Module 2

Data Communication Fundamentals

Lesson 3

Transmission Impairments and Channel Capacity

Specific Instructional Objectives

At the end of this lesson the students will be able to:

- Specify the Sources of impairments
- Explain Attenuation and unit of Attenuation
- Specify possible types of distortions of a signal
- Explain Data Rate Limits and Nyquist Bit Rate
- Distinguish between Bit Rate and Baud Rate
- Identify Noise Sources
- Explain Shannon Capacity in a Noisy Channel

2.3.1 Introduction

When a signal is transmitted over a communication channel, it is subjected to different types of impairments because of imperfect characteristics of the channel. As a consequence, the received and the transmitted signals are not the same. Outcome of the impairments are manifested in two different ways in analog and digital signals. These impairments introduce random modifications in analog signals leading to distortion. On the other hand, in case of digital signals, the impairments lead to error in the bit values. The impairment can be broadly categorised into the following three types:

- Attenuation and attenuation distortion
- Delay distortion
- Noise

In this lesson these impairments are discussed in detail and possible approaches to overcome these impairments. The concept of channel capacity for both noise-free and noisy channels have also been introduced.

2.3.2 Attenuation

Irrespective of whether a medium is guided or unguided, the strength of a signal falls off with distance. This is known as *attenuation.* In case of guided media, the attenuation is logarithmic, whereas in case of unguided media it is a more complex function of the distance and the material that constitutes the medium.

An important concept in the field of data communications is the use of on unit known as **decibel** (dB). To define it let us consider the circuit elements shown in Fig. 2.3.1. The elements can be either a transmission line, an amplifier, an attenuator, a filter, etc. In the figure, a transmission line (between points P_1 and P_2) is followed by an amplifier (between P_2 and P_3). The input signal delivers a power P_1 at the input of an communication element and the output power is P_2. Then the power gain G for this element in decibles is given by $G = 10\log_2 P_2/ P_1$. Here P_2/ P_1 is referred to as absolute power gain. When $P_2 > P_1$, the gain is positive, whereas if $P_2 < P_1$, then the power gain is negative and there is a power loss in the circuit element. For $P_2 = 5mW$, $P_1 = 10mW$, the power gain G = 10log 5/10 = 10 × -3 = -3dB is negative and it represents attenuation as a signal passes through the communication element.

Example: Let us consider a transmission line between points 1 and 2 and let the energy strength at point 2 is 1/10 of that of point 1. Then attenuation in dB is $10\log_{10}(1/10) = -10$ dB. On the other hand, there is an amplifier between points 2 and 3. Let the power is 100 times at point 3 with respect to point 2. Then power gain in dB is $10\log_{10}(100/1) = 20$ dB, which has a positive sign.

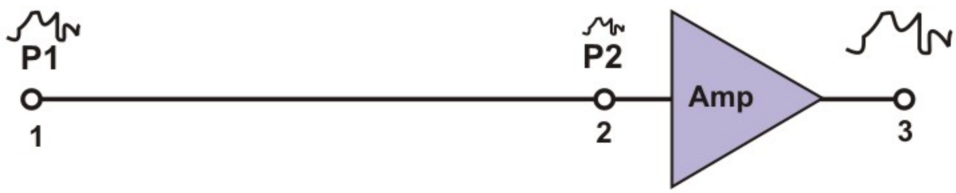

Figure 2.3.1 *Compensation of attenuation using an amplifier*

The attenuation leads to several problems:

Attenuation Distortion: If the strength of the signal is very low, the signal cannot be detected and interpreted properly at the receiving end. The signal strength should be sufficiently high so that the signal can be correctly detected by a receiver in presence of noise in the channel. As shown in Fig. 2.3.1, an amplifier can be used to compensate the attenuation of the transmission line. So, attenuation decides how far a signal can be sent without amplification through a particular medium.

Attenuation of all frequency components is not same. Some frequencies are passed without attenuation, some are weakened and some are blocked. This dependence of attenuation of a channel on the frequency of a signal leads to a new kind of distortion *attenuation distortion*. As shown in Fig. 2.3.2, a square wave is sent through a medium and the output is no longer a square wave because of more attenuation of the high-frequency components in the medium.

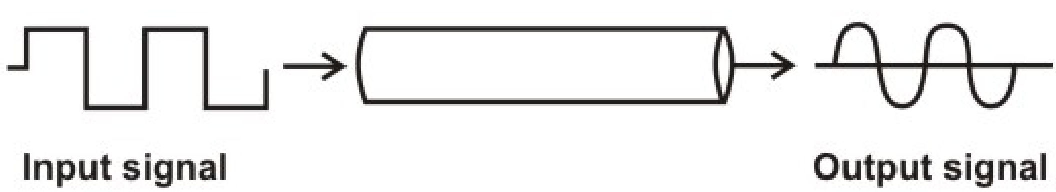

Figure 2.3.2 *Attenuation distortion of a square wave after passing through a medium.*

The effect of attenuation distortion can be reduced with the help of a suitable equalizer circuit, which is placed between the channel and the receiver. The equalizer has opposite attenuation/amplification characteristics of the medium and compensates higher losses of some frequency components in the medium by higher amplification in the equalizer. Attenuation characteristics of three popular transmission media are shown in Fig. 2.3.3. As shown in the figure, the attenuation of a signal increases exponentially as frequency is increased from KHz range to MHz range. In case of coaxial cable attenuation increases linearly with frequency in the Mhz range. The optical fibre, on the other hand, has attenuation characteristic similar to a band-pass filter and a small frequency band in the THz range can be used for the transmission of signal.

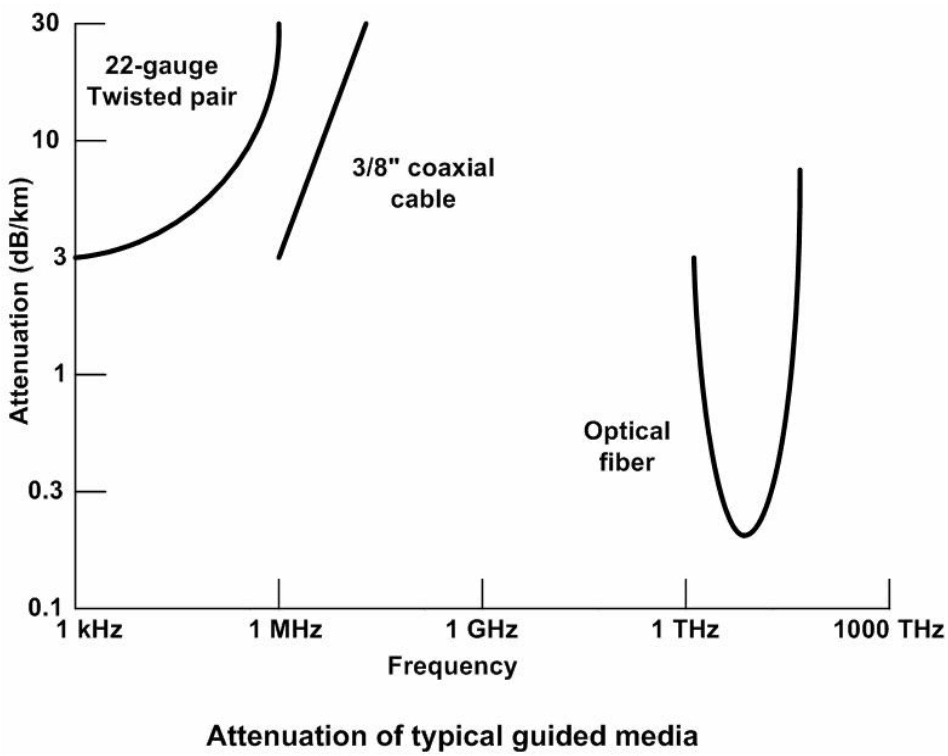

Figure 2.3.3 *Attenuation characteristics of the popular guided media*

2.3.3 Delay distortion

The velocity of propagation of different frequency components of a signal are different in guided media. This leads to delay distortion in the signal. For a bandlimited signal, the velocity of propagation has been found to be maximum near the center frequency and lower on both sides of the edges of the frequency band. In case of analog signals, the received signal is distorted because of variable delay of different components. In case of digital signals, the problem is much more severe. Some frequency components of one bit position spill over to other bit positions, because of delay distortion. This leads to intersymbol interference, which restricts the maximum bit rate of transmission through a particular transmission medium. The delay distortion can also be neutralised, like attenuation distortion, by using suitable equalizers.

2.3.4 Noise

As signal is transmitted through a channel, undesired signal in the form of noise gets mixed up with the signal, along with the distortion introduced by the transmission media. Noise can be categorised into the following four types:

- Thermal Noise
- Intermodulation Noise
- Cross talk
- Impulse Noise

The ***thermal noise*** is due to thermal agitation of electrons in a conductor. It is distributed across the entire spectrum and that is why it is also known as *white noise* (as the frequency encompass over a broad range of frequencies).

When more than one signal share a single transmission medium, ***intermodulation noise*** is generated. For example, two signals f_1 and f_2 will generate signals of frequencies (f_1 + f_2) and (f_1 - f_2), which may interfere with the signals of the same frequencies sent by the transmitter. Intermodulation noise is introduced due to nonlinearity present in any part of the communication system.

Cross talk is a result of bunching several conductors together in a single cable. Signal carrying wires generate electromagnetic radiation, which is induced on other conductors because of close proximity of the conductors. While using telephone, it is a common experience to hear conversation of other people in the background. This is known as *cross talk.*

Impulse noise is irregular pulses or noise spikes of short duration generated by phenomena like lightning, spark due to loose contact in electric circuits, etc. Impulse noise is a primary source of bit-errors in digital data communication. This kind of noise introduces burst errors.

2.3.5 Bandwidth and Channel Capacity

Bandwidth refers to the range of frequencies that a medium can pass without a loss of one-half of the power (-3dB) contained in the signal. Figure 2.3.4 shows the bandwidth of a channel. The points F_l and F_h points correspond to –3bB of the maximum amplitude A.

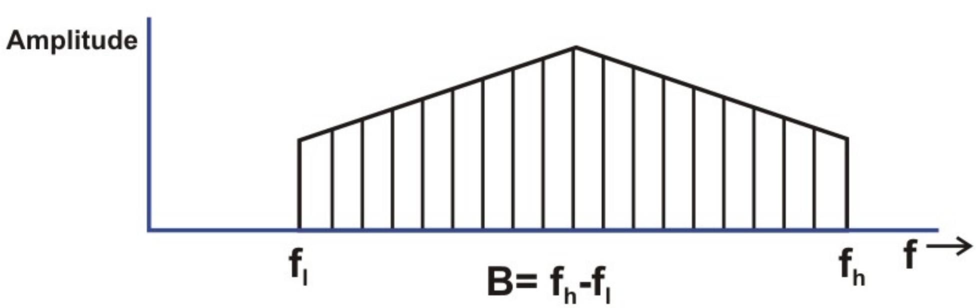

Figure 2.3.4 *Bandwidth of a channel*

Bandwidth of a medium decides the quality of the signal at the other end. A digital signal (usually aperiodic) requires a bandwidth from 0 to infinity. So, it needs a low-pass channel characteristic as shown in Fig. 2.3.5. On the other hand, a band-pass channel characteristic is required for the transmission of analog signals, as shown in Fig. 2.3.6.

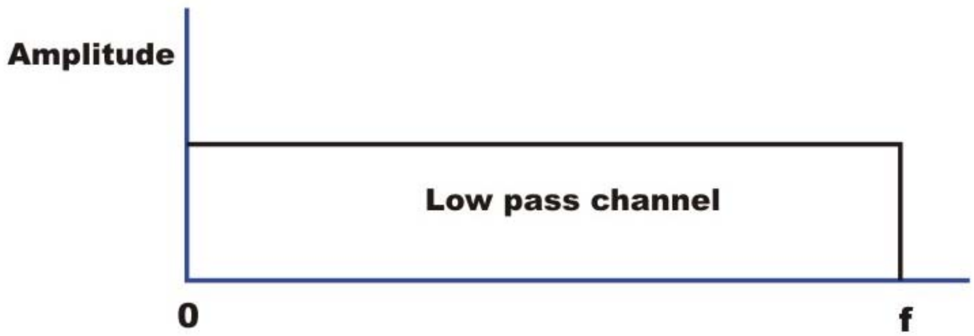

Figure 2.3.5 *Low-pass channel characteristic required for the transmission of digital signals*

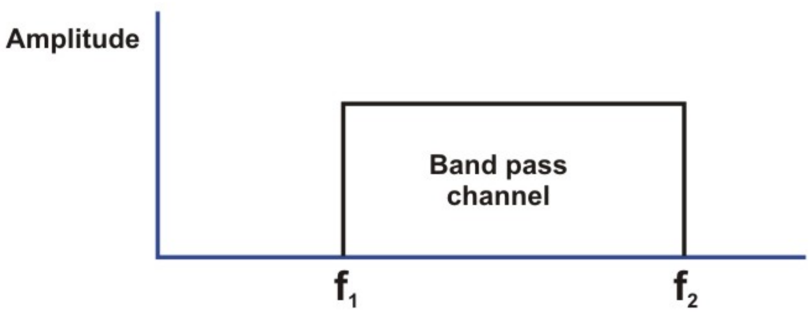

Figure 2.3.6 *Band-pass channel characteristic required for the transmission of analog signals*

Nyquist Bit Rate

The maximum rate at which data can be correctly communicated over a channel in presence of noise and distortion is known as its channel capacity. Consider first a noise-free channel of Bandwidth B. Based on Nyquist formulation it is known that given a bandwidth B of a channel, the maximum data rate that can be carried is 2B. This limitation arises due to the effect of intersymbol interference caused by the frequency components higher than B. If the signal consists of m discrete levels, then Nyquist theorem states:

Maximum data rate $C = 2 B \log_2 m$ bits/sec,

where C is known as the channel capacity,
B is the bandwidth of the channel
and m is the number of signal levels used.

Baud Rate: The baud rate or signaling rate is defined as the number of distinct symbols transmitted per second, irrespective of the form of encoding. For baseband digital transmission m = 2. So, the maximum baud rate = 1/Element width (in Seconds) = 2B
Bit Rate: The bit rate or information rate I is the actual equivalent number of bits transmitted per second. I = Baud Rate × Bits per Baud

= Baud Rate × N = Baud Rate × log2m

For binary encoding, the bit rate and the baud rate are the same; i.e., I = Baud Rate.
Example: Let us consider the telephone channel having bandwidth B = 4 kHz. Assuming there is no noise, determine channel capacity for the following encoding levels:
(i) 2, and (ii) 128.

Ans: (i) C = 2B = 2×4000 = 8 Kbits/s

(ii) C = 2×4000×log2128 = 8000×7 = 56 Kbits/s

Effects of Noise
When there is noise present in the medium, the limitations of both bandwidth and noise must be considered. A noise spike may cause a given level to be interpreted as a signal of greater level, if it is in positive phase or a smaller level, if it is negative phase. Noise becomes more problematic as the number of levels increases.

Shannon Capacity (Noisy Channel)
In presence of Gaussian band-limited white noise, Shannon-Hartley theorem gives the maximum data rate capacity

C = B log2 (1 + S/N),

where S and N are the signal and noise power, respectively, at the output of the channel. This theorem gives an upper bound of the data rate which can be reliably transmitted over a thermal-noise limited channel.

Example: Suppose we have a channel of 3000 Hz bandwidth, we need an S/N ratio (i.e. signal to noise ration, SNR) of 30 dB to have an acceptable bit-error rate. Then, the maximum data rate that we can transmit is 30,000 bps. In practice, because of the presence of different types of noises, attenuation and delay distortions, actual (practical) upper limit will be much lower.

In case of extremely noisy channel, C = 0
Between the Nyquist Bit Rate and the Shannon limit, the result providing the smallest channel capacity is the one that establishes the limit.

Example: A channel has *B* = 4 KHz. Determine the channel capacity for each of the following signal-to-noise ratios: (a) 20 dB, (b) 30 dB, (c) 40 dB.

Answer: (a) $C = B \log_2 (1 + S/N) = 4\times10^3\times\log_2 (1+100) = 4\times10^3\times3.32\times2.004 = 26.6$ kbits/s

b) $C = B \log_2 (1 + S/N) = 4\times10^3\times\log_2 (1+1000) = 4\times10^3\times3.32\times3.0 = 39.8$ kbits/s

(c) $C = B \log_2 (1 + S/N) = 4\times10^3\times\log_2 (1+10000) = 4\times10^3\times3.32\times4.0 = 53.1$ kbits/s

Example: A channel has B = 4 KHz and a signal-to-noise ratio of 30 dB. Determine maximum information rate for 4-level encoding.
Answer: For B = 4 KHz and 4-level encoding the *Nyquist Bit Rate* is 16 Kbps. Again for B = 4 KHz and S/N of 30 dB the *Shannon capacity* is 39.8 Kbps. The smallest of the two values has to be taken as the Information capacity I = 16 Kbps.

Example: A channel has B = 4 kHz and a signal-to-noise ratio of 30 dB. Determine maximum information rate for 128-level encoding.

Answer: The *Nyquist Bit Rate* for B = 4 kHz and M = 128 levels is 56 kbits/s. Again the *Shannon capacity* for B = 4 kHz and S/N of 30 dB is 39.8 Kbps. The smallest of the two values decides the channel capacity C = 39.8 kbps.

Example: The digital signal is to be designed to permit 160 kbps for a bandwidth of 20 KHz. Determine (a) number of levels and (b) S/N ratio.

(a) Apply *Nyquist Bit Rate* to determine number of levels.

$C = 2B \log_2 (M)$,

or $160\times10^3 = 2\times20\times10^3 \log_2 (M)$,

or $M = 2^4$, which means 4bits/baud.

(b) Apply *Shannon capacity* to determine the S/N ratio

$C = B \log_2 (1+S/N)$,

or $160\times10^3 = 20\times10^3 \log_2 (1+S/N) \times10^3 \log_2 (M)$,

or $S/N = 2^8 - 1$,

or $S/N = 255$,

or $S/N = 24.07$ dB.

Review Questions

Q-1. Distinguish between attenuation distortion and delay distortion.
Ans: Attenuation distortion arises because the attenuation of the signal in the transmitting media. Attenuation distortion is predominant in case of analog signals. Delay distortion arises because different frequency components of the signal suffer different delay as the signal passes through the media. This happens because the velocity of the signal varies with frequency and it is predominant in case of digital signals.

Q-2. How the effect of delay distortion can be minimized?
Ans: Delay distortion can be minimized by using an equalizer (a kind of filter).

Q-3. What is intermodulation noise?
Ans: When a signal (having different frequency components) passes through a transmitting media, then due to non-linearity, some of the frequency components may combine to generate a different frequency component. This leads to distortion in the signal, which is known as intermodulation noise. For example, a signal may be having frequency components f_1 and f_2, and due to non-linearity of the media they may generate a frequency component (f_1+f_2). Further a frequency of (f_1+f_2) may be already present in the original signal. This causes intermodulation noise.

Q-4. Why does impulse noise have more effect on digital signals rather than on analog signals?
Ans: Impulse noise is random in nature and arises due to random events like lightning, electrical sparks, etc. In case of digital signal, it makes a significant effect, as '0' may become '1' and vice versa. In analog signal the effect is not that serious as some portion of the signal gets affected.

Q-5. What is crosstalk?
Ans: Crosstalk refers to the picking up of electromagnetic signals from other adjacent wires by electromagnetic induction.

Q-6. Let the energy strength at point 2 is 1/50th with respect to the point 1. Find out the attenuation in dB.
Ans: Then attenuation in dB is $10\log_{10}(1/50) = -16.9$ dB.

Q-7. Assuming there is no noise in a medium of B = 4KHz, determine channel capacity for the encoding level 4.

Ans: $I = 2\times4000\times\log_2 4 = 16$ Kbps

Q-8. A channel has *B* = 10 MHz. Determine the channel capacity for signal-to-noise ratio 60 dB.
Ans: $C = B ´ \log_2(1 + S/N) = 10 \text{ x } \log_2(1 + 60)$

Q-9. The digital signal is to be designed to permit 56 kbps for a bandwidth of 4 KHz. Determine (a) number of levels and (b) S/N ratio.

Module 3

Data Link control

Lesson 3

Flow Control and Error Control

Special Instructional Objectives:

On completion, the student will be able to:

- State the need for flow and error control
- Explain how Stop-and-wait flow control works
- Explain how Sliding-window protocol is used for flow control
- Explain how Stop-and-wait ARQ works
- Explain how Go-back-N ARQ works
- Explain how Selective-repeat ARQ works

3.3.1 Introduction

As we have mentioned earlier, for reliable and efficient data communication a great deal of coordination is necessary between at least two machines. Some of these are necessary because of the following constraints:

- Both sender and receiver have limited speed
- Both sender and receiver have limited memory

It is necessary to satisfy the following requirements:

- A fast sender should not overwhelm a slow receiver, which must perform a certain amount of processing before passing the data on to the higher-level software.
- If error occur during transmission, it is necessary to devise mechanism to correct it

The most important functions of Data Link layer to satisfy the above requirements are **error control** and **flow control**. Collectively, these functions are known as **data link control**, as discussed in this lesson.

Flow Control is a technique so that transmitter and receiver with different speed characteristics can communicate with each other. Flow control ensures that a transmitting station, such as a server with higher processing capability, does not overwhelm a receiving station, such as a desktop system, with lesser processing capability. This is where there is an orderly flow of transmitted data between the source and the destination.

Error Control involves both error detection and error correction. It is necessary because errors are inevitable in data communication, in spite of the use of better equipment and reliable transmission media based on the current technology. In the preceding lesson we have already discussed how errors can be detected. In this lesson we shall discuss how error control is performed based on retransmission of the corrupted data. When an error is detected, the receiver can have the specified frame retransmitted by the sender. This process is commonly known as **Automatic Repeat Request (ARQ).** For example, Internet's Unreliable Delivery Model allows packets to be discarded if network resources are not available, and demands that ARQ protocols make provisions for retransmission.

3.3.2 Flow Control

Modern data networks are designed to support a diverse range of hosts and communication mediums. Consider a 933 MHz Pentium-based host transmitting data to a 90 MHz 80486/SX. Obviously, the Pentium will be able to drown the slower processor with data. Likewise, consider two hosts, each using an Ethernet LAN, but with the two Ethernets connected by a 56 Kbps modem link. If one host begins transmitting to the other at Ethernet speeds, the modem link will quickly become overwhelmed. In both cases, *flow control* is needed to pace the data transfer at an acceptable speed.

Flow Control is a set of procedures that tells the sender how much data it can transmit before it must wait for an acknowledgment from the receiver. The flow of data should not be allowed to overwhelm the receiver. Receiver should also be able to inform the transmitter before its limits (this limit may be amount of memory used to store the incoming data or the processing power at the receiver end) are reached and the sender must send fewer frames. Hence, **Flow control** refers to the set of procedures used to restrict the amount of data the transmitter can send before waiting for acknowledgment.

There are two methods developed for flow control namely **Stop-and-wait** and **Sliding-window**. Stop-and-wait is also known as Request/reply sometimes. Request/reply (Stop-and-wait) flow control requires each data packet to be acknowledged by the remote host before the next packet is sent. This is discussed in detail in the following subsection. **Sliding window** algorithms, used by TCP, permit multiple data packets to be in simultaneous transit, making more efficient use of network bandwidth as discussed in subsection 3.3.2.2.

3.3.2.1 Stop-and-Wait

This is the simplest form of flow control where a sender transmits a data frame. After receiving the frame, the receiver indicates its willingness to accept another frame by sending back an ACK frame acknowledging the frame just received. The sender must wait until it receives the ACK frame before sending the next data frame.This is sometimes referred to as *ping-pong* behavior, request/reply is simple to understand and easy to implement, but not very efficient. In LAN environment with fast links, this isn't much of a concern, but WAN links will spend most of their time idle, especially if several hops are required.

Figure 3.3.1 illustrates the operation of the stop-and-wait protocol. The blue arrows show the sequence of data frames being sent across the link from the sender (top to the receiver (bottom). The protocol relies on two-way transmission (full duplex or half duplex) to allow the receiver at the remote node to return frames acknowledging the successful transmission. The acknowledgements are shown in green in the diagram, and flow back to the original sender. A small processing delay may be introduced between reception of the last byte of a Data PDU and generation of the corresponding ACK.

Major drawback of Stop-and-Wait Flow Control is that only one frame can be in transmission at a time, this leads to inefficiency if propagation delay is much longer than the transmission delay.

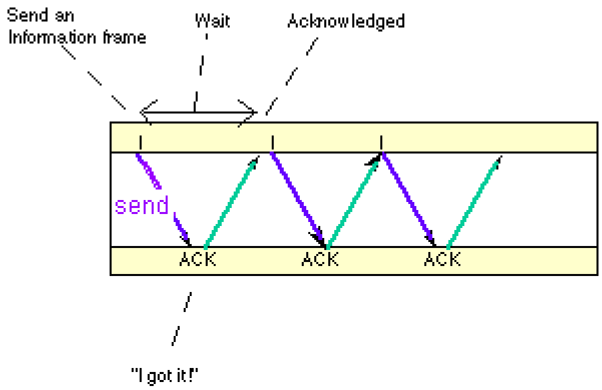

Figure 3. 3.1 Stop-and Wait protocol

Some protocols pretty much require stop-and-wait behavior. For example, Internet's Remote Procedure Call (RPC) Protocol is used to implement subroutine calls from a program on one machine to library routines on another machine. Since most programs are single threaded, the sender has little choice but to wait for a reply before continuing the program and possibly sending another request.

Link Utilization in Stop-and-Wait

Let us assume the following:

Transmission time: The time it takes for a station to transmit a frame (normalized to a value of 1).

Propagation delay: The time it takes for a bit to travel from sender to receiver (expressed as a).

- $a < 1$:The frame is sufficiently long such that the first bits of the frame arrive at the destination before the source has completed transmission of the frame.
- $a > 1$: Sender completes transmission of the entire frame before the leading bits of the frame arrive at the receiver.
- The link utilization U = 1/(1+2a),

  a = Propagation time / transmission time

It is evident from the above equation that the link utilization is strongly dependent on the ratio of the propagation time to the transmission time. When the propagation time is small, as in case of LAN environment, the link utilization is good. But, in case of long propagation delays, as in case of satellite communication, the utilization can be very poor. To improve the link utilization, we can use the following (sliding-window) protocol instead of using stop-and-wait protocol.

3.3.2.2 Sliding Window

With the use of multiple frames for a single message, the stop-and-wait protocol does not perform well. Only one frame at a time can be in transit. In stop-and-wait flow control, if $a > 1$, serious inefficiencies result. Efficiency can be greatly improved by allowing multiple frames to be in transit at the same time. Efficiency can also be improved by making use of the full-duplex line. To keep track of the frames, sender station sends sequentially numbered frames. Since the sequence number to be used occupies a field in the frame, it should be of limited size. If the header of the frame allows k bits, the sequence numbers range from 0 to $2^k - 1$. Sender maintains a list of sequence numbers that it is allowed to send (sender window). The size of the sender's window is at most $2^k - 1$. The sender is provided with a buffer equal to the window size. Receiver also maintains a window of size $2^k - 1$. The receiver acknowledges a frame by sending an ACK frame that includes the sequence number of the next frame expected. This also explicitly announces that it is prepared to receive the next N frames, beginning with the number specified. This scheme can be used to acknowledge multiple frames. It could receive frames 2, 3, 4 but withhold ACK until frame 4 has arrived. By returning an ACK with sequence number 5, it acknowledges frames 2, 3, 4 in one go. The receiver needs a buffer of size 1.

Sliding window algorithm is a method of flow control for network data transfers. TCP, the Internet's stream transfer protocol, uses a sliding window algorithm.

A sliding window algorithm places a buffer between the application program and the network data flow. For TCP, the buffer is typically in the operating system kernel, but this is more of an implementation detail than a hard-and-fast requirement.

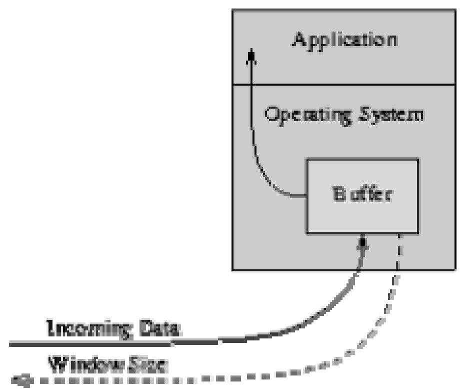

Figure 3.3.2 Buffer in sliding window

Data received from the network is stored in the buffer, from where the application can read at its own pace. As the application reads data, buffer space is freed up to accept more input from the network. The *window* is the amount of data that can be "read ahead" - the size of the buffer, less the amount of valid data stored in it. *Window announcements* are used to inform the remote host of the current *window size*.

Sender sliding Window: At any instant, the sender is permitted to send frames with sequence numbers in a certain range (the sending window) as shown in Fig. 3.3.3.

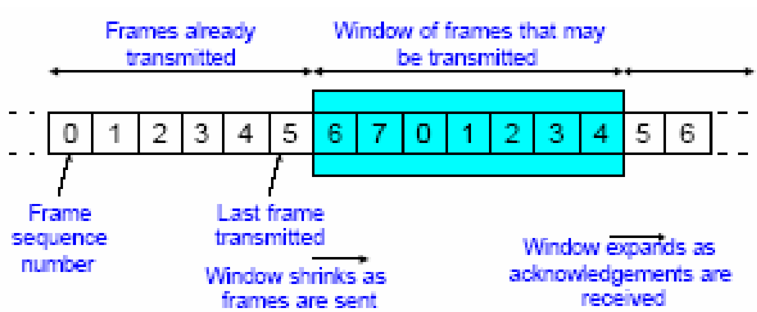

Figure 3.3.3 Sender's window

Receiver sliding Window: The receiver always maintains a window of size 1 as shown in Fig. 3.3.4. It looks for a specific frame (frame 4 as shown in the figure) to arrive in a specific order. If it receives any other frame (out of order), it is discarded and it needs to be resent. However, the receiver window also slides by one as the specific frame is received and accepted as shown in the figure. The receiver acknowledges a frame by sending an ACK frame that includes the sequence number of the next frame expected. This also explicitly announces that it is prepared to receive the next N frames, beginning with the number specified. This scheme can be used to acknowledge multiple frames. It could receive frames 2, 3, 4 but withhold ACK until frame 4 has arrived. By returning an ACK with sequence number 5, it acknowledges frames 2, 3, 4 at one time. The receiver needs a buffer of size 1.

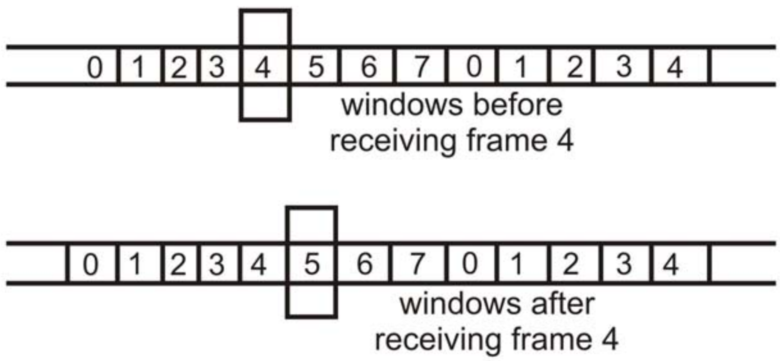

Figure 3.3.4 Receiver sliding window

On the other hand, if the local application can process data at the rate it's being transferred; sliding window still gives us an advantage. If the window size is larger than the packet size, then multiple packets can be outstanding in the network, since the sender knows that buffer space is available on the receiver to hold all of them. Ideally, a steady-state condition can be reached where a series of packets (in the forward direction) and window announcements (in the reverse direction) are constantly in transit. As each new window announcement is received by the sender, more data packets are transmitted. As the application reads data from the buffer (remember, we're assuming the application can keep up with the network), more window announcements are generated. Keeping a series of data packets in transit ensures the efficient use of network resources.

Hence, Sliding Window Flow Control

- Allows transmission of multiple frames
- Assigns each frame a k-bit sequence number
- Range of sequence number is $[0 \ldots 2^k-1]$, i.e., frames are counted modulo 2k.

The link utilization in case of Sliding Window Protocol

U = 1, for $N > 2a + 1$

$N/(1+2a)$, for $N < 2a + 1$

Where N = the window size,

and a = Propagation time / transmission time

3.3.3 Error Control Techniques

When an error is detected in a message, the receiver sends a request to the transmitter to retransmit the ill-fated message or packet. The most popular retransmission scheme is known as Automatic-Repeat-Request (ARQ). Such schemes, where receiver asks transmitter to re-transmit if it detects an error, are known as reverse error correction techniques. There exist three popular ARQ techniques, as shown in Fig. 3.3.5.

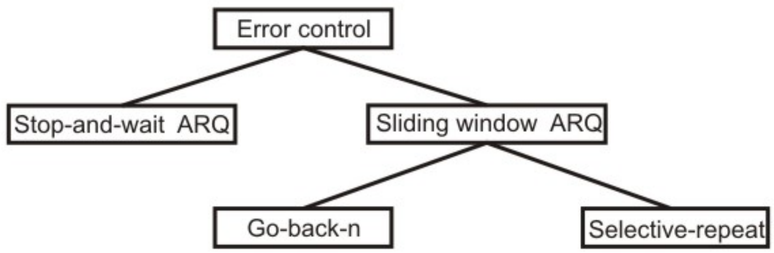

Figure 3.3.5 Error control techniques

3.3.3.1 Stop-and-Wait ARQ

In Stop-and-Wait ARQ, which is simplest among all protocols, the sender (say station A) transmits a frame and then waits till it receives positive acknowledgement (ACK) or negative acknowledgement (NACK) from the receiver (say station B). Station B sends an ACK if the frame is received correctly, otherwise it sends NACK. Station A sends a new frame after receiving ACK; otherwise it retransmits the old frame, if it receives a NACK. This is illustrated in Fig 3.3.6.

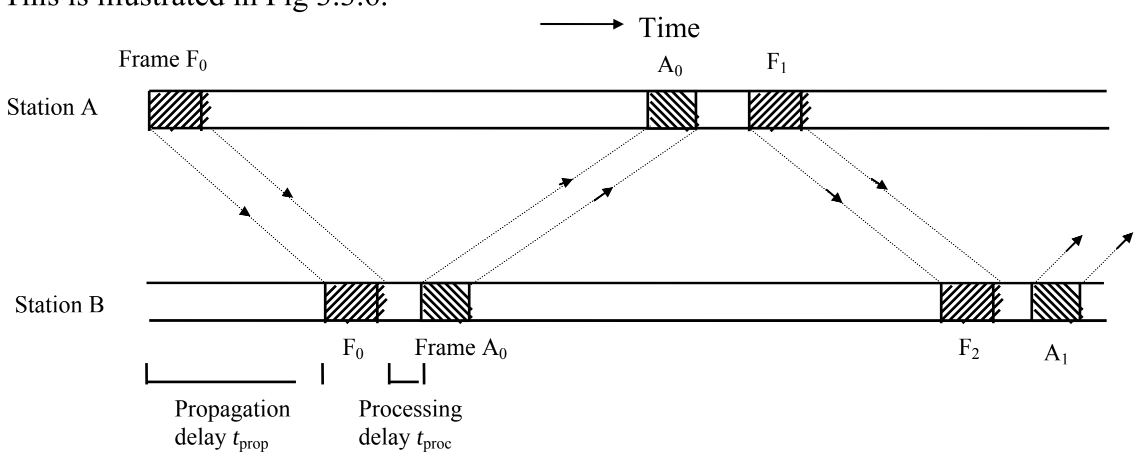

Figure 3.3.6 Stop-And-Wait ARQ technique

To tackle the problem of a lost or damaged frame, the sender is equipped with a timer. In case of a lost ACK, the sender transmits the old frame. In the Fig. 3.3.7, the second PDU of Data is lost during transmission. The sender is unaware of this loss, but starts a timer after sending each PDU. Normally an ACK PDU is received before the timer expires. In this case no ACK is received, and the timer counts down to zero and triggers retransmission of the same PDU by the sender. The sender always starts a timer following transmission, but in the second transmission receives an ACK PDU before the timer expires, finally indicating that the data has now been received by the remote node.

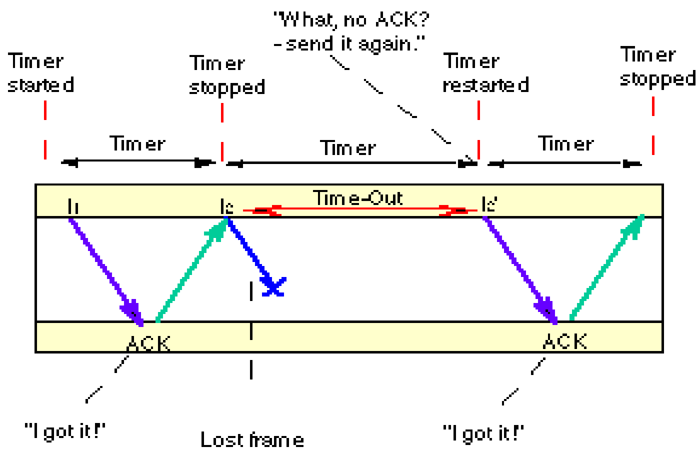

Figure 3.3.7 Retransmission due to lost frame

The receiver now can identify that it has received a duplicate frame from the label of the frame and it is discarded

To tackle the problem of damaged frames, say a frame that has been corrupted during the transmission due to noise, there is a concept of NACK frames, i.e. Negative Acknowledge frames. Receiver transmits a NACK frame to the sender if it founds the received frame to be corrupted. When a NACK is received by a transmitter before the time-out, the old frame is sent again as shown in Fig. 3.3.8.

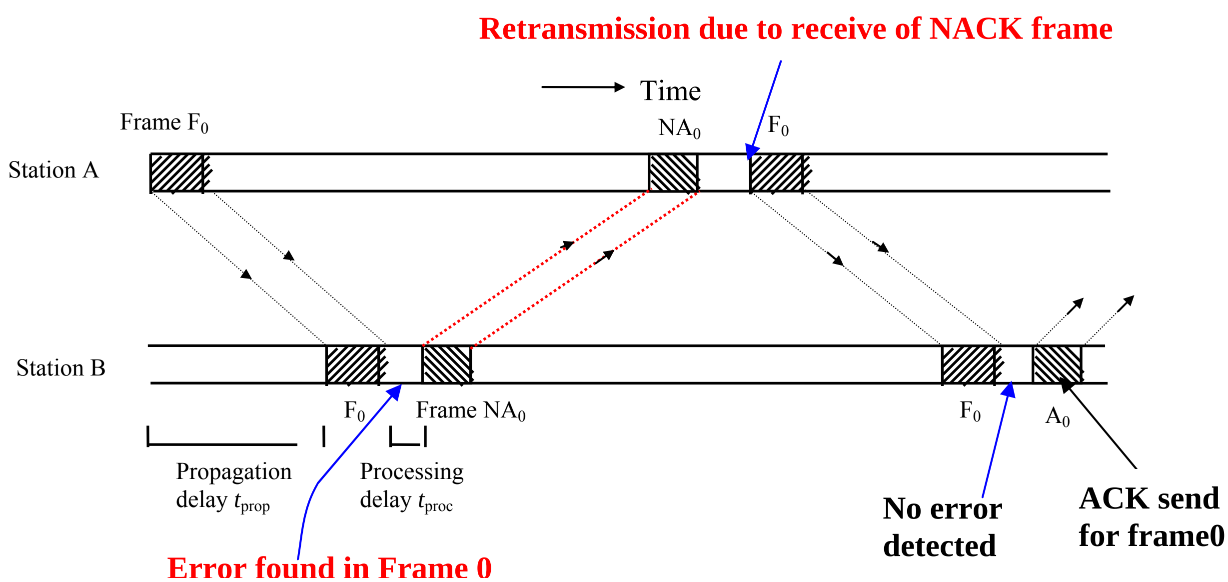

Figure 3.3.8 Retransmission due to damaged frame

The main advantage of stop-and-wait ARQ is its simplicity. It also requires minimum buffer size. However, it makes highly inefficient use of communication links, particularly when ‘a’ is large.

3.3.3.2 Go-back-N ARQ

The most popular ARQ protocol is the go-back-N ARQ, where the sender sends the frames continuously without waiting for acknowledgement. That is why it is also called as *continuous ARQ*. As the receiver receives the frames, it keeps on sending ACKs or a NACK, in case a frame is incorrectly received. When the sender receives a NACK, it retransmits the frame in error plus all the succeeding frames as shown in Fig.3.3.9. Hence, the name of the protocol is go-back-N ARQ. If a frame is lost, the receiver sends NAK after receiving the next frame as shown in Fig. 3.3.10. In case there is long delay before sending the NAK, the sender will resend the lost frame after its timer times out. If the ACK frame sent by the receiver is lost, the sender resends the frames after its timer times out as shown in Fig. 3.3.11.

Assuming full-duplex transmission, the receiving end sends piggybacked acknowledgement by using some number in the ACK field of its data frame. Let us assume that a 3-bit sequence number is used and suppose that a station sends frame 0 and gets back an RR1, and then sends frames 1, 2, 3, 4, 5, 6, 7, 0 and gets another RR1.This might either mean that RR1 is a cumulative ACK or all 8 frames were damaged. This ambiguity can be overcome if the maximum window size is limited to 7, i.e. for a k-bit sequence number field it is limited to 2^k-1. The number N (=2^k-1) specifies how many frames can be sent without receiving acknowledgement.

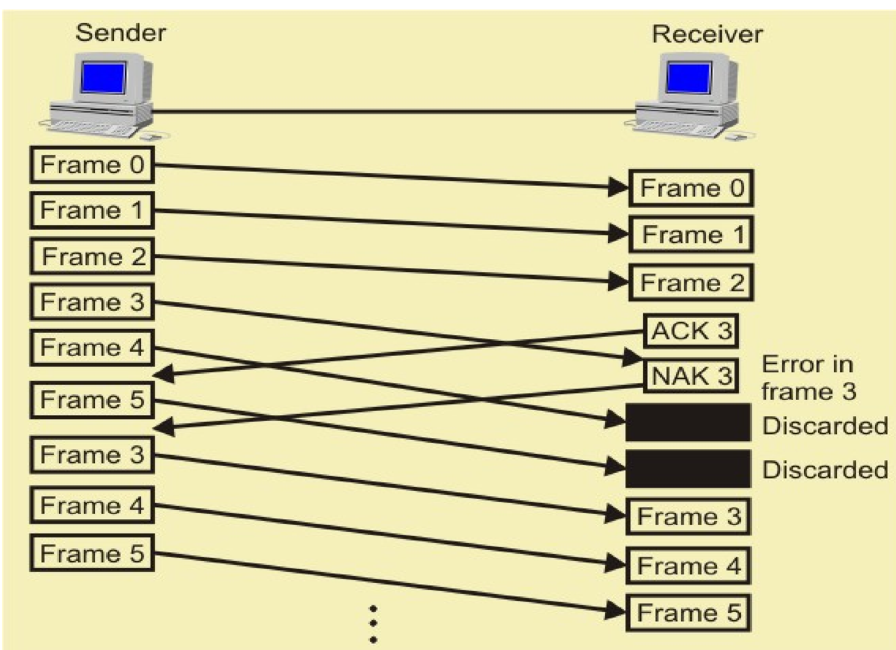

Figure 3.3.9 Frames in error in go-Back-N ARQ

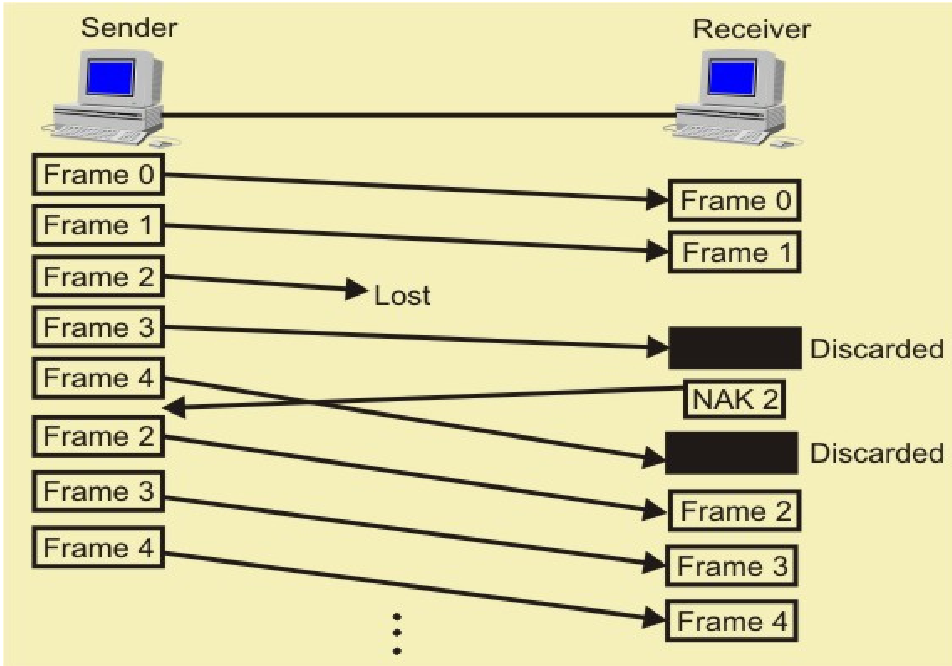

Figure 3.3.10 Lost Frames in Go-Back-N ARQ

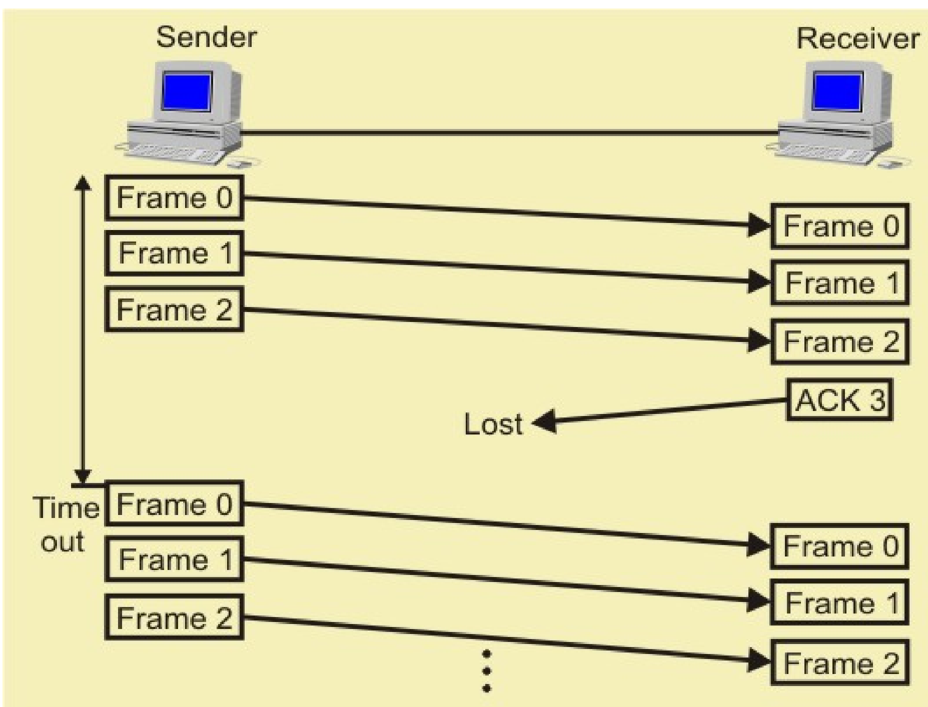

Figure 3.3.11 Lost ACK in Go-Back-N ARQ

If no acknowledgement is received after sending N frames, the sender takes the help of a timer. After the time-out, it resumes retransmission. The go-back-N protocol also takes care of damaged frames and damaged ACKs. This scheme is little more complex than the previous one but gives much higher throughput.

Assuming full-duplex transmission, the receiving end sends piggybacked acknowledgement by using some number in the ACK field of its data frame. Let us assume that a 3-bit sequence number is used and suppose that a station sends frame 0 and gets back an RR1, and then sends frames 1, 2, 3, 4, 5, 6, 7, 0 and gets another RR1.This might either mean that RR1 is a cumulative ACK or all 8 frames were damaged. This ambiguity can be overcome if the maximum window size is limited to 7, i.e. for a k-bit sequence number field it is limited to 2^k-1. The number N (=2^k-1) specifies how many frames can be sent without receiving acknowledgement. If no acknowledgement is received after sending N frames, the sender takes the help of a timer. After the time-out, it resumes retransmission. The go-back-N protocol also takes care of damaged frames and damaged ACKs. This scheme is little more complex than the previous one but gives much higher throughput.

3.3.3.3 Selective-Repeat ARQ

The selective-repetitive ARQ scheme retransmits only those for which NAKs are received or for which timer has expired, this is shown in the Fig.3.3.12. This is the most efficient among the ARQ schemes, but the sender must be more complex so that it can send out-of-order frames. The receiver also must have storage space to store the post-NAK frames and processing power to reinsert frames in proper sequence.

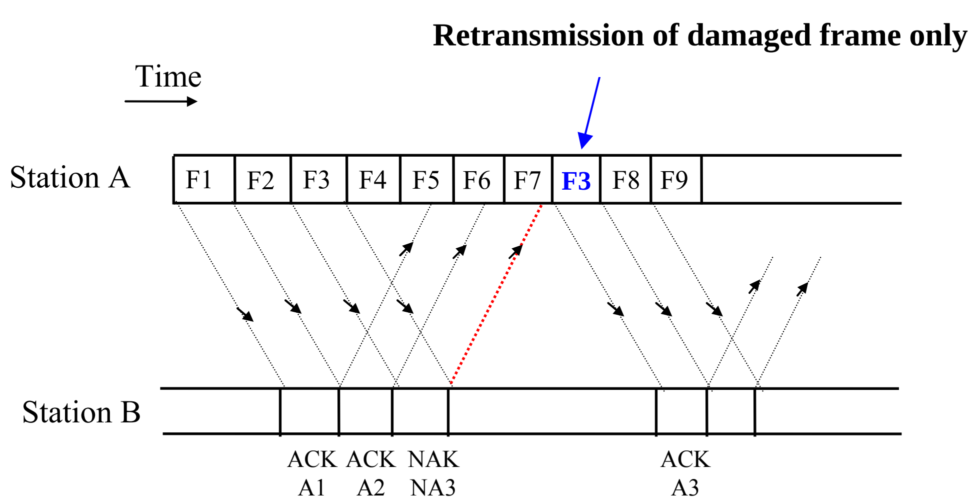

Figure 3.3.12 Selective-repeat Reject

Fill In The Blanks

1. _________Control is a technique for speed-matching of transmitter and receiver.
2. ________ Control provides for the addition of binary digits (redundant bits) that can be used to identify if there has been an error in the transmission of one or more bits.
3. There are two methods developed for flow control namely _____________ and ____________.
4. Stop-And-Wait is also known as ____________.
5. Sliding Window Flow Control allows transmission of ____________ frames.
6. Sliding Window Flow Control assigns each frame a _____-bit sequence number.
7. Sliding window ARQ can be of two types namely, ____________ and ____________.

Short Answer Questions

1. What are the key functions of error control techniques?

Ans: There are basically two types of errors, namely, (a) Damaged Frame (b) Lost Frame. The key functions for error control techniques are as follows:

- Error detection
- Sending of positive acknowledgement (ACK) by the receiver for no error
- Sending of negative acknowledgement (NAK) by the receiver for error
- Setting of timer for lost frame
- Numbering of frames

2. Why is flow control needed?

Ans: In case of data communication between a sender and a receiver, it may so happen that the rate at which data is transmitted by a fast sender is not acceptable by a slow receiver. IN such a situation, there is a need of flow control so that a fast transmitter does not overwhelm a slow receiver.

3. Mention key advantages and disadvantages of stop-and-wait ARQ technique?

Ans: Advantages of stop-and-wait ARQ are:

a. Simple to implement
b. Frame numbering is modulo-2, i.e. only 1 bit is required.

The main disadvantage of stop-and-wait ARQ is that when the propagation delay is long, it is extremely inefficient.

4. Consider the use of 10 K-bit size frames on a 10 Mbps satellite channel with 270 ms delay. What is the link utilization for stop-and-wait ARQ technique assuming P = 10^{-3}?

Ans: Link utilization = (1-P) / (1+2a)
Where a = (Propagation Time) / (Transmission Time)

Propagation time = 270 msec
Transmission time = (frame length) / (data rate)
= (10 K-bit) / (10 Mbps)
= 1 msec

Hence, a = 270/1 = 270
Link utilization = $0.999/(1+2*270) \approx 0.0018 = 0.18\%$

5. What is the channel utilization for the go-back-N protocol with window size of 7 for the problem 3?

Ans: Channel utilization for go-back-N

$$= N(1 - P) / (1 + 2a)(1-P+NP)$$

P = probability of single frame error $\approx 10^{-3}$
Channel utilization $\approx 0.01285 = 1.285\%$

6. In what way selective-repeat is better than go-back-N ARQ technique?

Ans : In selective-repeat scheme only the frame in error is retransmitted rather than transmitting all the subsequent frames. Hence it is more efficient than go-back-N ARQ technique.

7. In what situation Stop-and-Wait protocol works efficiently?

Ans: In case of Stop-and-Wait protocol, the transmitter after sending a frame waits for the acknowledgement from the receiver before sending the next frame. This protocol works efficiently for long frames, where propagation time is small compared to the transmission time of the frame.

8. How the inefficiency of Stop-and-Wait protocol is overcome in sliding window protocol?

Ans: The Stop-and-Wait protocol is inefficient when large numbers of small packets are send by the transmitter since the transmitter has to wait for the acknowledgement of each individual packet before sending the next one. This problem can be overcome by sliding window protocol. In sliding window protocol multiple frames (up to a fixed number of frames) are send before receiving an acknowledgement from the receiver.

9. What is piggybacking? What is its advantage?

Ans: In practice, the link between receiver and transmitter is full duplex and usually both transmitter and receiver stations send data to each over. So, instead of sending separate acknowledgement packets, a portion (few bits) of the data frames can be used for acknowledgement. This phenomenon is known as piggybacking.

The piggybacking helps in better channel utilization. Further, multi-frame acknowledgement can be done.

10. For a k-bit numbering scheme, what is the range of sequence numbers used in sliding window protocol?

Ans: For k-bit numbering scheme, the total number of frames, N, in the sliding window can be given as follows (using modulo-k).

$N = 2^k - 1$

Hence the range of sequence numbers is: 0, 1, 2, and 3 … $2^k - 1$

Module 5

Broadcast Communication Networks

Lesson 1

Network Topology

Specific Instructional Objectives

At the end of this lesson, the students will be able to:

- Specify what is meant by network topology
- Classify different Network topologies
- Categorize various Network topologies
- Explain the characteristics of the following topologies:
    - Mesh
    - Bus
    - Star
    - Ring
    - Tree
    - Unconstrained

5.1.1 Introduction

Topology refers to the way in which the network of computers is connected. Each topology is suited to specific tasks and has its own advantages and disadvantages. The choice of topology is dependent upon type and number of equipment being used, planned applications and rate of data transfer required, response time, and cost. Topology can also be defined as the *geometrically interconnection pattern* by which the stations (nodes/computers) are connected using suitable transmission media (which can be point-to-point and broadcast). Various commonly used topologies are discussed in the following sections.

5.1.2 Mesh Topology

In this topology each node or station is connected to every other station as shown in Fig. 5.1.1. The key characteristics of this topology are as follows:

Key Characteristics:

- Fully connected
- Robust – Highly reliable
- Not flexible
- Poor expandability

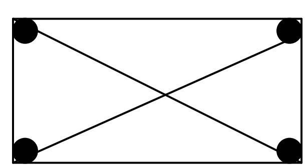

Figure 5.1.1 Mesh Topology

Two nodes are connected by dedicated point-point links between them. So the total number of links to connect n nodes = n(n-1)/2; which is proportional to n^2 . Media used for the connection (links) can be twisted pair, co-axial cable or optical fiber. With this topology there is no need to provide any additional information, that is from where the packet is coming, along with the packet because two nodes have a point-point dedicated link between them. And each node knows which link is connected to which node on the other end.

Mesh Topology is not flexible and has a poor expandability as to add a new node n links have to be laid because that new node has to be connected to each of the existing nodes via dedicated link as shown in Fig. 5.1.2. For the same reason the cost of cabling will be very high for a larger area. And due to these reasons this topology is rarely used in practice.

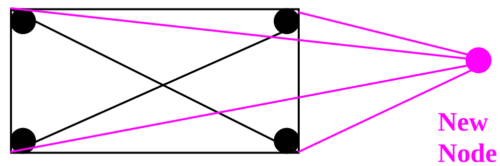

Figure 5.1.2 Adding a new node in Mesh Topology

5.1.3 Bus Topology

In Bus Topology, all stations attach through appropriate hardware interfacing known as a *tap*, directly to a linear transmission medium, or bus as shown in Fig. 5.1.3. 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 there is a *terminator*, which absorbs any signal, preventing reflection of signal from the endpoints. If the terminator is not present, the endpoint acts like a mirror and reflects the signal back causing interference and other problems.

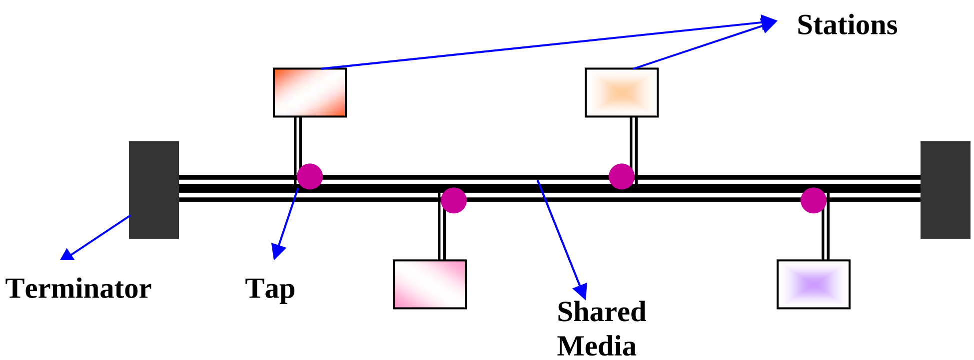

Figure 5.1.3 Bus Topology

Key Characteristics of this topology are:

- Flexible
- Expandable
- Moderate Reliability
- Moderate performance

A shared link is used between different stations. Hence it is very cost effective. One can easily add any new node or delete any node without affecting other nodes; this makes this topology easily expandable. Because of the shared medium, it is necessary to provide some extra information about the desired destination, i.e. to explicitly specify the destination in the packet, as compared to mesh topology. This is because the same medium is shared among many nodes. As each station has a unique address in the network, a station copies a packet only when the destination address of the packet matches with the self-address. This is how data communications take place among the stations on the bus.

As there are dedicated links in the mess topology, there is a possibility of transferring data in parallel. But in bus topology, only one station is allowed to send data at a time and all other stations listen to it, as it works in a broadcast mode. Hence, only one station can transfer the data at any given time. Suitable medium access control technique should be used so as to provide some way to decide "who" will go next to send data? Usually a distributed medium access control technique, as discussed in the next lesson, is used for this purpose.

As the distance through which signal traverses increases, the attenuation increases. If the sender sends data (signal) with a small strength signal, the farthest station will not be able to receive the signal properly. While on the other hand if the transmitter sends the signal with a larger strength (more power) then the farthest station will get the signal properly but the station near to it may face over-drive. Hence, delay and signal unbalancing will force a maximum length of shared medium, which can be used in bus topology.

5.1.4 STAR Topology

In the star topology, each station is directly connected to a common central node as shown in Fig. 5.1.4. Typically, each station attaches to a central node, referred to as the *star coupler*, via two point-to-point links, one for transmission and one for reception.

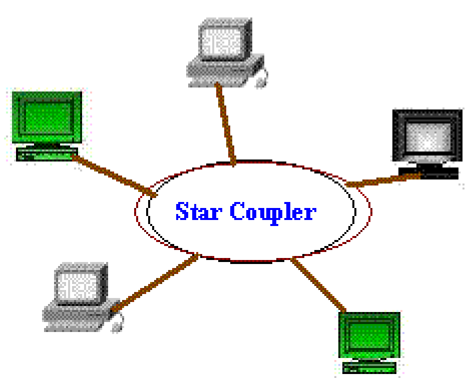

Figure 5.1.4 Star Topology

Key features:

- High Speed
- Very Flexible
- High Reliability
- High Maintainability

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. In this case the central node acts as a *repeater*.

- 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. In this approach, the central node acts as a *switch* and performs the switching or routing function. This mode of operation can be compared with the working of a telephone exchange, where the caller party is connected to a single called party and each pair of subscriber who needs to talk have a different connection.

Very High speeds of data transfer can be achieved by using star topology, particularly when the star coupler is used in the switch mode. This topology is the easiest to maintain, among the other topologies. As the number of links is proportional to n, this topology is very flexible and is the most preferred topology.

5.1.5 Ring topology

In the ring topology, the network consists of a set of repeaters joined by point-to-point links in a closed loop as shown in Fig. 5.1.5. 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).

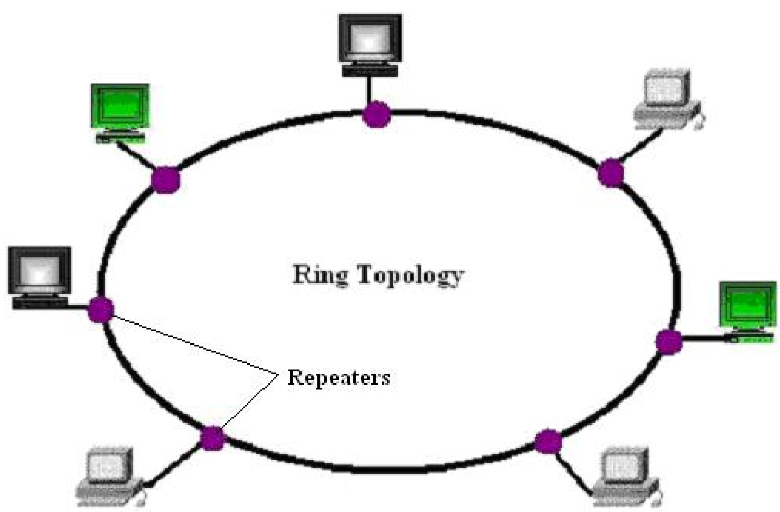

Figure 5.1.5 Ring Topology

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. Because multiple stations share the ring, medium access control is needed to determine at what time each station may insert frames.

How the source knows whether it has to transmit a new packet and whether the previous packet has been received properly by the destination or not. For this, the destination change a particular bit (bits) in the packet and when the receiver sees that packet with the changed bit, it comes to know that the receiver has received the packet.

This topology is not very reliable, because when a link fails the entire ring connection is broken. But reliability can be improved by using *wiring concentrator*, which helps in bypassing a faulty node and somewhat is similar to star topology.

Repeater works in the following three modes:

- **Listen mode**: In this mode, the station listens to the communication going over the shared medium as shown in Fig.5.1.6.

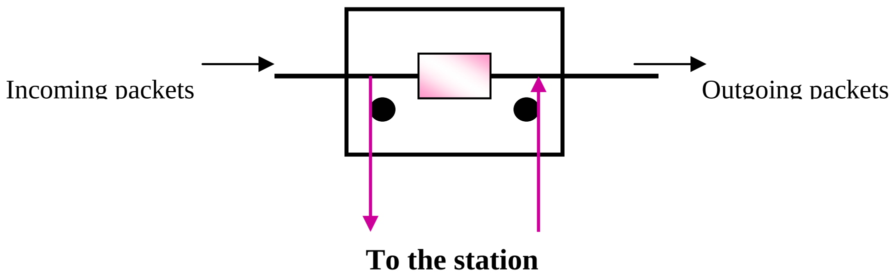

Figure 5.1.6 Repeater in Listen Mode

- **Transmit mode**: In this mode the station transmit the data over the network as shown in Fig. 5.1.7.

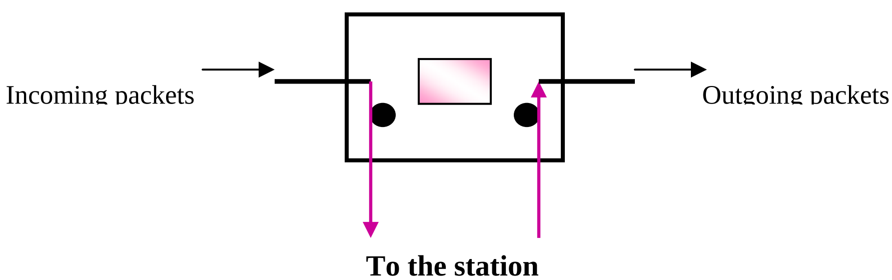

Figure 5.1.7 Repeater in Transmit Mode

- **By-Pass mode**: When the node is faulty then it can be bypassed using the repeater in bypass mode, i.e. the station doesn't care about what data is transmitted through the network, as shown in Fig. 5.1.8. In this mode there is no delay introduced because of this repeater.

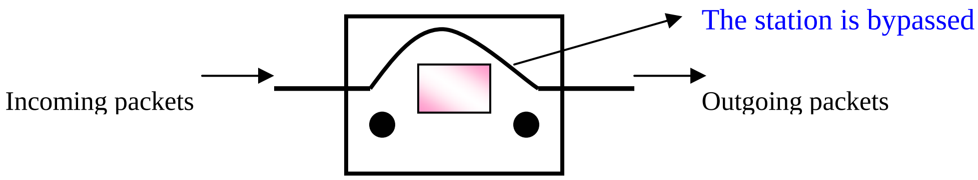

Figure 5.1.8 Repeater in Bypass Mode

5.1.6 Tree Topology

This topology can be considered as an extension to bus topology. It is commonly used in cascading equipments. For example, you have a repeater box with 8-port, as far as you have eight stations, this can be used in a normal fashion. But if you need to add more stations then you can connect two or more repeaters in a hierarchical format (tree format) and can add more stations. In the Fig. 5.1.9, R1 refers to repeater one and so on and each repeater is considered to have 8-ports.

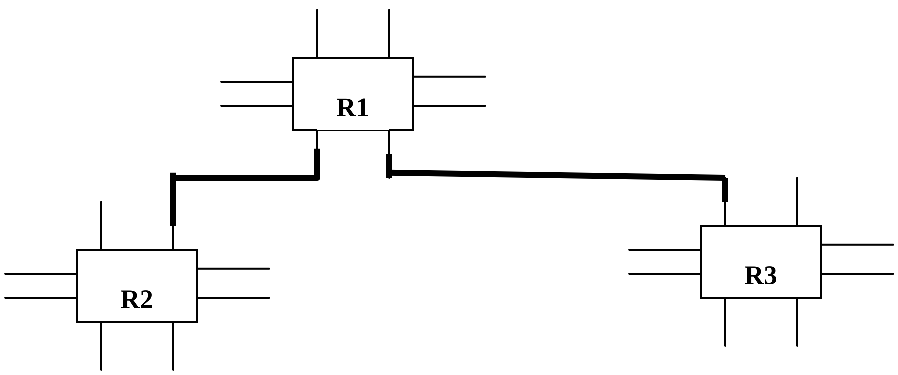

Figure 5.1.9 Tree Topology

This tree topology is very good in an organization as incremental expansion can be done in this way. Main features of this topology are scalability and flexibility. This is because, when the need arises for more stations that can be accomplished easily without affecting the already established network.

5.1.7 Unconstrained Topology

All the topologies discussed so far are symmetric and constrained by well-defined interconnection pattern. However, sometimes no definite pattern is followed and nodes are interconnected in an arbitrary manner using point-to-point links as shown in Fig 5.1.10. Unconstrained topology allows a lot of configuration flexibility but suffers from the complex routing problem. Complex routing involves unwanted overhead and delay.

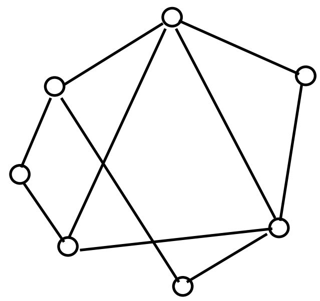

Figure 5.1.10 Unconstrained Topology

5.1.8 Combination of topology and transmission media

Topology and transmission media are interrelated. For example, all the important criteria of a network such as reliability, expandability and performance depend on both the topology and the transmission media used in the network. As a consequence, these two aspects are interrelated. Let us have a look at the various transmission media, which are used for different topologies.

- Twisted pair is suitable for use in star and ring topologies
  - *Cat 3*: voice grade UTP, data rate up to 10 Mbps
  - *Cat 5*: data grade UTP, data rate up to 100 Mbps
- Coaxial cable is suitable for use in bus topology
  - Baseband coaxial cable supports data rates of 20 Mbps at distances up to 2 Km.
- Fiber optics is suitable for use in ring and star topology
  - Gigabit data rates and longer distances.
- Unguided media are suitable for star topology

Fill In The Blanks.

1. Number of links to connect n nodes in a mesh topology is = _______.
2. Mesh Topology is _______ flexible and has a ________ expandability
3. In BUS topology, at each end of the bus is a ____________, which absorbs any signal, removing it from the bus.
4. In BUS topology, One can easily add any new node or delete any node with-out affecting other nodes; this makes this topology easily _____________.
5. _________ and ______________ will force a maximum length of shared medium which can be used in BUS topology.
6. The two alternatives for the operation of the central node in STAR topology are: __________ and ___________.
7. In Ring Topology, the links are _____________; that is, data are transmitted in __________ direction only and all are oriented in the same way

8. In Ring Topology, Repeater works in 3 modes: ____________, _____________ and _______________.
9. ____________ topology can be considered as an extension to BUS topology.
10. _______________ is suitable for use in star and ring topologies
11. Coaxial cable is suitable for use in ___________ topology.

Solutions.

1. n(n-1)/2
2. not, poor
3. terminator
4. expandable.
5. Delay, signal unbalancing
6. repeater, switch
7. unidirectional, one
8. Listen, Transmit, By-Pass
9. Tree
10. Twisted pair
11. BUS

Short Answer Questions:

Q-1. List out the advantages and drawbacks of bus topology.

Ans: Advantages:

i) Easy to implement
ii) It is very cost effective because only a single segment required
iii) It is very flexible
iv) Moderate reliability.
v) Can add new station or delete any station easily (scalable)

Disadvantages:

i) Required suitable medium access control technique.
ii) Maximum cable length restriction imposed due to delay and signal unbalancing problem.

Q-2. List out the advantages and drawbacks of ring topology.

Ans: Advantages:

i) Data insertion, data reception and data removal can be provided by repeater
ii) It can provide multicast addressing.
iii) Point-to-point links to its adjacent nodes (moderate cost)

Disadvantages:

i) The repeater introduces a delay
ii) The topology fails if any link disconnects or a node fails.
iii) Direct link not provided
iv) It provides complex management

Q-3. Why star topology is commonly preferred?

Ans: It gives high reliability, more flexible and higher bandwidth. Since there is a central control point, the control of network is easy and priority can be given to selected nodes.

Q-4. Is there any relationship between transmission media and topology?

Ans: Yes, medium should be selected based on the topology. For example, for bus topology coaxial cable medium is suitable, and for ring/star topology twisted-pair or optical fiber can be used.