SCHOOL OF ELECTRICAL AND ELECTRONICS

DEPARTMENT OF ELECTRONICS AND COMMMUNICATION ENGINEERING

UNIT-I TELECOMMUNICATION SYSTEMS AND SERVICES -SECA3002

UNIT 1 BASICS OF TELECOMMUNICATION

Switching system functions – strowger switching system – cross bar exchange – SPC exchange-End users, nodes and connectivity's, telephone numbering and Routing, use of Tandem switches in Local area connectivity, Busy Hour and Grade of Service, Simple, Half duplex and full duplex, One-way and two-way circuits, Network topologies, variations in traffic flow, quality of service.

Telecommunications

The electronic transmission of information over distances, called telecommunications, has become nearly inseparable from computers: Computers and telecommunications create value together. Components of a Telecommunications Network Telecommunications are the means of electronic transmission of information over distances. Telecommunication is the exchange of signs, signals, messages, words, writings, images and sounds or information of any nature by wire, radio, optical or other electromagnetic systems.

A complete, single telecommunications circuit consists of two stations, each equipped with a transmitter and a receiver. The transmitter and receiver at any station may be combined into a single device called a transceiver. The medium of signal transmission can be via electrical wire or cable ("copper"), optical fiber, electromagnetic fields or light. The free space transmission and reception of data by means of electromagnetic fields is called wirelesscommunications.

Types of telecommunications networks

The simplest form of telecommunications takes place between two stations, but it is common for multiple transmitting and receiving stations to exchange data among them. Such an arrangement is called a telecommunications network. The internet is the largest example of a telecommunications network. On a smaller scale, examples include:

- Corporate and academic wide-area networks(WANs)
- Telephonenetworks
- Cellularnetworks
- Police and fire communicationssystems
- Taxi dispatchnetworks
- Groups of amateur (ham) radiooperators
- Broadcastnetworks

Data is transmitted in a telecommunications circuit by means of an electrical signal called the carrier or the carrier wave. In order for a carrier to convey information, some form of modulation is required. The mode of modulation can be broadly categorized as either analog ordigital.

Telecommunication Network

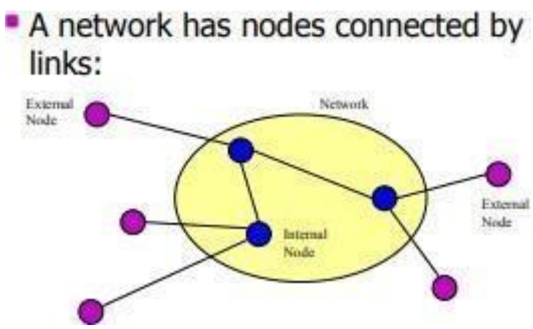


Figure 1 Telecommunication Network

External nodes are users and sometimes access points to other networks. ƒ Internal nodes are part of the network infrastructure and perform various tasks. ƒ Links provide interconnections between nodes. The goal is to have a path from any node to any other node without the need for an excessive number of links.

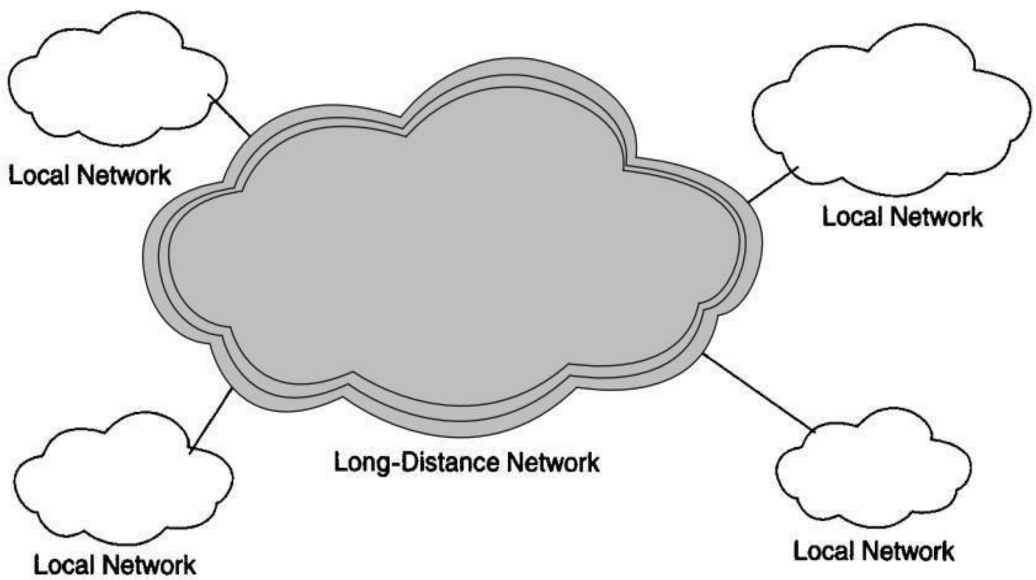
PUBLIC SWITCHED TELECOMMUNICATIONS NETWORK (PSTN)

The telephone is connected to the public switched telecommunications network (PSTN) for local, national, and international voice communications. These same telephone connections may also carry data and image information (e.g., television). In the United States the connection to the PSTN may be via a local exchange carrier (LEC) or by a competitive local exchange carrier (CLEC). The personal computer (PC) is beginning to take on a role similar to that of the telephone—namely, beingubiquitous.

In many situations, the PC uses telephone connectivity to obtain Internet and e-mail services. Cable television (CATV) offers another form of connectivity providing both telephone and Internet service.

- The PSTN has ever-increasing data communications traffic where the network is used as a channel for data. PSTN circuits may be rented or used in a dial-up mode for data connections.
- The Internet has given added stimulus to data circuit usage of the PSTN. The PSTN sees facsimile as just another data circuit, usually in the dial-upmode.
- Conference television traffic adds still another flavor to PSTN traffic and is also a main growthsegment.

- The trend for data is aloft where today data connectivity greatly exceeds telephone usage on the network. There is a growing trend for users to bypass the PSTN partially or completely.
- The use of satellite links in certain situations is one method for PSTNbypass.
- Other provider could be a power company with excess capacity on its microwave or fiber-opticsystem.
- There are other examples such as a railroad with extensive rights-of-way which may be used for a fiber-opticnetwork.
- Another possibility is to build a private network using any one or a combination of fiber optics, copper wire line, line-of-sight microwave, and satellite communications. Some private networks take on the appearance of amini-PSTN.



The PSTN consists of local networks interconnected by a long-distance network.

Figure 2 PSTN

It consists of local networks interconnected by one or more long-distance networks. The concept is illustrated in Figure. This is the PSTN, which is open to public correspondence. It is usually regulated by a government authority or may be a government monopoly, although there is a notable trend toward privatization.

End-Users

End-users, provide the inputs to the network and are recipients of network outputs. The end- user employs what is called an I/O, standing for input/output. An I/O may be a PC, computer, telephone instrument, cellular/PCS telephone or combined device, facsimile, or conference TV equipment. It may also be some type of machine that provides a stimulus to a coder or receives stimulus from a decoder in say some sort of SCADA(supervisory control and data acquisition) system.

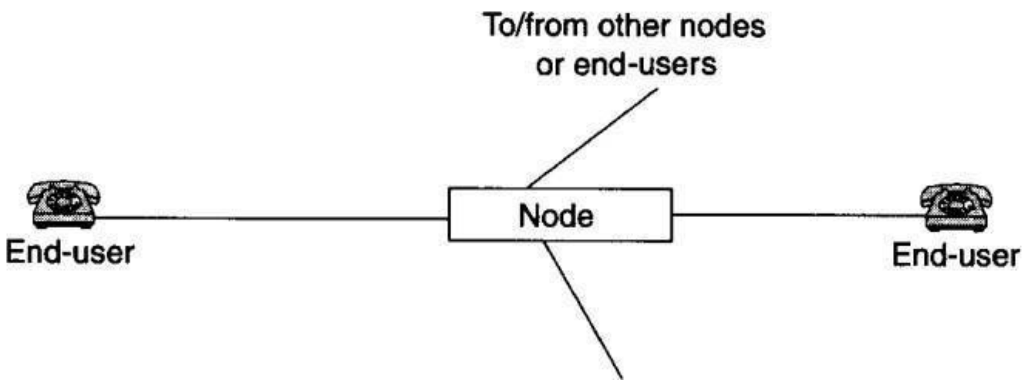


Figure 3 End User

Nodes

End-users usually connect to nodes. It is a node a point or junction in a transmission system where lines and trunks meet. A node usually carries out a switching function. In the case of the local area network (LAN).

Connectivity

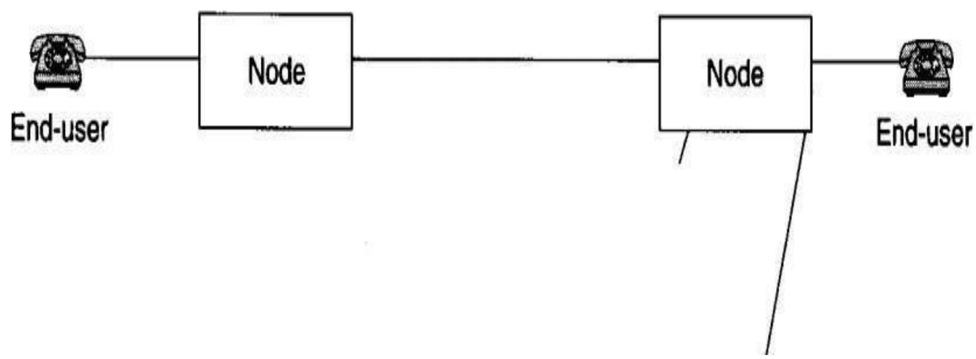


Figure 4 Connectivity

A network interface unit is used, through which one or more end-users may be connected. Connectivity links an end-user to a node and from there possibly through other nodes to some final end-user destination with which the initiating end-user wants to communicate. Figure 4 shows the IEEE defines a connection as “an association of channels, switching systems, and other functional units set up to provide means for a transfer of information between two or more points in a telecommunications network.” There would seem to be two interpretations of this definition.

First, the equipment, both switching and transmission facilities, are available to set up a path from, say, Socket A to Socket B. Assume A and B to be user end-points. The second interpretation would be that not only are the circuits available, but also they are connected and ready to pass information or are in the information passing mode.

End User as a Telephone User

The end-users are assumed to be telephone users, and the path that is set up is a speech path (it could, of course, be a data or video path).

There are three sequential stages to a telephone call.

1. Call setup

2. Informationexchange

3. Calltakedown

Call setup is the stage where a circuit is established and activated. The setup is facilitated by signaling,

It is initiated by the calling subscriber (user) going off-hook. This is a term that derives from the telephony of the early 1900s. It means "the action of taking the telephone instrument out of its cradle." Two little knobs in the cradle pop up, pushed by a spring action causing an electrical closure. If a light is turn on, an electrical closure allowing electrical current to pass. The same thing happens with our telephone set; it now passes current. The current source is a "battery" that resides at the local serving switch. It is connected by the subscriber loop. This is just a pair of copper wires connecting the battery and switch out to the subscriber premises and then to the subscriber instrument. The action of current flow alerts the serving exchange that subscriber requests service. When the current starts to flow, the exchange returns a dial tone, which is audible in the headset (of the subscriber instrument). The calling subscriber (user) now knows that she/he may start dialing digits or pushing buttons on the subscriber instrument. Each button is associated with a digit. There are 10 digits, 0 through 9.

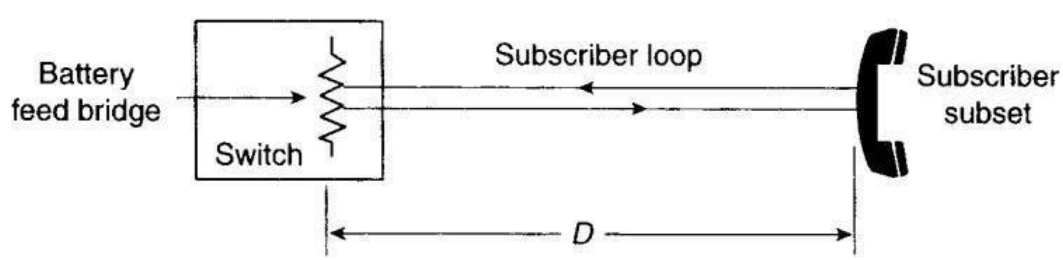


Figure 5 A subscriber set is connected to a telephone exchange by a subscriber loop
D-Distance (loop length)

Figure 5 shows a telephone end instrument connected through a subscriber loop to a local serving exchange. It also shows that all-important battery (battery feed bridge), which provides a source of current for the subscriberloop.

If the called subscriber and the calling subscriber are in the same local area, only seven digits need be dialed. These seven digits represent the telephone number of the called subscriber (user). This type of signaling, the dialing of the digits, is called **address signaling.** The digits actuate control circuits in the local switch, allowing a connectivity to be set up. If the calling and called subscribers reside in the serving area of that local switch, no further action need be taken. A connection is made to the called subscriber line, and the switch sends a special ringing signal down that loop to the called subscriber, and her/his telephone rings, telling her/him that someone wishes to talk to her/him on the telephone. This audible ringing is called alerting, another form of signaling. Once the called subscriber goes off-hook (i.e., takes the telephone out of its cradle), there is activated connectivity, and the call enters the information-passing phase or phase 2 of the telephonecall.

When the call is completed, the telephones at each end are returned to their cradle, breaking the circuit of each subscriber loop. This, of course, is analogous to turning off a light; the current stops flowing. Phase 3 of the telephone call begins. It terminates the call, and the connecting circuit in the switch is taken down and is then freed-up for another user. Both subscriber loops are now **idle**. If a third user tries to call either subscriber during stages 2 and 3, she/he is returned a busy-back by the exchange (serving switch). This is the familiar "**busy signal**," a tone with a particular cadence. The return of the busy-back is a form of signaling called **call- progress signaling.**

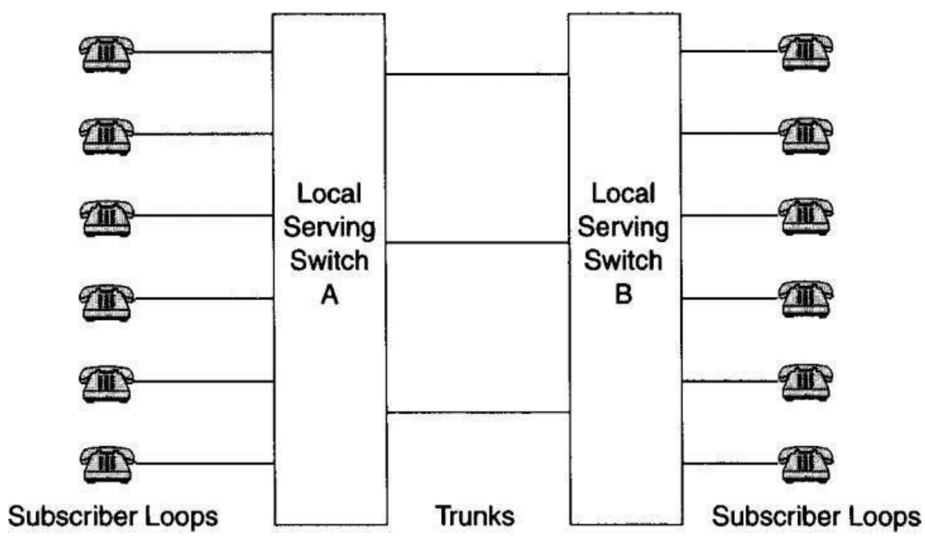


Figure 6. Subscriber loops connect telephone subscribers to their local serving exchange; trunks interconnect exchanges (switches).

Now that a subscriber wishes to call another telephone subscriber outside the local serving area of her/his switch. The call setup will be similar as before, except that at the calling subscriber serving switch the call will be connected to an outgoing trunk.

As shown in Figure 6, trunks are transmission pathways that interconnect switches. **Subscriber loops connect end-users (subscriber) to a local serving switch; trunks interconnect exchanges or switches**.

Trunk

The IEEE defines a trunk as “a transmission path between exchanges or central offices.” The word transmission in the IEEE definition refers to one (or several) transmission media. The medium might be wire-pair cable, fiber-optic cable, microwave radio, and, stretching the imagination, satellite communications. In the conventional telephone plant, coaxial cable has fallen out of favour as a transmission medium for this application.

Telephone Numbering and Routing

Every subscriber in the world is identified by a number, which is geographically tied to a physical location.4 This is the telephone number. The telephone number, as used it above, is seven digits long.

Example 234–5678

The last four digits identify the subscriber line;

the first three digits (i.e., 234) identify the serving switch (or exchange).

Numbering capacity

The subscriber number, consisting of the last four digits, has a theoretical numbering capacity of 10,000. The first telephone number issued could be 0000; the second number, if it were assigned in sequence, would be 0001, the third would be 0002, and so on. At the point where the numbers ran out, the last number issued would be 9999. The first three digits of the example above contain the exchange code (or central office code). These three digits identify the exchange or switch. The theoretical maximum capacity is 1000. If again assign numbers in sequence, the first exchange would have 001, the next 002, then 003, and finally 999. However, particularly in the case of the exchange code, there are blocked numbers. Numbers starting with 0 may not be desirable because in North America 0 is used to dial theoperator.

Example1

The numbering system for North America

The numbering system for North America (United States, Canada, and Caribbean islands) is governed by the North American Numbering Plan (NANP). It states that central office codes (exchange codes) are in the form NXX, where N can be any number from 2 through 9 and X can be any number from 0 through 9.

Numbers starting with 0 or 1 are blocked numbers in the case of the first digit N. This cuts the total exchange code capacity to 800 numbers. Inside these 800 numbers there are five blocked numbers such as 555 for directory assistance and 958/959 for local plant test. When long-distance service becomes involved, must turn to using still an additional three digits. Colloquially call these area codes. In the official North American terminology used in the NANP is "NPA" for **numbering plan area,** and call these area codes NPA codes. We try to assure that both exchange codes and NPA codes do not cross political/administrativeboundaries.

Example 2

For example, the exchange code 443 (in the 508 area code, middle Massachusetts) is exclusively for the use of the town of Sudbury, Massachusetts. Bordering towns, such as Framingham, shall not use that number. Of course, the 443 exchange code number is meant for Sudbury's singular central office (local serving switch). There is similar thinking for NPAs (area codes). In this case, these area codes may not cross state boundaries. For instance, 212 is for Manhattan and may not be used for northern New Jersey. Return now to our example telephone call. Here the calling party wishes to speak to a called party that is served by a different exchange (central office5).

Example 3

Assign the digits 234 for the calling party's serving exchange; for the called party's serving exchange we assign the digits 447. This connectivity is shown graphically in Figure 1.5. We described the functions required for the calling party to reach her/his exchange. This is the 234 exchange. It examines the dialed digits of the called subscriber, 447–8765.

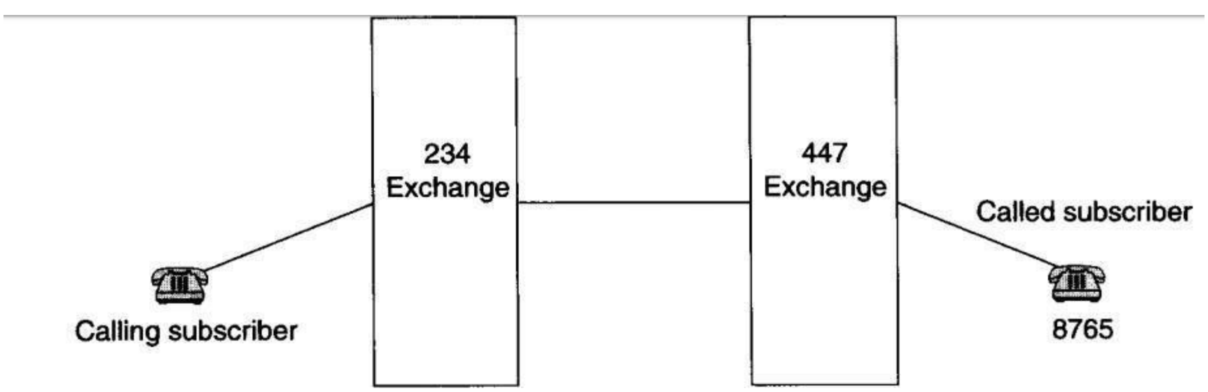


Figure 8 Connectivity subscriber to subscriber through two adjacent exchanges

To route the call, the exchange will only work upon the first three digits. It accesses its local look-up table for the routing to the 447 exchange and takes action upon that information. An appropriate vacant trunk is selected for this route, and the signaling for the call advances to the 447 exchange. Here this exchange identifies the dialed number as its own and connects it to the correct subscriber loop, namely the one matching the 8765 number. Ringing current is applied to the loop to alert the called subscriber. The called subscriber takes her/his telephone off hook, and conversation can begin.

Use of Tandem Switches in a Local Area Connectivity

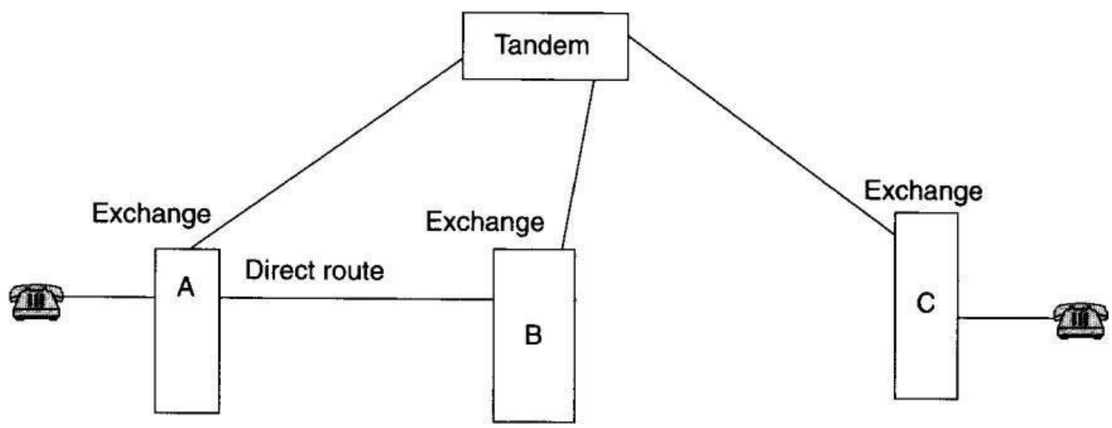


Figure 9 Direct route and tandem connectivities

Routing through a tandem switch is an important economic expedient for a telephone company or administration. A tandem switches a **traffic concentrator**. To employ a direct trunk circuit, there must be sufficient traffic to justify such a circuit. For connectivity with traffic intensity under 20 erlangs (The erlang is a unit of traffic intensity. One erlang represents one hour of line (circuit) occupancy.) For the busy hour (BH), the traffic should be routed through a tandem (exchange). For traffic intensities over that value, establish a directroute.

Busy Hour and Grade of Service

The PSTN is very inefficient. This inefficiency stems from the number of circuits and the revenue received per circuit. The PSTN would approach 100% efficiency if all the circuits were used all the time.

The facts are that the PSTN approaches total capacity utilization for only several hours during the working day. After 10 P.M. and before 7 A.M., capacity utilization may be 2% or 3%. The network is dimensioned (sized) to meet the period of maximum usage demand. This period is called the busy hour (BH). There are two periods where traffic demand on the PSTN is maximum: one in the morning and one in theafternoon.

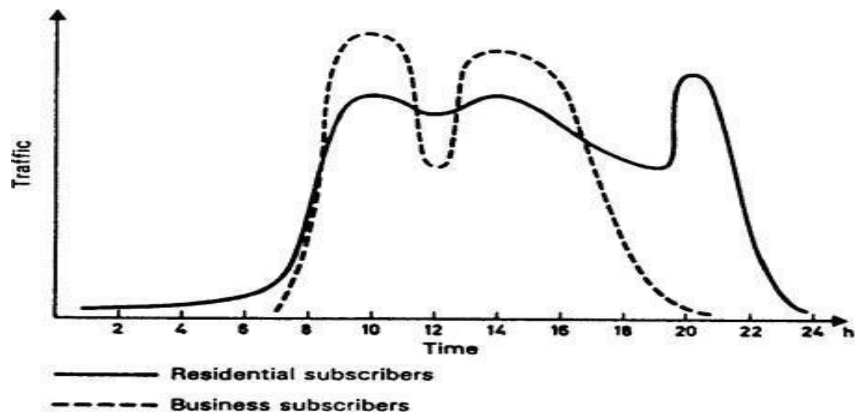


Figure 10 Busy hour

Note the two traffic peaks are caused by business subscribers. If the residential and business curves were combined, the peaks would be much sharper. Also note that the morning peak is somewhat more intense than the afternoon busy hour. In North America (i.e., north of the Rio Grande), the busy hour (BH) is between 9:30 A.M. and 10:30 A.M. Because it is more intense than the afternoon high-traffic period, it is called the busy hour. There are at least four distinct definitions of the busy hour. The IEEE gives several definitions. "That uninterrupted period of 60 minutes during the day when the traffic offered is maximum." Other definitions may be found in Ref. 4. BH traffic intensities are used to dimension the number of trunks required on connectivity as well as the size of (a) switch (es) involved. Now a PSTN company (administration) can improve its revenue versus expenditures by cutting back on the number of trunks required and making switches "smaller." Of course, network users will do a lot of complaining about poor service. Let's just suppose the PSTN does just that, cuts back on the number of circuits. Now, during the BH period, a user may dial a number and receive either a voice announcement or a rapid-cadence tone telling the user that **all trunks are busy (ATB)** and to try again later. From a technical standpoint, the user has encountered blockage. This would be due to one of two reasons, or may be due to both causes. These are**: insufficient switch capacity and not enough trunks to assign during the BH. Networks are sized /dimensioned for a traffic load expected during the busy hour. The sizing is based on probability, usually expressed as a decimal or percentage. That probability percentage or decimal is called the grade of service.** The IEEE (Ref. 2) defines grade of service as "the proportion of total calls, usually during the busy hour, that cannot be completed immediately or served within a prescribedtime.

Grade of service and blocking probability are synonymous. Blocking probability objectives are usually stated as B = 0.01% or 1%. This means that during the busy hour, 1 in 100 calls can be expected to meet blockage.

Simplex, Half-Duplex, and Full Duplex

Simplex is one way operation; there is no reply channel provided. Radio and television broadcasting are simplex. Certain types of data circuits might be based on simplex operation.

Half-duplex is a two-way service. It is defined as transmission over a circuit capable of transmitting in either direction, but only in one direction at a time.

Full duplex or just duplex defines simultaneous two-way independent transmission on a circuit in both directions. All PSTN-type circuits discussed in this text are considered using full- duplex operation unless otherwisespecified.

Basis for Comparison	Simplex	Half Duplex	Full Duplex
Direction of Communication	Unidirectional	Two-directional, one at a time	Two-directional, simultaneously
Send / Receive	Sender can only send data	Sender can send and receive data, but one a time	Sender can send and receive data simultaneously
Performance	Worst performing mode of transmission	Better than Simplex	Best performing mode of transmission
Example	Keyboard and monitor	Walkie-talkie	Telephone

One-Way and Two-Way Circuits

Trunks can be configured for either one-way or two-way operation. A third option is a hybrid where one-way circuits predominate and a number of two-way circuits are provided for overflow situations.

Figure 11 a shows two-way trunk operation. In this case, any trunk can be selected for operation in either direction. The incisive reader will observe that there is some fair probability that the same trunk can be selected from either side of the circuit. This is called double seizure. It is highly undesirable. One way to reduce this probability is to use normal trunk numbering (from top down) on one side of the circuit (at exchange A in the figure) and to reverse trunk numbering, from the bottom up at the opposite side of the circuit (exchangeB).

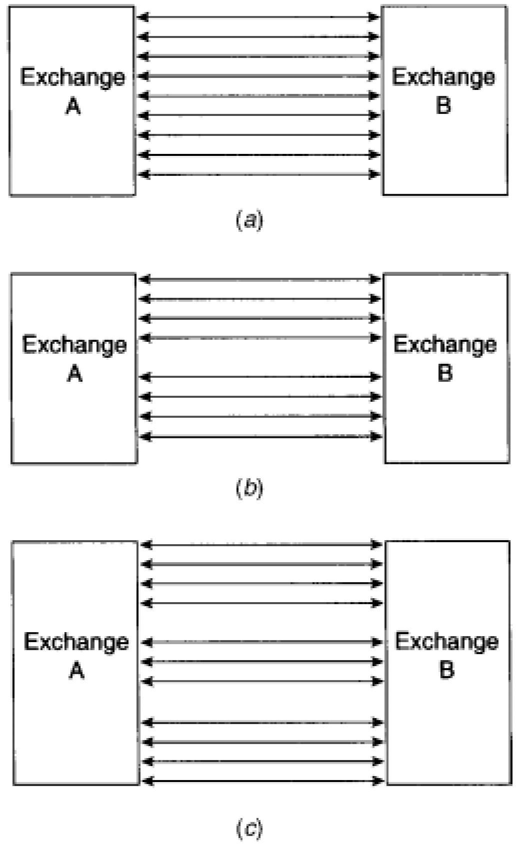


Figure 11 Two-way and one-way circuits: two-way operation (a), one-way operation (b), and a hybrid scheme, a combination of one-way and two-way operation (c).

Figure 11 b shows one-way trunk operation. The upper trunk group is assigned for the direction from A to B; the lower trunk group is assigned for the opposite direction, from exchange B to exchange A. Here there is no possibility of double seizure. Figure 11 c illustrates a typical hybrid arrangement. The upper trunk group carries traffic from exchange A to exchange B exclusively. The lowest trunk group carries traffic in the opposite direction. The small, middle trunk group contains two-way circuits. Switches are programmed to select from the one-way circuits first, until all these circuits become busy; then they may assign from the two-way circuit pool. Let us clear up some possible confusion here. Consider the one-way circuit from A to B, for example. In this case, calls originating at exchange A bound for exchange B in Figure 11b are assigned to the upper trunk group. Calls originating at exchange B destined for exchange A are assigned from the pool of the lower trunkgroup.

Network Topologies

The IEEE (Ref. 2) defines topology as “the interconnection pattern of nodes on a network.” A telecommunications network consists of a group of interconnected nodes or switching centers. There are a number of different ways, interconnect switches in a telecommunication network.

Mesh topology

If every switch in a network is connected to all other switches (or nodes) in the network, then it is called as “pattern” of a full-mesh network. The figure has 6 nodes. A full-mesh network is very survivable because of a plethora of possible alternativeroutes.

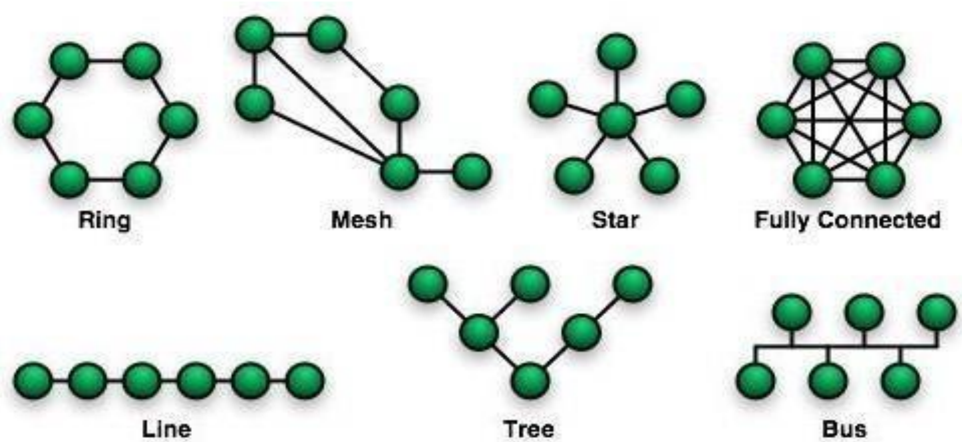


Figure 12 Different Network topologies

Star network

Figure 12 shows a star network. It is probably the least survivable. However, it is one of the most economic nodal patterns both to install and toadminister.

Tree Topology

The root node then communicates with a number of smaller nodes, and those in turn communicate with an even greater number of smaller nodes. A host that is a branch off from the main tree is called a leaf. If a leaf fails, its connection is isolated and the rest of the LAN can continue onwards.

Ring Topology

A ring topology (commonly known as a token ring topology) creates a network by arranging 2 or more hosts in a circle. Data is passed between hosts through a token. This token moves rapidly at all times throughout the ring in one direction. If a host desires to send data to another host, it will attach that data as well as a piece of data saying who the message is for to the token as it passes by. The other host will then see that the token has a message for it by scanning for destination addresses that match its own.

Line Topology

This rare topology works by connecting every host to the host located to the right of it. It is very expensive (due to its cabling requirements) and due to the fact that it is much more practical to connect the hosts on either end to form a ring topology, which is much cheaper and more efficient.

Bus Topology

A bus topology creates a network by connecting 2 or more hosts to a length of coaxial backbone cabling. In this topology, a terminator must be placed on the end of the backbone coaxial cabling.

Multiple star network

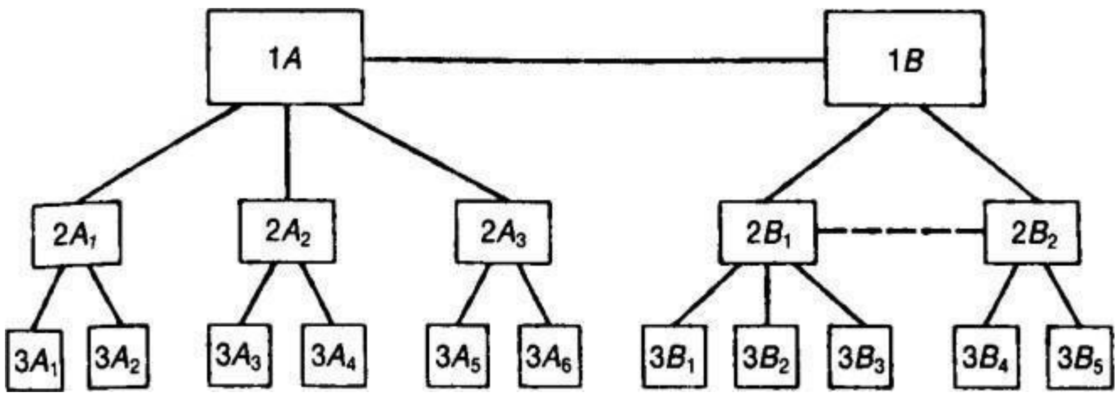


Figure 13 Multiple star network.

To modify such networks by adding direct routes, usually the 20-erlang rule is used in such situations. If a certain traffic relation has 20 erlangs or more of BH traffic, a direct route is usually justified. The term traffic relation simply means the traffic intensity (usually the BH traffic intensity).

Variations in Traffic Flow

In networks wrapper large geographic expanses and even in cases of certain local networks, there may be a variation of the time of day of the BH or in a certain direction of traffic flow. It should be pointy out that the busy hour is tied up with a country's culture. Countries have different working habits and standard business hours vary. In Mexico, for instance, the BH is more skewed toward noon because Mexicans eat lunch later than do people in the United States. In the United States, business traffic peaks during several hours beforehand and several hours after the noon lunch period on weekdays, and social calls peak in early evening. Traffic flow tends to be from residential living areas to an urban center in the morning, and the reverse occurs in the evening. In national networks covering several time zones where the difference in local time may be appreciable, long-distance traffic tends to be concentrated in a few hours common to BH peaks at both ends. In such cases it is conceivable to direct traffic so that peaks of traffic in one area (time zone) fall into valleys of traffic of another area. This is called taking advantage of the non-coincident busy hour. The network design can be made more optimal if configured to take advantage of these phenomena, particularly in the design of direct routes and overflow routes.

QUALITY OF SERVICE

Quality of service (QoS) appears at the outset to be an intangible concept. However, it is very tangible for a telephone subscriber unhappy with his or her service. The concept of service quality must be covered early in an all-encompassing text on telecommunications. System designers should never once lose sight of the concept, no matter what segment of the system they may be responsible for. Quality of service means how happy the telephone company (or other common carrier) is keeping the customer. For instance, might find that about half the time a customer dials, the call goes awry or the caller cannot get a dial tone or cannot hear what is being said by the party at the other end. All these have an impact on quality of service. The transmission engineer calls QoS customer satisfaction, which is commonly measured by how well the customer can hear the calling party. The unit for measuring how well, can hear a distant party on the telephone is loudness rating, measured in decibels (dB). From the network Other elements to be listed under QoSare:

- Can connectivity beachieved.
- Delay before receiving dial tone (dial tonedelay).
- Post dial(ing) delay (time from the completion of dialing the last digit of a number tothe first ring-back of the called telephone). This is the primary measure of signalingquality.
- Availability of service tones [e.g., busy tone, telephone out of order, time out, andall trunks busy(ATB)].
- Correctness ofbilling.
- Reasonable cost of service to thecustomer.
- Responsiveness to servicingrequests.
- Responsiveness and courtesy ofoperators.
- Time to installation of a new telephone, and, by some, the additional services offered by the telephone company. One way or another, each item, depending on the service quality goal, will have an impact on the design of a telecommunicationsystem.

TEXT / REFERENCE BOOKS

1. Roger L. Freeman, “Fundamentals of Telecommunications”, 4th Edition, John Wiley & Sons, 2010.
2. Jyrki T. J. Penttinen, “The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and SatelliteSystems”, 1st Edition, Wiley, 2015.
3. ThiagarajanViswanathan, ManavBhatnagar, “Telecommunication Switching Systems and Networks”, 2ndEdition, PHIlearning, 2015.

QUESTIONBANK:

PART-A

1: What is telecommunication network?
2: What is Node?
3: What is end user?
4: What is connectivity?
5: What is telephone numbering?
6: What is routing?
7: Define busy hour.
8: What is peak busy hour?
9: What is grade of service?
10: Define Simplex.
11: Define half duplex.
12: Define full duplex.
13: What is one way circuit?
14: What is two way circuits?
15: What is point to point network topology?
16: What is Bus/tree network topology?
17: What is Star topology?
18: What is Ring network?

PART-B

1.Explain the features of Telecommunication Network.
2.Elaborate about Network Topologies with diagrams.
3.Analyze the concepts involved in 0ne-way,Two way and Hybrid Trunk operations.
4.Discuss about sequential stages of telephone call.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS

DEPARTMENT OF ELECTRONICS AND COMMMUNICATION ENGINEERING

UNIT-II TELECOMMUNICATION SYSTEMS AND SERVICES -SECA3002

UNIT 2 SIGNALLING IN TELECOMMUNICATION SYSTEMS

Introduction, purpose of signaling, Defining the functional areas-supervisory signaling, address signaling and Call Progress-audio and visual. Signaling techniques - conveying signaling information, evolution of signaling subscriber call progress tones and push button codes, compelled signaling, concepts of Link-by-link and end-to-end signaling, effects of numbering on signaling, associated and disassociated channel signaling, signaling in the subscriber loop-background and purpose, metallic trunk signaling - basic loop signaling, reverse-battery signaling, stimulus signaling, functional signaling, Object-oriented signaling.

THE PURPOSE OF SIGNALING

The IEEE (Ref. 1) defines *signaling* as the "exchange of information specifically concerned with the establishment and control of connections and the transfer of user-to- user and management information in a telecommunication network. "Conventional signaling has evolved with the telephone network. Many of the techniques we deal with in this chapter are applicable to a telecommunication network which is principally involved with telephone calls. With telephony, signaling is broken down in three functional areas:

1. Supervisory;
2. Address;and
3. Call progressaudible-visual.

Another signaling breakdown is:

- Subscribersignaling;
- Interswitch (interregister)signaling.

DEFINING THE FUNCTIONAL AREAS

Supervisory Signaling

Supervisory signaling provides information on line or circuit condition. It informs a switch whether a circuit (internal to the switch) or a trunk (external to the switch) is busy or idle; when a called party is off-hook or on- hook, and when a calling party is on-hook or off-hook.Supervisory information (status) must be maintained end-to-end on a telephone call, whether voice, data, or facsimile is being transported. Itis necessary to know when a calling subscriber lifts her/ his telephone off-hook, thereby requesting service. It is equally important that we know when the called subscriber answers (i.e., lifts the tele-phone off-hook) because that is when we may start metering the call to establish charges. It is also important to know when the calling and called subscribers return their tele-phones to the on-hook condition. That is when charges stop, and the intervening trunks comprising the talk path as well as the switching points are then rendered idle for use by anotherpair of subscribers. During the period of occupancy of a speech path end- to-end, we must know that this particular path is busy (i.e., it is occupied) so no other call attempt can seizeit.

Address Signaling

Address signaling directs and routes a telephone call to the called subscriber. It originates as dialed digits or activated push-buttons from a calling subscriber. The local switch accepts these digits and, by using the information contained in the digits, directs the call to the called subscriber. If more than one switch is involved in the call setup, signaling is required between switches (both address and supervisory). Address signaling between switches is called *interregistersignaling*.

Call Progress—Audible-Visual

This type of signaling we categorize in the *forward direction* and in the *backward direction*. In the forward direction there is *alerting*. This provides some sort of audible-visual means of informing the called subscriber that there is a telephone call waiting. This is often done by ringing a telephone's bell. A buzzer, chime, or light may also be used for alerting.

The remainder of the techniques we will discuss are used in the backward direction. Among these are audible tones or voice announcements that will inform the calling subscriber of the following:

1. *Ringback.*Thistellsthecallingsubscriberthatthedistanttelephoneisringing.
2. *Busyback.*Thistellsthecallingsubscriberthatthecalledlineisbusy.
3. *ATB—All Trunks Busy.* There is congestion on the routing. Sometimesa recorded voice announcement is usedhere.
4. *Loud warble on telephone instrument—Timeout.* This occurs when a telephone instrument has been left off-hookunintentionally.

SIGNALING TECHNIQUES

Conveying Signaling Information

Signaling information can be conveyed by a number of means from a subscriber to the serving switch and between (among) switches. Signaling information can be transmitted by means such as:

- Durationofpulses(pulsedurationbearsaspecificmeaning);
- Combination ofpulses;
- Frequency ofsignal;
- Combination offrequencies;
- Presence or absence of asignal;
- Binary code;and
- Direction and or level of transmitted current(for dc systems).

Evolution of Signaling

Signaling and switching are inextricably tied together. Switching automated the network. But without signaling, switching systems could not function. Thus it would be better said that switching with signaling automated the network.

Conventional subscriber line signaling has not changed much over the years, with the exception of the push-button tones, which replaced the dial for address signaling. ISDN, being a full digital service to the subscriber, uses a unique digital signaling system called DSS-1 (Digital Subscriber Signaling No.1).

Nearly every international circuit required special signaling interfaces. The same was true, to a lesser extent, on the national level.

In this section we will cover several of the more common signaling techniques used on the analog network which operated with frequency division multiplex equipment (Section 4.5.2). Although these signaling systems are obsolete in light of the digital network, the concepts covered here will help in understanding how signalingworks.

Supervisory Line Signaling

*Introduction.*Line signaling on wire trunks was based essentially on the presence or absence of dc current. Such dc signals are incompatible with FDM equipment where the voice channel does not extend to 0 Hz. Remember; the analog voice channel occupies the band from 300 Hz to 3400 Hz. So the presence or absence of a dc current was converted to an ac tone for one of the states and no-tone for the other state. There were two ways to approach the problem. One was called *in-band signaling* and the other was called *out-of-bandsignaling*.

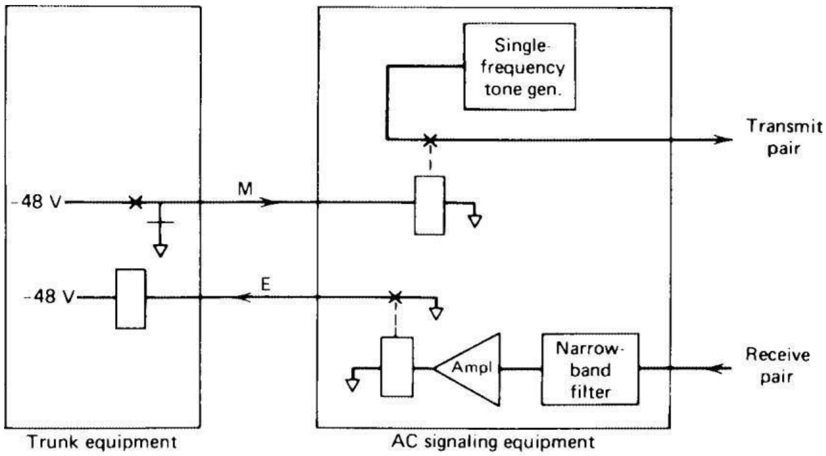
In-Band Signaling.In-band signaling refers to signaling systems using an audio tone, or tones, inside the conventional voice channel to convey signaling information. There are two such systems we will discuss here: (1) one-frequency (SF or single frequency), and
(2) two-frequency (2VF). These signaling systems used one or two tones in the 2000 Hz to 3000 Hz portion of the band, where less speech energy is concentrated.

Single-frequency (SF) signaling is used exclusively for supervision, often with its adjunct called *E&M signaling*, which we cover in Section 2.3.2.1.4. It is used with FDM equipment, and most commonly the tone frequency was 2600 Hz. Of course this would be in four-wire operation. Thus we would have a 2600-Hz tone in either both directions. The direction of the tone is important, especially when working with its E&M signaling adjunct. A diagram showing the application of SF signaling on a four-wire trunk is shown in Figure2.1.

Two-frequency (2VF) signaling can be used for both supervision (line signaling) and address signaling. Its application is with FDM equipment. Of course when discussing such types of line signaling (supervision), we know that the term *idle* refers to the on- hook condition, while *busy* refers to the off-hook condition. Thus, for such types of line signaling that are governed by audio tones of which SF and 2VF are typical, we have the conditions of “tone on when idle” and “tone on when busy.”

1 *Line signaling* is the supervisory signaling used among switches.

2 Called *out-band* by CCITT and in nations outside of North America.



Signal	Tone	Operation	Lead	Condition
On hook	On	Transmitting	M	Ground
		Receiving	E	Open
Off hook	Off	Transmitting	M	Battery
		Receiving	E	Ground

Figure 2.1 Functional block diagram of an SF signaling circuit. *Note:* Wire pairs "receive" and "transmit" derive from the FDM multiplex equipment. Note also the E-lead and M-lead.

It holds equally well for in-band and out-of-band signaling methods. However, for in-band signaling, supervision is by necessity tone-on idle; otherwise subscribers would have an annoying 2600-Hz tone on throughout thecall.

A major problem with in-band signaling is the possibility of "talk-down," which refers to the premature activation or deactivation of supervisory equipment by an inadvertent sequence of voice tones through the normal use of the channel. Such tones could simulate the SF tone, forcing a channel dropout (i.e., the supervisory equipment would return the channel to the idle state). Chances of simulating a 2VF tone set are much less likely. To avoid the possibility of talk-down on SF circuits, a time-delay circuit or slot filters to by-pass signaling tones may be used. Such filters do offer some degradation to speech unless they are switched out during conversation. They must be switched out if the circuit is going to be used for data transmission (Ref.2).

It becomes apparent why some administrations and telephone companies have turned to the use of 2VF supervision, or out-of-band signaling, for that matter. For example, a typical 2VF line signaling arrangement is the CCITT No. 5 code, where f_1 (one of the two VF frequencies) is 2400 Hz and f_2 is 2600 Hz. 2VF signaling is also used widely for addresssignaling.

Out-of-Band Signaling.With out-of-band signaling, supervisory information is transmitted out of band (i.e., above 3400 Hz). In all cases it is a single- frequency system. Some out-of-band systems use “tone on when idle,” indicating the on-hook condition, whereas others use “tone off.” The advantage of out-of-band signaling is that either system, tone on or tone off, may be used when idle. Talk-down cannot occur because all supervisory information is passed out of band, away from the speech-information portion of the channel. The preferred CCITT out-of-band frequency is 3825 Hz, whereas 3700 Hz is commonly used in the United States. It also must be kept in mind that out-of-bandsignaling.

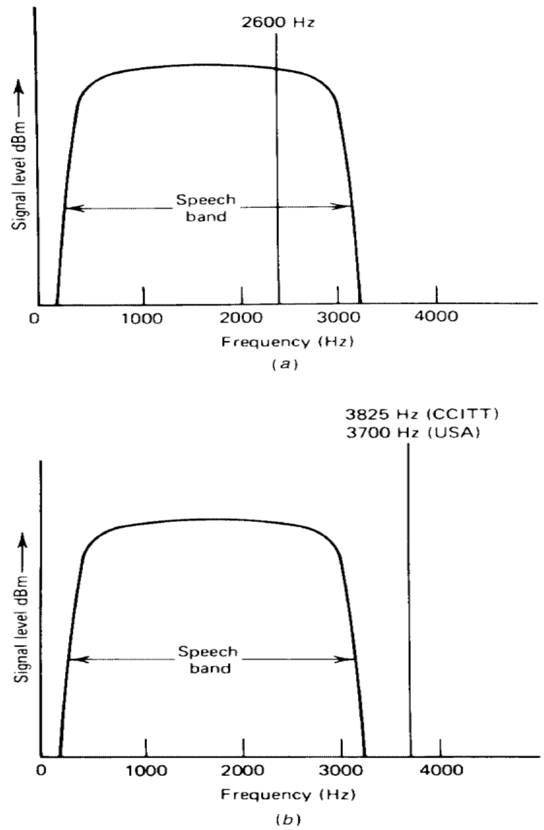


Figure 2.2 SF signaling (*a*) in-band and (*b*) out-of-band.

It is used exclusively on carrier systems, not on wire trunks. On the wire side, inside an exchange, its application is E&M signaling. In other words, out-of- band signaling is one method of extending E&M signaling over a carrier system. In the short run, out-of-band signaling is attractive in terms of both economy and design. One drawback is that when channel patching is required, signaling leads have to be patched as well. In the long run, the signaling equipment required may indeed make out-of-band signaling even more costly because of the extra supervisory signaling equipment and sig- naling lead extensions required at each end, and at each time that the carrier (FDM) equipment demodulates to voice.

The major advantage of out-of-band signaling is that continuous supervision is provided, whether tone on or tone off, during the entire telephone conversation. In-band SF signaling and out-of-band signaling are illustrated in Figure 2.2. An example of out-of-band signaling is the regional signaling system R-2, prevalent in Europe and nations under European heteronomy (see Table2.1.)

Table 2.1 R-2 Line Signaling (3825 Hz)

	Direction	
Circuit State	Forward (Go)	Backward (Return)
Idle	Toneon	Toneon
Seized	Toneoff	Tone on
Answered	Toneoff	Tone off
Clearback	Toneoff	Tone on
Release	Toneon	Tone onor off
Blocked	Toneon	Toneoff

E&M Signaling. The most common form of trunk supervision in the analog network was E&M signaling. It derived from the SF or 2VF equipment, as shown in Figure 2.1. It only becomes true E&M signaling where the trunk interfaces with the switch (see Figure 2.3). E-lead and M- lead signaling systems are semantically derived from the historical designation of signaling leads on circuit drawings covering these systems. Historically, the E&M interface provides two leads between the switch and what we call the *trunk signaling equipment* (signaling interface). One lead is called the E-lead, which carries signals *to* the switching equipment. Such signal directions are shown in Figure 2.3, where we see that signals from switch A and switch B leave A on the M-lead and are delivered to B on the E-lead. Likewise, from B to A, supervisory information leaves B on the M-lead and is delivered to A on the E-lead.

For the conventional E&M signaling (referring to the electromechanical exchanges), the following supervisory conditions are valid:

DIRECTION		**CONDITION AT*A***		**CONDITION AT*B***	
Signal A to B	*Signal B to A*	*M-Lead*	*E-Lead*	*M-Lead*	*E-Lead*
Onhook	On hook	Ground	Open	Ground	Open
Offhook	On hook	Battery	Open	Ground	Ground
Onhook	Off hook	Ground	Ground	Battery	Open
Off hook	Off hook	Battery	Ground	Battery	Ground

Source: Ref.8.

Address Signaling. Address signaling originates as dialed digits (or activated push buttons) from a calling subscriber, whose local switch accepts these digits and, using that information, directs the telephone call to the desired distant subscriber. If more than one switch is involved in the call setup, signaling is required between switches (both address and supervisory). Address signaling between switches in conventional

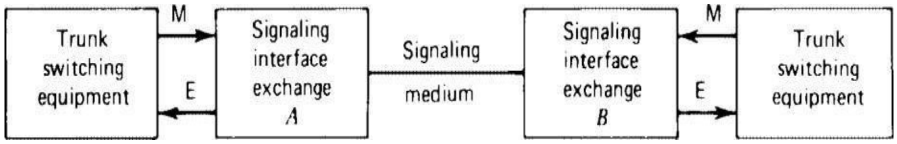


Figure 2.3 E&M signaling.

systems is called *interregistersignaling*.The paragraphs that follow discuss various more popular standard ac signaling techniques such as 2VF and MF tone. Although interregister signaling is stressed where appropriate, some supervisory techniques are also reviewed.

Multifrequency Signaling.

Multifrequency (MF) signaling has been in wide use around the world for interregister signaling. It is an in-band method using five or six tone frequencies, two tones at a time. It works well over metallic pair, FDM, and TDM systems. MF systems are robust and difficult to cheat. Three typical MF systems are reviewed in the following:

MULTIFREQUENCY SIGNALING IN NORTH AMERICA—THE R-1 SYSTEM.

The MF signaling system principally employed in the United States and Canada is recognized by the CCITT as the R-1 code (where R stands for “regional”). It is a two-out-of-five frequency pulse system. Additional signals for control functions are provided by frequency combination using a sixth basic frequency. Table 2.2 shows the ten basic digits (0–9) and other command functions with their corresponding two-frequency combinations, as well as a brief explanation of “other applications.” We will call this system a “spill forward” system. It is called this because few backward acknowledgment signals are required. This is in contraposition to the R-2 system, where every transmitted digit must be acknowledged.

Subscriber Call Progress Tones and Push-Button Codes (North America)

Table 2.7 shows the audible call progress tones commonly used in North America as presented to a subscriber. Subscriber subsets are either dial or push button, and they will probably be all push button in the next ten years. A push button actuates two audio tones simultaneously, similar to the multifrequency systems described previously with interregister signaling. However, the tone library used by the subscriber is different than the tone library used with interregistersignaling.

Table 2.8 compares digital dialed, dial pulses (breaks), and multifrequency (MF) push-button tones

Table 2.7 Audible Call Progress Tones Commonly Used in North America

Tone	Frequencies (Hz)	Cadence
Dial	350+440	Continuous
Busy(station)	480+620	0.5 s on, 0.5 s
Busy (network congestion)	480+620	0.2 s on, 0.3 s
Ringreturn	440+480	2 s on, 4 s off
Off-hookalert	Multifrequency	1 s on, 1 s off
Recordingwarning	1400	0.5 s on, 15 soff
Callwaiting	440	0.3 s on, 9.7off

Source: Ref. 7.

Clear Back. The incoming end restores the tone in the backward direction. When another link of the connection using tone-on-when-idle continuous signaling pre- cedes the outgoing exchange, the "tone-off" condition must be established on this link as soon as it is recognized in thisexchange.

Clear Forward. The outgoing end restores the tone in the forward direction.

*Blocked.*At the outgoing exchange the circuit stays blocked as long as the tone remains off in the backward direction.

COMPELLED SIGNALING

In many of the signaling systems discussed thus far, signal element duration is an important parameter. For instance, in a call setup an initiating exchange sends a 100-ms seizure signal.

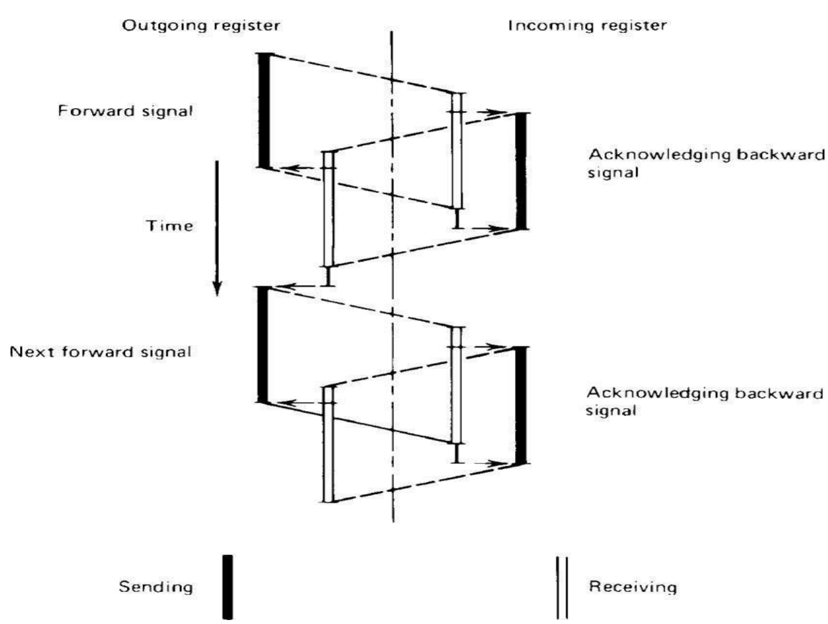


Figure 2.4 Fully compelled signaling procedure.

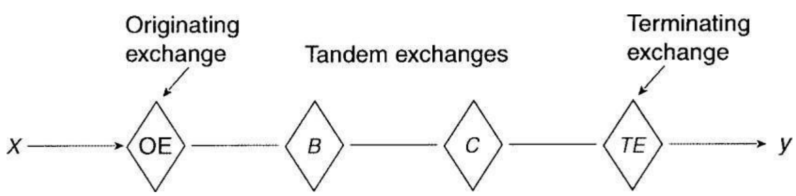
Once this signal is received at the distant end, the distant exchange sends a “proceed to send” signal back to the originating exchange; in the case of the R-1 system, this signal is 140 ms or more in duration. Then, on receipt of “proceed to send” the initiating exchange spills all digits forward. In the case of R-1, each digit is an MF pulse of 68-ms duration with 68 ms between each pulse. After the last address digit an ST (end-of- pulsing) signal is sent. In the case of R-1 the incoming (far-end) switch register knows the number of digits to expect. Consequently there is an explicit acknowledgment that the call setup has proceeded satisfactorily. Thus R-1 is a good example of noncompelled signaling.
A fully compelled signaling system is one in which each signal continues to be sent until an acknowledgment is received. Thus signal duration is not significant and bears no meaning. The R-2 and SOCOTEL are examples of fully compelled signaling systems. Figure 2.4 illustrates a fully compelled signaling sequence. Note the small overlap of signals, causing the acknowledging (reverse) signal to start after a fixed time on receipt of the forward signal. This is because of the minimum time required for recognition of the incoming signal. After the initial forward signal, further forward signals are delayed for a short recognition time (see Figure 2.4). Recognition time is normally less than 80 ms.

Fully compelled signaling adapts automatically to the velocity of propagation, to long circuits, to short circuits, to metallic pairs, or to carrier and is designed to withstand short interruptions in the transmission path. The principal drawback of compelled signaling is its inherent lower speed, thus requiring more time for setup. Setup time over space-satellite circuits with compelled signaling is appreciable and may force the system engineer to seek a compromise signaling system. There is also a partially compelled type of signaling, where signal duration is fixed in both forward and backward directions according to system specifications; and the forward signal is of indefinite duration and the backward signal is of fixed duration. The forward signal ceases once the backward signal has been received correctly. CCITT Signaling System No. 4 is an example of a partially compelled signalingsystem.

CONCEPTS OF LINK-BY-LINK VERSUS END-TO-END SIGNALING

An important factor to be considered in switching system design that directly affects both signaling and customer satisfaction is post dialing delay. This is the amount of time it takes after the calling subscriber completes dialing until ring-back is received. Ring- back is a backward signal to the calling subscriber indicating that the dialed number is ringing. Post dialing delay must be made as short as possible.
Another important consideration is register occupancy time for call setup as the set- up proceeds from originating exchange to terminating exchange. Call-setup equipment, that equipment used to establish a speech path through a switch and to select the proper outgoing trunk, is expensive. By reducing register occupancy per call, we may be able to reduce the number of registers (and markers) per switch, thus savingmoney.
Link-by-link and end-to-end signaling each affect register occupancy and post dialing delay, each differently. Of course, we are considering calls involving one or more tandem exchanges in a call setup, because this situation usually occurs on long- distance or toll calls.

Link-by-link signaling may be defined as a signaling system where *all* interregister address information must be transferred to the subsequent exchange in the call-setup routing. Once this information is received at this exchange, the preceding exchange control unit (register) releases. This same operation is carried on from the originating exchange through each tandem (transit) exchange to the terminating exchange of the call. The R-1 system is an example of link-by- link signaling. End-to-end signaling abbreviates the process such that tandem (transit) exchanges receive only the minimum information necessary to route the call. For instance, the last four digits of a seven-digit telephone number need be exchanged only between the originating exchange (e.g., the calling subscriber's local exchange or the first toll exchange in the call set-up) and the terminating exchange in the call setup. With this type of signaling, fewer digits are required to be sent (and acknowledged) for the overall call-setup sequence. Thus the signaling process may be carried out much more rapidly, decreasing post dialing delay. Intervening exchanges on the call route work much less, handling only the digits necessary to pass the call to the next exchange in the se- quence. The key to end-to-end signaling is the concept of "leading register." This is the register (control unit) in the originating exchange that controls the call routing until a speech path is setup to the terminating exchange before releasing to prepare for another call setup. For example, consider a call from subscriber *X* to subscriber*Y*:



The telephone number of subscriber *Y* is 345–6789. The sequence of events is as follows using end-to-end signaling:

- A register at exchange OE receives and stores the dialed number 345–6789 from subscriber*X*.
- Exchange OE analyzes the number and then seizes a trunk (junction) toexchange *B*. It then receives a "proceed-to-send" signal indicating that the register at *B* is ready to receive routing information (digits).
- Exchange OE then sends digits 34, which are the minimum necessary to effect correcttransit.
- Exchange *B* analyzes the digits 34 and then seizes a trunk to exchange *C*. Exchanges OE and *C* is now in direct contact and exchange B's register releases.
- Thus we see that a signaling path is opened between the leading register and the terminating exchange. To accomplish this, each exchange in the route must "know" its local routing arrangements and request from the leading register those digits it needs to route the call further along its propercourse.
- Again, the need for backward information becomes evident, and backward signaling capabilities must be nearly as rich as forward signaling capabilities when such a system isimplemented.

EFFECTS OF NUMBERING ON SIGNALING

Numbering, the assignment and use of telephone numbers, affects signaling as well as switching. It is the number or the translated number, as we found out in Section 1.3.2, that routes the call. There is “uniform” numbering and “nonuniform” numbering. How does each affect signaling? Uniform numbering can simplify a signaling system. Mostuniform systems in the nontoll or local-area case are based on seven digits, although some are based on six. The last four digits identify the subscriber. The first three dig- its (or the first two in the case of a six-digit system) identify the exchange. Thus the local exchange or transit exchanges know when all digits are received. There are two advantages to this sort ofscheme:

1. The switch can proceed with the call once all digits are received because it “knows” when the last digit (either the sixth or seventh)has beenreceived.
2. “Knowing” the number of digits to expect provides inherenterror control and makes “time out”simpler.

For nonuniform numbering, particularly on direct distance dialing in the international service, switches require considerably more intelligence built in. It is the initial digit or digits that will tell how many digits are to follow, at least in theory.

However, in local or national systems with nonuniform numbering, the originating register has no way of knowing whether it has received the last digit, with the exception of receiving the maximum total used in the national system. With nonuniform numbering, an incompletely dialed call can cause a useless call setup across a network up to the terminating exchange, and the call setup is released only after time out has run its course. It is evident that with nonuniform numbering systems, national (and international) networks are better suited to signaling systems operating end to end with good features of backward information, such as the R-2 system (Ref.5).

ASSOCIATED AND DISASSOCIATED CHANNEL SIGNALING

Here we introduce a new concept: disassociated channel signaling. Up to now we have only considered associated channel signaling. In other words, the signaling is carried right on its associated voice channel, whether in- band or out-of-band. Figure 2.5 illustrates two concepts: associated channel and separate channel signaling, but still associated. E-1 channel 16 is an example. It is indeed a separate channel, but associated with the 30-channel group of traffic channels. We will call this *quasi-associated channelsignaling*.

Disassociated channel signaling is when signaling travels on a separate and distinct route than the traffic channels for which it serves. CCITT Signaling System No. 7 uses either this type of signaling or quasi-associated channel signaling. Figure 2.6 illustrates quasi-associated channel signaling, whereas Figure 2.7 shows fully disassociated channel signaling.

SIGNALING IN THE SUBSCRIBER LOOP

When a subscriber takes a telephone off-hook (out of its cradle), there is a switch closure at the subset (see the hook-switch in Figure 5.3), current flows in the loop alerting the serving.

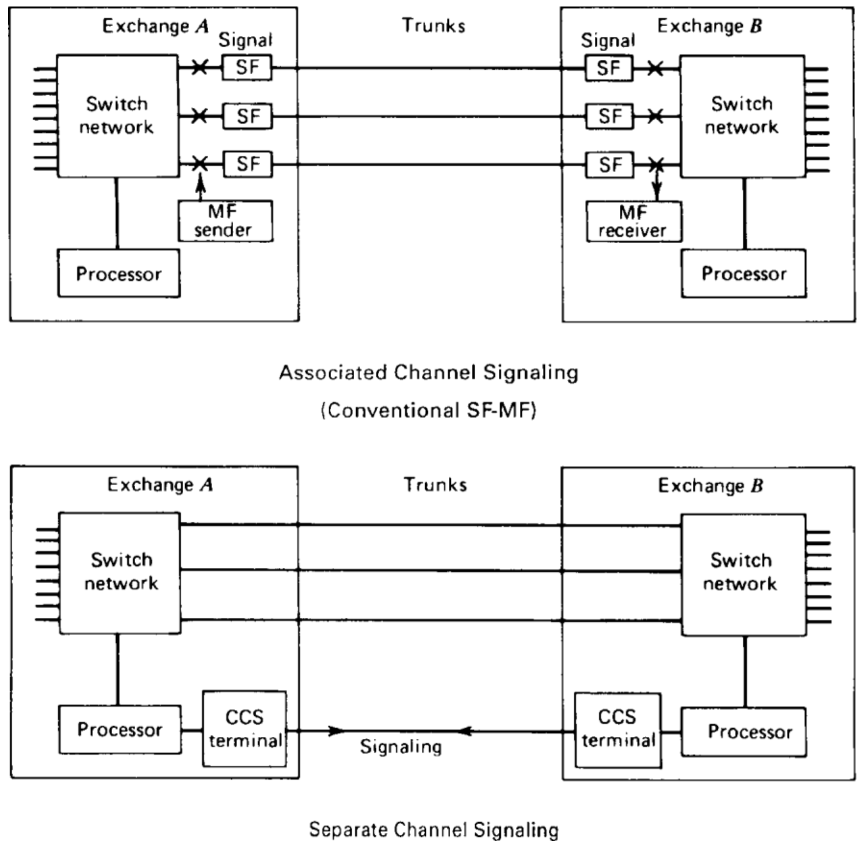


Figure 2.5 Conventional analog associated channel signaling (*upper*) versus separate channel signaling (which we call quasi-associated channel signaling) (*lower*).

Time out" is the resetting of call-setup equipment and return of dial tone to subscriber as a result of incomplete signaling procedure, subset left off hook, and so forth exchange that service is desired on that telephone. As a result, dial tone is returned to the subscriber. This is basic supervisory signaling on the subscriber loop.

A problem can arise from this form of signaling. It is called *glare*. Glare is the result of attempting to seize a particular subscriber loop from each direction. In this case it would be an outgoing call and an incoming call nearly simultaneously. There is a much greater probability of glare with a PABX than with an individual subscriber.

Ground-start signaling is the preferred signaling system when lines terminate in a switching system such as a PABX. It operates as follows: When a call is from the local serving switch to the PABX, the local switch immediately grounds the conductor tip to seize the line. With some several seconds delay, ringing voltage is applied to the line (where required). The PABX immediately detects the grounded tip conductor and will not allow an outgoing call from the PABX to use this circuit, thus avoidingglare.

In a similar fashion, if a call originates at the PABX and is outgoing to the local serving exchange, the PABX grounds the ring conductor to seize the line. The serving switch recognizes this condition and prevents other calls from attempting to terminate the circuit. The switch now grounds the tip conductor and returns dial tone after it connects a digit receiver. There can be a rare situation when double seizure occurs, causing glare. Usually one or the other end of the circuit is programmed to back down and allow the other call to proceed. A ground start interface is shown in Figure2.8.

Terminology in signaling often refers back to manual switchboards or, specifically, to the plug used with these boards and its corresponding jack as illustrated in Figure
2.9. Thus we have tip (T), ring (R), and sleeve (S). Often only the tip and ring are used, and the sleeve is grounded and has no real electrical function.

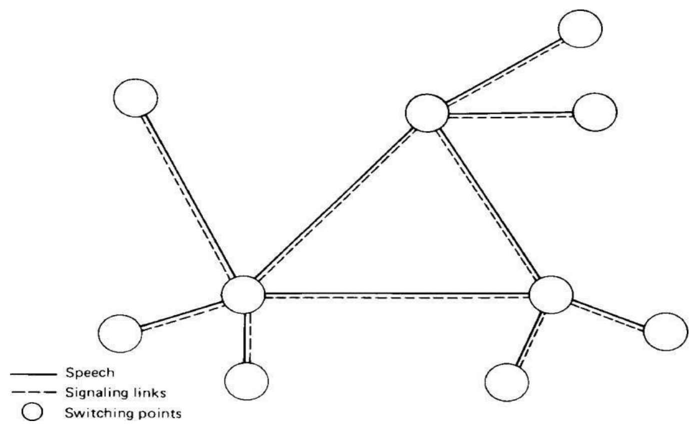


Figure 2.6 Quasi-associated channel signaling, typical of E-1 channel 16.

As shown, the signaling travels on a separate channel but associated with its group of traffic channels for which it serves. If it were conventional analog signaling, it would be just one solid line, where the signaling is embedded with its associatedtraffic.

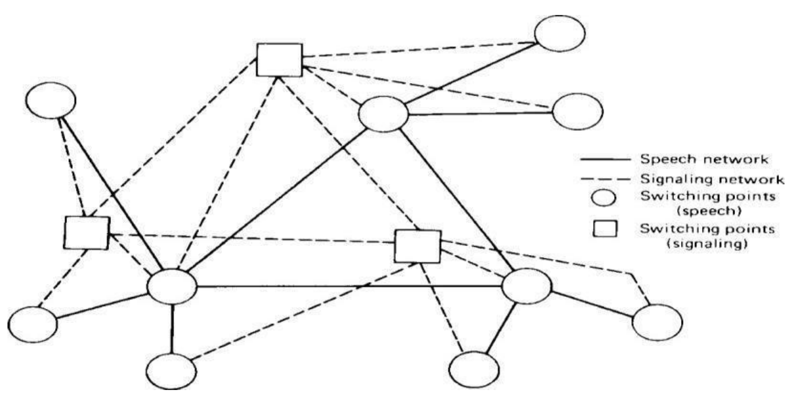


Figure 2.7 Fully disassociated channel signaling. This signaling may be used with CCITT Signaling System No. 7.

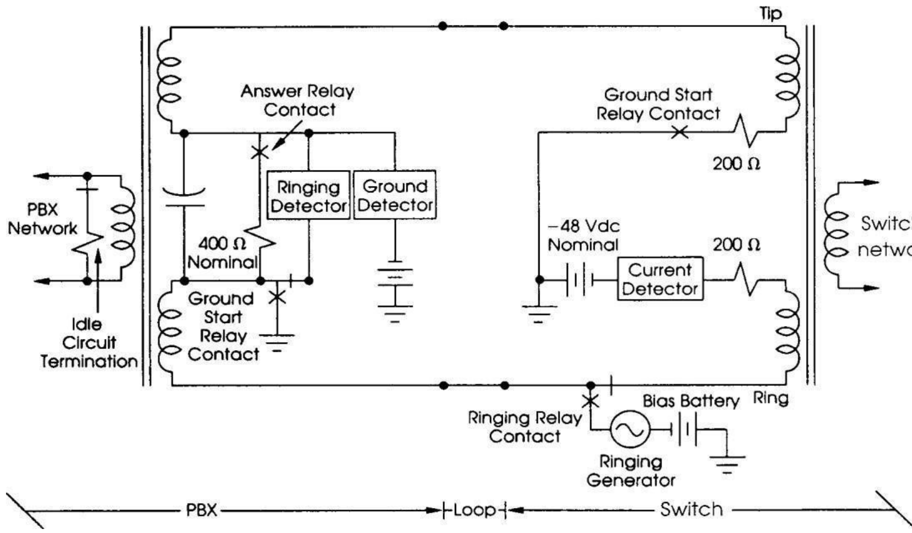


Figure 2.8 Ground-start interface block diagram. (From Figure 2-7 of Ref. 8, reprinted with permission.)

METALLIC TRUNK SIGNALING

Basic Loop Signaling

As mentioned earlier, many trunks serving the local area are metallic-pair trunks. They are actually loops much like the subscriber loop. Some still use dial pulses for address signaling along with some form of supervisory signaling.

Loop signaling is commonly used for supervision. As we would expect, it provides two signaling states: one when the circuit is opened and one when the circuit is closed. A third signaling state is obtained by reversing the direction or changing the magnitude of the current in the circuit. Combinations of (1) open close, (2) polarity reversal, and (3) high low current are used for distinguishing signals intended for one direction of signaling (e.g., dial-pulse signals) from those intended for the opposite direction (e.g.,answer signals). We describe the most popular method of supervision on metallic pair trunks below, namely, reverse-battery signaling.

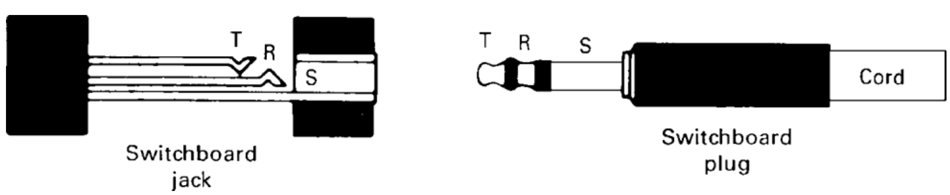


Figure 2.9 Switchboard plug with corresponding jack (R, S, and T are ring, sleeve, and tip, respectively).

Reverse-Battery Signaling

Reverse-battery signaling employs basic methods (1) and (2) just mentioned, and takes its name from the fact that battery and ground are reversed on the tip and ring to change the signal toward the calling end from on-hook to off- hook. Figure 2.10 shows a typical application of reverse-battery signaling in a common-controlpath.

In the idle or on-hook condition, all relays are unoperated and the switch (SW) con- tacts are open. Upon seizure of the outgoing trunk by the calling switch (exchange) (trunk group selection based on the switch or exchange code dialed by the calling sub- scriber), the following occur:

- SW1 and SW2 contacts close, thereby closing loop to calledoffice (exchange) and causing the A relay tooperate.
- Operation of the A relay signals off-hook (connect) indication to the called switch(exchange).
- Upon completion of pulsing between swtiches, SW3 contacts closeand the called subscriber is alerted. When the called subscriber answers, the S2 relay isoperated.
- Operation of the S2 relay operates the T relay, which reverses the voltage polarity on the loop to the callingend.
- The voltage polarity causes the CS relay to operate, transmitting an off- hook (answer) signal to the callingend.

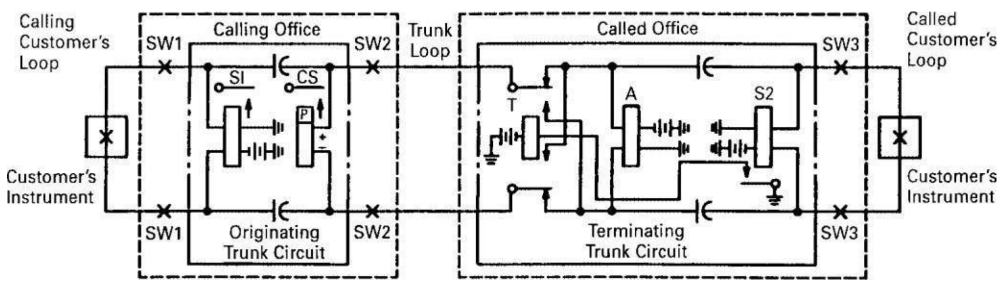


Figure 2.10 Reverse-battery signaling

When the calling subscriber hangs up, disconnect timing starts (between 150 ms and 400 ms). After the timing is completed, SW1 and SW2 contacts are released in the calling switch. This opens the loop to the A relay in the called switch and releases the calling subscriber. The disconnect timing (150-400 ms) is started in the called switch as soon as the A relay releases. When the disconnect timing is completed, the following occur: If the called subscriber has returned to on hook, SW3 contacts release. The called subscriber is now free to place another call. If the called subscriber is still off-hook, disconnect timing is started in the called switch. On the completion of the timing interval, SW3 contacts open. The called subscriber is then returned to dial tone. If the circuit is seized again from the calling switch during the disconnect timing, the disconnect timing is terminated and the called subscriber is returned to dial tone. The new call will be completed without interference from the previouscall.

When the called subscriber hangs up, the CS relay in the calling switch releases.
Then the following occur:

- If the calling subscriber has also hung up, disconnection takes place as previouslydescribed.
- If the calling subscriber is still off-hook, disconnect timing is started. On the completion of the disconnect timing, SW1 and SW2 contacts are opened. This returns the calling subscriber to dial tone and releases the A relay in the called switch. The calling subscriber is free to place a new call at this time. After the disconnect timing, the SW3 contacts are released, which releases the called subscriber. The called subscriber can place a new call at this time.

- -

TEXT / REFERENCE BOOKS

1. Roger L. Freeman, “Fundamentals of Telecommunications”, 4th Edition, John Wiley & Sons, 2010.
2. Jyrki T. J. Penttinen, “The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and SatelliteSystems”, 1st Edition, Wiley, 2015.
3. ThiagarajanViswanathan, ManavBhatnagar, “Telecommunication Switching Systems and Networks”, 2nd Edition,PHIlearning,2015.

QUESTIONBANK: PART-A

Q1:- What is inband and outband signaling?

Q2:- What is Address signaling

Q3:- What is interregister signaling?

Q4:- What is Link-by-link and end-to-end signaling?

Q5:- What is ASSOCIATED CHANNEL SIGNALING?

Q6:- What is DISASSOCIATED CHANNEL SIGNALING?

Q7:- What is Quasi-associated channel signaling?

Q8:- What is fully disassociated channel signaling?

Q9:- What is Basic Loop Signaling?

Q10:- What is Reverse-Battery Signaling?

PART-B

1. Analyze the concepts involved in Supervisory and Address signaling.
2. Explain the importance of Link by Link and End to End Signalling.
3. Develop short notes about Compelled Signalling technique.
4. Explain about reverse battery signalling.

SCHOOL OF ELECTRICAL AND ELECTRONICS

DEPARTMENT OF ELECTRONICS AND COMMMUNICATION ENGINEERING

UNIT-III TELECOMMUNICATION SYSTEMS AND SERVICES -SECA3002

UNIT 3 TELECOMMUNICATION TRAFFIC

Unit of Traffic, traffic measurement, a mathematical model, Lost- call systems: Theory, traffic performance, loss systems in tandem. Queuing systems - Erlang Distribution, probability of delay, Finite queue capacity, systems with a single server, Queues in tandem, delay tables and application of Delay formulae. Traffic Characteristics - arrival distributions, Holding time distribution. Loss Systems - Lost calls cleared, lost calls returning, lost calls Held, lost calls cleared. Overflow Traffic.

Unit of Traffic

Telecommunication traffic can be expressed in terms of the intensity of traffic and traffic volume. Measurement of traffic intensity associated with planning and technical performance of the network. Measurement of the volume of traffic associated with the revenue of telecommunication Operators andProviders.

Traffic unit of telecommunications traffic intensity on the telecommunications network based on IP (Internet Protocol) is still using Erlang. It should be noted that the Erlang which is dimensionless, suitable for use in modeling, telecommunications traffic, and can be used for all kinds of telecommunicationservices.

Traffic measurement

The volume of traffic is a traffic that is served by telecommunication systems (= Carried traffic) over a period of time measurement. A unit of telecommunications traffic volume depending on how revenues are calculated. The following gives examples of calculation and use of a unit of volume of traffic:

1. An Internet Service Provider determines that customers will have to pay $ 1 per 1 G byte of traffic used by customers, then in this case, the unit volume of traffic is G byte. If in a month, the Internet Service Provider to get the volume of traffic at 1 million G byte, then the income = 1 million G byte x $ 1 per G byte = $ 1million.
2. ATelephonyServiceProviderdeterminesthatcustomerswillhavetopay1cent

   $ every 1 minute of phone calls made by customers, so in this case, the unit volume of traffic is minute. If at one day, the traffic volume was 100 thousand minutes, then the income provider is = 100 thousand minutes x 1 cent $ per minute = $ 1,000.
3. 3. An optic communications network operator rents out bandwidth to its customers with a tariff of $ 5 per Mbps per month, then in this case, the unit volume of traffic is Mbps. If in a month, operator revenue = 5 million $, then the traffic volume in a month is = $ 5 million / $ 5 per Mbps = 1 millionMbps.

Grade of Service

For non-blocking service of an exchange, it is necessary to provide as many lines as there are subscribers. But it is not economical. So, some calls have to be rejected and retried when the lines are being used by other subscribers. The grade of service refers to the proportion of unsuccessful calls relative to the total number of calls. GOS is defined as the ratio of lost traffic to offeredtraffic.

GOS = Offered Busy Hour Calls / Blocked Busy Hour Calls

GOS = A / Ao□A Where Ao = carried traffic A = offered traffic A-Ao = loss traffic The smaller the value of grade of service, the better is the service. The recommended GOS is 0.002, i.e. 2 call per 1000 offered may lost. In a system, with equal no. of servers and subscribers, GOS is equal to zero. GOS is applied to a terminal to terminal connection. But usually a switching centre is broken into following components

Mathematical Model

Erlang-B formula is very famous in the telephony-switching circuit. For nearly a hundred years, Erlang B formula was used to calculate the traffic engineering in telecommunications networks. The results of calculations using the Erlang-B formula proved very accurate when compared with real measurement results on a telephone network that is LCC (loss-call-cleared).
There are two principles of the use of Birth and Death Process as a model of traffic in telecommunication networks, are summarized in the following:

1) The first is to determine the assumptions of the traffic coming. Erlang-B formula used in the telephone network telephone traffic is referring to the corresponding pointprocess.Therateofarrivalofcallisλandtheservicerateisμ.Traffic

   coming in and served on the telecommunications system is assumed to be a PCT-1 (Pure Chance Traffic Type 1). At traffic PCT-1 may be justified to use the average value as the basis, or known as PASTA (Poisson Arrival See Time Arrival). Traffic that comes can be expressed by erlang and written with the notation $A = \lambda / \mu$

The second is that we should be able to describe the transition of the state diagram. For that we need to know how many and how large state-coefficient birth and death-coefficient in each state.

2) The second is that we should be able to describe the transition of the state diagram. For that we need to know how many and how large state-coefficient birth and death-coefficient in eachstate.

Erlang-A Formula (also known as the first erlang formula) and Erlang-B formula known use in loss-network have been described before. Erlang is not just have been analyzed loss-network, but also have been analyzed queuing networks. Although many new queuing networks use today, but relevant formulas for queuing network has been derived by experts from 100 year ago. Formula developed by Erlang for queuing- network called the Erlang-C formula. At a later date, since D.G.Kendallpromote notation for the network queue, then the formula erlang-C is suitable for the use of the network queue M / M /n.

The formula consists of three variables, in contrast to the original standard, indicating that the variable S (number of users) is very much and formulas can be applied generally to all queuingdisciplines.

Blocking Probability

The value of the blocking probability is one aspect of the telephone company’s grade of service. The basic difference between GOS and blocking probability is that GOS is a measure from subscriber point of view whereas the blocking probability is a measure from the network or switching point of view. Based on the number of rejected 8 calls, GOS is calculated, whereas by observing the busy servers in the switching system, blocking probability will be calculated. The blocking probabilities can be evaluated by using various techniques. Lee graphs and Jacobaeus methods are popular and accurate methods . The blocking probability B is defined as the probability that all the servers in a system are busy. Congestion theory deals with the probability that the offered traffic load exceeds some value. Thus, during congestion, no new calls can be accepted. There are two ways of specifying congestion. They are time congestion and call congestion. Time congestion is the percentage of time that all servers in a group are busy. The call or demand congestion is the proportion of calls arising that do not find a free server. In general GOS is called call congestion or loss probability and the blocking probability is called time congestion. If the number of sources is equal to the number of servers, the time congestion is finite, but the call congestion is zero. When the number of sources is large, the probability of a new call arising is independent of the number already in progress and therefore the call congestion is equal to timecongestion.

Erlang Distribution

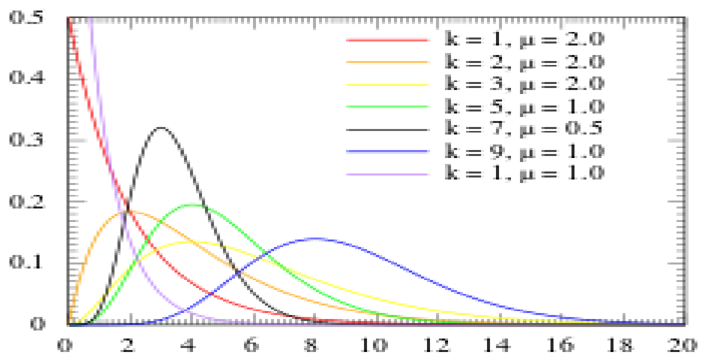
The Erlang distribution (sometimes called the Erlang-k distribution) was developed by

A.K. Erlang to find the number of phone calls which can be made simultaneously to switching station operators. Erlang was a telecommunications engineer for the Copehagen Telephone Company; his formulas for loss and waiting time were used by many telephone companies, including the British Post Office. Erlang's distribution has since been expanded for use in queuing theory, the mathematical study of waiting in lines. It is also used in stochastic processes and in mathematical biology.

The Erlang distribution is a specific case of the Gamma distribution. It is defined by two parameters, k and &u;, where:

k is the shape parameter. This must be a positive integer (an integer is a whole number without a fractional part). In the Gamma distribution, k can be any real number, includingfractions.

μ is the scale parameter. Must be a positive real number (a real number is any number found on the number line, including fractions).



The probability distribution function of the Erlang distribution is:

$$f(x;k,\mu) = \frac{x^{k-1}e^{-\frac{x}{\mu}}}{\mu^k (k-1)!} \quad \text{for } x, \mu \geq 0.$$

The factorial(!) in the denominator is the reason why the distribution is only defined for positive numbers. An equivalent form of the pdf for this distribution includes λ, a measure of rate, which is related to μ in the following way:
μ = 1/λ.

λ represents the number of items or calls expected in a given amount of time.

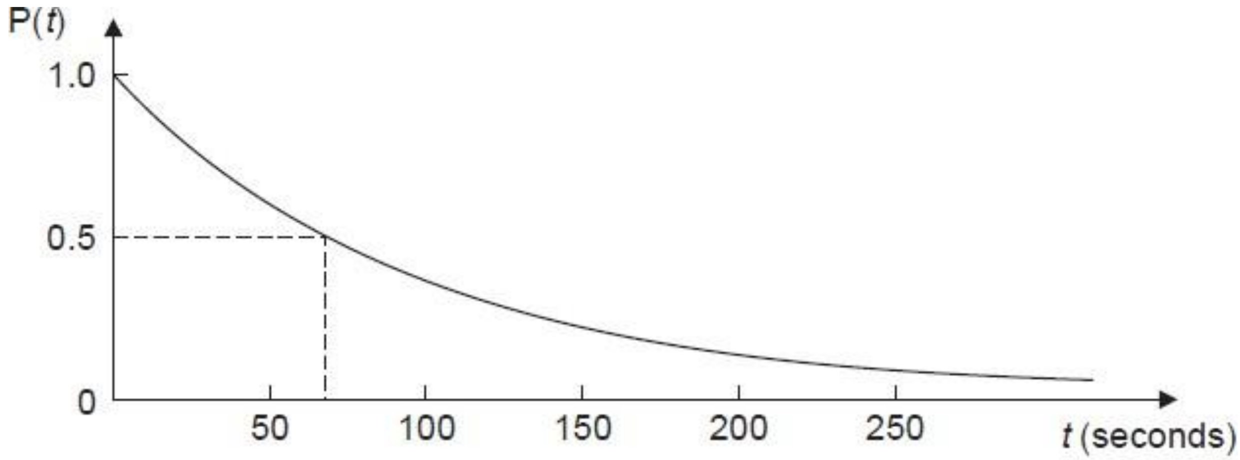
Holding time

- Holding time. The average holding time or service time 'h' is the averageduration of occupancy of a traffic path by a call. For voice traffic, it is the average holding time per call in hours or 100 seconds and for data traffic, average transmission per message inseconds.

- The reciprocal of the average holding time referred to as service rate (μ) incalls per hour is givenas
- Sometimes, the statistical distribution of holding time is needed. The distribution leads to a convenient analyticequation.
- The most commonly used distribution is the negative exponential distribution.The probability of a call lasting atleast t seconds is givenby

$$P(t) = \exp(-t/h)$$

For a mean holding time of h = 100 seconds, the negative exponential distribution function is shown in Fig



Negative distribution function for h = 100 sec.

Figure shows that, 50% probability call lasts longer than 70 sec.

Lost- call systems

The service of incoming calls depends on the number of lines. If number of lines equal to the number of subscribers, there is no question of traffic analysis. But it is not only uneconomical but not possible also. So, if the incoming call finds all available lines busy,thecallissaidtobeblocked.Theblockedcallscanbehandledintwoways.The

type of system by which a blocked call is simply refused and is lost is called loss system. Most notably, traditional analog telephone systems simply block calls from entering the system, if no line available. Modern telephone networks can statistically multiplex calls or even packetize for lower blocking at the cost of delay. In the case of data networks, if dedicated buffer and lines are not available, they block calls from entering the system. In the second type of system, a blocked call remains in the system and waits for a free line. This type of system is known as delay system. In this section loss system is described. These two types differs in network, way of obtaining solution for the problem and GOS. For loss system, the GOS is probability of blocking. For delay system,

GOS is the probability of waiting. Erlang determined the GOS of loss systems having N trunks, with offered traffic A, with the following assumptions. (a) Pure chance traffic (b) Statistical equilibrium (c) Full availablity and (d) Calls which encounter congestion are lost.The first two are explained in previous section. A system with a collection of lines is said to be a fully accessible system, if all the lines are equally accessible to all in arriving calls. For example, the trunk lines for inter office calls are fully accessible lines. The lost call assumption implies that any attempted call which encounters congestion is immediately cleared from the system. In such a case, the user may try again and it may cause more traffic during busyhour.

The Erlang loss system may be defined by the following specifications. Calls received as per 1.The arrival process of calls is assumed to be Poisson with a rate of hour. 2. The holding times are assumed to be mutually independent and identically distributed random variables following an exponential distribution with 1/ seconds. 4. Calls are served in the order of arrival. There are three models of loss systems. They are: 1. Lost calls cleared (LCC) 2. Lost calls returned (LCR) 3. Lost calls held(LCH)

Lost Calls Returned (LCR) System

In LCC system, it is assumed that unserviceable requests leave the system and never return. This assumption is appropriate where traffic overflow occurs and the other routes are in other calls service. If the repeated calls not exist, LCC system is used. But in many cases, blocked calls return to the system in the form of retries.

Some examples are subscriber concentrator systems, corporate tie lines and PBX trunks, calls to busy telephone numbers and access to WATS lines. Including the retried calls, the offered traffic now comprises two components viz., new traffic and retry traffic. The model used for this analysis is known as lost calls returned (LCR) model.

The following assumptions are made to analyze the CLR model.

1. All blocked calls return to the system and eventually get serviced, even if multiple retries arerequired.

2. Time between call blocking and regeneration is random statistically independent of each other. This assumption avoids complications arising when retries are correlated to each other and tend to cause recurring traffic peaks at a particular waiting timeinterval.

3. Time between call blocking and retry is somewhat longer than average holding time of a connection. If retries are immediate, congestion may occur or the network operation becomes delaysystem.

Consider a system with first attempt call arrival ratio of λ (say 100). If a percentage B (say 8%) of the calls blocked, B times λ retries (i.e. 8 calls retries). Of these retries, however a percentage B will be blocked again.

$\lambda' = \lambda + B\lambda + B2\lambda + B3\lambda + ...$

where B is the blocking probability from a lost calls cleared (LCC) analysis.

The effect of returning traffic is insignificant when operating at low blocking probabilities.

At high blocking probabilities, it is necessary to incorporate the effects of the returning traffic into analysis. The effect high blocking probability, is illustrated in the following example.

Lost Calls Cleared (LCC) System:

The LCC model assumes that, the subscriber who does not avail the service, hangs up the call,and tries later. The next attempt is assumed as a new call. Hence, the call is said to be cleared.

This also referred as blocked calls lost assumption. The first person to account fully and accurately for the effect of cleared calls in the calculation of blocking probabilities was A.K.Erlang in 1917.

Consider the Erlang loss system with N fully accessible lines and exponential holding times. The Erlang loss system can be modeled by birth and death process with birth and death rate as follows.

$$\lambda_k = \begin{cases} \lambda, k = 0, 1, \ldots\ldots N-1 \\ 0 \quad k \geq N \end{cases} \qquad \ldots(8.48)$$

$$\mu_k = \begin{cases} k\mu, k = 0, 1, \ldots\ldots N \\ 0, \quad k > N \end{cases} \qquad \ldots(8.49)$$

From 8.46, $\quad P(k) = \dfrac{\lambda_0 \lambda_1 \ldots\ldots \lambda_{k-1}}{\mu_1 \mu_2 \ldots\ldots \mu_k} P(0), k = 1, 2, 3$

Substituting equation (8.48 and 8.49) in the above equation, we get

$$P(k) = \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k P(0), k = 1, 2, 3, \ldots\ldots N \qquad \ldots(8.50)$$

From equation (8.4), the offered traffic is

$$A = \frac{\lambda}{\mu}$$

Substituting in (8.50), we get

- $$P(k) = \frac{1}{k!} (A)^k P(0), k = 1, 2, 3, \ldots\ldots N \qquad \ldots(8.51)$$

The probability P(0) is determined by the normalization condition

$$\sum_{k=0}^{N} P(k) = P(0) \sum_{k=0}^{N} \frac{A^k}{k!} = 1$$

$$P(0) = \frac{1}{\sum_{k=0}^{N} \frac{A^k}{k!}} \qquad \ldots(8.52)$$

Substituting (8.52) in (8.51), we get

$$P(k) = \frac{A^k / k!}{\sum_{k=0}^{N} \frac{A^k}{k!}} \qquad \ldots(8.53)$$

The probability distribution is called the truncated **Poisson distribution** or **Erlang's loss distribution.** In particular when $k = N$, the probability of loss is given by

$$P(N) = B(N, A) = \frac{A^N}{N! \sum_{k=0}^{N} \left(\frac{A^k}{k!}\right)} \qquad \ldots(8.54)$$

where $A = \lambda/\mu$.

This result is variously referred to as **Erlang's formula of the first kind, the Erlangs-B formula or Erlangs loss formula.**

Equation 8.54 specifies the probability of blocking for a system with random arrivals from an infinite source and arbitrary holding time distributions. The Erlang B formula gives the time congestion of the system and relates the probability of blocking to the offered traffic and the number of trunklines.

Values from B(N, A) obtained from equation 8.54, have been plotted against theoffered traffic 'A' erlangs for different values of the number of N lines in Fig. 8.6.

In design problems, it is necessary to find the number of trunk lines needed for a given offered traffic and a specified grade of service. The offered load generated by a Poisson input process with a rate λ calls per hour may be definedas

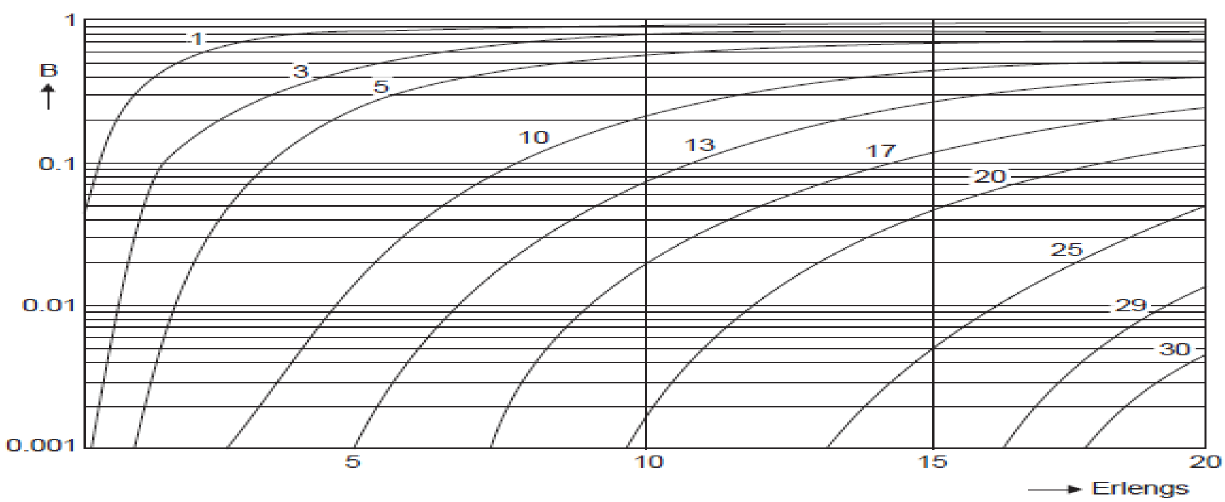
$$A = \int_0^{\infty} \lambda t \, dH(t) = \lambda h \qquad \ldots(8.55)$$

where $\lambda\, t$ = average number of increasing calls at fixed time internal

h = average holding time

The average number of occupied or busy trunks is defined as the carried load

$$A' = \sum_{k=1}^{N-1} k\, P(k) + N \sum_{k=N}^{\infty} P(k) \qquad \ldots(8.56)$$



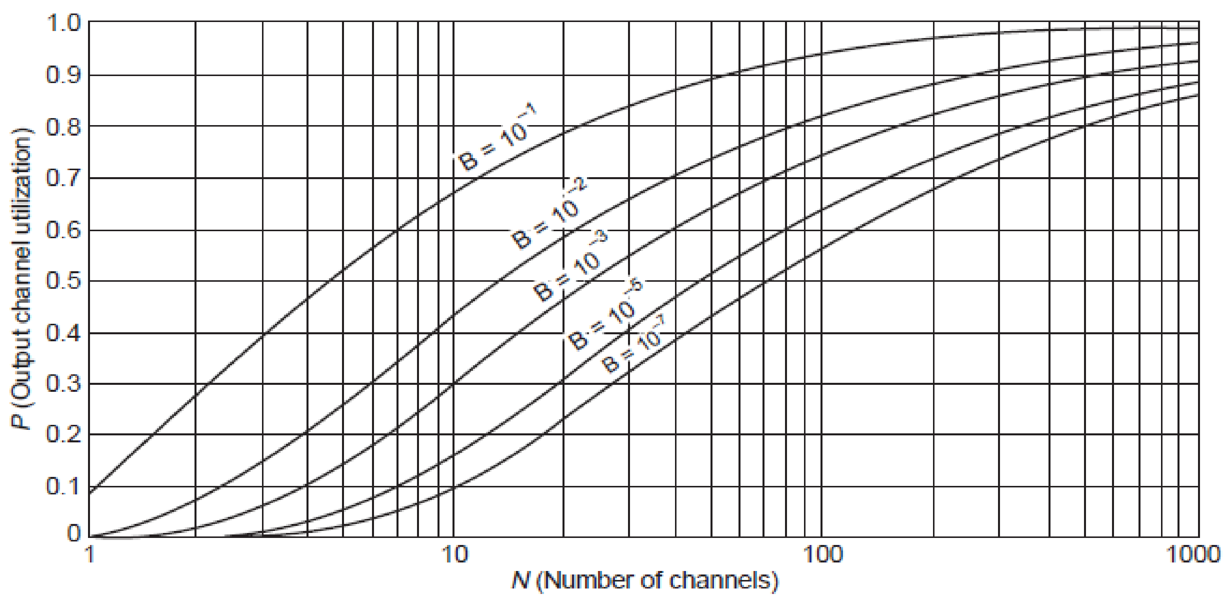
Plot of B(N, A).

$$A' = A\,[1 - B(N, A)] \qquad \ldots(8.57)$$

Thus, the carried load is the position of the offered load that is not lost from the system. The carried load per line is known as the trunk occupancy. From 8.57,

$$\rho = \frac{A'}{N} = \frac{A(1-B)}{N} \qquad \ldots(8.58)$$

The trunk occupancy ρ is a measure of the degree of utilization of a group of lines and is sometimes called the utilization factor. Fig. below presents the output channel utilization for various blocking probabilities and number of servers.



Output channel utilization of LCC system.

In designing a telephone system, it is necessary to ensure that the system will operate satisfactorily under the moderate overload condition.

Lost Calls Held (LCH) System

- In a lost calls held system, blocked calls are held by the system andserviced when the necessary facilities become available. The total time spend by a call is the sum of waiting time and the servicetime.
- Each arrival requires service for a continuous period of time andterminates its request independently of its being serviced or not. If number of calls blocked, a portion of it is lost until a server becomes free to service acall.
- An example of LCH system is the time assigned speech interpolation(TASI) system.
- LCH systems generally arise in real time applications in which the sources are continuously in need of service, whether or not the facilities are available. Normally, telephone network does not operate in a lost call heldmanner.

- The LCH analysis produces a conservative design that helps account for retries and day to day variations in the busy horn callingintensities.
- A TASI system concentrates some number of voice sources onto a smaller number of transmission channels. A source receives service only when it is active. If a source becomes active when all channels are busy, it is blocked and speech clippingoccurs.
- Each speech segment starts and stops independently of whether it is served or not. Digital circuit multiplication (DCM) systems in contrast with original TASI, can delay speech for a small amount of time, when necessary to minimize theclipping.
- LCH are easily analysed to determine the probability of the total number of calls in the system at any one time. The number of active calls in the system at any time is identical to the number of active sources in a system capable of carrying all traffic as it arises. Thus the distribution of the number in the system is the poisson distribution. The poisson distribution givenas
- The probability that k sources requesting service are being blocked is simply the probability that k + N sources are active when N is the number of servers.Based on the assumption that the routing is made only by direct routing or tandom routing, it is found that to route a stream of traffic, tandom route is more economical. In fact, even greater economics are often possible if just a proportion of the traffic is routed directly. This approach is known as alternativerouting.
- In alternative routing, connections should use the direct trunks (referred as high usage route), because direct route provides better transmissionquality and use fewer networkfacilities.

- If all the direct trunks are busy, calls are routed via a tandom exchanges or alternate routes to maintain suitably low blockingprobabilities.
- Thus, the networks are designed to allocate a limited number of heavily utilized trunks in the direct route and provide alternate routes for overflow.
- If the high usage route consists of N tunks and the offered traffic is A erlangs, the probability of all trunks busy is given by the Erlangs–Bformula

$$P(N) = B(N, A) = \frac{A^N}{N! \sum_{k=0}^{N} \left(\frac{A^k}{k!} \right)}$$

- The traffic carried on high usage route AH is given by AH =

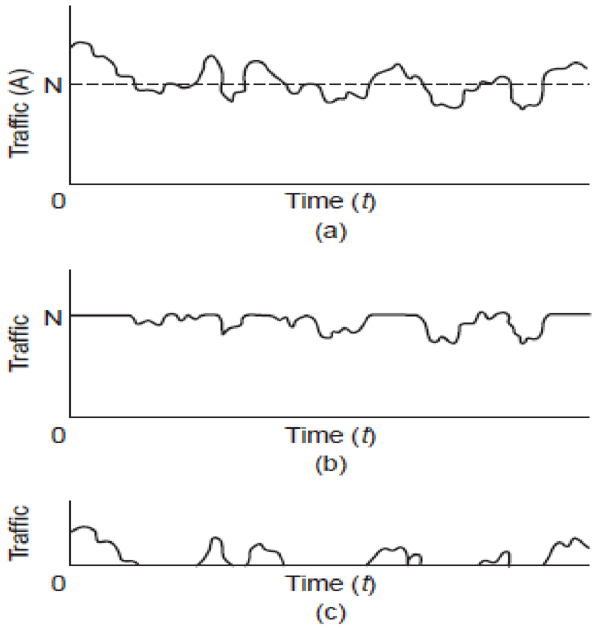
A(1 – B(N, A))erlangs

where $A = \lambda/\mu$.

The overflow traffic is A0 = AB(N, A) erlangs

- The Erlang-B formula is a good representation of the traffic on a high usage routebecause
- blocked calls are diverted to the alternative route and does not reappear. But the number of circuits required by a final route to carry the overflow traffic should not be calculatedfrom
- Erlang's-B formula, because this traffic is not poissonian. The characteristic of traffic for high usage route with overflow is shown inFig.

Characteristic traffic for high usage route with overflow



(a) shows the traffic offered to high usage route. The traffic carried by thehigh usage routes are shown in(*b*).

(b) depictsthatthetrafficcarriedisequaltothetrafficoffered,ifitislessthanor equal to number of high usage trunks. If the offered traffic is greater than the number of high usage routes, overflow occurs and the traffic carried equal to number oftrunks.

(c) shows the over flowtraffic.

The traffic offered to the final route is thus more peakly than poisonian traffic. The analysis of this traffic requires mean as well as variance.

The Wilkinson equivalent Random theory is the widely used method to analyse the random overflow traffic.

- Under static equilibrium condition A 5 N. But if A * N; i.e. traffic offeredis higher than nitmber of trunks then length of queue increasestowards infinity.

+ U total number of calls (x) in system is less than number of trunks then calls will be served without delay i.e. there is noqueue.

- The system behaves as lost call system without congestion.Then Their calls arriving is much greater than number of trunks. The incoming calls encounter delay as all the servers are busy.

Probability of Delay

- When incoming calls x are much more than available trunks N i.e. $x \geq N$. The delayeccun

c The probability that there are at least Zc6lJ81 3 ternis expressedUS '

$$P(x \geq Z) \quad \sum_{x=z}^{\infty} P(x)$$

N! o',$g M

$$\frac{N^N}{N!} P(0) \left(\frac{A}{N}\right)^z \sum_{k=0}^{\infty} \left(\frac{A}{N}\right)^k \quad \forall k = x - N$$

$$P(x \text{ ñ}Z) \quad \frac{N'}{M!} \left(\frac{A}{N}\right)^z P(0) \left[1 - \frac{A}{N}\right]^{-1}$$

$$9xzz) \quad \frac{N^N}{N!} \left(\frac{A}{N}\right)^z \left(\frac{N}{N-A}\right) P0)$$

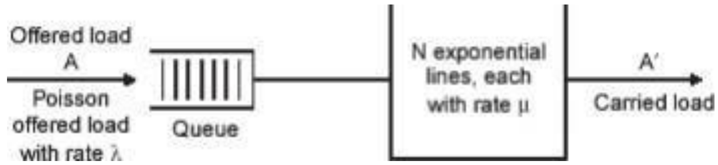
The probability o/ delay, Pp= P(x * N)

$$P_D = \frac{A^N}{N!} \cdot \frac{N}{N-A} P(0)$$

= *z ri(A)

Overflow Traffic

The delay system places the call or message arrivals in a queue if it finds all N servers (or lines) occupied. This system delays non-serviceable requests until the necessary facilities become available. These systems are variously referred to as delay system, waiting-call systems and queuing systems. The delay systems are analyzed using queuing theory which is sometimes known as waiting line theory. This delay system has wide applications outside the telecommunications. Some of the more common applications are data processing, supermarket checkout counters, aircraft landings, inventory control and various forms of services. Consider that there are k calls (in service and waiting) in the system and N lines to serve the calls. If $k \leq N$, k lines are occupied and no calls are waiting. If $k > N$, all N lines are occupied and $k - N$ calls waiting. Hence a delay operation allows for greater utilization of servers than does a losssystem.Eventhougharrivalstothesystemare22random,theserversseea somewhat regular arrival pattern. A queuing model for the Erlang delay system is shown in Fig



The basic purpose of the investigation of delay system is to determine the probability distribution of waiting times. From this, the average waiting time W as random variable can be easily determined. The waiting times are dependent on the following factors: 1. Number of sources 2. Number of servers 3.Intensity and probabilistic nature of the offered traffic 4. Distribution of service times 5. Service discipline of the queue. In a delay system, there may be a finite number of sources in a physical sense but an infinite number of sources in an operational sense because each source may have an arbitrary number of requests outstanding. If the offered traffic intensity is less than the servers, no statistical limit exists on the arrival of calls in a short period of time. In practice, only finite queue can be realized. There are two service time distributions. They are constant service times and exponential service times. With constant service times, the service time is deterministic and with exponential, it is random. The service discipline of the queue involves two important factors. 1. Waiting calls are selected on of first-come, first served (FCFS) or first in-first-out (FIFO) service. 2. The second aspect of the service discipline is the length of the queue. Under heavy loads, blocking occurs. The blocking probability or delay probability in the system is based on the queue size in comparison with number of effective sources. We can model the Erlang delay system by the birth and death process with the following birth and death ratesrespectively.

TEXT / REFERENCE BOOKS

1. Roger L. Freeman, “Fundamentals of Telecommunications”, 4th Edition, John Wiley & Sons, 2010.
2. Jyrki T. J. Penttinen, “The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and SatelliteSystems”, 1st Edition, Wiley, 2015.
3. ThiagarajanViswanathan, ManavBhatnagar, “Telecommunication Switching Systems and Networks”, 2ndEdition,PHIlearning,2015.

QUESTIONBANK:

PART-A

1. If we know that there are 354 seizures (lines connected for service) and 6 blocked calls (lost calls) during the BH, what is the grade of service?
2. A particular exchange has been dimensioned to handle 1000 calls during the busy hour. On a certain day during the BH 1100 calls are offered. What is the resulting grade of service?
3. 1400 calls were offered in busy hour with some trunk lines. Out of these 7 calls were lost. Average call duration was about 4 minutes. What is traffic offered, traffic carried, traffic lost, GOS and Congestion period.
4. If a telephone exchange serves 1500 users with the average BHCA of about 9000 and CCR is about 50%, what would be the busy hour calling rate?
5. Mention the relation between Erlang and CCS.
6. List out the models of loss systems.
7. What is Call Completion Rate?
8. What is meant by trunk occupancy?

PART-B

1. Determine the Erlangs loss formula for Lost Calls Cleared System (LCC).
2. If a group of 20 trunk carries 10 erlangs and the average call duration is 3 min (a) average number of calls in progress (b) total number of calls originating pe
3. Estimate (a) the grade of service (b) the probability that only one trunk is busy (c) the probability that only one trunk is free and (d) the probability that at least one trunk is free.Consider for a group of 7 trunks is offered 4E of traffic.
4. Discuss in detail about Lost Calls Held system.

SCHOOL OF ELECTRICAL AND ELECTRONICS

DEPARTMENT OF ELECTRONICS AND COMMMUNICATION ENGINEERING

UNIT-IV TELECOMMUNICATION SYSTEMS AND SERVICES – SECA3002

UNIT 4 TELECOMMUNICATION SERVICES ENGINEERING

Introduction, definition for service and service engineering. Telecommunication services engineering-Telecommunication services on broad band networks - basics of ATM, connection oriented and connectionless services.

Telecommunication services Engineering

Introduction:

The demand for advanced telecommunication services has led enormously over the last few years. This has led the situations where N/W operators must deploy new services at a rapid pace when satisfying customer needs.

The demand for ever more specialized end-user services keeps growing, along with the demand for having the new services deployed within shorter & shorter time frames. Structure & function of network must change, in order to scope with their new challenges.

A new discipline called "TSE" is emerging. It encompasses set of principles, architectures & tools required to tackle activities ranging from service specification to service implementation, service deployment & exploitation.

Definition for service:

The word "service" has become a magic word in the telecommunication world there last years.

This word is somewhat fuzzy & ambiguous there are so many aspects of services.

Telecommunication services-common name for all services offered by or over a telecommunication network.

In telecommunication, a telecommunication service is a service provided by a telecommunication provider or a specified set of user-information transfer capabilities provided to a group of users, by telecommunications systems.

The telecommunications service user is responsible for the information content of the messages.

Telecommunications service provider has responsibility for the acceptance, transmission & delivery of the messages.

Definition for service engineering:

It can be defined as the set of methods, techniques tools to specify, design, implement, verify & validate service that meet user needs & deploy & exploit their services over the current or future networks.

It is a young discipline, but is a discipline in itself, as is protocol engineering.

Telecommunication services engineering:

-3 important components are considered within the framework of service engineering as fig.

1) Service creationenvironment:

Software engineering platform specialized for the development of telecommunications services.

2) TelecommunicationNetwork:

It contains transmission & switching equipment. Each of these equipment may be seen as one black box that offers an application programming interface (API), this may be a signaling/management interface.

3) NetworkArchitecture:

It is responsible for controlling the N/W in such a way that a service's specific requirements get satisfied.
Three domains:

Service engineering covers 3 domains,

1) Service creation-where services is considered as a distributed application running onthe multiple nodes of a telecommunicationnetwork.

2) Service management-refers to the way of services is operated throughout itslifecycle.

3) Network management-refers to management of network resources used to provide telecommunicationservices.

2 kinds of services are involved, telecommunication services & management services.

Telecommunication services on broad band networks:

Broadband:

It commonly refers to high speed internet access that is always on & faster than the traditional dial-up access.

It includes several-high speed transmission technologies such as digital subscriber line (DSL) cable modern.

In telecommunication broadband is wide Bandwidth data transmission which transports multiple signals & traffic types.

The medium can be co-axial cable, optical fiber, radio/twisted pair.

Definition for connectionless:

In telecommunication, connectionless describes communication between two network end points in which a message can be sent from one end point to another without prior arrangement. The device at one end of the communication transmits data to the other, without first ‘ensuring that the recipient is available and ready to receive the data. The device sending a message simply sends it addressed to the intended recipient. If there are problems with the transmission, it may be necessary to resend the data several times. The Internet Protocol (IP) and User Datagram Protocol (UDP) are connectionless protocols.

Definition for connection-oriented service:

A connection-oriented service is a technique used to transport data at the session layer. Unlike its opposite, connectionless service, connection-oriented service requires that a session connection be established between the sender and receiver, analogous to a phone call. This method is normally considered to be more reliable than a connectionless service, although not all connection-oriented protocols are considered reliable.

A connection-oriented service can be a circuit-switched connection or a virtual circuit connection in a packet-switched network. For the latter, traffic flows are identified by a connection identifier, typically a small integer of 10 to 24 bits. This is used instead of listing the destination and sourceaddresses.

Parameter	Connection Less Service	Connection Oriented Service
Definition	It is the Communication System in which there is no need to establish virtual connection between sender and receiver.	It is the communication system in which virtual connection is established between sender and receiver before the communication beings.
Data Acknowledge	No data acknowledge is used, sender cannot be sure about the accurate delivery of themessage.	Receiver can acknowledge the data send by the sender and can re request the data if any packet fails or getsdamaged.
Connection Termination	No Need of Connection termination.	Connection needs to be terminated after completion of communication.
Packet Route	Packets follow different path to reach destination and ma reach in an order.	All the frames are sent through same route or path.
Example	Postal System	Telephone Call

Key terms Used:

Sender: The individual that initiates a message in communication system is called sender. Anybody e.g. A speaker, Writer or somebody who merely gestures can be a sender.

Receiver: Reader, Observer or Listener in the communication system is called receiver.

Channel: Physical transmission medium through which message is passed from sender to receiver is called communication channel. That may be wired or wireless.

Protocol: These are the pre-defined, pre-agreed rules between sender and receiver for the communication.

Transmission: It is the process of sending and receiving digital and analog signals over wired and unwired network media.

Packets: It is the formatted data unit used in packet switched networks. Packets contain two types of data:

Control Information: It Provides information like sender, receiver, communication path, error control and detection bits etc.

User Data: Actual data to be sent.

Asynchronous Transfer Mode (ATM):

By the mid l980s, three types of communication networks had evolved.

The telephone network carries voice calls, television network carries video transmissions, and newly emerging computer network carries data.

Telephone companies realized that voice communication was becoming a commodity service and that the profit margin would decrease over time.

They realized that data communication was increasing.

The telecommunication industry decided to expand its business by developing networks to carry traffic other thanvoice.

Goal of ATM (extremely

ambitious) UniversalService

Support for all users
Single, unified infrastructure - Service

guarantees Support for low-cost Devices

ATM

The phone companies created Integrated Service Digital Network (ISDN) and Asynchronous Transfer Mode (ATM).

ATM is intended as a universal networking technology that handles voice, video, and data transmission.

ATM uses a connection--oriented paradigm in which an application first creates a virtual channel (VC), uses the channel for communication, and then terminates it.

The communication is implemented by one or more ATM switches, each places an entry for the VC in its forwarding table.

ATM

There are two types of ATM VCs: a PVC is created manually and survive power failures, and an SVC is created on demand.

When creating a VC, a computer must specify quality of service (QoS)

requirements. The ATM hardware either reserves the requested resources or denies the request.

Development of ATM:

ATM designers faced a difficult challenge because the three intended uses (voice, video, and data) have different sets of requirements.

For example, both voice and video require low delay and low jitter (i.e. low variance in delay) that make it possible to deliver audio and video smoothly with gaps or delays in the output.

Video requires a substantially higher data rate than

audio. Most data networks introduce jitter as they handle

packets.

To allow packet switches to operate at high speeds and to achieve low delay, low jitter, and echo cancellation, ATM technology divides all data into small, fixed-size packets called cells.

Each ATM cell contains exactly 53

octets. 5 octets for header

48 octets for data

ATM Cell

Structure

Bits: 07

Flow Control	VPI (First 4 bits)	
VPI (Last 4 bits)	VCI (First 4 bits)	
VCI(Middle 8 bits)		
VCI (Last 4 bits)	Payload type	PRIO
Cyclic Redundancy Check		

48 Data Octets start here

ATM design and cells:

ATM was designed to be completely general. We will large cell for data and small cell for voice.

In ATM, cell size is chosen as a compromise between large cells and small cells. Header is l0% of the payload area.

In Ethernet: data => 1500 octets

header => 14 octets

cell tax =>l%

In ATM: data m> 48

octets header => 5 octets

cell tax => 10%

ATM : Connection oriented

After the establishment of a connection between sender and receiver, the network hardware returns a connection identifier (a binary value) to each of the two computers.

When sender sends cells, it places the connection identifier in each cell header.

When it receives a cell, an ATM switch extracts the connection identifier and consults a table to determine how to forward the cell.

VPI/VCI:

Formally, an ATM connection is known as a Virtual channel (VC).

ATM assigns each VC a 24-bit identifier that is divided into 2 parts to produce a hierarchy.

The first part, a Virtual path identifier (VPI), specifies the path the VC follows through the network.

A VPI is 8 bits long.
The second part, a Virtual Channel Identifier (VCI), specifies a single VC within the

path. A VCI is 16 bits long.

ATM ProtocolLayer:

Physical Layer: The lowest layer in the ATM protocol. It describes the physical transmission media. We can use shielded and unshielded twisted pair, coaxial cable, and fiber-optic cable.

ATM Layer: It performs all functions relating to the routing and multiplexing of cells over VCs. It generates a header to the segment streams generated by the AAL. Similarly, on receipt of a cell streams, it removes the header from the cell and pass the cell contents to the AAL protocol. To perform all these functions, the ATM layer maintains a table which contains a list of VCIS.

ATM Adaptation Layer: Top layer in the ATM protocol Model. It converts the submitted information into streams of 48-octet segments and transports these in the payload field of multiple ATM cells. Similarly, on receipt of the stream of cells relating to the same call, it converts the 48-octet information field into required form for delivery to the particular higher protocol layer. Currently five service types have been defined. They are referred to as AALl-5. AALI and AAL2 are connection oriented. AALI provides a constant bit rate (CBR) service, whereas AAL2 provides a variable bit rate (VBR) service. Initially, AAL 3 was defined to provide connection oriented and VBR service. Later, this service type was dropped and it is now merged with AAL 4. Both AAL % and AAL 5 provide a similar connectionless VBRservice.

Disadvantages:

ATM has not been widely accepted. Although some phone companies still use it in their backbone networks.

The expense, complexity and lack of interoperability with other technologies have prevented ATM from becoming more prevalent.

Expense: ATM technology provides a comprehensive list of services, even a moderate ATM switch costs much more than inexpensive LAN hardware. In addition, the network interface card needed to connect a computer to an ATM network is significantly more expensive than a corresponding Ethernet NIC.

Connection Setup Latency: ATM's connection-oriented paradigm introduces significant delay for distant communication. The time required to set up and tear down the ATM VC for distant communication is significantly larger than the time required using it.

Cell Tax: ATM cell headers impose a 10% tax on all data transfer. In case of Ethernet, cell tax is 1%. **Lack of Efficient Broadcast**: Connection-oriented networks like ATM are sometimes called Non Broadcast Multiple Access (NBMA) networks because the hardware does not support broadcast or multicast. On an ATM network, broadcast to a set of computers is 'simulated' by arranging for an application program to pass a copy of the data to each computer in the set. As a result, broadcast is in efficient.

Complexity of Quality of service: The complexity of the specification makes implementation cumbersome and difficult. Many implementations do not support the full standard.

Assumption of Homogeneity: ATM is designed to be a single, universal networking system. There is minimal provision for interoperating with other technologies.

TEXT / REFERENCE BOOKS

1. Roger L. Freeman, “Fundamentals of Telecommunications”, 4th Edition, John Wiley & Sons, 2010.
2. Jyrki T. J. Penttinen, “The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and SatelliteSystems”, 1st Edition, Wiley, 2015.
3. ThiagarajanViswanathan, ManavBhatnagar, “Telecommunication Switching Systems and Networks”, 2ndEdition,PHIlearning,2015.

QUESTION BANK:

PART-A

1. Name the types of network services.
2. Identify the components of service engineering.
3. Compare the connection oriented and connectionless services.

PART-B

1. Compare telecommunication services and service engineering.
2. Develop short notes about features of ATM.
3. Explain following terms with example: (i) Connection oriented (ii)Connection less service.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS

DEPARTMENT OF ELECTRONICS AND COMMMUNICATION ENGINEERING

UNIT –V TELECOMMUNICATION SYSTEMS AND SERVICES -SECA3002

UNIT 5 QUALITY OF SERVICE AND TELECOMMUNICATION IMPAIRMENTS

QoS (voice, data and image) - signal-to-noise ratio, voice transmission, data circuits, video.Basic impairments - amplitude distortion, phase distortion and noise. Level - typical levels, echo and singing. QoS issues in video transmission - problems and solutions. Protocols for QoS support for audio and video applications - RSVP applications, Real-Time Streaming Protocol Applications and Active Streaming Format, Internet stream protocol (version 2), IP Multicast.

OBJECTIVE Quality of service (QoS)

There are a number of generic impairments that will directly or indirectly affect quality of service. An understanding of these impairments and their underlying causes is extremely important if one wants to grasp the entire picture of a telecommunication system.

QUALITY OF SERVICE: VOICE, DATA, AND IMAGE

Introduction to Signal-to-Noise Ratio

Signal-to-noise ratio (S/N or SNR) is the most widely used parameter for measurement of signal quality in the field of transmission. Signal-to-noise ratio expresses in decibels the amount by which signal level exceeds the noise level in a specified bandwidth.

The following are S/N guidelines at the corresponding receiving devices:

Voice: 40 dB;

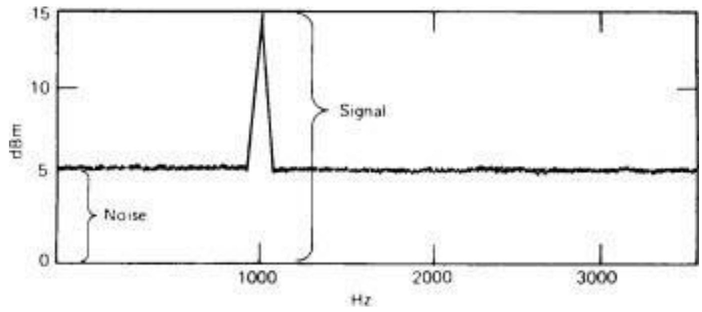
Video (TV): 45 dB;

Data: ∼15 dB, based upon the modulation type and specified error performance.

To demonstrate the concept of S/N, consider Figure 1. This oscilloscope presentation shows a nominal analog voice channel (300 –3400 Hz) with a 1000-Hz test signal. The vertical scale is signal power measured in dBm and the horizontal scale is frequency, 0 Hz to 3400 Hz. The S/N as illustrated is 10 dB. We can derive this by inspection or by reading the levels on the oscilloscope presentation. The signal level is +15 dBm; the noise is +5 dBm,then:

(S/N)dB =level(signal in dBm) - level(noiseindBm)

(1)



Signal to Noise Ratio

Inserting the values given in the oscilloscope example,

: S/N c +15 dBm ` (+5 dBm) c 10 dB. This expression is set up as shown because we are dealing withlogarithms.Whenmultiplying inthedomainoflogarithms,weadd.When dividing, we subtract. We are dividing because on the left side of the equation we have S/N or S divided byN.

Signal-to-noise ratio really has limited use in the PSTN for characterizing speech transmission because of the “spurtiness” of the human voice. We can appreciate that individual talker signal power can fluctuate widely so that the S/N ratio is far from constant during a telephone call and from one telephone call to thenext.

In lieu of actual voice, we use a test tone to measure level and S/N. A test tone is a single frequency, usually around 800 or 1000 Hz, generated by a signal generator and inserted in the voice channel. The level of the tone (often measured in dBm) can be easily measured with the appropriate test equipment. Such a tone has constant amplitude and no silent intervals, which is typical of voice transmission .

Voice Transmission

Loudness Rating and Its Predecessors.

“Hearing sufficiently well” on a telephone connection is a subjective matter. This is a major element of QoS. Any method to measure “hearing sufficiently well” should incorporate intervening losses on a telephone connection.

Losses are conventionally measured in dB. Thus the unit of measure of “hearing sufficiently well” is the decibel. From the present method of measurement we derive the loudness rating, abbreviated LR.

It had several predecessors: reference equivalent and corrected reference equivalent.

Reference Equivalent.

The reference equivalent value, called the overall reference equivalent (ORE), was indicative of how loud a telephone signal is. How loud is a subjective matter. Given a particular voice level, for some listeners it would be satisfactory, others unsatisfactory.

The ITU in Geneva brought together a group of telephone users to judge telephone loudness. A test installation was set up made up of two standard telephone subsets, a talker’s simulated subscriber loop and a listener’s simulated loop. An adjustable attenuating network was placed between the two simulated loops. The test group, on an individual basis, judged level at the receiving telephone earpiece. At a 6-dB setting of the attenuator or less, calls were judged too loud. Better than 99% of the test population judged calls to be satisfactory with an attenuator setting of 16 dB; 80% rated a call satisfactory with an ORE 36 dB or better, and 33.6% of the test population rated calls with an ORE of 40 dB as unsatisfactory, and so on.

To calculate overall reference equivalent (ORE) we summed the three dB values (i.e., the transmit reference equivalent of the telephone set, the intervening network losses, and the receive reference equivalent of the same type subset).

Corrected Reference Equivalent. Because difficulties were encountered in the use of reference equivalents, the ORE was replaced by the corrected reference equivalent (CRE) around 1980. The concept and measurement technique of the CRE was essentially the same as RE (reference equivalent) and the dB remained the measurement unit. CRE test scores variedsomewhatfromitsREcounterparts.Lessthan5dB(CRE)wastooloud;anoptimum

connection had an RE value of 9 dB and a range from 7 dB to 11 dB for CRE. For a 30-dB value of CRE, 40% of a test population rated the call excellent, whereas 15% rated it poor or bad.

Loudness Rating.

Table gives opinion results for various values of OLR in dB. These values are based upon representative laboratory conversation test results for telephone connections in which other characteristics such as circuit noise have little contribution to impairment.

Determination of Loudness Rating. The designation with notations of loudness rating concept for an international connection is given in Figure2. It is assumed that telephone sensitivity, both for the earpiece and microphone, have been measured. OLR is calculated using the following formula:

OLR = SLR + CLR+RLR. (2)

Table 3.1 Overall Loudness Rating Opinion Results

Overall Loudness Rating (dB)	Representative Opinion Results[a] Percent "Good plus Excellent"	Percent "Poor plus Bad"
5–15	<90	<1
20	80	4
25	65	10
30	45	20

[a]Based on opinion relationship derived from the transmission quality index (see Annex A, ITU-T Rec. P.11).
Source: ITU-T Rec. P.11, Table 1/P.11, p. 2, Helsinki, 3/93.

The measurement units in Eq. (2) are dB. OLR is defined as the loudness loss between the speaking subscriber's mouth and the listening subscriber's ear via a telephone connection.

The send loudness rating (SLR) is defined as the loudness loss between the speaking subscriber's mouth and an electrical interface in the network.

The receive loudness rating (RLR) is the loudness loss between an electrical interface in the network and the listening subscriber's ear.

The circuit loudness rating (CLR) is the loudness loss between two electrical interfaces in a connection or circuit, each interface terminated by its nominal impedance .

Data Circuits Bit error rate (BER) is the underlying QoS parameter for data circuits. BER is not subjective; it is readily measurable. Data users are very demanding of network operators regarding BER. If a network did not ever carry data, BER requirements could be much lessstringent.

CCITT/ITU-T recommends a BER of 1 × 10`6 for at least 80% of a month.1 Let us assume that these data will be transported on the digital network, typical of a PSTN. Let us further assume that conventional analog modems are not used, and the data is exchanged bit for bit with “channels” on the digital network.

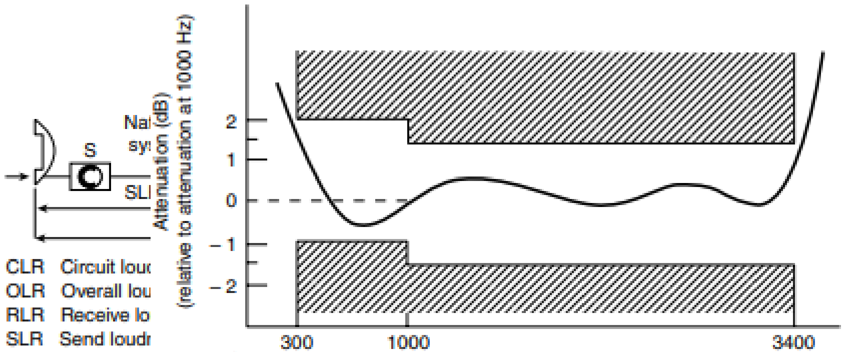


Figure 3.2 Designation of LRs in an international connection.

Video (Television) Television picture quality is subjective to the viewer. It is based on the S/N of the picture channel. The S/N values derived from two agencies are provided below. The TASO (Television Allocations Study Organization) ratings follow:

TASO Picture Rating

Quality	S/N
1. Excellent (no perceptible snow)	45 dB
2. Fine (snow just perceptible)	35 dB
3. Passable (snow definitely perceptible but not objectionable)	29 dB
4. Marginal (snow somewhat objectionable)	25 dB

Snow is the visual perception of high levels of thermal noise typical with poorer S/N values.

CCIR developed a five-point scale for picture quality versus impairment. This scale s shown in the table below:

CCIR Five Grade Scale

Quality	Impairment
5. Excellent	5. Imperceptible
4. Good	4. Perceptible, but not annoying
3. Fair	3. Slightly annoying
2. Poor	2. Annoying
1. Bad	1. Very annoying

Later CCIR/ITU-R documents steer clear of assigning S/N to such quality scales. In fact, when digital compression of TV is employed, the use of S/N to indicate picture quality is deprecated.

THREE BASIC IMPAIRMENTS AND HOW THEY AFFECT THE END-USER There are three basic impairments found in all telecommunication transmission systems. Theseare:

1. Amplitude (or attenuation)distortion;

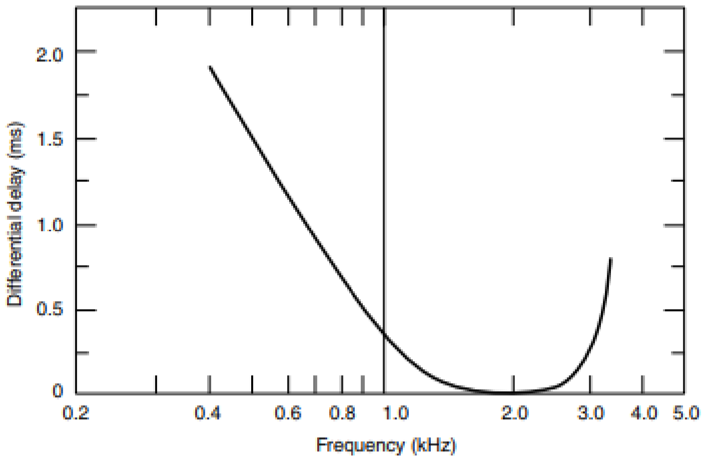
2. Phase distortion;and

3. Noise.

Amplitude Distortion The IEEE defines attenuation distortion (amplitude distortion) as the change in attenuation at any frequency with respect to that of a reference frequency. For the discussion, it is the measured response. This section, we’ll narrow the subject to the (analog) voice channel. In most cases a user is connected, through his/ her metallic subscriber loop, to the local serving exchange. This circuit is analog. Based upon the CCITT definition, the voice channel occupies the band from 300 Hz to 3400 Hz. We call this the passband. Attenuation distortion can be avoided if all frequencies within the passband are subjected to the same loss (or gain).

Whatever the transmission medium, however, some frequencies are attenuated more than others. Filters are employed in most active circuits (and in some passive circuits) and are major causes of attenuation distortion. Figure 3 is a response curve of a typical bandpass filter with voice channel application. For example, one European requirement may state that between 600 Hz and 2800 Hz the level will vary no more than -1 to +2 dB, where the plus sign means more loss and the minus sign means less loss.

Phase Distortion We can look at a voice channel as a band-pass filter. A signal takes a finite time to pass through the telecommunication network. This time is of propagation for the medium and, of course, the length of the medium. The value can vary from 10,000 mi/sec (16,000 k/sec) to 186,000 mi/sec (297,600 km/sec). The former value is for heavily loaded subscriber pair cable.4 This latter value is the velocity of propagation in free space, namely, radio propagation.



The velocity of propagation also tends to vary with frequency because of the electrical characteristics associated with the network. Again, the biggest offender is filters. Considering the voice channel, therefore, the velocity of propagation tends to increase toward band center and decrease toward band edge.

This is illustrated in Figure 4. The finite time it takes a signal to pass through the total extension of the voice channel or through any network is called delay. Absolute delay is the delay a signal experiences while passing through the channel end-to-end at a reference frequency. But we have learned that propagation time is different for different frequencies with the wavefront of one frequency arriving before the wavefront of another frequency in thepassband.

A modulated signal will not be distorted on passing through the channel if the phase shift changes uniformly with frequency, whereas if the phase shift is nonlinear with respect to frequency, the output signal is distorted with respect to frequency. In essence, we are dealing with phase linearity of a circuit. If the phase–frequency relationship over a passband is not linear, phase distortion will occur in the transmitted signal. Phase distortion is often measured by a parameter called envelope delay distortion (EDD).

Mathematically, EDD is the derivative of the phase shift with respect to frequency. The maximum variation in the envelope over a band of frequencies is called envelope delay distortion. Therefore EDD is always a difference between the envelope delay at one frequency and that at another frequency of interest in thepassband.

Figure 4 shows that absolute delay is minimum around 1700 Hz and 1800 Hz in the voice channel. The figure also shows that around 1700 Hz and 1800 Hz, envelope delay distortion is flattest.6 Itis for this reason that so many data modems use 1700 Hz or 1800 Hz for the characteristic tone frequency, which is modulated by the data. A data modem is a device that takes the raw electrical baseband data signal and makes it compatible for transmission over the voice channel. This brings up an important point. Phase distortion (or EDD) has little effect on speech communications over the telecommunications network. However, regarding data transmission, phase distortion is the greatest bottleneck for data rate (i.e., the number of bits per second that a channel can support). It has probably more effect on limiting data rate that any otherparameter.

Noise

Noise, in its broadest definition, consists of any undesired signal in a communication circuit. The subject of noise and noise reduction is probably the most important single consideration in transmission engineering. It is the major limiting factor in overall system performance. For our discussion in this text, noise is broken down into four categories:

1. Thermalnoise;
2. Intermodulationnoise;
3. Impulsenoise;
4. Crosstalk.

Thermal Noise. Thermal noise occurs in all transmission media and all communication equipment, including passive devices such as waveguide. It arises from random electron motion and is characterized by a uniform distribution of energy over the frequency spectrum with a Gaussian distribution of levels.

Gaussian distribution tells us that there is statistical randomness. For those of you who have studied statistics, this means that there is a “normal” distribution with standard deviations. Because of this, we can develop a mathematical relationship to calculate noise levels given certain key parameters. Every equipment element and the transmission medium itself contributes thermal noise to a communication system if the temperature of that element or medium is above absolute zero on the Kelvin temperature scale. Thermal noise is the factor that sets the lower limit of sensitivity of a receiving system and is often expressed as a temperature, usually given in units referred to absolute zero. These units are called kelvins

(K), not degrees.

Thermal noise is a general term referring to noise based on thermal agitations of electrons. The term “white noise” refers to the average uniform spectral distribution of noise energy with respect to frequency. Thermal noise is directly proportional to bandwidth and noise temperature.

Work of the Austrian scientist, Ludwig Boltzmann, who did landmark work on the random motion of electrons. From Boltzmann’s constant, we can write a relationship for the thermal noise level (Pn) in 1 Hz of bandwidth at absolute zero (Kelvin scale) or

Pn = -228.6 dBW per Hz of bandwidth for a perfect receiver at absolute zero. (3a)

At room temperature (290 K or 178C)

we have: Pn =-204 dBW per Hz of bandwidth for a perfect receiver. (3b)

or =-174 dBm/Hz of bandwidth for a perfect receiver.

A perfect receiver is a receiving device that contributes no thermal noise to the communication channel. Of course, this is an idealistic situation that cannot occur in real life. It does provide us a handy reference, though. The following relationship converts Eq. (3b) for a real receiver in a real-life setting.

Pn =-204 dBW/Hz + NFdB + 10 log B, (4)

where B is the bandwidth of the receiver in question. The bandwidth must always be in Hz or converted to Hz.

NF is the noise figure of the receiver. It is an artifice that we use to quantify the amount of thermal noise a receiver (or any other device) injects into a communication channel. The noise figure unit is thedB.

An example of application of Eq. (4) might be a receiver with a 3-dB noise figure and a 10-MHz bandwidth. What would be the thermal noise power (level) in dBW of the receiver? Use Eq. (4).

Pn = -204 dBW/Hz + 3 dB + 10 log(10 × 106)

= -204 dBW/Hz + 3 dB + 70 dB

=-131 dBW

Intermodulation Noise.

Intermodulation (IM) noise is the result of the presence of intermodulation products. If two signals with frequencies F1 and F2 are passed through a nonlinear device or medium, the result will contain IM products that are spurious frequency energy components. These components may be present either inside and/ or outside the frequency band of interest for a particular device or system. IM products may be produced from harmonics of the desired signal in question, either as products between harmonics, or as one of the basic signals and the harmonic of the other basic signal, or between both signals themselves. The products result when two (or more) signals beat together or “mix.” These products can be sums and/or differences. Look at the mixing possibilities when passing F1 and F2 through a nonlinear device. The coefficients indicate the first, second, or third harmonics.

• Second-order products F1 ±F2;

• Third-order products 2F1 ± F2; 2F2 ± F1;and

• Fourth-order products 2F1 ± 2F2; 3F1 ± F2 . . . .

Devices passing multiple signals simultaneously, such as multichannel radio equipment, develop IM products that are so varied that they resemble white noise. Intermodulation noise may result from a number of causes:

• Improper level setting. If the level of an input to a device is too high, the device is driven into its nonlinear operating region(overdrive).

• Improper alignment causing a device to functionnonlinearly.

• Nonlinear envelopedelay.

• Devicemalfunction.

To summarize, IM noise results from either a nonlinearity or a malfunction that has the effect of nonlinearity. The causes(s) of IM noise is (are) different from that of thermal noise. However, its detrimental effects and physical nature can be identical with those of thermal noise, particularly in multichannel systems carrying complex signals.

Impulse Noise. Impulse noise is noncontinuous, consisting of irregular pulses or noise spikes of short duration and of relatively high amplitude. These spikes are often called hits, and each spike has a broad spectral content (i.e., impulse noise smears a broad frequency bandwidth). Impulse noise degrades voice telephony usually only marginally, if at all. However, it may seriously degrade error performance on data or other digital circuits. The causes of impulse noise are lightning, car ignitions, mechanical switches (even light switches), flourescent lights, and soon.

Crosstalk. Crosstalk is the unwanted coupling between signal paths. There are essentially three causes of crosstalk:

1. Electrical coupling between transmission media, such as between wire pairs on a voice-frequency (VF) cable system and on digital (PCM) cablesystems;
2. Poor control of frequency response (i.e., defective filters or poor filter design);and
3. Nonlinear performance in analog frequency division multiplex (FDM)system.

Excessive level may exacerbate crosstalk. By “excessive level” we mean that the level or signal intensity has been adjusted to a point higher than it should be. In telephony and data systems, levels are commonly measured in dBm. In cable television systems levels are measured as voltages over a common impedance (75 Q).

There are two types of crosstalk:

1. **Intelligible**, where at least four words are intelligible to the listener from extraneous conversation(s) in a seven-second period;and

2. **Unintelligible**, crosstalk resulting from any other form of disturbing effects of one channel onanother.

Intelligible crosstalk presents the greatest impairment because of its distraction to the listener. Received crosstalk varies with the volume of the disturbing talker, the loss from the disturbing talker to the point of crosstalk, the coupling loss between the two circuits under consideration, and the loss from the point of crosstalk to thelistener.

LEVEL Level is an important parameter in the telecommunications network, particularly in the analog network or in the analog portion of a network. Level could be comparative. The output of an amplifier is 30 dB higher than the input. But more commonly, we mean absolute level, and in telephony it is measured in dBm (decibels referenced to 1 milliwatt) and in radio systems we are more apt to use dBW (decibels referenced to 1 watt). Television systems measure levels in voltage, commonly the dBmV (decibels referenced to 1 millivolt).

In the telecommunication network, if levels are too high, amplifiers become overloaded, resulting in increases in intermodulation noise and crosstalk. If levels are too low, customer satisfaction suffers (i.e., loudness rating). In the analog network, level was a major issue; in the digital network, somewhat less so.

System levels are used for engineering a communication system. On the chart, a 0 TLP (zero test level point) is established. A TLP is a location in a circuit or system at which a specified test-tone level is expected during alignment. A 0 TLP is a point at which the test-tone level should be 0 dBm. A test tone is a tone produced by an audio signal generator, usually 1020 Hz. Note that these frequencies are inside the standard voice channel which covers the range of 300 –3400 Hz. In the digital network, test tones must be applied on the analog side.

From the 0 TLP other points may be shown using the unit dBr (decibel reference). A minus sign shows that the level is so many decibels below reference and a plus sign, above. The unit dBm0 is an absolute unit of power in dBm referred to the 0 TLP. The dBm can be related to the dBr and dBm0 by the following formula:

dBm c dBm0 + dBr. (5)

For instance, a value of -32 dBm at a -22 dBr point corresponds to a reference level of -10 dBm0. A -10-dBm0 signal introduced at the 0-dBr point (0 TLP) has an absolute signal level of -10 dBm .

Typical Levels Earlier measurements of speech level used the unit of measure VU, standing for volume unit. For a 1000-Hz sinusoid signal (simple sine wave signal), 0 VU = 0 dBm. When a VU meter is used to measure the level of a voice signal, it is difficult to exactly equate VU and dBm. However, a good approximation relating VU to dBm is the following formula:

Average power of a telephone talker ≈ VU -1.4(dBm). (6)

Voice channel inputs were standardized with a level of either `15 dBm or `16 dBm, and the outputs of demultiplexers were +7 dBm. These levels, of course, were test-tone levels. In industrialized and post industrialized nations, in nearly every case, multiplexers are digital. These multiplexers have an overload point at about +3.17 dBm0. The digital reference signal is 0 dBm on the analog side using a standard test tone between 1013 Hz and 1022 Hz.

ECHO AND SINGING Echo and singing are two important impairments that impact QoS. Echo is when a talker hears her/ his own voice delayed. The annoyance is a function of the delay time (i.e., the time between the launching of a syllable by a talker and when the echo of that syllable is heard by the same talker). It is also a function of the intensity (level) of the echo, but to some lesser extent. Singing is audio feedback. It is an "ear-splitting" howl, much like the howl one gets by placing a public address microphone in front of a loudspeaker.

QoS Issues and Video over the Internet

IP/Internet Background Network communications can de-categorized into two basic types: circuit-switched (sometimes called connection-oriented) and packet/fastpacket-switched (these can be connectionless or connection-oriented)networks.

Circuit-switched networks operate by forming a dedicated connection (circuit) between two points. In packet-switched network, data to be transferred across a network is segmented into small blocks called packets (also known as data grams or protocol data units) that are multiplexed onto high capacity inter switch trunks. A packet, which usually contains few hundred bytes of data, carries routing information that enables the network hardware to know how to send it forward to the specified destination. In frame relay, the basic transfer unit is the (data link layer) frame; in cell relay this basic unit is the (data link layer)cell.

Internet Protocol Suite TCP/IP is the name for a family of over 100 data communications protocols used in the Internet and in intranets. In addition to the communication functions supported by TCP (end-to-end reliability over a connection-oriented session) and IP (subnetwork-level routing and forwarding in a connectionless manner), the other protocols in the suite support specific application-oriented tasks, e.g., transferring files between computers, sending mail, or logging in to a remote host. **The Internet** The same IP technology now used extensively in corporate intranets is used in (in fact, originated from) the Internet. The Internet is a collection of interconnected government, education, and business computer networks — in effect, a network of networks. Recently there has been a near-total commercialization of the Internet, allowing it to be used for pure business applications. (The original roots of the Internet were in the research and education arena.) Communications software in routers in the intervening networks between the source and destination networks "read" the addresses on packets moving through the Internet and forward the packets toward their destinations. TCP guarantees end-to-endintegrity.

Network interface layer	This layer is responsible of accepting and transmitting IP datagrams. This layer may consist of a device driver (e.g., when the network is a local network to which the machine attaches directly) or a complex subsystem that uses its own data link protocol.
Network layer (Internet layer)	This layer handles communication from one machine to the other. It accepts a request to send data from the transport layer along with the identification of the destination. It encapsulates the transport layer data unit in an IP datagram, uses the datagram routing algorithm to determine whether to send the datagram directly onto a router. The Internet layer also handles the incoming datagrams, and uses the routing algorithm todetermine whether the datagram is to be processed locally or is to be forwarded.
Transport layer	In this layer the software segments the stream of data being transmitted into small data units and passes each packet along with a destination address to the next layer for transmission. The software adds information to the packets including codes that identify which application program sent it, as well as a checksum. This layer also regulates the flow of information and providesreliable transport, ensuring that data arrives in sequence and with no errors
Application layer	At this level, users invoke application programs to access available services across the TCP/IP Internet. The application program chooses the kind of transport needed, which can either be messages or stream of bytes, and passes it to the transportlevel.

The term Internet is defined as “a mechanism for connecting or bridging different networks so that two communities can mutually interconnect.” The ARPA-developed technology included a set of network standards that specified the details of the computers that would be able to communicate, as well as a set of conventions for interconnecting networks and routing traffic. ARPA was also interested in integrated voice and data. While the ARPANet was growing into a national network, researchers at the Xerox Corporation Palo Alto Research Center were developing one of the technologies that would be used in local-area networking, namely, the ethernet. Ethernet became one of the important standards for how to implement building and campus data communicationsnetworks.

At about the same time, ARPA funded the integration of TCP/IP support into the version of the UNIX operating system that the University of California at Berkeley was developing.

Itfollows that when companies began marketing non-host-dependent workstations that ran

UNIX, TCP/IP was already built into the operating system software, and vendors such as Sun Microsystems included an ethernet port on the device. Consequently, TCP/IP over ethernet became a common way for workstations to interconnect. The same technology that made PCs and workstations possible made it possible for vendors to offer relatively inexpensive add-on cards to allow a variety of PCs to connect to ethernet LANs. Software vendors took the TCP/IP software from Berkeley UNIX and ported it to the PC, making it possible for PCs and UNIX machines to use the same protocol on the same network.

TCP and IP were developed for basic control of information delivery across the Internet. Application layer protocols, such as TELNET (Network Terminal), FTP (File Transfer Protocol), SMTP (Simple Mail Transfer Protocol), HTTP (HyperText Transfer Protocol), have been added to the TCP/IP suite of protocols to provide specific network services. Access and backbone speeds have increased from 56 kbps, to 1.5 Mbps (most common now), to 45 Mbps and beyond, for most of the backbones. Voice applications over IP have to ride over the Internet systems developed for traditional data services. Most problematic is the lack of QoS support; this, however, is expected to slowly change. Nonetheless, in spite of the emergence of new technologies, such as RSVP and RTP, a retarding factor to true QoS support is the very success of the Internet: the number of people using it is increasing at such a rapid rate that it is difficult to add enough resource and protocol improvements to keep up with the demand. Intranets use the same WWW/HTML/HTTP and TCP/IP technology used for the Internet. When the Internet caught on in the early-to-mid-1990s, planners were not looking at it as a way to run their businesses. But just as the action of putting millions of computers around the world on the same protocol suite fomented the Internet revolution, so connecting islands of information in a corporation via intranets is now sparking a corporate-based information revolution. Thousands of corporations now have intranets. Across the business world, employees from engineers to office workers are creating their own home pages and sharing details of their projects with the rest of thecompany.

QoS — Problems and Solutions Voice and video over IP is impacted by network congestion. QoS encompasses various levels of bandwidth reservation and traffic prioritization for multimedia and other bandwidth-intensive applications. The specific QoS solutions depend on the applications and circumstances at hand. QoS is generally not required for batch applications; it is needed for most if not all real-time applications. See Table.

For non-multimedia applications, QoS in enterprise networks is useful for allocating and prioritizing bandwidth to specific users.

For example, accounting departments may need more bandwidth when they are closing the books each month and a CEO needs more bandwidth during an extensive videoconferencing session. QoS is also important to supply streams of data that continuously move across the user's computer screen, such as stock tickers, real-time news, or viable data. Various QoS solutions for intranets and the Internet are available, beginning at the low end with more bandwidth to the LAN desktop via Layer 2 switching. New protocols and standards offer the next level of QoS for enterprise network environments, including 802.1p, 802.1q, and RSVP.48 Using ATM as a backbone improves bandwidth between subnetworks, and Layer

3 switching adds performance improvements in environments where IP dominates.

Finally, end-to-end ATM provides many levels of built-in QoS. Besides the capability for bandwidth reservation, QoS is affected by abilities of switches to perform real-time IP routing. Advances in silicon integration are being brought to bear for optimizing the performance of third-wave switches and paving the way for wire-speed IP routing capabilities.

	QoS Required	Applications
Non-real-time data	Little or none	Data file transfer, imaging, simulation, and modeling
Non-real-time multimedia	Little or none	Exchange text E-mail, exchange audio/video E-mail, Internet browsing with voice and video, intranetbrowsing with voice and video
Real-time one-way	Various QoS levels	Multimedia playback from server, broadcast video, distance learning, surveillance video, animation playback
Real-time interactive	Various medium or high QoSlevels	Videoconferencing, audioconferencing, process control

Third wave switches are optimized for switching at Gbps speeds. This is possible in high-performance custom ASICs that can process packets simultaneously and in real-time across multiple ports in a switch. Furthermore, the design of ultrawide data paths and multigigabit switching backplanes enable third-wave switches to perform at gigabit speeds through full-duplex connections on all ports without blocking.

Protocols for QoS Support for Audio and Video Applications

RSVP Applications RSVP, along with available network bandwidth, is required to ameliorate the overall quality in IP networks. New applications are now emerging that requires such capabilities.

For example, some companies are adding Web telephone access to their call centers, letting customers reach the carrier's customer service agent by clicking an icon at their Web site that reads "speak to the agent." But in order to scale this on a broad scale, standards are required so that QoS can be supported and made available as a network service. RSVP is based on receiver-controlled reservation requests for unicast or multicast communication. RSVP carries a specific QoS through the network, visiting each node the network uses to carry the stream. At each node, RSVP attempts to make a resource reservation for the stream. To make a resource reservation at a node, the RSVP daemon communicates with two local decision modules, admission **control and policy control**. Admission control determines whether the node has sufficient available resources to supply the requested QoS. Policy control determines whether the user has administrative permission to make the reservation. If either check fails, the RSVP program returns an error notification to the application process that originated the request. If both checks succeed, the RSVP daemon sets parameters in a packet classifier and packet scheduler to obtain the desired QoS. The packetclassifier determines the QoS class for each packet and the scheduler orders packet transmission to achieve the promised QoS for each stream (Figure).

A receiver-controlled reservation allows scaling of RSVP for large multicast groups. This support is based on the ability of RSVP to merge reservation requests as they progress up the multicast tree. The reservation for a single receiver does not need to travel to the source of a multicast tree; rather, it travels only until it reaches a reserved branch of the tree.

RSVP does not perform its own routing; instead, it uses underlying routing protocols. There is vendor interest in delivering RSVP on routers. A draft version of RSVP was approved by the IETF in 1996, and by 1997 vendors such as Cisco and Bay Networks was expressing interest, although they were being quoted as stating that "**there is little demand for RSVP from applications at themoment.**"

To ensure delivery through the network, RSVP allows listeners to request a specific QoS for a particular data flow. Listeners can specify how much bandwidth they will need and what maximum delay they can tolerate; internetworking devices then set aside the bandwidth for that flow. Users are either granted the channel they have requested or are given a "busy signal." RSVP hosts and networks interact to achieve a guaranteed end-to-end QoS transmission. All the hosts, routers, and other network infrastructure elements between the receiver and sender must support RSVP. They each reserve system resources such as bandwidth, CPU, and memory buffers to satisfy arequest.

RSVP operates on top of IP (either IPv4 or IPv6), occupying the place of a transport protocol in the protocol stack, but provides session-layer services (it does not transport any data). The RSVP protocol is used by routers to deliver control requests to all nodes along the paths of theflows.

Vendors have implemented RSVP both above and below Winsock. RSVP-aware applications can be developed with Winsock 2, which has a QoS-sensitive API. Another approach is to use an RSVP proxy that runs independently of the real application, making RSVP reservations.

RSVP raises questions about billing for Internet bandwidth. In the current model, ISPs oversell their available capacity, and customers accept slowdowns. Since resource reservation puts a specified demand on bandwidth, overselling would result in unacceptable performance (by the admission control module). ISPs will probably offer different service levels, and premiums will be charged for RSVP reservations. Billing across multiple carriers will also have to be resolved, as will the allocation of computational resources to routers to inspect and handle packets on a prioritized basis. It is unclear whether existing routers would be able to handle widescale implementation of RSVP across the whole Internet.

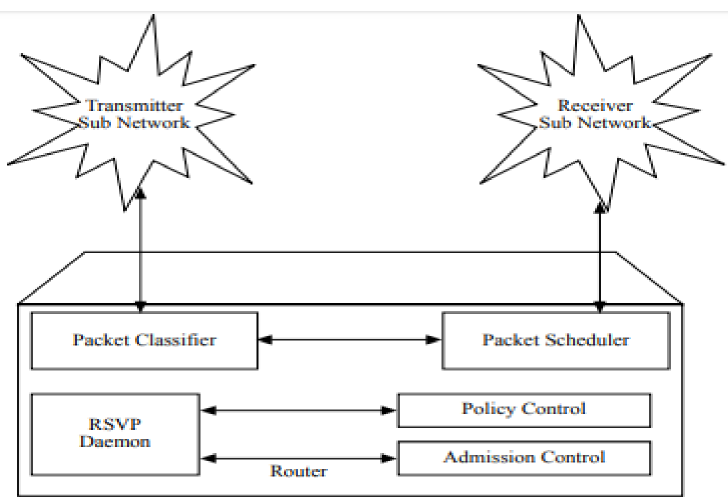
Developers now see the use of RSVP at the edges, as a signaling protocol; MPLS is likely to emerge in the core of a large IP-based network. Typically such network would be powered by gigarouters.

Real-Time Streaming Protocol Applications and Active Streaming Format

The Internet provided the impetus for the development of streaming technologies. The growth of realtime media on the Internet has stretched the HTTP capabilities used for downloading files to their maximum. The IETF is now attempting to standardize functions such as starting and stopping data streams, synchronizing multiple media elements, and implementing other controls. To this end, the main IETF work is embodied in RTSP, which was jointly proposed by Progressive Networks, Netscape, and Columbia University toward the end of 1996.

The latest version of RTSP essentially provides HTTP-level services to real-time streaming data types. However, RTSP differs from HTTP in that data delivery takes place out-of-band utilizing a distinct protocol. RTSP establishes and controls either single or several time- synchronized streams of continuous media. RTSP is expected to use TCP as the transport layer (for control only), but UDP may be optionally supported. Although in draft specification, RTSP implementations are availabletoday

In parallel with RTSP development, Microsoft has implemented a proprietary protocol called Active Streaming Format (ASF) in its new Netshow server platform. While it offers capabilities similar to RTSP, Microsoft documents refer to ASF as a “file format” and describe it as a component of the Microsoft overall ActiveX strategy. It is a kind of metafile that packages multiple “media objects” into a unified framework. Like RTSP, it may be used to synchronize a URLs, HTML pages, script commands, and executable programs. Unlike RTSP, an ASF stream includes both control and content elements. ASF content may either be constructed off line or captured in real time. This multimedia content is stored into ASF as



objects. These elements may be combined into a single ASF file.number of multimedia objects including audio, video, still images, events,ASF retains all forms of media (e.g., audio and video compression) and (optional) synchronization information so that when the file is played over a network, the user sees and hears the file exactly as the file creator intended.50 ASF data objects are stored within an ASF file as “packets.” Each packet is designed to be directly inserted “as is” into the data field of a data communication transport protocol. These packets are designed to be streamed across a network at a specific bit rate. The packet structure contains one or more payloads (i.e., distinct media streams) of data. Each packet may contain the data from a single media stream, or interleaved data from several media streams.

A "packet" is a collection of multimedia data that is ready to be streamed "as is" over the Internet/intranet. Ideally, the packet has been correctly sized so that all that needs to be done to ship it "over the wire" is to append the appropriate data communication protocol headers. ASF does not impose a packet size limitation; however, in practice, the packet sizes generally run from 512 bytes to the data communication's maximum transmissionunit (MTU)size.

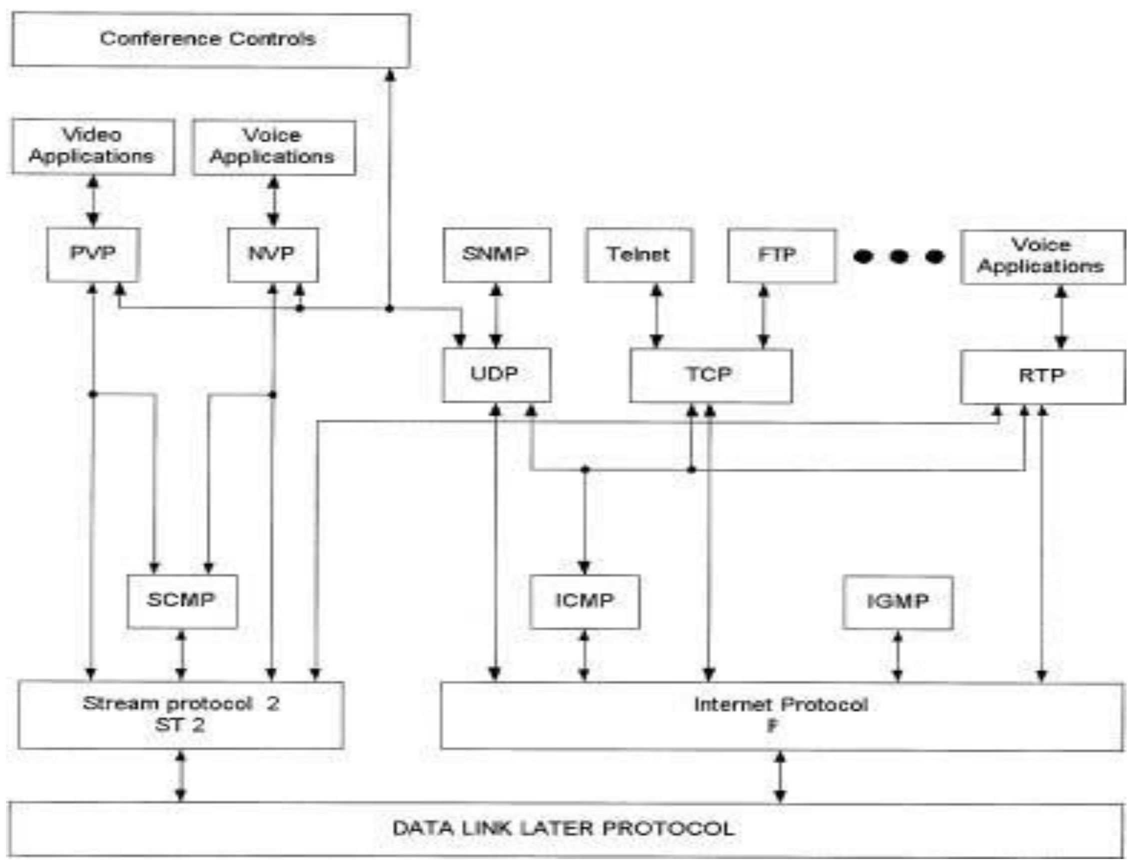
Each packet may contain interleaved data (i.e., composed of data from multiple multimedia streams). The format of the packet data is fairly complex in order to ensure that the packet data is as dense as possible for efficient transmission over a network.50 ASF data can be tailored to satisfy a variety of network requirements. For example, the data in each ASF file have been designed to stream at a distinct bit rate. The actual streaming bit rate is determined by the file's creator. The file's content creator has a range of streaming bit rates to choose between (e.g., 14.4 kbps to 6 Mbps).

ASF content can thus be flexibly targeted to specific network environments with distinct capacity requirements. Similarly, there are no data communications dependencies within ASF: ASF data can be carried over a wide variety of differing transport protocols. ASF multimedia streams can be stored on traditional file servers, HTTP servers, or specialized media servers, and can be transmitted efficiently over a variety of different network transports. These transports include TCP/IP, UDP/IP, RTP, and IPX/SPX. This data may be sent as either unicast- (point-to-point) or multicast streams.

Internet Stream Protocol Version 2 The first version of the Stream Protocol (ST) was published in the late 1970s and was used throughout the 1980s for experimental transmission of voice, video, and distributed simulation. The experience gained in these applications led to the development of the revised protocol version ST2. The revision extends the original protocol to make it more complete and more applicable to emerging multimedia environments. The specification of this protocol version is contained in RFC 1190 that was published in October 1990. With more and more developments of commercial distributed multimedia applications under way, and with a growing dissatisfaction at the network- secured QoS for audio and video over IP (particularly in the MBONE context), interest in ST2 has grown over the last few years. Companies have products available incorporating the protocol. Implementations of ST2 for Digital Equipment, IBM, NeXT, Macintosh, PC, Silicon Graphics, and Sun platforms are available.52 ST2 is an experimental resource reservation protocol intended to provide end- to-end real-time guarantees over the Internet or intranet. It allows applications to build multidestination simplex data streams with a desired QoS. The ST2 is an connection- oriented internetworking protocol that operates at the same layer as IP. It has been developed to support the efficient delivery of data streams to single or multiple destinations in applications that require guaranteed quality of service. ST2 is part of the IP protocol family and serves as an adjunct to, not a replacement for, IP. The main application areas of the protocol are the real-time transport of multimedia data, e.g., digital audio and video packet streams. ST2 can be used to reserve bandwidth for real-time streams. This reservation, together with appropriate network access and packet scheduling mechanismsin

all nodes running the protocol, guarantees a welldefinedQoS to ST2 applications.

It ensures that real-time packets are delivered within their performance targets, that is, at the time where they need to be presented. This facilitates a smooth delivery of data that is essential for time-critical applications, but cannot typically be provided by best-effort IP communication. ST2 consists of two protocols: ST (stream transport) for the data transport and SCMP (stream control message protocol), for all control functions. ST is simple and contains only a single PDU format that is designed for fast and efficient data forwarding in order to achieve low communication delays. SCMP packets are transferred within ST packets. For comparison, SCMP is more complex than ICMP. ST2 is designed to coexist with IP on each node. A typical distributed multimedia application would use both protocols: IP for the transfer of traditional data and control information, and ST2 for the transfer of real- time data. Whereas IP typically will be accessed from TCP or UDP, ST2 will be accessed via new end-to-end real-time protocols. The position of ST2 with respect to the other protocols of the Internet family is represented inFigure.



Both ST2 and IP apply the same addressing schemes to identify different hosts. ST2 and IP packets differ in the first four bits, which contain the internetwork protocol version number: number 5 is reserved for ST2 (IP itself has version number

4). As a network layer protocol, like IP, ST2 operates independently of its underlying subnets. As a special function, ST2 messages can be encapsulated in IP packets. This link allows ST2 messages to pass through routers, which do not run ST2. Resource management is typically not available for these IP route segments. IP encapsulation is, therefore, suggested only for portions of the network that do not constitute a system bottleneck.

In Figure, the RTP protocol is shown as an example of transport layer on top of ST2. Others include the packet video protocol (PVP), and the network voice protocol(NVP).

ST2 proposes a two-step communication model. In the first step, the real-time channels for the subsequent data transfer are built. This is called stream setup; it includes selecting the routes to the destinations and reserving the correspondent resources. In the second step, the data are transmitted over the previously established streams; this is called data transfer. While stream setup does not have to be completed in real time, data transfer has stringent real-time requirements. The architecture used to describe the ST2 communication model includes,

- Data transfer protocol for the transmission of real-time data over the establishedstreams
- Setup protocol to establish real-time streams based on the flowspecification
- Flow specification to express user real-timerequirements
- Routing function to select routes in theInternet
- Local resource manager to handle resources involved in the communicationappropriately

IP Multicast

For the Internet to be a viable real-time audio/video medium, it needs a method for serving a community of users. IP multicast is a suite of tools that addresses the bandwidth cost, availability, and service-quality problems facing real-time, large-scale Webcasting Rather than duplicating data, multicast sends the same information just once to multiple users. When a listener requests a stream, the Internet routers find the closest node that has the signal and replicates it, making the model scalable. IP multicast can run over just about any network that can carry IP, including ATM, frame relay, dial-up, and even satellite links. Originally developed in the late 1980s, it is now supported by virtually all major internetworking vendors, and its implementation and usage is picking up speed. Reliability is a challenge with multicast because there is not necessarily a bidirectional path from the server to the user to support retransmission of lost packets. A string of lost packets could create enough return traffic to negate multicast bandwidth savings. For this reason, TCP/IP cannot be used. Among the transport protocols developed for IP multicast, RTP and RTCP are the main ones for real- time multimedia delivery. RTP adds to each packet header the timing information necessary for data sequencing and synchronization. It does not provide mechanisms to ensure timely delivery or provide QoS guarantees; it does not guarantee delivery, nor does it assume that the underlying network is reliable. RTP and RTCP are currently in draft status; both were expected to be final in 1998. Uninterrupted audio requires a reliable transport layer; nevertheless, existing basic concealment techniques such as frequency domain repetition combined with packet interleaving work reasonably well if packet loss is minimal and occasional departures from perfection can be tolerated. One approach is to use FEC. Adding some redundant data improves performance considerably; combined with interleaving, this can be a good strategy, but it requires more bandwidth for a given quality level. This can be a challenge on a 28.8 kbps modemconnection.

Reliable multicast can be used to increase the performance of many applications that deliver information or live events to large numbers of users, such as financial data or video streaming. Reliable multicast creates higher-value application services for today's IP-based networks. According to a study recently conducted by the IP Multicast Initiative (IPMI), 54% of information systems managers stated that IP multicast had created new business opportunities for their companies and these numbers are likely to grow from year to year.

TEXT / REFERENCE BOOKS

1. Roger L. Freeman, "Fundamentals of Telecommunications", 4th Edition, John Wiley & Sons, 2010.
2. Jyrki T. J. Penttinen, "The Telecommunications Handbook: Engineering Guidelines for Fixed, Mobile and SatelliteSystems", 1st Edition, Wiley, 2015.
3. ThiagarajanViswanathan, ManavBhatnagar, "Telecommunication Switching Systems and Networks", 2nd Edition, PHIlearning,2015.

QUESTIONBANK:

PART – A

1. What are the basic impairments of telecommunication transmission systems?
2. Summarize the causes of cross talk.
3. State the term: “Quality of Service”.
4. Define Signal to Noise ratio. Mention its formula.
5. Differentiate between delay and absolute delay.
6. List out the factors responsible for overall loudness rating.

PART - B

1. Discuss about voice transmission in terms of loudness rating.
2. Categorize the types of noise in telecommunication systems.
3. Criticize the QoS issues in networks for video transmission.