively narrow band of frequencies, which is known as **bandwidth**.

### 2.2 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 2.6a. 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 2.6b. 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.
(a) Analog transmission: all details must be reproduced accurately

\[ \text{Sent} \quad \text{Received} \]

- e.g. AM, FM, TV transmission

(b) Digital transmission: only discrete levels need to be reproduced

\[ \text{Sent} \quad \text{Received} \]

- e.g. digital telephone, CD Audio

*Figure 2.6 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 2.7. Such signal regeneration is fundamentally different for analog and digital transmissions.

*Figure 2.7 Typical long-distance link*

**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 2.8 shows the basic functions carried out by the analog repeater.
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 2.7. 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 2.9.
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.

### 2.3 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:
**Data rate.** This is the rate, in bits per second (bps), at which data can be communicated.

**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).

**Noise.** The average level of noise over the communications path.

**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.

### 2.3.1 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, the data rate can be increased by increasing the number of different signals. 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.

2.3.2 Shannon Channel Capacity

All of these concepts can be tied together neatly in a formula developed by the mathematician Claude Shannon. As we have just illustrated, the higher the data rate, the more damage that unwanted noise can do. For a given level of noise, we would expect that greater signal strength would improve the ability to correctly receive data in the presence of noise. The key parameter involved in this reasoning is the signal-to-noise ratio (S/N), which is the ratio of the power in a signal to the power contained in the noise that is present at a particular point in the transmission. Typically, this ratio is measured at a receiver, as it is at this point that an attempt is made to process the signal and eliminate the unwanted noise. For convenience, this ratio is often reported in decibels:

\[ (S/N)_{db} = 10 \log \left( \frac{\text{signal power}}{\text{noise power}} \right) \]

This expresses the amount, in decibels, that the intended signal exceeds the noise level. A high S/N will mean a high-quality signal and a low number of required intermediate repeaters. The effect of noise on signal is shown in the Figure 2.10.

The signal-to-noise ratio is important in the transmission of digital data because it sets the upper bound on the achievable data rate. Shannon’s result is that the theoretical maximum channel capacity, in bits per second, obeys the equation

\[ C = W \log_2(1 + (S/N)) \]

where C is the capacity of the channel in bits per second and W is the bandwidth of the channel in hertz.
As an example, consider a voice channel being used, via modem, to transmit digital data. Assume a bandwidth of 3100 Hz. A typical value of S/N for a voice-grade line is 30 dB, or a ratio of 1000:1. Thus,

\[ C = 3100 \log_2(1 + 1000) \]

\[ C = 30,894 \text{ bps} \]

This represents the theoretical maximum channel capacity that can be achieved.

### 2.4 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.
Figure 2.11 Line coding methods

Figure 2.11 shows various line codes that are used in practice. The figure shows the digital signals that are produced by the line codes for the binary sequence 101011100. The simplest scheme is the unipolar non-return-to-zero (NRZ) encoding in which a binary 1 is transmitted by sending a +A voltage level, and a 0 is transmitted by sending a 0 voltage. If binary 0s and 1s both occur with probability 1/2, then the average transmitted power for this line code is

\[(1/2)A^2 + (1/2)0^2 = A^2/2\]

The polar NRZ encoding method that maps a binary 1 to +A/2 and binary 0 to −A/2 is more efficient than unipolar NRZ in terms of average transmitted power. Its average power is given by

\[(1/2)(+A/2)^2 + (1/2)(-A/2)^2 = A^2/4\]

The spectrum that results from applying a given line code is of interest. We usually assume that the binary information is equally likely to be 0 or 1 and that they are statistically independent of each other, much as if they were produced by a sequence of independent coin flips. The unipolar and the polar NRZ encoding methods have the same 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 as shown in Figure 2.12. 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.
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 as shown in Figure 2.12.

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.

A problem with polar coding is that a systematic error in polarity can cause all 0s to be detected as 1s and all 1s as 0s. The problem can be avoided by mapping the binary information into transitions at the beginning of each interval. A binary 1 is transmitted by enforcing a transition at the beginning of a bit time, and a 0 by having no transition. The signal level within the actual bit time remains constant. Figure 2.11 shows an example of how differential encoding, or NRZ inverted, carries out this mapping. Starting at a given level, the sequence of bits determines the subsequent transitions at the beginning of each interval.

Figure 2.12 Spectra for different line coding

[Diagram showing spectra for bipolar, Manchester, and NRZ codes]
Note that differential encoding will lead to the same spectrum as binary and polar encoding. However, errors in differential encoding tend to occur in pairs. An error in one bit time will provide the wrong reference for the next time, thus leading to an additional error in the next bit.

Note that 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.

### 2.5 Modems and Digital Modulation

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 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 which 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 modulator-demodulator); 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.

![Diagram](image)

*Figure 2-13  Use of both analog and digital transmission for a computer to computer call. Conversion is done by the modems*

Various parameters of sine wave carrier like amplitude, frequency, or phase can be modulated to transmit information. In **amplitude modulation**, two different voltage levels are used to represent bits 0 and 1. In **frequency modulation**, also known as **frequency shift keying**, two (or more) different tones or frequencies are used. 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. Figure 2.14 illustrates the three forms of modulation.

Figure 2-14 (a) A binary signal. (b) Amplitude modulation (c) Frequency Modulation  
(d) Phase Modulation

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 2.15a we see dots at 0, 90, 180, and 270 degrees, with two amplitude levels per phase shift. Amplitude is indicated by the distance from the origin. In Figure 2.15b we see a different modulation scheme, in which 16 different combinations of amplitude and phase shift are used.
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 $16 = 2^4$ different combination each baud represents 4 bits. The scheme of Figure 2.15b when used to transmit 9600 bps over a 2400 baud line is called **QAM** (Quadrature Amplitude Modulation).

### 2.6 Transmission Media

The transmission medium is the physical path between transmitter and receiver in a data transmission system. Transmission media can be classified as guided or unguided. In both cases, communication is in the form of electromagnetic waves. With guided media, the waves are guided along a solid medium, such as copper twisted pair, copper coaxial cable, and optical fiber. The atmosphere and outer space are examples of unguided media that provide a means of transmitting electromagnetic signals but do not guide them; this form of transmission is usually referred to as **wireless transmission**.

The characteristics and quality of a data transmission are determined both by the characteristics of the medium and the characteristics of the signal. In the case of guided media, the medium itself is more important in determining the limitations of transmission.

For unguided media, the bandwidth of the signal produced by the transmitting antenna is more important than the medium in determining transmission characteristics. One key property of signals transmitted by antenna is directionality. In general, signals at lower frequencies are omnidirectional; that is, the signal propagates in all directions from the antenna. At higher frequencies, it is possible to focus the signal into a directional beam.

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

- **Bandwidth.** All other factors remaining constant, greater the bandwidth of the signal allowed over the transmission line, higher the data rate that can be achieved.

- **Transmission impairments.** Impairments, such as attenuation, limit the distance. For guided media, twisted pair generally suffers more impairment than coaxial cable, which in turn suffers more than optical fiber.

- **Interference.** Interference from competing signals in overlapping frequency bands can distort or wipe out a signal. Interference is of particular concern for unguided media, but it is also a problem with guided media. For guided media, interference can be caused by emanations from nearby cables. For example, twisted pair cables are often bundled together, and conduits often carry multiple cables. Interference can also be experienced from unguided transmissions. Proper shielding of a guided medium can minimize this problem.

- **Number of receivers.** A guided medium can be used to construct a point-to-point link or a shared link with multiple attachments. In the latter case, each attachment introduces some attenuation and distortion on the line, limiting distance and/or data rate.

![Figure 2.16 Electromagnetic spectrum and its uses for communication](image)

Figure 2.16 depicts the electromagnetic spectrum and indicates the frequencies at which various guided media and unguided transmission techniques operate. In this section, we examine these guided and unguided alternatives.

### 2.6.1 Guided Transmission Media

For guided transmission media, the transmission capacity, in terms of either data-rate or bandwidth, depends critically on the distance and on whether the medium is point-to-point or multipoint such as in a local area network (LAN). Table 3.1 indicates the type of performance typical for the common guided medium for long-distance point-to-point applications.

**TABLE 2.1 Point-to-point transmission characteristics of guided media.**
The three, guided media commonly used for data transmission are *twisted pair*, *coaxial cable*, and *optical fiber*.

**Twisted Pair**

The oldest and still most common transmission medium is **twisted pair**. It is also the least-expensive one.

A twisted pair consists of two insulated copper wires arranged in a regular spiral pattern, just like a DNA molecule. A wire pair acts as a single communication link. Typically, a number of these pairs are bundled together into cable by wrapping them in a tough protective sheath. Over longer distances, cables may contain hundreds of pairs. The twisting tends to decrease the cross-talk interference between adjacent pairs in a cable. Neighboring pairs in a bundle typically have somewhat different twist lengths to enhance the cross-talk interference.

It is the most commonly used medium in the telephone network as well as being the workhorse for communications within buildings. In the telephone system, individual residential telephone sets are connected to the local telephone exchange, or *end office*, by twisted-pair wire. These are referred to as *subscriber loops*. Within an office building, each telephone is also connected to a twisted pair, which goes to the in-house private branch exchange (PBX) system. Twisted pair is much less expensive than the other commonly used guided transmission media (coaxial cable, optical fiber) and is easier to work with. It is more limited in terms of data rate and distance.

Twisted pair may be used to transmit both analog and digital signals. For analog signals, amplifiers are required about every 5 to 6 km. For digital signals, repeaters are required every 2 or 3 km. For point-to-point analog signaling, a bandwidth of up to about 250 kHz is possible. This accommodates a number of voice channels. For long-distance digital point-to-point signaling, data rates of up to a few Mbps are possible; for very short distances, data rates of up to 100 Mbps have been achieved in commercially available products.

**Coaxial Cable**
Coaxial cable, like twisted pair, consists of two conductors, but is constructed differently to permit it to operate over a wider range of frequencies. It consists of a hollow outer cylindrical conductor that surrounds a single inner wire conductor. The inner conductor is held in place by either regularly spaced insulating rings or a solid dielectric material. The outer conductor is covered with a jacket or shield. A single coaxial cable has a diameter of from 0.4 to about 1 in. Because of its shielding, concentric construction, coaxial cable is much less susceptible to interference and cross-talk than is twisted pair. Coaxial cable can be used over longer distances and supports more stations on a shared line than twisted pair.

![Coaxial cable diagram](image)

**Figure 2-18  Coaxial cable.**

Coaxial cable is perhaps the most versatile transmission medium and is enjoying widespread use in a wide variety of applications; the most important of these are

- Television distribution
- Long-distance telephone transmission
- Short-run computer system links
- Local Area Networks

Coaxial cable is spreading rapidly as a means of distributing TV signals to individual homes - cable TV. A cable TV system can carry dozens or even hundreds of TV channels at ranges up to a few tens of miles.

Coaxial cable has traditionally been an important part of the long-distance telephone network. Today, it is getting replaced by optical fiber, terrestrial microwave, and satellite. Using frequency-division multiplexing, a coaxial cable can carry over 10,000 voice channels simultaneously. Coaxial cable is also commonly used for short-range connections between devices. Using digital signaling, coaxial cable can be used to provide high-speed I/O channels on computer systems.

Another application area for coaxial cable is local area networks. Coaxial cable can support a large number of devices with a variety of data and traffic types, over distances that encompass a single building or a complex of buildings.

Coaxial cable is used to transmit both analog and digital signals. Coaxial cable has frequency characteristics that are superior to those of twisted pair, and can hence be used effectively at higher frequencies and data rates. The principal constraints on performance are attenuation, thermal noise, and inter modulation noise.

For long-distance transmission of analog signals, amplifiers are needed every few kilometers, with closer spacing required if higher frequencies are used. The usable spectrum for analog signaling extends to about 400 MHz. For digital signaling, repeaters are needed every kilometer or so, with closer spacing needed for higher data rates.
**Optical transmission system and Optical Fiber.**
An optical transmission system has three components; the light source, the transmission medium, and the detector. Conventionally, a pulse of light indicates a bit 1 and absence of light indicates bit 0. Transmission medium is an ultra thin fiber of glass. The transmitter generates the light pulses based on the input electrical signal. The detector regenerates the electrical signal based on the light signal it detects on the transmission medium. By attaching a light source to one end of an optical fiber and a detector to the other, we have an *unidirectional data transmission system* that accepts an electrical signal, converts and transmits it by light pulse, and then reconverts the output to an electrical signal at the receiving end.

![Figure 2-19 Optical fiber](image)

An optical fiber is a thin (2 to 125 nm – nana meter – $10^{-9}$ meter), flexible medium capable of conducting an optical ray. Various glasses and plastics can be used to make optical fibers. The lowest losses have been obtained using fibers of ultrapure fused silica. Ultrapure fiber is difficult to manufacture; higher-loss multicomponent glass fibers are more economical and still provide good performance. Plastic fiber is even less costly and can be used for short-haul links, for which moderately high losses are acceptable.

An optical fiber cable has a cylindrical shape and consists of three concentric sections: the core, the cladding, and the jacket. The core is the innermost section and consists of one or more very thin strands, or fibers, made of glass or plastic. Each fiber is surrounded by *its own cladding*, a glass or plastic coating that has optical properties different from those of the core. The outermost layer, surrounding one or a bundle of cladded fibers, is the jacket. The jacket is composed of plastic and other material layered to protect against moisture, abrasion, crushing and other environmental dangers.

One of the most significant technological breakthroughs in data transmission has been the development of practical fiber optic communications systems. Optical fiber already enjoys considerable use in *long-distance telecommunications*. The continuing improvements in performance and decline in prices, together with the inherent advantages of optical fiber, have made it increasingly attractive for *local area networking* and *metropolitan networks*. The following characteristics distinguish optical fiber from twisted pair or coaxial cable:

**Greater capacity.** The potential bandwidth, and hence data rate of optical fiber is
immense; data rates of 2 Gbps over tens of kilometers have been demonstrated. Compare this capability to the practical maximum of hundreds of Mbps over about 1 km for coaxial cable and just a few Mbps over 1 km or up to 100 Mbps over a few tens of meters for twisted pair.

**Smaller size and lighter weight.** Optical fibers are considerably thinner than coaxial cable or bundled twisted-pair cable - at least an order of magnitude thinner for comparable information-transmission capacity. For cramped conduits in buildings and underground along public rights-of-way, the advantage of small size is considerable. The corresponding reduction in weight reduces structural support requirements.

**Lower attenuation.** Attenuation is significantly lower for optical fiber than for coaxial cable or twisted pair and is constant over a wide range.

**Electromagnetic isolation.** Optical fiber systems are not affected by external electromagnetic fields. Thus, the system is not vulnerable to interference, impulse noise, or cross-talk. By the same token, fibers do not radiate energy, thereby causing little interference with other equipment and thus providing a high degree of security from eavesdropping. In addition, fiber is inherently difficult to tap.

**Greater repeater spacing.** Fewer repeaters mean lower cost and fewer sources of error. The performance of optical fiber systems from this point of view has been steadily improving. For example, AT&T has developed a fiber transmission system that achieves a data rate of 3.5 Gbps over a distance of 318 km without repeaters. Coaxial and twisted-pair systems generally have repeaters every few kilometers.

### 2.6.2 Wireless (Radio) Transmission

Radio encompasses the electromagnetic spectrum in the range of 3 KHz to 300 GHz. In radio communications the signal is transmitted into the air or space, using an antenna that radiates energy at some carrier frequency. For example, in QAM modulation the information sequence determines a point in the signal constellation that specifies the amplitude and phase of the cosine wave that is transmitted. Depending on the frequency and the antenna, this energy can propagate in either a unidirectional or omni-directional fashion. In the unidirectional case a properly aligned antenna receives the modulated signal, and an associated receiver in the direction of the transmission recovers the original information. In the omnidirectional case any receiver with an antenna in the area of coverage can pick up the signal.

Radio communication systems are subject to a variety of transmission impairments. The attenuation in radio links varies logarithmically with the distance. Attenuation for radio systems also increases with rainfall. Radio systems are subject to multipath fading and interference. Multipath fading refers to the interference that results at a receiver when two or more versions of the same signal arrive at slightly different times because of reflections from different objects. If the arriving signals differ in polarity, then they will cancel each other. Multipath fading can result in wide fluctuations in the amplitude and phase of the received signal. Interference refers to energy that appears at the receiver from sources other than the transmitter. Interference can be generated
by other users of the same frequency band or by equipment that inadvertently transmits energy outside its band and into the bands of adjacent channels. Interference can seriously affect the performance of radio systems, and for this reason regulatory bodies apply strict requirements on the emission properties of electronic equipment.

![Radio Spectra Diagram]

**Figure 2.20 Radio Spectra**

Figure 2.20 gives the range of various frequency bands and their applications. The frequency bands are classified according to wavelengths. Thus the low frequency (LF) band spans the range 30 kHz to 300 kHz, which corresponds to a wavelength of 1 km to 10 km, whereas the extremely high frequency (EHF) band occupies the range from 30 to 300 GHz corresponding to wavelengths of 1 millimeter to 1 centimeter. Note that the progression of frequency bands in the logarithmic frequency scale have increasingly larger bandwidths, for example, the band from $10^{11}$ to $10^{12}$ Hz has a bandwidth of $0.9 \times 10^{12}$ Hz, whereas the band from $10^5$ to $10^6$ Hz has a bandwidth of $0.9 \times 10^6$ Hz.

The propagation properties of radio waves vary with the frequency. Radio waves at the VLF, LF, and MF bands follow the surface of the earth in the form of ground waves. VLF waves can be detected at distances up to about 1000 km, and MF waves, for example, AM radio, at much shorter distances. Radio waves in the HF band are reflected by the ionosphere and can be used for long-distance communications. These waves are detectable only within certain specific distances from the transmitter. Finally, radio waves in the VHF band and higher are not reflected back by the ionosphere and are detectable only within line-of-sight.

In general, radio frequencies below 1 GHz are more suitable for omnidirectional applications. For example, paging systems (beepers) are an omnidirectional application that provides one-way communications. Cordless telephones are example of an
omnidirectional application that provides two-way communications. Here a simple base station connects to a telephone outlet and relays signaling and voice information to a cordless phone. This technology allows the user to move around in an area of a few tens of meters while talking on the phone. Other applications of wireless transmission are given below.

**Cellular Communications**

Cellular telephone networks extend the basic telephone service to mobile users with portable telephones. Unlike conventional telephone service where a cable carries signal between the telephone exchange and the telephone equipment at the user end, here the radio signal is used for transmission. Both portable telephone equipment and the interface at the telephone exchange are capable of broadcasting and receiving the voice signal which modulates the radio carrier signal.

There are two standards in cellular telephone systems. GSM - Global System for Mobile (GSM), Code Division Multiple Access (CDMA)

**Wireless LANs**

Wireless LANs are another application of omnidirectional wireless communications. A Wireless LAN (WLAN) is a flexible data communication system implemented as an extension to, or as an alternative for, a wired LAN within a building or campus. Using electromagnetic waves, WLANs transmit and receive data over the air, minimizing the need for wired connections. Thus, WLANs combine data connectivity with user mobility, and, through simplified configuration, enable movable LANs. In recent past, WLANs have gained strong popularity.

Wireless LANs use electromagnetic airwaves (radio and infrared) to communicate information from one point to another without relying on any physical connection. The data being transmitted is superimposed on the radio carrier so that it can be accurately extracted at the receiving end.

Multiple radio carriers can exist in the same space at the same time without interfering with each other if the radio waves are transmitted on different radio frequencies. To extract data, a radio receiver tunes in (or selects) one radio frequency while rejecting all other radio signals on different frequencies.
In a typical WLAN configuration, a transmitter/receiver (transceiver) device, called an access point, connects to the wired network from a fixed location using standard Ethernet cable. At a minimum, the access point receives, buffers, and transmits data between the WLAN and the wired network infrastructure. A single access point can support a small group of users and can function within a range of less than one hundred to several hundred feet. The access point (or the antenna attached to the access point) is usually mounted high but may be mounted essentially anywhere that is practical as long as the desired radio coverage is obtained. End users access the WLAN through wireless LAN adapters, which are implemented as PC cards in notebook computers, or use ISA or PCI adapters in desktop computers, or fully integrated devices within hand-held computers.

**WLAN technology - Spread Spectrum**

Most wireless LAN systems use spread-spectrum technology, a wideband radio frequency technique developed by the military for use in reliable, secure, mission-critical communications systems. Spread-spectrum is designed to trade off bandwidth efficiency for reliability, integrity, and security.

**Satellite Communications**

Early satellite communications systems can be viewed as microwave systems with a single repeater in the sky. A (Geostationary) satellite is placed at an altitude of about 36,000 km above the equator where its orbit is stationary relative to the rotation of the earth. A modulated microwave radio signal is beamed to the satellite on an uplink carrier frequency. A transponder in the satellite receives the uplink signal, regenerates
it, and beams it down back to earth on a downlink carrier frequency. A satellite typically contains 12 to 20 transponders so it can handle a number of simultaneous transmissions. Each transponder typically handles about 50 Mbps. Satellites operate in the 4/6, 11/14, and 20/30 GHz bands, where the first number indicates the downlink frequency and the second number the uplink frequency.

Figure 2.22 satellite Communication

Geostationary satellite systems have been used to provide point-to-point digital communications to carry telephone traffic between two points. Satellite systems have an advantage over fiber systems in situations where communications needs to be established quickly or where deploying the infrastructure is too costly. Satellite systems are inherently broadcast in nature, so they are also used to simultaneously beam television, and other signals, to a large number of users.

Satellite systems are also used to reach mobile users who roam wide geographical areas. Constellations of low-earth orbit satellites (LEOS) are deployed. These include the Iridium and Teledesic systems. The satellites are not stationary with respect to the earth, but they rotate in such a way that there is continuous coverage of the earth. The component satellites are interconnected by high-speed links forming a network in the sky.

2.7 Multiplexing

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 2.23  Multiplexing

Figure 2.23a 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 2.23b shows 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.

2.7.1 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 2.24 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.
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 2.24b. 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.
2.7.2 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 2.25a 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 2.25b 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.

(a) Each signal transmits 1 unit every $3T$ seconds

(b) Combined signal transmits 1 unit every $T$ seconds

Figure 2.25 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 2.26

Figure 2.26 Time Division Multiplexing – T1 Carrier System

2.7.3 Wavelength Division Multiplexing
For fiber optic channels, a variation of frequency division multiplexing is used. It is called **WDM (Wavelength Division Multiplexing)**. A simple way of achieving FDM on fibers is depicted in Figure 2.27. Here two fibers come together at a prism (or more likely, a diffraction grating), each with its energy in a different band. The two beams are passed through the prism or grating, and combined onto a single shared fiber for transmission to a distant destination, where they are split again.

![WDM Diagram](image)

There is really nothing new here. As long as each channel has its own frequency range, and all the ranges are disjoint, they can be multiplexed together on the long-haul fiber. The only difference with electrical FDM is that an optical system using a diffraction grating is completely passive, and thus highly reliable.

It should be noted that the reason WDM is popular is that the energy on a single fiber is typically only a few Gigahertz ($10^9$ Hz) wide because it is currently impossible to convert between electrical and optical media any faster and since the bandwidth of a single fiber band is about 25,000 GHz, there is great potential for multiplexing many channels together over long-haul routes. A necessary condition, however, is that the incoming channels use different frequencies.

### 2.8 Circuit Switching

A network is frequently represented as a cloud that connects multiple users as shown in Figure 2.28a. A circuit-switched network is a generalization of a physical cable in the sense that it provides connectivity that allows information to flow between inputs and outputs to the network. Unlike a cable, however, a network is geographically distributed and consists of a graph of transmission lines (that is, links) interconnected by switches (nodes).
As shown in Figure 2.28b, the function of a circuit switch is to transfer the signal that arrives at a given input to an appropriate output. The interconnection of a sequence of transmission links and circuit switches enables the flow of information between inputs and outputs in the network.

2.8.1 Space-Division Switches

Space-division switches provide a separate physical connection between inputs and outputs so the different signals are separated in space. Figure 2.29 shows the crossbar switch, which is an example of this type of switch. The crossbar switch consists of an N x N array of crosspoints that can connect any input to any available output. When a request comes in from an incoming line for an outgoing line, the corresponding crosspoint is closed to enable information to flow from the input to the output. The crossbar switch is said to be nonblocking; in other words, connection requests are never denied because of lack of connectivity resources, that is, crosspoints. Connection requests are denied only when the requested outgoing line is already engaged in another connection.
The complexity of the crossbar switch as measured by the number of cross-points is \( N^2 \). This number grows quickly with the number of input and output ports. Thus a 1000-input-by-1000-output switch requires \( 10^6 \) crosspoints, and a 100,000 by 100,000 switch requires \( 10^{10} \) crosspoints. In the next section we show how the number of crosspoints can be reduced by using multistage switches.

**Multistage Switches**

Figure 2.30 shows a **multistage switch** that consists of three stages of smaller space-division switches. The \( N \) inputs are grouped into \( N/n \) groups of \( n \) input lines. Each group of \( n \) input lines enters a small switch in the first stage that consists of an \( n \times n \) array of crosspoints. Each input switch has one line connecting it to each of \( k \) intermediate stage \( N/n \times N/n \) switches. Each intermediate switch in turn has one line connecting it to each of the \( N/n \) switches in the third stage. The latter switches are \( k \times n \). In effect each set of \( n \) input lines *shares k possible paths to any one of the switches at the last stage*, that is, the first path goes through the first intermediate switch, the second path goes through the second intermediate switch, and so on. The resulting multistage switch is not necessarily nonblocking. For example, if \( k < n \), then as soon as a switch in the first stage has \( k \) connections, all other connections will be blocked.
The number of crosspoints required in a three-stage switch is the sum of the following components:

- \( \frac{N}{n} \) input switches \( \times nk \) crosspoints/input switch.
- \( k \) intermediate switches \( \times (\frac{N}{n})^2 \) crosspoints/intermediate switch.
- \( \frac{N}{n} \) output switches \( \times nk \) crosspoints/output switch.

In this case the total number of crosspoints is
\[
2Nk + k(\frac{N}{n})^2.
\]

The number of crosspoints required to make the switch nonblocking is
\[
2N(2n - 1) + (2n-1)(\frac{N}{n})^2
\]

The number of crosspoints can be minimized through the choice of group size \( n \). By differentiating the above expression with respect to \( n \), we find that the number of crosspoints is minimized if \( n \sim = (\frac{N}{2})^{1/2} \). The minimum number of crosspoints is then
\[
4N((2N)^{1/2} - 1).
\]

We then see that the minimum number of crosspoints grows at a rate proportional to \( N^{1.5} \) which is less than the \( N^2 \) growth rate of a crossbar switch.

### 2.8.2 Time-Division Switches

In the previous section, we explained how Time Division Multiplexing (TDM) could replace multiple physical lines by a single high-speed line. In TDM a slot within a frame corresponds to a single connection. The time-division switch uses time-slot interchange (TSI) technique to do the switching operation. It basically switches the time slots of the input frame in the output frame.
Consider users A & B who wishes to talk to each other. Suppose User A is assigned slot 3 and user B is assigned slot 5 then the time division switch will take the data information in the slot 3 in the input frame and puts it into the slot 5 in the output frame. In effect information bits in the slot 3 which came from user A is going into the slot allocated to the user B in the output frame. This results in the one way connection from A to B. Similarly the time division switch put the information in the slot 5 in the input frame to the slot 3 in the output frame resulting in connection from B to A.

The development of the TSI technique was crucial in completing the digitization of the telephone network. Starting in 1961 digital transmission techniques were introduced in the trunks that interconnected telephone central offices. Initially, at each office the digital streams would be converted back to analog form and switched by using space switches of the type discussed above. The introduction of TSI in digital time-division switches led to significant reductions in cost and to improvements in performance by obviating the need to convert back to analog form. Most modern telephone backbone networks are now entirely digital in terms of transmission and switching.

**Exercise**

1) Differentiate between a continuous signal and discrete signal.
2) What is a periodic signal? Explain.
3) Discuss the frequency domain concepts of electromagnetic signal.
4) Explain analog and digital repeaters.
5) Briefly discuss why digital transmission is preferred over analog transmission.
6) What is Nyquist sampling rate?
7) Discuss Shannon’s channel capacity.
8) Discuss various line coding techniques.
9) What are modems? Explain how higher data rates are achieved in modems.
10) Explain different modulation techniques.
11) List different guided transmission mediums. Discuss properties of optical fiber cable.
Paper Name: Data Communication

12) Discuss propagation properties of various radiowaves.
13) List the application of wireless transmission of signal in data communication. Discuss two of them.
14) What is multiplexing? List the different kind of multiplexing schemes.
15) How Wavelength Division Multiplexing is different from Frequency Division Multiplexing.
16) What is Circuit Switching?
17) Discuss Crossbar switch. How the cross-points required are reduced in the Multistage Switches?
18) Discuss the working of Time Division Multiplexing.
UNIT 3
Data Link Control

3.0 INTRODUCTION
3.1 ERROR DETECTION
   3.1.1 SINGLE PARITY CHECK
   3.1.2 CYCLIC REDUNDANCY CHECK (CRC)
3.2 ERROR CONTROL MECHANISMS
   3.2.1 STOP-AND-WAIT ARQ
   3.2.2 GO-BACK-NARQ
   3.2.3 SELECTIVE REPEAT ARQ
3.3 FLOW CONTROL
3.4 HIGH-LEVEL DATA LINK CONTROL (HDLC)
   3.4.1 BASIC CHARACTERISTICS OF HDLC
   3.4.2 HDLC FRAME STRUCTURE
   3.4.3 HDLC OPERATION

3.0 Introduction
Discussion so far was focused towards sending signals over a transmission link. For effective digital data communications, much more is needed to control and manage the exchange. In this chapter, we shift our emphasis to that of 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.

3.1 **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 3.1 General Error detection system](image)

As illustrated in Figure 3.1, 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.

### 3.1.1 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:

\[
b_{k+1} = (b_1 + b_2 + b_3 + \ldots + b_k) \mod 2
\]

where \( b_1, b_2, b_3, \ldots, b_k \) are the information bits. And \( b_{k+1} \) is the parity bit generated.

The modulo 2 arithmetic is given below:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0</td>
<td>+0</td>
<td>+1</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

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 3.2 Error detection system using check bits

Figure 3.2 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 \( 1/(k+1) \) 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 codewords 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 3.3, with codeword shown by \( \times \)s in the space and noncodeword by \( 0 \) s.
To minimize the probability of error-detection failure, we want the codewords to be selected so that they are spaced as far away from each other as possible. Thus the code in Figure 3.3a is a poor code because the codewords are close to each other or in other words distance between two valid codewords are low. On the other hand, the code in Figure 3.3b is good because the distance between codewords is maximized.

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

### 3.1.2 Cyclic Redundancy Check (CRC)

One of the most common, and most powerful, error-detecting code 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.

### 3.2 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.

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 3.4.
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, as discussed in previous section. 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 Figure 3.4. 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.

3.2.1 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 3.5a 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 3.5a 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.

(a) Frame 1 lost
(b) ACK lost

In parts (a) and (b) transmitting station A acts the same way, but part (b) receiving station B accepts frame 1 twice.

**Figure 3.5 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 3.5b shows the situation that begins as in Figure 3.5a, 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 3.5. 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 3.6 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.

![Diagram showing time-out and frames 0, 1, and 2 with ACKs]

**Figure 3.6 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.

### 3.2.2 Go-Back-N 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 $W$, 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 $(W_s - 1)$ 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 Figure 3.7. 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 3.7 Basic Go-Back-N ARQ](image)

Go-Back-N ARQ gets its name from the action that is taken when an error occurs. As shown in Figure 3.7, 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 $N = W_s$, 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 $(W_s - 1)$ 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 3.8 Go-Back-N ARQ

Figure 3.8 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.

3.2.3 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.
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 3.9 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 $R_{next}$ to $(R_{next} + W_r - 1)$, where $W_r$ 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 $R_{next}$ and $S_{last}$ through the delivery of the oldest outstanding frame. Thus ACK frames carry $R_{next}$ 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 $R_{next}$.

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 $R_{next}$ have already been received correctly and are buffered in the receiver. $R_{next}$ 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.

### 3.3 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.
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 $T_{prop}$ entity A stops transmitting as shown in Figure 3.10.

From the instant that B sent its signal, it receives an additional $2T_{prop} R$ 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.

**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 $W_s$ is made equal to the number of buffers that are available at the receiver. Because $W_s$ is the maximum number of outstanding frames from the transmitter, buffer overflow cannot occur at the receiver.

Figure 3.11 shows an example where receiver sends an acknowledgement after the last frame in a window has been received. In this figure $t_{cycle}$ is the basic delay that elapses from the time the first frame is transmitted to the receipt of its acknowledgement.
The delay in sending the acknowledgements has the effect of pacing or controlling the rate at which the transmitter sends frames to the receiver.

3.4 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.

3.4.1 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 multilink 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.

3.4.2 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 3.12a 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 3.12(a), (b) HDLC Frame Structure](image)

**Figure 3.12(a), (b) HDLC Frame Structure**
Figure 3.12(c), (d) HDLC frame structure

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.
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 3.13 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.

**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 (Figure 3.12b). 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.

**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 3.12c 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.

**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.

**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.

**3.4.3 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.

**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.

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.

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.

Exercise
1) List some of the requirements and objectives of data link control, for effective data communication between two directly connected transmitting-receiving stations.
2) Discuss Single parity check error detection.
3) Discuss error detection system. What are the limitations of error detection systems? What is the objective in selecting an error-detection codeword?
4) What is Cyclic Redundancy Check? Discuss.
5) What are the types of error possible in a frame transmission over a channel?
6) What is ARQ protocol? What are the mechanisms involved in it? Discuss basic elements of ARQ mechanism.
7) List various ARQ protocols. Discuss at least two of them.
8) Why Flow control is required? Discuss Sliding Window Flow Control mechanism.
9) What is HDLC? Discuss the basic characteristics of HDLC.
10) Discuss the HDLC frame structure.
11) Discuss different HDLC operations.
4.0 Introduction

Basically there are two types of Networks Technology: Switched Networks and Broadcast Networks. In Switched Networks, systems are interconnected by means of point-to-point transmission lines, multiplexers and switches, which are covered in Chapter 2. In Switched Networks, the transfer of packet across networks requires routing, to direct packets from source to destination, as the source and destination stations will not be connected by a single transmission link most of the cases. But in Broadcast Networks, all the systems are connected to common transmission medium, which acts as a broadcast medium making routing unnecessary. In its simplest form, a transmission from any one system is broadcast to and received by all other systems and
if the received packet is meant for the system, it will process it, otherwise discards. Since the transmission medium is shared by all the connected systems, there is a need for additional layer, consisting of **Medium Access Control (MAC) Protocol** to orchestrate the transmissions from various systems. The role of the MAC protocol is to *co-ordinate the access to the transmission medium* so that packets transmitted from the systems do not interfere with other packets.

There are two networks *based on broadcast medium*. Local Area Networks (LANs) and Metropolitan Area Networks (MANs). The definitions of LANs and MANs, taken from one of the IEEE 802 standards documents are as follows.

The **LANs** are distinguished from other types of data networks in that they are optimized for a moderate size geographic area such as a single office building or a campus. The LAN is a shared medium peer-to-peer communications network that broadcasts information for all stations to receive. As a consequence, it does not inherently provide privacy. The LAN enables stations to communicate directly using a common physical medium on a point-to-point basis without any intermediate switching node being required. There is always need for an access sub-layer in order to arbitrate the access to the shared medium. The network is generally owned, used, and operated by a single organization.

A **MAN** is optimized for a larger geographical area than a LAN, ranging from several blocks of buildings to entire city. As with local networks, MANs can also depend on communications channels of moderate-to-high data rates. Error rates and delay may be slightly higher than might be obtained on a LAN. A MAN might be owned and operated by a single organization, but usually will be used by many individuals and organizations. They will often provide means for internetworking of local area networks.

The key technology ingredients that determine the nature of a LAN or MAN are
- **Topology**
- **Transmission medium**
- **Medium Access Control technique**

Various topologies and transmission medium that are most commonly used for LANs and MANs, various Medium Access Control Protocols and the concept of a bridge, which play a critical role in extending LAN coverage, are discussed in this Chapter. However LAN is the most widely implemented network amongst the two and with the usage of bridges they cover wider area also and underlying technologies used in LAN and MAN bearing many similarities we concentrate here on LAN.

**4.1 LAN Architecture**

The architecture of a LAN is best described in terms of a layering of protocols that organize the basic functions of a LAN. This section opens with a description of the standardized protocol architecture for LANs, which encompasses physical, medium access control, and logical link control layers. Each of these layers is then examined in
LAN Protocol Architecture

Protocols defined specifically for LAN and MAN transmission address issues relating to the transmission of blocks of data over the network. In OSI terms, higher layer protocols (layer 3 or 4 and above) are independent of network architecture and are applicable to LANs, MANs, and WANs. Thus, a discussion of LAN protocols is concerned principally with lower layers of the OSI model.

Figure 4.1 relates the LAN protocols to the OSI architecture. This LAN architecture was developed by the IEEE 802 committee and has been adopted by all organizations working on the specification of LAN standards. It is generally referred to as the IEEE 802 reference model.

Working from the bottom up, the lowest layer of the IEEE 802 reference model corresponds to the physical layer of the OSI model, and includes such functions as

- Encoding / decoding of signals
- Preamble generation / removal (for synchronization)
- Bit transmission / reception

In addition, the physical layer of the 802 models includes a specification of the transmission medium and the topology. Generally, this is considered below the lowest layer of the OSI model. However, the choice of transmission medium and topology...
Figure 4.1 IEEE 802 protocol layers compared to OSI mode

is critical in LAN design, and so a specification of the medium is included. Above the physical layer are the functions associated with providing service to LAN users. These include:

- On transmission, assemble data into a frame with address and error-detection fields.
- On reception, disassemble frame, perform address recognition and error detection.
- Govern access to the LAN transmission medium.
- Provide an interface to higher layers and perform flow and error control.

These are functions typically associated with OSI layer 2 and are grouped into a logical link control (LLC) layer. The functions in the first three items are treated as a separate layer, called medium access control (MAC). The separation is done for the following reasons:

- The logic required to manage access to a shared-access medium is not found in traditional layer-2 data link control.
- For the same LLC, several MAC options may be provided.

Some of the standards that have been issued by IEEE are listed below

<table>
<thead>
<tr>
<th>Standard</th>
<th>Topology</th>
<th>Network</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.3 CSMA/CD</td>
<td>Bus / Tree / Star</td>
<td>LAN</td>
<td>Co-axial cable, Twisted Pair cable, Optical Fiber</td>
</tr>
<tr>
<td>IEEE 802.4 Token Bus</td>
<td>Bus / Tree / Star</td>
<td>LAN</td>
<td>Co-axial cable, Optical Fiber</td>
</tr>
<tr>
<td>IEEE 802.5 Token Ring</td>
<td>Ring topology</td>
<td>LAN</td>
<td>Twisted Pair cable</td>
</tr>
<tr>
<td>FDDI Token Ring</td>
<td>Ring topology</td>
<td>LAN</td>
<td>Optical Fiber, Unshielded Twisted Pair</td>
</tr>
<tr>
<td>IEEE 802.6 DQDB</td>
<td>Dual Bus topology</td>
<td>MAN</td>
<td>Optical Fiber</td>
</tr>
<tr>
<td>IEEE 802.11 CSMA Polling</td>
<td>Wireless Infrared</td>
<td>LAN</td>
<td>Spread Spectrum</td>
</tr>
</tbody>
</table>

Most of the standards were developed by a committee IEEE 802, sponsored by the Institute for Electrical and Electronics Engineers. All of these standards have subsequently been adopted as international standards by the International Organization for Standardization (ISO).
Figure 4.2 LAN protocols in context

Figure 4.2 illustrates the relationship between the levels of the architecture. User data are passed down to LLC, which appends control information as a header, creating an LLC protocol data unit (PDU). This control information is used in the operation of the LLC protocol. The entire LLC PDU is then passed down to the MAC layer, which appends control information header and trailer at the front and back of the packet, forming a MAC frame. Again, the control information in the frame is needed for the operation of the MAC protocol. For context, the figure also shows the use of TCP/IP and an application layer above the LAN protocols.

4.2 LAN Topologies
The common topologies for LANs are bus, tree, ring, and star (Figure 4.3). The bus is a special case of the tree, with only one trunk and no branches.
4.2.1 Bus and Tree Topologies

Both bus and tree topologies are characterized by the use of a multipoint medium. For the bus, all stations attach, through appropriate hardware interfacing known as a tap, directly to a linear transmission medium, or bus. Full-duplex operation between the station and the tap allows data to be transmitted onto the bus and received from the bus. A transmission from any station propagates the length of the medium in both directions and can be received by all other stations. At each end of the bus is a terminator, which absorbs any signal, removing it from the bus.

The tree topology is a generalization of the bus topology. The transmission medium is a branching cable with no closed loops. The tree layout begins at a point known as the headend, where one or more cables start, and each of these may have branches. The branches in turn may have additional branches to allow quite complex layouts. Again, a transmission from any station propagates throughout the medium and can be received by all other stations.
Two problems present themselves in this arrangement. First, because a transmission from any one station can be received by all other stations, there needs to be some way of indicating for whom the transmission is intended. Second, a mechanism is needed to regulate transmission. To see the reason for this, consider that if two stations on the bus attempt to transmit at the same time, their signals will overlap and become garbled. Or, consider that one station decides to transmit continuously for a long period of time.

To solve these problems, stations transmit data in small blocks, known as frames. Each frame consists of a portion of the data that a station wishes to transit, plus a frame header that contains control information. Each station on the bus is assigned a unique address, or identifier, and the destination address for a frame is included in its header.
Figure 4.4 illustrates the scheme. In this example, station C wishes to transmit a frame of data to A. The frame header includes A’s address. As the frame propagates along the bus, it passes B, which observes the address and ignores the frame. A, on the other hand, sees that the frame is addressed to itself and therefore copies the data from the frame as it goes by.

So the frame structure solves the first problem mentioned above: It provides a mechanism for indicating the intended recipient of data. It also provides the basic tool for solving the second problem, the regulation of access. In particular, the stations take turns sending frames in some cooperative fashion; this involves putting additional control information into the frame header.

With the bus or tree, no special action needs to be taken to remove frames from the medium. When a signal reaches the end of the medium, it is absorbed by the terminator.

### 4.2.2 Ring Topology

In the ring topology, the network consists of a set of repeaters joined by point-to-point links to form a closed loop. The repeater is a comparatively simple device, capable of receiving data on one link and transmitting them, bit by bit, on the other link as fast as they are received, with no buffering at the repeater. The links are unidirectional; that is, data are transmitted in one direction only and all are oriented in the same way. Thus, data circulate around the ring in one direction (clockwise or counterclockwise). Each station attaches to the network at a repeater and can transmit data onto the network through that repeater.
As with the bus and tree, data are transmitted in frames. As a frame circulates past all the other stations, the destination station recognizes its address and copies the frame into a local buffer as it goes by. The frame continues to circulate until it returns to the source station, where it is removed (Figure 4.5). Because multiple stations share the ring, medium access control is needed to determine at what time each station may insert frames.

4.2.3 Star Topology

In the star LAN topology, each station is directly connected to a common central node. Typically, each station attaches to a central node, referred to as the star coupler, via two point-to-point links, one for transmission in each direction.

In general, there are two alternatives for the operation of the central node. One approach is for the central node to operate in a broadcast fashion. A transmission of a frame from one station to the node is retransmitted on all of the outgoing links. In this case, although the arrangement is physically a star, it is logically a bus; a transmission from any station is received by all other stations, and only one station at a time may successfully transmit.

Another approach is for the central node to act as a frame-switching device. An incoming frame is buffered in the node and then retransmitted on an outgoing link to the destination station.

4.3 Medium Access Control

All LANs and MANs consist of collections of devices that must share the network’s transmission capacity. Some means of controlling access to the transmission medium is needed to provide for an orderly and efficient use of that capacity. This is the function of a Medium Access Control (MAC) protocol.

The key parameters in any medium access control technique are where and how. Where refers to whether control is exercised in a centralized or distributed fashion. In a
centralized scheme, a controller is designated that has the authority to grant access to
the network. A station wishing to transmit must wait until it receives permission from
the controller. In a decentralized network, the stations collectively perform a medium
access control function to dynamically determine the order in which stations transmit.
A centralized scheme has certain advantages, such as the following:
- It may afford greater control over access for providing such things as priorities,
  overrides, and guaranteed capacity.
- It enables the use of relatively simple access logic at each station.
- It avoids problems of distributed coordination among peer entities.
The principal disadvantages of centralized schemes are
- It creates a single point of failure; that is, there is a point in the network that, if it
  fails, causes the entire network to fail.
- It may act as a bottleneck, reducing performance.
The second parameter, how, is constrained by the topology and is a trade-off between
competing factors, including cost, performance, and complexity. In general, we can
categorize access control techniques as being either synchronous or asynchronous.
With synchronous techniques, a specific capacity is dedicated to a connection; this is
the same approach used in circuit switching, frequency-division multiplexing (FDM),
and synchronous time-division multiplexing (TDM). Such techniques are generally not
optimal in LANs and MANs because the needs of the stations are unpredictable. It is
preferable to be able to allocate capacity in an asynchronous (dynamic) fashion, more or
less in response to immediate demand. The asynchronous approach can be further
subdivided into three categories: round robin, reservation, and contention.

4.3.1 Round Robin
With round robin, each station in turn is given the opportunity to transmit. During that
opportunity, the station may decline to transmit or may transmit subject to a specified
upper bound, usually expressed as a maximum amount of data transmitted or time for
this opportunity. In any case, the station, when it is finished, relinquishes its turn, and
the right to transmit passes to the next station in logical sequence. Control of sequence
may be centralized or distributed. Polling is an example of a centralized technique.

When many stations have data to transmit over an extended period of time, round
robin techniques can be very efficient. If only a few stations have data to transmit over
an extended period of time, then there is a considerable overhead in passing the turn
from station to station, as most of the stations will not transmit but simply pass their
turns. Under such circumstances, other techniques may be preferable, largely
depending on whether the data traffic has a stream or bursty characteristic. Stream
traffic is characterized by lengthy and fairly continuous transmissions; examples are
voice communication, telemetry, and bulk file transfer. Bursty traffic is characterized by
short, sporadic transmissions; interactive terminal-host traffic fits this description.

4.3.2 Reservation
For stream traffic, reservation techniques are well suited. In general, for these
techniques, time on the medium is divided into slots, much as with synchronous TDM. A station wishing to transmit reserves future slots for an extended or even an indefinite period. Again, reservations may be made in a centralized or distributed fashion.

### 4.3.3 Contention

For bursty traffic, contention techniques are usually appropriate. With these techniques, no control is exercised to determine whose turn it is; all stations contend for time in a way that can be, as we shall see, rather rough and tumble. These techniques are, of necessity, distributed by nature. Their principal advantage is that they are simple to implement and, *under light to moderate load of data traffic, they are efficient.* For some of these techniques, however, performance tends to collapse under heavy load.

Although both centralized and distributed reservation techniques have been implemented in some LAN products, round robin and contention techniques are the most common.

The discussion above has been somewhat abstract and should become clearer as specific techniques are discussed in later sections.

The following Table lists the MAC protocols that are defined in LAN and MAN standards.

<table>
<thead>
<tr>
<th>Bus Topology</th>
<th>Ring Topology</th>
<th>Switched Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Round Robin</strong></td>
<td>Token Bus (IEEE 802.4)</td>
<td>Token Ring (IEEE 802.5)</td>
</tr>
<tr>
<td><strong>Reservation</strong></td>
<td>DQDB (IEEE 802.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Contention</strong></td>
<td>CSMA/CD (IEEE 802.3)</td>
<td>CSMA (IEEE 802.11)</td>
</tr>
</tbody>
</table>

### 4.4 MAC Frame Format

The MAC layer receives a block of data from the LLC layer and is responsible for performing functions related to medium access and for transmitting the data. As with other protocol layers, MAC implements these functions, making use of a protocol data unit at its layer; in this case, the PDU is referred to as a MAC frame.

The exact format of the MAC frame differs somewhat for the various MAC protocols in use. In general, all of the MAC frames have a format similar to that of Figure 4.6.
Figure 4.6 LLC PDU with generic MAC frame format

The fields of this frame are

- **MAC control.** This field contains any protocol control information needed for the functioning of the MAC protocol. For example, a priority level could be indicated here.

- **Destination MAC address.** The address of the destination device on the LAN for this frame.

- **Source MAC address.** The address of the source device on the LAN from which this frame is being transmitted.

- **LLC.** The LLC data from the next higher layer.

- **CRC.** The cyclic redundancy check field (also known as the frame check sequence, FCS, field). This is an error-detecting code as we have seen in previous chapter.

In most data link control protocols, the data link protocol entity is responsible not only for detecting errors using the CRC, but for recovering from those errors by retransmitting damaged frames. In the LAN protocol architecture, these two functions are split between the MAC and LLC layers. The MAC layer is responsible for detecting errors and discarding any frames that are in error. The LLC layer optionally keeps track of which frames have been successfully received and retransmits unsuccessful frames.

### 4.5 Logical Link Control

The LLC layer for LANs is similar in many respects to other link layers in common use. Like all link layers, LLC is concerned with the transmission of a link-level protocol data unit (PDU) between two stations, without the necessity of an intermediate switching node. LLC has two characteristics not shared by most other link control protocols:

1. It must support the multi-access, shared-medium nature of the link.
2. It is relieved of some details of link access by the MAC layer.

Addressing in LLC involves specifying the source and destination LLC users. Typically,
a user is a higher-layer protocol or a network management function in the station. These LLC user addresses are referred to as service access points (SAPs), in keeping with OSI terminology for the user of a protocol layer. We look first at the services that LLC provides to a higher-level user, then at the LLC protocol.

LLC Services

LLC specifies the mechanisms for addressing stations across the medium and for controlling the exchange of data between two users. The operation and format of this standard is based on HDLC. Three services are provided as alternatives for attached devices using LLC:

- **Unacknowledged connectionless service.** This service is a datagram style service. It is a very simple service that does not involve any of the flow and error-control mechanisms. Thus, the delivery of data is not guaranteed. However, in most devices, there will be some higher layer of software that deals with reliability issues.
- **Connection-mode service.** This service is similar to that offered by HDLC. A logical connection is set up between two users exchanging data, and flow control and error control are provided.
- **Acknowledged connectionless service.** This is a cross between the previous two services. It provides that datagrams are to be acknowledged, but no prior logical connection is set up.

Typically, a vendor will provide these services as options that the customer can select when purchasing the equipment. Alternatively, the customer can purchase equipment that provides two or all three services and select a specific service based on application.

The unacknowledged connectionless service requires minimum logic and is useful in two contexts. First, it will often be the case that higher layers of software will provide the necessary reliability and flow-control mechanism, and it is efficient to avoid duplicating them. For example, either TCP or the ISO transport protocol standard would provide the mechanisms needed to ensure that data are delivered reliably. In a monitoring application, the loss of an occasional data unit would not cause distress, as the next report should arrive shortly. Thus, in most cases, the unacknowledged connectionless service is the preferred option.

The connection-mode service could be used in very simple devices, such as terminal controllers, that have little software operating above this level. In these cases, it would provide the flow control and reliability mechanisms normally implemented at higher layers of the communications software.

The acknowledged connectionless service is useful in several contexts. With the connection-mode service, the logical link control software must maintain some sort of table for each active connection, so as to keep track of the status of that connection. If the user needs guaranteed delivery, but there are a large number of destinations for data, then the connection-mode service may be impractical because of the large number of tables required. Another use is the handling of important and time-critical alarm or
emergency control signals in a factory. Because of their importance, an acknowledgment is needed so that the sender can be assured that the signal got through. Because of the urgency of the signal, the user might not want to take the time to first establish a logical connection and then send the data.

4.6 LAN Systems
We now consider specific LAN systems. As was mentioned earlier, the medium access control technique and topology are key characteristics used in the classification of LANs and in the development of standards. Following systems are discussed here:

- Ethernet and Fast Ethernet (CSMA/CD)
- Token Ring
- FDDI

4.6.1 Ethernet and Fast Ethernet (CSMA/CD)
The most commonly used medium access control technique for bus/tree and star topologies is **Carrier Sense Multiple Access with Collision Detection (CSMA/CD)**. The original baseband version of this technique was developed by Xerox as part of the Ethernet LAN. The original broadband version was developed by MITRE as part of its MITREnet LAN. All of this work formed the basis for the IEEE 802.3 standard.

This Ethernet system derived its name from the *luminiferous ether*, through which electromagnetic radiations were once thought to propagate. (When British physicist James Clerk Maxwell discovered in nineteenth century that electromagnetic radiation could be described by the wave equation, scientists assumed that space must be filled with some *ethereal medium* in which medium the radiation was propagating. Only later they found that electromagnetic radiation could propagate in a vacuum.)

**Precursors – History of CSMA/CD**
The 802.3 has an interesting history. The real beginning was the ALOHA system developed to allow radio communication between machines scattered over Hawaiian Islands. CSMA/CD and its precursors can be termed random access, or contention, techniques. They are random access in the sense that there is no predictable or scheduled time for any station to transmit; station transmissions are ordered randomly. They exhibit contention in the sense that stations contend for time on the medium. The earliest of these techniques, known as ALOHA, was developed for packet radio networks. However, it is applicable to any shared transmission medium. ALOHA, or pure ALOHA as it is sometimes called, is a true free-for-all.

Whenever a station has a frame to send, it does so. The station then listens for an amount of time equal to the maximum possible round-trip propagation delay on the network (twice the time it takes to send a frame between the two most widely separated stations) plus a small fixed time increment. If the station hears an acknowledgment during that time, fine; otherwise, it resends the frame. If the station fails to receive an acknowledgment after repeated transmissions, it gives up. A receiving station determines the correctness of an incoming frame by examining a frame-check-sequence field, as in HDLC. If the frame is valid and if the destination address in the frame header matches the receiver’s address, the station immediately sends an
acknowledgment. The frame may be invalid due to noise on the channel or because another station transmitted a frame at about the same time. In the latter case, the two frames may interfere with each other at the receiver so that neither gets through; this is known as a collision. If a received frame is determined to be invalid, the receiving station simply ignores the frame.

ALOHA is as simple as can be, and pays a penalty for it. Because the number of collisions rise rapidly with increased load, the maximum utilization of the channel is only about 18%. To improve efficiency, a modification of ALOHA, known as slotted ALOHA, was developed. In this scheme, time on the channel is organized into uniform slots whose size equals the frame transmission time. Some central clock or other technique is needed to synchronize all stations. Transmission is permitted to begin only at a slot boundary. Thus, frames that do overlap will do so totally. This increases the maximum utilization of the system to about 37%.

Both ALOHA and slotted ALOHA exhibit poor utilization. Both fail to take advantage of one of the key properties of both packet radio and LANs, which is that propagation delay between stations is usually very small compared to frame transmission time. Consider the following observations. If the station-to-station propagation time is large compared to the frame transmission time, then, after a station launches a frame, it will be a long time before other stations know about it. During that time, one of the other stations may transmit a frame; the two frames may interfere with each other and neither gets through. Indeed, if the distances are great enough, many stations may begin transmitting, one after the other, and none of their frames get through unscathed.

Suppose, however, that the propagation time is small compared to frame transmission time. In that case, when a station launches a frame, all the other stations know it almost immediately. So, if they had any sense, they would not try transmitting until the first station was done. Collisions would be rare because they would occur only when two stations began to transmit almost simultaneously. Another way to look at it is that a short delay time provides the stations with better feedback about the state of the network; this information can be used to improve efficiency.

The foregoing observations led to the development of carrier-sense multiple access (CSMA). With CSMA, a station wishing to transmit first listens to the medium to determine if another transmission is in progress (carrier sense). If the medium is in use, the station must wait. If the medium is idle, the station may transmit. It may happen that two or more stations attempt to transmit at about the same time. If this happens, there will be a collision; the data from both transmissions will be garbled and not received successfully. To account for this, a station waits a reasonable amount of time, after transmitting, for an acknowledgment, taking into account the maximum round-trip propagation delay, and the fact that the acknowledging station must also contend for the channel in order to respond. If there is no acknowledgment, the station assumes that a collision has occurred and retransmits.
One can see how this strategy would be effective for networks in which the average frame transmission time is much longer than the propagation time. Collisions can occur only when more than one user begins transmitting within a short time (the period of the propagation delay). If a station begins to transmit a frame, and there are no collisions during the time it takes for the leading edge of the packet to propagate to the farthest station, then there will be no collision for this frame because all other stations are now aware of the transmission.

The maximum utilization achievable using CSMA can far exceed that of ALOHA or slotted ALOHA. The maximum utilization depends on the length of the frame and on the propagation time; the longer the frames or the shorter the propagation time, the higher the utilization.

With CSMA, an algorithm is needed to specify what a station should do if the medium is found busy. The most common approach, and the one used in IEEE 802.3, is the 1-persistent technique. A station wishing to transmit listens to the medium and obeys the following rules:
1. If the medium is idle, transmit; otherwise, go to step 2.
2. If the medium is busy, continue to listen until the channel is sensed idle; then transmit immediately.

If two or more stations are waiting to transmit, a collision is guaranteed. Things get sorted out only after the collision.

**Description of CSMA/CD**

CSMA, although more efficient than ALOHA or slotted ALOHA, still has one glaring inefficiency: When two frames collide, the medium remains unusable for the duration of transmission of both damaged frames. For long frames, compared to propagation time, the amount of wasted capacity can be considerable. This waste can be reduced if a station continues to listen to the medium while transmitting. This leads to the following rules for CSMA/CD:
1. If the medium is idle, transmit; otherwise, go to step 2.
2. If the medium is busy, continue to listen until the channel is idle, then transmit immediately.
3. If a collision is detected during transmission, transmit a brief Jamming signal to assure that all stations know that there has been a collision and then cease transmission.
4. After transmitting the jamming signal, wait a *random amount of time*, and then attempt to transmit again. (Repeat from step 1.)

Figure 4.7 illustrates the technique for a baseband bus. At time to, station A begins transmitting a packet addressed to D. At t1, both B and C are ready to transmit. B senses a transmission and so defers. C, however, is still unaware of A's transmission and begins its own transmission. When A's transmission reaches C, at t2, C detects the collision and ceases transmission. The effect of the collision propagates back to A, where it is detected some time later, t3, at which time A ceases transmission.
With CSMA/CD, the amount of wasted capacity is reduced to the time it takes to detect a collision. Question: how long does that take? Let us consider first the case of a baseband bus and consider two stations as far apart as possible. For example, in Figure 4.7, suppose that station A begins a transmission and that just before that transmission reaches D, D is ready to transmit. Because D is not yet aware of A’s transmission, it begins to transmit. A collision occurs almost immediately and is recognized by D. However, the collision must propagate all the way back to A before A is aware of the collision. By this line of reasoning, we conclude that the amount of time that it takes to detect a collision is no greater than twice the end-to-end propagation delay.

An important rule followed in most CSMA/CD systems, including the IEEE standard, is that frames should be long enough to allow collision detection prior to the end of transmission. If shorter frames are used, then collision detection does not occur, and CSMA/CD exhibits the same performance as the less efficient CSMA protocol.

Although the implementation of CSMA/CD is substantially the same for baseband and broadband, there are differences. One is the means for performing carrier sense; for
baseband systems, this is done by detecting a voltage pulse train and for broadband, the RF carrier is detected.

**CSMA/CD MAC Frame**

Figure 4.8 depicts the frame format for the 802.3 protocol; it consists of the following fields:

- **Preamble.** A 7-octet pattern of alternating 0s and 1s used by the receiver to establish bit synchronization.
- **Start frame delimiter.** The sequence 10101011, which indicates the actual start of the frame and enables the receiver to locate the first bit of the rest of the frame.
- **Destination address (DA).** Specifies the station(s) for which the frame is intended. It may be a unique physical address, a group address, or a global address. The choice of a 16 or 48-bit address length is an implementation decision, and must be the same for all stations on a particular LAN.
- **Source address (SA).** Specifies the station that sent the frame.
- **Length.** Length of the LLC data field.
- **LLC data.** Data unit supplied by LLC.
- **Pad.** Octets added to ensure that the frame is long enough for proper CD operation.

<table>
<thead>
<tr>
<th>Octets</th>
<th>Preamble</th>
<th>SFD</th>
<th>DA</th>
<th>SA</th>
<th>Length</th>
<th>LLC Data</th>
<th>Pad</th>
<th>FCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>2 or 6</td>
<td>2 or 6</td>
<td>2</td>
<td>&gt;= 0</td>
<td>&gt;= 0</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4.8 IEEE 802.3 frame format.*

Legend

SFD = Start-frame delimiter  SA = Source address
DA = Destination address  FCS = Frame-check sequence

**4.6.2 Token Ring**

Token ring is the most commonly used MAC protocol for ring-topology LANs. In this section, we examine two standard LANs that use token ring: IEEE 802.5 and FDDI.

**4.6.2.1 IEEE 802.5 Medium Access Control**

**MAC Protocol**

The token ring technique is based on the use of a small frame, called a *token*, which circulates when all stations are idle. A station wishing to transmit must wait until it detects a token passing by. It then seizes the token by changing one bit in the token, which transforms it from a token into a start-of-frame sequence for a data frame. The station then appends and transmits the remainder of the fields needed to construct a data frame.

When a station seizes a token and begins to transmit a data frame, there is no token on the ring, so other stations wishing to transmit must wait. The frame on the ring will make a round trip and be absorbed by the transmitting station. The transmitting station will insert a new token on the ring when both of the following conditions have been met:
The station has completed transmission of its frame.

The leading edge of the transmitted frame has returned (after a complete circulation of the ring) to the station.

If the bit length of the ring is less than the frame length, the first condition implies the second; if not, a station could release a free token after it has finished transmitting but before it begins to receive its own transmission. The second condition is not strictly necessary, and is relaxed under certain circumstances. The advantage of imposing the second condition is that it ensures that only one data frame at a time may be on the ring and that only one station at a time may be transmitting, thereby simplifying error-recovery procedures.

Once the new token has been inserted on the ring, the next station downstream with data to send will be able to seize the token and transmit. Figure 4.9 illustrates the technique. In the example, A sends a packet to C, which receives it and then sends its own packets to A and D.

Note that under lightly loaded conditions, there is some inefficiency with token ring because a station must wait for the token to come around before transmitting. However, under heavy loads, which is when it matters, the ring functions in a round-robin fashion, which is both efficient and fair. To see this, consider the configuration in Figure 4.9. After station A transmits, it releases a token. The first station with an opportunity to transmit is D. If D transmits, it then releases a token and C has the next opportunity, and so on.

The principal advantage of token ring is the flexible control over access that it provides. In the simple scheme just described, the access is fair. As we shall see, schemes can be used to regulate access to provide for priority and for guaranteed bandwidth services.
Figure 4.9 Token ring operation

The principal disadvantage of token ring is the requirement for token maintenance. Loss of the token prevents further utilization of the ring. Duplication of the token can also disrupt ring operation. One station must be selected as a monitor to ensure that exactly one token is on the ring and to ensure that a free token is reinserted, if necessary.

Token ring MAC frame format

Figure 4.10 depicts the frame format for the 802.5 protocol. It consists of the following fields:

- **Starting delimiter (SD).** Indicates start of frame. The SD consists of signaling patterns that are distinguishable from data. It is coded as follows: JKOJK000, where J and K are nondata symbols. The actual form of a nondata symbol depends on the signal encoding on the medium.
• Access control (AC). Has the format PPPTMRRR, where PPP and RRR are 3-bit priority and reservation variables, and M is the monitor bit; their use is explained below. T bit indicates whether this is a token or data frame. In the case of a token frame, the only remaining field is ED.

• Frame control (FC). Indicates whether this is an LLC data frame. If not, bits in this field control operation of the token ring MAC protocol.

<table>
<thead>
<tr>
<th>Octets</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>2 or 6</th>
<th>2 or 6</th>
<th>≥0</th>
<th>4</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>AC</td>
<td>FC</td>
<td>BA</td>
<td>SA</td>
<td>Data unit</td>
<td>FCS</td>
<td>ED</td>
<td>FS</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>starting delimiter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>frame control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>destination address</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>source address</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data unit</td>
<td>frame check sequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCS</td>
<td>ending delimiter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>frame status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) General Frame Format

<table>
<thead>
<tr>
<th>Octets</th>
<th>SD</th>
<th>AC</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>K</td>
<td>I</td>
<td>L</td>
</tr>
<tr>
<td>P</td>
<td>P</td>
<td>P</td>
<td>T</td>
</tr>
<tr>
<td>M</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

(b) Token Frame Format

PPP = priority bits
T = token bit
M = monitor bit
RRR = reservation bits

(c) Access Control Field

FF = frame-type bits
ZZZZZ = control bits

(d) Frame Control Field

Figure 4.10 IEEE 802.5 frame format.

• Destination address (DA). As with 802.3 CSMA/CD.

• Source address (SA). As with 802.3 CSMA/CD.

• Data unit. Contains LLC data unit.

• Frame check sequence (FCS). As with 802.3 CSMA/CD.

• End delimiter (ED). Contains the error-detection bit (E), which is set if any repeater detects an error, and the intermediate bit (I), which is used to indicate that this is a frame other than the final one of a multiple-frame transmission.

• Frame status (FS). Contains the address recognized (A) and frame-copied (C) bits, whose use is explained below. Because the A and C bits are outside the scope of the FCS, they are duplicated to provide a redundancy check to detect erroneous settings.

We can now restate the token ring algorithm for the case when a single priority is used. In this case, the priority and reservation bits are set to 0. A station wishing to transmit waits until a token goes by, as indicated by a token bit of 0 in the AC field. The station seizes the token by setting the token bit to 1. The SD and AC fields of the received token now function as the first two fields of the outgoing frame. The station transmits one or more frames, continuing until either its supply of frames is exhausted or a token-
holding timer expires. When the AC field of the last transmitted frame returns, the station sets the token bit to 0 and appends an ED field, resulting in the insertion of a new token on the ring. Stations in the receive mode listen to the ring. Each station can check passing frames for errors and can set the E bit to 1 if an error is detected. If a station detects its own MAC address, it sets the A bit to 1; it may also copy the frame, setting the C bit to 1. This allows the originating station to differentiate three results of a frame transmission:

- Destination station nonexistent or not active ($A = 0, C = 0$)
- Destination station exists but frame not copied ($A = 1, C = 0$)
- Frame received ($A = 1, C = 1$)

### 4.6.2.2 FDDI

The **Fiber Distributed Data Interface (FDDI)** is a token-based LAN standard developed by the American National Standards Institute. FDDI uses a ring topology network in which station interfaces are interconnected by *optical fiber transmission links* operating at 100 Mbps in a ring that spans up to 200 kilometers and accommodates up to 500 stations. FDDI can also operate over twisted-pair cable at lengths of less than 100 meters.

![Figure 4.11 FDDI Token-ring network](image-url)
FDDI can be used in the same way as any of the 802 LANs, as shown in Figure 4.11, but with its high bandwidth it is more often used as a backbone network to interconnect various Ethernet LAN sub-networks as shown in Figure 4.12.

**Optical fiber Taps**

With an optical fiber network, either an active or passive tap can be used. In the case of an active tap (Figure 4.13a), the following steps occur:

1. Optical signal energy enters the tap from the bus.
2. Clocking information is recovered from the signal, and the signal is converted to an electrical signal.
3. The converted signal is presented to the node and perhaps modified by the latter.
4. The optical output (a light beam) is modulated according to the electrical signal and launched into the bus.

In effect, the ring consists of a chain of point-to-point links, and each node acts as a repeater. Each tap actually consists of two of these active couplers and requires two fibers; this is because of the inherently unidirectional nature of the device of Figure 4.13a.
Optical fiber bus taps

In the case of a passive tap (Figure 4.13b) the tap extracts a portion of the optical energy from the bus for reception and it injects optical energy directly into the medium for transmission. Thus, there is a single run of cable rather than a chain point-to-point link. The electronic complexity and interface cost are drawbacks for the implementation of the active tap. Also, each tap will add some increment of delay. For passive taps the lossy nature of pure optical taps limits the number of devices and the length of the medium. However, the performance of such taps has improved sufficiently in recent years so to make fiber bus networks practical.

To provide reliability with respect to link breakages and interface failure, a dual-ring arrangement is used, as shown in Figure 4.14. The transmission on one ring will be clockwise while on the other ring it will be anti-clockwise. A break in the ring is handled by redirecting the flow in the opposite direction at the last station before the break. This action has the effect of converting the dual ring into a single ring. Figure 4.14(a) illustrates a normal FDDI ring while Figure 4.14(b) illustrate the arrangement in
the case of breakage of links at a point.

From Figure 4.15 we can see that the FDDI frame structure is very similar to that of IEEE 802.5 Token Ring frame structure. The frame begins with 16 or more idle control signals that generate a square wave signal that serves to synchronize the receiver. The SD and ED fields contain distinct signal violations that help identify them. The FDDI frame does not contain an AC field or a token bit. Instead the FC field is used to indicate the presence of a token and to provide information about the type of frame. A token frame is indicated by either 10000000 or 11000000 in the FC field. The capture of the token is done, not by flipping a bit, but by removing the token transmission from the ring and replacing it with a data frame. The other remaining fields function as in IEEE 802.5.

\[
\begin{array}{c|c|c|c|c|c|c|c}
\hline
\text{PRE} & \text{SD} & \text{FC} & \text{ED} \\
\hline
\text{PRE} & \text{SD} & \text{FC} & \text{Destination Address} & \text{Source Address} & \text{Information} & \text{FCS} & \text{ED} & \text{FS} \\
\hline
\end{array}
\]

Preamble

Frame Control CLFFZZZZ 
C = Synch/Asynch 
L = Address length (16 or 48 bits) 
FF = LLC/MAC control/reserved frame type

Figure 4.15 FDDI frame structure

4.7 Interconnection of LANs - Bridges

There are several ways of interconnecting networks. When two or more networks interconnected at the physical layer, the type of device is called a \textit{repeater}. When two or more networks are interconnected at the MAC or data link layer, the device is called a \textit{bridge}. When two or more networks are interconnected at the network layer, the device is called a \textit{router}. Interconnection at higher layers is done less frequently. The device that interconnects networks at higher level is usually called a \textit{gateway}. Gateways usually perform some protocol conversion and security functions. In this section we will focus on bridges as they are used to interconnect LANs at MAC layer.

\textbf{Bridges}

When range extension is the only problem, repeaters may solve the problem as long as the maximum distance between two stations is not exceeded. Local area networks (LANs) that involve sharing of media, such as Ethernet and Token ring, can only handle up to some maximum level of traffic. As the number of stations in the LAN increases, or as the traffic generated per station increases, amount of activity in the LAN medium increases until it reaches a saturation point. A typical approach is to segment the user
At times it becomes essential to have multiple LANs. There are several reasons for the use of multiple LANs interconnected:

- **Geography.** Clearly, two separate LANs are needed to support devices clustered in two geographically distant locations. Even in the case of two buildings separated by a highway, it may be far easier to use a microwave bridge link than to attempt to string coaxial cable between the two buildings.

- **Performance.** When range extension is the only problem, repeaters may solve the problem as long as the maximum distance between two stations is not exceeded. Local area networks (LANs) that involve sharing of media, such as Ethernet and Token ring, can only handle up to some maximum level of traffic. As the number of stations in the LAN increases, or as the traffic generated per station increases, amount of activity in the LAN medium increases until it reaches a saturation point. In general, performance on a LAN or MAN declines with an increase in the number of devices or with the length of the medium. A typical approach is to segment the user group into two or, more LANs and to use bridges to interconnect the LANs. LANs will often give improved performance if devices can be clustered so that intra-network traffic significantly exceeds internetwork traffic.

- **Reliability.** The danger in connecting all data processing devices in an organization to one network is that a fault on the network or even a fault in a single device may disable communication for all devices in the network. The network may be required to be partitioned into self-contained units.

- **Security.** LANs originally assume an element of trust between the users in the LAN. As LAN grows, this assumption breaks down, and security concerns become prominent. The fact that most LANs are broadcast in nature implies that eavesdropping can be done easily. This behavior opens the door for the various security threats. Also it is desirable to keep different types of traffic (e.g., accounting, personnel, strategic planning) that have different security needs on physically separate media. At the same time, the different types of users with different levels of security need to communicate through controlled and monitored mechanisms.

The establishment of multiple LANs and the filtering mechanism at the bridges will improve security of communications.
Functions of a Bridge
Figure 4.16 illustrates the operation of a bridge between two LANs, A and B. The bridge performs the following functions:

- Reads all frames transmitted on A, and accepts those addressed to stations on B.
- Using the medium access control protocol for B, retransmits the frames onto B.
- Does the same for B-to-A traffic.

Several design aspects of a bridge are worth highlighting:
1. The bridge makes no modification to the content of the frames it receives.
2. The bridge should contain enough buffer space to meet peak demands. Over a short period of time, frames may arrive faster than they can be retransmitted.
3. The bridge must contain addressing and routing intelligence. At a minimum, the bridge must know which addresses are on each network in order to know which frames to pass. Further, there may be more than two LANs interconnected by a number of bridges. In that case, a frame may have to be routed through several bridges in its journey from source to destination.
4. A bridge may connect more than two LANs. The bridge provides an extension to the LAN that requires no modification to the communications software in the stations attached to the LANs. It appears to all stations on the two (or more) LANs that there is a single LAN on which each station has a unique address. The station uses that unique address and need not explicitly discriminate between stations on the same LAN and stations on other LANs; the bridge takes care of that.

The description above has applied to the simplest sort of bridge. More sophisticated
bridges can be used in more complex collections of LANs. These constructions would include additional functions, such as,

- Each bridge can maintain status information on other bridges, together with the cost and number of bridge-to-bridge hops required to reach each network. This information may be updated by periodic exchanges of information among bridges; this allows the bridges to perform a dynamic routing function.
- A control mechanism can manage frame buffers in each bridge to overcome congestion. Under saturation conditions, the bridge can give precedence to en route packets over new packets just entering the internet from an attached LAN, thus preserving the investment in line bandwidth and processing time already made in the en route frame.

**Bridge Protocol Architecture**

The IEEE 802.1D specification defines the protocol architecture for MAC bridges. In addition, the standard suggests formats for a globally administered set of MAC station addresses across multiple homogeneous LANs. In this subsection, we examine the protocol architecture of these bridges.

Within the 802 architecture, the endpoint or station address is designated at the MAC level. Thus, it is at the MAC level that a bridge can function. Figure 4.17 shows the simplest case, which consists of two LANs connected by a single bridge.

![Figure 4.17 Operation of a LAN Bridge from 802.3 to 802.4](image)

The LANs employ the same MAC and LLC protocols. The bridge operates as previously described. A MAC frame whose destination is not on the immediate LAN is captured by the bridge, buffered briefly, and then transmitted on the other LAN. As far as the LLC layer is concerned, there is a dialogue between peer LLC entities in the two endpoint stations. The bridge need not contain an LLC layer, as it is merely serving to relay the MAC frames.
Exercise

1) Discuss Switched Networks and Broadcast Networks.
2) What is a LAN? Mention the key technology ingredients that determine the nature of a LAN.
3) What are the functions of MAC layer in a LAN?
4) List the standards issued by IEEE 802 committee for LAN and MAN. Also mention what topology and transmission medium each of them relies on.
5) List the various network topologies. Discuss one of them
6) What protocol controls access to the transmission media in a LAN? Discuss key parameters in medium access control mechanisms.
7) Discuss different asynchronous medium access control mechanisms.
8) Discuss MAC frame format.
9) What are the types of services provided by LLC layer? Discuss them.
10) Discuss the development of CSMA/CD technology.
11) Describe how frames are transmitted in CSMA/CD LAN.
12) Discuss various fields in the CSMA/CD MAC frame format.
13) Discuss Token ring MAC protocol. What is the principal disadvantage of the token ring.
14) Discuss various fields in Token ring MAC frame format.
15) What is FDDI, in practice how it is used? Discuss different types of optical taps.
16) What are the devices available to interconnect LANs? How do they differ?
17) Discuss the reasons for having interconnected multiple LANs instead of one single LAN.
18) List the design aspects of a bridge.