CSC3317: DATA COMMUNICATION AND NETWORKS

Overview

Data communications and networking are changing the way we do business and the way we live. Business decisions have to be made ever more quickly, and the decision makers require immediate access to accurate information. Why wait a week for that report from Germany to arrive by mail when it could appear almost instantaneously through computer networks? Businesses today rely on computer networks and internetworks.

But before we ask how quickly we can get hooked up, we need to know how networks operate, what types of technologies are available, and which design best fills which set of needs.

The development of the personal computer brought about tremendous changes for business, industry, science, and education. A similar revolution is occurring in data communications and networking. Technological advances are making it possible for communications links to carry more and faster signals. As a result, services are evolving to allow use of this expanded capacity. For example, established telephone services such as conference calling, call waiting, voice mail, and caller ID have been extended.

Research in data communications and networking has resulted in new technologies. One goal is to be able to exchange data such as text, audio, and video from all points in the world. We want to access the Internet to download and upload information quickly and accurately and at any time.

DATA COMMUNICATIONS

When we communicate, we are sharing information. This sharing can be local or remote. Between individuals, local communication usually occurs face to face, while remote communication takes place over distance. The term *telecommunication*, which includes telephony, telegraphy, and television, means communication at a distance *(tele* is Greek for "far").

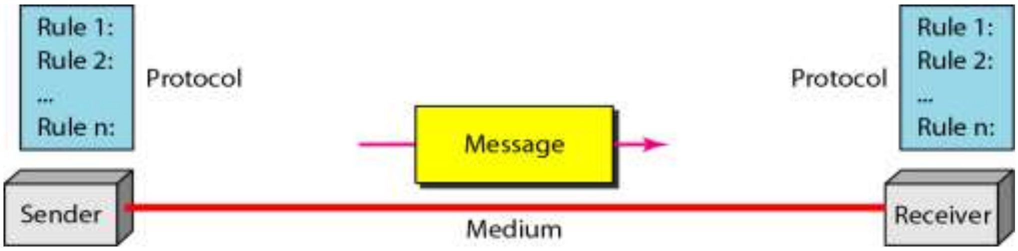
The word *data* refers to information presented in whatever form is agreed upon by the parties creating and using the data. The word *data* refers to information presented in whatever form is agreed upon by the parties creating and using the data

Data communications are the exchange of data between two devices via some form of transmission medium such as a wire cable.

For data communications to occur, the communicating devices must be part of a communication system made up of a combination of hardware (physical equipment) and software (programs). The effectiveness of a data communications system depends on four fundamental characteristics: delivery, accuracy, timeliness, and jitter.

1. Delivery. The system must deliver data to the correct destination. Data must be received by the intended device or user and only by that device or user.
2. Accuracy. The system must deliver the data accurately. Data that have been altered in transmission and left uncorrected are unusable.
3. Timeliness. The system must deliver data in a timely manner. Data delivered late are useless. In the case of video and audio, timely delivery means delivering data as they are produced, in the same order that they are produced, and without significant delay. This kind of delivery is called *real-time* transmission.
4. Jitter. Jitter refers to the variation in the packet arrival time. It is the uneven delay in the delivery of audio or video packets. For example, let us assume that video packets are sent every 3D ms. If some of the packets arrive with 3D-ms delay and others with 4D-ms delay, an uneven quality in the video is the result

A data communications system has five components



1. Message. The message is the information (data) to be communicated. Popular forms of information include text, numbers, pictures, audio, and video.
2. Sender. The sender is the device that sends the data message. It can be a computer, workstation, telephone handset, video camera, and so on.
3. Receiver. The receiver is the device that receives the message. It can be a computer, workstation, telephone handset, television, and so on.
4. Transmission medium. The transmission medium is the physical path by which a message travels from sender to receiver. Some examples of transmission media include twisted-pair wire, coaxial cable, fiber-optic cable, and radio waves.
5. Protocol. A protocol is a set of rules that govern data communications. It represents an agreement between the communicating devices. Without a protocol, two devices may be connected but not communicating, just as a person speaking French cannot be understood by a person who speaks only Japanese.

Physical Structures

Before discussing networks, we need to define some network attributes.

Type of Connection

A network is two or more devices connected through links. A link is a communications pathway that transfers data from one device to another. For visualization purposes, it is simplest to imagine any link as a line drawn between two points. For communication to occur, two devices must be connected in some way to the same link at the same time.

There are two possible types of connections: point-to-point and multipoint.

Point-to-Point: A point-to-point connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices. Most point-to-point connections use an actual length of wire or cable to connect the two ends, but other options, such as microwave or satellite links, are also possible (see Figure 1.3a). When you change television channels by infrared remote control, you are establishing a point-to-point connection between the remote control and the television's control system.

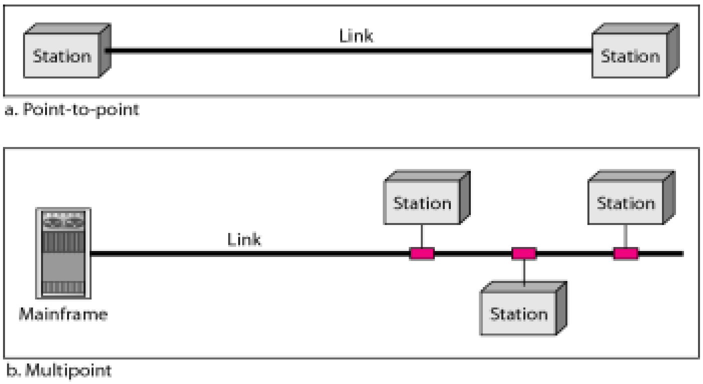


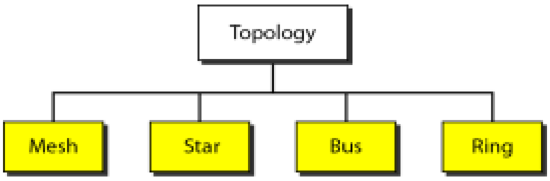
Figure : *Types of connections: point-to-point and multipoint*

Multipoint: A multipoint (also called multidrop) connection is one in which more than two specific devices share a single link (see Figure 1.3b).

In a multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a *spatially shared* connection. If users must take turns, it is a *timeshared* connection.

Physical Topology

The term *physical topology* refers to the way in which a network is laid out physically: two or more devices connect to a link; two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called nodes) to one another. There are four basic topologies possible: mesh, star, bus, and ring



In a mesh topology, every device has a dedicated point-to-point link to every other device. The term *dedicated* means that the link carries traffic only between the two devices it connects. To find the number of physical links in a fully connected mesh network with *n* nodes, we first consider that each node must be connected to every other node. Node 1 must be connected to *n* - I nodes, node 2 must be connected to *n* – 1 nodes, and finally node *n* must be connected to *n* - 1 nodes. We need *n(n* - 1) physical links. However, if each physical link allows communication in both directions (duplex mode), we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need

$$n(n -1) /2$$

Advantages: Fast communication, Robust and Privacy (Security)

Disadvantages: Cabling Space and cost

Star: Devices are connected Point to Point to a central "Hub" (Controller Exchanger)

Advantages: Less cabling and H/W ports, two hops only.

Disadvantage: Not robust

Bus Topology: The preceding examples all describe point-to-point connections. A bus topology, on the other hand, is multipoint. One long cable acts as a backbone to link all the devices in a network.

Advantage: Less Cabling

Disadvantages: Topology dependent, limit number of nodes on the bus due to signal power loss with distance, not so robust.

Ring Topology In a ring topology, each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.

Advantages: easy installation, better fault isolation and robustness.

Disadvantages: N/2 hops communication

Network Models

Computer networks are created by different entities. Standards are needed so that these heterogeneous networks can communicate with one another. The two best-known standards are the OSI model and the Internet model. In Chapter 2 we discuss these two models. The OSI (Open Systems Interconnection) model defines a seven-layer network; the Internet model defines a five-layer network. This book is based on the Internet model with occasional references to the OSI model.

Categories of Networks

Today when we speak of networks, we are generally referring to two primary categories: local-area networks and wide-area networks. The category into which a network falls is determined by its size. A LAN normally covers an area less than 2 mi; aWAN can be worldwide. Networks of a size in between are normally referred to as metropolitan area networks and span tens of miles.

Local Area Network

A local area network (LAN) is usually privately owned and links the devices in a single office, building, or campus (see Figure 1.10). Depending on the needs of an organization and the type of technology used, a LAN can be as simple as two PCs and a printer in someone's home office; or it can extend throughout a company and include audio and video peripherals. Currently, LAN size is limited to a few kilometres.

Wide Area Networks (WAN):

WAN span a large geographical area about 100’s – 1000’s of Km

1) Switched: End users connected via a cloud of switches (subnet).

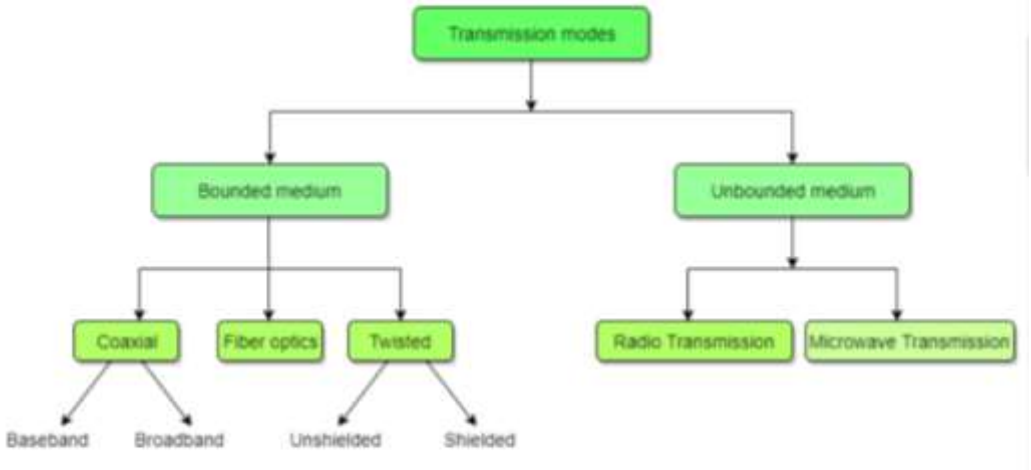
2) Point-to Point: Line leased from Telephone Company/ TV connecting users to the ISP for Internet access

Transmission Mediums in Computer Networks

Data is represented by computers and other telecommunication devices using signals. Signals are transmitted in the form of electromagnetic energy from one device to another. Electromagnetic signals travel through vacuum, air or other transmission mediums to move from one point to another(from sender to receiver).

Electromagnetic energy (includes electrical and magnetic fields) consists of power, voice, visible light, radio waves, ultraviolet light, gamma rays etc.

Transmission medium is the means through which we send our data from one place to another. The first layer (physical layer) of Communication Networks OSI Seven layer model is dedicated to the transmission media, we will study the OSI Model later.



Factors to be considered while selecting a Transmission Medium

1. Transmission Rate
2. Cost and Ease of Installation
3. Resistance to Environmental Conditions
4. Distances

Bounded or Guided Transmission Media

Guided media, which are those that provide a conduit from one device to another, include **Twisted-Pair Cable**, **Coaxial Cable**, and **Fibre-Optic Cable**.

A signal travelling 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 fibre** is a cable that accepts and transports signals in the form of light.

Twisted Pair Cable

This cable is the most commonly used and is cheaper than others. It is lightweight, cheap, can be installed easily, and they support many different types of network. Some important points :

- Its frequency range is 0 to 3.5 kHz.
- Typical attenuation is 0.2 dB/Km @ 1kHz.
- Typical delay is 50 µs/km.
- Repeater spacing is 2km.

A twisted pair consists of two conductors(normally copper), each with its own plastic insulation, twisted together. One of these wires is used to carry signals to the receiver, and the other is used only as 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. This results in a difference at the receiver.

Twisted Pair is of two types:

- **Unshielded Twisted Pair (UTP)**
- **Shielded Twisted Pair (STP)**

Unshielded Twisted Pair Cable

It is the most common type of telecommunication when compared with Shielded Twisted Pair Cable which consists of two conductors usually copper, each with its own colour plastic insulator. Identification is the reason behind coloured plastic insulation.

UTP cables consist of 2 or 4 pairs of twisted cable. Cable with 2 pair use **RJ-11** connector and 4 pair cable use **RJ-45** connector.

Advantages of Unshielded Twisted Pair Cable

- Installation is easy
- Flexible
- Cheap
- It has high speed capacity,
- 100 meter limit
- Higher grades of UTP are used in LAN technologies like Ethernet.

It consists of two insulating copper wires (1mm thick). The wires are twisted together in a helical form to reduce electrical interference from similar pair.

Disadvantages of Unshielded Twisted Pair Cable

- Bandwidth is low when compared with Coaxial Cable
- Provides less protection from interference.

Shielded Twisted Pair Cable

This cable has a metal foil or braided-mesh covering which encases each pair of insulated conductors. Electromagnetic noise penetration is prevented by metal casing. Shielding also eliminates crosstalk (explained in KEY TERMS Chapter).

It has same attenuation as unshielded twisted pair. It is faster the unshielded and coaxial cable. It is more expensive than coaxial and unshielded twisted pair.

Advantages of Shielded Twisted Pair Cable

- Easy to install
- Performance is adequate
- Can be used for Analog or Digital transmission

- Increases the signalling rate
- Higher capacity than unshielded twisted pair
- Eliminates crosstalk

Disadvantages of Shielded Twisted Pair Cable

- Difficult to manufacture
- Heavy

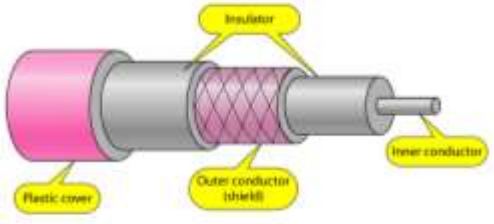
Coaxial Cable

Coaxial is called by this name because it contains two conductors that are parallel to each other. Copper is used in this as centre conductor which can be a solid wire or a standard one. It is surrounded by PVC installation, a sheath which is encased in an outer conductor of metal foil, barid or both.

Outer metallic wrapping is used as a shield against noise and as the second conductor which completes the circuit. The outer conductor is also encased in an insulating sheath. The outermost part is the plastic cover which protects the whole cable.

Here the most common coaxial standards.

- 50-Ohm RG-7 or RG-11 : used with thick Ethernet.
- 50-Ohm RG-58 : used with thin Ethernet
- 75-Ohm RG-59 : used with cable television
- 93-Ohm RG-62 : used with ARCNET.



Coaxial Cable Standards

Coaxial cables are categorized by their Radio Government(RG) ratings. Each RG number denotes a unique set of physical specifications, including the wire gauge of the inner conductor, the thickness and the type of the inner insulator, the construction of the shield, and the size and type of the outer casing.

There are two types of Coaxial cables:

1. BaseBand

This is a 50 ohm (Ω) coaxial cable which is used for digital transmission. It is mostly used for LAN's. Baseband transmits a single signal at a time with very high speed. The major drawback is that it needs amplification after every 1000 feet.

2. BroadBand

This uses analog transmission on standard cable television cabling. It transmits several simultaneous signal using different frequencies. It covers large area when compared with Baseband Coaxial Cable.

Advantages of Coaxial Cable

- Bandwidth is high

- Used in long distance telephone lines.
- Transmits digital signals at a very high rate of 10Mbps.
- Much higher noise immunity
- Data transmission without distortion.
- The can span to longer distance at higher speeds as they have better shielding when compared to twisted pair cable

Disadvantages of Coaxial Cable

- Single cable failure can fail the entire network.
- Difficult to install and expensive when compared with twisted pair.
- If the shield is imperfect, it can lead to grounded loop.

Fiber Optic Cable

A fibre-optic cable is made of glass or plastic and transmits signals in the form of light.

For better understanding we first need to explore several aspects of the **nature of light**.

Light travels in a straight line as long as it is mobbing through a single uniform substance. If ray of light travelling through one substance suddenly enters another substance (of a different density), the ray changes direction.

Advantages of Fibre Optic Cable

Fibre optic has several advantages over metallic cable:

- Higher bandwidth
- Less signal attenuation
- Immunity to electromagnetic interference
- Resistance to corrosive materials
- Light weight
- Greater immunity to tapping

Disadvantages of Fibre Optic Cable

There are some disadvantages in the use of optical fibre:

- Installation and maintenance
- Unidirectional light propagation
- High Cost

UnBounded or UnGuided Transmission Media

Unguided medium 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.

We can divide wireless transmission into three broad groups:

1. Radio waves
2. Micro waves
3. Infrared waves

Radio Waves

Electromagnetic waves ranging in frequencies between 3 KHz and 1 GHz are normally called radio waves.

Radio waves 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 send waves that can be received by any receiving antenna. The omnidirectional property has disadvantage, too. The radio waves transmitted by one antenna are susceptible to interference by another antenna that may send signal suing the same frequency or band.

Applications of Radio Waves

- 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, and paging are examples of multicasting.

Micro Waves

Electromagnetic waves having frequencies between 1 and 300 GHz are called micro waves. Micro waves are unidirectional. When an antenna transmits microwaves, they can be narrowly focused. This means that the sending and receiving antennas need to be aligned.

Applications of Micro Waves

Microwaves, due to their unidirectional properties, are very useful when unicast(one-to-one) communication is needed between the sender and the receiver. They are used in cellular phones, satellite networks and wireless LANs.

There are 2 types of Microwave Transmission:

1. Terrestrial Microwave
2. Satellite Microwave

Advantages of Microwave Transmission

- Used for long distance telephone communication
- Carries 1000's of voice channels at the same time

Disadvantages of Microwave Transmission

- It is very costly

Infrared Waves

Infrared waves, with frequencies from 300 GHz to 400 THz, 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 on room cannot be affected by another system in the next room.

When we use infrared remote control, we do not interfere with the use of the remote by our neighbours. 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.

Applications of Infrared Waves

- 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.
- The Infrared Data Association (IrDA), an association for sponsoring the use of infrared waves, has established standards for using these signals for communication between devices such as keyboards, mouse, PCs and printers.

- Infrared signals can be used for short-range communication in a closed area using line-of-sight propagation.

Data Representation

Information today comes in different forms such as text, numbers, images, audio, and video.

Text

In data communications, text is represented as a bit pattern, a sequence of bits (Os orIs). Different sets of bit patterns have been designed to represent text symbols. Each set is called a code, and the process of representing symbols is called coding. Today, the prevalent coding system is called Unicode, which uses 32 bits to represent a symbol or character used in any language in the world. The American Standard Code for Information

Interchange (ASCII), developed some decades ago in the United States, now constitutes the first 127 characters in Unicode and is also referred to as Basic Latin.

Numbers

Numbers are also represented by bit patterns. However, a code such as ASCII is not used to represent numbers; the number is directly converted to a binary number to simplify mathematical operations. Appendix B discusses several different numbering systems.

Images

Images are also represented by bit patterns. In its simplest form, an image is composed of a matrix of pixels (picture elements), where each pixel is a small dot. The size of the pixel depends on the *resolution.* For example, an image can be divided into 1000 pixels or 10,000 pixels. In the second case, there is a better representation of the image (better resolution), but more memory is needed to store the image.

After an image is divided into pixels, each pixel is assigned a bit pattern. The size and the value of the pattern depend on the image. For an image made of only blackand-white dots (e.g., a chessboard), a I-bit pattern is enough to represent a pixel.

If an image is not made of pure white and pure black pixels, you can increase the size of the bit pattern to include gray scale. For example, to show four levels of grayscale, you can use 2-bit patterns. A black pixel can be represented by 00, a dark gray pixel by 01, a light gray pixel by 10, and a white pixel by 11.

There are several methods to represent color images. One method is called RGB, so called because each color is made of a combination of three primary colors: *red*, green, and blue. The intensity of each color is measured, and a bit pattern is assigned to it. Another method is called YCM, in which a color is made of a combination of three other primary colors: yellow, cyan, and magenta.

Audio

Audio refers to the recording or broadcasting of sound or music. Audio is by nature different from text, numbers, or images. It is continuous, not discrete. Even when we use a microphone to change voice or music to an electric signal, we create a continuous signal.

Video

Video refers to the recording or broadcasting of a picture or movie. Video can either be produced as a continuous entity (e.g., by a TV camera), or it can be a combination of images, each a discrete entity, arranged to convey the idea of motion.

NETWORKS

Modern world scenario is ever changing. Data Communication and network have changed the way business and other daily affair works. Now, they highly rely on computer networks and internetwork. A set of devices often mentioned as nodes connected by media link is called a Network. A network is a set of devices (often referred to as *nodes)* connected by communication links.

A node can be a device which is capable of sending or receiving data generated by other nodes on the network like a computer, printer etc. These links connecting the devices are called **Communication channels**. Computer network does not mean a system with one Control Unit connected to multiple other systems as its slave. That is Distributed system, not Computer Network..

Distributed Processing

Most networks use distributed processing, in which a task is divided among multiple computers. Instead of one single large machine being responsible for all aspects of a process, separate computers (usually a personal computer or workstation) handle a subset.

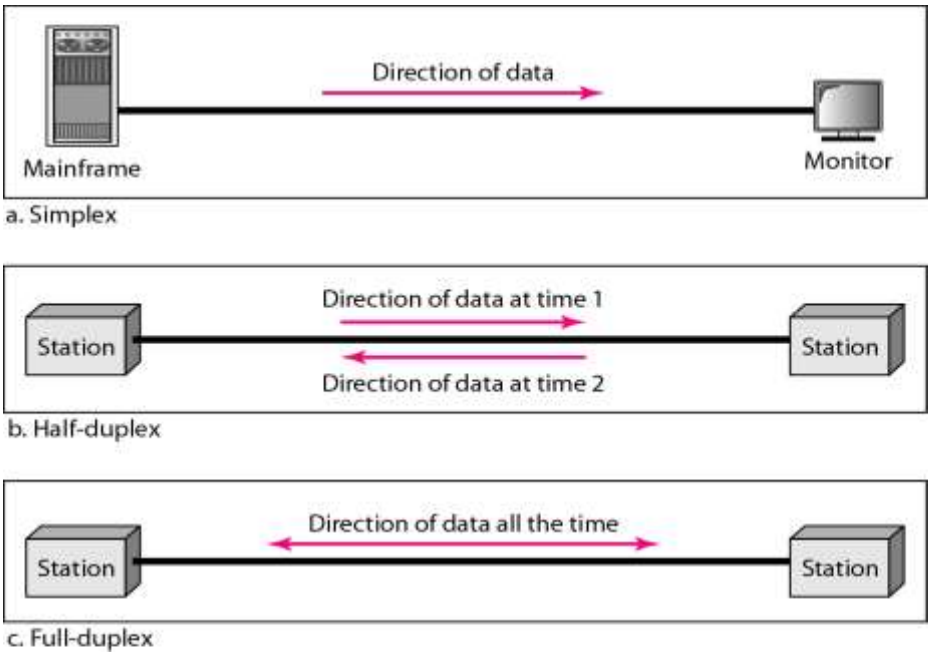
Communication Modes in Computer Networks

Transmission mode refers to the mechanism of transferring of data between two devices connected over a network. It is also called **Communication Mode**. These modes direct the direction of flow of information

Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex as shown in Figure below;

Data flow (simplex, half-duplex, and full-duplex)



Simplex

In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive (see Figure 1.2a).

Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

Half-Duplex

In half-duplex mode, each station can both transmit and receive, but not at the same time. : When one device is sending, the other can only receive, and vice versa.

The half-duplex mode is like a one-lane road with traffic allowed in both directions.

When cars are traveling in one direction, cars going the other way must wait. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time. Walkie-talkies and CB (citizens band) radios are both half-duplex systems.

The half-duplex mode is used in cases where there is no need for communication in both directions at the same time; the entire capacity of the channel can be utilized for each direction.

Full-Duplex

In full-duplex mode (also called duplex), both stations can transmit and receive simultaneously. The full-duplex mode is like a two-way street with traffic flowing in both directions at the same time. In full-duplex mode, signal going in one direction share the capacity of the link: with signals going in the other direction. This sharing can occur in two ways: Either the link must contain two physically separate transmission ID n paths, one for sending and the other for receiving; or the capacity of the channel is divided between signals traveling in both directions.

One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.

The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

Transmission Mode

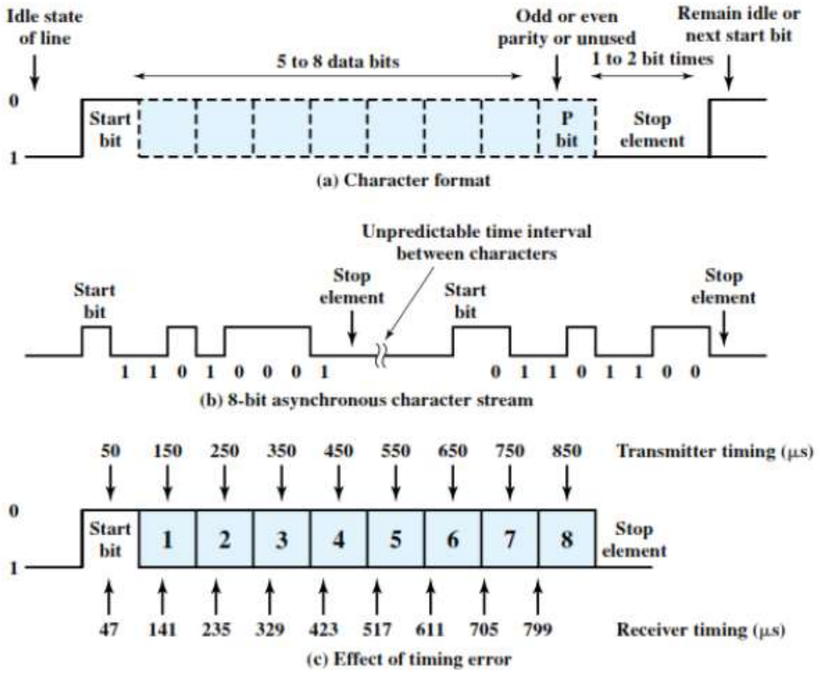
The transmission of a stream of bits from one device to another across a transmission link involves a great deal of cooperation and agreement between the two sides. One of the most fundamental requirements is ***synchronization***. The receiver must know the rate at which bits are being received so that it can sample the line at appropriate intervals to determine the value of each received bit. Two techniques are in common use for this purpose. In ***asynchronous transmission***, each character of data is treated independently. Each character begins with a start bit that alerts the receiver that a character is arriving. The receiver samples each bit in the character and then looks for the beginning of the next character. *This technique would not work well for long blocks of data because the receiver's clock might eventually drift out of synchronization with the transmitter's clock.* However, *sending data in large blocks is more efficient than sending data one character at a time.*

Asynchronous and Synchronous Transmission

In order for the receiver to sample the incoming bits properly, it must know the arrival time and duration of each bit that it receives. Support that the sender simply wants to transmit a stream of data bits. The sender has a clock that governs the timing of the transmitted bits. *For example*, if data are to be transmitted at one million bits per second (1 Mbps), then one bit will be transmitted every 1/10 6=1 microsecond (μm) as measured by the sender's clock. Typically, the receiver will attempt to sample the medium at the center of each bit time. The receiver will time its samples at intervals of one bit time. In our example, the sampling would occur once every 1 μm. If the receiver times its samples based on its own clock, then there will be a problem if the transmitter's and receiver's clocks are not precisely aligned. If there is a drift of 1% (the receiver's clock is 1% faster or slower than the transmitter's clock), then the first sampling will be 0.01 of a bit time (0.01 (μm)) away from the center of the bit (center of bit is 0.5 μm from beginning and end of bit). After 50 or more samples, the receiver may be in error because it is sampling in the wrong bit time (50 * 0.01 (μm)).

Asynchronous Transmission

Two approaches are common for achieving the desired synchronization. The first is called, ***oddly enough, asynchronous transmission***. The strategy with this scheme is to avoid the *timing problem* by *not sending long, uninterrupted streams of bits*. Instead, data are transmitted one character at a time, where each character is five to eight bits in length. Timing or synchronization must only be maintained within each character; the receiver has the opportunity to resynchronize at the beginning of each new character.



(a) Character format

(b) 8-bit asynchronous character stream

(c) Effect of timing error

Figure 1. Asynchronous Transmission

Figure 1 illustrates this technique. When no character is being transmitted, the line between transmitter and receiver is in an *idle state*. The definition of ***idle*** *is equivalent to the signaling element for binary 1*. The beginning of a character is signaled by a ***start bit*** with a value *of binary 0*. This is followed by the 5 to 8 bits that actually make up the character. The bits of the character are transmitted beginning with the ***least significant bit***. For example, the data bits are usually followed by a parity bit, which therefore is in the most significant bit position. *The* ***parity bit*** is set by the transmitter such that the total number of ones in the character, ***including the parity bit, is even or odd, depending on the convention being used.*** The receiver uses this bit for error detection, which will be discussed later. The final element is a ***stop element***, which is a binary 1. A minimum length for the stop element is specified, and this is usually 1, 1.5, or 2 times the duration of an ordinary bit. No maximum value is specified. Because the stop element is the same as the idle state the transmitter continue to transmit the stop element until it is ready to send the next character.

An error such as just described actually results in two errors.

1. The last sampled bit is incorrectly received.

2. The bit count may now be out of alignment. If bit 7 is a 1 and bit 8 is a 0, bit 8 could be mistaken for a start bit.

This condition is termed a ***framing error***, as the character plus start bit and stop element are sometimes referred to as ***a frame***. A framing error can also occur if some noise condition causes the false appearance of a start bit during the idle state. Asynchronous transmission is simple and cheap but requires an **overhead** of *two to three bits per character*. For example, for an 8-bit character with no parity bit, using a 1-bit-long stop element, two out of every ten bits convey no information but are there merely for synchronization; thus the overhead is 20%. Of course, the *percentage overhead could be reduced by sending larger blocks of bits between the start bit and stop element.* However, as Figure 1c indicates, the larger the block of bits, the greater the cumulative timing error. To achieve greater efficiency, a different form of synchronization, known as synchronous transmission, is used.

Synchronous Transmission

With synchronous transmission, a block of bits is transmitted in a steady stream without start and stop codes. *The block may be many bits in length*. To prevent ***timing drift*** between transmitter and receiver, their clocks must somehow be synchronized. One possibility is to provide a separate clock line between transmitter and receiver. One side (transmitter or receiver) pulses the line regularly with one short pulse per bit time. The other side uses these regular pulses as a clock. This technique works well over short distances, but over longer distances the clock pulses are subject to the same **impairments** as the data signal, and **timing errors** can occur. The other alternative is to **embed the clocking information** in the data signal. For digital signals, this can be accomplished with *Manchester or differential Manchester encoding*. For analog signals, a number of techniques can be used; for example, the *carrier frequency* itself can be used to synchronize the receiver based on the phase of the carrier.

With synchronous transmission, there is another level of synchronization required, to allow the receiver to determine the beginning and end of a block of data. To achieve this, each block begins with a ***preamble bit pattern*** and generally ends with a ***postamble bit pattern***. In addition, other bits are added to the block that convey control information used in the data link control procedures. *The data plus preamble, postamble, and control information are called a* ***frame***. The exact format of the frame depends on which data link control procedure is being used.

Figure 2 shows, in general terms, a typical frame format for synchronous transmission. Typically, the frame starts with a preamble called ***a flag***, which is 8 bits long. The same flag is used as a postamble. The receiver looks for the occurrence of the flag pattern to signal the start of a frame. This is followed by some number of control fields (containing data link control protocol information), then a data field (variable length for most protocols), more control fields, and finally the flag is repeated. For sizable blocks of data, synchronous transmission is far more efficient than asynchronous. Asynchronous transmission requires *20% or more overhead*. The control information, preamble, and postamble in synchronous transmission are typically less than *100 bits*.

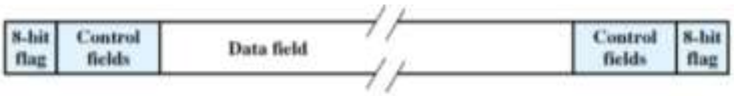


Figure 2: Synchronous Frame Format

Error Detection

Regardless of the design of the transmission system, there will be errors, resulting in the change of one or more bits in a transmitted frame. In what follows, we assume that data are transmitted as one or more contiguous sequences of bits, called frames. This is the kind of result that motivates the use of error-detecting techniques.

- **Parity Check**

The simplest error-detecting scheme is to append a parity bit to the end of a block of data. A typical example is character transmission, in which a parity bit is attached to each 7-bit International Reference Alphabet (IRA) character. The value of this bit is selected so that the character has an even number of 1s (even parity) or an odd number of 1s (odd parity). Note, however, that if two (or any even number) of bits are inverted due to error, an undetected error occurs. Typically, even parity is used for synchronous transmission and odd parity for asynchronous transmission. The use of the parity bit is not foolproof, as noise impulses are often long enough to destroy more than one bit, particularly at high data rates.

- **Cyclic Redundancy Check (CRC)**

One of the most common, and one of the most powerful, error-detecting codes is the Cyclic Redundancy Check (CRC), which can be described as follows. Given a *k-bit* block of bits, or message, the transmitter generates an *(n-k)-bit* sequence, known as a FCS, such that the resulting frame, consisting of *n* bits, is exactly divisible by some predetermined number. The receiver then divides the incoming frame by that number and, if there is no remainder, assumes there was no error. CRC presents the procedure *in three equivalent ways: modulo 2 arithmetic, polynomials, and digital logic*.

Error Correction

Error detection is a useful technique, found in data link control protocols, such as HDLC, and in transport protocols, such as TCP. However, correction of errors using an ***error-detecting code***, requires that block of data be retransmitted. For wireless applications this approach is inadequate for two reasons.

 The bit error rate on a wireless link can be quite high, which would result in a large number of retransmissions.

 In some cases, especially satellite links, the propagation delay is very long compared to the transmission time of a single frame. The result is a very inefficient system.

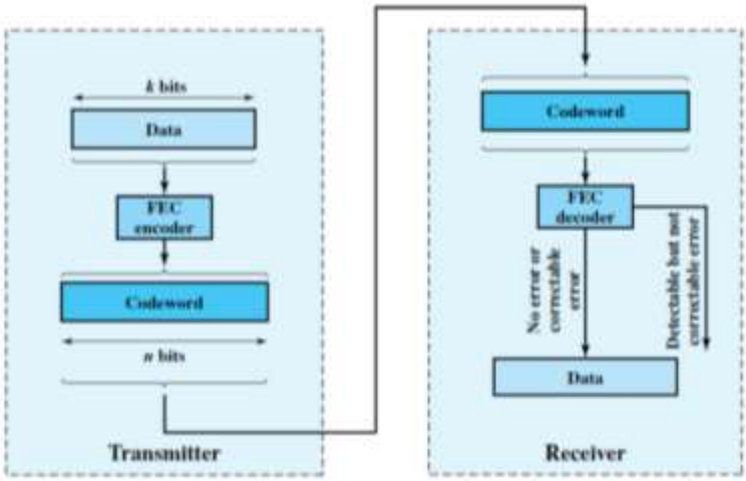


Figure 3. Error Correction Process

Instead, it would be desirable to enable the receiver to correct errors in an incoming transmission on the basis of the bits in that transmission. Figure 3 shows in general how this is done. On the transmission end, each *k-bit* block of data is mapped into an *n-bit* block called a ***codeword***, using a Forward Error Correction (FEC) encoder. The codeword is then transmitted. During transmission, the signal is subject to impairments, which may produce *bit errors* in the signal. At the receiver, the incoming signal is demodulated to produce a bit string that is similar to the original code word but may contain errors. This block is passed through an FEC decoder, with one of four possible outcomes:

1. If there are no bit errors, the input to the FEC decoder is identical to the original *codeword*, and the decoder produces the original data block as output.

2. For certain error patterns, it is possible for the decoder to detect and correct those errors. Thus, even though the incoming data block differs from the transmitted *codeword*, the FEC decoder is able to map this block into the original data block.

3. For certain error patterns, the decoder can detect but not correct the errors. In this case, the decode simply reports an uncorrectable error.

4. For certain, typically rare, error patterns, the decoder does not detect that any errors have occurred and maps the incoming n-bit data block into a k-bit block that differs from the original k-bit block.

How is it possible for the decoder to correct bit errors? In essence, error correction works by adding redundancy to the transmitted message. The redundancy makes it possible for the receiver to deduce what the original message was, even in the face of a certain level of error rate. In this section we look at a widely used form of error-correcting code known as a block error-correcting code. Our discussion only deals with basic principles; a discussion of specific error-correcting codes is b

Tools to understand Data Communications

There are two basic tools that will be utilized throughout this book in order to understand data communication principles:

a. **The Data Communications Model**

   The Data Communications Model describes the equipment, interfaces and communication medium that make up the communication path from the source to the destination. It concentrates on the physical layer.

b. **The OSI Model**

   The OSI model describes 7 basic layers that are required for data communications. The 7 layers make up what is called the protocol stack and describe the logical operation of a device. The term logical operation refers to the logic behind the communication process. It typically resides in software programming or rules. Generally it is not something physical that you can hold. Having said that, be aware that there is a component of the OSI model that does indeed describes the physical connection and rules.

Introduction to the ISO - OSI Model

The ISO (International Standards Organization) has created a layered model called the OSI (Open Systems Interconnect) model to describe defined layers in a network operating system. The purpose of the layers is to provide clearly defined functions to improve internetwork connectivity between "computer" manufacturing companies. Each layer has a standard defined input and a standard defined output.

Understanding the function of each layer is instrumental in understanding data communication within networks whether Local, Metropolitan or Wide.

This is a top-down explanation of the OSI Model, starting with the user's PC and what happens to the user's file as it passes though the different OSI Model layers. The top-down approach was selected specifically (as opposed to starting at the Physical Layer and working up to the Application Layer) for ease of understanding of how the user's files are transformed through the layers into a bit stream for transmission on the network.

There are 7 Layers of the OSI model and they are always presented in this manner starting with layer 7:

- 7. Application Layer (Top Layer)
- 6. Presentation Layer
- 5. Session Layer
- 4. Transport Layer
- 3. Network Layer
- 2. Data Link Layer
- 1. Physical Layer (Bottom Layer)

There are a few ways of remembering the OSI layers, one is the phrase "Please Do Not Take Salami Pizza Away".

- 7. Application Layer - Away
- 6. Presentation Layer - Pizza
- 5. Session Layer - Salami
- 4. Transport Layer - Take

- 3. Network Layer - Not
- 2. Data Link Layer - Do
- 1. Physical Layer - Please

Physical Layer - OSI Reference Model

Physical layer is the lowest layer of the OSI reference model. It is responsible for sending bits from one computer to another. This layer is not concerned with the meaning of the bits and deals with the setup of physical connection to the network and with transmission and reception of signals.

Functions of Physical Layer

Following are the various functions performed by the Physical layer of the OSI model.

1. **Representation of Bits:** Data in this layer consists of stream of bits. The bits must be encoded into signals for transmission. It defines the type of encoding i.e. how 0's and 1's are changed to signal.
2. **Data Rate:** This layer defines the rate of transmission which is the number of bits per second.
3. **Synchronization:** It deals with the synchronization of the transmitter and receiver. The sender and receiver are synchronized at bit level.
4. **Interface:** The physical layer defines the transmission interface between devices and transmission medium.
5. **Line Configuration:** This layer connects devices with the medium: Point to Point configuration and Multipoint configuration.
6. **Topologies:** Devices must be connected using the following topologies: Mesh, Star, Ring and Bus.
7. **Transmission Modes:** Physical Layer defines the direction of transmission between two devices: Simplex, Half Duplex, Full Duplex.
8. Deals with baseband and broadband transmission.

Design Issues with Physical Layer

- The Physical Layer is concerned with transmitting raw bits over a communication channel.
- The design issue has to do with making sure that when one side sends a 1 bit, it is received by the other side as a 1 bit and not as a 0 bit.
- **Typical questions here are:**
  - How many volts should be used to represent a 1 bit and how many for a 0?
  - How many nanoseconds a bit lasts?
  - Whether transmission may proceed simultaneously in both directions?
  - Whether transmission may proceed simultaneously in both directions?
  - How many pins the network connector has and what each pin is used for?
- The design issues here largely deal with mechanical, electrical and timing interfaces, and the physical transmission medium, which lies below the physical layer.

Data Link Layer - OSI Model

Data link layer performs the most reliable node to node delivery of data. It forms frames from the packets that are received from network layer and gives it to physical layer. It also synchronizes the information which is to be transmitted over the data. Error controlling is easily done. The encoded data are then passed to physical.

Error detection bits are used by the data link layer. It also corrects the errors. Outgoing messages are assembled into frames. Then the system waits for the acknowledgements to be received after the transmission. It is reliable to send message.

The main task of the **data link layer** is to transform a raw transmission facility into a line that appears free of undetected transmission errors to the network layer. It accomplishes this task by having the sender break up the input data into **data frames**(typically a few hundred or few thousand bytes) and transmit the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by send back an **acknowledgement frame**.

Functions of Data Link Layer

1. **Framing:** Frames are the streams of bits received from the network layer into manageable data units. This division of stream of bits is done by Data Link Layer.
2. **Physical Addressing:** The Data Link layer adds a header to the frame in order to define physical address of the sender or receiver of the frame, if the frames are to be distributed to different systems on the network.
3. **Flow Control:** A flow control mechanism to avoid a fast transmitter from running a slow receiver by buffering the extra bit is provided by flow control. This prevents traffic jam at the receiver side.
4. **Error Control:** Error control is achieved by adding a trailer at the end of the frame. Duplication of frames are also prevented by using this mechanism. Data Link Layers adds mechanism to prevent duplication of frames.
5. **Access Control:** Protocols of this layer determine which of the devices has control over the link at any given time, when two or more devices are connected to the same link.

Design Issues with Data Link Layer

- The issue that arises in the data link layer(and most of the higher layers as well) is how to keep a fast transmitter from drowning a slow receiver in data. Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Frequently, the flow regulation and the error handling are integrated.
- Broadcast networks have an additional issue in the data link layer: How to control access to the shared channel. A special sublayer of the data link layer, the Medium Access Control(MAC) sublayer, deals with this problem.

Network Layer - OSI Model

The network Layer controls the operation of the subnet. The main aim of this layer is to deliver packets from source to destination across multiple links (networks). If two computers (system) are connected on the same link, then there is no need for a network layer. It routes the signal through different channels to the other end and acts as a network controller.

It also divides the outgoing messages into packets and to assemble incoming packets into messages for higher levels.

In broadcast networks, the routing problem is simple, so the network layer is often thin or even non-existent.

Functions of Network Layer

1. It translates logical network address into physical address. Concerned with circuit, message or packet switching.
2. Routers and gateways operate in the network layer. Mechanism is provided by Network Layer for routing the packets to final destination.
3. Connection services are provided including network layer flow control, network layer error control and packet sequence control.
4. Breaks larger packets into small packets.

Design Issues with Network Layer

- A key design issue is **determining how packets are routed from source to destination**. Routes can be based on static tables that are wired into the network and rarely changed. They can also be highly dynamic, being determined anew for each packet, to reflect the current network load.
- If **too many packets** are present in the subnet at the same time, they will get into one another's way, forming **bottlenecks**. The **control of such congestion** also belongs to the network layer.
- Moreover, the **quality of service** provided(delay, transmit time, jitter, etc) is also a network layer issue.
- When a packet has to **travel from one network to another to get to its destination**, many problems can arise such as:
  - The addressing used by the second network may be different from the first one.
  - The second one may not accept the packet at all because it is too large.
  - The protocols may differ, and so on.
- It is up to the network layer to overcome all these problems to allow heterogeneous networks to be interconnected.

Transport Layer - OSI Model

The basic function of the Transport layer is to accept data from the layer above, split it up into smaller units, pass these data units to the Network layer, and ensure that all the pieces arrive correctly at the other end.

Furthermore, all this must be done efficiently and in a way that isolates the upper layers from the inevitable changes in the hardware technology.

The Transport layer also determines what type of service to provide to the Session layer, and, ultimately, to the users of the network. The most popular type of transport connection is an **error-free point-to-point channel** that delivers messages or bytes in the order in which they were sent.

The Transport layer is a true end-to-end layer, all the way from the source to the destination. In other words, a program on the source machine carries on a conversation with a similar program on the destination machine, using the message headers and control messages.

Functions of Transport Layer

1. **Service Point Addressing:** Transport Layer header includes service point address which is port address. This layer gets the message to the correct process on the computer unlike Network Layer, which gets each packet to the correct computer.

2. **Segmentation and Reassembling:** A message is divided into segments; each segment contains sequence number, which enables this layer in reassembling the message. Message is reassembled correctly upon arrival at the destination and replaces packets which were lost in transmission.
3. **Connection Control:** It includes 2 types:
   - Connectionless Transport Layer : Each segment is considered as an independent packet and delivered to the transport layer at the destination machine.
   - Connection Oriented Transport Layer : Before delivering packets, connection is made with transport layer at the destination machine.
4. **Flow Control:** In this layer, flow control is performed end to end.
5. **Error Control:** Error Control is performed end to end in this layer to ensure that the complete message arrives at the receiving transport layer without any error. Error Correction is done through retransmission.

Design Issues with Transport Layer

- Accepting data from Session layer, split it into segments and send to the network layer.
- Ensure correct delivery of data with efficiency.
- Isolate upper layers from the technological changes.
- Error control and flow control.

Session Layer - OSI Model

The Session Layer allows users on different machines to establish active communication sessions between them.

It's main aim is to establish, maintain and synchronize the interaction between communicating systems. Session layer manages and synchronize the conversation between two different applications. In Session layer, streams of data are marked and are resynchronized properly, so that the ends of the messages are not cut prematurely and data loss is avoided.

Functions of Session Layer

1. **Dialog Control :** This layer allows two systems to start communication with each other in half-duplex or full-duplex.
2. **Token Management:** This layer prevents two parties from attempting the same critical operation at the same time.
3. **Synchronization :** This layer allows a process to add checkpoints which are considered as synchronization points into stream of data. Example: If a system is sending a file of 800 pages, adding checkpoints after every 50 pages is recommended. This ensures that 50 page unit is successfully received and acknowledged. This is beneficial at the time of crash as if a crash happens at page number 110; there is no need to retransmit 1 to100 pages.

Design Issues with Session Layer

- To allow machines to establish sessions between them in a seamless fashion.
- Provide enhanced services to the user.
- To manage dialog control.

- To provide services such as **Token management** and **Synchronization**.

Presentation Layer - OSI Model

The primary goal of this layer is to take care of the **syntax** and **semantics** of the information exchanged between two communicating systems. Presentation layer takes care that the data is sent in such a way that the receiver will understand the information(data) and will be able to use the data. Languages(syntax) can be different of the two communicating systems. Under this condition presentation layer plays a role translator.

In order to make it possible for computers with different data representations to communicate, the data structures to be exchanged can be defined in an **abstract** way. The presentation layer manages these **abstract data structures** and allows higher-level data structures(eg: banking records), to be defined and exchanged.

Functions of Presentation Layer

1. **Translation:** Before being transmitted, information in the form of characters and numbers should be changed to bit streams. The presentation layer is responsible for interoperability between encoding methods as different computers use different encoding methods. It translates data between the formats the network requires and the format the computer.
2. **Encryption:** It carries out encryption at the transmitter and decryption at the receiver.
3. **Compression:** It carries out data compression to reduce the bandwidth of the data to be transmitted. The primary role of Data compression is to reduce the number of bits to be 0transmitted. It is important in transmitting multimedia such as audio, video, text etc.

Design Issues with Presentation Layer

- To manage and maintain the **Syntax** and **Semantics** of the information transmitted.
- **Encoding data** in a standard agreed upon way. Eg: String, double, date, etc.
- Perform **Standard Encoding** on wire.

Application Layer - OSI Model

It is the top most layer of OSI Model. Manipulation of data(information) in various ways is done in this layer which enables user or software to get access to the network. Some services provided by this layer includes: E-Mail, transferring files, distributing the results to user, directory services, network resources, etc.

The Application Layer contains a variety of protocols that are commonly needed by users. One widely-used application protocol is **HTTP(HyperText Transfer Protocol)**, which is the basis for the World Wide Web. When a browser wants a web page, it sends the name of the page it wants to the server using HTTP. The server then sends the page back.

Other Application protocols that are used are: **File Transfer Protocol(FTP)**, **Trivial File Transfer Protocol(TFTP)**, **Simple Mail Transfer Protocol(SMTP)**, **TELNET**, **Domain Name System(DNS)** etc.

Functions of Application Layer

1. **Mail Services:** This layer provides the basis for E-mail forwarding and storage.
2. **Network Virtual Terminal:** It allows a user to log on to a remote host. The application creates software emulation of a terminal at the remote host. User's computer talks to the software terminal which in turn talks to the host and vice versa. Then the remote host believes it is communicating with one of its own terminals and allows user to log on.
3. **Directory Services:** This layer provides access for global information about various services.

4. **File Transfer, Access and Management (FTAM):** It is a standard mechanism to access files and manages it. Users can access files in a remote computer and manage it. They can also retrieve files from a remote computer.

Design Issues with Application Layer

There are commonly reoccurring problems that occur in the design and implementation of Application Layer protocols and can be addressed by patterns from several different pattern languages:

- Pattern Language for Application-level Communication Protocols
- Service Design Patterns
- Patterns of Enterprise Application Architecture
- Pattern-Oriented Software Architecture

Channel Characteristics:

With any communications system, the signal that is received may differ from the signal that is transmitted, due to various transmission impairments. For analog signals, these impairments introduce various random modifications that degrade the signal quality. For digital signals, bit errors may be introduced, such that a binary 1 is transformed into a binary 0 or vice versa. In this section, we examine the various impairments and how they may affect the information-carrying capacity of a communication link. Moreover, we have discussed the tools of transmitting data (signals) over a network and how the data behave. One important issue in networking is the performance of the network—how good is it? In this section, also, we introduce terms that we need in this part.

Transmission Impairments

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 (see Figure 1).

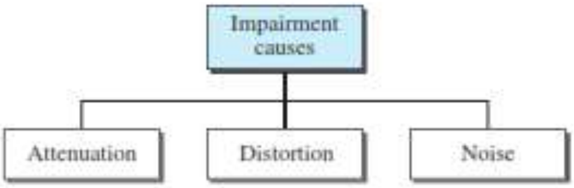


Figure : Causes of impairment

- **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. Figure 2 shows the effect of attenuation and amplification.

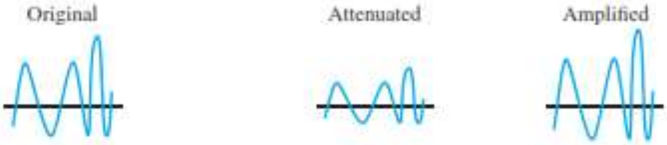


Figure 2. Attenuation

- 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 (see the next section) 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 3 shows the effect of distortion on a composite signal.

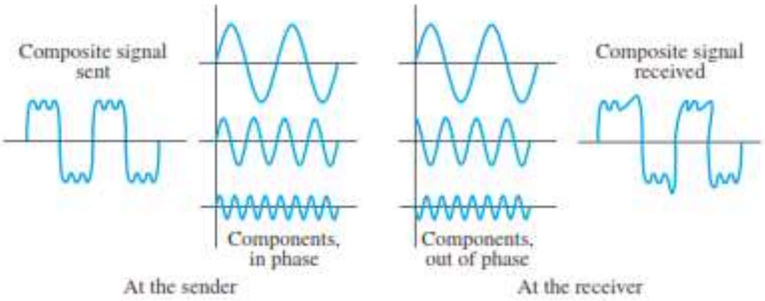


Figure 3. Distortion

- 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 acts 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 4 shows the effect of noise on a signal.

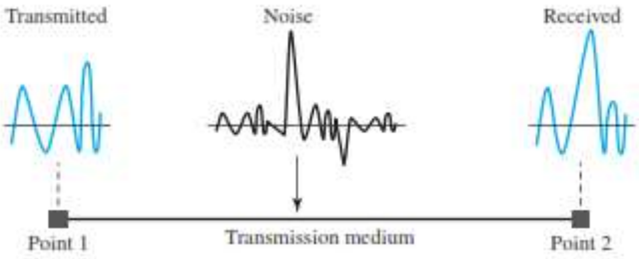


Figure 4. Noise

Performance

Performance refers to measures of service quality of a network as seen by the customer. Also, it is the analysis and review of collective network statistics, to define the quality of services offered by the underlying computer network. It is a qualitative and quantitative process that measures and defines the performance level of a given network. There are many different ways to measure the performance of a network, as each network is different in nature and design.

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***.

Communication companies such as American Telephone and Telegraph (AT&T) and Western Union are called common carriers, and they provide three general classes of service for both voice and data communication:

- Narrowband handles low data volumes. Data transmission rates are from 45 to 300 baud. The low-speed devices might use narrow band communications.

- Voiceband handles moderate data transmission volumes between 300 and 9600 baud. They are used for applications ranging from operating a CRT to running a line printer. Their major application is for telephone voice communication hence, the term voiceband.

- Broadband handles very large volumes of data. These systems provide data transmission rates of 1 million baud or more. High-speed data analysis and satellite communications are examples of broadband communication systems.

Bandwidth in Hertz

It is the range of frequencies contained in a composite signal or the range of frequencies a channel can pass. For example, we can say the bandwidth of a subscriber telephone line is 4 kHz.

Bandwidth in Bits per Seconds

The term bandwidth can also refer to the number of bits per second (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.

Note:- *There is an explicit relationship between the bandwidth in hertz and bandwidth in bits per second. Basically, an increase in bandwidth in hertz means an increase in bandwidth in bits per second. The relationship depends on whether we have baseband transmission or transmission with modulation.*

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. Imagine a highway designed to transmit 1000 cars per minute from one point to another. However, if there is congestion on the road, this figure may be reduced to 100 cars per minute. The bandwidth is 1000 cars per minute; the throughput is 100 cars per minute.

Example: A network with bandwidth of 10 Mbps can pass only an average of 12,000 frames per minute with each frame carrying an average of 10,000 bits. What is the throughput of this network?

We can calculate the throughput as:

Throughput = (12000 * 10000)/60 = 2000000 bps = 2 Mbps

The throughput is almost one-fifth of the bandwidth in this case.

Latency (Delay)

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)

The propagation speed of electromagnetic signals depends on the medium and on the frequency of the signal. For example, in a vacuum, light is propagated with a speed of 3 × 108 m/s. It is lower in air; it is much lower in cable.

Example: What is the propagation time if the distance between the two points is 12,000 km? Assume the propagation speed to be 2.4 × 108 m/sin cable.

We can calculate the propagation time as

Propagation time = (12000 * 1000) / (2.4 × 108) = 50 ms

The example shows that a bit can go over the Atlantic Ocean in only 50 ms if there is a direct cable between the source and the destination.

□ ***Transmission Time***: 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 transmission time of a message depends on the size of the message and the bandwidth of the channel.

Transmission time = (Message size) / Bandwidth

Example: What are the propagation time and the transmission time for a 2.5-KB (kilobyte) message (an email) if the bandwidth of the network is 1 Gbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4 * 108 m/s.

We can calculate the propagation and transmission time as:

Propagation time = (12000 * 1000) / (2.4 * 108) = 50 ms

Transmission time = (2500 * 8) / 10^9 = 0.02 ms

Note that in this case, because the message is short and the bandwidth is high, the dominant factor is the propagation time, not the transmission time. The transmission time can be ignored.

Example: What are the propagation time and the transmission time for a 5-MB (megabyte) message (an image) if the bandwidth of the network is 1 Mbps? Assume that the distance between the sender and the receiver is 12,000 km and that light travels at 2.4 * 10^8 m/s.

We can calculate the propagation and transmission times as:

Propagation time = (12000 * 1000) / (2.4 * 10^8) = 50 ms

Transmission time = (5000000 * 8) / 10^6 = 40 s

Note that in this case, because the message is very long and the bandwidth is not very high, the dominant factor is the transmission time, not the propagation time. The propagation time can be ignored.

□ ***Queuing Time:*** 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.

□ ***Processing Delay:*** processing delay is the time it takes nodes to process the packet header. Processing delay is a key component in network delay. During processing of a packet, nodes may check for bit-level errors in the packet that occurred during transmission as well as determining where the packet's next destination is.

Jitter

Another performance issue that is related to delay is jitter. Jitter is defined as a variation in the delay of received packets (see Figure 5). The sending side transmits packets in a continuous stream and spaces them evenly apart. Because of network congestion, improper queuing, or configuration errors, the delay between packets can vary instead of remaining constant. This variation causes problems for audio playback at the receiving end. Playback may experience gaps while waiting for the arrival of variable delayed packets. 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.

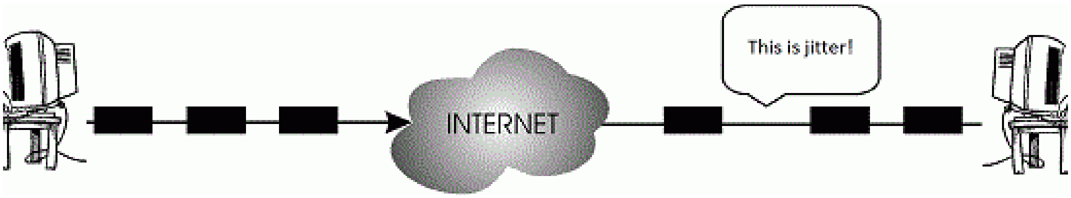


Figure 5. Jitter

MODEM

A modem (modulator-demodulator) is a device that modulates an analog carrier signal to encode digital information, and also demodulates such a carrier signal to decode the transmitted information. The goal is to produce a signal that can be transmitted easily and decoded to reproduce the original digital data. Modems can be used over any means of transmitting analog signals, from light emitting diodes to radio.

A modem modulates outgoing digital signals from a computer or other digital device to analog signals for a conventional copper twisted pair telephone line and demodulates the incoming analog signal and converts it to a digital signal for the digital device.

In recent years, the 2400 bits per second modem that could carry e-mail has become obsolete. 14.4 Kbps and 28.8 Kbps modems were temporary landing places on the way to the much higher bandwidth devices and carriers of tomorrow. From early 1998, most new personal computers came with 56 Kbps modems. By comparison, using a digital Integrated Services Digital Network adapter instead of a conventional modem, the same telephone wire can now carry up to 128 Kbps. With Digital Subscriber Line (DSL) systems, now being deployed in a number of communities, bandwidth on twisted-pair can be in the megabit range.

2.6.1 Types of Modems

□ Landline Modems

□ Wireless Modems

□ LAN Modems

A. Landline Modems:

Landline modems are modems which connect to the public switched telephone network (PSTN). To connect to PSTN, these modems have a jack known as RJ-11, or regular phone jack. A telephone cable with a RJ-11 plug connects the modem to the nearest phone jack, which also conforms to the RH-11standard. Landline modems can be further classified into the followings types:

1. Internal modems: This device is a circuit board that plugs into one of the expansion slots of the computer. Internal modems usually are cheaper than external modems, but when problems occur, fixing and troubleshooting the modem can sometimes prove to be quite difficult. The telephone line plugs into the modem port in the back of the computer. Most internal modems come installed in the computer you buy. Internal modems are more directly integrated into the computer system and, therefore, do not need any special attention. Internal modems are activated when you run a communications program and are turned off when you exit the program. This convenience is especially useful for novice users.

Internal modems usually cost less than external modems, but the price difference is usually small. The major disadvantage with internal modems is their location: inside the computer. When you want to replace an internal modem you have to go inside the computer case to make the switch.

2. External modems: This device is attached to the back of the computer by way of a cable that plugs into the modem port. It is usually less expensive and very portable. It can be used with other computers very easily by unplugging it and plugging it into another computer. This is the simplest type of modem to install because you don’t have to open the computer. External modems have their own power supply and connect with a cable to a computer’s serial port.

The telephone line plugs into a socket on the rear panel of the modem. Because external modems have their own power supply, you can turn off the modem to break an online connection quickly without powering down the computer. Another advantage over an internal modem is that an external modem’s separate power supply does not drain any power from the computer. You also can monitor your modem’s connection activity by watching the status lights.

3. Voice/data/fax modems: This device can be hooked up to your telephone and used to send information to your computer. Your computer can also send information to a fax machine. Most computer modems are modems with faxing capabilities.

4. PC Card modem: These modems, designed for portable computers, are the size of a credit card and fit into the PC Card slot on notebook and handheld computers. These modems are removed when the odem is not needed.

Except for their size, PC Card modems are like a combination of external and internal modems. These devices are plugged directly into an external slot in the portable computer, so no cable is required other than the telephone line connection. The cards are powered by the computer, which is fine unless the computer is battery-operated.

Running a PC Card modem while the portable computer is operating on battery power drastically decreases the life of your batteries.

B. Wireless Modems:

Wireless modems are radio transmitters/receivers installed into mobile computing devices (i.e. devices that are used while you are moving such as mobile phones, laptops etc.) Using wireless modems, one can connect to a network while being mobile. Unlike landline modems, wireless modems do not plug into an RJ-11 jack.

C. LAN Modems:

LAN modems allow shared remote access to LAN (Local Area Network) resources. LAN modem comes fully preconfigured for single particular network architecture such as Ethernet or Token Ring and/or particular network software such as IPX, NetBIOS, NetBEUI etc.

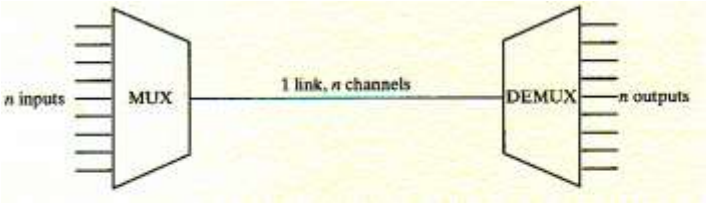
Working of Modem

Modems convert analog data transmitted over phone lines into digital data so that computers can read; they also convert digital data into analog data so it can be transmitted. This process involves modulating and demodulating the computer's digital signals into analog signals that travel over the telephone lines. In other words, the modem ranslates computer data into the language used by telephones and then reverses the process to translate the responding data back into computer language.

Multiplexing:

Multiplexing is the transmission of multiple signals on one medium. For a medium to transmit multiple signals simultaneously, the signals must be altered so that they do not interfere with one another.

Generally, two communicating stations will not utilize the full capacity of a data link. Example: transmit a voice signal over a optical fiber. Multiplexing allows multiple users sharing the capacity of a transmission link.



Components

– Multiplexer: Combines data from the n input lines

– Link: with *n* separate channels, • example: optical fiber or microwave link

– Demultiplexer: Separates the data according to channel that delivers them to the appropriate output lines

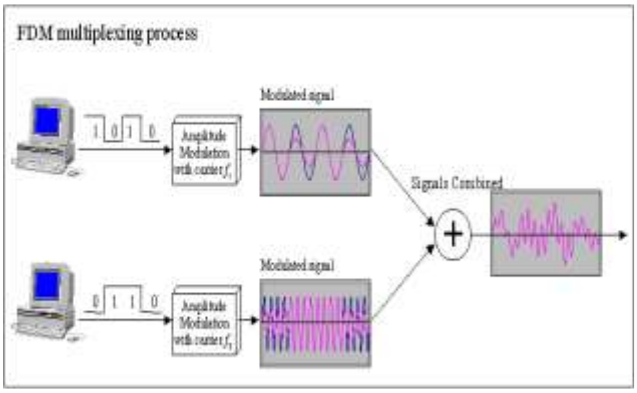
Multiplexing Techniques

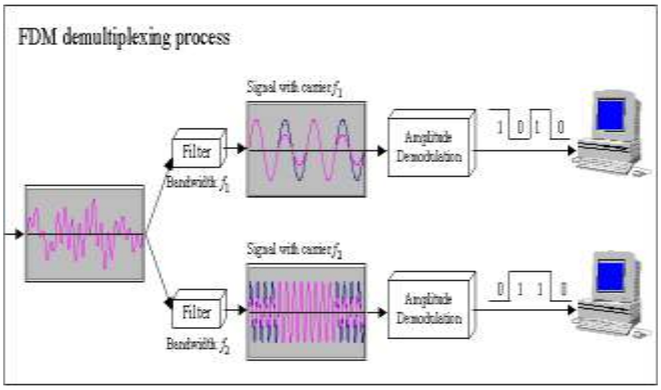
The set of techniques that allows the simultaneous transmission of multiple signals across a single data link.

1. Frequency-Division Multiplexing (FDM)
2. Wavelength-Division Multiplexing (WDM)
3. Time-Division Multiplexing (TDM)
4. Code-Division Multiplexing (CDM)

Frequency-division multiplexing (FDM):– It is most popular and is used extensively in radio and TV transmission. Here the frequency spectrum is divided into several logical channels, giving each user exclusive possession of a particular frequency band. each channel occupies a fraction of the bandwidth of the link whereby a channel is defined by its center frequency, and its bandwidth. Example: radio and television signal transmission.

- Each logical channel is transmitted on a separate frequency.
- Television and radio uses FDM to broadcast many channels over the same media.
- Filters separate the multiplexed signal back into its constituent component signals





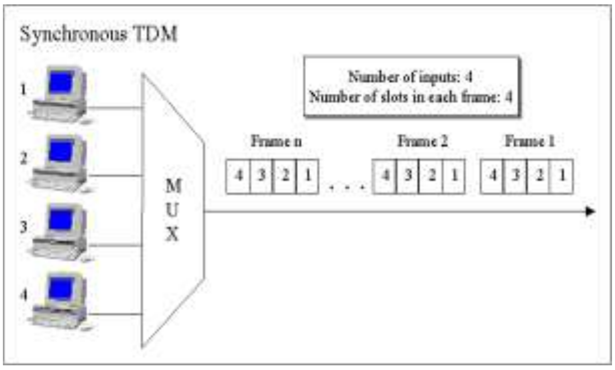
Wavelength-Division Multiplexing

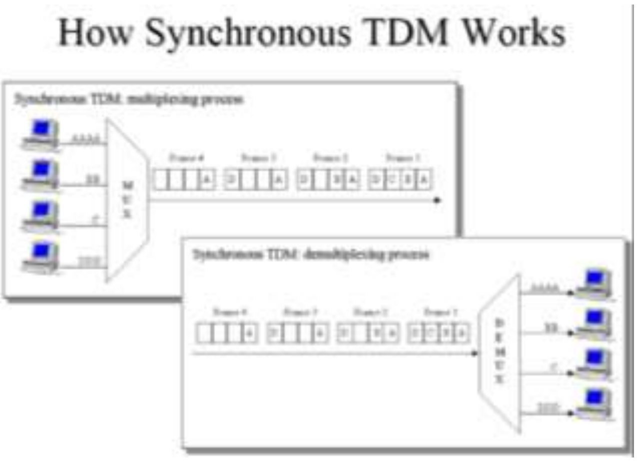
Wavelength-division multiplexing (WDM) is conceptually same as the FDM, except that the multiplexing and demultiplexing involves light signals transmitted through fibre-optic channels. The idea is the same: we are combining different frequency signals. However, the difference is that the frequencies are very high. It is designed to utilize the high data rate capability of fibre-optic cable

Time-division multiplexing (TDM): It is also called synchronous TDM, which is commonly used for multiplexing digitized voice stream. The users take turns using the entire channel for short burst of time. In frequency division multiplexing, all signals operate at the same time with different frequencies, but in Time-division multiplexing all signals operate with same frequency at different times.

- Each channel occupies the entire bandwidth of the link for a very short period of time
- A channel is made up of a sequence of time slot
- Example: multiplexing digitalized voice signals and data streams
- TDM can be implemented in two ways
  - Synchronous TDM
  - Asynchronous TDM

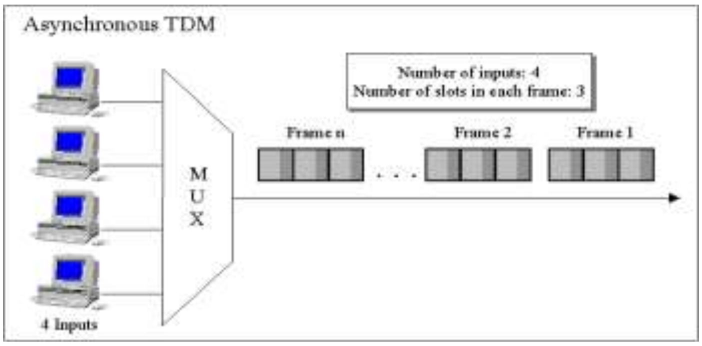
  a. Synchronous TDM
    - The multiplexer allocates exactly the same time slot to each device at all times, whether or not a device has anything to transmit
    - A frame consists of one complete cycle of time slots. Thus the number of slots in frame is equal to the number of inputs.

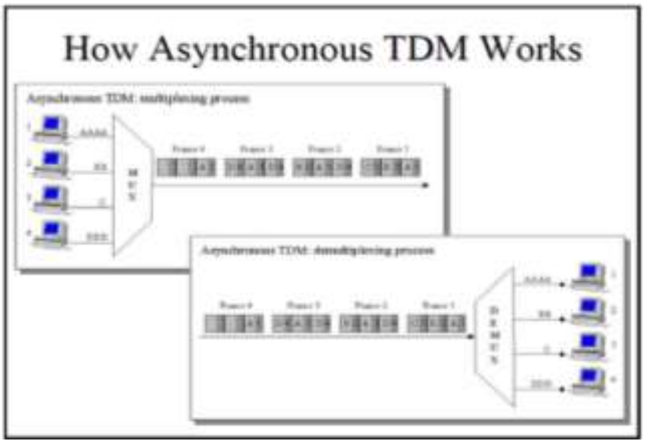



How Synchronous TDM Works
Synchronous TDM: multiplexing process
M U X
A
D A
D B A
D C B A
Synchronous TDM: demultiplexing process
D E M U X

  b. Asynchronous TDM (or statistical time-division multiplexing)
    - Each slot in a frame is not dedicated to the fix device
    - The number of slots in a frame is not necessary to be equal to the number of input devices. More than one slots in a frame can be allocated for an input device.

- – Allows maximum utilization of the link. It allows a number of lower speed input lines to be multiplexed to a single higher speed line




How Asynchronous TDM Works
Asynchronous TDM: multiplexing process
MUX
Asynchronous TDM: demultiplexing process
DEMUX

In asynchronous TDM, a frame contains a fix number of time slots. Each slot has an index of which device to receive.

Code Division Multiplexing

• Sends many signals or "chips" per bit.
• Each sender uses a unique pattern of chips.
• May use multiple frequencies for spread spectrum communication.
• Common with wireless systems.

Internet Protocol (IP)

Internet Protocol is a set of technical rules that defines how computers communicate over a network. There are currently two versions: IP version 4 (IPv4) and IP version 6 (IPv6).

What is IPv4? IPv4 was the first version of Internet Protocol to be widely used, and accounts for most of today's Internet traffic. There are just over 4 billion IPv4 addresses. While that is a lot of IP addresses, it is not enough to last forever.

IP v4 Problems

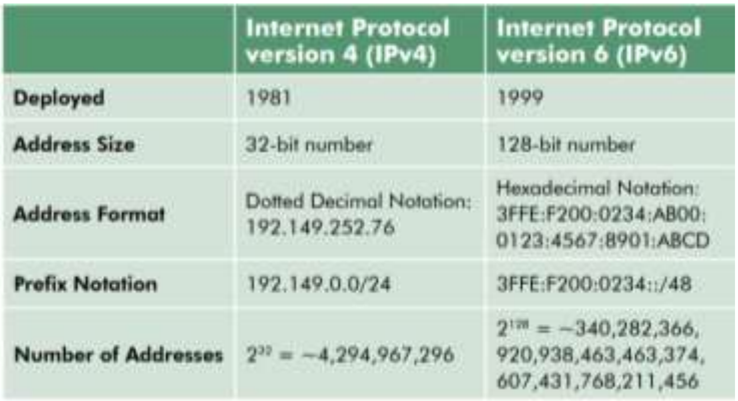
- IP address starvation Distribution of addresses (USA >50%)
- Routing is complicated
- Realization of new technologies (Mobile computing, real time services, multicast, security, QOS, etc.)

What is IPv6? IPv6 is a newer numbering system that provides a much larger address pool than IPv4. It was deployed in 1999 and should meet the world's IP addressing needs well into the future.

IP version 6: advantages

- much more adresses available (2^128)
- no fragmentation in routers
- efficient routing
- no checksum in header security functions (e.g. IPSEC)
- auto-configuration

What is the major difference? The major difference between IPv4 and IPv6 is the number of IP addresses. There are 4,294,967,296 IPv4 addresses. In contrast, there are 340,282,366,920,938,463,463,374, 607,431,768,211,456 IPv6 addresses. The technical functioning of the Internet remains the same with both versions and it is likely that both versions will continue to operate simultaneously on networks well into the future. To date, most networks that use IPv6 support both IPv4 and IPv6 addresses in their networks.



	Internet Protocol version 4 (IPv4)	Internet Protocol version 6 (IPv6)
Deployed	1981	1999
Address Size	32-bit number	128-bit number
Address Format	Dotted Decimal Notation: 192.149.252.76	Hexadecimal Notation: 3FFE:F200:0234:AB00: 0123:4567:8901:ABCD
Prefix Notation	192.149.0.0/24	3FFE:F200:0234::/48
Number of Addresses	2^{32} = ~4,294,967,296	2^{128} = ~340,282,366, 920,938,463,463,374, 607,431,768,211,456

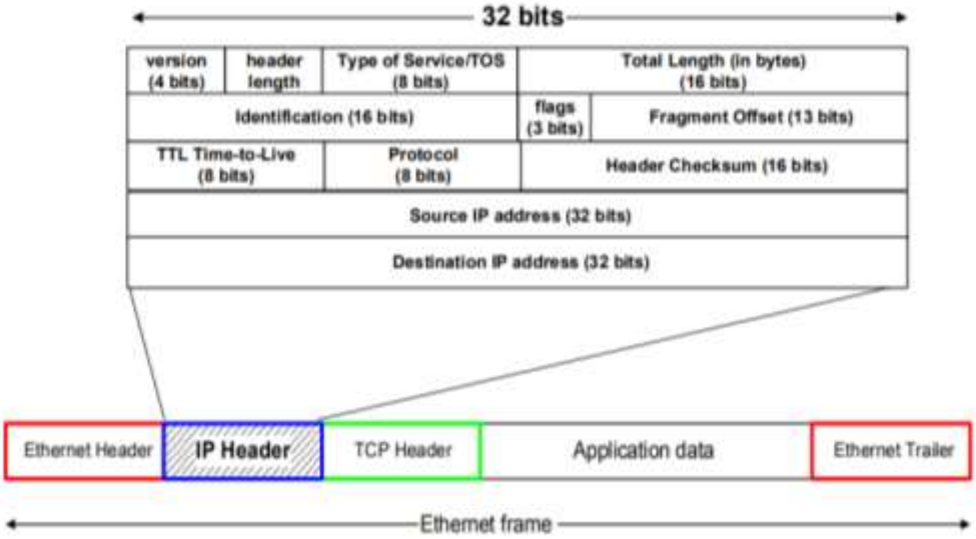
Is IPv4 or IPv6 better?

IPv4 is the fourth version of the Internet Protocol (IP), while IPv6 is the most recent version of the Internet Protocol. Therefore, IPv6 is more advanced, secure, and faster compared to IPv4.

What Is an IP Address?

IP address stands for internet protocol address; is a unique global address for a network interface for identifying number that is associated with a specific computer or computer network. When connected to the internet, the IP address allows the computers to send and receive information.

• An IP address: - is a 32 bit long identifier - encodes a network number (network prefix) and a host number


32 bits
version (4 bits)
header length
Type of Service/TOS (8 bits)
Total Length (in bytes) (16 bits)
Identification (16 bits)
flags (3 bits)
Fragment Offset (13 bits)
TTL Time-to-Live (8 bits)
Protocol (8 bits)
Header Checksum (16 bits)
Source IP address (32 bits)
Destination IP address (32 bits)
Ethernet Header
IP Header
TCP Header
Application data
Ethernet Trailer
Ethernet frame

KEY TAKEAWAYS

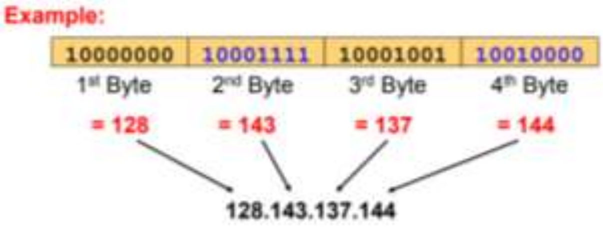
- An internet protocol (IP) address allows computers to send and receive information.
- There are four types of IP addresses: public, private, static, and dynamic.
- An IP address allows information to be sent and received by the correct parties, which means it can also be used to track down a user's physical location in some instances.

Why do we need IP addresses?

- A MAC address has no structure, so it tells a switch the identity (who) of the destination interface but not its location (where).
- With just MAC addresses, switches would have to resort to broadcast the first time they encounter a new address.
- Switch forwarding table sizes would be on the order of the total number of MAC addresses**.**

Dotted Decimal Notation

IP addresses are written in a so-called dotted decimal notation • Each byte is identified by a decimal number in the range [0..255]:



How an IP Address Works

An IP address allows computers to send and receive data over the internet. Most IP addresses are purely numerical, but as internet usage grows, letters have been added to some addresses.

There are four different types of IP addresses: public, private, static, and dynamic. While the public and private are indicative of the location of the network—private being used inside a network while the public is used outside of a network—static and dynamic indicate permanency.

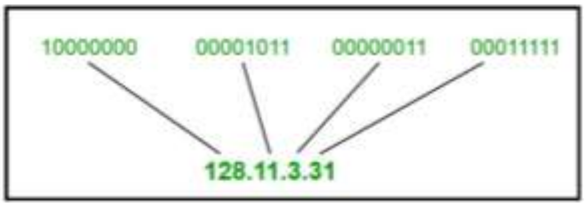
A static IP address is one that was manually created, as opposed to having been assigned. A static address also does not change, whereas a dynamic IP address has been assigned by a Dynamic Host Configuration Protocol (DHCP) server and is subject to change. Dynamic IP addresses are the most common type of internet protocol addresses. Dynamic IP addresses are only active for a certain amount of time, after which they expire. The computer will either automatically request a new lease, or the computer may receive a new IP address.

An IP address can be compared to a Social Security Number (SSN) since each one is completely unique to the computer or user it is assigned to. The creation of these numbers allows routers to identify where they are sending information on the internet. They also make sure that the correct devices are receiving what is being sent. Much like the post office needs a mailing address to deliver a package, a router needs an IP address to deliver to the web address requested.

Classful IP Addressing

IP address is an address having information about how to reach a specific host, especially outside the LAN. An IP address is a 32 bit unique address having an address space of 2^{32}. Generally, there are two notations in which IP address is written, dotted decimal notation and hexadecimal notation.

Dotted Decimal Notation:



The 32 bit IP address is divided into five sub-classes. These are:

- Class A
- Class B

- Class C
- Class D
- Class E

Each of these classes has a valid range of IP addresses. Classes D and E are reserved for multicast and experimental purposes respectively. The order of bits in the first octet determine the classes of IP address.

IPv4 address is divided into two parts:

- **Network ID**
- **Host ID**

The class of IP address is used to determine the bits used for network ID and host ID and the number of total networks and hosts possible in that particular class. Each ISP or network administrator assigns IP address to each device that is connected to its network.

Class A:

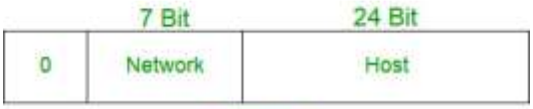
IP address belonging to class A are assigned to the networks that contain a large number of hosts.

- The network ID is 8 bits long.
- The host ID is 24 bits long.

The higher order bit of the first octet in class A is always set to 0. The remaining 7 bits in first octet are used to determine network ID. The 24 bits of host ID are used to determine the host in any network. The default subnet mask for class A is 255.x.x.x. Therefore, class A has a total of:

- 2^7-2= 126 network ID(Here 2 address is subtracted because 0.0.0.0 and 127.x.y.z are special address.)
- 2^24 – 2 = 16,777,214 host ID

IP addresses belonging to class A ranges from 1.x.x.x – 126.x.x.xx



Class A

Class B:

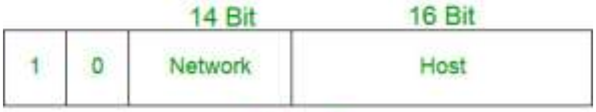
IP address belonging to class B are assigned to the networks that ranges from medium-sized to large-sized networks.

- The network ID is 16 bits long.
- The host ID is 16 bits long.

The higher order bits of the first octet of IP addresses of class B are always set to 10. The remaining 14 bits are used to determine network ID. The 16 bits of host ID is used to determine the host in any network. The default sub-net mask for class B is 255.255.x.x. Class B has a total of:

- 2^14 = 16384 network address
- 2^16 – 2 = 65534 host address

IP addresses belonging to class B ranges from 128.0.x.x – 191.255.x.x.



Class B

Class C:

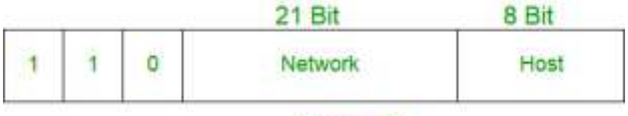
IP address belonging to class C are assigned to small-sized networks.

- The network ID is 24 bits long.
- The host ID is 8 bits long.

The higher order bits of the first octet of IP addresses of class C are always set to 110. The remaining 21 bits are used to determine network ID. The 8 bits of host ID is used to determine the host in any network. The default sub-net mask for class C is 255.255.255.x. Class C has a total of:

- 2^21 = 2097152 network address
- 2^8 – 2 = 254 host address

IP addresses belonging to class C ranges from 192.0.0.x – 223.255.255.x.



Class C

Class D:

IP address belonging to class D are reserved for multi-casting. The higher order bits of the first octet of IP addresses belonging to class D are always set to 1110. The remaining bits are for the address that interested hosts recognize.

Class D does not posses any sub-net mask. IP addresses belonging to class D ranges from 224.0.0.0 – 239.255.255.255

Class E:

IP addresses belonging to class E are reserved for experimental and research purposes. IP addresses of class E ranges from 240.0.0.0 – 255.255.255.254. This class doesn't have any sub-net mask. The higher order bits of first octet of class E are always set to 1111.

Rules for assigning Host ID:

Host ID's are used to identify a host within a network. The host ID are assigned based on the following rules:

- Within any network, the host ID must be unique to that network.
- Host ID in which all bits are set to 0 cannot be assigned because this host ID is used to represent the network ID of the IP address.
- Host ID in which all bits are set to 1 cannot be assigned because this host ID is reserved as a broadcast address to send packets to all the hosts present on that particular network.

Rules for Assigning Network ID:

Hosts that are located on the same physical network are identified by the network ID, as all host on the same physical network is assigned the same network ID. The network ID is assigned based on the following rules:

- The network ID cannot start with 127 because 127 belongs to class A address and is reserved for internal loop-back functions.
- All bits of network ID set to 1 are reserved for use as an IP broadcast address and therefore, cannot be used.
- All bits of network ID set to 0 are used to denote a specific host on the local network and are not routed and therefore, aren't used.

Problems with Classful Addressing:

- The problem with this classful addressing method is that millions of class A address are wasted, many of the class B address are wasted, whereas, number of addresses available in class C is so small that it cannot cater the needs of organizations. Class D addresses are used for multicast routing and are therefore available as a single block only. Class E addresses are reserved.
- Since there are these problems, Classful networking was replaced by Classless Inter-Domain Routing (CIDR) in 1993.

Unit 4

Mobile Internet Protocol (or Mobile IP)

Mobile IP is a communication protocol (created by extending Internet Protocol, IP) that allows the users to move from one network to another with the same IP address. It ensures that the communication will continue without user's sessions or connections being dropped. This is an IETF (Internet Engineering Task Force) standard communications protocol designed to allow mobile devices' (such as laptop, PDA, mobile phone, etc.) users to move from one network to another while maintaining their permanent IP (Internet Protocol) address.

Defined in RFC (Request for Comments) 2002, mobile IP is an enhancement of the internet protocol (IP) that adds mechanisms for forwarding internet traffic to mobile devices (known as mobile nodes) when they are connecting through other than their home network.

Terminologies:

• Mobile Node (MN):

It is the hand-held communication device that the user caries e.g. Cell phone.

• Home Network:

• It is a network to which the mobile node originally belongs to as per its assigned IP address (home address).

• Home Agent (HA):

It is a router in home network to which the mobile node was originally connected

• Home Address:

It is the permanent IP address assigned to the mobile node (within its home network).

• Foreign Network:

It is the current network to which the mobile node is visiting (away from its home network).

• Foreign Agent (FA):

It is a router in foreign network to which mobile node is currently connected. The packets from the home agent are sent to the foreign agent which delivers it to the mobile node.

- Correspondent Node (CN):

It is a device on the internet communicating to the mobile node.

- Care of Address (COA):

It is the temporary address used by a mobile node while it is moving away from its home network.

Working:

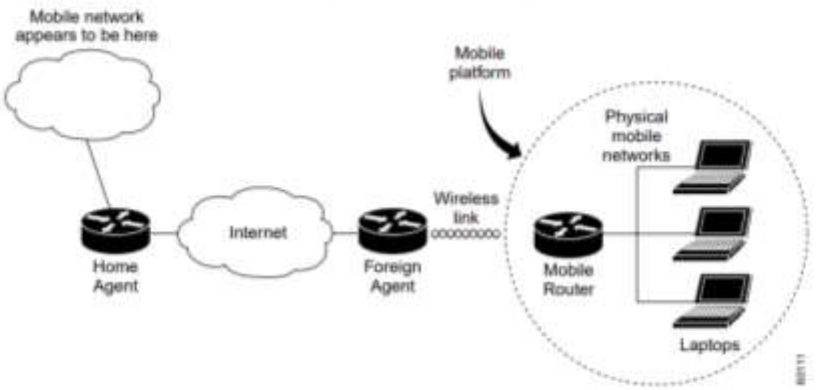
Correspondent node sends the data to the mobile node. Data packets contains correspondent node's address (Source) and home address (Destination). Packets reaches to the home agent. But now mobile node is not in the home network, it has moved into the foreign network. Foreign agent sends the care-of-address to the home agent to which all the packets should be sent. Now, a tunnel will be established between the home agent and the foreign agent by the process of tunneling.

Tunneling establishes a virtual pipe for the packets available between a tunnel entry and an endpoint. It is the process of sending a packet via a tunnel and it is achieved by a mechanism called encapsulation.

Now, home agent encapsulates the data packets into new packets in which the source address is the home address and destination is the care-of-address and sends it through the tunnel to the foreign agent.

Foreign agent, on other side of the tunnel receives the data packets, decapsulates them and sends them to the mobile node. Mobile node in response to the data packets received, sends a reply in response to foreign agent. Foreign agent directly sends the reply to the correspondent node.

Figure 3-1 Cisco Mobile Network Components and Relationships



`

The mobile access router functions similarly to the mobile node with one key difference—the mobile access router allows entire networks to roam. For example, an airplane with a mobile access router can fly around the world while passengers stay connected to the Internet. This communication is accomplished by Mobile IP aware routers tunneling packets, which are destined to hosts on the mobile networks, to the location where the mobile access router is visiting. The mobile access router then forwards the packets to the destination device.

Key Mechanisms in Mobile IP:

1. AgentDiscovery:

Agents advertise their presence by periodically broadcasting their agent advertisement messages. The mobile node receiving the agent advertisement messages observes whether the message is from its own home agent and determines whether it is in the home network or foreign network.

1. Agent Registration:

Mobile node after discovering the foreign agent, sends registration request (RREQ) to the foreign agent. Foreign agent in turn, sends the registration request to the home agent with the care-of-address. Home agent sends registration reply (RREP) to the foreign agent. Then it forwards the registration reply to the mobile node and completes the process of registration.

2. Tunneling:

It establishes a virtual pipe for the packets available between a tunnel entry and an endpoint.

It is the process of sending a packet via a tunnel and it is achieved by a mechanism called encapsulation. It takes place to forward an IP datagram from the home agent to the care-ofaddress. Whenever home agent receives a packet from correspondent node, it encapsulates

the packet with source address as home address and destination as care-of-address.

Route Optimization in Mobile IP:

The route optimization adds a conceptual data structure, the binding cache, to the correspondent

node. The binding cache contains

2. bindings for mobile node’s home address and its current care-of-address. Every time the

home agent receives a IP datagram that is destined to a mobile node currently away from the

home network, it sends a binding update to the correspondent node to update the

information in the correspondent node’s binding cache. After this the correspondent node

can directly tunnel packets to the mobile node.

Process of Mobile IP

The mobile IP process has following three main phases, which are:

1. Agent Discovery

During the agent discovery phase the HA and FA advertise their services on the network by

using the ICMP router discovery protocol (IROP).

Mobile IP defines two methods: agent advertisement and agent solicitation which are in fact

router discovery methods plus extensions.

o Agent advertisement: For the first method, FA and HA advertise their presence

periodically using special agent advertisement messages. These messages advertisement

can be seen as a beacon broadcast into the subnet. For this advertisement internet control

message protocol (ICMP) messages according to RFC 1256, are used with some mobility

extensions.

o Agent solicitation: If no agent advertisements are present or the inter arrival time is too

high, and an MN has not received a COA, the mobile node must send agent solicitations.

These solicitations are again bases on RFC 1256 for router solicitations.

2. Registration

The main purpose of the registration is to inform the home agent of the current location for correct forwarding of packets.

Registration can be done in two ways depending on the location of the COA.

o If the COA is at the FA, the MN sends its registration request containing the COA to the FA which is forwarding the request to the HA. The HA now set up a mobility binding containing the mobile node's home IP address and the current COA.

Additionally, the mobility biding contains the lifetime of the registration which is negotiated during the registration process. Registration expires automatically after the lifetime and is deleted; so a mobile node should register before expiration. After setting up the mobility binding, the HA send a reply message back to the FA which forwards it to the MN.

o If the COA is co-located, registration can be very simpler. The mobile node may send the request directly to the HA and vice versa. This by the way is also the registration procedure for MNs returning to their home network.

3. Tunneling

A tunnel is used to establish a virtual pipe for data packets between a tunnel entry and a tunnel endpoint. Packets which are entering in a tunnel are forwarded inside the tunnel and leave the tunnel unchanged. Tunneling, i.e., sending a packet through a tunnel is achieved with the help of encapsulation.

Tunneling is also known as "port forwarding" is the transmission and data intended for use only within a private, usually corporate network through a public network.