UNIT 1 INTRODUCTION TO LAYER FUNCTIONALITY AND DESIGN ISSUES

1.0 INTRODUCTION

This chapter discusses about the network layer, which is the third layer of the OSI model. Job of this layer is to send the packets from a source to destination. This layer responds to the service requests of the transport layer and takes the services from the data link layer. This chapter starts with an overview of the services of the network layer. Switching is the backbone of network architecture. This important concept of packet switching is elaborated with diagrams in section 1.3. How the address will be assigned to a host and the different concepts of addressing are discussed in further section. This is followed by congestion and routing concepts. Delay is an important concept in packet switched networks. The various types of delay have been discussed. The delay computation in different scenarios is illustrated with various examples in further section.

1.1 OBJECTIVES

After completing this unit, a student will be able to

- Explain the basic concepts and services of network layer.
- define the concepts of packet switching
- Differentiatebetween virtual circuit and datagram approach.

- Elaborate and utilize the concepts of addressing
- define congestion and policies to overcome the congestion in the network layer
- explain the concept of routing;
- calculate the delay in a given network scenario

1.2 SERVICES OF THE NETWORK LAYER

The third layer of the protocol stack is the network Layer. This layer is responsible for delivering the data from source machine to the destination machine that is end to end communication. At the source machine, it takes the services from the transport layer. Whereas on the destination side, the network layer provides the services to the upper layer that is transport layer. The important services provided by the network layer are

a) **Routing** – As discussed above, this layer is responsible for machine to machine communication. Thus, this layer decides the route that a packet has to follow from source to destination. There could be various possible paths from a source to a destination. Based on the chosen metric like delay, number of hops, a particular path would be selected as the best route.

b) **Packetization**–At the source machine, network layer receives the segment from the transport layer and send further to the data link layer. The received segment needs to be divided further into small packets or send as a whole packet, this decision is to be taken by the network layer by visualizing the maximum transmission unit of the data link layer. Control information i.e. header is to be added at the sending side so at the receiving side, packet is to be reassembled or decapsulated correctly.

c) **Forwarding** – when a packet is send from a source to a destination, this packet pass through a number of routers along the path. A router has a number of interfaces. Which interface is to be selected for the packet is decided by the network layer.

Let us understand the clear distinction between routing and forwarding with an example. We are planning a drive from JIIT, Noida to IGNOU. There are various possible paths like one is via GT road, another is via Indirapuram and so- on. Which path is the best one as per the time taken or road conditions? This decision process is routing and here our metric to decide the best route could be any one like traffic conditions on the road, infrastructure of the road, etc. Suppose the selected route is via Indirapuram, and the person started the journey. At one of the intermediate junctions, there are various directions. Which direction to be chosen at the junction is the forwarding decision taken by the router.

1.3 PACKET SWITCHING

There are two switching mechanisms that work in the backbone of the network, circuit switching and packet switching. In today's Internet, packet switching is utilized where

telephone networks is an example which best describes the concept of circuit switching. In circuit switching, the resources are reserved for a user where as in packet switching, the resources are shared among different users on demand basis. Network layer utilizes the packet switching as packet is the basic data unit used at this layer.

Let us understand the concept of packet switching more clearly with the following scenario. Consider two banks where bank 1 requirement is book an appointment before coming to the bank. If you reach directly, you would not be entertained.If already booked an appointment, your waiting time is negligible.There is no such requirement for 2nd bank. As soon as you reach to the bank, you will be entertained based on the number of people already waiting. The services will be provided to you without any hassle, if no one is there. If already a large number of people are waiting, then your waiting time would be large or in some situations, bank will say it's already full, kindly come on next day. But on the other hand, there is no hassle of calling before leaving from home. The scenario of 2nd bank describes how packets will be handled during packet switching.

Network layer receives the data from the transport layer and divide into manageable units known as packets. Based on different forwarding mechanisms used by connected devices to forward the packets from a given source to a particular destination, packet switched networksare further divided into two categories: virtual circuit approach and datagram approach.

1.3.1 Virtual Circuit Approach (Connection-oriented Service)

Before going into the detail of virtual circuit approach, first let us understand the meaning of **connection oriented service**. Connection oriented service in which an end to end logical connection would be established between the source machine and destination machine. All the data between a source destination pair would be sent through the same connection. After sending the data, connection would be terminated. A connection oriented service has the following properties

a) All data would be sent in order and without any error to the destination machine.
b) All the received data would be acknowledged by the destination machine.
c) The underlying service guarantees the in-order delivery of packets without any loss or duplication of packets.
d) There is a retransmission policy which will handle the lost packets.

Due to all these properties, connection oriented service is also known as **reliable** service. A connection oriented service is a three step process which are described as follows

a) **Connection establishment:** This is a handshaking process which needs to be executed before any data exchange among two entities. Suppose, person A wants to talk to other unknown person B. Before any informal or formal talk, they will exchange formal hello messages. Similarly, here in connection oriented service, source machine will send the connection request message (control message) to the intended destination machine. In receipt of this,

destination machine will send an acknowledgement message to the source machine. Source machine will send a confirmation of the received acknowledgment message. Purpose of this 3 step control messages exchange is to prepare both the entities for handling the data transfer further.

b) **Data transfer:**Once the connection gets establishment, data could be transferred among the two entities.

c) **Connection termination:** Once the data transfer is over, a connection termination request will be send by the source machine to the destination machine. Connection termination is also a three step process similar to the connection establishment phase. Source machine will send the connection termination request to the destination machine. Destination machine will send the ACK of termination request and its own termination request. In the last step, sender sends the confirmation of received acknowledgement and termination packet.

Transmission Control Protocol (TCP) is an example of connection oriented protocol which works at the transport layer. Connection oriented service is provided at the transport layer as well as the network layer. However, there are some subtle differences. At the transport layer, only two end systems are involved in connection establishment and setting of parameters, where as in the network layer, along with end systems, connecting devices i.e. routers along the path are also involved in setup process. Connection oriented service at the transport layer is implemented in the two end systems where as in the network layer it is implemented in all the routers in the chosen path along with the end systems.

In virtual circuit approach, a virtual connection would be established from source to destination on which all packets among this source destination pair would be sent. Therefore, it is also known as connection oriented service.

A network layer packetcontains the source and destination address as a part of header information because it provides logical communication among the machines.In virtual circuit approach, along with the source and destination addresses, packet contains a VC-ID. It is necessary to mention VC-ID in the packet as all packets has to follow the same virtual connection. When packet reaches to a router it will consult the forwarding table on the basis of VC-ID mentioned in the packet and decide the output port.

A virtual circuit approach involves three phases. The phases are a) setup phase b) data transfer and then connection termination. All are explained as follows:

a) **Setup phase:** Establishing a virtual circuit implies the following needs to be done.
   i. Deciding the path between a source and destination
   ii. Assigning a virtual circuit identifier (VC-ID) to each link along the path
   iii. Change in the forwarding table of all intermediate routers along the path with respect to virtual circuit

Let us understand this process with the figure 1.

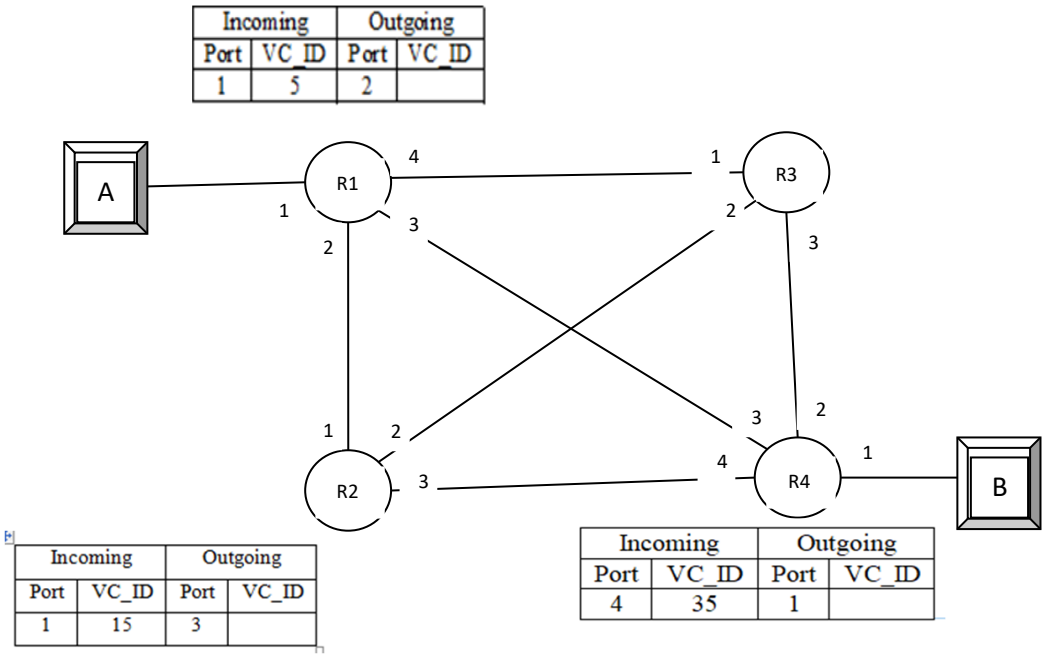


Figure 1: Sending of Request packet in virtual connection establishment process

As shown in figure 1, machine A wants to send the data to machine B. Let us chosen path between machine A and B is A-R1-R2-R4-B. Thus, a virtual connection needs to be established among A and B by involving all the intermediate routers. The process is as follows

1) Machine A will chose a VC-ID from its available list of VC-ID's and send the request packet to R1. As shown in figure 1, chosen VC-ID by A is 5.
2) As soon as Router R1 will receive this request packet, it will create an entry for this virtual circuit in its forwarding table as shown in figure 1. In this entry, Router R1 notes that the packet is coming from incoming port 1 and incoming VC-ID 5. Outgoing port is 2 and leave blank in place of outgoing VC-ID.
3) Now, R1 will forward this request packet to R2. In the similar manner, R2 will create an entry of this virtual circuit request in its forwarding table. Suppose, the chosen VC-ID by R1 is 15, thus the values of incoming port, incoming VC-ID, outgoing port and outgoing VC-ID are 1, 15, 3 and blank respectively.
4) R2 will forward the packet to R4. R4 complete the three entries of its forwarding table in the similar manner as shown in figure 1.
5) R4 sends the packet further to machine B. Machine B will chose a VC-ID and let this value is 60. In future communications, the VC-ID 60 is an indication for B that this packet comes from machine A.

All these five steps show the forwarding of request packet for setting the virtual connection from source machine A to destination machine B. But this forwarding completes the only three entries in the forwarding table. To complete the 4th entry of forwarding table, B will send an acknowledgment packet back to A via same path that is B-R4-R2-R1-A. The process can be visualized in figure 2 and explained as follows.

1) Destination machine B sends an acknowledgement packet carrying VC-ID 60 to R4. By knowing this value, Router R4 will complete the 4th column i.e. outgoing VC-ID of its forwarding table as shown in figure 2.
2) Router R4 will forwardthis acknowledgement packet to router R2. This packet contains the incoming VC-ID 35 which will be copied at the place of outgoing VC-ID in the table of R2.
3) Similar process will happen at R1. Router R1 receives the incoming VC-ID 15 from the R2 table. It will be copied at the place of outgoing VC-ID in the table of R1.
4) Finally, R1 forwards the acknowledgement packet to machine A which carries incoming VC-ID as 5. This VC-ID is chosen by A only in the initial process. Machine A knows that this VC-ID is to be used for communication to B.

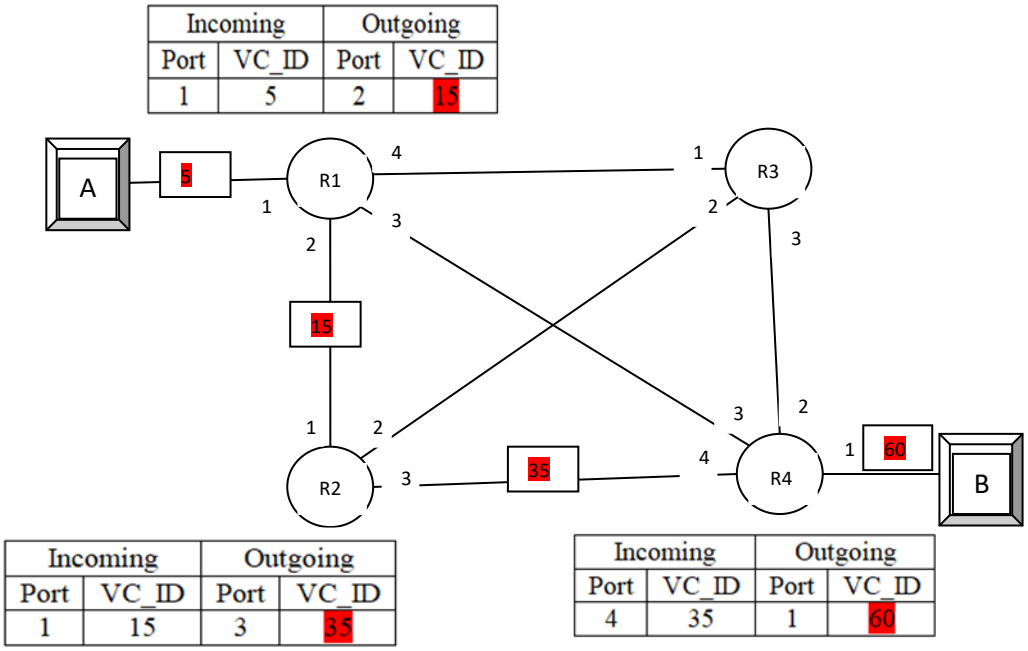


Figure 2: Sending of Acknowledgment packet in virtual connection establishment process

As discussed initially in the setup phase the virtual circuit establishment implies three works (deciding the path, assignment of VC-ID to each link, change in forwarding table) to be done. So, as explained in figure 1 and 2, all three mentioned works has been completed.

b) Data transfer:All the packets between A and B will be sent through the same established virtual circuit between them. As a result, thus all reach to the destination in order. Each intermediate router changes the value of VC-ID by seeing the forwarding table as shown in figure 3. As soon as the packet is reached to a router, it will see the VC_ID of this packet. In this example it is 5. Thus, R1 will see its forwarding table for the VC_ID 5 and incoming port

1. It can be visualized from figure 3, for these values as an index; the outgoing port is 2 and VC_ID is 15. R1 will change the VC_ID value in the packet and forward it further. Similar process will be followed at the other routers as well and finally the packet will be delivered to the destination machine B via established virtual connection. The figure3 shows the process for one packet. The same process would be followed by all the packets.

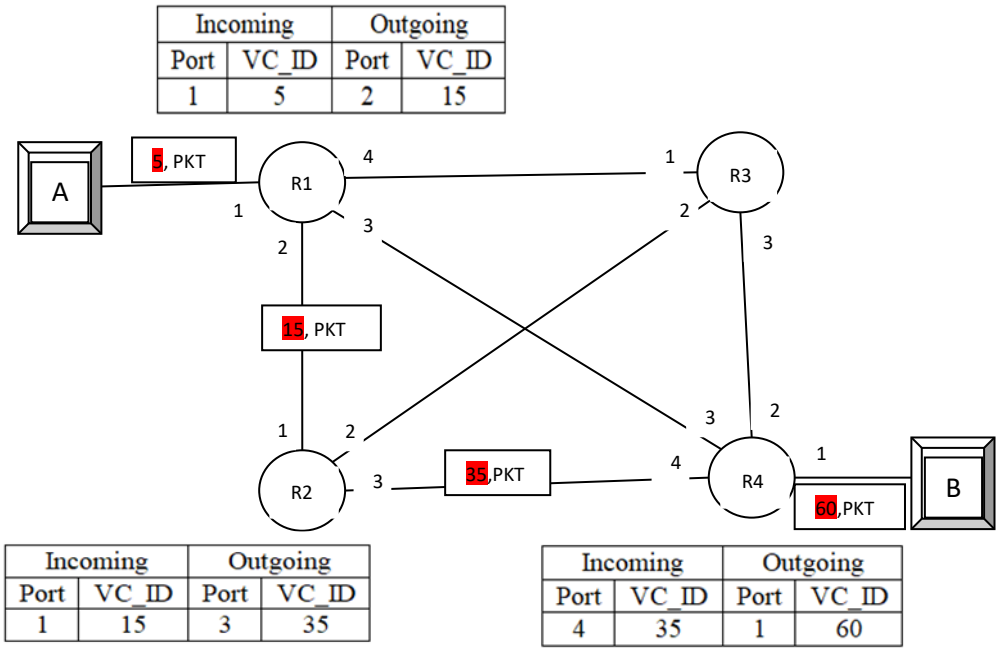


Figure 3: Packet transfer over an established virtual connection

c) **Connection termination:** Once A has sent all the packets to B, machine A will send a termination request packet to B and in return, B will send an acknowledgment of the same. As a result, all the routers delete the entry from the routing table.

1.3.2 Datagram Approach (Connection-less Service)

Datagram approach is used in today's Internet scenario. This approach follows the concepts of connection-less service. So, let us first understand the basic concepts of connection less service.

Connection less service, as its name implies no virtual connection would be made between source and destination. Both the entities does not do any handshaking. When a machine wants to send the packet to another machine, it simply starts sending. The message is divided into manageable units called packets where each packet is treated individually. Each packet can follow same or different paths, thus they may reach out of order or can be lost in between. Sender machine does not have any clue regarding the loss of packets as there is no provision of acknowledgement of packets. Due to all these properties, connection less service is an unreliable service.

Although this service is unreliable but it is required in some situations like where we want immediate transfer of data or where loss of some packets does not affect the overall quality of message or the situation where less overhead is required. The overhead of handshaking or sending acknowledgements is not present in connection less service.**User Datagram protocol (UDP)** and **Internet protocol (IP)** are the examples of connection less protocols which works at the transport layer and network layer respectively.

As **datagram approach** is a connection less service, thus all packets either belonging to same source destination pair or different, are treated individually. Here, packet is called as a datagram. Datagram contains the source and destination address. Forwarding decision is taken individually for each packet on the basis of destination address. Each router looks into the forwarding table for the mentioned destination address in the datagram. It returns the output interface based on matching on which the datagram will be forwarded further. If more than one entry are matched then based on the principle of longest prefix matching, the output interface would be selected.

In datagram approach, routing tables were modified by the routing algorithms. Routing algorithms is an important aspect of network layer which will be discussed later.

Destination address follows hierarchical addressing structure. How the destination address is actually extracted and processed to decide the path, let us understand it more clearly with an analogy. Suppose we want to go to a particular location 116-A, H Block, Vikaspuri, New Delhi. We started our journey from Noida and at the first junction if we ask for 116-A, H Block, Vikaspuri, New Delhi. Nobody will be able to tell the exact location or if somebody tried to do so, he or she will tell you only the road going towards Delhi. After reaching to Delhi, if u ask now for Vikaspuri, someone can instruct you by taking this path, you can reach to the desired location. Again the same situation, when you enter into Vikaspuri and you ask for 116-A, H Block , the person can tell you only about the path directions to reach H-block, not exactly 116-A. In the similar manner, part of destination address will be extracted and used to decide the output interface.

1.3.3 Comparison of Virtual Circuit and Datagram Approach

Both the approaches havetheir own advantages and disadvantages. Both the approaches can be compared on the basis of following points.

a) **Setup time:** Connection is to be setup in case of virtual circuit approach where as in datagram approach, no setup phase is required. Due to setup phase in virtual circuit approach, the sequencing of packets can be easily maintained. On the other hand, in datagram approach there is no setup overhead so sending of packets can be started immediately.

b) **Routing decision:**As the virtual circuit has been established between a specified source and destination, so no routing decision has to be taken for individual packets. Thus, packets can be forwarded more quickly in virtual circuit approach whereas in datagram approach, for every packet, the routing

decision is to be made. In virtual circuit approach, output port is decided by looking into the VC_ID of a packet where as in datagram approach, output port is decided by looking into the destination address.

c) **Reliability:**In virtual circuit approach, if a router gets failed, all the connection passing through that router or whose state information is maintained in this routergets lost. However, in datagram approach only the packets waiting in the queue of that router gets lost.

d) **Routing tables:** While establishing a virtual circuit, state of a connection needs to be updated in all the intermediate routers. Routing table is indexed by VC_ID where as in datagram approach, routing table is indexed by destination address and routing algorithms update the routing tables.

e) **Load balancing:** In datagram approach, packets may travel different paths. Thus traffic can be balanced over multiple routes.

f) **Reservation of resources:** In virtual circuit approach, resources are reserved so delivery can be guaranteed. If there are packets then allocated resources (like buffers, bandwidth, etc.) would be directly used. Whereas in datagram approach, the resources are shared on the demand basis. If everyone is trying to use the resources and resources are limited, then it may lead to congestion or packet loss. Congestion can be easily avoided in virtual circuit approach.

Table 1 provides a brief overview about the difference in virtual circuit and datagram approach.

Table 1: Comparison of virtual circuit and datagram approach

Virtual Circuit Approach	**Datagram approach**
Route is decided for all packets of a conversation between S and D	Route is decided for each packet
Overload may block connection setup and increase packet delay	Overload increase packet delay
Connection set up delay along with packet transmission delay	Only packet transmission delay
Forwarding decision based on VC_ID	Forwarding decision based on destination address
Congestion avoidance is easy	Congestion avoidance is difficult

1.3.4 A view of some Network Service models

Till now, we had discussed the overview of network layer services. This section discusses some of the network architectures to get an idea about their services.

Internet is most widely used network architecture. Internet's network layer provides best effort service. Best effort service implies it will try but does not guarantee anything. Therefore, it can be visualized from table 2, Internet network service model does not guarantee on any issue like ordering of packets, packet loss, bandwidth etc. It does not even preserve the timings difference among packets when the packets reach

at the receiver side. But, there are other network architectures which provide more than best effort service. Table 2 compares three network service architectures on the basis of their services. For more details please refer [1]

Table 2: Comparison of Network Service models

	Internet	**ATM**	**ATM**
Service model	Best effort	CBR	ABR
Guaranteed Bandwidth	No	Constant rate Guarantee	Minimum Guarantee
Delay Guarantee	No	Yes	No
Sequencing of packets	Any order	In order	In order
Packet Loss Guarantee	No	Yes	No
Congestion indication	No	No chances of congestion	Provides congestion indication

ATM CBR (constant bit rate) works on the principle of a virtual pipe between source and destination. Thus, it is able to provide some of the services like ordering of packets, guaranteed bandwidth to each user, etc. No packet would be lost and there are no chances of congestion as the resources are reserved while establishing the connection. ATM ABR (available bit rate) provides a minimum amount of bandwidth guarantee and delivers packets in order. But it does not provide any guarantee about the loss of packets and jitter among packets. Thus, ATM ABR is a little bit better than best effort service model of Internet.

➢ Check Your Progress 1

Choose the correct option.

Q1. _____ Approach takes the forwarding decision based on destination address.

a) Virtual circuit
b) Datagram

Q2. _____ is a connection less protocol used at the transport layer.

a) IP
b) TCP
c) UDP

Q3. _____ is an example of packet switched datagram networks.

a) Internet
b) Telephone networks

Q4. Compare Virtual circuit approach with the datagram approach. Provide at least two differences.

.....

……

……

1.4 NETWORK ADDRESSING

Network layer provides end to end communication i.e. it delivers the packets from source machine to the destination machine. This communication could be at a global level, thus, a unique identifier for every machine is required. This identifier is the logical address of machine, also known as Internet address or IP address. In actual terms, this address is not associated to the machine; it is associated to the interface. The portion between machine and the link is called as an interface. Generally, a host is connected to a single network through a link so it has one interface thus one IP address. On the other hand, a router is connected to many networks or hosts so it has multiple interfaces and each interface will have an IP address.

1.4.1 IP address

IP address is a 32 bit address. Along with the property of uniqueness, IP address should be universally acceptable also, that is who so ever wants to communicate, follow a common format. As the IP address is of 32 bits, thus 2^{32} unique addresses are possible. This much amount of address space impliesat an instant of time, approximately 4 billion machines with unique addresses could be connected.

IP address is generally written in dotted decimal notation (base 256). The other two notations are binary (base 2) and hexadecimal (base 16). As shown in figure 4, binary notation is just writing of all 32 bits in binary form. However, to increase its readability, the bits are written in a group of 8 bits that is a byte and some space will be provided between each byte. If we write decimal value of each byte and put a dot to separate the group is referred as dotted decimal notation. This is most commonly used notation. If we write hexadecimal value with respect to a group of 4 bits, then that notation is called as hexadecimal notation.

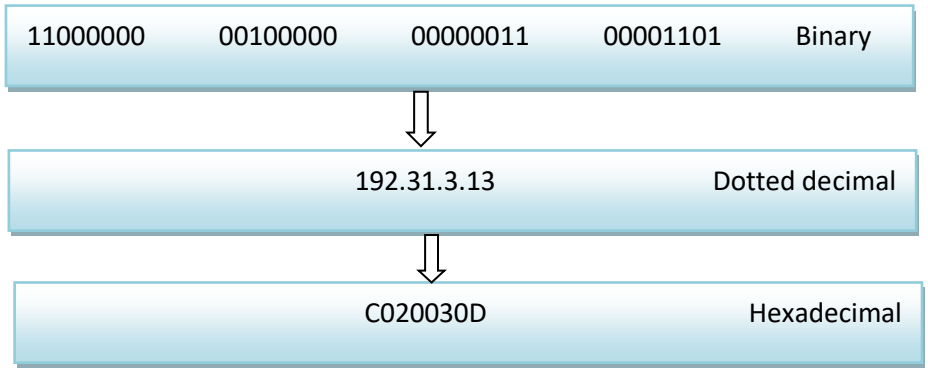


Figure 4: IP Address notations

1.4.2 Hierarchy in addressing

Network addressing follows hierarchical structure. The hierarchical structure can be easily visualized in our daily life examples like postal service, telephone networks. In postal service, posts have been distributed on the basis of written postal address. Postal address is extracted on the basis of country, state, district, city, street, building and house number. The address is always extracted in the reverse order. First, all the posts have been separated on the basis of country then state will be looked upon and so on. Similarly, the telephone networks also follow the hierarchy in telephone number. First few digits signify the country code which is followed by the area code and then the connection number itself.

In the similar manner, IP address is divided into two parts where first part signifies the network portion and second part is the host address. Network portion can be fixed or variable. If the network portion is fixed, it refers to **classful addressing** which is widely used in earlier days. But nowadays people switch onto the concepts of variable network portion which refers to **classless addressing**. The next chapter discusses the concept of classful and classless addressing in complete detail. Suppose b bits are used to denote network address, then the remaining $(32 - b)$ bits would be used to denote host address.

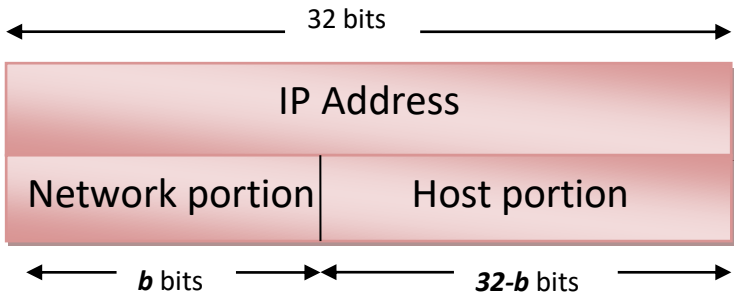


Figure 5: Parts of IP address

1.4.3 Getting an IP address

An IP address can be assigned to a host in two ways either manually or dynamically.

a) **Manual Assignment:** Network administrator allocates an IP address to the host from the available block of addresses. It would not be changed until the administrator himself changes it. This is also called static assignment.

b) **Dynamic Assignment:** When a host machine joins the network, an IP address would be automatically assigned by some protocol like Dynamic Host Configuration Protocol (DHCP). DHCP works like a plug and play protocol in the sense that as soon as someone joins the network, an address would be allocated and free the address when the host machine leaves the network. This is a dynamic assignment in the sense that every time a host joins a network, it will get a new address.

➢ Check Your Progress 2

Choose the correct option.

Q1. The _____ protocol is used to assign dynamic IP address.

a) Internet protocol
b) Transmission Control Protocol
c) Dynamic Host Configuration Protocol

Q2. If a host portion is of 8 bits then how many bits denote the network portion?

a) 32 bits
b) 24 bits
c) 16 bits

Q3. IP address is generally written in _____ notation.

a) Dotted decimal
b) Binary
c) Hexadecimal

Q4. How a host gets an IP address?

.....

.....

.....

1.5 CONGESTION

If the packets are coming at a faster rate than the handling capacity of the network, this leads to a situation known as congestion. Initially, when the packet arrival rate starts getting higher than the packet processing rate, queue starts filling up. As a result, packet delivery time gets increased. If the same situation continues, queue becomes full and packet drop starts. In this situation, source does not receive any acknowledgement and for a large number of the packets, the timer is up; which leads to unnecessary retransmissions. Sometimes the situation becomes worse and reaches to a deadlock point and whole system gets collapsed.

To understand the situation of congestion, let us see the behaviour of two important performance metrics i.e. delay and throughput with respect to the capacity of the network. Figure 6 shows the delay and throughput as a function of load.

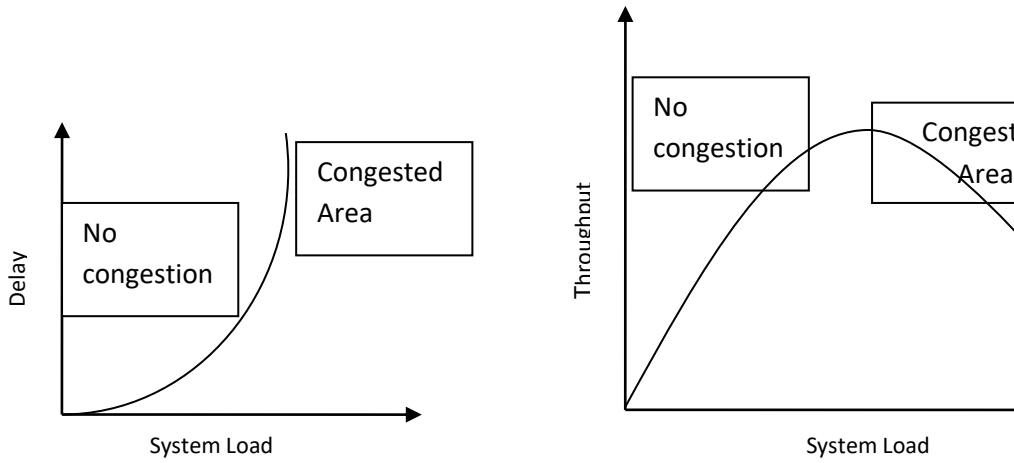


Figure 6: Delay and Throughput vs Load

Initially, when the less number of packets, they would be delivered without any delay and throughput is good. As the system load increases, packets experience queuing delay which in turn affects throughput as well. When the system load reaches the capacity, queue becomes full and some packets get discarded. As a result, delay approaches towards infinite value and throughput starts decreasing.

The issue of congestion is not only handled at the network layer; it is also handledat the transport layer. The main idea behind congestion is try to avoid the situation i.e. take preventive measures before reaching to a threshold and if the situation happens, try to come out of that situation. First one is known as congestion avoidance phase and second one is congestion removal. The policies used for congestion avoidance phase is called as open loop congestion control policies. Closed loop congestion control policies are for congestion removal phase.

Some of the **open loop congestion control policies** are as follows

a) Admission policy: A router can visualize the possibility of congestion and if there are chances of congestion then the new virtual connection request can be rejected.

b) Retransmission policy: When the sender does not receive any acknowledgement of the sent packet, it does the retransmission. For how long the sender has to wait or after how many lost packets, the packet needs to be retransmitted, all these kinds of retransmission policies should be designed in such a way that it will not add more congestion in the network.

c) Acknowledgement policy: Receiver's acknowledgement policy can also control the congestion at some level. For example, if receiver sends the acknowledgement packet after receiving some packets, it will slow the sender as well as not add a burden of sending acknowledgement packets.

d) Discard policy: If a router implements the good discarding policy then it can also prevent congestion at some level. The good discarding policy which does

not impact the overall quality of transmission. For example, in a multimedia transmission, some less priority packets gets discarded at the time of chances of congestion then it will not impact the overall quality of transmission.

Sometimes, even after taking all the preventive measures, congestion occurred. In this situation, **closed loop congestion control policies** to be used to avoid stucking into deadlock. These policies are

a) Sending of Choke packet: A choke packet is a control packet sent by the router to the source node. This packet informs the sender about congestion occurrence. This method is implemented by protocols like ICMP, etc.

b) Signaling: In this method a signal would be sent by the congested node to inform the sender about congestion. Rather than sending an explicit packet like choke packet, here, a signal will be sent in the existing packets carrying data.

c) Implicit signaling: In this method, no specific information is sent to the source rather sender itself guesses about congestion. For example, if the sender does not receive any acknowledgements of several packets with in timeout period is a signal for the sender that there is congestion in the network.

1.6 ROUTING

Job of the network layer is to send the datagram from a source end system to destination end system. Data may travel through different paths or through multiple hops to reach to the destination. The process of deciding about the path to reach to the destination is known as **routing**. There are routing protocols or which helps in constructing the forwarding table. The forwarding table or routing table is stored at every end system and router. Whenever a router receives a packet, router consults its routing table to decide the output interface. Looking into the routing table and choosing the output interface, this process is known as **forwarding**. Filling up of routing tables, their maintenance and regular updation is done at continuous intervals by **routing protocols or routing algorithms.**Routing algorithms is a part of network layer software.

There are various desirable properties ofa routing algorithm. These are as follows

1) Routing protocols decide the best route from a source S to destination D. This best route can be best in terms of any of the metrics like delay, throughput, packet loss, etc. This is similar to the analogy when we decide our travelling route from one city to another. There are various paths as well as various modes of transportation. We chose the one on the basis of cost,

comfort, time etc. Similarly, here in communication networks, the path is decided by looking into various metrics.

2) Other than the metrics, paths should be decided or updated in between by looking into the conditions of congestion into the network.
3) Routing is successful only when all the nodes are cooperative with each other. So, cooperation among the nodes is must.
4) Suppose, a link gets down or a router become fail, then all the routing tables to be quickly updated. This knowledge should be reflected in all the routing tables, so that packet loss will be minimal. Quick convergence and stability is very important in a routing algorithm.

To solve the routing problems, network isrepresented in terms of graphs. While formulations as a graph, routers are represented as nodes and the links connecting the routers are represented as edges. A graph G is represented as (V, E). Where, V is a set of nodes and E is a set of edges connecting those nodes. Each edge is labeled with a value representing its cost. This cost could be directly or indirectly related to the link type or metric value. For example, cost could be directly proportional to the congestion or inversely proportional to the bandwidth. Higher bandwidth link or less congested link gives a low cost value of that link. If the nodes are connected by an edge then the cost is associated with that edge else it is infinity. Undirected graphs are considered for formulation. Thus, the cost associated with edge (a, b) is same as edge (b, a). A path is a sequence of edges traversed from chosen starting node to chosen destination node. The cost of a path is sum of all the edges cost of that path. Let us understand this formation more clearly with the figure 7.

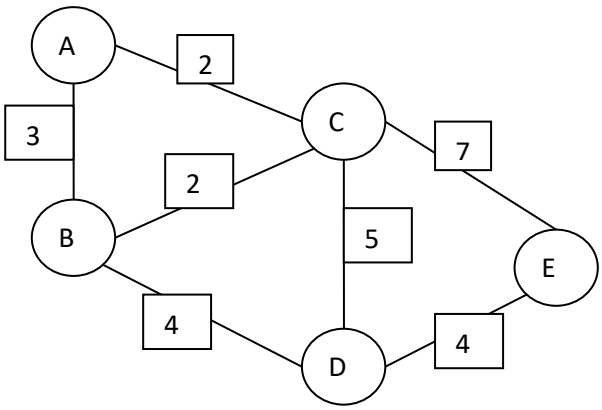


Figure 7: Graph representation of a network

It can be visualized from the figure 7, set V consist of routers {A, B, C, D, E} and set E consist of edges such as (A,B), (A,C), (D, E), etc. Each edge is labeled with a value known as cost. Let us calculate the path from node A to node E. It can be visualized from the figure there are multiple paths from A to E. Some of the paths and associated costs are as follows

Path	Cost
A-C-E	2+7=9
A-B-C-E	3+2+7=12

A-B-D-E	3+4+4 = 11

The cost of a path is sum of the costs of all traversed edges from A to E. On the basis of low cost, routing algorithm chose the path A-C-E. Suppose, all the edges have same unit cost, then the path with less number of hops is chosen. For this scenario, again path A-C-E would be selected because it has less number of hops in comparison to other paths.

Now, revise all this scenario of finding the paths in your mind and tell me, how you have chosen the path. Have you tried all the combinations? I think no, you have tried just 2-3 paths and convinced yourself that this is the least cost path. All this work of your mind is done actually by a routing algorithm but to do this job, routing algorithm shouldknow the complete knowledge about the network.

1.6.1 Classification of Routing Algorithms

Routing algorithms can be classified into different types like centralized or decentralized based on how the routes would be computed. Another categorization is adaptive or non-adaptive routing algorithms which are based on strategy of updation of routes. All these different types are discussed as follows

a) **Global Routing Algorithm**: These routing algorithms compute the best path (least cost one or shortest one) by gathering the complete knowledge about the network. It is also known as centralized routing algorithm. It can be run at one central location or a replica to be run at multiple locations. All the information about the nodes and links to be collected at the central algorithm and then this algorithm computes the routes. The computed routing tables would be distributed to all nodes. In this way, global optimal routes will be computed and distributed to all. It will reduce the burden on each node. **Link state algorithms** are a kind of global routing algorithms.

b) **Decentralized Routing Algorithm:** These routingalgorithms works in an iterative, asynchronous and distributive manner. Complete network knowledge is not available at a single location rather each node interacts with its neighbors about the cost and their knowledge. Now, these neighbors interact further with their neighbors and this process goes on. In this way, the path to one or more destinations would be computed. **Distance vector routing algorithm** is an example of decentralized routing algorithm.

c) **Adaptive Routing Algorithm:**These routing algorithms reflect the change in the routing tables whenever there is a change in topology. That's why these algorithms are also known as **Dynamic routing algorithms.** These algorithms also update the paths with respect to network traffic conditions.

d) **Non-adaptive Routing Algorithm:** These routing algorithms do not update the routing table periodically or at the instance of change in topology. Manually, the routes would be computed and stored. These routing algorithms

do not respond to failures automatically, that's why these algorithms are also known as **Static algorithms**.

➢ Check Your Progress 3

Choose the correct option.

Q1._____ process creates and maintains the routing table.

Q2. Which routing algorithm works in an iterative, asynchronous and distributive manner?

a) Link state routing algorithm
b) Distance Vector routing algorithm
c) Static algorithm

Q3. Define congestion.

.....

.....

Q4. What are the various policies that can be used to avoid congestion?

.....

.....

.....

1.7 Delay in Packet Switched Networks

The packet transmission process starts from a source and ends at a desired destination. In this transmission process, packet travels through a number of intermediate routers and paths. Thus, a packet will not reach immediately to the destination. Rather it experiences a number of delays.

1.7.1 Types of delay

A packet experience four types of delays which are explained as follows:

a) **Transmissiondelay**: A source machine transmits a packet means that source machine put one by one bit of that packet on the link. A packet has certain length thus, it can't be put on the link in one go. Total time experienced by the source machine in this process is known as transmission delay. If a packet length is denoted as L and transmission rate is denoted by R, then transmission delay is calculated as L/R .

b) **Propagation delay:**As soon asthe bit is put on the link, this bit has to travel through a number of intermediate links. For a single intermediate link, propagation delay is calculated as distance of this link divided by the speed of

the link. Speed of the link depends on the physical type of link. Generally, speed is considered as $3*10^8$ m/s, which is propagation speed of the vacuum.

c) **Processing delay:** This is the amount of time taken by a router or destination machine to process a packet. Packet content would be checked for error detection. Packet would be processed to the upper layer protocol if it is a destination machine. If it is a router, it would be processed to the selected outgoing port. Generally, the value of processing delay depends on the speed of the router.

d) **Queuing delay:** As its name suggests, this is the amount of time a packet waits for its turn to get to be transmitted. Each router has an input queue for incoming port and an output queue for outgoing port. Summation of both the waiting times is known as queuing delay. Queuing delay mainly depends on the packets already waiting for their turn. If there is no packet in the queue, queuing delay is zero.

1.7.2 Computation of delay

Total delay is the summation of all the above types of delays defined in above subsection. The following notations represent four delays.

- $d_t - Transmission\ delay$
- $d_p - Propagation\ delay$
- $d_{proc} - Processing\ delay$
- $d_q - Queuing\ delay$

To compute the total delay experienced by a packet from source to destination, it has to been how many links and routers in between. If there are k links in the path from source to destination, it implies there are $k - 1$ routers in between. For all the k links, transmission, propagation and processing delays to be computed but queuing delay is to be computed at router only. Total delay is computed as follows

$$Total\ delay = k\left(d_T + d_P + d_{proc}\right) + (k-1)d_q$$

1.7.3 Numerical

Q1. Suppose two hosts Y and Z are directly connected by a link. Length of this link is 10,000 Km and this link transmission rate is 1Mbps. The propagation speed of the link is $2.5*10^8$ m/s. Based on this information answer the following parts

a) Y sends a file of 400K bits to Z. How long does it take to send the file assuming it is sent continuously?

b) Suppose now the file is broken up into 10 packets with each packet containing 40K bits. Z sends an ACK for each packet and Y cannot send a packet until the preceding one is acknowledged. Transmission time of an ACK packet is negligible. How long does it take to send the file?

Answer:

a) File size = 400,000 bits , Transmission rate = 1 Mbps

Thus, transmission delay is

$$= \frac{file\ size}{trans\ rate} = \frac{400000}{10^6} = 400\ msec$$

Distance = 10,000 Km, Speed = 2.5 *10^8m/s

Thus, propagation delay is

$$= \frac{distance}{speed} = \frac{10 * 10^6}{2.5 * 10^8} = 40\ msec$$

In this scenario, there is no processing delay or queuing delay. Therefore, total delay is

Total delay = Trans delay + Propagation delay = 400 + 40 = **440 msec.**

b) Packet size = 40,000 bits , Transmission rate = 1 Mbps

Thus, transmission delay for one packet is

$$= \frac{file\ size}{trans\ rate} = \frac{40000}{10^6} = 40\ msec$$

Distance = 10,000 Km, Speed = 2.5 *10^8m/s

Thus, propagation delay is

$$= \frac{distance}{speed} = \frac{10 * 10^6}{2.5 * 10^8} = 40\ msec$$

There are two important points to be note down

i. Second packet is sent only when Y receives the acknowledgement of first packet, thus twice of propagation delay is used.
ii. Acknowledgement can be sent only when the first packet is received completely at the receiver end. Thus, twice of transmission delay is used.

Therefore, for one packet total delay is = 2*Trans delay + 2*Propagation delay

And the total delay for all 10 packets is = 10*(2*Trans. delay + 2*Propagation delay)

= 10*(2*40 + 2*40) =**1600 msec.**

Q2. Compute the end to end delay for circuit switching and packet switching for a network. This network is having 5 hops to switch a message of 1200 bits where all the links have a data rate of 4800bps. Size of the packet is 1024 bits along with a header of 32 bits. In case of circuit switching, consider 0.5sec as a call setup time. Hop to hop delay is .02 sec. Assume zero processing delay.

Answer: This answer is divided into two parts a) Computation of delay in circuit switching scenario b) Computation of delay in packet switching scenario

a) Computation of delay in circuit switching scenario

Call set up time is = 0.5 sec
Propagation delay is = .02 sec
Transmission delay is = $\frac{1200}{4800} = 0.25$ sec
Total delay = call set up time + message delivery time
= 0.5 + 5*(propagation delay) + transmission delay
=0.5+ 5*(0.02) + 0.25 = **0.85 sec**

b) Computation of delay in packet switching scenario

The given packet size is 1024 bits. It implies 32 bits of header and the leftover 992 bits are data bits. Thus, to send the total message of 1200 bits, two packets are required.

- First packet is of 1024 bits (992 bits of data and 32 bits of header).
- Second packet is of 240 bits (1200 – 992= 208 bits of data and 32 bits of header).

Propagation delay is = 5*0.02 = 0.1 sec
Transmission delay at first hop = $\frac{1024}{4800} + \frac{240}{4800} = 0.213 + 0.05 = 0.263$ sec
Transmission delay at rest of the hops is $= 4 * \frac{1024}{4800} = 4 * 0.213 = 0.852$ sec. because at the rest of the hops there is no transmission delay for 2nd packet or any other number of packets.

The total delay is $0.1 + 0.263 + 0.852 = \mathbf{1.215}$ **sec.**

➢ Check Your Progress 4

Choose the correct option.

Q1. If there is no buffer at the router, each incoming packet directly forwarded further onto the outgoing port. In this situation which kind of delay is negligible?
a) Processing delay
b) Queuing delay
c) Transmission delay
d) Propagation delay

Q2. Host X is connected to Y via switch S. The link bandwidth is 10Mbps and propagation delay on each link is 20μs. S is a store and forward switch, it begins retransmitting a received packet 35μs after it has finished receiving it. Calculate the total time required to transmit 10,000 bits from X to Y.

a) As a single packet
b) As two 5,000 bit packets sent on right after the other

.....

.....

.....

.....

1.8 SUMMARY

In this unit, we understood the concepts of packet switching. Network layer follows the concept of packet switching as packet is the basic data unit used at this layer. There are two types of packet switching techniques, virtual circuit and datagram approach. In virtual circuit approach, before sending any data between a source destination pair, end to end logical connection needs to be established between them. Datagram approach is a connection less service. There is no handshaking between source and destination and each packet follows its own route.

Each machine is identified by a logical address, known as IP address. It is a 32 bit address which is unique for an individual. Due to overload of the network, network layer faces a problem which is known as congestion. Open loop congestion control policies for avoiding the congestion and Closed loop congestion control policies for congestion removal phase has been studied. A very important job of network layer is route the packets, thus concepts of routing has been discussed. It is followed by the computation of delay metric which is a very important in network layer as packet travels through a number of paths and routers.

1.9 SOLUTIONS/ANSWERS

Check your progress 1

1) b

2) c

3) a

4) Virtual circuit approach decides the output port on the basis of VC_ID of a packet whereas datagram approach decides the output port on the basis of destination address mentioned in the packet. If a router gets failed, only the packets waiting in the queue of that router gets lost in datagram approach. However, in virtual circuit approach, all the connection passing through that router or whose state information is maintained in this router gets lost.

Check your progress 2

1) c

2) b

3) a

4) A host gets an IP address either manually which is assigned by a network administrator or dynamically assigned by the DHCP protocol. Manual assignment is also known as static assignment as the allocated address can't be changed until the administrator himself wants to change the same. In manual assignment, network administrator allocates an IP address to the host from the available block of addresses.
During dynamic assignment, Dynamic Host Configuration Protocol assigns the IP address to the machine as soon as the host joins the network and frees the address when the host machine leaves the network. Therefore, whenever host joins the network, it will get a new IP address.

Check your progress 3

1) Routing

2) b

3) Network Layer faces this problem of congestion when the number of packets sent to the network is greater than the capacity of the network. Network is not able to handle the packets as packets are coming at a faster rate. Then, packet loss starts and sometimes it leads to a deadlock situation and the whole system gets collapsed.

4) There are policies used to avoid the congestion known as open loop congestion control policies. But if congestion occurred, then some of the policies are used to remove the congestion.

The following **open loop congestion control policies** can be used to avoid congestion. Receiver's acknowledgement policy can control the congestion at some level. For example, if receiver sends the acknowledgement packet after receiving some packets, it will slow the sender as well as not add a burden of sending acknowledgement packets. Another policy implemented by router that the router can visualize the possibility of congestion and if there are chances of congestion then the new virtual connection request can be rejected. Sender can implement the retransmission policy to avoid the problem of congestion. For how long the sender has to wait or after how many lost packets, the packet needs to be retransmitted, all these kinds of retransmission policies should be designed in such a way that it will not add more congestion in the network. Sometimes, even after taking all the preventive measures, congestion occurred. In this situation, **closed loop congestion control policies** to be used to avoid stucking into deadlock.

These policies are mainly about informing all about the situation of congestion. One of the policy is sending a choke packet. A choke packet is a control packet sent by the router to the source node. This packet informs the sender about congestion occurrence. Another is signaling in which a signal would be sent by the congested node to inform the sender about congestion. Rather than sending an explicit packet like choke packet, here, a signal will be sent in the existing packets carrying data.

Check your progress 4

1) b

2) Propagation delay =20μs
   Transmission delay $=\frac{L}{R}=\frac{10,000}{10*10^6}=1000\mu s$

   a) Total time (as a single packet) =20+1000+35+20+1000=2075μs
   b) Transmission delay $=\frac{L}{R}=\frac{5,000}{10*10^6}=500\mu s$
      Total time for first packet = 20+500+35+20+500=1075μs
      Total time for second packet = 500μs

      Thus, total delay = 1075+500 = 1575μs

1.10 FURTHER READINGS

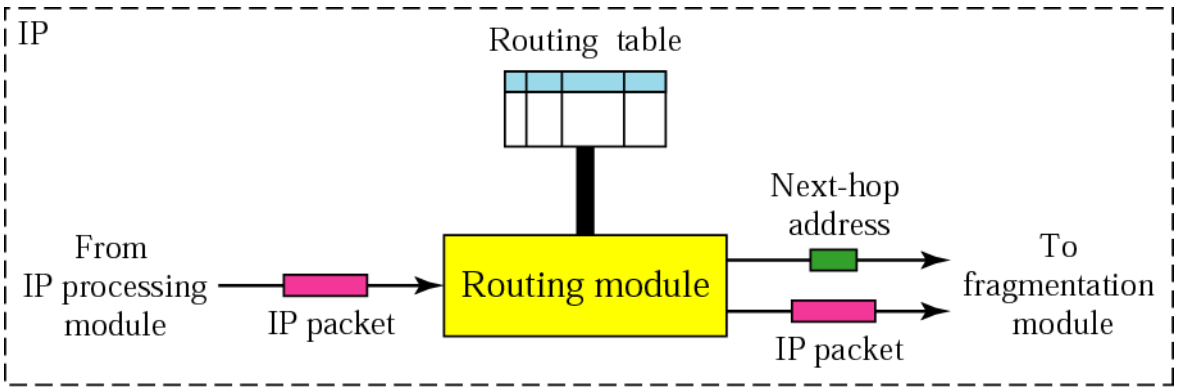
[1] Kurose, J. F., & Ross, K. W. (2012). "Computer networking: A top-down approach featuring the Internet", Boston: Addison-Wesley.

[2] Forouzan, B.A., & Mosharraf, F. (2012), "Computer Networks: A top-down approach", McGraw Hill.

UNIT 2 ROUTING ALGORITHMS

2.0 INTRODUCTION

Network layer is responsible for finding the optimal route from source to a destination. Multiple paths may exist between a pair of source and destination. A path with minimum cost is considered to be the optimal route. Routing algorithms construct and maintain a table called Routing Table which is referred while looking for a route. A routing algorithm is responsible for selecting the most appropriate route in the network between source and destination. Router is the network layer device which is responsible for performing routing for the network. A cost is associated with each path in the form of bandwidth, delay, congestion, security etc. Router performs routing and selects the minimum cost path for all the remote networks. A router is implemented with a number of algorithms to find the optimal routes.Based on the criteria of optimal path according to the requirement of the network traffic, appropriate routing algorithm can be chosen.



In this unit section 2.3 is about flooding which uses broadcasting. In section 2.4 shortest path routing algorithmi.e. Dijkstra's algorithm is discussed. In section 2.5 Distance vector routing algorithm: Bellman-Ford Algorithm is discussed. In this section comparison between Dijkstra's algorithm and Bellman-Ford Algorithm and count-to-infinity problemis also discussed in this section. Section 2.6 covers a link state routing protocol and its working. In the section 2.7 hierarchical routing is discussed. Section 2.8 deals with the Internet Protocol (IP). In this section Ipv4 and IPv6 along with ICMP, DHCP and IP security are covered. Section 2.9 discusses routing in the Internet and protocols RIP, OSPF and BGP. Section 2.10 is about multicast routing. In section 2.11 Mobile IP is introduced. Section 2.12 summarizes the chapter. In Section 2.13 review questions and their solutions are covered. Section 2.14 lists further readings.

2.1 OBJECTIVES

After completing this section, one should be able to:

- understand how the working of shortest path routing algorithm;
- construct a spanning tree;
- understand the working of distance vector routing and link state routing;
- understand hierarchical routing;
- understand Internet Protocols IPv4 and IPv6, and
- understand and implement multicast routing.

2.2 FLOODING

Whenever any link's or router's state is changed either up or down, leads to change in the topology. And whenever a topology change happens the same is required to be communicated to all the nodes of the topology. This is done by sending a topology change message to all the nodes (very large in numbers) in the network using **broadcasting**. In networking such type of broadcasting is known as **flooding**. Flooding is of two types: uncontrolled flooding and controlled flooding. Uncontrolled flooding does not restrict sender to send the packet to the node from which it received the same packet. In controlled

flooding a node does not forward the packet back to the node from which it has received the packet. Another technique to reduce the duplicate packets in the network is to make the provisions so that a node relays a packet only once. To implement this the sender adds it ID and a unique auto increment sequence number to each packet. Whenever a packet is received by a node, it stores the sequence number and the ID of origin node. Before relaying the packet, node checks the sequence number of the newly received packet and the sequence number stored for the origin sender. A packet will only be relayed if the sequence number of the packet is greater than the sequence number stored for the origin sender. By doing so the duplicate packets in the network are reduced to a great extent. Now days uncontrolled flooding is not used in general. In flooding the source node sends a packet (of information to be shared) to all its neighbours: the nodes connected with a direct link. These nodes further send the packet totheirneighbours, and this process continues till each node in the network receives the packet.

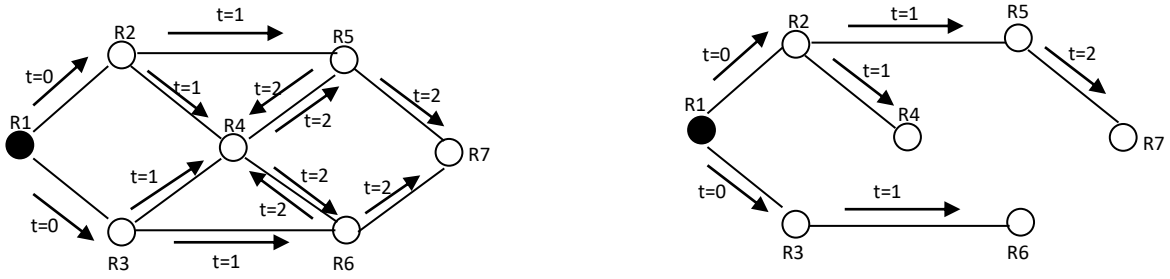


Figure 1: Packet Flooding (a) without spanning tree (b) With spanning tree

Considering the *figure 1(a).*Suppose a change in topology is observed by node R1. R1 will send the notification packets toR2 and R3. R2 will send the packet to R4 and R5. R2 will send the packet to R1 (as it has received the same from R1). In similar way node R3will send the packet to R4 and R6. Node R4 has received the same notification packet originated by same origin with same sequence number. So R4 will further send the packet arrives first and will discard the later one. Similarly R6 will discard the later one and forwards the first received packet to R7. Node R5 receives packets both from R2 and R4, so in same way the packet arrives first will be forwarded and later one will be discarded. R7 will receive the packet from both R5 and R6.

Another way to reduce the redundancy of packets in the network and to avoid the cyclic forwarding of the packets is to construct a logical spanning tree of the topology.

Considering L as the number of bi-directional links of the network, for a packet to be broadcasted the total number ofpacket transmission lies between L and 2L. Arrows on the links show packet transmissions with the time of

transmission (assumed to be 01 unit for each packet) shown. In figure 1, the flooding is shown with both the methods without and with spanning tree construction. In flooding without spanning tree the number of packets transmitted is in generally many more as in the case of with spanning tree. Also, in both the cases, the broadcast packet reaches to all nodes within same time. For a graph many spanning trees are possible, hence the flooding time depends on the spanning tree constructed.

2.3 SHORTEST PATH ROUTING ALGORITHM

- In Shortest path routing method router builds a graph of the network, where node of the graph represents the router and connecting links are the lines joining the nodes. Shortest path between two nodes is calculated by considering parameters as the number of hops, or the geographic distance between them etc. Shortest path routing technique is based on the greedy approach by considering the next best possible option without considering the overall best option. Hence, sometimes it may be the case that overall there may exist some other shortest path between two nodes as selected by the algorithm.In this method least cost paths from one node ('source") to all other nodes are computed.

Many algorithms are proposed based on this technique,**Dijkstra algorithm** is one of the widely popular shortest path routing algorithm. Dijkstra algorithm is also based on greedy approach as is also known as the least cost path approach. The working of Dijkstra algorithm is as follows:

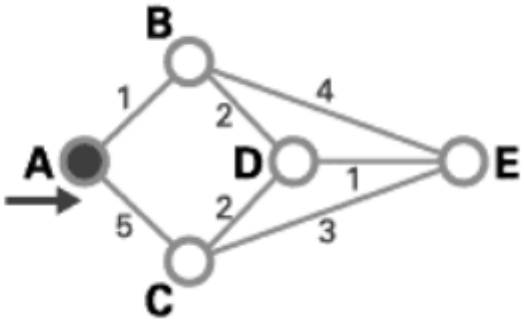
1) Select the source node as the start node (S).

2) Mark the direct neighbor nodes as tentative nodes (Initially all the nodes are considered as tentative nodes).

3) Choose the node from tentative nodes list with lowest cost from the source node and mark it as permanent nodes and make it as source node.

4) If, the destination node is covered or there is no more nodes in the tentative list(i.e. no more nodes to be explored) then stop, otherwise go to step number 2.

Considering the figure below and applying **Dijkstra** routing algorithm for finding the best path or the shortest path fromsource node A tothe destination node E. The steps followed are as follows:



1) Node A is selected as source node.
2) Direct neighbor nodes B and C are marked as tentative node.
(3) Node B has the lowest cost path (of cost 1) from source node A, so it is marked as permanent node.
3) Node B is not the destination node and node D, E are available nodes in the tentative nodes lists, so
(4) Make node B as the source node now and explore the direct neighbors of it.
(5) Nodes D and E are the direct neighbors of B,
(6) Node D has minimum cost path from B so, node D is selected and marked as permanent node and D is not the destination node and also the tentative list is not empty.
(7) Make D as the source node now and explore its direct neighbors.
(8) Node C and E are the direct neighbors of D.
(9) Node E has the less path cost than C from node D, so E is marked as permanent node. Node E is destination node, so stop here.
(10) The shortest path from A to E is: A – B – D – E.

2.4 DISTANCE VECTOR ROUTING

In today's scenario where the number of Internet users increased manifolds', static routing is not feasible as the static routing algorithms are not adaptive in nature. Under such dynamic scenario dynamic routing algorithms performs very well.
Considering dynamic algorithms: **Distance vector routing** and **link state routing** are most widely known and used dynamic algorithms. Distance Vector Routing Algorithm has got following properties:

- **Distributed** – Each node receives some information from one or more of its ***direct*** neighbours and performs path calculation at its own.
- **Iterative** – iterative in nature means exchange of information among neighbours continues until no more (new/updated) information available to exchanged.

Here, in this section distance vector routing algorithm is discussed. Many timesdistance vector algorithm is known as Bellman-Ford algorithm because vector algorithm is based on Bellman-Ford Equation.

Bellman-Ford Algorithm

Each router also constructs/ maintains routing table known as Distance Vector table storing the path cost (in terms of distance or the hop count) to reach ALL feasible destination nodes from itself. The path cost iscalculated using information received from the neighbour's distance vectors.

A router maintains following information for Distance Vector table -

- Router ID (each router has an ID)
- Link cost for each link connected to a router
- Intermediate hops

Initially the Distance Vector table is initialized as follows:

- Cost to itself (C = 0
- Cost to all other routers = infinity.

Algorithm: Distance Vector Routing-

1. A router shares its routing table/Distance Vector to its neighbours (directly connected routers).
2. Each router on receipt of Distance Vector from its neighbours, it saves the most recently received Distance Vectors.
3. A router updates its routing information/ Distance Vector according to the Bellman-Ford equation, when:
   - A neighbour shares a Distance Vector with routing information different than before.
   - The status of a link to its any of the neighbour changes (up or down).

The Distance Vector is measured while minimizing the cost to each destination router.

Notations:

$D_x[y]$ = Estimate of minimum distance from node x to node y

C[x,v] = Node x has cost to each neighbour v

D_x= $[D_x[y]: y \in N]$ = Node x records distance vector

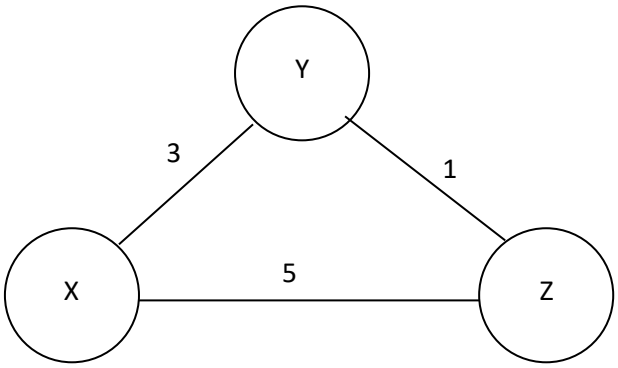
Node x also maintains its neighbours' distance vectors

Note:

- For each neighbour v, x maintains $D_v = [D_v[y]: y \in N]$

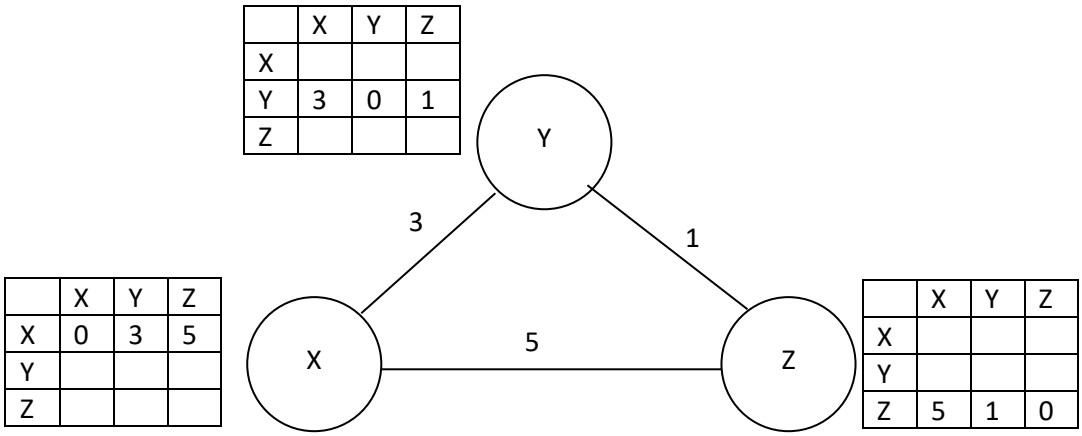
- Each node sends its own routing information/ Distance Vector information estimate to its neighbours after certain time interval.
- When any node x shared with a new Distance vector information from any of its neighbor v, it stores the distance vector of node v and updates its own routing information using Bellman-Ford equation:
  $D_x[y] = MIN\{ C[x,v] + D_v[y], D_x[y] \}$ for each node $y \in N$

Example – Considering the figure below, construct the routing tables for routers X, Y and Z.



Step 1: Each node knows the distance to reach to its direct neighbour nodes. The distance to itself is 0 and the distance to nodes which are known discovered yet is considered as ∞.

Initially, the DV/ routing table of each node can be as:



	X	Y	Z
X			
Y	3	0	1
Z			

	X	Y	Z
X	0	3	5
Y			
Z			

	X	Y	Z
X			
Y			
Z	5	1	0

Step 2: For router X:

Router X will share its distance vector (DV) information to its direct neighbours and neighbours (Y and Z) will share their DV information/routing table to X.

The distance from source node X to destination node will be determined using bellmen- ford equation.

$D_x(y) = \min \{ C(x,v) + D_v(y)\}$ for each node $y \in N$

When node Y shares its DV information with X, X will recalculate its DV information using Bellman Ford equation.

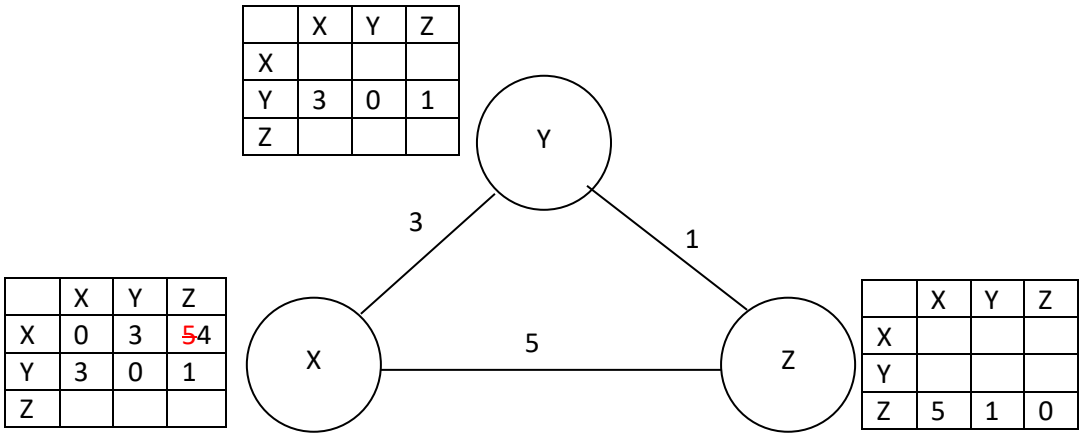
Given data:

Clearly, $C_v(z) = 1$, $C_x(z) = 3$, $C_x(Y) = 3$, $d_x(x) = 0$

Now, using the bellman Ford equation and the DV information shared by node Y to node X.

$d_x(z) = min\ \{\ \mathbf{[c(x,y) + d_v(z)], [c(x,z) + d_z(z)]}\}$

$= min\ \{[3 + 1], [5 + 0]\} = 4$



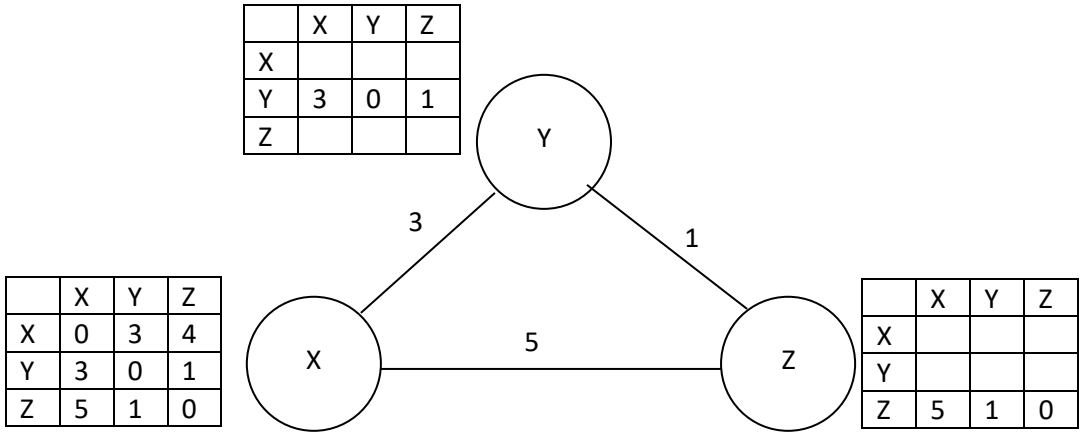
	X	Y	Z
X			
Y	3	0	1
Z			

	X	Y	Z
X	0	3	~~5~~4
Y	3	0	1
Z			

	X	Y	Z
X			
Y			
Z	5	1	0

Similarly, when node Z shares its DV with node X:

No updates in X's DV information. [Note that Z still have a distance 5 to node X]



	X	Y	Z
X			
Y	3	0	1
Z			

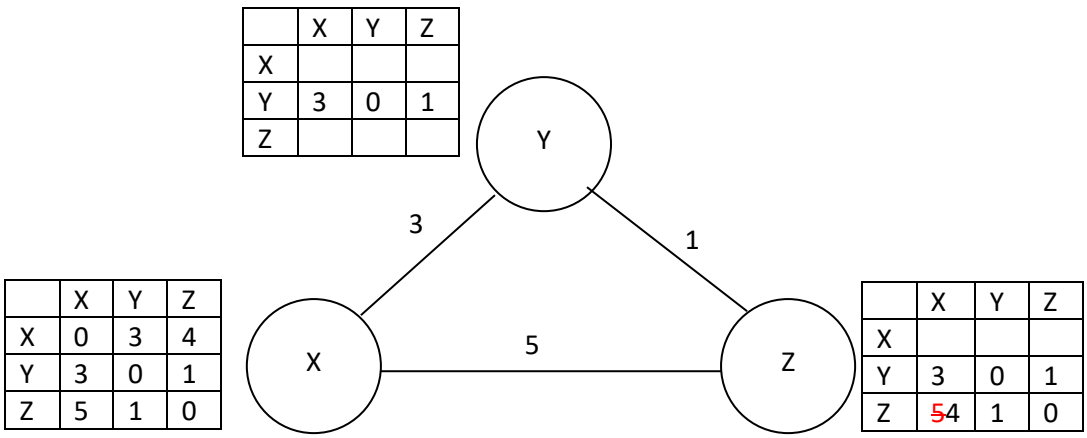
	X	Y	Z
X	0	3	4
Y	3	0	1
Z	5	1	0

	X	Y	Z
X			
Y			
Z	5	1	0

Similarly, when Y shares its DV information with node Z:

$d_z(x) = min\ \{\ \mathbf{[c(z,y) + d_v(x)], [c(z,x) + d_x(x)]}\}$

$= min\ \{[1 + 3], [5 + 0]\} = 4$

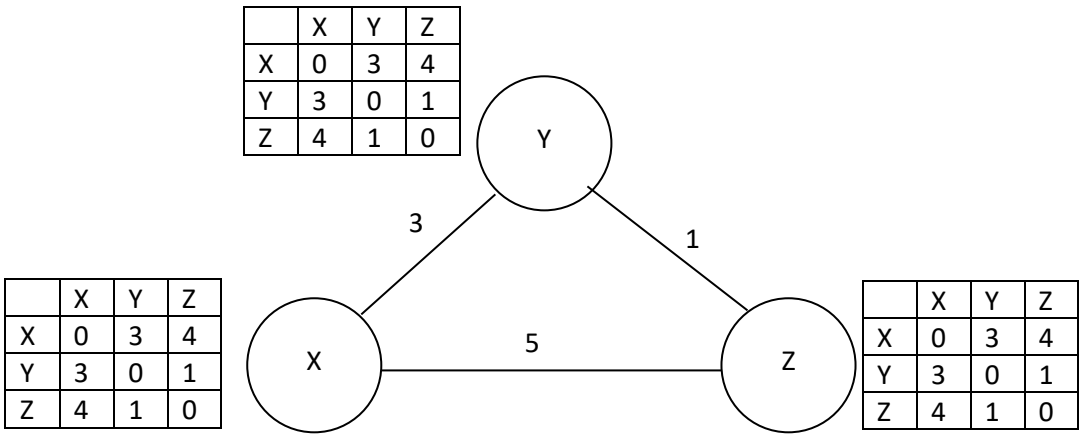


	X	Y	Z
X			
Y	3	0	1
Z			

	X	Y	Z
X	0	3	4
Y	3	0	1
Z	5	1	0

	X	Y	Z
X			
Y	3	0	1
Z	~~5~~4	1	0

Finally the routing table for all –



At the end of convergence process, the DV information of all the nodes are same until a new change in the topology occurs.

2.4.1 Comparison

The comparison of the two routing algorithms approach should be based on the new path processing time and the traffic generated by these for the routing information convergence process.

The evaluation of an algorithm is moreover depends onthe implementation approach and the specific implementation.

The discussed routing algorithms can be compared on following points:

1. Message complexity

- Link State algorithm:Link state algorithm sends order of O(nE) messages with n nodes, E links.
- Distance Vector algorithm: Messages areexchange between directly connected neighbors only

2. Speed of Convergence

- Link State algorithm:It takes order of $O(n^2)$ to converge, where n is the number of routing nodes
- Distance Vector algorithm: Convergence time is not standard with this and varies due to following situations:
  - may be routing loops
  - count-to-infinity problem

3. Robustness: Robustness is the confidence of getting the correct result under any circumstance.

- Link State algorithm:Link State algorithm can face issues like:
  - node can advertise incorrect *link* cost
  - each node computes only its *own* table
- Distance Vector algorithm:Distance Vector algorithm can face issues like:
  - Distance Vector node can advertise incorrect *path* cost

- each node's table used by others, so if a node shares incorrect path it could be propagated further to the network on a large scale.

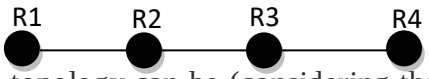
Further, both can be compared on some more points as follows:

Distance Vector Routing	Link State Routing
• Tell neighbors about distance of all the destination. • Node's computation depends on neighbors. • Each router constructs a distance vector table with (dist/cost, exit interface) tuple for eachdestination. • A node shares copy of distance vector table to all its direct neighbors.	• Tell about distance to each neighbor to all routers • Each router computes its best paths

2.4.2 The Count-to-Infinity Problem

One of the serious drawbacks in Bellman Ford algorithm is Count to Infinity Problem. It is also known as routing loop problem. In Bellman Ford algorithm implementation count to infinity is also known as routing loop problem. In Bellman Ford algorithm, routing loops problem is faced when a link/interface shutdown/goes off. This issue may also occur, when two routers exchange routing information to each other at the same time.

Consider the below figure to discuss it in detail.



The routing table for above topology can be (considering the distance of each link is to be 1 unit). Each cell shows the pair (distance, predecessor node):

	R1	R2	R3	R4
R1	0, -	1, R1	2, R2	3, R3
R2	1, R2	0, -	1, R2	2,R3
R3	2, R2	1, R3	0, -	1, R3
R4	3, R2	2, R3	1, R4	0, -

As it is visible in the figure, node R1 is connected to rest of the topology with only single link i.e R1will be cut off to rest of the topology if the link is down.

Now, say the link between R1 and R2 goes down.

As a result, routers R1 and R2 are the source to know this change in the topology. Rest all other routers will be able to get this information only when R1 and R2 share their routing information with them. The router R2 updates this change in topology in its routing information.

Since R3 is unaware of the link down between R1 and R2, R3 has the information that R1 is reachable with a cost of 2 via R2 (1 for R3 to R2 and 1 for R2 to R1), as it is not aware of the link break between R1 and R2.

Routers share their routing information after regular time interval, resulting the router R2 receives router R3's routing information.

When R2 receives R3's table, it assumes that R1 is reachable via R3 and it updates its routing information with infinity to 3 (1 for R2 to R3 and R3 shared the cost as 2 for reaching to R1).

Further, after certain time interval when routing information is shared again among routers.

The node R3 receives R2's routing information, it sees that R2 has updated the cost to reach to R1 from 1 to 3, so R3 also changes its routing table with cost 4 to R1(as R3 discovered the node R1 very first time from R2 only).

This process continues until all routers reaches to the cost of link to A is infinity. To stop this loop route poisoning technique is used, in which a number is considered as the limit of the distance for a path, beyond that cost the path is considered as unreachable and it is called as Route Poisoning. For RIP routing protocol infinity is defined as 16, that is beyond the cost value 16 the path is considered to beunreachable.

This situation is shown in table below (each entry shows [distance, predecessor node])

	R2	R3	R4
Sum of cost to R1 after link down	∞, R2	2, R2	3, R3
Sum of cost to R1 after 1st updating	3, R3	2, R2	3, R3
Sum of cost to R1 after 2nd updating	3, R3	4, R2	3, R3
Sum of cost to R1 after 3rd updating	5, R3	4,R2	5,R3
Sum of cost to R1 after 4th updating	5, R3	6, R2	5, R3
Sum of cost to R1 after 5th updating	7, R3	6, R2	7, R3
Sum of cost to R1 after nth updating	....	....	....
∞	∞	∞	∞

In the above table it is understood that the network will not be able to converge ever. The root cause of this issue is the sharing of routing information with the node (R2) from which it (R3) first discovered that node (R1).

As discussed above one of the possible resolution of this problem is Rout Poisoning and another resolution is with Split Horizon.

In Split horizon Rule says that, the information about the path for a destination (say for R1) is never sent back in the direction from which it was received i.e. R3 discovered node R1 through R2 so, R3 will not send back the same path information which was received from R2 about R1 to R2.

2.5 LINK STATE ROUTING

The Distance Vector routing algorithm is driven by the sharing of self routing information with neighbours which leads to routing challenges as the count to infinity problem. In DV routing algorithm rumors can be spread in a many fold speed and strength.

For these reasons, a new routing algorithm introduced namely: Link State Routing algorithm also known as shortest path first.

Link state routing approach is inspired by road navigation map. In contrast to DV approach, in LS each router has a complete view of thenetwork topology. In link state protocol each routershares information about itself, its directly connected links, and the state of thoselinks. Instead of sharing routing information (routing table containing cost) as in DV, in link state information regarding the status of the link is shared. On receiving information shared by other routers, each router keeps a copy of it and further passes it without any change in it. Each router independently computes the best route (route with minimum cost) to reach to every possible destination in the topology. That is, after convergence each router has the map (topology) of the entire network designated to it. In link state routing each router has the same routing information. When there is a change in the topology, directly affected router send the change in routing information to all routers in the topology.

Link state routing protocol maintains three tables namely: neighbor table, topology table and actual routing table to perform routing.

Link state routing protocols are the most widely used protocols in the Internet. Some of the widely used link state protocols are: Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS).

Working of the Link State Routing protocol:

Link state routing protocol can be divided into five parts as written by Tanenbaum. Each router of the topology uses link state routing protocol performs following actions:

1. Building of Neighbour table:
   In link state algorithm a special type message/packet namely: HELLO is used to discover neighbour nodes in the network. A router sends HELLO message on each of its connected link. Neighbor routers reply with their network addresses. The router uses this information and the port on which it received this information to build up its neighbor table.
2. Path cost measurement Measure to neighbour nodes:

A path cost is determined by sending an ECHO message/packet. A link state routing node on receipt of this ECHO message (with no payload, a very small size message) it immediately (with no time delay) sends back the ECHO message. The node calculates the duration starting from the sending of ECHO message to receipt of the reply. This duration is called as the round trip time for the node replied. The path cost for this destination node will be half of the round trip time with assumption that the delays are symmetric (same delay from sender to receiver and receiver to sender, which is not always true). The

path cost may be a composite metrics with factors like the end-to-end delay, throughput, or a combination of these.

3. Once the node has discovered the neighbours and their path cost it constructs a packet called as link state packet (LSP) including the link cost to these neighbours. The structure of the LSP is shown in table below. This packet is broadcasted in the network.

Advertiser ID	Network ID	Cost	Neighbour ID
…… ……	…… ……	…… … …… …	…… ……

4. Each node constructs routing information to all possible destination nodes with the help of routing information received from other nodes of the topology by applying link state algorithm like Dijkstra's algorithm.

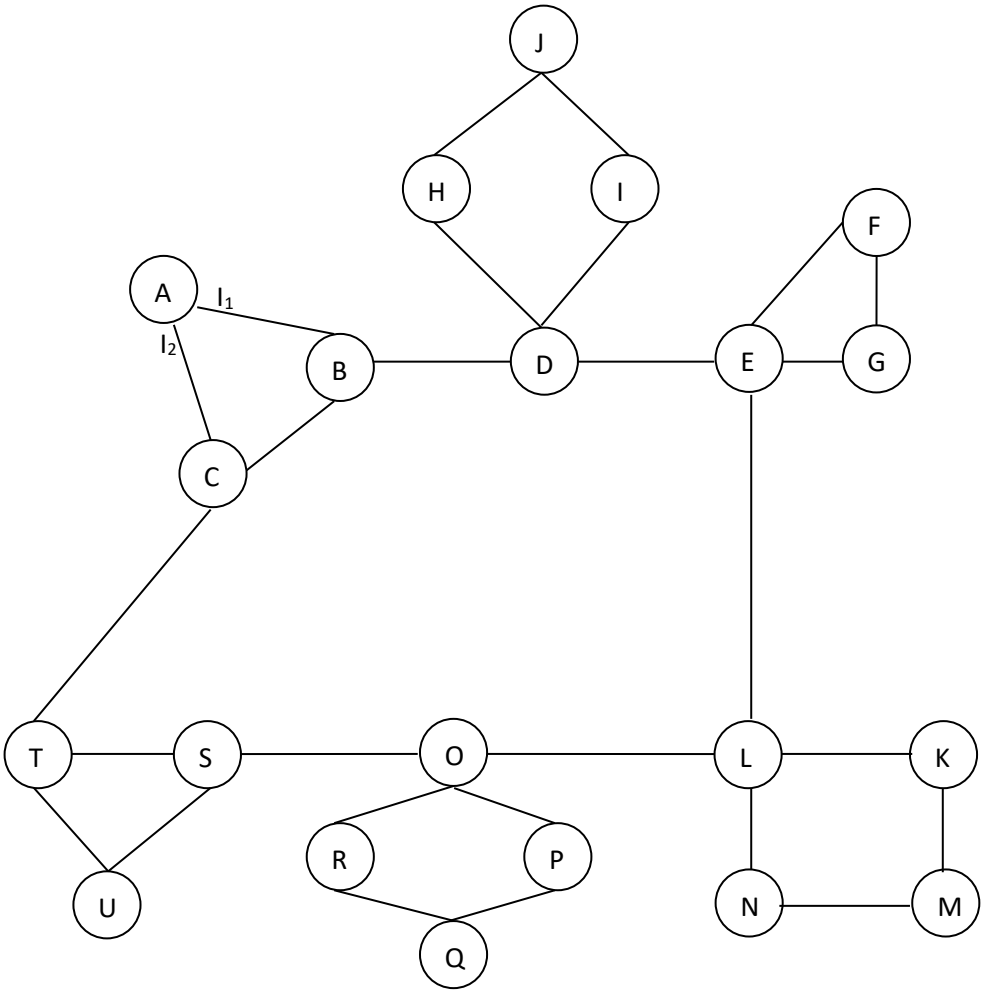
Problems in Link State Routing

One of the major drawbacks to the **link state routing protocols** is that the CPU overhead to recalculate the route due to change in topology is very high.
Another drawback is the amount of memory required to store the routing information i.e. the neighbor tables, routing table and the full map of the topology.
If a node advertises wrong neighbor information, the error is propagated to the whole topology.

2.6 HIERARCHICAL ROUTING

As discussed in link state and distance vector routing algorithms, each router has to store routing information in the form of routing table. In the routing table router stores information of path for remote networks i.e. the path cost, the exit interface.
The amount of routing information to be stored is directly proportional to the number of routers in the network. That is for a small size network with few numbers of routers the routing information to be stored by routers can be handled easily. Whereas, for a network with large number of routers, the size of routing information to be stored is highly voluminous in size. The purpose of routers is to route (finding the path) the packet to destination. As a result the routing tables will become big in size and will consume more space on router as well as more bandwidth in the network when shared.
To overcome this problem, instead of a flat structure the network can be designed as a hierarchical structure.
Considering the following example with distance vector routing algorithm node A has to store 21 entries into its routing table (considering each path

having a cost of 1 unit). Exit interface is the interface through which the destination is connected (here I_1 and I_2 are considered as the interfaces of A) .



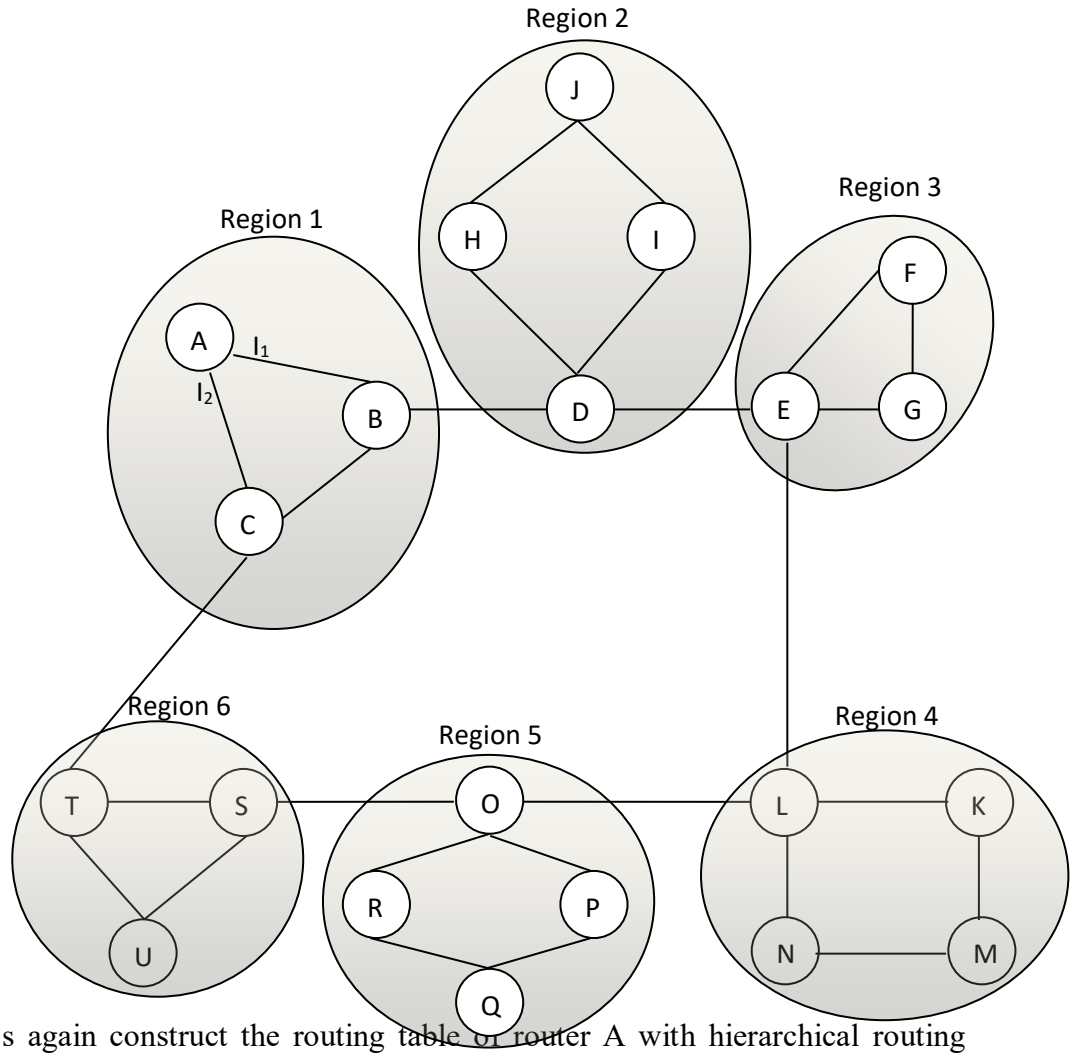
Destination	Exit Interface	Cost
A	--	0
B	I_1	1
C	I_2	1
D	I_1	2
E	I_1	3
F	I_1	4
G	I_1	4
H	I_1	3
I	I_1	3
J	I_1	4
K	I_1	5
L	I_1	4
M	I_1	6

N	I_1	5
O	I_2	4
P	I_2	5
Q	I_2	6
R	I_2	5
S	I_2	3
T	I_2	2
U	I_2	3

From table and figure above, it is visible that the traffic due to exchange of these routing tables will be high.

One possible solution of this can be, if routers are divided into small groups(called as Regions) in which they have to store routing information of the routers of the region they belongs to. A router stores only one entry collectively for all routers of a region.

In the example discussed here, the complete network can be classified into 6 regions as shown below:



Let's again construct the routing table of router A with hierarchical routing approach:

A's Routing table for Hierarchical routing

Destination	Exit Interface	Cost
A	---	---
B	I_1	1
C	I_2	1
Region 2	I_1	2
Region 3	I_1	3
Region 4	I_1	4
Region 5	I_2	3
Region 6	I_2	2

If A wants to send packets to any router in region 3 (E, F or G), it sends them to the interface I_1. From the above table it is clear that the routing table size is reduced leading to improved efficiency due to less overhead of the traffic in the network.

Hierarchical routing further can be classified into levels. In the example discussed above a two-level hierarchical routing is implemented. The level of hierarchical routing is chosen according to the size (number of routers) of the network. A three or four level hierarchical routing can also be used. In a three-level hierarchical routing, the network is classified into a number of *clusters*. Where, each cluster contains a number of regions, and each region contains a number of routers. In Internet at a wide scale commonly hierarchical routing is used.

2.7 THE INTERNET PROTOCOL (IP)

2.7.1 IPV4 addressing

IP addresses are used to uniquely identify and locate any system connected in the Internet i.e. two networked systems cannot be assigned identical IP address (although private IP addresses are reusable among private networks, will be discussed later in this chapter). Internet Protocol version 4 (**IPv4**) is the 4thversion of the Internet Protocol (IP). IPv4 uses a 32-bit address space with total number of 2^{32} unique IP addresses, but from these large number of IP addreses are reserved for special purpose in networking. While performing routing, IPv4 addresses are used by routers. Some of the IPv4 addresses are reserved (as shown in table below) for private networks and multicast addresses and for future/ scientific purpose.

Address range	Reserved as	Description
10.0.0.0– 10.255.255.255	Private network	These addresses are assigned to systems connected within a private network.

Address range	Reserved as	Description
127.0.0.0–127.255.255.255	Host	Used for local host or the loopback addresses.
169.254.0.0–169.254.255.255	Subnet	Assigned to hosts connected with a link directly, when there is no other device to allocate the IP address. Known as link local address.
172.16.0.0–172.31.255.255	Private network	These addresses are assigned to systems connected within a private network.
192.168.0.0–192.168.255.255	Private network	Used for communications within a private network.
224.0.0.0–239.255.255.255	Multicast	Used to send message to a group of hosts.

Address representations

IPv4 addresses are commonly written in dot-decimal notation. In this 32-bits are divided into 4 octets separated by periods(.). In dot-decimal notation each octet is written in decimal format.

For example, IP address *172.16.1.32*. For computing purpose, sometimes it is convenient to use binary notation of IPv4 addresses.

The 32 bits of the IPv4 addresses are divided into two parts: network portion and host portion. Five classes of IPv4 addresses are defined

Private networks

Private IP addresses are used in private networks managed by single authority. Private IP addresses are reusable among private networks i.e. private IP address used in one private network can be reused in another private network. Private IP addresses are not routable in the public Internet; that is they are not recognized by public routers. Therefore, hosts with private IP address cannot communicate with public networks directly, there is a need of network address translation (NAT) system for this purpose.

IPv4 address classes:

IPv4 address range is classified into five classes; A through E. These classes are identified by the first octet of the IP address. The details are as shown in table below:

	1st Byte	2nd Byte	3rd Byte	4th Byte	Description
Class A	0 to 127 (in decimal) 00000000 to 01111111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The 1st MSB bit of the 1stoctet is always set to 0 (zero) (as shown in red color)
Class B	128 to 191 (in decimal) 10000000 to 10111111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The first two MSB bits of the 1stoctet are always set to 10 (as shown in red color)
Class C	192 to 223 (in decimal) 11000000 to 11011111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The 1st three bits of the 1stoctet are always set to 110 (as shown in red color)
Class D	224 to 239 (in decimal) 11100000 to 11101111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The 1st four bits of the 1stoctet are always set to 1110 (as shown in red color)
Class E	240 to 255 (in decimal) 11110000 to 11111111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The 1st four bits of the 1stoctet are always set to 1111 (as shown in red color)

2.7.2 Datagram Format

The data unit of IP is named as packet. For each of the IP packet, control information is added which is used by intermediate nodes to make it delivered successfully to the destination and also used by end to end nodes for confirmation of correctness of the message. This control information is added at the starting of the content called as header of the packet. An IP packet has two sections: a header section (with control information of IP) and a data section (payload handed over by upper layer).

Header **of IP packet**

The header of the IPv4 packet consists of 14 fields, of which first thirteen field are necessary to be included and the last 14thfield is (options) optional to add. The header of the IP packet is formed as big endian format (most significant byte first). The most significant bit is numbered 0. The version of the IP protocol is the first field of the header (four bits of the 1st byte). The structure of the IP header is shown in figure below.

Offsets	Octet	0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	0	Version				IHL				DSCP						ECN		Total Length															
4	32	Identification																Flags			Fragment Offset											
8	64	Time To Live								Protocol								Header Checksum															
12	96	Source IP Address																															
16	128	Destination IP Address																															
20	160	Options (if IHL > 5)																															
24	192																																
28	224																																
32	256																																

Version

The first field of the IP header the protocol version of the Internet Protocol (IP). This field is of four bits length. For IPv4, as it is the 4th version of IP so version field contains 4 (0100).

Internet Header Length (IHL)

As the header of the IPv4 packet is not of the fixed size due to the 14th field options, it is necessary to include the size of the header so that the receiver is able to separate the header fields from the payload of the packet. This field is of 4 bits. The minimum value for IHL is 5 and the maximum is 15. The value of this field is calculated by multiplying IHL value with 32 and the result will be in bit, i.e. IHL field value 5 means: 5 x 32 bits =160 bits = 20 Bytes. The maximum value of IHL can be 15 (4 bit length), that is the maximum size of the IPv4 header can be 15 × 32 bits = 480 bits = 60 bytes.

Differentiated Services Code Point (DSCP)

This field is used to specify the type of service (ToS) for the packet in transmission. At present this field specifies differentiated services (DiffServ). This field is commonly used by the real-time data streaming applications. An example is Voice over IP (VoIP), which is used for interactive voice services.

Explicit Congestion Notification (ECN)

This field is used to provide end-to-end congestion control mechanism to avoid dropping of packets.

Total Length

This field is of the size 16-bits. This field defines the size of the entire packet in bytes, that is header and data. This field can take a value between 20 bytes (only header with no data) and 65,535 bytes.

Identification

This field is used for the purpose of identification of the packets for uniquely identifying the group of fragments (breaking the packet into smaller size due to constraints of routers or the network links) of a single IP datagram.

Flags

A total of 3 flags are defined each with 1 bit. The purpose of these flag values is to control or identify fragments. These flags are as follows:

- bit 0: Reserved; must be zero. [most significant bit of flag field]
- bit 1: Donot Fragment (DF)
- bit 2: More Fragments (MF) [least significant bit of flag field]

A packet can be fragmented iff, the DF flag is cleared (assigned a value 0). If the DF flag is set (assigned a value 1), and the size of this packet is more than

than the MTU (Maximum Transferable Unit]) value, that is fragmentation is required, the packet will be dropped.

The MF field denotes that there are more fragments available after this one of the original packet. MF flag is 0 (zero) for unfragmented packets, and the last fragment of a packet. The last fragment of a packet has a non-zero Fragment Offset field, differentiating it from an unfragmented packet.

Fragment Offset

This filed is of the size 13bits. Fragment offset of a fragment is measured in units of 8 byte blocks. This field represents the position of a particular fragment with respect to the beginning of the unfragmented (original) IP packet. The first fragment has an offset of zero. The maximum value of this field can be $(2^{13} – 1) × 8 = 65,528$ bytes, which would exceed the maximum IP packet length of 65,535 bytes with the header length included (65,528 + 20 = 65,548 bytes).

Time To Live (TTL)

Under some circumstances many of the times packets got stuck in loops in the network and consumes the bandwidth of the network unnecessarily. The size of this field is 8 bit. This field restricts packets to enter to live infinitely long time in the Internet. This value is in seconds, but time intervals less than 1 second are rounded up to 1. The value of this field is set to a number (known as maximum hop count limit). When the packet arrives at a router, the TTL field value is decrement by one. The packet is dropped by the router if it encounters TTL field value to be 0 (zero).

Protocol

This field contains the protocol used at the upper layer (transport layer) in the data portion of the IP datagram.

Header Checksum

IP supports error-checking of the header. On arrival of the packet at a router, the checksum is calculated again of the header and compared with the checksum field of the header. If both values do not match, the router discards the packet. On each of the intermediate router the TTL field value is decreased by one so the router must recalculate the checksum value of the header.

Source address

IPv4 address of the sender of the packet is included in this field. The size of this field is 32 bit. This field is the IPv4 address of the sender of the packet. If the sender belongs to the private network (having private IP address), this has to be changed in transit by a network address translation (NAT) device.

Destination address

IPv4 address of the receiver of the packet is included in this field. If the receiver belongs to the private network (having private IP address). If the receiver belongs to the private network (having private IP address), this has to be changed in transit by a network address translation (NAT) device.

Options

In general this field is not used while forming IP packets.

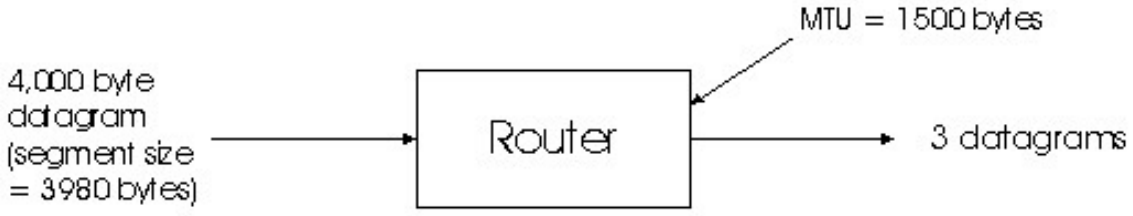
2.7.3 IP Datagram Fragmentation

- Each IP datagram is supposed to be encapsulated within the link-layer frame for transportation from one router to other. In the Internet different types of links are used to connect various routers operating

with differentlink layer protocols. Each link layer protocol sets the limit on length of an IP datagram known as MTU (maximum transferable unit). MTU: Maximum amount of data that a Link-Layer frame can carry.

When a packet arrives at the router, its destination address is examined to get the outgoing link on which it has to be forwarded. Once the outgoing link is identified its MTU is determined. If packet size is more than the MTU value, and the IP packet header flag field 'Do not Fragment (DF)' value is 0 (zero), then the router fragments the packet into smaller parts and sent one by one on the link. The allowed maximum size of any fragment can be MTU value minus the IP header size i.e. it ranges from 20 bytes to 60 bytes).

Considering the following example, the MTU of the exit link is 1500 bytes. A datagram of the size 4000 bytes with identification number 777 and DF flag bit set to 0 is received (given that including 20 bytes of transport layer header).



Here, MTU size is 1500 bytes that is the maximum size of the packet (header (20 bytes) + payload 1480 bytes can be allowed on the link.
So, for the 1st fragment:

- Payload/ data in the packet =1480 bytes
- offset = 0 (data of this packet should be inserted at byte 0 at the time of reassembling)
- identification number = 777
- MF flag value= 1 (there are more fragments after this fragment of the original packet)

2nd fragment

- Payload/ data in the packet =1480 bytes
- offset = 1480 (data of this packet should be inserted after byte 1480 at the time of reassembling)
- identification number = 777
- MF flag value= 1 (there are more fragments after this fragment of the original packet)

3rd fragment

- Payload/ data in the packet =1020 byte (=3980-1480-1480) information field
- offset = 2,960 (data of this packet should be inserted after byte 2960 at the time of reassembling)
- identification number = 777
- MF flag = 0 (this is the last fragments of the original packet)

Reassembly of the fragmented parts of the packet is performed only at the end system, routers are not allowed to reassemble the fragments in transit.

2.7.4 IPv6

Due to increase in users of the Internet, IPv4 addresses are exhausted. Hence, the size of IP address required to be increased to accommodate all the users of the Internet. IPv6 is the 6th version of Internet protocol and the size of the IPv6 address is 128 bit. Interoperability between IPv4 and the IPv6 is not provided, and thus shifting to IPv6 was not easy at all. There was a need of intermediate system which sits in between these two protocols and acts as a convertor for both. Several transition mechanisms have been proposed to make the communication between these two protocols possible.

IPv6 not only provides a large addressing space but also permits hierarchical address allocation methods that facilitate route aggregation across the Internet, and thus limit the size of routing tables even in a very large network.

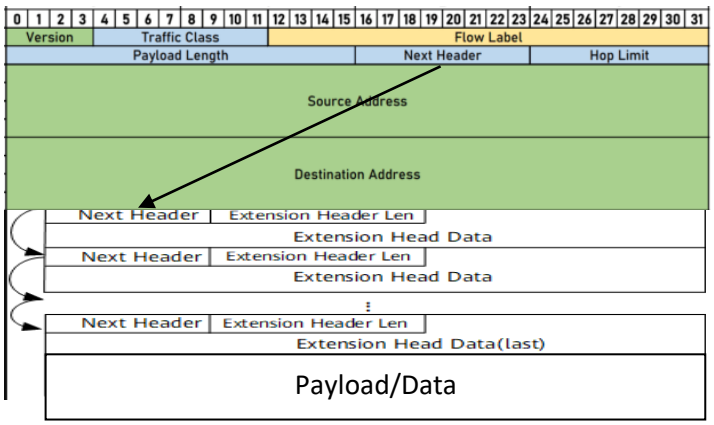
Address Representation

IPv6 addresses are represented in hexadecimal format. 128 bits of the address are divided into eight groups, separated by colons, each of size 16 bits represented into 4 hexadecimal digits. Example of IPv6 address is 2001:0db8:0000:0000:0000:8a2e:0370:7334. Further, the IPv6 address can be shortened by omitting continuous 0(zero) groups and placing double colon (::) instead. The leading 0 (zeros) in a group can also be omitted. For example, 2001:0db8:0000:0000:0000:8a2e:0370:7334 address can be written as - 2001:db8::8a2e:370:7334.

The host portion of IPv6 address is fixed of the size 64 leaving remaining 64 bits for the subnet size.

IPv6 packets

An IPv6 packet has two parts: a header and payload.



The header consists of fixed portion and the variable/ optional portion. The fixed portion is of size 40 octets and consists of eight compulsory fields to support minimal functionality required for all packets and optional extensions to implement special features.

Version field is to define the version of the IP protocol, here its value will always be 0110 (version 6). Traffic class and the flow label fields are used to provide traffic specific QoS. As the fixed header length is not variable so header length field is not necessary here. But there is need to identify the last bit of the packet, hence the field Payload length is added which tells the size of the payload (total size = header size + payload size). After the header either the optional header fields will be inserted or the payload is inserted. Next Header field is used to help the receiver to interpret the data which follows the header. The "Next Header" field of the last option points to the upper-layer protocol that is carried in the packet's payload.Without options, at max 64kB of the payload can be inserted.

In contrast to IPv4 packet, IPv6 packets are not allowed to be fragmented in transit by routers. Hop Limit field of size 8 bits is used to protect looping of a packet in the Internet.

2.7.5 Internet control message protocol

Internet control message protocol (ICMP) is a network layer protocol. Since Internet Protocol(IP) does not have a built-in mechanism for sending error and control messages, a separate protocol namely: ICMP is used to support error control. It is responsible to report errors and management queries. ICMP is a implemented in the networks devices like routers for sending the error messages and operations information. e.g. host is unreachable. Another use of the ICMP is in congestion control in the network and flow control between sender and receiver by sending a request to decrease traffic rate. ICMP packet is also used in defining the TTL value of a path.

2.7.6 Dynamic host configuration protocol

Dynamic Host Configuration Protocol (DHCP) is a network management protocol. For a network with large number of IP based systems (Systems assigned an IP address) it is almost impractical to assign IP addresses manually. To automate this process Dynamic Host Configuration protocol was designed. DHCP server dynamically assigns an IP address and other network configuration parameters (like: default gateway, subnet mask, DNS server IP etc) to each device on a network.

2.7.7 IP Security

The IP security (IPSec) protocol is used services likeconfidentiality, integrity and authentication. IPSec protocol can be used to encrypted, decrypted and

authenticated packets. IPsec was introduced to provide encryption services for application layer data. It can also be used to provide security for routing data generated by transmitted across the public internet. Authentication services are also provided by IPSec without encryptionlike to authenticate that the data originates from a known sender. In IPv6 IPSec is made available as optional header of the IP packet. IPSec also provides facility to establish a circuit using IPsec tunneling to transmit the data in encrypted form between two endpoints.

2.8 ROUTING WITH INTERNET

2.8.1 Intra Autonomous System Routing in the Internet: RIP & OSPF

An intra-autonomous system routing protocols are responsible to provide routing capabilities to routers within an autonomous system (An autonomous system (AS) is a very large network or group of networks with a single routing policy.). Intra-AS routing protocols are also known as **interior gateway protocols. RIP** (the Routing Information Protocol), and **OSPF** (Open Shortest Path First) are the most widely used Intra-AS routing protocols.

1. RIP

RIP (Routing Information Protocol) is based on distance vector routing algorithm. RIP uses Bellman Ford routing algorithm. RIP is a proprietary routing protocol of Cisco and available with Cisco routers. In RIPv2 is capable of preventing routing loops in the network. A maximum Hop count value 15 is used for this purpose. RIPv2 is implemented with mechanism like split horizon, route poisoning and hold-downto prevent spreading of rumours and routing loops. RIP is suitable for a small size network. RIP uses UDP as transport layer protocol making it light weight protocol.

2. Open Shortest Path First (OSPF) :

OSPF is based on link-state routing algorithm. This is an open protocol, i.e anyone can use it freely. OSPF protocol uses Dijkstra's algorithm. It finds shortest path for each source to all destination pair. One of the advantages of it is that it uses multicast routing in a broadcast domain. Another advantage of it is that it can handle the error detection by itself.

- **Difference Between RIP and OSPF**

SR.NO	RIP	OSPF
1	RIP is Routing Information Protocol.	OSPF is Open Shortest Path First protocol.
2	Routing Information Protocol is based on the Bellman Ford algorithm.	Open shortest path first protocol is based on Dijkstra algorithm.
3	RIP is a DV protocol.	OSPF is a link state

SR.NO	RIP	OSPF
	Distance or hops count are the metric used in measuring the path cost.	protocol. It uses bandwidth, congestion metric while identifying the best path.
4	This protocol is suitable for small size networks.	This protocol is best suitable for large size network.
5	A maximum hop count of 15 is allowed in RIP.	Hop count restriction is not there in OSPF.
7	RIP uses User Datagram Protocol.	OSPF works for Internet Protocol.

2.8.2 Inter Autonomous System Routing protocol: BGP

An autonomous system (AS) is a network managed and controlled by a single authority. Inter Autonomous system routing protocols are responsible to route a packet among various autonomous systems. Border Gateway Protocol (BGP) is one of the inter autonomous system routing protocol which is used as the routing protocol in the Internet to exchange routing information among ISPsmanaged by different Authorities.

It can connect any network of ASirrespective of the topology used. The only requirement to connect many AS together is that each of the AS should have at least one router running with BGP. BGP's is responsible to exchange network reach ability information with other BGP systems. BGP constructs an ASs' graph based on the information exchanged between BGP routers.

2.9 MULTICAST ROUTING

In the Internet at many instances there is a need to send same information to a group of clients at the same time. In such cases if, the unicast routing is used the data has to send to individual client and the server has to connect with each client independently to send the same information leading to overburden the server as well as the network. Instead of this if broadcasting is used to send the information, even it is also a wastage of resources computing as well as networking as not all the clients are interested in sent information or sometimes the information is not supposed to be disclosed to them. Hence, in both situations broadcasting approach is not suitable.

Multicast routing algorithm is used to handle such cases (sending a message to a group of users/ clients).

In multicast routing the most important part is the group management.Under group management tasks like group creation, deletion and management of

membership to the group(like join and leave)are performed. When a host joins/leaves a group, the information is propagated to its router. A router may be a member of one group, many groups or not a member of any group. While performing routing to send a message to a group, router requires the information about the members of the group. Hence, the multicast routing algorithm has to maintain this information of group membership. This information can be maintained and propagated in two ways: either the host informs to router about their membership, or routers send a query to their hosts periodically. Each router periodically shares their group management information to theirneighbors, and like this the information propagated through the subnet.

In addition to group management, a logical spanning tree is constructed of the topology to perform the multicast routing. Once the spanning tree is constructed pruning is performed for each of the group. For there may exist more than one spanning tree of a graph, so each node will construct its own spanning tree.

Once the spanning tree is constructed it is then pruned for a group. Pruningis a technique to preserve links connecting hosts that are member of the group only.

One of the methods of pruning the spanning tree can be: starting from the last node of the path and moving towards the root, remove all routers that do not belong to the group under consideration.

Let's consider an example to understand the working of multicast routing.

Here, there are two groups 1 and 2. Some of the hosts belong to group 1 and some to group 2 and some belongs to both 1 and 2 simultaneously. Here to perform multicast routing, each router constructs a *spanning tree* covering all otherrouters. *Figure below* shows one of the spanning trees for the router R1.

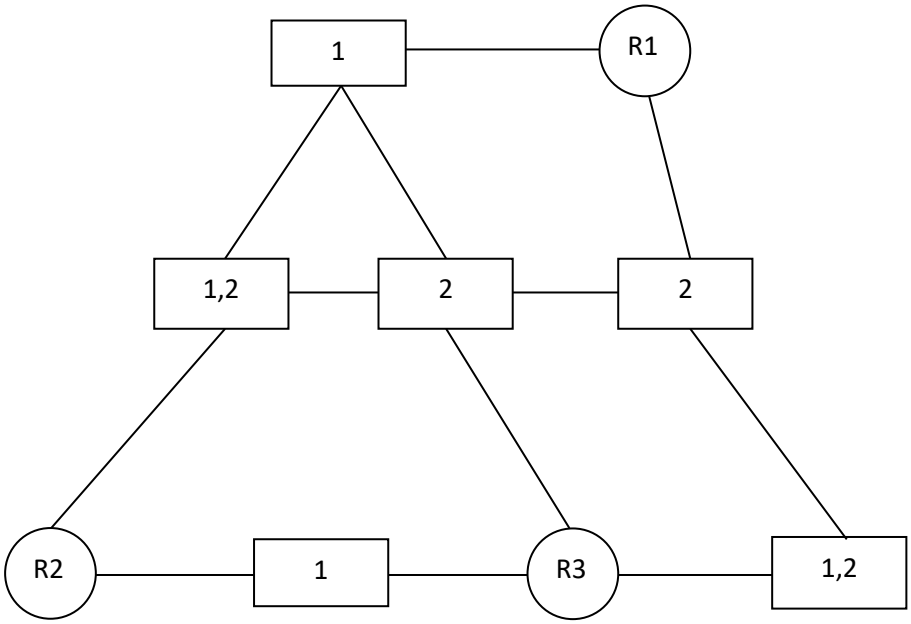
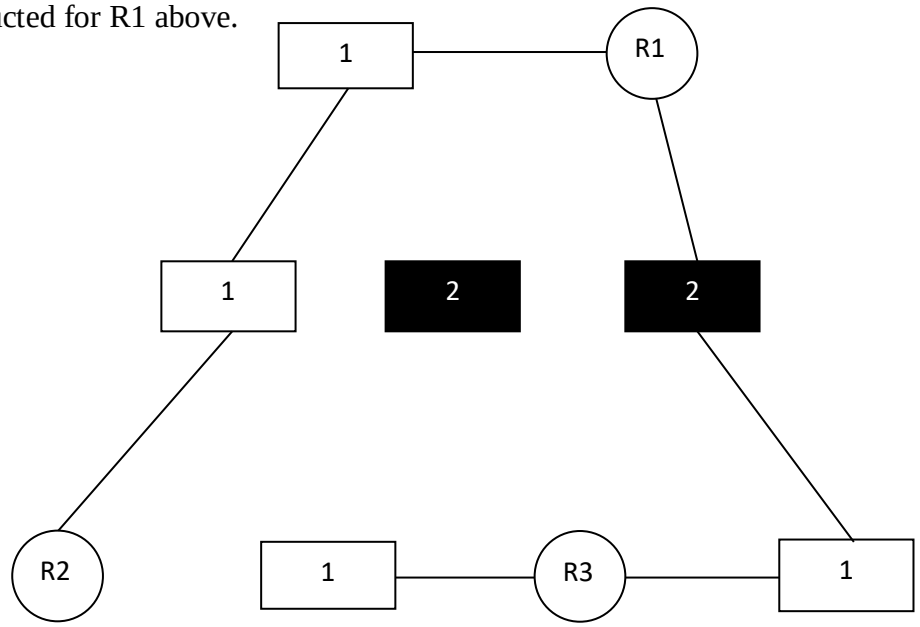
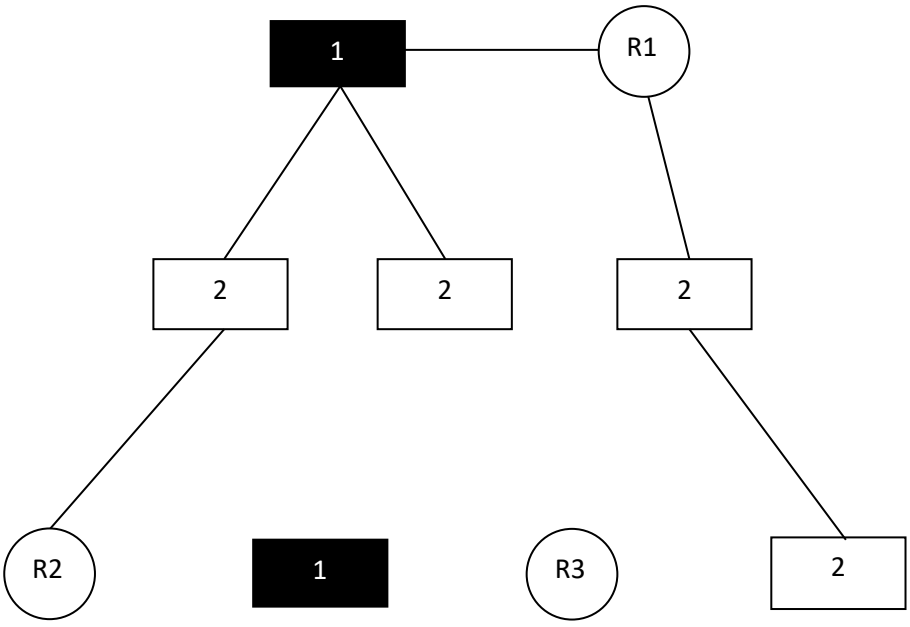


Figure below shows the pruned spanning tree for group 1 for the spanning tree constructed for R1 above.



Similarly, the pruned spanning tree for the group 2 of spanning tree of router R1 is as shown in figure below:



After pruning is completed, the multicast packets are forwarded only along the corresponding spanning tree. As the basic requirement of this algorithm is to store separate pruned spanning tree for each member of every group. Hence this method is not suitable for large networks.

2.10 MOBILE IP

Increasing the number of mobile devices with Internet access leads to invent the new modified protocol for these devices namely: Mobile IP. This protocol is designed by extending standard Internet Protocol. It is designed by keeping in mind the mobility of the devices and it provides the ability to users that they canswitch to another network with the same IP address without dropping out the connection.
This protocol allows location-independent routing of IP packets throughout the Internet. A mobile device is always recognized by the home address assigned irrespective of its current location in the Internet.

2.11 SUMMARY

In this unit we have learnt about the routing of a packet. That, how a packet reaches to destination from the source following the best route. Shortest path routing is simple to understand and implement hop count based routing approach. It is static in nature, means a graph is constructed by the source node for all the destination nodes in the network. Dijkstra's algorithm is one of the widely used shortest path approach based routing algorithm. It is also known as the greedy approach based routing algorithm. Distance Vector routing algorithm is another solution for routing in the network. DV approach is dynamic in nature that individual node constructs a complete map of the network topology. Each node constructs the routing table with cost and the vector component for each of the remote network. Distance vector routing algorithm is based on the Bellman-Ford equation. Distance Vector based approach face the count to infinity problem which is addressed by implementing split horizon and route poisoning together. Above both methods of routing are not suitable in a large network due to huge traffic generated by routers to exchange routing information. Hierarchical routing is one of the possible solutions to perform routing in large networks with huge number of routers. Complete network is divided into smaller sub-networks in the form of hierarchical network. Further,we have learnt about Internet protocol and its two versions: IPv4 and ipv6. The address space of IPv4 is of the size 32 bits and that is of IPv6 is of 128 bits. IPv6 also include optional headers fields. One of the features provided by IPv6 as optional header is the IPSec. ICMP and DHCP are major network management protocols. ICMP is used for many services like, congestion control, flow control, network diagnosis etc. DHCP is responsible for assigning IP address, subnet mask, and gateway and DNS information to the clients in a network. It is very difficult to manage this information manually so managed efficiently by inserting a DHCP server.

2.12 SOLUTIONS/ANSWERS

Review Questions:

1) Which of the following statements are True or False.

A	Distance vector routing is a static routing algorithm.	T	F
B	Dijkstra's algorithms can be run locally in link state routing to constructthe shortest path.	T	F
C	Flooding discovers only the optimal routes.	T	F
D	Flooding generates lots of redundant packets.	T	F
E	In hierarchical routing, each router has no information about routers in other regions.	T	F
F	A spanning tree is a subset of a graph that includes some of the nodes of that graph.	T	F
G	Sending a message to a group of users in a network is known as broadcasting.	T	F

2) What are the problems with distance vector routing algorithm?

3) Explain the spanning tree.

Solution:

1)

A. False, B. True, C. False, D. True, E. True, F. True, G. False

2) Distance vector routing algorithm faces the count to infinity problem. The convergence is slow. Routing information isexchange among direct neighbors only. Chances of rumors about false routing information are always there in distance vector routing. Due to which a packet may enter into routing loop.

3) A spanning tree is a tree with no cycles and constructed such that all the vertices are covered with minimum possible number of edges.

2.13 FURTHER READINGS

1. *Computer Network*, S. Tanenbaum, 4th edition, Prentice Hall of India, New Delhi 2002.

2. *Data Network,* Drnitri Berteskas and Robert Galleger, Second edition, Prentice Hall of India, 1997, New Delhi.

3. *Data and Computer Communication,* William Stalling, Pearson Education, 2nd Edition, Delhi.

UNIT 3 CONGESTION CONTROL ALGORITHMS

3.0 INTRODUCTION

In the Internet nodes acting as transmitting nodes are inserting packets into the Internet and nodes acting as receiving nodes consume the packets from the Internet. Internet has a capacity to handle the traffic load (packets). When the rate of insertion of packets into the Internet is higher than the rate of consumption of the packets from the Internet at last Internet is unable to handle the traffic and the performance of the resources of the Internet is degraded. This situation is termed as congestion.

Hence, the goal of congestion control algorithms is torefrain the transmitter from inserting packets in the network more than the handling capacity of the Internet.

In this unit section 3.3 discusses about the reasons for congestion in the network. Section 3.4 differentiates congestion control from flow control. Congestion prevention mechanisms are covered in section 3.5. In section 3.6 general principles of congestion control are elaborated. Further in section 3.7 congestion control technique namely: Open loop control is discussed. Section 3.8 is about the congestion control in packet-switched networks. Section 3.9 summarizes the unit. Problems and their solutions covering the entire unit are discussed in the section 3.10. Section 3.11 enlists further readings.

3.1 OBJECTVIES

After completing this unit, one should be able to:

- Identify the reasons of the congestion in the Internet;
- Differentiate congestion control and the flow control;
- Devise the preventive measures of the congestion;

- understand the congestion control techniques;
- understand the close loop and open loop congestion control techniques, and
- understand congestion control in packet-switched networks.

3.2 REASONS FOR CONGESTION IN THE NETWORK

In the Internet there can be several reasons to occur the congestion. When many transmitters insert data packets on to input lines at a time and are to be sent on the same output line, assuming that the capacity of the output line is much less than that of the packets received then a long queue will be build up for that output line. In this situation if the buffer memory if not big enough to hold all these packets, then extra packets will be dropped. To stop dropping of the packets, if the memory available is made infinitely large even then the congestion may be reduced but the overall quality of the service of the traffic will be worse; because by the time packets reach to the output line to get dispatched, there TTL (time to live) value gets expired and their duplicate packets been already inserted into the network. If all the packets carried to the final destination, these duplicate packets will increase the traffic load only in the Internet and will be discarded, due to time out. So, it will be good for the Internet to drop these packets as soon as their TTL value gets expired.

Another reason of the congestion in the Internet is the slugging performance of the processors of the intermediate devices. If any of the intermediate router's CPU is performing slower than expected speed, their jobs (i.e.Queuing buffers, routing packets, updating tables, reporting any exceptions etc.), will be slowed down. The arrival rate of the packets at input line is greater than the processing and removal of the packet from output line. This again creates a situation of congestion.

Another point of issue is the LowBandwidth. Due to low bandwidth capacity of the lines amount of the traffic increases in the network causing congestion.

Resolution of any one of the issues discussed above will not handle the congestion;instead it will just shift the bottleneck to some other point. The root cause of the real problem is the mismatch of the capacity (computing or the carrying) of various components of the system. Once congestion happens in the network the routers respond to overloading by simply dropping the packets.

The bursty nature of traffic is one of the major causes of congestion. This could be controlled by restricting the insertion of the traffic at a uniform rate.

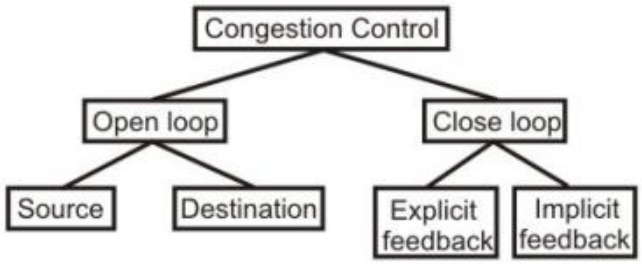
3.3 CONGESTION CONTROL VS. FLOW CONTROL

Congestion control and flow control are two different things, which are mixed up at times. As discussed earlier congestion control is the entity of the network whereas flow control is about regulating the transmission of data between devices on the connection/link between them, and not what is happening in devices between them.

Flow control is about point-to-point traffic control between sender and receiver for a specific transmission to avoid packet drop at receiver. If the incoming traffic rate is higher than the processing rate at receiver, the receiver is flooded and packet will be dropped. To overcome this situation the receiver should send some kind of feedback to sender to inform the sender about the drop of packets and to slow down the sending speed. This is called flow control between sender and receiver and is handled at transport layer responsible for end-to-end data delivery. Congestion is a situation when the traffic in the network is higher than the handling capacity of the network.

3.4 CONGESTION PREVENTION MECHANISM

Congestion control is to restrict the traffic load below the handling capacity of the network. As shown in Figure below, the congestion control techniques can be broadly classified into two categories:



• Open loop: Methods to prevent or avoid congestion are classified as open loop techniques. Open loop methods ensure that congestion state never exists in the network. Open loop policies are applied in the network to prevent congestion before it happens. The congestion control policies are applied either at the source or the destination.

• Close loop: these methods acts to treat or alleviate the congestion once it happens. Once the system enters to congestion state, closed loop techniques used to detect it, and then take action to bring the system out of it.

Open Loop solutions are static in nature. These policies are not adaptive in nature and do not change according to the present state of the system. These methods take decisions about when to accept packets, when to dropthem etc. These methods make decision without taking into consideration the present state of the system. The open loop congestion control methods are further classified on the basis of whether these are applied on source or on destination.

Close loop congestion control techniques are based on the concept of feedback. These techniques are dynamic in nature and actions are taken during transmission. Some system parameters are continuously measured in the network and whenever a congestion state is observed, feedback system is used to take action to reduce the congestion. Open loop techniques work as per the following 3 steps:

Step 1: Continuous monitoring of the network to detect the congestion state, the actual location of the congestion and devices involved.

Step 2: Sending thefeedbackabout congestion state to the devices where actions can be taken

Step 3: Take the necessary actions to remove the located congestion.

Some of congestion control algorithms based on these techniques are discussed in following sections.

3.5 GENERAL PRINCIPLES OF CONGESTION CONTROL

Congestion in the network can be measured in terms of various Metrics like: the average queue length, timed-out packets, delay, packets dropped due to unavailability of buffer space, etc.

As discussed previously, congestion occurs in the network when senders insert packets more than the handling capacity of the network. The responsible entities for the congestion in the network are: the sender: sending packets without considering the status of the network, the intermediate resources: speed of the router, bandwidth of the bottleneck link, control messages generated by intermediate device etc. The congestion in the network can be controlled by either dropping the excess packets or to restrict the sender by inserting packets into the network at a lower speed. In general TCP provides mechanism to control the congestion. Internet protocol header also provides ECN field to notify the sender about the congestion in the network. The congestion happens in the network and happened due to sender, hence the network entities are required to notify the sender about the congestion and accordingly the sender takes necessary steps to control it.

3.6 OPEN LOOP CONTROL

As discussed in previous section, open loop methods are preventive measure for congestion control.

Some of the open loopmethod based policies of congestion control are discussed here:–

Retransmission Policy :

A packet transmitted with reliable data delivery protocol, if it’s TTL value expired before it reaches to destination, gets dropped. The sender has to retransmit such packets until get delivered successfully. More the congestion in the network leads to more packet drops which leads to retransmission of these packets leading to more traffic in the network. This retransmission leads to congestion in the system.

To prevent congestion due to this issue, value of timers used by retransmission policy must be set such that state of congestion in the network is prevented and also able to optimize the efficiency of the network.

Window Policy:

In Go-Back-N window if a packet is lost/received out of order, several packets are resent, although some packets may be received successfully at the receiver side. This

may increase the congestion in the network. Therefore, selective repeat window should be preferred instead of Go-Back-N. In selective repeat window packets dropped arethat may have been lost.

Discarding Policy :

Routers has to adapt a good discarding policy such that congestion in the network is prevented at the same time a router must attempt to discard corrupted or packets of unreliable services (i.e. UDP) by maintaining the quality of the messages with reduced number of retransmission of dropped packets.

Packets transmitted with UDP services may be discarded before the packets transmitted with reliable services i.e. TCP. The video streaming over the internet may tolerate some loss of packets while text messages may not tolerate loss of any packet.

Acknowledgment Policy :

Acknowledgment is sent by receiver to notify the sender about receipt of the packet or not. Even though the size of acknowledgement packets issmall in comparison to the data packets but still they also offer traffic load in network. In order to reduce the number of acknowledgement packets sent, the receiver should wait for the next incoming packet and if it is in sequence with the previous packet instead of sending the acknowledgement of individual packet a cumulative acknowledgement is sent for both the packets. That is sending a cumulative acknowledgement of packets received in sequence can save the bandwidth of the network.

Admission Policy :

Admission policy is applied in a network to prevent the congestion. Before allowing the traffic to enter into the network, switches first check whether the resources required are available or not. The traffic is only admitted if the available resources are more than the requirement of the traffic. The requested virtual circuit is established iff there is no chance of congestionafter reserving the resources for this.

The policies discussed above can be used to prevent congestion before it happens in the network.

3.6.1 Admission Control

In this section, how the congestion is handled in a virtualcircuit network is discussed. Admission control is a closed-loop technique, in which action istaken once network enters into a congestion state.Admission control policy is the first and very simple rule to prevent the congestion in network. It says that admit the traffic on the condition that it will not be responsible for congestion in the network. If there is congestion in the network, the first attempt to recover the network from congestion state could be to stop new packets to be inserted into the network. The thumb rule to prevent the traffic is “admit traffic only when the network can handle it without congestion”.Admission control policy is feasible and can be applied successfully in the circuit-switched datagram networks. This is not easy to identify the source of congestion hence could not be applied in packet-switched networks.

Some of the admission control systems are as follows:

Admission Control Methods:

The goal of admission control methods is to estimate the expected bandwidth requirement for the incoming traffic and determine whether this traffic can be allocated the needed bandwidth, such that congestion state does not occur. Admission control methods are widely used for real time application sensitive towards delay and jitter. Many admission control methods are available. Some admission control methods are based on mathematical calculations and statistical indicators, and others are based on measuring traffic state.

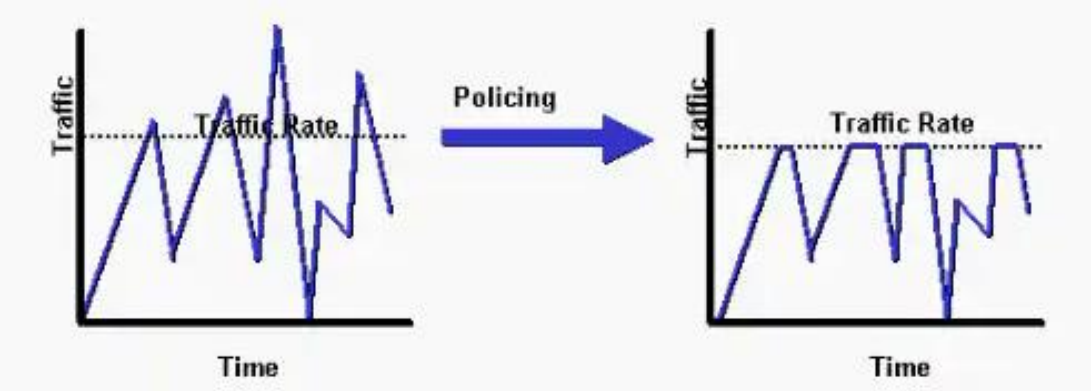
Admission control methods are generally classified into two categories: parameter based and measurement based admission control. Parameter based methods are based on the characteristics of the active traffic and do not consider new incoming traffic, hence are not optimal. Measurement based methods consider the real time network conditions by serving new incoming traffic, hence a higher network utilization may be achieved.

Each admission control method follows the principle that, allocatethe availablebandwidth tothe incoming traffic flows only in case of not exceeding the capacity of the line. For a node to implement admission control policy, it should have access to QoS parameters i.e delay, packet drop rate etc.By doing so the traffic can achieve the QoS as desired, but should be independent of the type of traffic underwent.

Similarly, in case of virtual circuit subnets, no more new virtual circuits are accepted once congestion state in the network is identified.

3.6.2 Traffic Policing and its Implementation

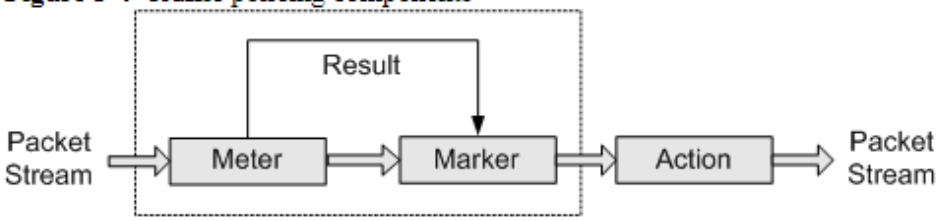
Traffic policing is to monitor the traffic flow in the network. If the traffic flow rate is greater than the specified rate, traffic policing methods simply discards the overflow packets.Traffic policing can be used to control both inbound and outbound traffic. Traffic policing methods maintain a constant flow (pre-defined) of traffic.



Traffic policing does not hold packets received above the allowed flow rate, hence does not require buffer. It is easy to implement traffic policing in comparison to traffic shaping as it does not require maintaining packet buffers. Traffic policing does not cause delay, and queuing, rather it simply discards the packets.

Components of an implementation of traffic policing system are as follows:

Figure 5-4 Traffic policing components

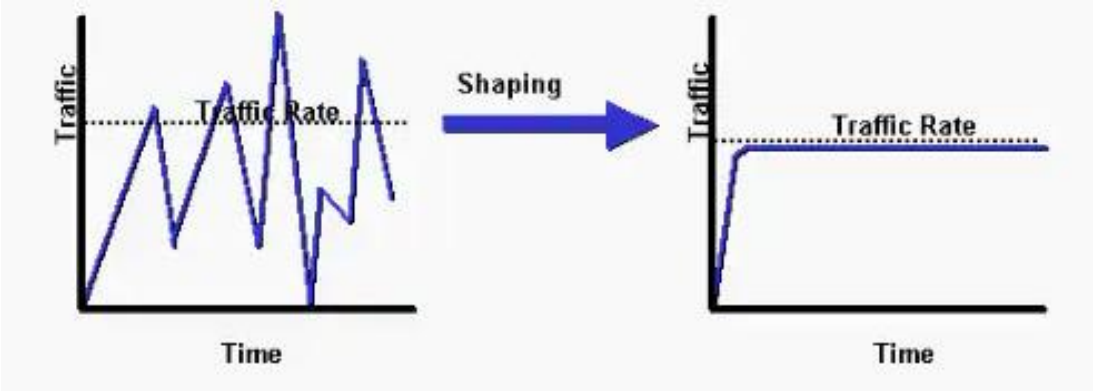


- Meter: this component measures the traffic and provides the measurement result to the next component (marker) for further action.
- Marker: marker assigns colorsto packets out of green, yellow, or red based on the measurement result provided by the meter. Marker provides this coloring information to the next component namely: Action.
- Action: this component performs actions based on packet coloring results received from the marker. This component performs following actions in accordance to the pre-defined rules:
  - Pass: a packet will be forwarded further if it meets network requirements.
  - Re-mark + pass: local priority of the packets not meets the network requirements are changed and forwarded.
  - Discard: packets not meeting network requirements are dropped.

If the rate of traffic is below the threshold value, packets are marked with green and yellow color and forwarded, whereas if the rate of traffic exceeds the threshold value, packets are either marked with yellow, lowers the priority and forwarded or marked with red color and dropped according to the traffic policing configuration.

3.6.3 Traffic Shaping and its Implementation

In contrast to traffic policing, traffic shaping tries to adjust the rate of outgoing traffic instead of dropping the packets to ensure an even transmission rate. Traffic shaping makes use of a buffer to hold bursty traffic for a while to control the traffic. Packets are delayed if the system is unable to forward all of them at a time and will be forwarded as the link is found free. It is a congestion control technique which delays some packets to remove the congestion state. Traffic shaping is not practically applicable to traffic of real time applications. Traffic shaping can be used to control the outbound traffic only.



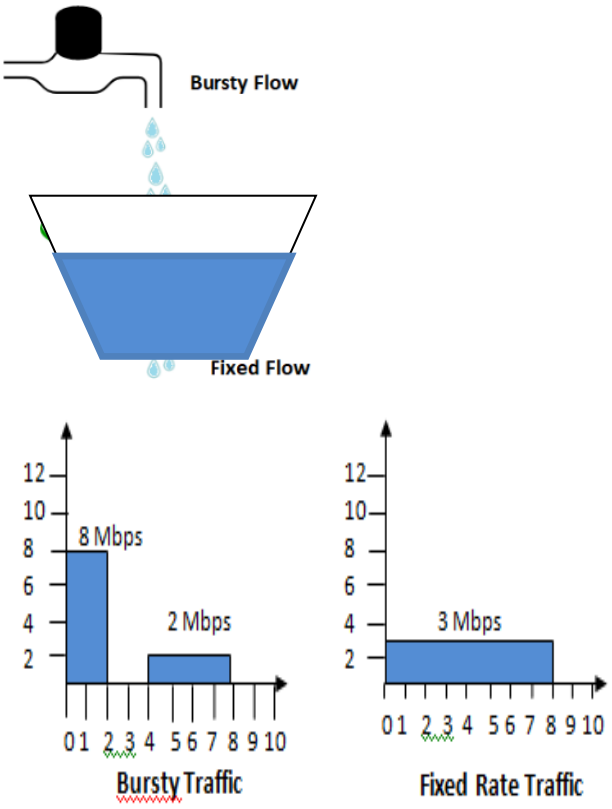
Further, Traffic shapers can be classified into two categories based on their capabilities; simple traffic shaper and advanced or more sophisticated traffic shapers. Simple traffic shapers shape all traffic uniformly. Whereas advanced traffic shaper can classify the traffic and can be used as a technique to provide Quality of Service

(QoS) to a traffic category by delayingother category of trafficto bring them into compliance with a desired traffic profile.

Two of the widely known traffic-shaping algorithms are leaky bucket and token bucket, discussed in next section in detail.

3.6.3.1 Leaky Bucket Shaper

Leaky bucket shaper as its name says, is based on the way aleaky bucket functions. It sends out the traffic at a fix rate even if the incoming traffic is bursty in nature. Bursty traffic could not be sent out at a time,, will be stored in the buffer (called the leaky bucket) and will be sent out once the outgoing line is free.



In the figure above, it is considered that the capacity of the network to carry the traffic is of 3 Mbps. The leaky bucket traffic shaper will not send traffic above 3 mbps in the network. Here, the host inserts a burst of data at a rate of 8 Mbps for 2 sec, and sends 16Mbits of data. Further, it does not send any data for next2 sec and then sends data at a rate of 2 Mbps for 4 sec, by sending 8Mbits of data. The host inserts a total data of 24Mbits in a duration of a 8 sec. After applying the leaky bucket traffic shaping policy the traffic is sent out at a rate of 3 Mbps for the duration of 8 sec. Here, traffic shaping policy smooth the traffic in the network. There are not data in the duration 2 to 3 sec, and a burst of data during the interval 0 to 2 forcing the network to congestion state. Leaky bucket policy can transmit this whole data in a smooth manner without any congestion in the network.

3.6.3.2 Token Bucket Shaper

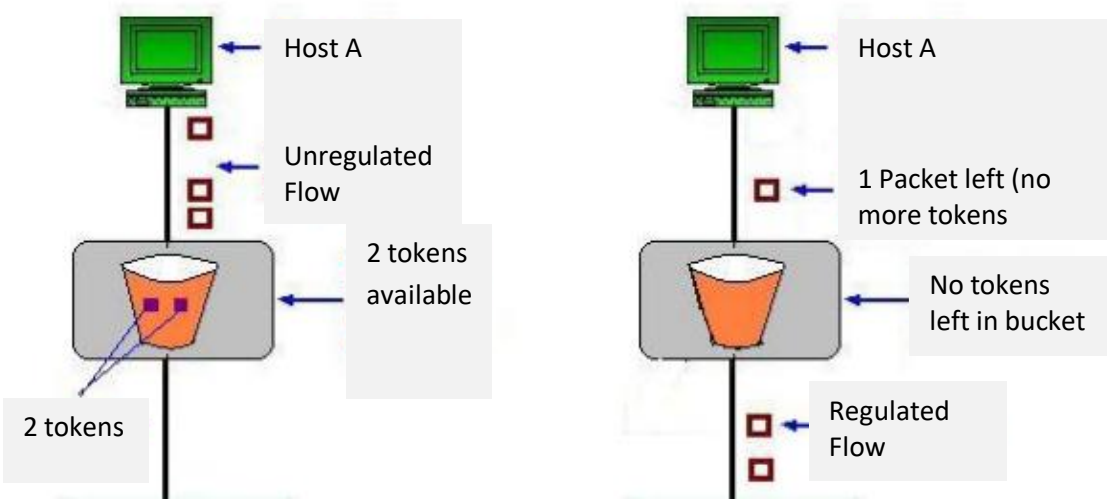
The leaky bucket traffic shaping policy discussed in previous section, does not consider the input traffic pattern. It shapes the traffic with a fixed defined rate.

Token bucket traffic shaping policy considers the input traffic bursts and allows sending the traffic on a higher rate also to prevent the drop of packets.

Token bucket policy uses the leaky bucket which holds tokens generated at regular intervals (one policy to add tokens is to generate a token per clock tick). Token bucket policy works as follows:

- Token are generated at regular intervals and placed into the bucket.
- The bucket has a maximum capacity of holding the tokens.
- A packet can be sent to the output line only if a token is available in the bucket.
- Once a packet is sent on output line, is removed from the bucket.
- As many tokens are removed from the buckets as number of packets are sent from the bucket.
- If there is no token available in the bucket, the packet cannot be sent.

Token bucket policy shapes the bursty traffic by allowing bursty traffic on output line but to a limit of available number of tokens. Figure below shown the working of the token bucket traffic shaping mechanism. In the figure host A sent 3 packets and there are only 2 tokens available in the bucket, hence only 2 of these packets are transmitted on the output line and 1 is hold back and will be sent once a token is placed in the bucket.



3.6.4 Difference between Leaky Bucket Traffic Shaper and token Bucket Traffic

Shaper

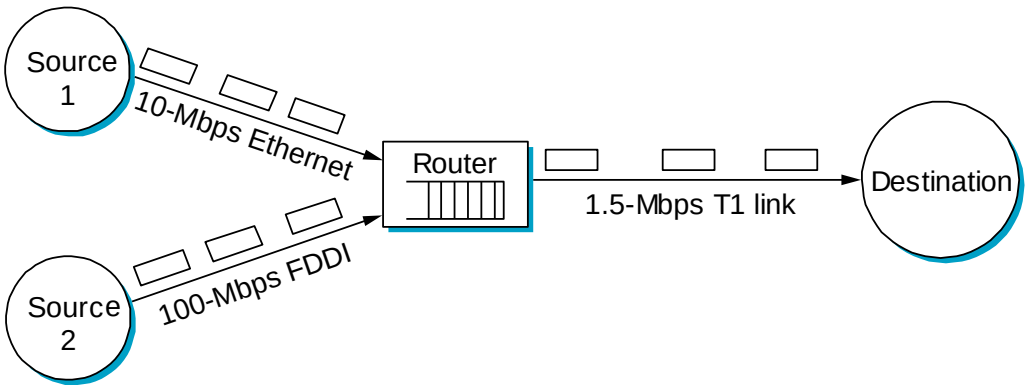
Difference between Leaky and Token buckets –

Leaky Bucket	Token Bucket
Host places the packet to be sent in the bucket.	Tokens are generated in fix intervals and placed into the bucket.
The traffic is sent onto the output link at a constant	Output traffic rate is regulated by the number of

rate	tokens available in the bucket.
Leaky bucket shapes the bursty traffic into uniform traffic.	The output traffic may be bursty (to a limit of tokens available in the bucket).
The bucket can hold a finite number of packets in a queue and outputs at finite rate	A packet can be sent only when a token is available in bucket.

3.7 CONGESTION CONTROL IN PACKET-SWITCHED NETWORKS

Congestion control is one of the very important parts to be considered while designing any packet-switching networks due to rapidly growing network and bandwidth intensive network applications. Various methods are proposed for congestion control.



In the figure above, source 1 and 2 inserts traffic at a rate of 10 and 100 Mbps respectively. The router can transmit the traffic on output link limited to 1.5 Mbps. The packets will start dropping at router once the buffer is full and the state is kown as the congestion state. Congestion control mechanism in packet switched network can be applied on either transport layer or network layer. Flow is a sequence of packets flowing between a source/destination pair and following the same route through the network.TCP provides connection oriented reliable service at transport layer while Internet protocol (IP) provides connectionless packet delivery service.Routers does not maintain any state of the flow for connectionless service whereas state of the flow is maintained at routers for the connection oriented service.**The Internet Protocol** (**IP**) provides the basis for packet delivery and **the Transmission Control Protocol** (**TCP**) provides a best-effort delivery mechanism.**Best-effort** delivery **service** is the basic packet delivery**services**without guarantee of delivering it. The best efforts are made to deliver packets to the destination, but there is no mechanism to recover lost packets. At transport layer a TCP window is used to control the transmission rate according to feedback received from the sub network. As the congestion is a network layer issue and happens in the network, the routers play crucial role in handling the congestion state. Each router is installed with certain

buffers to hold the incoming packets could not be sent at the moment due to congestion. Many policies are applied to these incoming packets to handle them in the buffer queuing. Some of the possible choices in queuing algorithms are:FIFO *also called* Drop-Tail, Fair Queuing (FQ), Weighted Fair Queuing (WFQ), Random Early Detection (RED) etc. Routers also send a special type of packet namely: choke packet for the purpose of congestion handling. Routers monitor the utilization of their output line and send choke packets back to hosts using output lines whose utilization has exceeded some warning level. Another solution frequently used to control the congestion state is Explicit Congestion Notification (ECN) used by routers at network layer to notify the sender about the congestion state.An ECN-aware router sets a field in the header of the IP instead of dropping a packet to signal about the congestion. The receiver of the packet notifies about the congestion to the sender, which reduces its transmission rate.

3.8 SUMMARY

In this section we have discussed about the congestion control state in the network. A network is congested when traffic in the network is more than its capacity to handle it. The congestion occurs when the number of packets into the network is more than its handling capacity. The bursty nature of traffic is the root cause of congestion.When part of the network no longer can cope with a sudden increase of traffic, congestion builds upon. Other factors, such as lack of bandwidth, ill-configuration and slow routers can also bring up congestion.

Flow control is an issue of data link layer whereas congestion control is an issue of network layer. Flow control is meant to prevent a fast sender from crushing a slow receiver. Flow control can be helpful at reducing congestion, but it can't really solve the congestion problem. Many congestion control techniques are applied in the network to avoid the congestion state in the network. Open loop and closed loop congestion control techniques are the broad categories of the congestion control algorithms. Traffic policing and traffic shaping are the main techniques of open loop congestion control. Traffic policing is, sending the traffic at a fix rate irrespective of the incoming traffic pattern. In contrast to traffic policing, traffic shaping tries to adjust the rate of outgoing traffic instead of dropping the packets to ensure an even transmission rate.

3.9 SOLUTIONS/ANSWERS

Q1. What is congestion?

Ans :In the Internet nodes acting as transmitting nodes are inserting packets into the Internet and nodes acting as receiving nodes consume the packets from the Internet. Internet has a capacity to handle the traffic load (packets). When the rate of insertion of packets into the Internet is higher than the rate of consumption of the packets from the Internet at last Internet is unable to handle the traffic and the performance of the resources of the Internet is degraded. This situation is termed as congestion.

Q2. Why congestion occurs?

Ans: In a packet switched network, every intermediate device maintains buffers/ queues to hold packets while processing them to forward further. Under the situations of receipt of the bursty traffic these buffers gets full and packets are dropped. As a result as per the quality of service of the dropped packets, they may require to be retransmitted, further increasing the traffic in the network. At last the system enters to congestion state.

Q3. What are the two basic mechanisms of congestion control?

Ans :Congestion in the network can be addressed in two ways: preventive method and recovery method. In preventive method,actions are taken such that congestion doesn't occur and recovery method allows the system to enter in congestion state and then it tries to remove it.

Q4. How congestion control is performed by leaky bucket algorithm?

Ans : In leaky bucket algorithm, packets are inserted into the bucket. In case of bucket overflow, packets are dropped. The packets are exited from the bucket at a constant rate allowing bursty incoming traffic into the network at a constant rate.

Q4. In what way token bucket algorithm is superior to leaky bucket algorithm?

Ans : The leaky bucket algorithm is very conservative in nature in the sense that it is not adaptive to the incoming traffic. Token bucket algorithm is made sensitive towards incoming traffic. The output rate is not dependent on the predefined upper limit rather, it depends on the availability of the tokens in the bucket. In the starting if the tokens are available in enough quantity the rate can be more and once there are no tokens available after that the output is limited by the rate of token generation.

Q5. Differentiate traffic policing and traffic shaping.

Ans. **Difference between Traffic Policing and Traffic Shaping:**

S.NO.	Traffic Policing	Traffic Shaping
1.	Traffic policing is a mechanism which monitors the traffic in any network.	Traffic Shaping is a congestion control mechanism that brings delays in packets.
2.	The packets with rates that are greater than the traffic policing rate are discarded.	It buffers the packets with rates that are greater than the traffic shaping rate.
3.	Traffic policing doesn't cause delay.	Traffic shaping causes delay of packets.
4.	The token values are calculated in bytes per second.	The token values are calculated in bits per second.
5.	In traffic policing queuing of	Queuing of traffic is not

S.NO.	Traffic Policing	Traffic Shaping
	traffic is not performed.	performed in traffic shaping.
6.	Traffic policing supports traffic remarking.	Traffic shaping doesn't supports traffic remarking.
7.	Traffic policing can be used to control outbound or inbound traffic.	Traffic policing can used to control outbound traffic only.

3.10 FURTHER READINGS

Computer Network, S. Tanenbaum, 4th edition, Prentice Hall of India,New Delhi 2002.
Data Network, DrnitriBerteskas and Robert Galleger, Second edition,Prentice Hall of India, 1997, New Delhi.
Data and Computer Communication, William Stalling, Pearson Education,2nd Edition, Delhi.