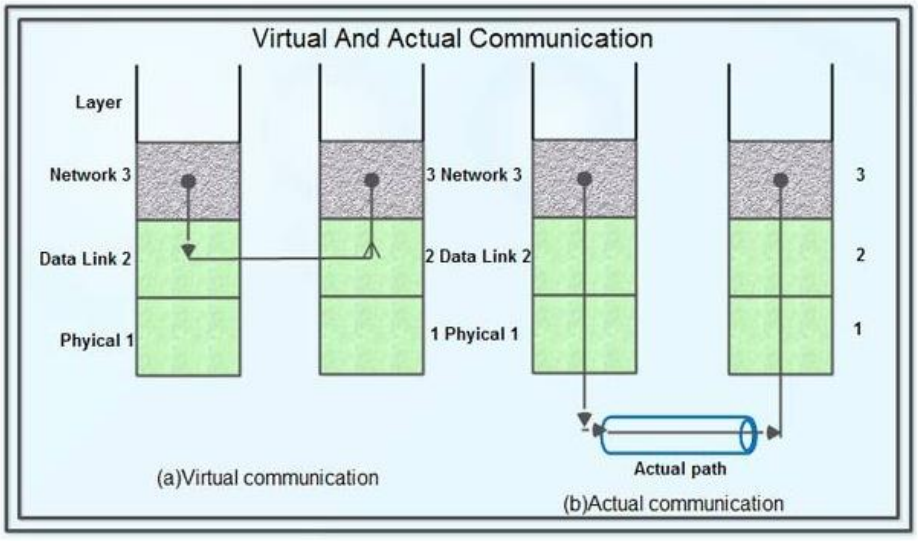
Unit 3: The Data Link Layer

Functions of Data Link Layer:

- The data link layer transforms the physical layer, a raw transmission facility, to a link responsible for node-to-node (hop-to-hop) communication.
- Specific responsibilities of the data link layer include framing, addressing, flow control, error control, and media access control.
- The data link layer divides the stream of bits received from the network layer into manageable data units called frames. The data link layer adds a header to the frame to define the addresses of the sender and receiver of the frame.
- If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- The data link layer also adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged, duplicate, or lost frames.
- When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

Services Provided to Network Layer:

- Data link layer provides several services to the network layer. The one of the major services provided is the transferring the data from network layer on the source machine to the network layer on destination machine.
- On source machine data link layer receives the data from network layer and on destination machine pass on this data to the network layer as shown in Figure. The path shown in fig (a) is the virtual path.
- But the actual path is Network layer -> Data link layer -> Physical layer on source machine, then to physical media and thereafter physical layer -> Data link layer -> Network layer on destination machine.

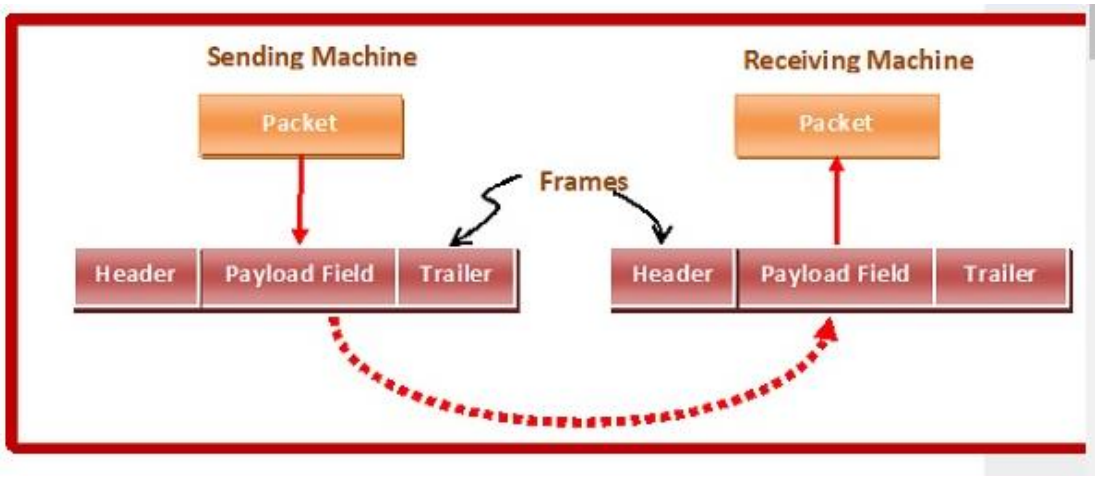


The three major types of services offered by data link layer are:

1. Unacknowledged connectionless service.
2. Acknowledged connectionless service.
3. Acknowledged connection-oriented service.

Framing:

- In the physical layer, data transmission involves synchronized transmission of bits from the source to the destination. The data link layer packs these bits into frames.
- Data-link layer takes the packets from the Network Layer and encapsulates them into frames.
- If the frame size becomes too large, then the packet may be divided into small sized frames.
- Smaller sized frames make flow control and error control more efficient. Then, it sends each frame bit-by-bit on the hardware.
- At receiver's end, data link layer picks up signals from hardware and assembles them into frames.



Parts of a Frame

A frame has the following parts −

Frame Header − It contains the source and the destination addresses of the frame.

Payload field − It contains the message to be delivered.

Trailer − It contains the error detection and error correction bits.

Flag − It marks the beginning and end of the frame.

Types of Framing

Framing can be of two types, fixed sized framing and variable sized framing.

Fixed-sized Framing

Here the size of the frame is fixed and so the frame length acts as delimiter of the frame. Consequently, it does not require additional boundary bits to identify the start and end of the frame.

Example − ATM cells.

Variable – Sized Framing

Here, the size of each frame to be transmitted may be different. So additional mechanisms are kept to mark the end of one frame and the beginning of the next frame.

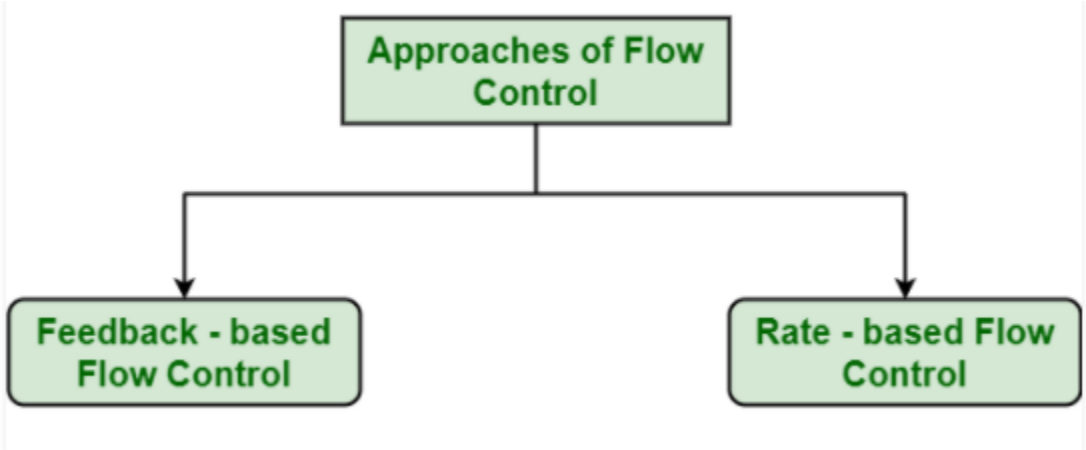
It is used in local area networks.

Flow Control:

- Data link layer protocols are mainly responsible for flow control. When a data frame is sent from one host to another over a single medium, it is required that the sender and receiver should work at the same speed.
- That is, sender sends at a speed on which the receiver can process and accept the data. If sender is sending too fast, the receiver may be overloaded and data may be lost.
- Flow control is basically technique that gives permission to two of stations that are working and processing at different speeds to just communicate with one another.
- Flow control in Data Link Layer simply restricts and coordinates number of frames or amount of data sender can send just before it waits for an acknowledgment from receiver.
- Flow control is actually set of procedures that explains sender about how much data or frames it can transfer or transmit before data overwhelms receiver.

Approaches to Flow Control:

Flow Control is classified into two categories:



Feedback-based Flow Control:

In these protocols, the sender sends frames after it has received acknowledgments from the user. This is used in the data link layer.

Rate-based Flow Control:

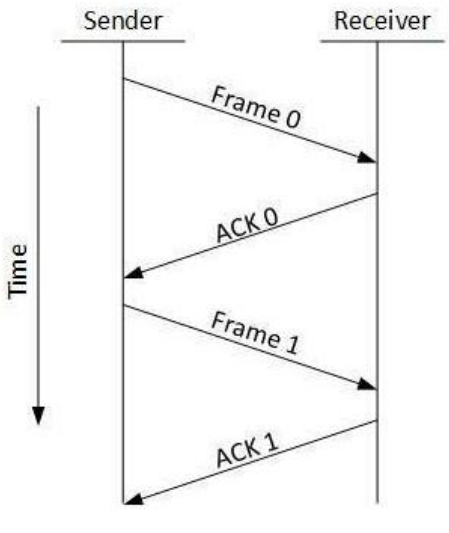
These protocols have built in mechanisms to restrict the rate of transmission of data without requiring acknowledgment from the receiver. This is used in the transport layer.

Two types of mechanisms can be deployed to control the flow based on the feedback:

- A simple stop and wait Protocol
- Sliding Window Protocol

Simplex Stop and Wait

- This flow control mechanism forces the sender after transmitting a data frame to stop and wait until the acknowledgement of the data-frame sent is received.
- The sender sends the next frame only when it has received a positive acknowledgement from the receiver that it is available for further data processing.
- Data transmission is one directional, but must have bidirectional line.



Sliding Window

- In this flow control mechanism, both sender and receiver agree on the number of data-frames after which the acknowledgement should be sent.
- As we learnt, stop and wait flow control mechanism wastes resources, this protocol tries to make use of underlying resources as much as possible.
- In this protocol, multiple frames can be sent by a sender at a time before receiving an acknowledgment from the receiver.
- This protocol improves the efficiency of stop and wait protocol by allowing multiple frames to be transmitted before receiving an acknowledgment.
- The working principle of this protocol can be described as follows −
- Both the sender and the receiver have finite sized buffers called windows. The sender and the receiver agree upon the number of frames to be sent based upon the buffer size.
- The sender sends multiple frames in a sequence, without waiting for acknowledgment. When its sending window is filled, it waits for acknowledgment. On receiving acknowledgment, it advances the window and transmits the next frames, according to the number of acknowledgments received.

Error Control:

- Error control in data link layer is the process of detecting and correcting data frames that have been corrupted or lost during transmission.
- In case of lost or corrupted frames, the receiver does not receive the correct data-frame and sender is ignorant about the loss.
- Data link layer follows a technique to detect transit errors and take necessary actions, which is retransmission of frames whenever error is detected or frame is lost. The process is called Automatic Repeat Request (ARQ).

Phases in Error Control

The error control mechanism in data link layer involves the following phases −

- Detection of Error: Transmission error, if any, is detected by either the sender or the receiver.
- Acknowledgment: acknowledgment may be positive or negative.
    - Positive ACK − On receiving a correct frame, the receiver sends a positive acknowledge.
    - Negative ACK − On receiving a damaged frame or a duplicate frame, the receiver sends a negative acknowledgment back to the sender.
- Retransmission: The sender maintains a clock and sets a timeout period. If an acknowledgment of a data-frame previously transmitted does not arrive before the timeout, or a negative acknowledgment is received, the sender retransmits the frame.

Error Control Techniques

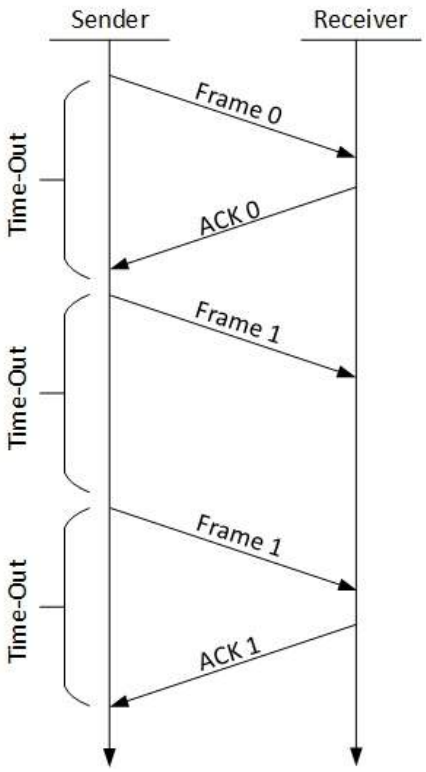
Error Control Techniques in data link layer are:

- Stop and Wait ARQ
- Sliding Window ARQ
    - Go-Back-N ARQ
    - Selective Repeat ARQ

Stop and Wait ARQ:

The following transition may occur in Stop-and-Wait ARQ:

- The sender maintains a timeout counter.
- When a frame is sent, the sender starts the timeout counter.
- If acknowledgement of frame comes in time, the sender transmits the next frame in queue.
- If acknowledgement does not come in time, the sender assumes that either the frame or its acknowledgement is lost in transit. Sender retransmits the frame and starts the timeout counter.
- If a negative acknowledgement is received, the sender retransmits the frame.



Sliding Window ARQ:

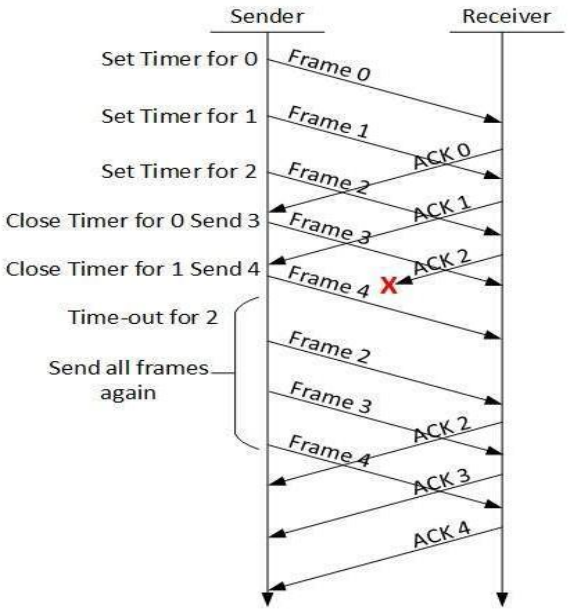
This technique is generally used for continuous transmission error control. It is further categorized into two categories as given below:

Go-Back-N ARQ:

In this protocol, we can send several frames before receiving acknowledgements; we keep a copy of these frames until the acknowledgements arrive. Stop and wait mechanism does not utilize the resources at their best. When the acknowledgement is received, the sender sits idle and does nothing. In Go-Back-N method, both sender and receiver maintain a window.

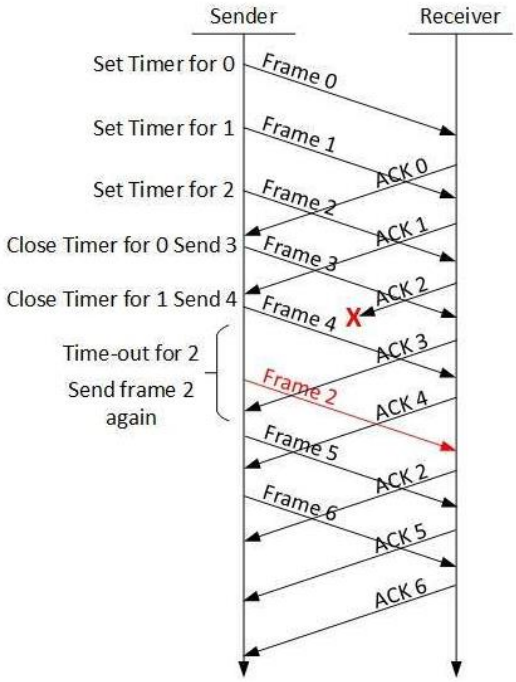
- The sending-window size enables the sender to send multiple frames without receiving the acknowledgement of the previous ones.
- The receiving-window enables the receiver to receive multiple frames and acknowledge them. The receiver keeps track of incoming frame's sequence number.

When the sender sends all the frames in window, it checks up to what sequence number it has received positive acknowledgement. If all frames are positively acknowledged, the sender sends next set of frames. If sender finds that it has received NACK or has not receive any ACK for a particular frame, it retransmits all the frames after which it does not receive any positive ACK.



Selective Repeat ARQ:

- In Go-back-N ARQ, it is assumed that the receiver does not have any buffer space for its window size and has to process each frame as it comes.
- This enforces the sender to retransmit all the frames which are not acknowledged.
- In Selective-Repeat, the receiver while keeping track of sequence numbers, buffers the frames in memory and sends NACK for only frame which is missing or damaged.
- The sender in this case, sends only packet for which NACK is received.



Error Detection and Correction Techniques:

- Networks must be able to transfer data from one device to another with acceptable accuracy.
- For most applications, a system must guarantee that the data received are identical to the data transmitted.
- Any time data are transmitted from one node to the next, they can become corrupted in passage. Many factors can alter one or more bits of a message.
- Some applications require a mechanism for detecting and correcting errors.
- Some applications can tolerate a small level of error. For example, random errors in audio or video transmissions may be tolerable, but when we transfer text, we expect a very high level of accuracy.

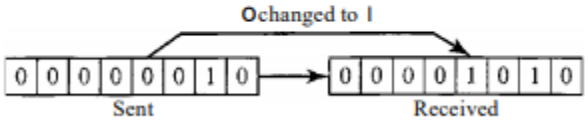
Types of Errors:

- Whenever bits flow from one point to another, they are subject to unpredictable changes because of interference. This interference can change the shape of the signal.
- In a single-bit error, a 0 is changed to a 1 or a 1 to a 0.
- In a burst error, multiple bits are changed.

Single-Bit Error:

- The term single-bit error means that only 1 bit of a given data unit (such as a byte, character, or packet) is changed from 1 to 0 or from 0 to 1.

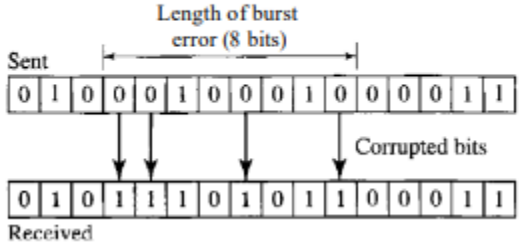
Single-bit error



Burst Error:

- The term burst error means that 2 or more bits in the data unit have changed from 1 to 0 or from 0 to 1.

Burst error of length 8



- A burst error is more likely to occur than a single-bit error.

- The duration of noise is normally longer than the duration of 1 bit, which means that when noise affects data, it affects a set of bits.
- The number of bits affected depends on the data rate and duration of noise.

Redundancy

- The central concept in detecting or correcting errors is redundancy. To be able to detect or correct errors, we need to send some extra bits with our data.
- These redundant bits are added by the sender and removed by the receiver. Their presence allows the receiver to detect or correct corrupted bits.

Detection Versus Correction

- The correction of errors is more difficult than the detection.
- In error detection, we are looking only to see if any error has occurred.
- The answer is a simple yes or no. We are not even interested in the number of errors. A single-bit error is the same for us as a burst error.
- In error correction, we need to know the exact number of bits that are corrupted and more importantly, their location in the message.
- The number of the errors and the size of the message are important factors.
- If we need to correct one single error in an 8-bit data unit, we need to consider eight possible error locations; if we need to correct two errors in a data unit of the same size, we need to consider 28 possibilities.

Forward Error Correction Versus Retransmission

- There are two main methods of error correction. Forward error correction is the process in which the receiver tries to guess the message by using redundant bits. This is possible if the number of errors is small.
- Correction by retransmission is a technique in which the receiver detects the occurrence of an error and asks the sender to resend the message. Resending is repeated until a message arrives that the receiver believes is error-free.

Error Detecting Codes:

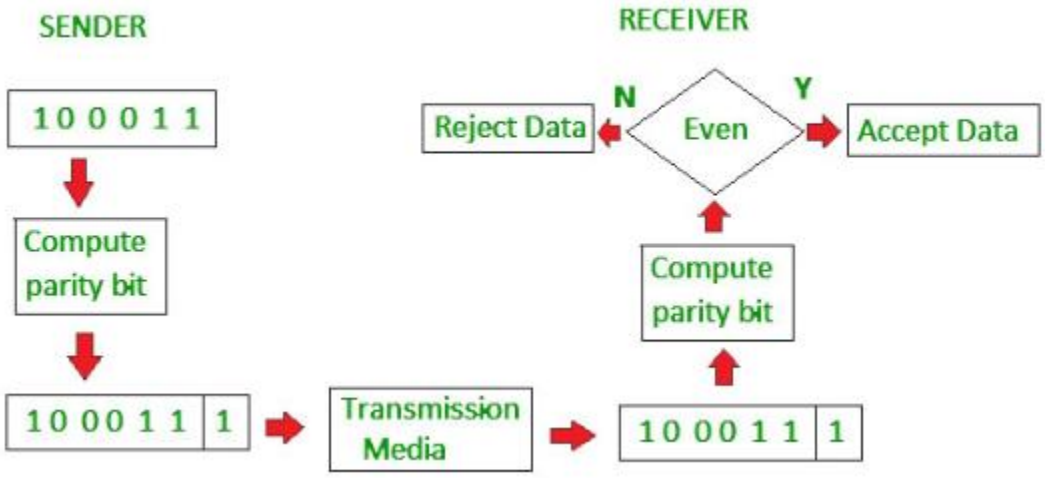
- Whenever a message is transmitted, it may get scrambled by noise or data may get corrupted.
- To avoid this, we use error-detecting codes which are additional data added to a given digital message to help us detect if any error has occurred during transmission of the message.

Some popular techniques for error detection are:

- Parity
- Checksum
- Cyclic redundancy check

Parity check

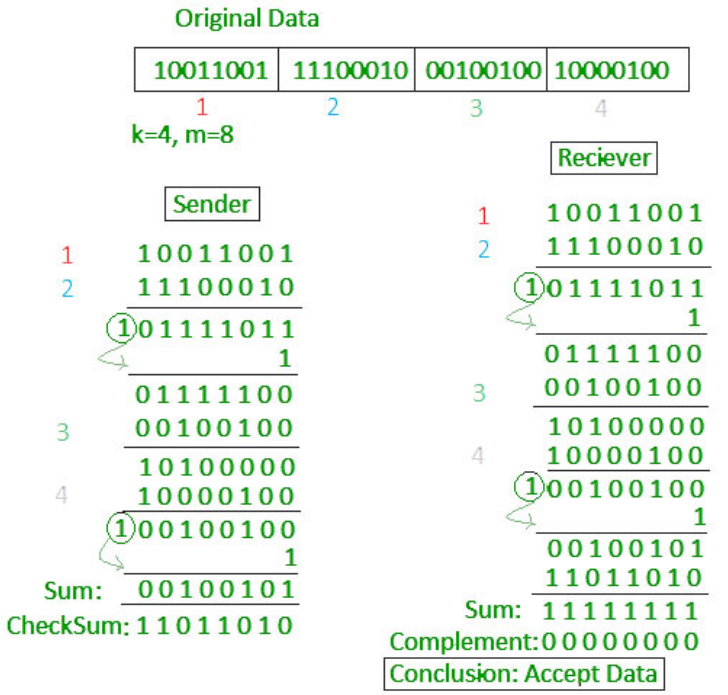
- The most common and least expensive mechanism for error- detection is the parity check. The parity check is done by adding an extra bit, called parity bit to the data to make a number of 1s either even in case of even parity or odd in case of odd parity. While creating a frame, the sender counts the number of 1s in it and adds the parity bit in the following way:
- In case of even parity: If a number of 1s is even then parity bit value is 0. If the number of 1s is odd then parity bit value is 1.
- In case of odd parity: If a number of 1s is odd then parity bit value is 0. If a number of 1s is even then parity bit value is 1.
- On receiving a frame, the receiver counts the number of 1s in it. In case of even parity check, if the count of 1s is even, the frame is accepted, otherwise, it is rejected. A similar rule is adopted for odd parity check.
- The parity check is suitable for single bit error detection only.



Checksum

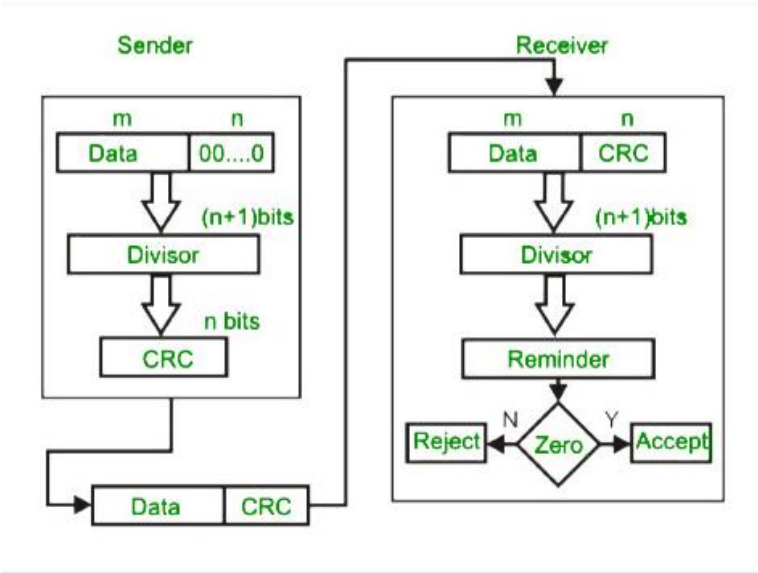
In this error detection scheme, the following procedure is applied:

- Data is divided into fixed sized frames or segments. (k segments each of m bits)
- The sender adds the segments using 1's complement arithmetic to get the sum. It then complements the sum to get the checksum and sends it along with the data frames.
- The receiver adds the incoming segments along with the checksum using 1's complement arithmetic to get the sum and then complements it.
- If the result is zero, the received frames are accepted; otherwise, they are discarded.

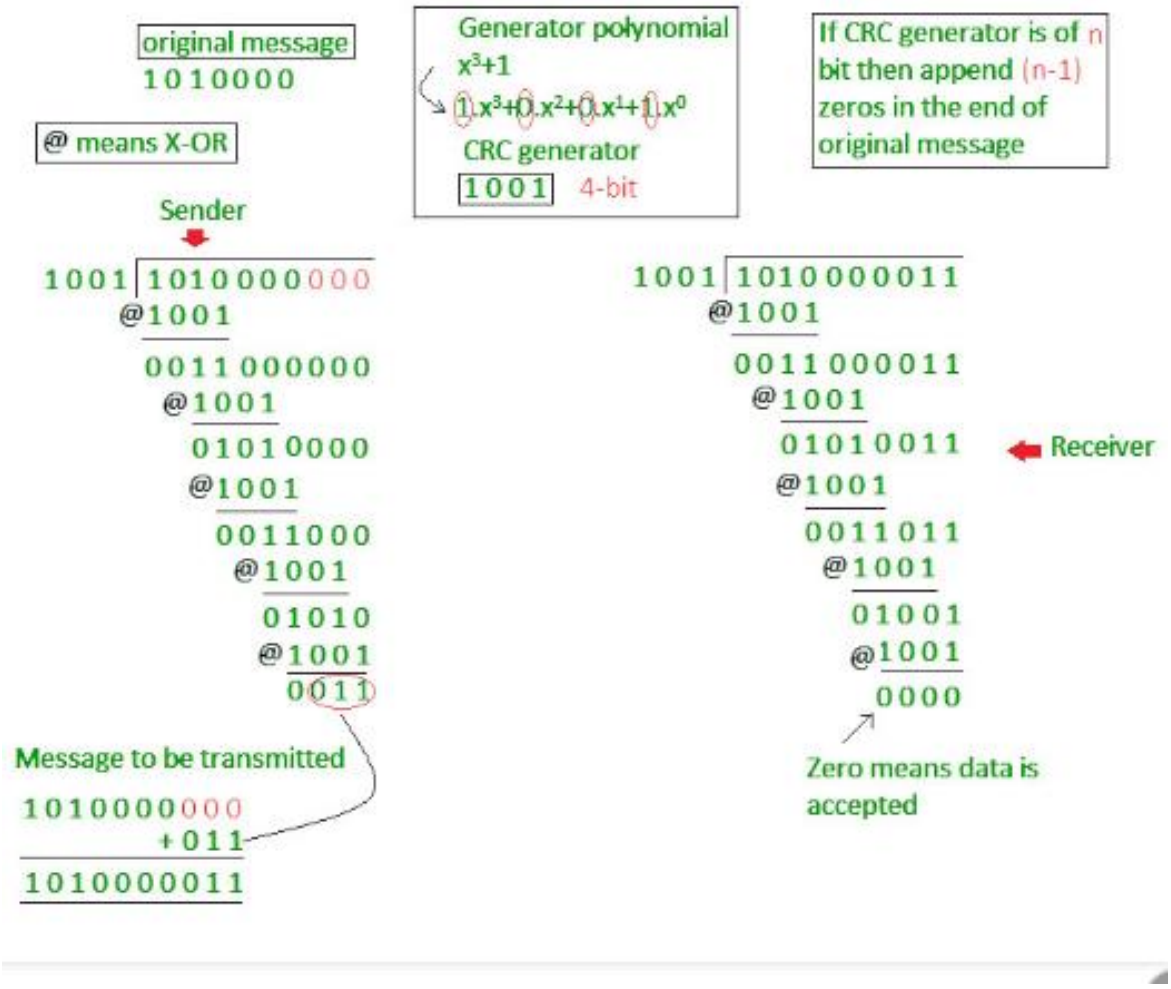


Cyclic Redundancy Check:

A cyclic redundancy check (CRC) is an error-detecting code commonly used in digital networks and storage devices to detect accidental changes to raw data. Unlike checksum scheme, which is based on addition, CRC is based on binary division. In CRC, a sequence of redundant bits, called cyclic redundancy check bits, are appended to the end of data unit so that the resulting data unit becomes exactly divisible by a second, predetermined binary number. At the destination, the incoming data unit is divided by the same number. If at this step there is no remainder, the data unit is assumed to be correct and is therefore accepted. A remainder indicates that the data unit has been damaged in transit and therefore must be rejected.



- At the sender side, the data unit to be transmitted is divided by a predetermined divisor (binary number) in order to obtain the remainder. This remainder is called CRC.
- The CRC has one bit less than the divisor. It means that if CRC is of n bits, divisor is of n+ 1 bit.
- The sender appends this CRC to the end of data unit such that the resulting data unit becomes exactly divisible by predetermined divisor i.e., remainder becomes zero.
- At the destination, the incoming data unit i.e., data + CRC is divided by the same number (predetermined binary divisor).
- If the remainder after division is zero then there is no error in the data unit & receiver accepts it.
- If remainder after division is not zero, it indicates that the data unit has been damaged in transit and therefore it is rejected.
- This technique is more powerful than the parity check and checksum error detection.



High Level Data Link Control (HDLC):

High-level Data Link Control (HDLC) is a group of communication protocols of the data link layer for transmitting data between network points or nodes. Since it is a data link protocol, data is organized into

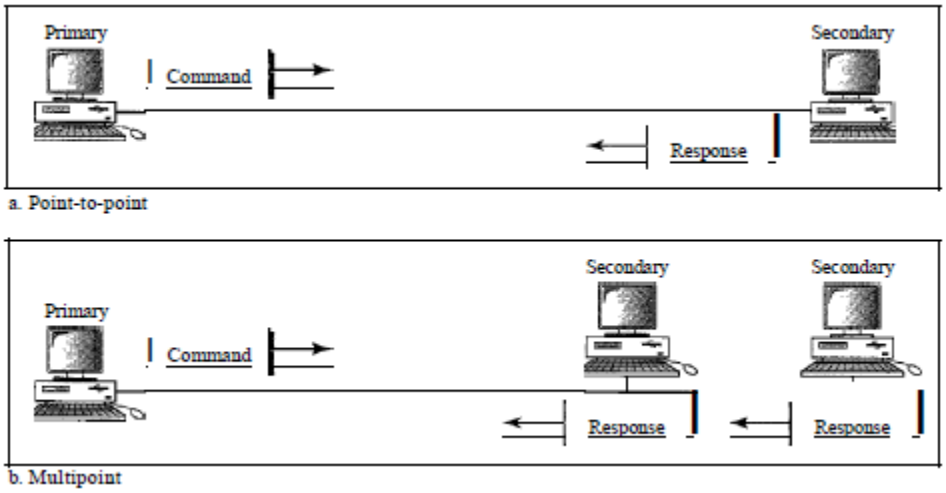
frames. A frame is transmitted via the network to the destination that verifies its successful arrival. It is a bit - oriented protocol that is applicable for both point - to - point and multipoint communications.

Configurations and Transfer Modes

HDLC provides two common transfer modes that can be used in different configurations: normal response mode (NRM) and asynchronous balanced mode (ABM).

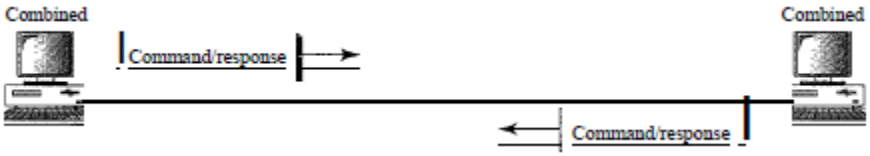
Normal Response Mode

- In normal response mode (NRM), the station configuration is unbalanced. We have one primary station and multiple secondary stations.
- A primary station can send commands; a secondary station can only respond. The NRM is used for both point-to-point and multiple-point links.



Asynchronous Balanced Mode

- In asynchronous balanced mode (ABM), the configuration is balanced.
- The link is point-to-point, and each station can function as a primary and a secondary (acting as peers). This is the common mode today.

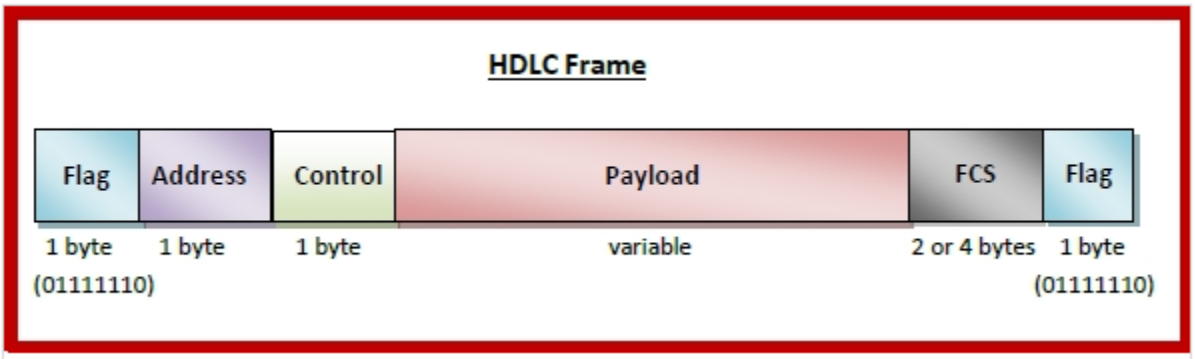


HDLC Frame

HDLC is a bit - oriented protocol where each frame contains up to six fields. The structure varies according to the type of frame. The fields of a HDLC frame are –

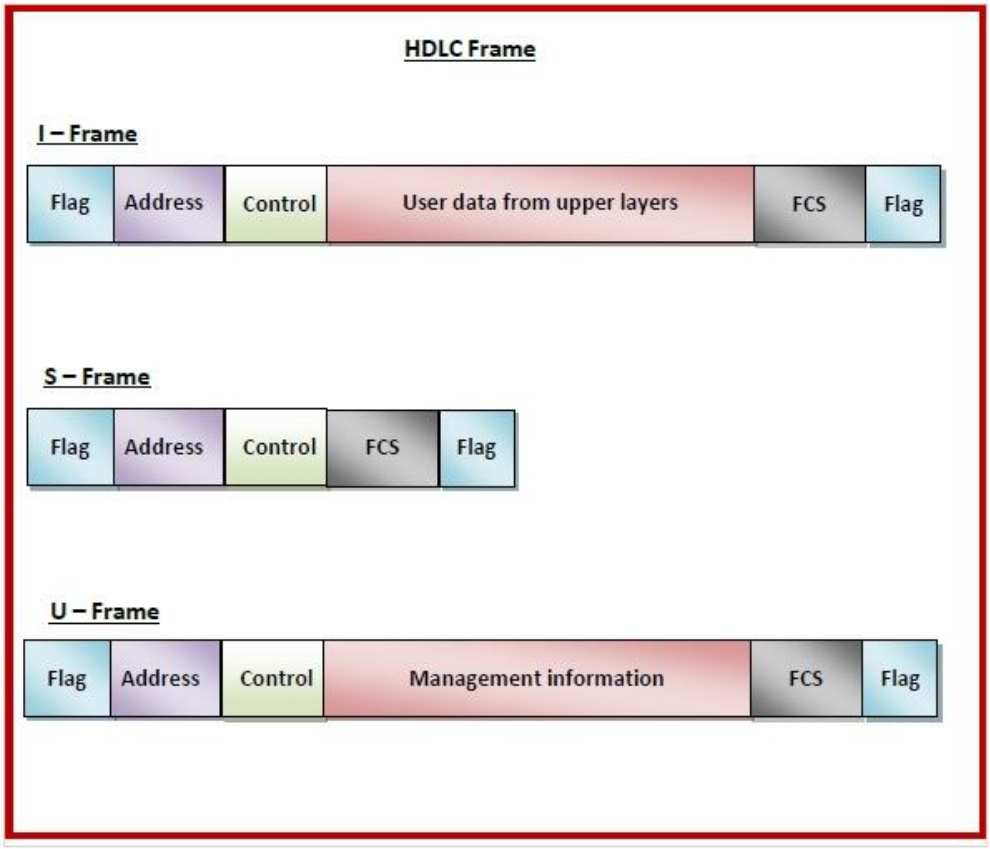
- Flag – It is an 8-bit sequence that marks the beginning and the end of the frame. The bit pattern of the flag is 01111110.

- Address − It contains the address of the receiver. If the frame is sent by the primary station, it contains the address(es) of the secondary station(s). If it is sent by the secondary station, it contains the address of the primary station. The address field may be from 1 byte to several bytes.
- Control − It is 1- or 2-bytes containing flow and error control information.
- Payload − This carries the data from the network layer. Its length may vary from one network to another.
- FCS − It is a 2 byte or 4 bytes frame check sequence for error detection. The standard code used is CRC (cyclic redundancy code)



Types of HDLC Frames

There are three types of HDLC frames. The type of frame is determined by the control field of the frame:



- **I-frame** – I-frames or Information frames carry user data from the network layer. They also include flow and error control information that is piggybacked on user data. The first bit of control field of I-frame is 0.
- **S-frame** – S-frames or Supervisory frames do not contain information field. They are used for flow and error control when piggybacking is not required. The first two bits of control field of S-frame is 10.
- **U-frame** – U-frames or Un-numbered frames are used for myriad miscellaneous functions, like link management. It may contain an information field, if required. The first two bits of control field of U-frame is 11.

Point-to-Point Protocol (PPP):

- Although HDLC is a general protocol that can be used for both point-to-point and multipoint configurations, one of the most common protocols for point-to-point access is the Point-to-Point Protocol (PPP).
- Today, millions of Internet users who need to connect their home computers to the server of an Internet service provider use PPP.
- The majority of these users have a traditional modem; they are connected to the Internet through a telephone line, which provides the services of the physical layer.
- But to control and manage the transfer of data, there is a need for a point-to-point protocol at the data link layer. PPP is by far the most common.

PPP provides several services:

- PPP defines the format of the frame to be exchanged between devices.
- PPP defines how two devices can negotiate the establishment of the link and the exchange of data.
- PPP defines how network layer data are encapsulated in the data link frame.
- PPP defines how two devices can authenticate each other.
- PPP provides multiple network layer services supporting a variety of network layer protocols.
- PPP provides connections over multiple links.
- PPP provides network address configuration. This is particularly useful when a home user needs a temporary network address to connect to the Internet.

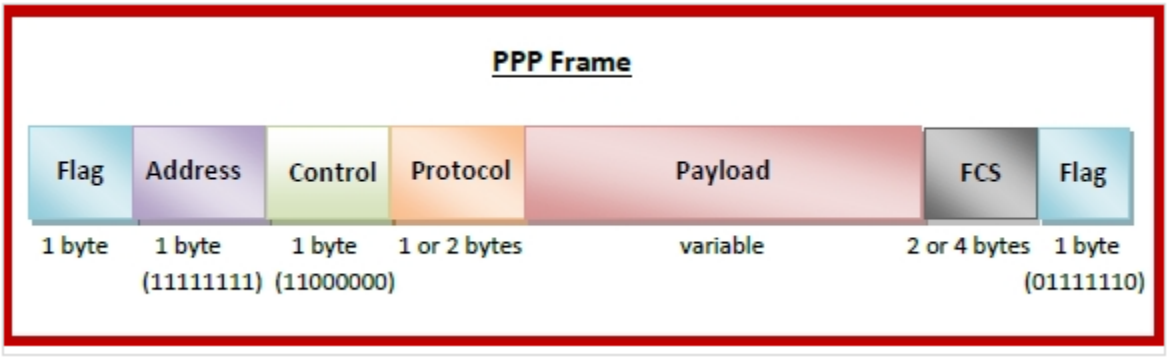
On the other hand, to keep PPP simple, several services are missing:

- PPP does not provide flow control. A sender can send several frames one after another with no concern about overwhelming the receiver.
- PPP has a very simple mechanism for error control. A CRC field is used to detect errors. If the frame is corrupted, it is silently discarded; the upper-layer protocol needs to take care of the problem. Lack of error control and sequence numbering may cause a packet to be received out of order.
- PPP does not provide a sophisticated addressing mechanism to handle frames in a multipoint configuration.

PPP Frame

PPP is a byte - oriented protocol where each field of the frame is composed of one or more bytes. The fields of a PPP frame are:

- Flag – 1 byte that marks the beginning and the end of the frame. The bit pattern of the flag is 01111110.
- Address – 1 byte which is set to 11111111 in case of broadcast.
- Control – 1 byte set to a constant value of 11000000.
- Protocol – 1 or 2 bytes that define the type of data contained in the payload field.
- Payload – This carries the data from the network layer. The maximum length of the payload field is 1500 bytes. However, this may be negotiated between the endpoints of communication.
- FCS – It is a 2 byte or 4 bytes frame check sequence for error detection. The standard code used is CRC (cyclic redundancy code)



Byte Stuffing in PPP Frame:

- Byte stuffing is used is PPP payload field whenever the flag sequence appears in the message, so that the receiver does not consider it as the end of the frame.
- The escape byte, 01111101, is stuffed before every byte that contains the same byte as the flag byte or the escape byte.
- The receiver on receiving the message removes the escape byte before passing it onto the network layer.

Channel Allocation Problem:

- When there are more than one user who desire to access a shared network channel, an algorithm is deployed for channel allocation among the competing users.
- The network channel may be a single cable or optical fiber connecting multiple nodes, or a portion of the wireless spectrum.
- Channel allocation algorithms allocate the wired channels and bandwidths to the users, who may be base stations, access points or terminal equipment.

Channel Allocation Schemes

Channel Allocation may be done using two schemes:

- Static Channel Allocation
- Dynamic Channel Allocation

Static Channel Allocation

- In static channel allocation scheme, a fixed portion of the frequency channel is allotted to each user.
- For N competing users, the bandwidth is divided into N channels using frequency division multiplexing (FDM), and each portion is assigned to one user.
- This scheme is also referred as fixed channel allocation or fixed channel assignment.
- In this allocation scheme, there is no interference between the users since each user is assigned a fixed channel.
- However, it is not suitable in case of a large number of users with variable bandwidth requirements.

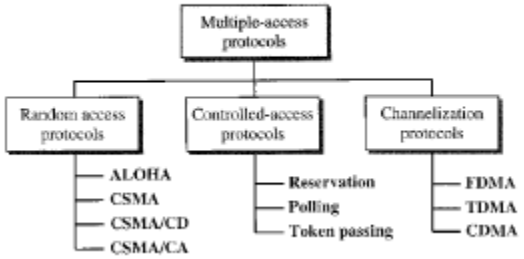
Dynamic Channel Allocation

- In dynamic channel allocation scheme, frequency bands are not permanently assigned to the users.
- Instead, channels are allotted to users dynamically as needed, from a central pool.
- The allocation is done considering a number of parameters so that transmission interference is minimized.
- This allocation scheme optimizes bandwidth usage and results is faster transmissions.
- Dynamic channel allocation is further divided into centralized and distributed allocation.

Multiple Access Protocols:

- When nodes or stations are connected and use a common link, called a multipoint or broadcast link, we need a multiple access protocol to coordinate access to the link.
- Many formal protocols have been devised to handle access to a shared link. We categorize them into three groups.

Taxonomy of multiple-access protocols discussed in this chapter



Random access:

- In random access or contention methods, no station is superior to another station and none is assigned the control over another.
- No station permits or does not permit another station to send.
- At each instance, a station that has to send uses a procedure defined by the protocol to make a decision on whether or not to send.
- This decision depends on the state of the medium (idle or busy). In other words, each station can transmit when it desires on testing of the state of the medium.

ALOHA:

- ALOHA is the earliest random-access method developed for wireless LAN but can be used on any shared medium.
- In this, multiple stations can transmit data at the same time and can hence lead to collision. The data from the two stations collide.

Pure ALOHA:

- The idea behind this protocol is that each station sends a frame whenever it has a frame to send.
- However, since there is only one channel to share, there is possibility of collision between frames from different stations.
- Even if one bit of a frame coexists on the channel with one bit from another frame, there is a collision and both will be destroyed.

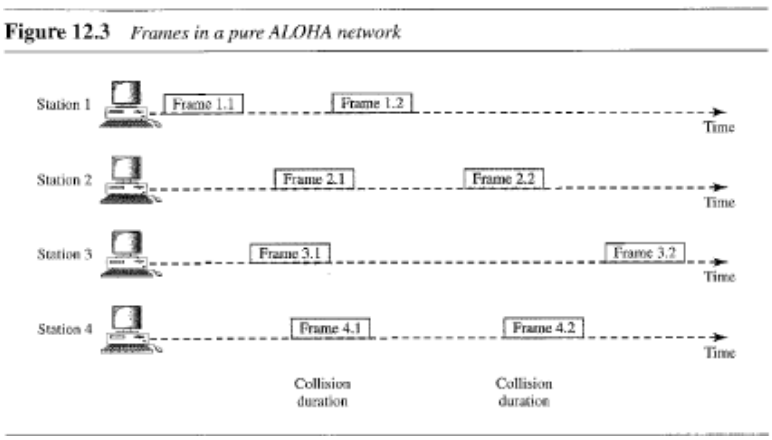


Figure 12.3 *Frames in a pure ALOHA network*

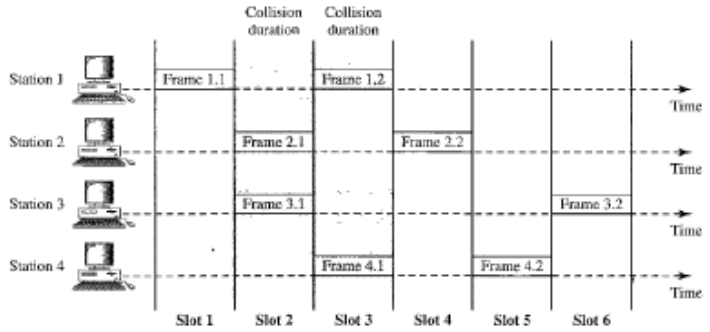
- The pure ALOHA protocol relies on acknowledgements from the receiver. When a station sends a frame, it expects the receiver to send an acknowledgement.
- If the acknowledgement does not arrive after the time out period, the station assumes that the frame (or the acknowledgement) has been destroyed and resends the frame.
- A collision involves two or more stations. If all these stations try to resend their frames after the time out period passes, each station waits a random amount of time before resending its frame.
- The randomness will help avoid more collisions, called back-off time.

- Since different stations may wait for different amount of time, the probability of further collision decreases.

Slotted ALOHA:

- In pure ALOHA, there is no rule that defines when the station can send.
- A station may send soon after another station has started or soon before another station has finished. So, still the collision may occur.
- Slotted ALOHA is similar to pure ALOHA, except that we divide time into slots and sending of data is allowed only at the beginning of these slots.
- If a station misses out the allowed time, it must wait for the next slot. This reduces the probability of collision.

Figure 12.6 *Frames in a slotted ALOHA network*



- Allowing a station to send only at the beginning of the time slot means that the station sending in the previous slot has finished sending its frame already.
- However, there is still possibility of collision if two stations try to send at the beginning of the same time slot.

Carrier Sense Multiple Access (CSMA):

- The chance of collision can be reduced if a station senses the medium before trying to use it.
- Carrier Sense Multiple Access ensures fewer collisions as the station is required to first sense the medium (for idle or busy) before transmitting data.
- If it is idle then it sends data, otherwise it waits till the channel becomes idle. (Listen before talk)
- However, there is still chance of collision in CSMA due to propagation delay. For example, if station A wants to send data, it will first sense the medium.
- If it finds the channel idle, it will start sending data. However, by the time the first bit of data is transmitted (delayed due to propagation delay) from station A, if station B requests to send data and senses the medium it will also find it idle and will also send data.
- This will result in collision of data from station A and B.

CSMA access modes-

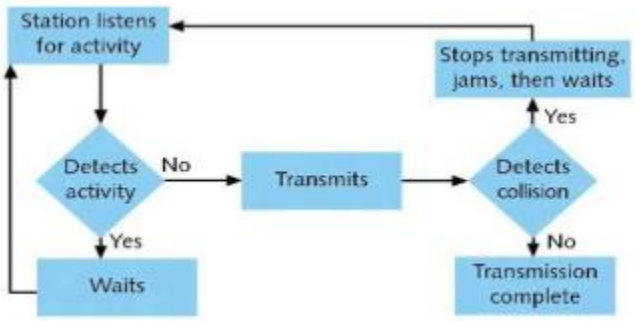
- 1-persistent: The node senses the channel, if idle it sends the data, otherwise it continuously keeps on checking the medium for being idle and transmits unconditionally (with 1 probability) as soon as the channel gets idle.

- Non-Persistent: The node senses the channel, if idle it sends the data, otherwise it checks the medium after a random amount of time (not continuously) and transmits when found idle.
- P-persistent: The node senses the medium, if idle it sends the data with p probability. If the data is not transmitted ((1-p) probability) then it waits for some time and checks the medium again, now if it is found idle then it sends with p probability. This repeat continues until the frame is sent. It is used in Wi-Fi and packet radio systems.
- O-persistent: Superiority of nodes is decided beforehand and transmission occurs in that order. If the medium is idle, node waits for its time slot to send data.

Carrier sense multiple access with collision detection (CSMA/CD):

- The CSMA method does not specify the procedure following a collision.
- In Carrier sense multiple access with collision detection method, a station monitors the medium after it sends a frame to see if the transmission was successful. If so, the transmission is completed. However, if there is a collision, the frame is sent again.

The basic idea behind CSMA/CD is that a station needs to be able to receive while transmitting, to detect a collision. When there is no collision, the station receives one signal; its own signal. When there is a collision, the station receives two signals: its own signal and the signal transmitted by a second station. To distinguish between these two cases, the received signals in these two cases must be significantly different. In other words, the signal from the second station needs to add a significant amount of energy to the one created by the first station.

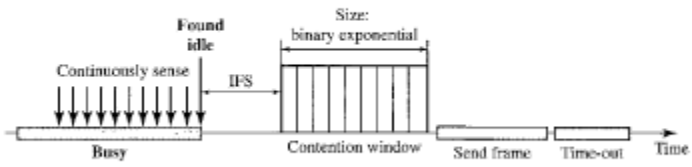


Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA):

- The process of collision detection involves sender receiving acknowledgement signals.
- If there is just one signal (its own), then the data is successfully sent but if there are two signals (its own and the one with which it has collided), then it means a collision has occurred.
- To distinguish between these two cases, collision must have a lot of impact on received signal. The second signal adds significant amount of energy to the first signal.
- However, this applies only to the wired networks since the received signal has almost the same energy as the sent signal.
- In wireless networks, much of the sent energy is lost in transmission. The received signal has very little energy.

- Therefore, a collision may add only 5 to 10 percent additional energy. This is not useful for effective collision detection.
- We need to avoid collisions on wireless networks because they cannot be detected. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) was invented for this network.
- In contrast to the Carrier Sense Multiple Access/Collision Detect (CSMA/CD) protocol, which handles transmissions only after a collision has taken place, CSMA/CA works to avoid collisions prior to their occurrence.
- Collisions are avoided through the use of CSMA/CA's three strategies as shown in figure below.

Figure 12.16 *Timing in CSMA/CA*



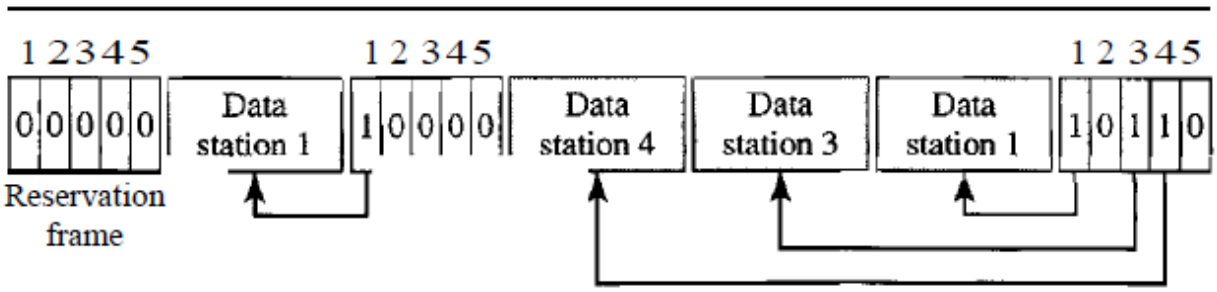
- Interframe space – Station waits for medium to become idle and if found idle it does not immediately send data (to avoid collision due to propagation delay) rather it waits for a period of time called Interframe space or IFS. After this time, it again checks the medium for being idle. IFS can also be used to define the priority of a station or a frame. Higher the IFS lower is the priority.
- Contention Window –It is the amount of time divided into slots. A station which is ready to send frames chooses random number of slots as wait time.
- Acknowledgement – The positive acknowledgements and time-out timer can help guarantee a successful transmission of the frame.

Controlled Access:

- In controlled access, the stations consult one another to find which station has the right to send. A station cannot send unless it has been authorized by other stations.
- We discuss three popular controlled-access methods: Reservation, Polling & Token Passing

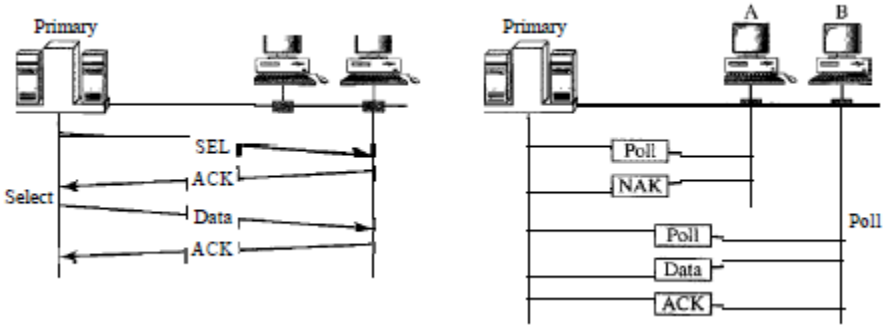
Reservation

- In the reservation method, a station needs to make a reservation before sending data.
- Time is divided into intervals. In each interval, a reservation frame precedes the data frames sent in that interval.
- If there are N stations in the system, there are exactly N reservation minislots in the reservation frame. Each minislot belongs to a station.
- When a station needs to send a data frame, it makes a reservation in its own minislot.
- The stations that have made reservations can send their data frames after the reservation frame.
- Figure below shows a situation with five stations and a five-minislot reservation frame.
- In the first interval, only stations 1, 3, and 4 have made reservations. In the second interval, only station 1 has made a reservation.



Polling:

- Polling works with topologies in which one device is designated as a primary station and the other devices are secondary stations.
- All data exchanges must be made through the primary device even when the ultimate destination is a secondary device.
- The primary device controls the link; the secondary devices follow its instructions. It is up to the primary device to determine which device is allowed to use the channel at a given time.
- The primary device, therefore, is always the initiator of a session.
- If the primary wants to receive data, it asks the secondaries if they have anything to send; this is called poll function.
- If the primary wants to send data, it tells the secondary to get ready to receive; this is called select function.



Select

- The select function is used whenever the primary device has something to send. Remember that the primary controls the link.
- If the primary is neither sending nor receiving data, it knows the link is available. If it has something to send, the primary device sends it.
- What it does not know, however, is whether the target device is prepared to receive. So, the primary must alert the secondary to the upcoming transmission and wait for an acknowledgment of the secondary's ready status.
- Before sending data, the primary creates and transmits a select (SEL) frame, one field of which includes the address of the intended secondary.

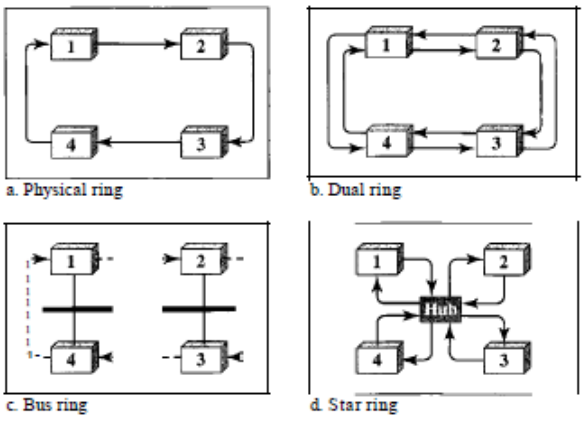
Poll

- The poll function is used by the primary device to solicit transmissions from the secondary devices.
- When the primary is ready to receive data, it must ask (poll) each device in turn if it has anything to send.
- When the first secondary is approached, it responds either with a NAK frame if it has nothing to send or with data (in the form of a data frame) if it does.
- If the response is negative (a NAK frame), then the primary polls the next secondary in the same manner until it finds one with data to send.
- When the response is positive (a data frame), the primary reads the frame and returns an acknowledgment (ACK frame), verifying its receipt.

Token Passing

- In the token-passing method, the stations in a network are organized in a logical ring. In other words, for each station, there is a predecessor and a successor.
- The predecessor is the station which is logically before the station in the ring; the successor is the station which is after the station in the ring.
- The current station is the one that is accessing the channel now. The right to this access has been passed from the predecessor to the current station.
- The right will be passed to the successor when the current station has no more data to send.
- But how is the right to access the channel passed from one station to another? In this method, a special packet called a token circulates through the ring.
- The possession of the token gives the station the right to access the channel and send its data.
- When a station has some data to send, it waits until it receives the token from its predecessor.
- It then holds the token and sends its data. When the station has no more data to send, it releases the token, passing it to the next logical station in the ring.
- The station cannot send data until it receives the token again in the next round.
- In this process, when a station receives the token and has no data to send, it just passes the data to the next station.

Logical ring and physical topology in token-passing access method



a. Physical ring
b. Dual ring
c. Bus ring
d. Star ring

Channelization:

- Channelization is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between different stations.
- There are three channelization protocols: FDMA, TDMA, and CDMA.

Frequency-Division Multiple Access (FDMA)

- In frequency-division multiple access (FDMA), the available bandwidth is divided into frequency bands. Each station is allocated a band to send its data.
- In other words, each band is reserved for a specific station, and it belongs to the station all the time.
- Each station also uses a bandpass filter to confine the transmitter frequencies. To prevent station interferences, the allocated bands are separated from one another by small guard bands.
- FDMA specifies a predetermined frequency band for the entire period of communication. This means that stream data (a continuous flow of data that may not be packetized) can easily be used with FDMA.
- We need to emphasize that although FDMA and FDM (Frequency Division Multiplexing) conceptually seem similar, there are differences between them.
- FDM is a physical layer technique that combines the loads from low-bandwidth channels and transmits them by using a high-bandwidth channel. FDMA, on the other hand, is an access method in the data link layer.

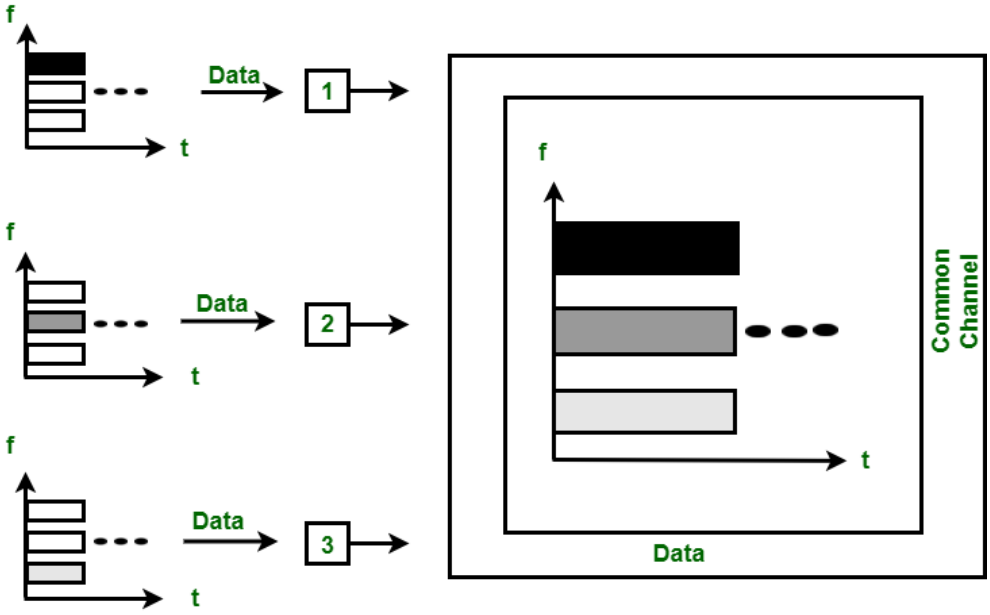


Fig: Frequency-Division Multiple Access

Time-Division Multiple Access (TDMA)

- In time-division multiple access (TDMA), the stations share the bandwidth of the channel in time.
- Each station is allocated a time slot during which it can send data. Each station transmits its data in is assigned time slot.
- The main problem with TDMA lies in achieving synchronization between the different stations. Each station needs to know the beginning of its slot and the location of its slot.
- This may be difficult because of propagation delays introduced in the system if the stations are spread over a large area. To compensate for the delays, we can insert guard times.
- Synchronization is normally accomplished by having some synchronization bits (normally referred to as preamble bits) at the beginning of each slot.
- We also need to emphasize that although TDMA and TDM conceptually seem the same, there are differences between them.
- TDM is a physical layer technique that combines the data from slower channels and transmits them by using a faster channel.
- The process uses a physical multiplexer that interleaves data units from each channel.
- TDMA, on the other hand, is an access method in the data link layer. The data link layer in each station tells its physical layer to use the allocated time slot. There is no physical multiplexer at the physical layer.

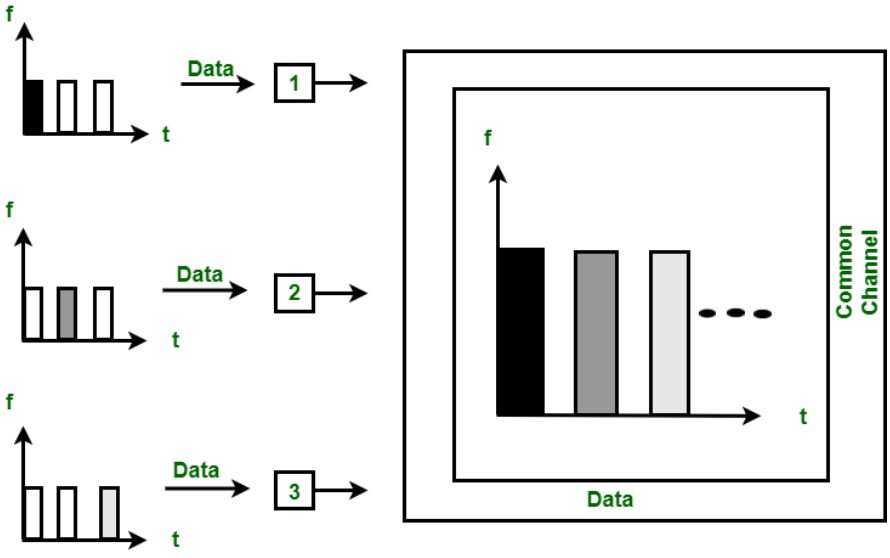


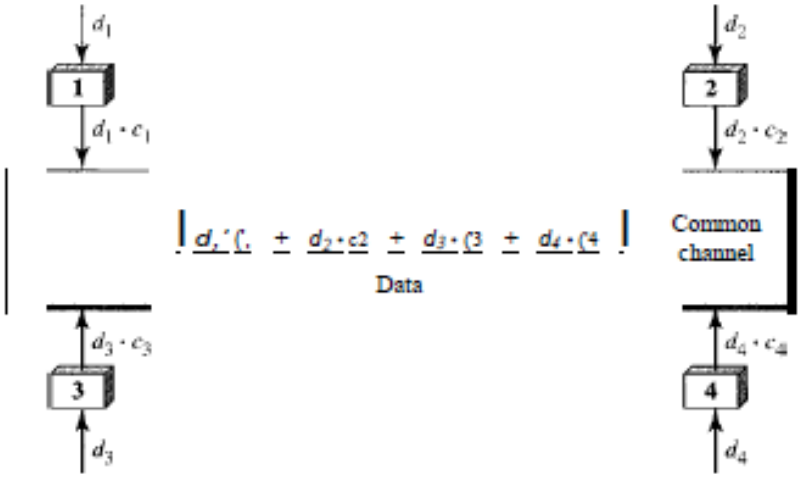
Fig: Time Division Multiple Access

Code Division Multiple Access (CDMA):

- Code-division multiple access (CDMA) was conceived several decades ago. Recent advances in electronic technology have finally made its implementation possible.

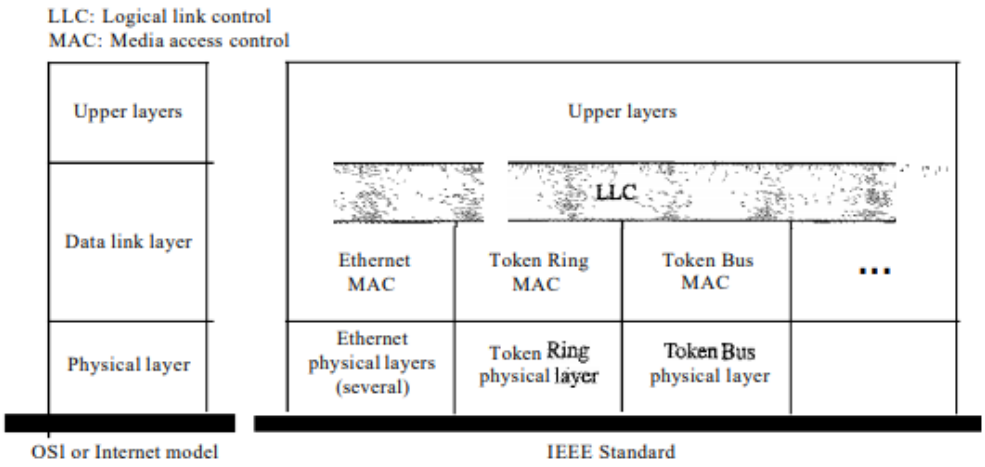
- CDMA differs from FDMA because only one channel occupies the entire bandwidth of the link. It differs from TDMA because all stations can send data simultaneously; there is no timesharing.
- CDMA simply means communication with different codes. For example, in a large room with many people, two people can talk in English if nobody else understands English. Another two people can talk in Chinese if they are the only ones who understand Chinese, and so on. In other words, the common channel, the space of the room in this case, can easily allow communication between several couples, but in different languages (codes).
- Let us assume we have four stations 1, 2, 3, and 4 connected to the same channel. The data from station 1 are d1, from station 2 are d2, and so on. The code assigned to the first station is c1, to the second is c2, and so on.
- Station 1 multiplies its data by its code to get d1. c1. Station 2 multiplies its data by its code to get d2.c2 and so on.
- The data that go on the channel are the sum of all these terms, as shown in the box. Any station that wants to receive data from one of the other three multiplies the data on the channel by the code of the sender.
- For example, suppose stations 1 and 2 are talking to each other. Station 2 wants to hear what station 1 is saying. It multiplies the data on the channel by c1 the code of station 1.

Simple idea ofcommunication with code



Wired LAN: IEEE Standards

In 1985, the Computer Society of the IEEE (Institute of Electrical and Electronics Engineers) started a project, called Project 802, to set standards to enable intercommunication among equipment from a variety of manufacturers. Project 802 does not seek to replace any part of the OSI or the Internet model. Instead, it is a way of specifying functions of the physical layer and the data link layer of major LAN protocols.

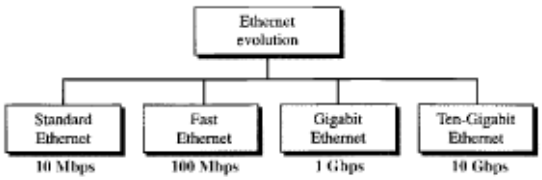


IEEE 802 is comprised of standards with separate working groups that regulate different communication networks, including IEEE 802.1 for bridging (bottom sublayer), 802.2 for Logical link (upper sublayer), 802.3 for Ethernet, 802.5 for token ring, 802.11 for Wi-Fi, 802.15 for Wireless Personal area networks, 802.15.1 for Bluetooth, 802.16 for Wireless Metropolitan Area Networks etc.

Ethernet:

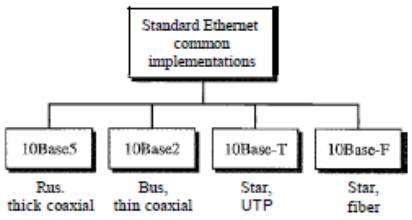
The original Ethernet was created in 1976 and since then, it has gone through four generations. Ethernet is most widely used LAN Technology, which is defined under IEEE standards 802.3. The reason behind its wide usability is Ethernet is easy to understand, implement, maintain and allows low-cost network implementation. Also, Ethernet offers flexibility in terms of topologies which are allowed. Ethernet generally uses Bus Topology. Ethernet operates in two layers of the OSI model, Physical Layer, and Data Link Layer.

Ethernet evolution through four generations

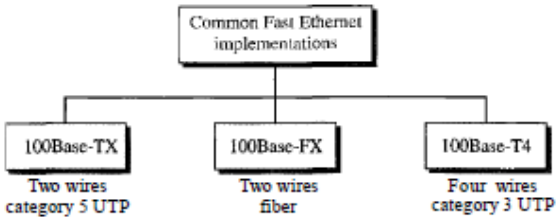


Ethernet is a family of computer networking technologies commonly used in local area networks (LAN), metropolitan area networks (MAN) and wide area networks (WAN). Systems using Ethernet communication divide data streams into packets, which are known as frames. Frames include source and destination address information, as well as mechanisms used to detect errors in transmitted data and retransmission requests. An Ethernet cable is the physical, encased wiring over which the data travels. Compared to wireless LAN technology, Ethernet is typically less vulnerable to disruptions. It can also offer a greater degree of network security and control than wireless technology, as devices must connect using physical cabling, making it difficult for outsiders to access network data or hijack bandwidth for unsanctioned devices.

Categories of Standard Ethernet



Fast Ethernet implementations



Gigabit Ethernet implementations

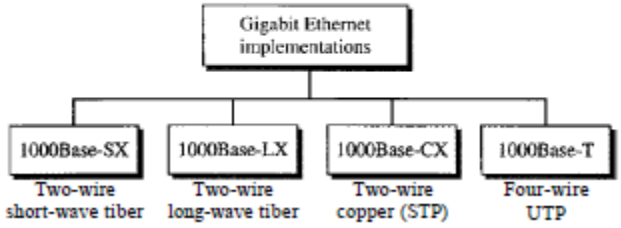


Table 13.4 *Summary of Ten-Gigabit Ethernet implementations*

Characteristics	*10GBase-S*	*10GBase-L*	*10GBase-E*
Media	Short-wave 850-nm multimode	Long-wave 1310-nm single mode	Extended 1550-nm single mode
Maximum length	300m	10km	40km

Fiber Distributed Data Interface (FDDI):

Fiber Distributed Data Interface (FDDI) is a standard for transmission of data in local area network (LAN) over fiber optic cables. It is applicable in large LANs that can extend up to 200 kilometers in diameter.

Features

- FDDI uses optical fiber as its physical medium.
- It operates in the physical and medium access control (MAC layer) of the Open Systems Interconnection (OSI) network model.
- It provides high data rate of 100 Mbps and can support thousands of users.
- It is used in LANs up to 200 kilometers for long distance voice and multimedia communication.
- It uses ring-based token passing mechanism and is derived from IEEE 802.4 token bus standard.
- It contains two token rings, a primary ring for data and token transmission and a secondary ring that provides backup if the primary ring fails.
- FDDI technology can also be used as a backbone for a wide area network (WAN).

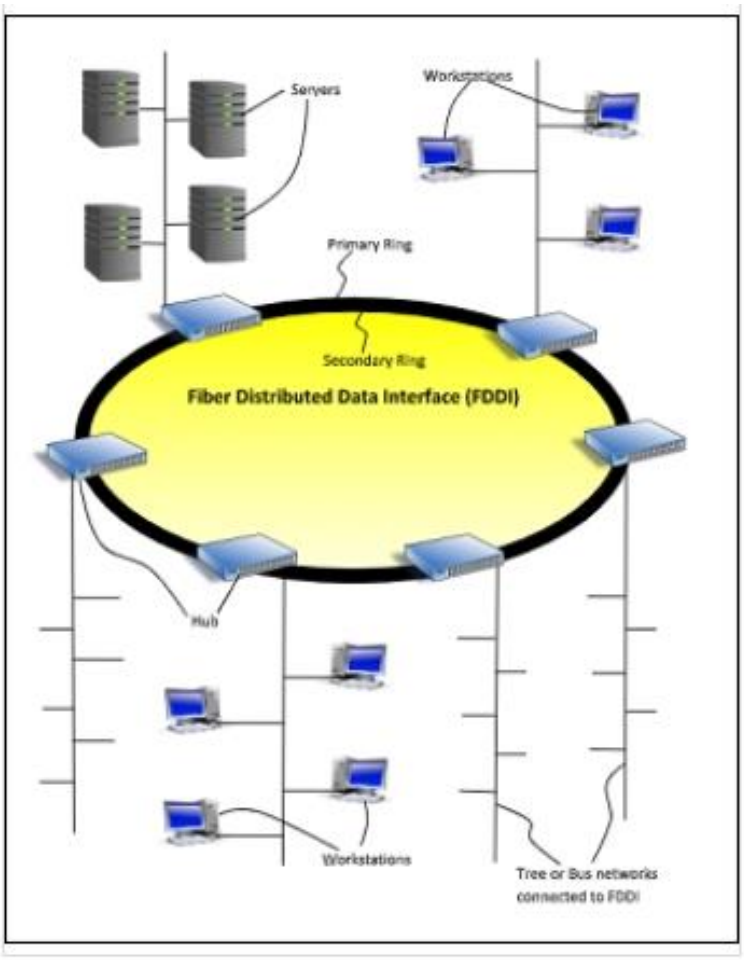


Fig: FDDI Implementation

Wireless LANs: IEEE 802.11x

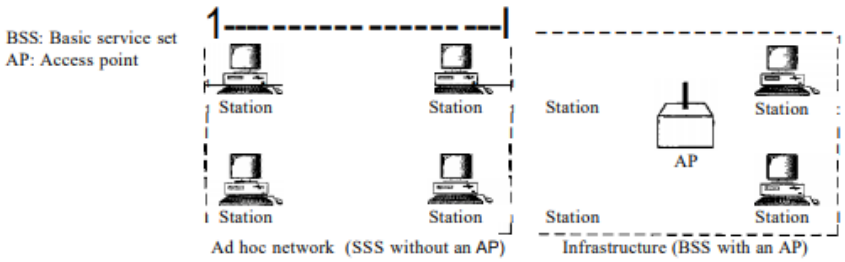
IEEE has defined the specifications for a wireless LAN, called IEEE 802.11, which covers the physical and data link layers. IEEE 802.11, commonly known as **Wi-Fi**, specifies an over-the-air interface between a wireless client and a base station or between two wireless clients.

The standard defines two kinds of services: the basic service set (BSS) and the extended service set (ESS).

Basic Service Set

IEEE 802.11 defines the basic service set (BSS) as the building block of a wireless LAN. A basic service set is made up of stationary or mobile wireless stations and an optional central base station, known as the access point (AP). Figure below shows two sets in this standard.

Figure 14.1 *Basic service sets (BSSs)*

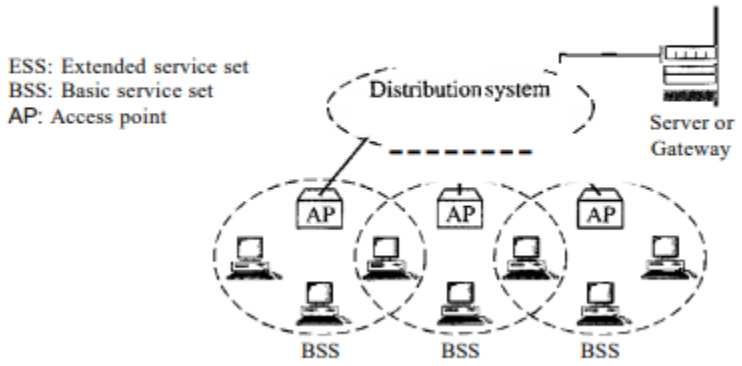


The BSS without an AP is a stand-alone network and cannot send data to other BSSs. It is called an ad hoc architecture. In this architecture, stations can form a network without the need of an AP; they can locate one another and agree to be part of a BSS. A BSS with an AP is sometimes referred to as an infrastructure network.

Extended Service Set

An extended service set (ESS) is made up of two or more BSSs with APs. In this case, the BSSs are connected through a distribution system, which is usually a wired LAN. The distribution system connects the APs in the BSSs. IEEE 802.11 does not restrict the distribution system; it can be any IEEE LAN such as an Ethernet. Note that the extended service set uses two types of stations: mobile and stationary. The mobile stations are normal stations inside a BSS. The stationary stations are AP stations that are part of a wired LAN.

Extended service sets (ESSs)



When BSSs are connected, the stations within reach of one another can communicate without the use of an AP. However, communication between two stations in two different BSSs usually occurs via two APs. The idea is similar to communication in a cellular network if we consider each BSS to be a cell and each AP to be a base station. Note that a mobile station can belong to more than one BSS at the same time.

Wi-Fi:

The IEEE 802.11 wireless LAN, also known as Wi-Fi, is the name of a popular wireless networking technology that uses radio waves to provide wireless high-speed Internet and network connections. Wi-

Fi networks have no physical wired connection between sender and receiver, by using radio frequency (RF) technology (a frequency within the electromagnetic spectrum associated with radio wave propagation). When an RF current is supplied to an antenna, an electromagnetic field is created that then is able to propagate through space.

There are several 802.11 standards for wireless LAN technology, including 802.11b, 802.11a, and 802.11g. Table below summarizes the main characteristics of these standards. 802.11g is by far the most popular technology.

Standard	Frequency Range (United States)	Data Rate
802.11b	2.4–2.485 GHz	up to 11 Mbps
802.11a	5.1–5.8 GHz	up to 54 Mbps
802.11g	2.4–2.485 GHz	up to 54 Mbps

Wi-Fi uses multiple parts of the IEEE 802 protocol family and is designed to seamlessly interwork with its wired sister protocol Ethernet. Devices that can use Wi-Fi technologies include desktops and laptops, smartphones and tablets, smart TVs, printers, digital audio players, digital cameras, cars and drones. Compatible devices can connect to each other over Wi-Fi through a wireless access point as well as to connected Ethernet devices and may use it to access the Internet. Such an access point (or hotspot) has a range of about 20 meters (66 feet) indoors and a greater range outdoors. Hotspot coverage can be as small as a single room with walls that block radio waves, or as large as many square kilometers achieved by using multiple overlapping access points.

Wi-Fi is potentially more vulnerable to attack than wired networks because anyone within range of a network with a wireless network interface controller can attempt access. Wi-Fi Protected Access (WPA) is a family of technologies created to protect information moving across Wi-Fi networks and includes solutions for personal and enterprise networks.

Bluetooth:

Bluetooth is a short-range wireless communication technology that allows devices such as mobile phones, computers, and peripherals to transmit data or voice wirelessly over a short distance. The purpose of Bluetooth is to replace the cables that normally connect devices, while still keeping the communications between them secure. It creates a 10-meter radius wireless network, called a personal area network (PAN) or piconet, which can network between two and eight devices. Bluetooth uses less power and costs less to implement than Wi-Fi. Its lower power also makes it far less prone to suffering from or causing interference with other wireless devices in the same 2.4GHz radio band.

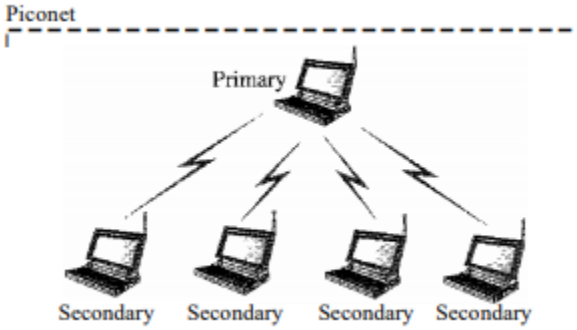
There are some downsides to Bluetooth. The first is that it can be a drain on battery power for mobile wireless devices like smartphones, though as the technology (and battery technology) has improved, this problem is less significant than it used to be. Also, the range is fairly limited, usually extending only about 30 feet, and as with all wireless technologies, obstacles such as walls, floors, or ceilings can reduce this

range further. The pairing process may also be difficult, often depending on the devices involved, the manufacturers, and other factors that all can result in frustration when attempting to connect.

Bluetooth defines two types of networks: piconet and scatternet.

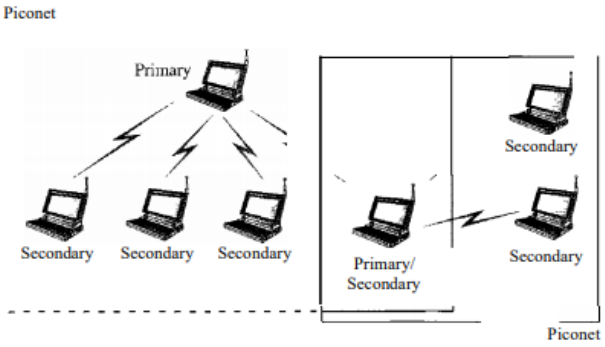
Piconets

A Bluetooth network is called a piconet, or a small net. A piconet can have up to eight stations, one of which is called the primary; t the rest are called secondaries. All the secondary stations synchronize their clocks and hopping sequence with the primary. Note that a piconet can have only one primary station. The communication between the primary and the secondary can be one-to-one or one-to-many. Figure below shows a piconet.



Scatternet

Piconets can be combined to form what is called a scatternet. A secondary station in one piconet can be the primary in another piconet. This station can receive messages from the primary in the first piconet (as a secondary) and, acting as a primary, deliver them to secondaries in the second piconet. A station can be a member of two piconets. Figure below illustrates a scatternet.



Token Ring:

Token ring is the IEEE 802.5 standard for a token-passing ring in Communication networks. A ring consists of a collection of ring interfaces connected by point-to-point lines i.e. ring interface of one station is connected to the ring interfaces of its left station as well as right station. Internally, signals travel around the Communication network from one station to the next in a ring. These point-to-point links can be created with twisted pair, coaxial cable or fiber optics. Each bit arriving at an interface is copied into a 1-

bit buffer. In this buffer the bit is checked and may be modified and is then copied out to the ring again. This copying of bit in the buffer introduces a 1-bit delay at each interface.

Token Ring is a LAN protocol defined in the IEEE 802.5 where all stations are connected in a ring and each station can directly hear transmissions only from its immediate neighbor. Permission to transmit is granted by a message (token) that circulates around the ring. A token is a special bit pattern (3 bytes long). There is only one token in the network. Token-passing networks move a small frame, called a token, around the network. Possession of the token grants the right to transmit. If a node receiving the token in order to transmit data, it seizes the token, alters 1 bit of the token (which turns the token into a start-of-frame sequence), appends the information that it wants to transmit, and sends this information to the next station on the ring. Since only one station can possess the token and transmit data at any given time, there are no collisions.

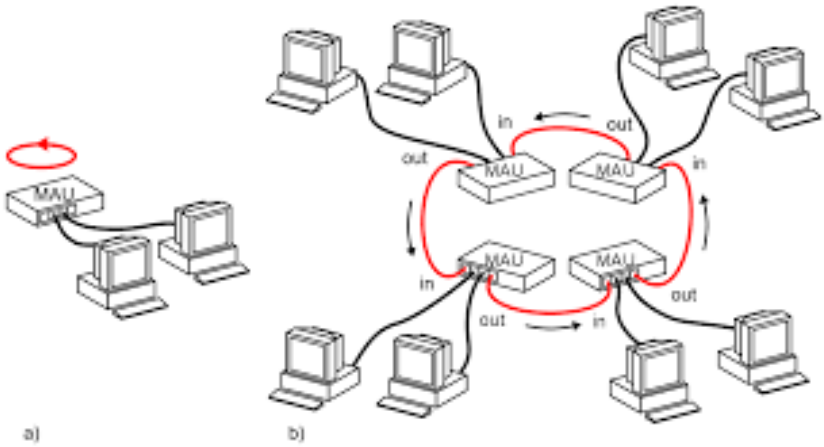


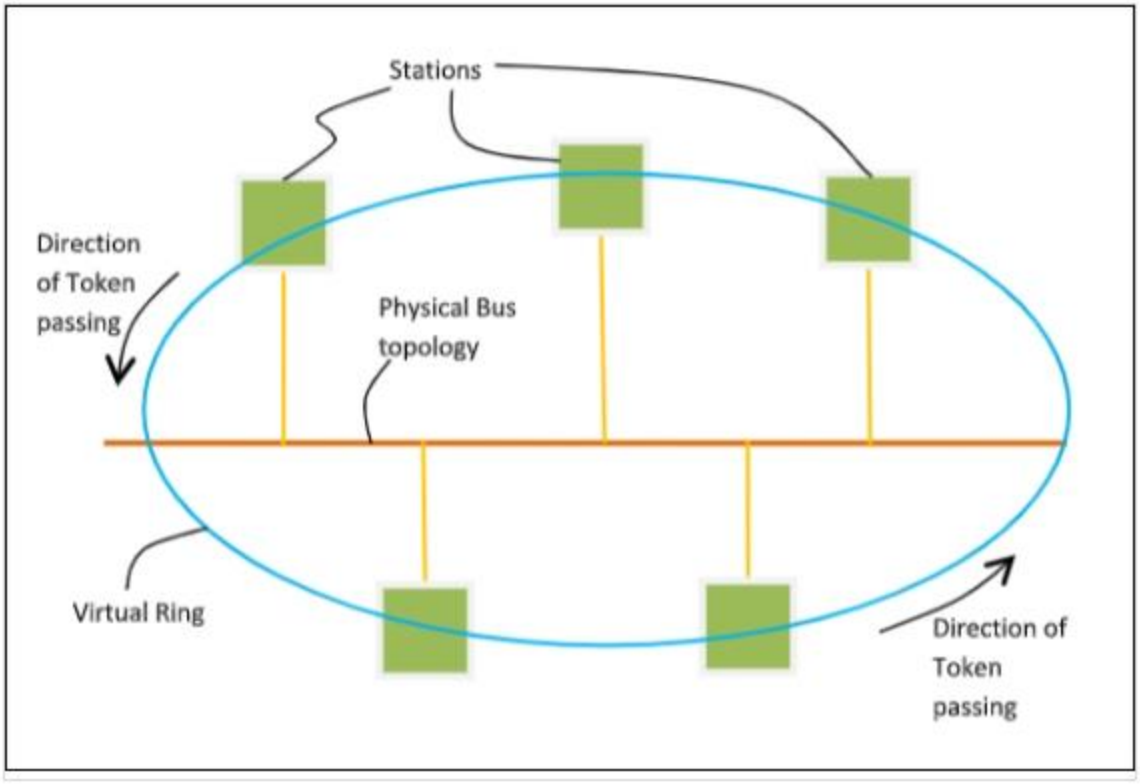
Fig: Two examples of Token Ring networks a) Using a single MAU b) Using several MAUs connected to each other, MAU (Media Access Unit)

Token Bus:

Token Bus (IEEE 802.4) is a standard for implementing token ring over virtual ring in LANs. The physical media has a bus or a tree topology and uses coaxial cables. A virtual ring is created with the nodes/stations and the token is passed from one node to the next in a sequence along this virtual ring. Each node knows the address of its preceding station and its succeeding station. A station can only transmit data when it has the token. The working principle of token bus is similar to Token Ring.

Token Passing Mechanism in Token Bus

A token is a small message that circulates among the stations of a computer network providing permission to the stations for transmission. If a station has data to transmit when it receives a token, it sends the data and then passes the token to the next station; otherwise, it simply passes the token to the next station. This is depicted in the following diagram:



Differences between Token Ring and Token Bus

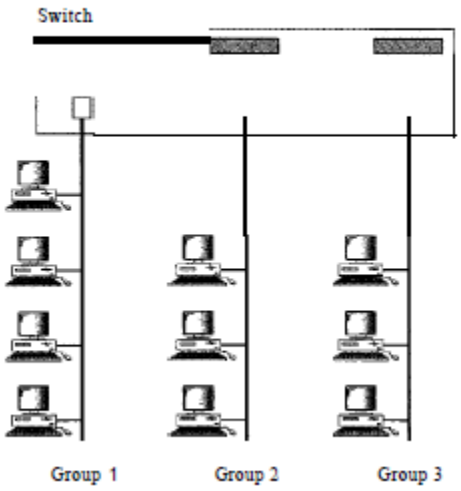
Token Ring	Token Bus
The token is passed over the physical ring formed by the stations and the coaxial cable network.	The token is passed along the virtual ring of stations connected to a LAN.
The stations are connected by ring topology, or sometimes star topology.	The underlying topology that connects the stations is either bus or tree topology.
It is defined by IEEE 802.5 standard.	It is defined by IEEE 802.4 standard.
The maximum time for a token to reach a station can be calculated here.	It is not feasible to calculate the time for token transfer.

Virtual LANs:

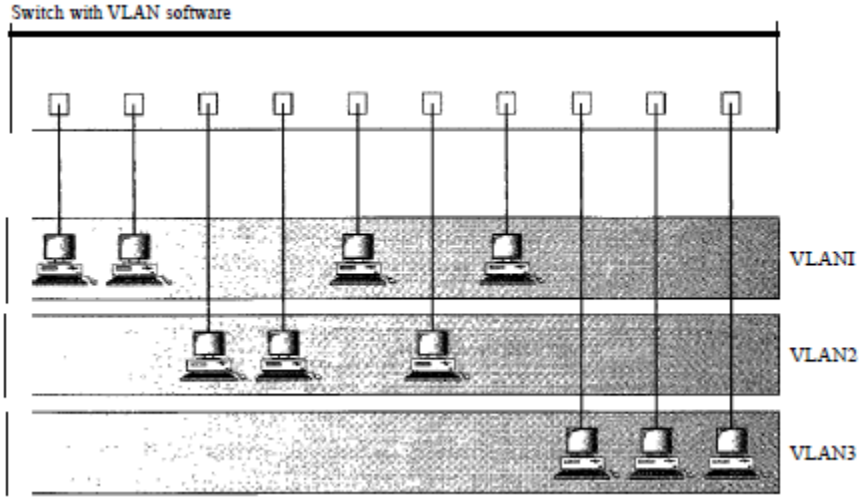
- A station is considered part of a LAN if it physically belongs to that LAN. The criterion of membership is geographic.
- What happens if we need a virtual connection between two stations belonging to two different physical LANs?
- We can roughly define a virtual local area network (VLAN) as a local area network configured by software, not by physical wiring.

- Figure below shows a switched LAN in an engineering firm in which 10 stations are grouped into three LANs that are connected by a switch.

A switch connecting three LANs



- The first four engineers work together as the first group, the next three engineers work together as the second group, and the last three engineers work together as the third group. The LAN is configured to allow this arrangement.
- But what would happen if the administrators needed to move two engineers from the first group to the third group, to speed up the project being done by the third group?
- The LAN configuration would need to be changed. The network technician must rewire.
- The problem is repeated if, in another week, the two engineers move back to their previous group.
- In a switched LAN, changes in the work group mean physical changes in the network configuration.
- Figure below shows the same switched LAN divided into VLANs. The whole idea of VLAN technology is to divide a LAN into logical, instead of physical, segments.



- A LAN can be divided into several logical LANs called VLANs. Each VLAN is a work group in the organization.

- If a person moves from one group to another, there is no need to change the physical configuration. The group membership in VLANs is defined by software, not hardware.
- Any station can be logically moved to another VLAN. All members belonging to a VLAN can receive broadcast messages sent to that particular VLAN.
- This means if a station moves from VLAN 1 to VLAN 2, it receives broadcast messages sent to VLAN 2, but no longer receives broadcast messages sent to VLAN 1.
- It is obvious that the problem in our previous example can easily be solved by using VLANs. Moving engineers from one group to another through software is easier than changing the configuration of the physical network.