LESSON 2:
EVOLUTION OF SWITCHING AND MESSAGE SWITCHING

Objective

To provide a detailed understanding of the concepts, principles, architectures, and issues associated with transmission, switching, and signaling systems, with emphasis on digital communication**.**

Introduction

The concept of the switchboard became prevalent in the 1880's, and as the system grew exponentially, the way calls were routed had to become increasingly automated. At first, there was no dial pad, no electronic switches, no signaling system. The caller would crank a handle on the phone to call the local switchboard operator who would connect the caller to the other party. If the call was long distance, the operator would have to contact another operator over an external line, then perhaps another, until the desired party was reached.

In the first year of service, the signaling system underwent a radical transformation. The first telephones did not have bells to signal an incoming call. Subscribers would have to shout in the mouthpiece or thump on it with a pencil, frequently breaking the diaphragm. J.C. Watson developed the "thumper", then the "magento-generator" which drove a "magneto-bell." The problem was that this bell would ring indiscriminantly each subscriber's phone in the system whenever a call was placed. The pressing need for an individual signaling system became quickly apparent and was solved within the first year of service.

The operator could easily listen in to any conversation, so the personal nature of their service made the operator the town message center. That is, at least until there were too many telephone subscribers to be handled efficiently by operators. The first telephone operators were undertrained, unsupervised telegraph boys but were soon replaced by young women who proved to be more pleasant, reliable and submissive.

The ever increasing number of callers soon transformed the profession. By 1946, nearly a quarter-million young women were employed by AT&T, but their numbers would soon dwindle due to the implimentation of automatic call switching. Almon Stowger in 1889 invented the automatic switch to connect two parties without the aid of an operator. An undertaker in Kansas City, Stowger suspected that local operators were routing the calls to his rivals, so he wanted to cut the women out of the calling process. The rotary dial - which routes a call via a set of switches activated by a series of pulse signals - was developed in 1900 but was not installed on a large scale until 1914 in Newark, New Jersey. In 1921, Omaha, Nebraska opened the first all-automatic exchange. By 1926, only 20 percent of the system used the rotary dial and automatic switches. Ironically, because other countries were slower to install their telephone systems, they were able to install automatic switchboards earlier than AT&T. Direct dial long distance took a little longer to develop. The first computerized switchboard was put into commercial service in 1976, and by 1982, half of all calls were switched electronically. The electronic switchboards used tones to route calls. This was the origin of touch tone keypad.

At the turn of the century, when Bell's 17 year patent protection ran out, around 6000 independent phone companies sprang up. These local switchboards would connect a few hundred households but could not be linked to a national system. AT&T would refuse to link these companies into their network monopoly and would buy them out when they were on the verge of bankruptcy.

AT&T, bought out in 1907 by New York Bankers like JP Morgan, had effectively overwhelmed the independent telephone companies and had over 7 million phones in their telephone system. By 1930, there were 20 million subscribers. As technology improved and the system became immense, calling costs went down. In 1927, a three minute call from New York to London cost $75, took at least eight operators and fourteen minutes to effect the connection. In 1945, the same call would cost $12 and take only 90 seconds to effectuate, due to improvements in technology. This trend has been ongoing and now long distance calls cost a tiny fraction of what it once did.

Many people were rather concerned about AT&T's monopoly. AT&T agreed to Federal Communication Commission regulation in 1956 in exchange for maintaining the status quo and promising not to enter the emerging computer industry. The AT&T monopoly lasted until a ten year lawsuit to break up the telecommunication giant was settled in 1984. AT&T persuaded the courts to let it get out of the local phone service industry in favor of joining the deregulated long distance competition, where it believed the real money was. AT&T got out of the local phones service and seven "baby" bells were formed. The dissolution of AT&T also meant that people could now own their telephones and freely hook up peripheral devices such as faxes and modems.

Switching System

Consequently, practical use of Bell's invention on a large scale or even on a moderate scale demanded not only the telephone sets and the pairs of wires, but also the switching system. Switching system is also called the switching office or the exchange. With the introduction of switching systems, the subscribers are not connected directly to one another. But the subscribers are connected to the switching system shown in. Fig2.1.

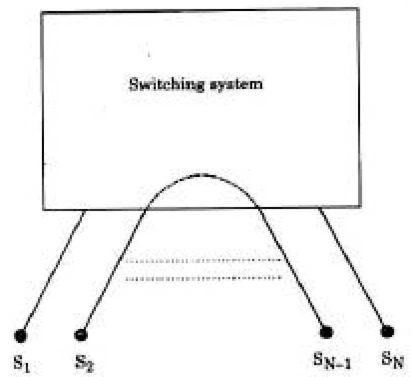


Fig.2.1 Illustration of subscriber interconnection using a switching system

When a subscriber wants to communicate with another, a connection is established between the two at the switching system. A connection between subscriber S_2 and S_{N-1} is shown in Fig. 2.1. In this configuration, only one link per subscriber is required between the subscriber and the switching system. In this configuration, the total number of such links is equal to the number of subscribers connected to the system. Signalling is now required to draw the attention of the switching system to establish or release a connection. Signalling also enables the switching system to detect whether a called subscriber in busy and, if so, indicate the same to the calling subscriber. The functions performed by a switching system in establishing and releasing connections are known as control functions.

Early switching systems were manual and operator oriented. Limitations of manual switching systems were quickly recognised and automatic exchanges came into existence.

Automatic switching systems are classified as electromechanical and electronic. Electromechanical automatic switching systems include step-by-step and crossbar systems. The step-by-step system is popularly known as Strowger switching system after its inventor A.B. Strowger. The control functions in a Strowger switching system are performed by circuits associated with the switching elements in the system.

Crossbar switching systems have hard-wired control sub-systems, which use relays and latches. These subsystems have limited capability and it is virtually impossible to modify them to provide additional functionalities.

In electronic switching systems, the control functions are performed by a computer or a processor. Hence, these switching systems are also called stored program control (SPC) systems. We can easily add new facilities to a SPC system -by changing the control program. The switching scheme used by electronic switching systems may be either Space Divisions Switching (SDS) or Time Division Switching (TDS).

In Space Division Switching, a dedicated path is established between the calling and called subscribers for entire duration of the call. Space Division Switching is also the technique used in Strowger and crossbar systems.

An electronic exchange may use a crossbar-switching matrix for space division switching.

In time division switching, the sampled values of speech signals are transferred at fixed intervals. Time Division Switching may be analog or digital.

In analog time division switching, the sampled voltage levels are transmitted as they are, whereas in digital time division switching, they are binary coded and transmitted.

If the coded values are transferred during the same interval from input to output, the technique is known as Space Switching. If the values are stored and transferred to the output at a later time interval, the technique is known as Time Switching. A time-division digital switch may be designed by using a combination of space and time switching techniques.

Classification of switching systems are summarized in Fig.2.2

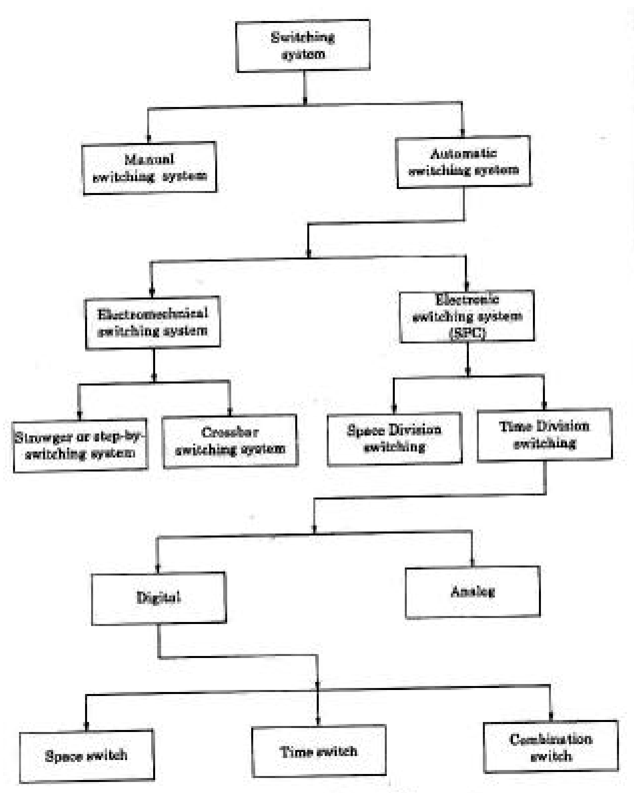


Fig. 2.2 Classification of switching system

Fundamentals of switching system

A major component of a switching system (or an exchange) is the set of input and output circuits. These set of input and output circuits are called inlets and outlets, respectively. The primary function of the switching system is to establish an electrical path between a given inlet-outlet pair switching matrix or switching network is a hardware that is used to establish a connection between a given inlet-outlet pair. It is important to note that the switching network is a component of the switching system and should not be confused with telecommunication network. A model of a switching network with *N* inlets and *M* outlets in shown in Fig. 2.3 (a). When $M = N$, the switching network is called a symmetric network. The inlets/ outlets may be connected to local subscriber lines or to trunk from/ to other exchanges. It is shown in Fig. 2.3 (b). When all the inlets/ outlets are connected to the subscriber lines, then logical connection appears as shown in Fig. 2.3 (c).

In this case, output lines are folded back to the input and hence the network is called a folded network.

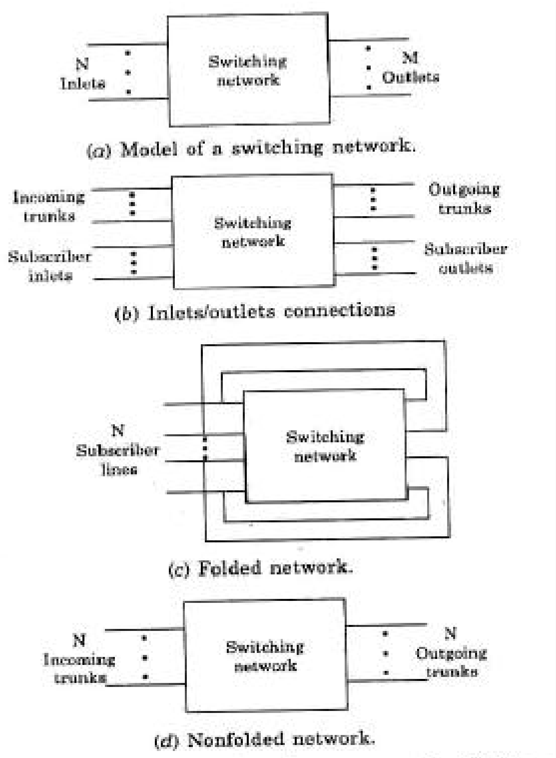


Fig.2.3 illustration of various types of switching network configurations.

There are four types of connections which are possible in inlets / outlet connection.

1. Local call connection between two subscribers in the system.
2. Outgoing call connection between a subscriber and an outgoing trunk.
3. Incoming call connection between an incoming trunk and a local subscriber.
4. Transit calls connection between an incoming trunk and an outgoing trunk.

In a folded switching network with N subscribers, there can be a maximum of (N / 2) simultaneous calls or information interchanges.

The switching network may be designed to provide (N / 2) simultaneous switching paths. In this case, it is called non-blocking switching network. In a non-blocking switching network, as long as a called subscriber will always be able to establish a connection to the called subscriber. In other words, a subscriber will not be denied a connection for want of switching resources.

But, in general, it rarely happens that all the possible conversations take place simultaneously. It may, hence, be economical to design a switching network that has as many simultaneous switching paths as the average number of conversations expected. In this case, it may happen occasionally that when a subscriber requests a connection, there are no switching paths free in the network, and hence he is denied connection. In such a connection the subscriber is called blocked and the switching network is called a blocking network. In a blocking network, the number of simultaneous switching paths is less than the maximum number of simultaneous conversations that can take place. The probability that a user may get blocked is called blocking probability.

All the switching systems or exchanges are designed to meet an estimated maximum average simultaneous traffic. This traffic is known as Busy hour traffic. Past records of the telephone traffic show that even in a busy exchange, not more than 20 to 30% of the subscribers are active at the same time.

Hence, switching systems are designed such that all the resources in a system are treated as common resources and the required resources are allocated to conservation as long as it lasts. The quantum of common resources is determined based on the estimated busy hour traffic.

When the traffic exceeds the limit to which the switching system is designed, a subscriber suffers blocking. A good design of switching system generally ensures a low blocking probability.

The traffic in a telecommunication network is determined or measured by an internationally accepted unit of traffic intensity. It is known as erlang (E). It is named after an early contributor to traffic theory in telecommunication. A switching resource is said to carry one erlang of traffic if it is continuously occupied throughout a given period of observation.

In a switching network, all the inlet/outlet connections can be used for inter exchange transmission. In such a case, the exchange does not support local subscribers. It is called a transit exchange. A switching network of this kind is shown in Fig. 2.3 (d). It is also known as a non-folded network. In non-folded switching network with N inlets and N outlets, N simultaneous information transfers are possible. Consequently, for a non-folded network to be non- blocking, the network should support N simultaneous switching paths.

While the switching network provides the switching paths, it is the control subsystem of the switching system that actually establishes the path. The switching network does not distinguish between inlets / outlets that are connected to the subscribers or to the trunks. Control subsystem distinguishes these lines and interprets correctly the signalling information received on these lines. It senses the end of information transfer and releases connections. A connection is established on the basis of signalling information received on the inlet lines. The control subsystem sends out signalling information to the subscriber and other exchanges connected to the outgoing trunks. In addition, signalling is also needed between different subsystems within an exchange. The signalling formats and the requirements for the subscriber, the trunks and the subsystems differ significantly.

Three different forms of signalling are provided by a switching system.

1. Subscriber loop signalling
2. Inter-exchange signalling
3. Intra-exchange or register signalling.

A switching system is composed of three types of elements that perform switching, control and signalling functions. Different elements of a switching system and their logical connections are shown in Fig. 2.4. The subscriber lines are terminated at the subscriber line interface circuits. But trunks are terminated at the trunks are terminated at the trunk interface circuits. There are some service lines used for maintenance and testing purposes. Junctor circuits are used for a folded connection for the local subscribers and the service circuits. It is possible that some switching system provide on internal mechanism for local connections without using the junctor circuits.

Line scanning units sense and obtain signalling information from the respective lines. Distributor units send out signalling information on the respective lines.

Operators console permits interaction with the switching system for maintenance and administrative purposes. In some switching systems, the control subsystem may be an integral part of the switching network itself. These systems are called Direct Control Switching Systems. Switching systems in which the control subsystem is outside the switching network are called Common Control Switching Systems. Strowger exchanges are usually direct control switching systems. But the crossbar and electronic exchanges are common control switching systems. Common control is also known as Indirect Control or Register Control.

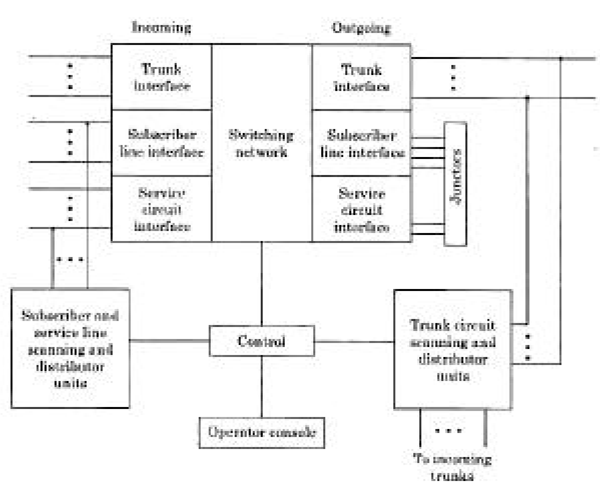


Fig.2.4 Illustration of elements of a switching system

Functions of a switching system

The basic functions that a switching system must perform are as follows:

1. Attending
2. Information receiving
3. Information processing
4. Busy testing
5. Interconnection
6. Altering
7. Supervision
8. Information sending.

Now we discuss each function one by one.

i) Attending: - The system must be continually monitoring all lines to detect call requests. The calling signal is sometimes known as a "seize" signal because it obtains a resource from the exchange.

ii) Information Receiving: - In addition to receiving call and clear signals system must receive information from the caller as to the called line required, this is called address signal.

iii) Information Processing: - The switching system must process the information received. Both originating and terminating calls are handled differently for different customers; class of information service must be processed in addition to the address information.

iv) Busy Testing: - Having processed the receiving information to determine the required outgoing circuit the system must make a busy test to determine whether it is free or already engaged on another call. If a call is to customer with a group of lines to a PBX (private branch exchange) or to an outgoing junction route each line in the group is tested until a free one found.

v) Interconnection: - For a call between two customers, their connections are made in the following sequence:

a) A connection to the calling terminal.

b) A connection to the called terminal.

c) A connection between the two terminals.

vi) Alerting: - Having made the connection, the system sends s signal to alert the called customer to the call, i.e. ringing current to a customer's telephone.

vii) Supervision: - After the called terminal has answered, the switching system continues to monitor the connection in order to be able to clear it down when the call has ended. When charge for the call is made by metering, the supervision circuit sends pulses over the P wire to operate a meter in the line circuit of the calling customer. When the automatic ticketing is employed, the switching system must send the number of caller to the supervisory circuit when the connection is set up. This process known as Calling Line Identification (CLI), or Automatic Number Identification (ANI).

viii) Information Sending: - If the called customer's line is located on another exchange the additional function of information sending is required. The originating exchange must signal the required address to the terminating exchange.

Switching System	Operation	Method of Switching	Type of Control	Type of Network
1878 manual operator	manual	space/analog	human	plug/cord/jack
1892 step-by-step	electromechanical	space/analog	distributed stage-by-stage	stepping switch train
1918 cross-bar	electromechanical	space/analog	common control	X-bar switch
1960 ESS—first generation	semielectronic	space/analog	common control	reed switch
1972 ESS—second generation	semielectronic	space/analog	stored program control	reed switch
1976 ESS—third generation	electronic	time/digital	stored program common control	pulse code modulation

Table 1. Types of End-Office Switching and Their Evolution

In the following paragraphs we get into the details of different switching techniques.

Message Switching

In the early days of telegraphy a customer might wish to send message from town A to town B although there was no telegraph circuit between these towns. However if there was a circuit between A and C and one between C and B. This could be achieved by the process known as message switching. The operator at A sends the message to C, where it was written down by the receiving operator. This operator recognized the address of the message as being at B and then retransmitted the message over the circuit to B.

Subsequent technical developments enabled improvements to be made in message switching. The first message received at C was automatically recorded on punched tape and subsequently toned off the receiver by the operator who read the address from the tape. The message was then transmitted automatically from the same tape, this was known as tone – tape relay system. Later the outgoing route was also selected automatically. Finally the paper tape was eliminated by storing the messages electronically and analyzing their address by electronic logic. A modern message-switching center is thus a special purpose electronic computer. In a message switching center an incoming message is not lost when the required outgoing route is busy. It is stored in a queue. Message switching is still used for telegraph traffic and a modified form of it is known as packet switching, used in data communication.

" Practice of transporting complete messages from a source to a destination in non-real time and without interaction between source and desitination, usually in a store-and-forward fashion is called as Message Switching"

Store and Forward Switching.

A Store and Forward (S and F) network configuration is shown in

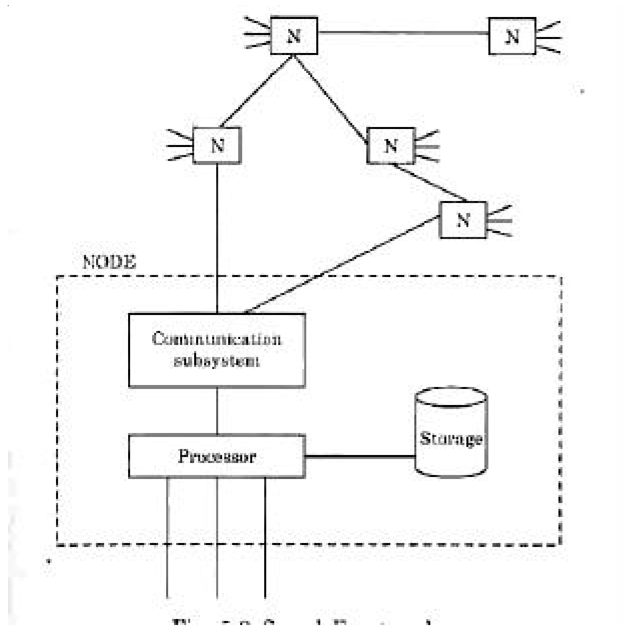


Fig.2.5 S And F network

In Sand F switching, the switching nodes have the ability to store user messages and forward the same towards the destination as and when the links become available. For this purpose, each node is equipped with a processor and some buffer storage. No end· to-end link is set up prior to data transmission. The user deposits his/her message to the nearest switching node and then on, network takes the responsibility for delivering the message to the destination user or host. S and F switching functions in a fashion that is analogous to postal or telegraphy system.

The network moves the user information from node to node. One such movement is called hop. Since communication links are used one at a time between any two nodes, line speeds can be utilized efficiently.

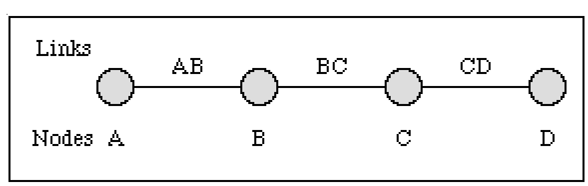
S and F Switching is of two types:

(a) Message Switching

(b) Packet Switching.

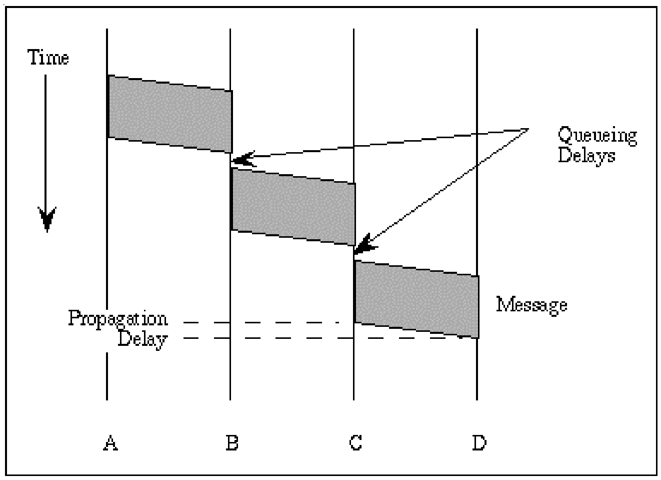
Message Switching

Sometimes there is no need for a circuit to be established all the way from the source to the destination. Consider a connection between the users (A and D) in the figure below (i.e. A and D) is represented by a series of links (AB, BC, and CD).



A connection between two systems A & D formed from 3 links

For instance, when a telex (or email) message is sent from A to D, it first passes over a local connection (AB). It is then passed at some later time to C (via link BC), and from there to the destination (via link CD). At each message switch, the received message is stored, and a connection is subsequently made to deliver the message to the neighboring message switch. Message switching is also known as store-and-forward switching since the messages are stored at intermediate nodes en route to their destinations.

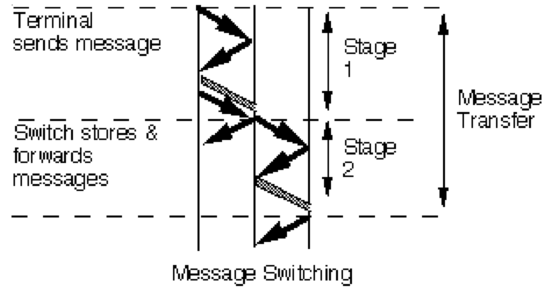


The use of message switching to communicate between A and D

The figure illustrates message switching; transmission of only one message is illustrated for simplicity. As the figure indicates, a complete message is sent from node A to node B when the link interconnecting them becomes available. Since the message may be competing with other messages for access to facilities, a queuing delay may be incurred while waiting for the link to become available. The message is stored at B until the next link becomes available, with another queuing delay before it can be forwarded. It repeats this process until it reaches its destination.

Circuit setup delays are replaced by queuing delays. Considerable extra delay may result from storage at individual nodes. A delay for putting the message on the communications link (message length in bits divided by link speed in bps) is also incurred at each node en route. Message lengths are slightly longer than they are in circuit switching, after establishment of the circuit, since header information must be included with each message; the header includes information identifying the destination as well as other types of information.

Most message switched networks do not use dedicated point-to-point links and therefore a call must be set-up using a circuit switched network. The figure below illustrates the use of message switching over a circuit switched network, in this case using one intermediate message switch.



Message switching using circuit switched connections between message switches.

Although message switching is still used for electronic mail and telex transmission, it has largely been replaced by packet switching (in fact, most electronic mail is carried using message switching with the links between message switches provided by packet or circuit-switched networks).

In message switching, once the transmission is initiated, a message is transmitted in its entirety without a break from one node to another.

Here, node processor performs following functions:

1. It receives the full user message and store the same
2. Check the entire message for data transmission errors and perform error recovery if required.
3. Determine the destination address from the user message
4. Choose an appropriate link towards destination based on certain routing criterion.
5. Forward the message to the next node on the chosen link.

Message Switching has some drawbacks

1. For long messages, it becomes important to ensure that there is adequate storage space on the receiving node' before the transmission is initiated. Otherwise, the buffer storage may become full, and part of the message may not be stored. Thus it requires retransmission of the entire message.
2. If an error occurs during transmission, the entire message may have to be retransmitted. Retransmission of long messages results a large communication overheads in the network.

If a high priority short message arrives while a long message is in transmission, it wi1l have to wait until the transmission of the long message ends. Such drawbacks are overcome in Packet Switching.

Packet Switching

In this Switching, messages are split into a number of packets. Often packets are fixed in size. These packets are transmitted in a stored and forward (S and F) fashion. Messages are split at the source host and reassembled at the destination host. Each packet transmission is independent of the others. The packets of a single message may travel via different routes and arrive at the destination with different transmission delays. This leads to a situation where the packets of the same message arrive out of sequence at the destination node. Every packet needs to carry the complete address information, wiz destination identifier (id), source id, message id, and packet id, in addition to the actual user data. A typical packet format is shown in Fig 2.6 (a).

Schematic diagram of a packet switching is shown in Fig. 2.6 (b).

In this schematic diagram, a packet is numbered using a four subscript quantity where the first subscript is the destination host id, the second the source host id, the third message id, and the fourth the packet id of the message. For example, P_{7543} indicates that this is the third packet of the fourth message originating from the host 5 and destined to host 7. It can be represented as,

P [Destination Host id] [Source Host id][Message id] [Packet id]

We can observe that the source host delivers the packets of a message in sequence to the network node and it is natural to expect that the packets are delivered to the destination host in proper sequence. However, as the packets may arrive out of sequence at the destination node, it becomes the responsibility of the network to re sequence the· packets before delivery to the destination host.

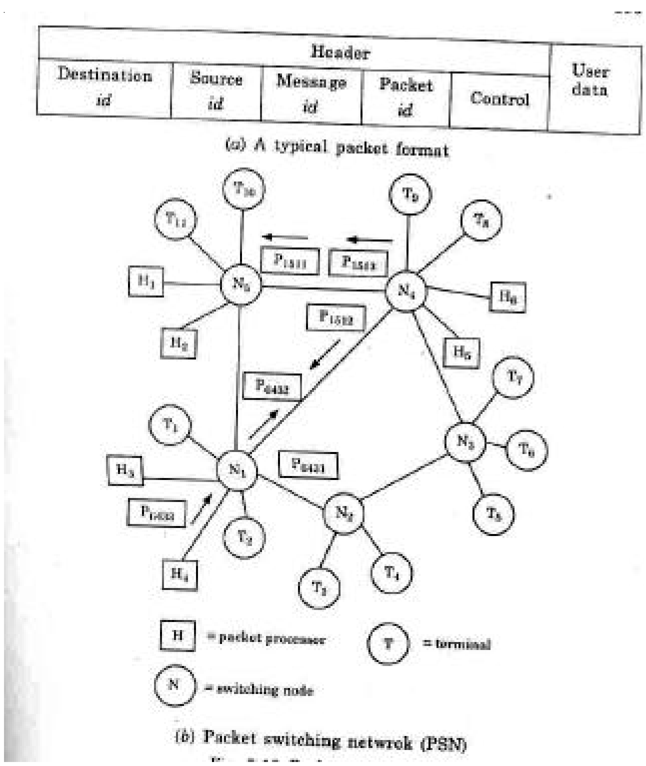


Fig.2.6 Packet Switching

The resequencing of the packets before delivery to the destination calls for considerable overhead in terms of buffer storage and processing power at the network nodes. Hence, turns out to be somewhat an expensive service.

On the basis of cost effectiveness of the packet switching, the packet networks offer two different forms of services:

(i) Datagram Service

(ii) Virtual Circuit Service

In these services, no re sequencing of packets is required Datagram Service is normally used for transmitting short messages of one or two packet lengths.

In the case of virtual circuit service, the route from the source to destination is fixed for all packets of a message, since the packets in virtual-circuit service are delivered to the network in sequence and they follow the same route. On a First-Come-First-Serve (FCFS) basis, they arrive in sequence at the destination. However, the transmission of packets may not start until and unless a route/circuit is chosen and finalised between the source and the destination. In virtual-circuit-service, no dedicated circuit path is used for a particular connection, as the same route and the circuit may be used for transmitting packets from different sources. Hence, the term ‘Virtual Circuit’ is used to denote such connections. In virtual circuit connection, the packets of a message need not carry the full address information as the packets follow the same route and are delivered in sequence. As soon as a virtual circuit is established between a communicating pair, the same is given an identifier.

The identifier is sufficient as address in the packet. Optionally, a user may want to use the same virtual circuit for transmitting a number of messages over a period of time. The circuit established for this purpose is called permanent virtual circuit. In this case, the user needs to identify the different messages suitably.

Delays in store and forward (S & F) networks arise from two factors:

1. Storage Delay (Ts)
2. Forwarding Delay (Tf)

Hence, the (S & F) delay is the sum of storage delay and forwarding delay,

$$T_{sf} = T_s + T_f$$

A node at any time may have a number of packets to be sent on an outgoing line. These packets are queued and transmitted according to some queuing rule or discipline (FCFS or priority etc.). The data forwarding delay T_f is largely due to this queuing delay. A small component comes from the processing which needs to be performed by the node processors for routing etc.

T_f can be expressed as,

$$T_f = (N - 1)(T_q + T_m) + T_t$$

Where,

T_q = Average queuing delay in each node

T_m = Processing delay in each node

T_t = Message or packet transmission delay

N = Number of nodes involved in the transfer.

We assume that the destination node does not contribute to any delay the destination host is directly attached to destination node and no other network traffic is on this link.

If the transfer rate R is uniform on all the links, then the data transmission delay T_t is given by,

$$T_t = (N + 1)(M/R) + T_p$$

There are (N + 1) data links or hops from the source to destination.

M = Message length in case of message switching or packet size in case of packet switching.

T_p = Total propagation delay from source to destination.

If the data transfer rates are non-uniform, Tt is given by

$$T_t = M\{ 1/R_1 + 1/R_2 + \ldots\ldots\ldots + 1/R_{N+1}\}$$

Queuing delay and in effect the data forwarding delay T_f is a function of load on the network. T_f increases and becomes asymptotic when the network is overloaded.

On the other hand, the data transmission delay remains almost constant irrespective of the load. Variations of delay components in S&F switching is shown in fig.2.7

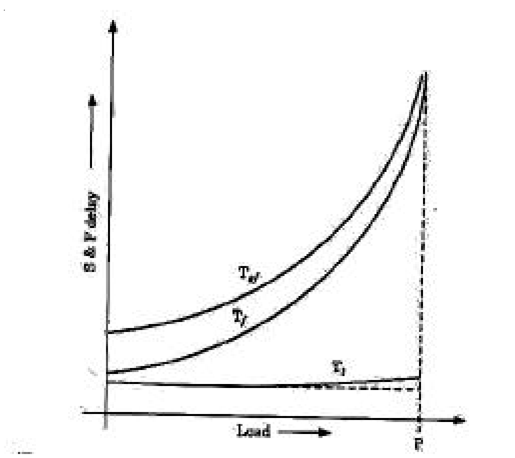


Fig.2.7 Variation of delay components in S & F switching

Notes.