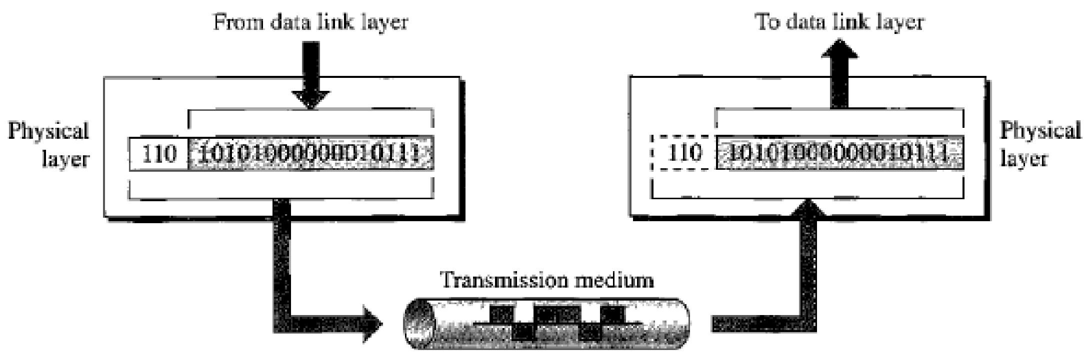
Unit 2: The Physical Layer

Functions of Physical Layer:

The physical layer coordinates the functions required to carry a bit stream over a physical medium. It deals with the mechanical and electrical specifications of the interface and transmission medium. It also defines the procedures and functions that physical devices and interfaces have to perform for transmission to occur. Figure below shows the position of the physical layer with respect to the transmission medium and the data link layer.

Figure 2.5 *Physical layer*



The physical layer is responsible for movements of individual bits from one hop (node) to the next.

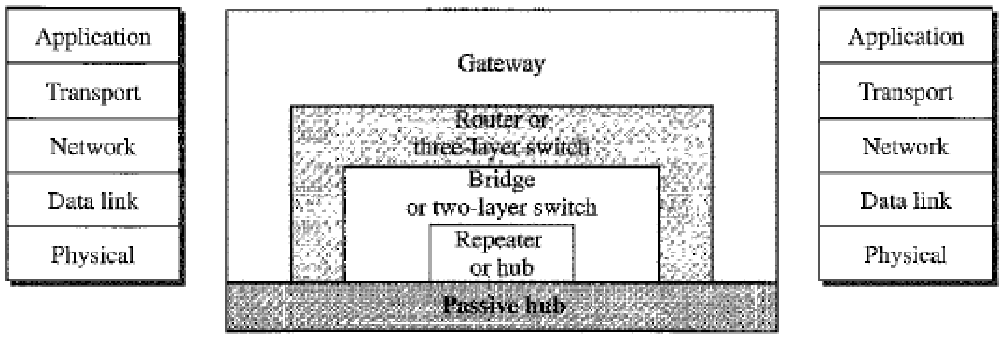
Specific responsibilities of phycial layer include:

- Physical characteristics of interfaces and medium
- Representation of bits
- Data rate
- Synchronization of bits
- Line configuration
- Physical topology
- Transmission mode

Network Devices (Connecting Devices):

We divide connecting devices into five different categories based on the layer in which they operate in a network, as shown in Figure below.

Figure 15.1 *Five categories of connecting devices*



The five categories contain devices which can be defined as:

1. Those which operate below the physical layer such as a passive hub.
2. Those which operate at the physical layer (a repeater or an active hub).
3. Those which operate at the physical and data link layers (a bridge or a two-layer switch).
4. Those which operate at the physical, data link, and network layers (a router or a three-layer switch).
5. Those which can operate at all five layers (a gateway).

Data and Signals:

Signal:

A signal is an electrical or electromagnetic current that is used for carrying data from one device or network to another. It is the key component behind data communication and networking. Signals can be periodic and nonperiodic. A periodic signal repeats the pattern over identical periods. A nonperiodic signal changes without repeating a pattern or cycle over time.

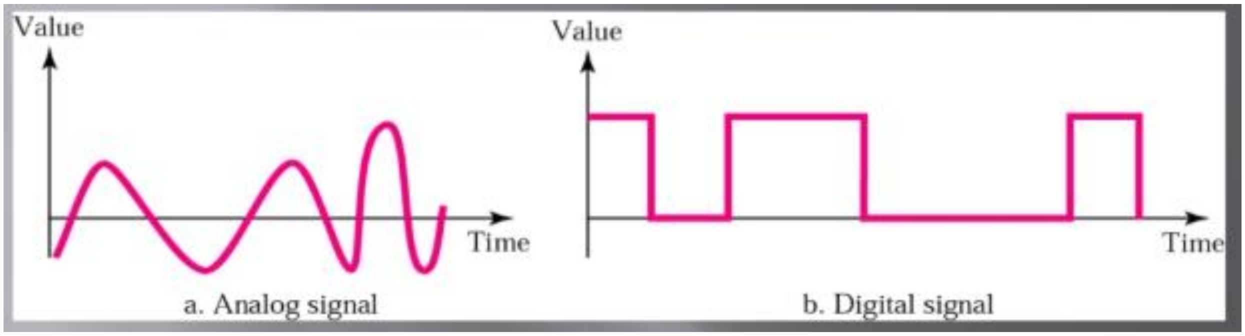
A signal can be either analog or digital.

Analog Signal:

Analog signal is a continuous wave that keeps on changing over a time period. In other words, an analog signal is a continuous wave denoted by a sine wave and may vary in signal strength (amplitude) or frequency (waves per unit time). Analog signals can be classified as simple or composite. A simple analog signal or sine wave cannot be further decomposed into simpler signals. A composite analog signal is composed of multiple sine waves.

Digital Signal:

Digital signals also carry information like analog signals but is somewhat is different from analog signals. Digital signal is noncontinuous, discrete time signal. Digital signal carries information or data in the binary form i.e. a digital signal represent information in the form of bits (0s and 1s). Digital signals are easier to transmit and are more reliable when compared to analog signals.



Key Differences Between Analog and Digital Signal:

- An analog signal represents a continuous wave that keeps changing over a time period. On the other hand, a digital signal represents a noncontinuous wave that carries information in a binary format and has discrete values.
- An analog signal is always represented by the continuous sine wave whereas, a digital signal is represented by square waves.
- While talking of analog signal, we describe the behavior of the wave in respect of amplitude, period or frequency, and phase of the wave. On the other hand, while talking of discrete signals we describe the behavior of the wave in respect of bit rate and bit interval.
- The range of an analog signal is not fixed whereas the range of the digital signal is finite and which can be 0 or 1.
- An analog signal is more prone to distortion in response to noise, but a digital signal has immunity in response to noise hence it rarely faces any distortion.
- An analog signal transmits data in the form of wave whereas, a digital signal transmits the data in the binary form i.e. in the form of bits.
- The best example of an analog signal is a human voice, and the best example of a digital signal is the transmission of data in a computer.

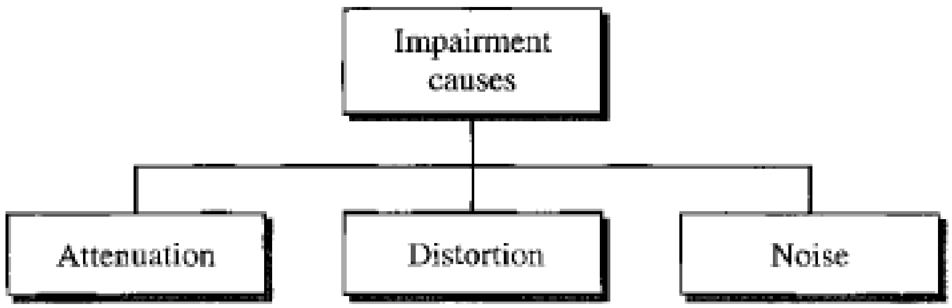
Analog Data:
The term analog data refers to information that is continuous; For example, an analog clock that has hour, minute, and second hands gives information in a continuous form; the movements of the hands are continuous. Analog data, such as the sounds made by a human voice, take on continuous values. When someone speaks, an analog wave is created in the air. This can be captured by a microphone and converted to an analog signal or sampled and converted to a digital signal.

Digital Data:
Digital data refers to information that has discrete states. For example, a digital clock that reports the hours and the minutes will change suddenly from 8:05 to 8:06. Digital data takes on discrete values. For example, data are stored in computer memory in the form of 0s and 1s. They can be converted to a digital signal or modulated into an analog signal for transmission across a medium.

Transmission Impairment:

Signals travel through transmission media, which are not perfect. The imperfection causes signal impairment. This means that the signal at the beginning of the medium is not the same as the signal at the end of the medium. What is sent is not what is received. Three causes of impairment are attenuation, distortion, and noise.

Causes of impairment



Attenuation

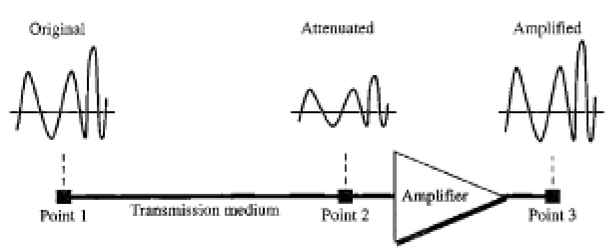
Attenuation means a loss of energy. When a signal, simple or composite, travels through a medium, it loses some of its energy in overcoming the resistance of the medium. That is why a wire carrying electric signals gets warm, if not hot, after a while. Some of the electrical energy in the signal is converted to heat. To compensate for this loss, amplifiers are used to amplify the signal. Attenuation is measured in terms of Decibels.

The decibel (dB) measures the relative strengths of two signals or one signal at two different points. Note that the decibel is negative if a signal is attenuated and positive if a signal is amplified.

$dB=10\log_{10} P2/P1$

Variables P1 and P2 are the powers of a signal at points 1 and 2, respectively.

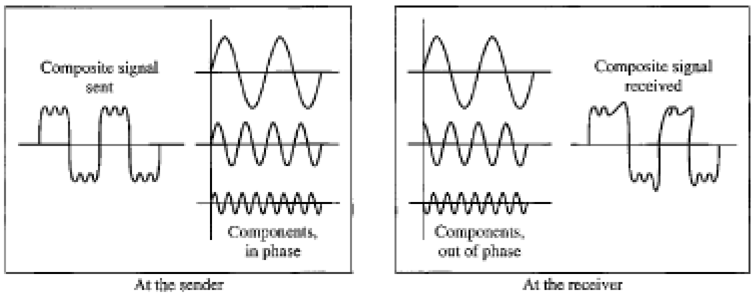
Figure 3.26 *Attenuation*



Distortion:

Distortion means that the signal changes its form or shape. Distortion can occur in a composite signal made of different frequencies. Each signal component has its own propagation speed through a medium and, therefore, its own delay in arriving at the final destination. Differences in delay may create a difference in phase if the delay is not exactly the same as the period duration. In other words, signal components at the receiver have phases different from what they had at the sender. The shape of the composite signal is therefore not the same. Figure shows the effect of distortion on a composite signal.

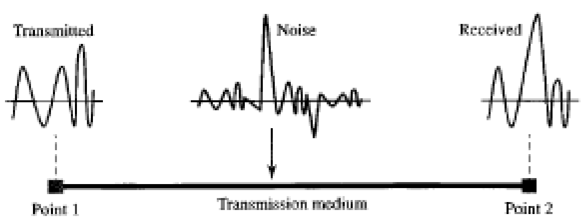
Figure 3.28 *Distortion*



Noise

Noise is another cause of impairment. Several types of noise, such as thermal noise, induced noise, crosstalk, and impulse noise, may corrupt the signal. Thermal noise is the random motion of electrons in a wire which creates an extra signal not originally sent by the transmitter. Induced noise comes from sources such as motors and appliances. These devices act as a sending antenna, and the transmission medium acts as the receiving antenna. Crosstalk is the effect of one wire on the other. One wire act as a sending antenna and the other as the receiving antenna. Impulse noise is a spike (a signal with high energy in a very short time) that comes from power lines, lightning, and so on.

Figure 3.29 *Noise*



Signal-to-Noise Ratio (SNR)

The signal-to-noise ratio is defined as

SNR= Average Signal power / Average Noise Power

SNR is actually the ratio of what is wanted (signal) to what is not wanted (noise). A high SNR means the signal is less corrupted by noise; a low SNR means the signal is more corrupted by noise. Because SNR is the ratio of two powers, it is often described in decibel units, SNR dB, defined as

$SNR_{dB} = 10\log_{10} SNR$

Data Rate Limits

Data rate can be defined as how fast can we send the data, in bits per second, over a channel. Maximum Data Rate (Channel Capacity) is the tight upper bound on the rate at which information can be reliably transmitted over a communication channel.
Data rate depends on three factors:

- The bandwidth available
- The level of the signals we use
- The quality of the channel (the level of noise)

There are two theoretical formulas to calculate the data rate:

- Nyquist for a noiseless channel
- Shannon for a noisy channel

Noiseless Channel: Nyquist Bit Rate

For a noiseless channel, the Nyquist bit rate formula defines the theoretical maximum bit rate
BitRate = 2 * Bandwidth * $\log_2(L)$
In the above equation, bandwidth is the bandwidth of the channel, L is the number of signal levels used to represent data, and BitRate is the bit rate in bits per second.
Bandwidth is a fixed quantity, so it cannot be changed. Hence, the data rate is directly proportional to the number of signal levels.
Increasing the levels of a signal may reduce the reliability of the system.

Noisy Channel: Shannon Capacity

In reality, we cannot have a noiseless channel; the channel is always noisy. Shannon capacity is used, to determine the theoretical highest data rate for a noisy channel:
Capacity = bandwidth * $\log_2(1 + SNR)$
In the above equation, bandwidth is the bandwidth of the channel, SNR is the signal-to-noise ratio, and capacity is the capacity of the channel in bits per second.
Bandwidth is a fixed quantity, so it cannot be changed. Hence, the channel capacity is directly proportional to the power of the signal, as SNR = (Power of signal) / (power of noise).
The signal-to-noise ratio (S/N) is usually expressed in decibels (dB) given by the formula:
SNR dB=10*$\log_{10}$ (SNR)

Performance

One important issue in networking is the performance of the network-how good it is? It can be referred as Quality of Service (QoS) that is an overall measurement of the network performance. There are four factors of determining network performance.

- Bandwidth
- Throughput

- Delay
- Jitter

Bandwidth

One characteristic that measures network performance is bandwidth. However, the term can be used in two different contexts with two different measuring values: bandwidth in hertz and bandwidth in bits per second.

- The first, bandwidth in hertz, refers to the range of frequencies in a composite signal or the range of frequencies that a channel can pass. For example, we can say the bandwidth of a subscriber telephone line is 4 kHz.
- The second, bandwidth in bits per second, refers to the speed of bit transmission in a channel or link. For example, one can say the bandwidth of a Fast Ethernet network (or the links in this network) is a maximum of 100 Mbps. This means that this network can send 100 Mbps.

Throughput

- The throughput is a measure of how fast we can actually send data through a network.
- Although, at first glance, bandwidth in bits per second and throughput seem the same, they are different.
- A link may have a bandwidth of B bps, but we can only send T bps through this link with T always less than B.
- In other words, the bandwidth is a potential measurement of a link; the throughput is an actual measurement of how fast we can send data.
- For example, we may have a link with a bandwidth of 1 Mbps, but the devices connected to the end of the link may handle only 200 kbps. This means that we cannot send more than 200 kbps through this link.

Delay (Latency)

- The latency or delay defines how long it takes for an entire message to completely arrive at the destination from the time the first bit is sent out from the source.
- We can say that latency is made of four components: propagation time, transmission time, queuing time and processing delay.

  Latency =propagation time +transmission time +queuing time + processing delay

- Propagation time measures the time required for a bit to travel from the source to the destination. The propagation time is calculated by dividing the distance by the propagation speed.

  Propagation Time= Distance / Propagation Speed

- In data communications we don't send just 1 bit, we send a message. The first bit may take a time equal to the propagation time to reach its destination; the last bit also may take the same amount of time. However, there is a time between the first bit leaving the sender and the last bit arriving at the receiver. The first bit leaves earlier and arrives earlier; the last bit leaves later and arrives later. The time required for transmission of a message depends on the size of the message and the bandwidth of the channel.

Transmission Time= Message Size / Bandwidth

- The third component in latency is the queuing time, the time needed for each intermediate or end device to hold the message before it can be processed. The queuing time is not a fixed factor; it changes with the load imposed on the network. When there is heavy traffic on the network, the queuing time increases. An intermediate device, such as a router, queues the arrived messages and processes them one by one. If there are many messages, each message will have to wait.

Jitter

- Another performance issue that is related to delay is jitter. We can roughly say that jitter is a problem if different packets of data encounter different delays and the application using the data at the receiver site is time-sensitive (audio and video data, for example).
- For Example, If the delay for the first packet is 20 ms, for the second is 45 ms, and for the third is 40 ms, then the real-time application that uses the packets endures jitter.

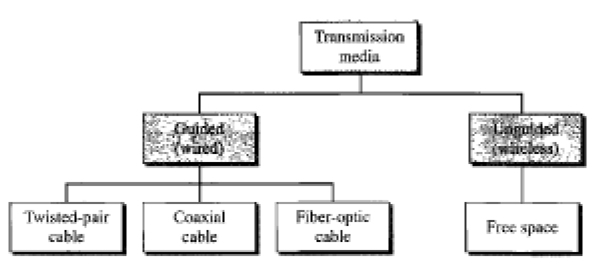
TRANSMISSION MEDIA:

A transmission medium can be broadly defined as anything that can carry information from a source to a destination. For example, the transmission medium for two people having a dinner conversation is the air. For a written message, the transmission medium might be a mail carrier, a truck, or an airplane.

In data communications, the definition of the information and the transmission medium is more specific. The transmission medium is usually free space, metallic cable, or fiber-optic cable. The information is usually a signal that is the result of a conversion of data from another form.

Classification of Transmission Media:

Figure 7.2 *Classes of transmission media*



Guided Media

Guided media, which are those that provide a channel from one device to another, include twisted-pair cable, coaxial cable, and fiber-optic cable. A signal traveling along any of these media is directed and contained by the physical limits of the medium. Twisted-pair and coaxial cable use metallic (copper) conductors that accept and transport signals in the form of electric current. Optical fiber is a cable that accepts and transports signals in the form of light.

1. Twisted-Pair Cable

A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together, as shown in Figure below.

Figure 7.3 *Twisted-pair cable*

One of the wires is used to carry signals to the receiver, and the other is used only as a ground reference. The receiver uses the difference between the two. In addition to the signal sent by the sender on one of the wires, interference (noise) and crosstalk may affect both wires and create unwanted signals. If the two wires are parallel, the effect of these unwanted signals is not the same in both wires because they are at different locations relative to the noise or crosstalk sources (e,g., one is closer and the other is farther). This results in a difference at the receiver.

By twisting the pairs, a balance is maintained. For example, suppose in one twist, one wire is closer to the noise source and the other is farther; in the next twist, the reverse is true. Twisting makes it probable that both wires are equally affected by external influences (noise or crosstalk). This means that the receiver, which calculates the difference between the two, receives no unwanted signals. The unwanted signals are mostly canceled out. From the above discussion, it is clear that the number of twists per unit of length (e.g., inch) has some effect on the quality of the cable.

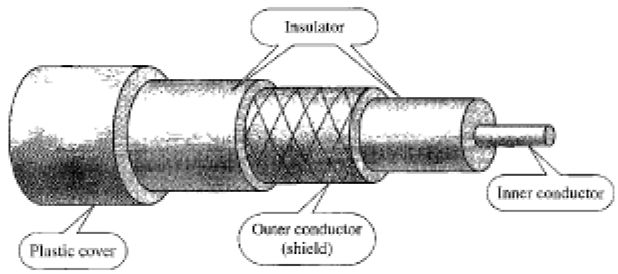
Applications

Twisted-pair cables are used in telephone lines to provide voice and data channels. The local loop-the line that connects subscribers to the central telephone office-commonly consists of unshielded twisted-pair cables. The DSL lines that are used by the telephone companies to provide high-data-rate connections also use the high-bandwidth capability of unshielded twisted-pair cables. Local-area networks also use twisted-pair cables.

2. Coaxial Cable

Coaxial cable (or coax) carries signals of higher frequency ranges than those in twisted pair cable, in part because the two media are constructed quite differently. Instead of having two wires, coax has a central core conductor of solid or stranded wire (usually copper) enclosed in an insulating sheath, which is, in turn, encased in an outer conductor of metal foil, braid, or a combination of the two. The outer metallic wrapping serves both as a shield against noise and as the second conductor, which completes the circuit. This outer conductor is also enclosed in an insulating sheath, and the whole cable is protected by a plastic cover (see Figure below).

Figure 7.7 *Coaxial cable*



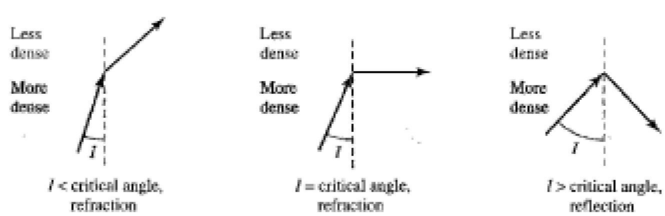
Applications

Coaxial cable was widely used in analog telephone networks where a single coaxial network could carry 10,000 voice signals. Later it was used in digital telephone networks where a single coaxial cable could carry digital data up to 600 Mbps. However, coaxial cable in telephone networks has largely been replaced today with fiber-optic cable. Cable TV networks also use coaxial cables. In the traditional cable TV network, the entire network used coaxial cable. Later, however, cable TV providers replaced most of the media with fiber-optic cable; hybrid networks use coaxial cable only at the network boundaries, near the consumer premises. Cable TV uses RG-59 coaxial cable. Another common application of coaxial cable is in traditional Ethernet LANs. Because of its high bandwidth, and consequently high data rate, coaxial cable was chosen for digital transmission in early Ethernet LANs.

3. Fiber Optic Cable:

A fiber-optic cable is made of glass or plastic and transmits signals in the form of light. To understand optical fiber, we first need to explore several aspects of the nature of light. Light travels in a straight line as long as it is moving through a single uniform If a ray of light traveling through one substance suddenly enters another substance (of a different density), the ray changes direction. Figure below shows how a ray of light changes direction when going from a more dense to a less dense substance.

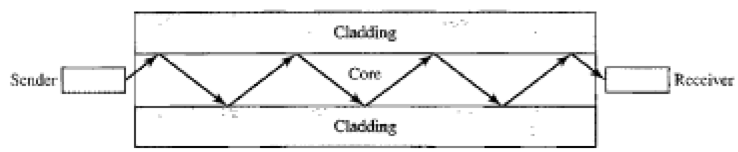
Figure 7.10 *Bending of light ray*



As the figure shows, if the angle of incidence I (the angle the ray makes with the line perpendicular to the interface between the two substances) is less than the critical angle, the ray refracts and moves closer to the surface. If the angle of incidence is equal to the critical angle, the light bends along the interface. If the angle is greater than the critical angle, the ray reflects (makes a turn) and travels again in the denser substance. Note that the critical angle is a property of the substance, and its value differs from one substance to another.

Optical fibers use reflection to guide light through a channel. A glass or plastic core is surrounded by a cladding of less dense glass or plastic. The difference in density of the two materials must be such that a beam of light moving through the core is reflected off the cladding instead of being refracted into it. See Figure below.

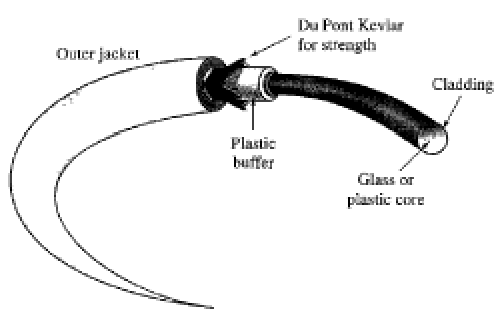
Figure 7.11 *Optical fiber*



Cable Composition

Figure below shows the composition of a typical fiber-optic cable. The outer jacket is made of either PVC or Teflon. Inside the jacket are Kevlar strands to strengthen the cable. Kevlar is a strong material used in the fabrication of bulletproof vests. Below the Kevlar is another plastic coating to cushion the fiber. The fiber is at the center of the cable, and it consists of cladding and core.

Figure 7.14 *Fiber construction*



Applications

Fiber-optic cable is often found in backbone networks because its wide bandwidth is cost-effective. Today, with wavelength-division multiplexing (WDM), we can transfer data at a rate of 1600 Gbps. Some cable TV companies use a combination of optical fiber and coaxial cable, thus creating a hybrid network. Optical fiber provides the backbone structure while coaxial cable provides the connection to the user premises. This is a cost-effective configuration since the narrow bandwidth requirement at the user end does not justify the use of optical fiber. Local-area network such as Fast Ethernet uses fiber-optic cable.

Advantages and Disadvantages of Optical Fiber

Advantages

Fiber-optic cable has several advantages over metallic cable (twisted pair or coaxial).

a. Higher bandwidth. Fiber-optic cable can support dramatically higher bandwidths (and hence data rates) than either twisted-pair or coaxial cable. Currently, data rates and bandwidth utilization over fiber-optic cable are limited not by the medium but by the signal generation and reception technology available.

b. Less signal attenuation. Fiber-optic transmission distance is significantly greater than that of other guided media. A signal can run for 50 km without requiring regeneration. We need repeaters every 5 km for coaxial or twisted-pair cable.

c. Immunity to electromagnetic interference. Electromagnetic noise cannot affect fiber-optic cables.

d. Resistance to corrosive materials. Glass is more resistant to corrosive materials than copper.

e. Light weight. Fiber-optic cables are much lighter than copper cables.

f. Greater immunity to tapping. Fiber-optic cables are more immune to tapping than copper cables. Copper cables create antenna effects that can easily be tapped.

Disadvantages

There are some disadvantages in the use of optical fiber.

a. Installation and maintenance. Fiber-optic cable is a relatively new technology. Its installation and maintenance require expertise that is not yet available everywhere.

b. Unidirectional light propagation. Propagation of light is unidirectional. If we need bidirectional communication, two fibers are needed.

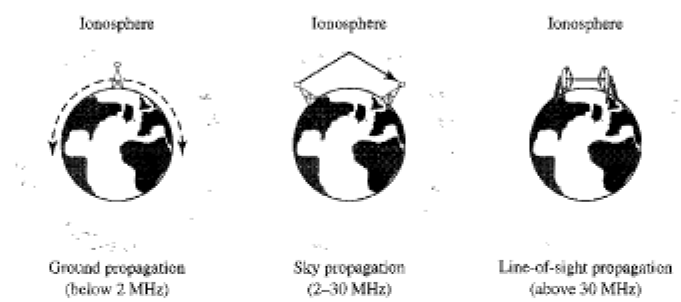
c. Cost. The cable and the interfaces are relatively more expensive than those of other guided media. If the demand for bandwidth is not high, often the use of optical fiber cannot be justified.

UNGUIDED MEDIA: WIRELESS

Unguided media transport electromagnetic waves without using a physical conductor. This type of communication is often referred to as wireless communication. Signals are normally broadcast through free space and thus are available to anyone who has a device capable of receiving them.

Unguided signals can travel from the source to destination in several ways: ground propagation, sky propagation, and line-of-sight propagation, as shown in Figure below.

Figure 7.18 *Propagation methods*

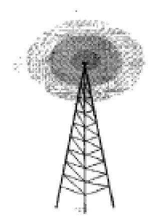


In ground propagation, radio waves travel through the lowest portion of the atmosphere, hugging the earth. These low-frequency signals emanate in all directions from the transmitting antenna and follow the curvature of the planet. Distance depends on the amount of power in the signal: The greater the power, the greater the distance. In sky propagation, higher-frequency radio waves radiate upward into the ionosphere where they are reflected back to earth. This type of transmission allows for greater distances with lower output power. In line-or-sight propagation, very high-frequency signals are transmitted in straight lines directly from antenna to antenna. Antennas must be directional, facing each other, and either tall enough or close enough together not to be affected by the curvature of the earth. Line-of-sight propagation is tricky because radio transmissions cannot be completely focused.

1. Radio Waves

Waves ranging in frequencies between 3 kHz and 1 GHz are called radio waves. Radio waves, for the most part, are omnidirectional. When an antenna transmits radio waves, they are propagated in all directions. This means that the sending and receiving antennas do not have to be aligned. A sending antenna sends waves that can be received by any receiving antenna. The omnidirectional property has a disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signals using the same frequency or band. Radio waves, particularly those waves that propagate in the sky mode, can travel long distances. This makes radio waves a good candidate for long-distance broadcasting such as AM radio. Radio waves, particularly those of low and medium frequencies, can penetrate walls. This characteristic can be both an advantage and a disadvantage. It is an advantage because, for example, an AM radio can receive signals inside a building. It is a disadvantage because we cannot isolate a communication to just inside or outside a building. The radio wave band is relatively narrow, just under 1 GHz, compared to the microwave band. When this band is divided into sub bands, the sub bands are also narrow, leading to a low data rate for digital communications.

Figure 7.20 *Omnidirectional antenna*

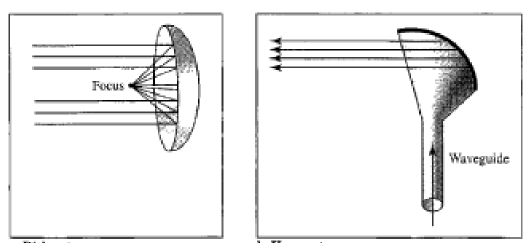


Applications: The omnidirectional characteristics of radio waves make them useful for multicasting, in which there is one sender but many receivers. AM and FM radio, television, maritime radio, cordless phones are examples of multicasting.

2. Microwaves

Electromagnetic waves having frequencies between 1 and 300 GHz are called microwaves. Microwaves are unidirectional. When an antenna transmits microwave waves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned. The unidirectional property has an obvious advantage. A pair of antennas can be aligned without interfering with another pair of aligned antennas.

Figure 7.21 *Unidirectional antennas*



a. Dish antenna　　b. Horn antenna

Microwaves need unidirectional antennas that send out signals in one direction. Two types of antennas are used for microwave communications: the parabolic dish and the horn (see Figure). A parabolic dish antenna is based on the geometry of a parabola: Every line parallel to the line of symmetry (line of sight) reflects off the curve at angles such that all the lines intersect in a common point called the focus. Outgoing transmissions are broadcast through a horn aimed at the dish. The microwaves hit the dish and are deflected outward in a reversal of the receipt path.

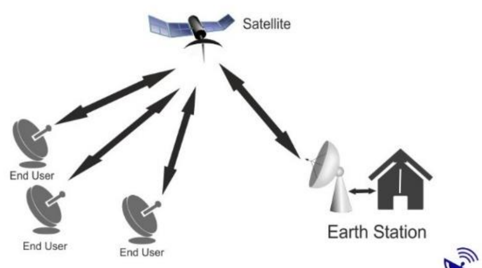
3. Infrared

Infrared waves, with frequencies from 300 GHz to 400 THz (wavelengths from 1 mm to 770 nm), can be used for short-range communication. Infrared waves, having high frequencies, cannot penetrate walls. This advantageous characteristic prevents interference between one system and another; a short-range communication system in one room cannot be affected by another system in the next room. When we use our infrared remote control, we do not interfere with the use of the remote by our neighbors. However, this same characteristic makes infrared signals useless for long-range communication. In addition, we cannot use infrared waves outside a building because the sun's rays contain infrared waves that can interfere with the communication.

The infrared band, almost 400 THz, has an excellent potential for data transmission. Such a wide bandwidth can be used to transmit digital data with a very high data rate.

VSAT:

A very small aperture terminal (VSAT) is a small telecommunication earth station that receives and transmits real-time data via satellite. VSAT is a satellite communications system that serves home and business users. A VSAT end user needs a box that interfaces between the user's computer and an outside antenna with a transceiver. The tranceiver receives or sends a signal to a satellite transponder in the sky. The satellite sends and receives signals from an earth station computer that acts as a hub for the system. For one end user to communicate with another, each transmission has to first go to the hub station which retransmits it via the satellite to the other end user's VSAT.

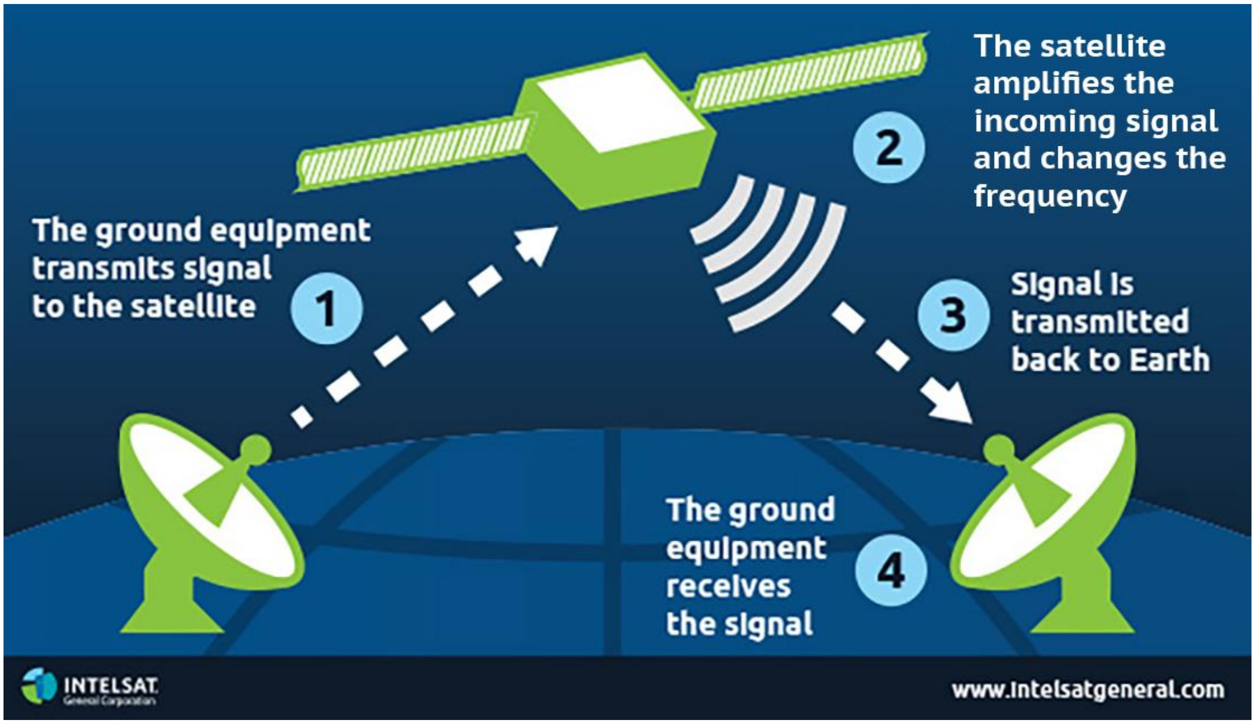


VSAT Communication System

VSAT is designed to serve both businesses and individuals and involves the use of specific technology and devices that are designed to facilitate effective telecommunications and Internet connectivity. When the system is comprised of multiple users, in order to establish communications with one another the data must be transmitted to the station-based PC which sends the signal to the sky satellite. The satellite sky transponder then forwards the data transmission to the end user’s VSAT antenna and finally to the end user’s device. VSAT can be used by both home users who sign up with a primary VSAT service and by private organizations and companies that lease or operate their own VSAT infrastructure. A main advantage of VSAT is it provides companies with complete control over their own communications infrastructure without having to depend upon third party sources.

Satellite:

A communications satellite is an artificial satellite that relays and amplifies radio telecommunications signals via a transponder; it creates a communication channel between a source transmitter and a receiver at different locations on Earth. Communications satellites are used for television, telephone, radio, internet, and military applications. The purpose of communications satellites is to relay the signal around the curve of the Earth allowing communication between widely separated geographical points.



Applications: Television, Internet, Military

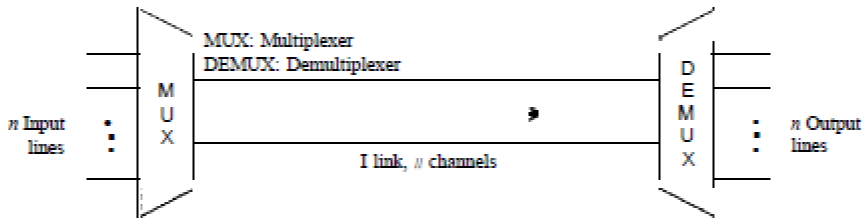
Bandwidth Utilization: Multiplexing and Spreading

- Bandwidth utilization is the wise use of available bandwidth to achieve specific goals.
- There are two broad categories of bandwidth utilization: multiplexing and spreading.
    - In multiplexing, our goal is efficiency; we combine several channels into one.
    - In spreading, our goals are privacy and antijamming; we expand the bandwidth of a channel.

Multiplexing

- Whenever the bandwidth of a medium linking two devices is greater than the bandwidth needs of the devices, the link can be shared.
- Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link.
- As data and telecommunications use increases, so does traffic.
- We can accommodate this increase by continuing to add individual links each time a new channel is needed; or we can install higher-bandwidth links and use each to carry multiple signals.
- If the bandwidth of a link is greater than the bandwidth needs of the devices connected to it, the bandwidth is wasted.
- An efficient system maximizes the utilization of all resources; bandwidth is one of the most precious resources we have in data communications.
- In a multiplexed system, n lines share the bandwidth of one link.

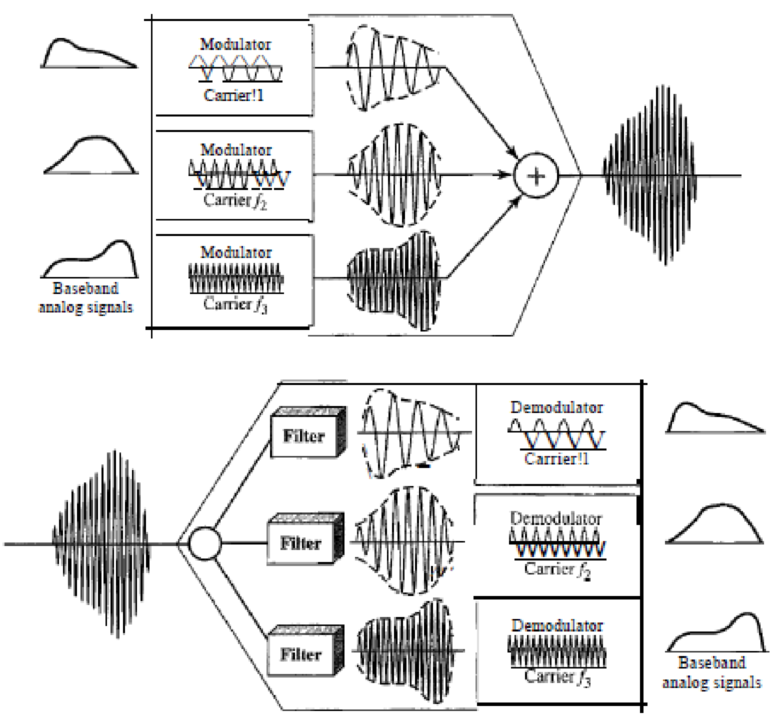
- MUX combines the streams into a single stream at the sender side whereas the DEMUX separates the streams back into its component transmissions.



- There are three basic multiplexing techniques: frequency-division multiplexing, wavelength-division multiplexing, and time-division multiplexing.
- The first two are techniques designed for analog signals, the third, for digital signals.

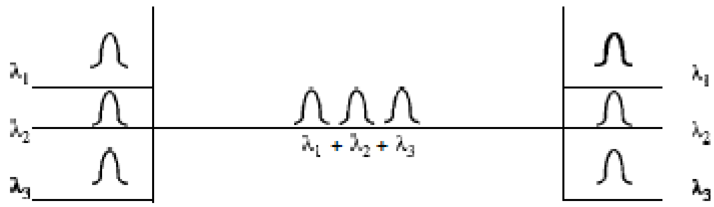
<u>Frequency-Division Multiplexing</u>

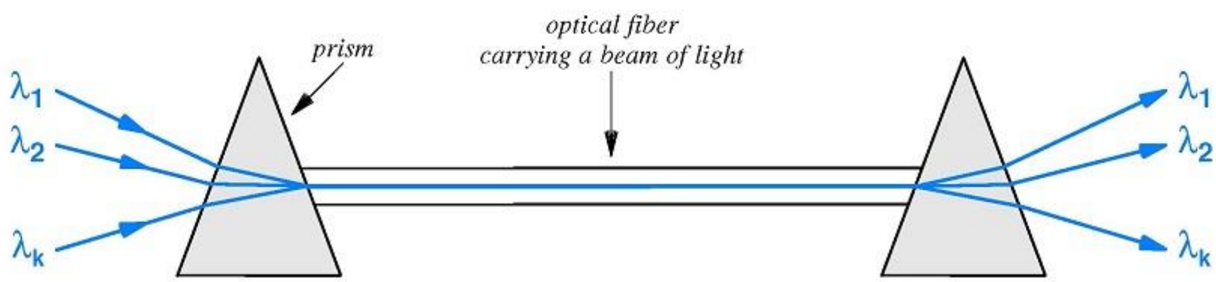
- Frequency-division multiplexing (FDM) is an analog technique that can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted.
- In FDM, signals generated by each sending device modulate different carrier frequencies. These modulated signals are then combined into a single composite signal that can be transported by the link.
- Carrier frequencies are separated by sufficient bandwidth to accommodate the modulated signal. These bandwidth ranges are the channels through which the various signals travel.
- Channels can be separated by strips of unused bandwidth-guard bands-to prevent signals from overlapping. In addition, carrier frequencies must not interfere with the original data frequencies.



Wavelength-Division Multiplexing

- Wavelength-division multiplexing (WDM) is designed to use the high-data-rate capability of fiber-optic cable.
- The optical fiber data rate is higher than the data rate of metallic transmission cable. Using a fiber-optic cable for one single line wastes the available bandwidth.
- Multiplexing allows us to combine several lines into one.
- WDM is conceptually the same as FDM, except that the multiplexing and demultiplexing involve optical signals transmitted through fiber-optic channels.
- The idea is the same: We are combining different signals of different frequencies.
- The difference is that the frequencies are very high.



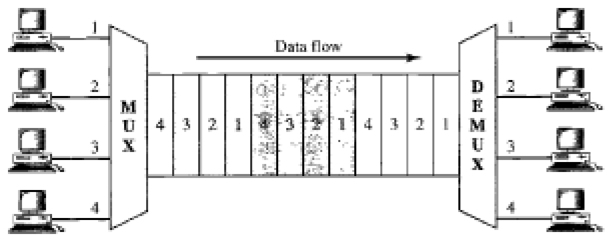


- Although WDM technology is very complex, the basic idea is very simple.
- We want to combine multiple light sources into one single light at the multiplexer and do the reverse at the demultiplexer. The combining and splitting of light sources are easily handled by a prism.
- A prism bends a beam of light based on the angle of incidence and the frequency.
- Using this technique, a multiplexer can be made to combine several input beams of light, each containing a narrow band of frequencies, into one output beam of a wider band of frequencies.
- A demultiplexer can also be made to reverse the process.

Time-Division Multiplexing (TDM)

- TDM is a digital process that allows several connections to share the high bandwidth of a line. Instead of sharing a portion of the bandwidth as in FDM, time is shared.
- Each connection occupies a portion of time in the link.
- Note that the same link is used as in FDM; here, however, the link is shown sectioned by time rather than by frequency.
- In the figure below, portions of signals 1,2,3, and 4 occupy the link sequentially.

Figure 6.12 *TDM*

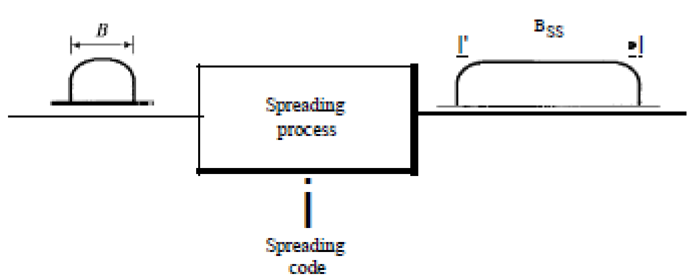


- Note that in Figure above, we are concerned with only multiplexing, not switching. This means that all the data in a message from source 1 always go to one specific destination, be it 1, 2, 3, or 4. The delivery is fixed and unvarying.

Spread Spectrum:

- Multiplexing combines signals from several sources to achieve bandwidth efficiency; the available bandwidth of a link is divided between the sources.
- In spread spectrum, we also combine signals from different sources to fit into a larger bandwidth, but our goals are somewhat different.
- Spread spectrum is designed to be used in wireless applications (LANs and WANs).
- In these types of applications, we have some concerns that outweigh bandwidth efficiency.
- In wireless applications, all stations use air (or a vacuum) as the medium for communication.
- Stations must be able to share this medium without interception by an eavesdropper and without being subject to jamming from a malicious intruder (in military operations, for example).
- To achieve these goals, spread spectrum techniques add redundancy; they spread the original spectrum needed for each station.
- If the required bandwidth for each station is B, spread spectrum expands it to B_{SS} such that B_{SS} » B.
- The expanded bandwidth allows the source to wrap its message in a protective envelope for a more secure transmission.

Spread spectrum



- Spread spectrum achieves its goals through two principles:

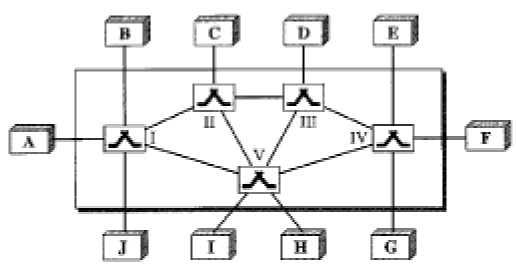
- The bandwidth allocated to each station needs to be, by far, larger than what is needed. This allows redundancy.
- The expanding of the original bandwidth B to the bandwidth B_{SS} must be done by a process that is independent of the original signal. In other words, the spreading process occurs after the signal is created by the source.

Switching:

A network is a set of connected devices. Whenever we have multiple devices, we have the problem of how to connect them to make one-to-one communication possible. One solution is to make a point-to-point connection between each pair of devices (Mesh topology) or between a central device and every other device (a star topology). These methods are not applicable for very large networks. Other topologies employing multipoint connections are also not efficient due to the distances between devices and the total number of devices increase beyond the capacities of the media and equipment.

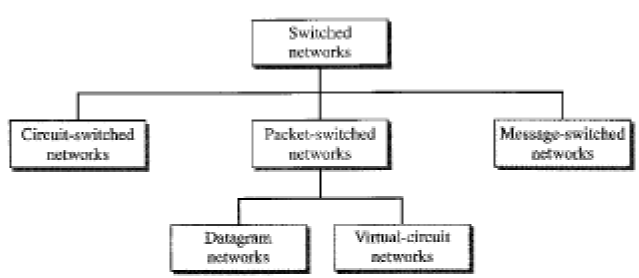
A better solution is **switching**. A switched network consists of a series of interlinked nodes, called switches. Switches are devices capable of creating temporary connections between two or more devices linked to the switch.

Switched network



We can divide today's network into three broad categories: circuit-switched networks, packet-switched networks and message-switched networks.

8.2 *Taxonomy of switched networks*

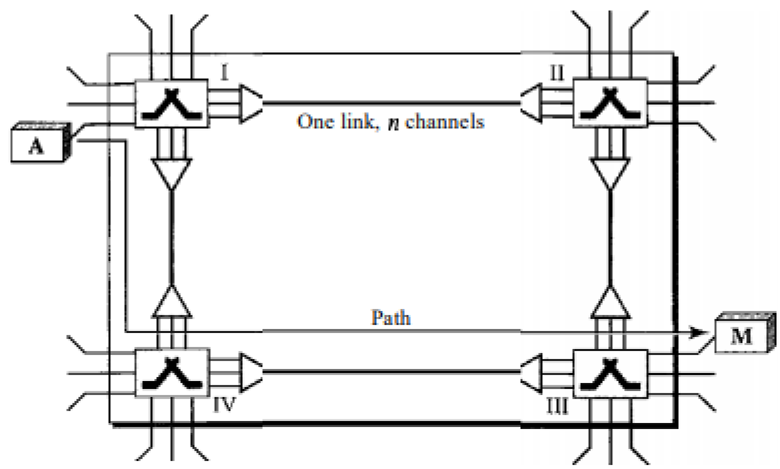


Circuit Switched Networks:

A circuit-switched network consists of a set of switches connected by physical links. A connection between two stations is a dedicated path made of one or more links. However, each connection uses only one dedicated channel on each link. Each link is normally divided into n channels by using FDM or TDM.

Figure below shows a trivial circuit-switched network with four switches and four links. Each link is divided into n (n is 3 in the figure) channels by using FDM or TDM.

A trivial circuit-switched network



We have explicitly shown the multiplexing symbols to emphasize the division of the link into channels even though multiplexing can be implicitly included in the switch fabric.

The end systems, such as computers or telephones, are directly connected to a switch. We have shown only two end systems for simplicity. When end system A needs to communicate with end system M, system A needs to request a connection to M that must be accepted by all switches as well as by M itself. This is called the setup phase; a circuit (channel) is reserved on each link, and the combination of circuits or channels defines the dedicated path. After the dedicated path made of connected circuits (channels) is established, data transfer can take place. After all data have been transferred, the circuits are torn down.

We need to emphasize several points here:

- Circuit switching takes place at the physical layer.
- Before starting communication, the stations must make a reservation for the resources to be used during the communication. These resources, such as channels (bandwidth in FDM and time slots in TDM), switch buffers, switch processing time, and switch input/output ports, must remain dedicated during the entire duration of data transfer until the teardown phase.
- Data transferred between the two stations are not packetized (physical layer transfer of the signal). The data are a continuous flow sent by the source station and received by the destination station, although there may be periods of silence.

- There is no addressing involved during data transfer. The switches route the data based on their occupied band (FDM) or time slot (TDM). Of course, there is end-to-end addressing used during the setup phase.

Three Phases:

The actual communication in a circuit-switched network requires three phases: connection setup, data transfer, and connection teardown.

Setup Phase

Before the two parties (or multiple parties in a conference call) can communicate, a dedicated circuit (combination of channels in links) needs to be established. The end systems are normally connected through dedicated lines to the switches, so connection setup means creating dedicated channels between the switches.

Data Transfer Phase

After the establishment of the dedicated circuit (channels), the two parties can transfer data.

Teardown Phase

When one of the parties needs to disconnect, a signal is sent to each switch to release the resources.

Packet Switched Networks:

A packet switched network (PSN) is a type of computer communications network that groups and sends data in the form of small packets. It enables the sending of data or network packets between a source and destination node over a network channel that is shared between multiple users and/or applications. A packet switched is also known as a connectionless network, as it does not create a permanent connection between a source and destination node. Packet-switched describes the type of network in which packets are routed through a network based on the destination address contained within each packet. Breaking communication down into packets allows the same data path to be shared among many users in the network.

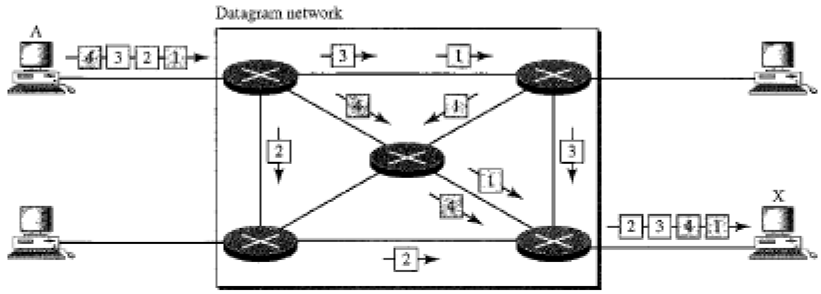
In packet switching, there is no resource allocation for a packet. This means that there is no reserved bandwidth on the links, and there is no scheduled processing time for each packet. Resources are allocated on demand. The allocation is done on a first-come, first serve basis. When a switch receives a packet, no matter what is the source or destination, the packet must wait if there are other packets being processed.

Packet switching may be classified into connectionless packet switching, also known as datagram switching, and connection-oriented packet switching, also known as virtual circuit switching.

Datagram Approach:

In a datagram network, each packet is treated independently of all others. Even if a packet is part of a multipacket transmission, the network treats it as though it existed alone. Packets in this approach are referred to as datagrams. Datagram switching is normally done at the network layer.

Figure 8.7 *A datagram network with four switches (routers)*



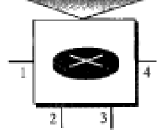
In this example, all four datagrams belong to the same message, but may travel different paths to reach their destination. This is so because the links may be involved in carrying packets from other sources and do not have the necessary bandwidth available to carry all the packets from A to X. This approach can cause the datagrams of a transmission to arrive at their destination out of order with different delays between the packets. Packets may also be lost or dropped because of a lack of resources.

The datagram networks are sometimes referred to as connectionless networks. The term connectionless here means that the switch does not keep information about the connection state. There are no setup and or teardown phases. Each packet is treated the same by a switch regardless of its source or destination.

Since there are no setup or teardown phases, each switch has a routing table which is based on the destination address. The routing tables are dynamic and are updated periodically. The destination address and the corresponding forwarding output ports are recorded in the tables. This is different from circuit-switched network in which each entry is created when the setup phase is completed and deleted when the teardown phase is over.

Routing table in a datagram network

Destination address	Output port
1232	1
4150	2
⋮	⋮
9130	3



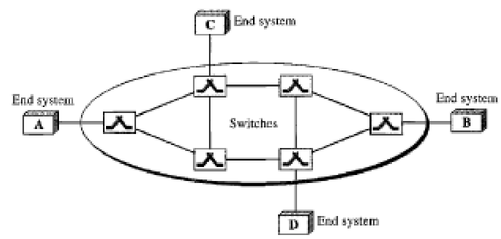
Virtual Circuit Networks:

A virtual circuit network is a cross between a circuit-switched network and a datagram network. It has some characteristics of both.

- As in a circuit-switched network, there are setup and teardown phases in addition to the data transfer phase.

- Resources can be allocated during the setup phase, as in a circuit-switched network, or on demand, as in a datagram network.
- As in a datagram network, data are packetized and each packet carries an address in the header. However, the address is of the next hop to be reached towards the destination.
- As in a circuit-switched network, all packets follow the same path established during the connection.
- A virtual-circuit network is normally implemented in the data link layer, while a circuit-switched network is implemented in the physical layer and a datagram network is implemented in the network layer.

10 *Virtual-circuit network*



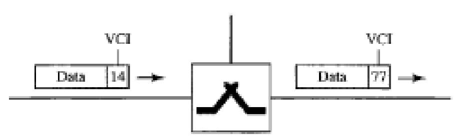
Addressing:

In a virtual-circuit network, two types of addressing are involved: global and local.

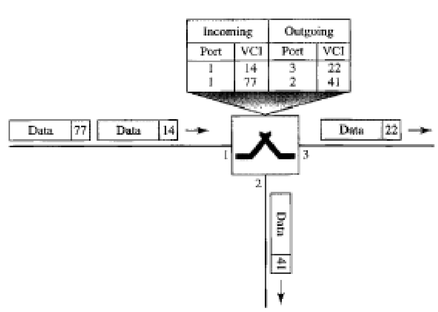
Global addressing: A source or a destination needs to have a global address-an address that can be unique in the scope of the network or internationally.

Local Addressing: The address that is actually used for data transfer is called the virtual-circuit identifier (VCI). A VCI, unlike a global address, is a smaller number that has only one switch scope; it is used by a frame between two switches.

Virtual-circuit identifier



Switch and tables in a virtual-circuit network



Three phases:

As in circuit-switched network, a source and destination need to go through three phases in a virtual circuit network: setup, data transfer, and teardown.

In the setup phase, the source and destination use their global addresses to help switches make table entries for the connection.

In the teardown phase, the source and destination inform the switches to delete the corresponding entry.

Data transfer occurs between these two phases.

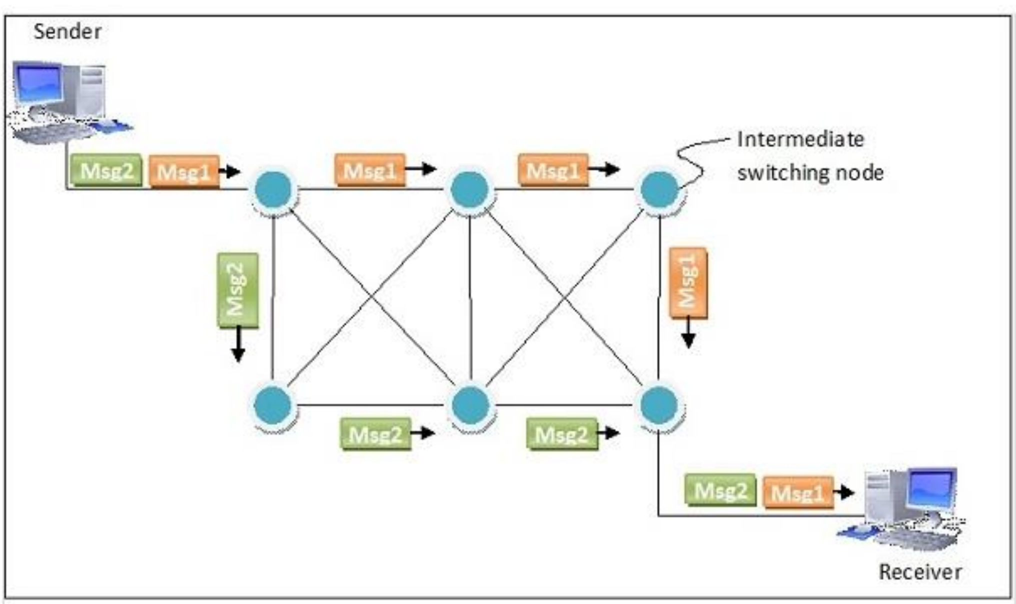
Message Switched Networks:

Message switching is a network switching technique in which data is routed in its entirety from the source node to the destination node, one hop at a time. During message routing, every intermediate switch in the network stores the whole message. If the entire network's resources are engaged or the network becomes blocked, the message-switched network stores and delays the message until ample resources become available for effective transmission of the message.

Before the advancements in packet switching, message switching acted as an efficient substitute for circuit switching.

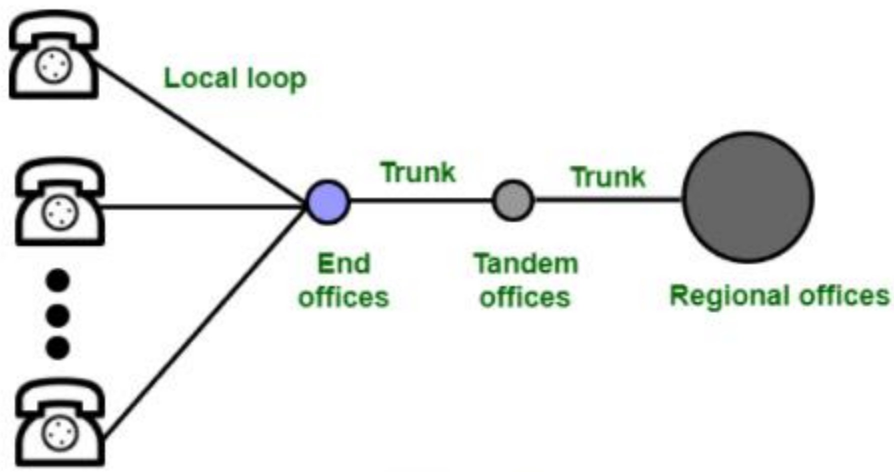
In message switching, the source and destination nodes are not directly connected. Instead, the intermediary nodes (mainly switches) are responsible for transferring the message from one node to the next. Thus, every intermediary node inside the network needs to store every message prior to retransferring the messages one-by-one as adequate resources become available. If the resources are not available, the messages are stored indefinitely. This characteristic is known as store and forward.

The following diagram represents routing of two separate messages from the same source to same destination via different routes, using message switching.



Telephone Network:

- Telephone Network is used to provide voice communication which uses Circuit Switching.
- Originally, the entire network was referred to as a plain old telephone system (POTS) which used analog signals.
- With the advancement of technology, i.e., in the computer era, there comes a feature to carry data in addition to voice. Today's network is both analogous and digital.
- The telephone network is made of three major components: local loops, trunks, and switching offices.
- The telephone network has several levels of switching offices such as end offices, tandem offices, and regional offices.



A telephone system

Local Loops

- One component of the telephone network is the local loop, a twisted-pair cable that connects the subscriber telephone to the nearest end office or local central office.
- The local loop, when used for voice, has a bandwidth of 4000 Hz (4 kHz). It is interesting to examine the telephone number associated with each local loop.
- The first three digits of a local telephone number define the office, and the next four digits define the local loop number.

Trunks

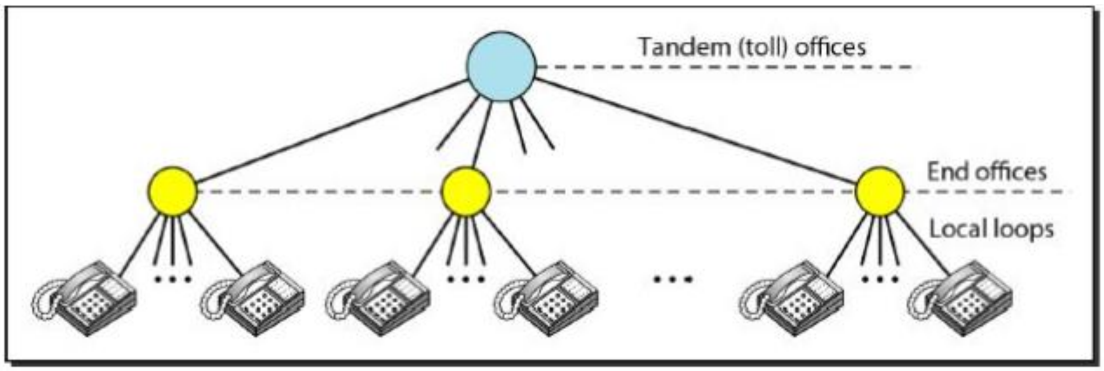
- Trunks are transmission media that handle the communication between offices.
- A trunk normally handles hundreds or thousands of connections through multiplexing.
- Transmission is usually through optical fibers or satellite links.

Switching Offices

- To avoid having a permanent physical link between any two subscribers, the telephone company has switches located in a switching office.
- A switch connects several local loops or trunks and allows a connection between different subscribers.

LATAs

- The Local-Access Transport Areas (LATAs) are the local telephone networks covering small or large metropolitan area.
- A small state or province may have one single LATA; a large state/province may have several LATAs. A LATA boundary may overlap the boundary of a state; part of a LATA can be in one state, part in another state.



- Communication inside a LATA is handled by end switches and tandem switches.
- A call that can be completed by using only end offices is considered toll-free. A call that has to go through a tandem office is charged.

Intra-LATA Services

The services offered by the common carriers (telephone companies) inside a LATA are called intra-LATA services. The carrier that handles these services is called a local exchange carrier (LEC).

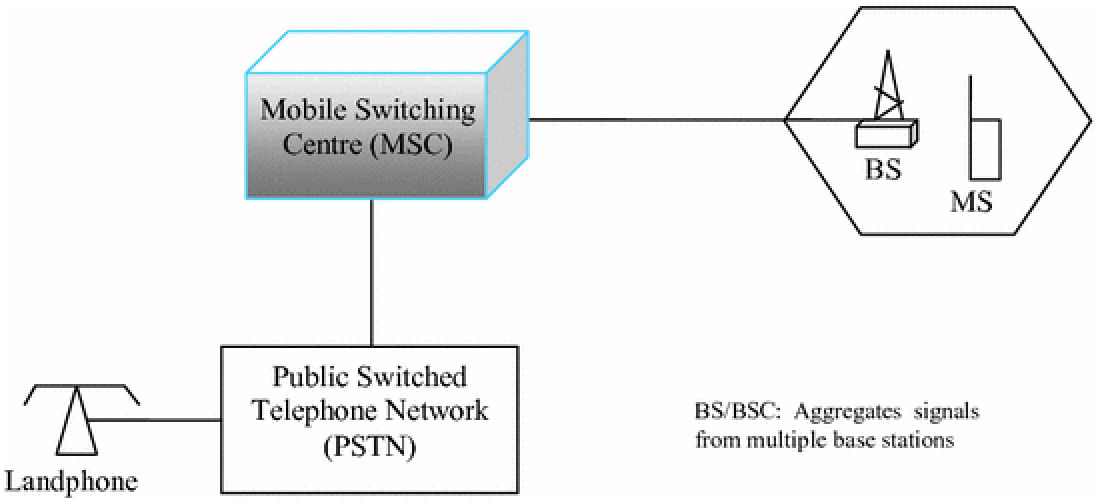
Inter-LATA Services

The services between LATAs are handled by interexchange carriers (IXCs). These carriers, sometimes called long-distance companies, provide communication services between two customers in different LATAs.

Mobile Networks:

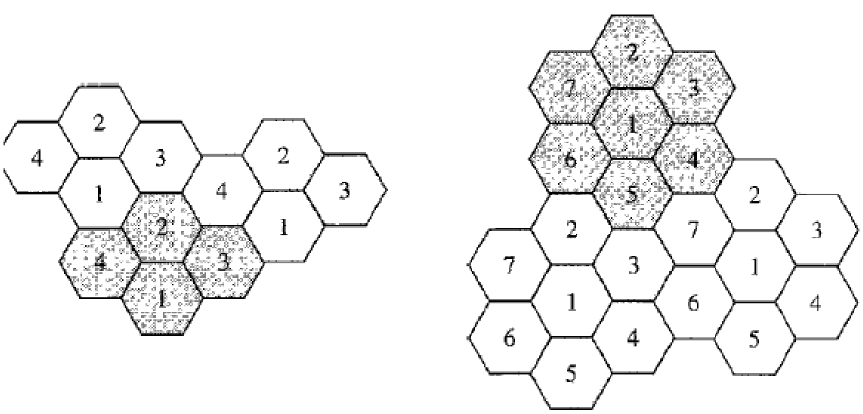
- Mobile Networks or Cellular networks are high-speed, high-capacity voice and data communication networks with enhanced multimedia and seamless roaming capabilities for supporting cellular devices (wireless end devices).
- With the increase in popularity of cellular devices, these networks are used for more than just entertainment and phone calls.
- Cellular telephony is designed to provide communications between two moving units, called mobile stations (MSs), or between one mobile unit and one stationary unit, often called a land unit.
- A service provider must be able to locate and track a caller, assign a channel to the call, and transfer the channel from base station to base station as the caller moves out of range.
- To make this tracking possible, each cellular service area is divided into small regions called cells.

- Each cell contains an antenna and is controlled by a solar or AC powered network station, called the base station (BS).
- Each base station, in turn, is controlled by a switching office, called a mobile switching center (MSC).
- The MSC coordinates communication between all the base stations and the telephone central office.
- It is a computerized center that is responsible for connecting calls, recording call information, and billing.



- In general, neighboring cells cannot use the same set of frequencies for communication because it may create interference for the users located near the cell boundaries.
- However, the set of frequencies available is limited, and frequencies need to be reused.
- A frequency reuse pattern is a configuration of N cells, N being the reuse factor, in which each cell uses a unique set of frequencies.
- When the pattern is repeated, the frequencies can be reused. There are several different patterns.

Frequency reuse patterns



a. Reuse factor of 4

b. Reuse factor of 7

Transmitting

- To place a call from a mobile station, the caller enters a code of 7 or 10 digits (a phone number) and presses the send button.
- The mobile station then scans the band, seeking a setup channel with a strong signal, and sends the data (phone number) to the closest base station using that channel.
- The base station relays the data to the MSC. The MSC sends the data on to the telephone central office.
- If the called party is available, a connection is made and the result is relayed back to the MSC.
- At this point, the MSC assigns an unused voice channel to the call, and a connection is established.
- The mobile station automatically adjusts its tuning to the new channel, and communication can begin.

Receiving

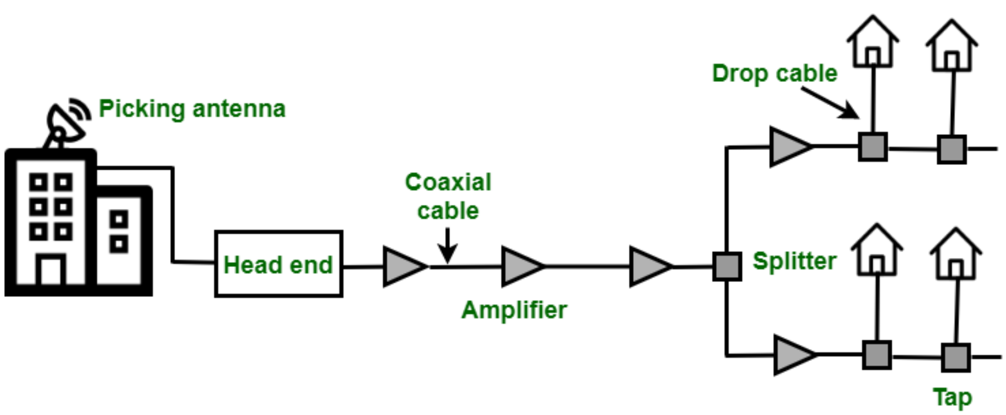
- When a mobile phone is called, the telephone central office sends the number to the MSC.
- The MSC searches for the location of the mobile station by sending query signals to each cell in a process called paging.
- Once the mobile station is found, the MSC transmits a ringing signal and, when the mobile station answers, assigns a voice channel to the call, allowing voice communication to begin.

Cable Networks:

- The cable TV network started as a video service provider, but it has moved to the business of Internet access.

Traditional Cable Networks:

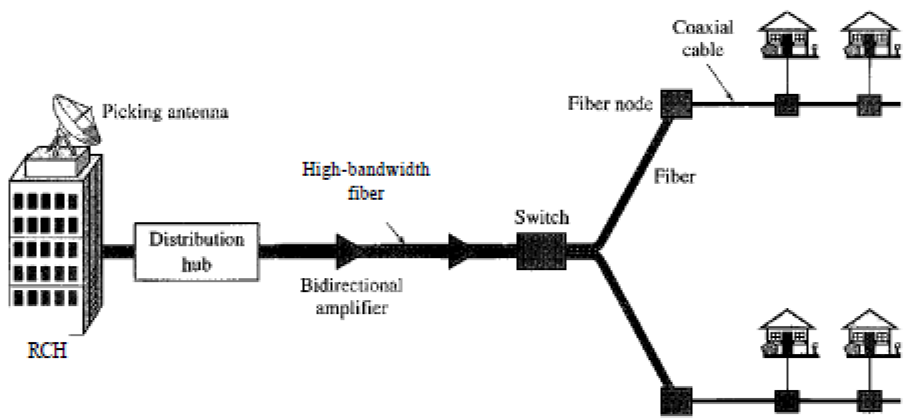
- Cable TV started to distribute broadcast video signals to locations with poor or no reception in the late 1940s.
- It was called community antenna TV (CATV) because an antenna at the top of a tall hill or building received the signals from the TV stations and distributed them, via coaxial cables, to the community.



- The cable TV office, called the head end, receives video signals from broadcasting stations and feeds the signals into coaxial cables.
- The signals became weaker and weaker with distance, so amplifiers were installed through the network to renew the signals. There could be up to 35 amplifiers between the head end and the subscriber premises.
- At the other end, splitters split the cable, and taps and drop cables make the connections to the subscriber premises.
- The traditional cable TV system used coaxial cable end to end.
- Due to attenuation of the signals and the use of a large number of amplifiers, communication in the traditional network was unidirectional (one-way).
- Video signals were transmitted downstream, from the head end to the subscriber premises.

Hybrid Fiber-Coaxial Network:

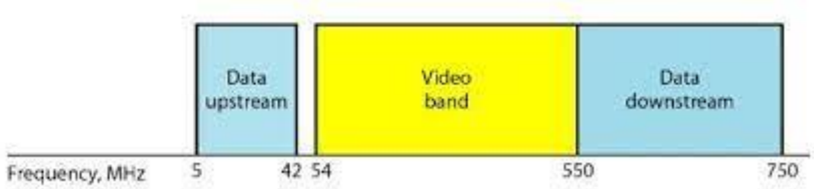
- Hybrid Fiber-Coaxial Network is that the second generation of the cable network. which is a combination of fiber-optic and coaxial cable is used in this type of network.
- The transmission mode is used is fiber node i.e., fiber mode. The schematic diagram of the HFC network is as follows:



- There are nearly 400, 000 subscribers served by Regional Cable Head (RCH). The RCHs feed the distribution hubs, each of which serves up to 40,000 subscribers. The distribution hub plays an important role in the new infrastructure.
- Modulation and demodulation of the signal are done through the distribution hubs after these signals are sent to the fiber nodes through fiber-optic cables.
- The fiber node split the analog signal so that the same signal is sent to each coaxial cable. Approx. 1000 subscribers are served by coaxial cable.
- The use of fiber-optic cable reduces the need for amplifiers down to eight or less.
- One reason for moving from traditional to hybrid infrastructure is to make the cable network bidirectional (two-way).

Cable TV for Data Transfer:

- Cable companies are now competing with telephone companies for the residential customer who wants high-speed data transfer.
- DSL technology provides high-data-rate connections for residential subscribers over the local loop. However, DSL uses the existing unshielded twisted-pair cable, which is very susceptible to interference. This imposes an upper limit on the data rate.
- Another solution is the use of the cable TV network.
- Even in an HFC system, the last part of the network, from the fiber node to the subscriber premises, is still a coaxial cable.
- This coaxial cable has a bandwidth that ranges from 5 to 750 MHz (approximately).
- To provide Internet access, the cable company has divided this bandwidth into three bands: video, downstream data, and upstream data.



Downstream Video Band:

- The downstream video band occupies frequencies from 54 to 550 MHZ. Since each TV channel occupies 6 MHz, this can accommodate more than 80 channels.

Downstream Data Band:

- The downstream data (from the Internet to the subscriber premises) occupies the upper band, from 550 to 750 MHZ. This band is also divided into 6-MHz channels.

Upstream Data Band:

- The upstream data (from the subscriber premises to the Internet) occupies the lower band, from 5 to 42 MHZ. This band is also divided into 6-MHz channels.