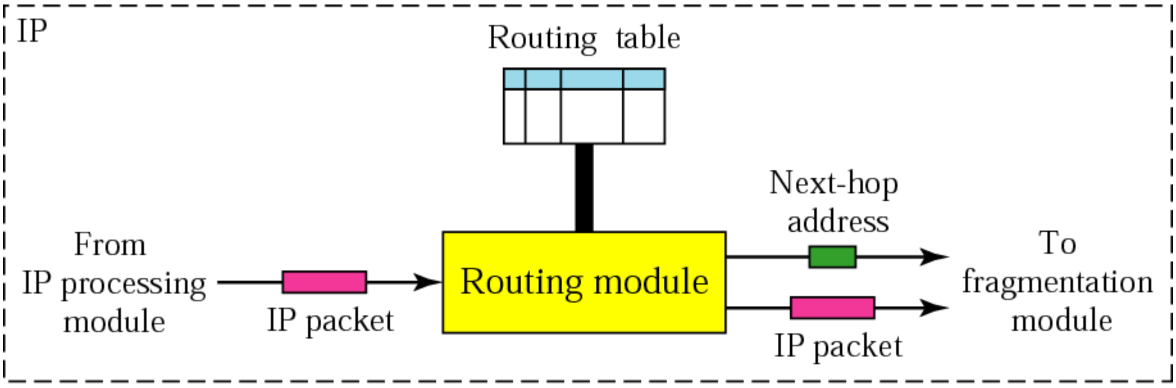
UNIT 2 ROUTING ALGORITHMS

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.

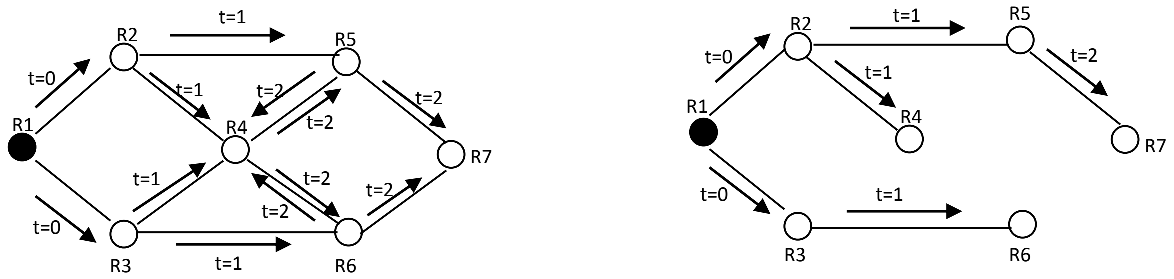
2.1 OBJECTIVES

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.



R1 to R7: Routers, Pkt: Packet, R1: Origin node

(a)

(b)

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.

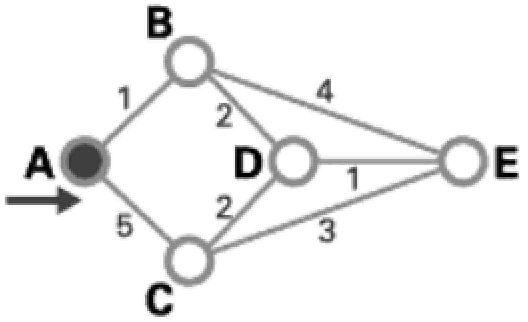
Another way to reduce the redundancy of packets in the network and to avoid the cyclic forwarding of the packets is to construct a logical spanning tree of the topology.

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 ∈ N] = Node x records distance vector

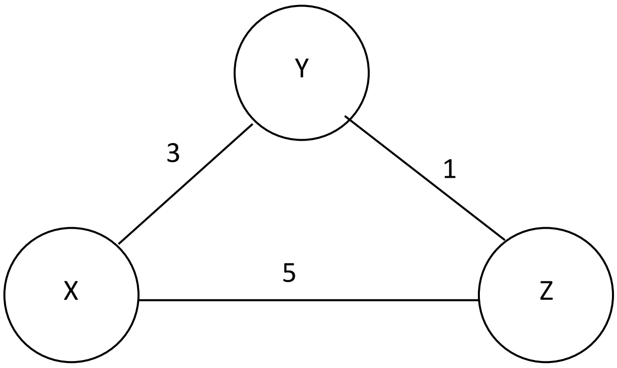
Node x also maintains its neighbours' distance vectors

Note:

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

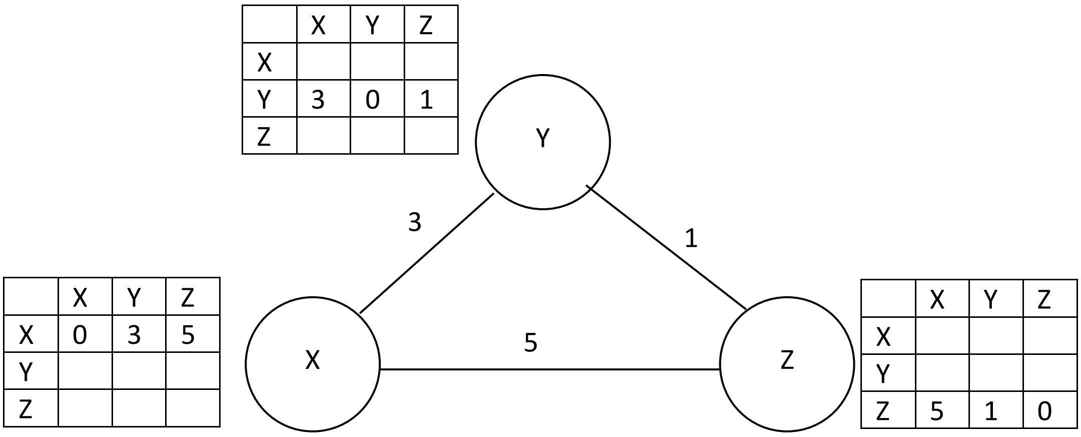
- Each node sends its own routing information/ Distance Vector information estimate to its neighbours after certain time interval.
- When any node x shared with a new Distance vector information from any of its neighbor v, it stores the distance vector of node v and updates its own routing information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

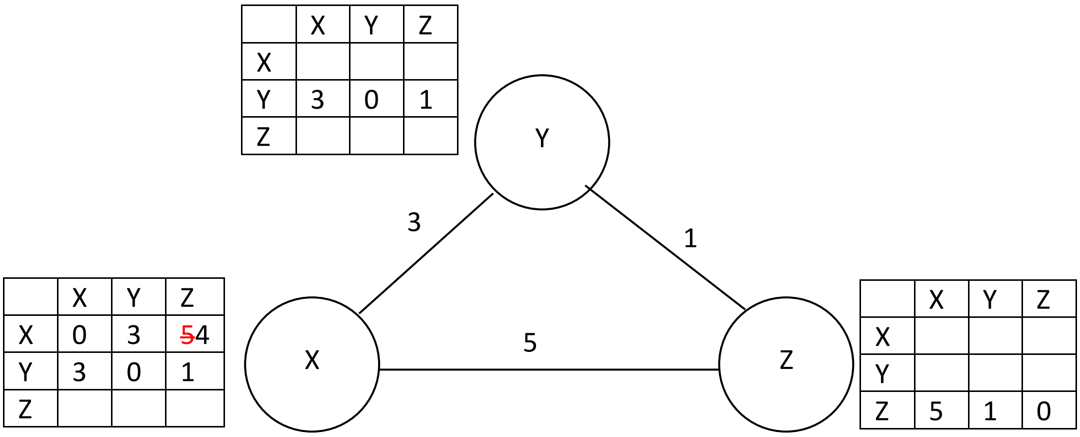
Given data:

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.

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

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



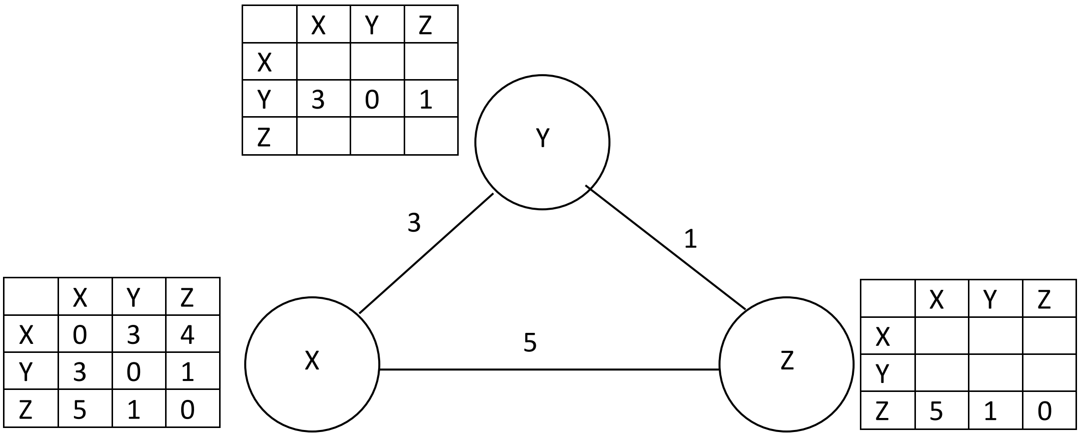
	X	Y	Z
X			
Y	3	0	1
Z			

	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			

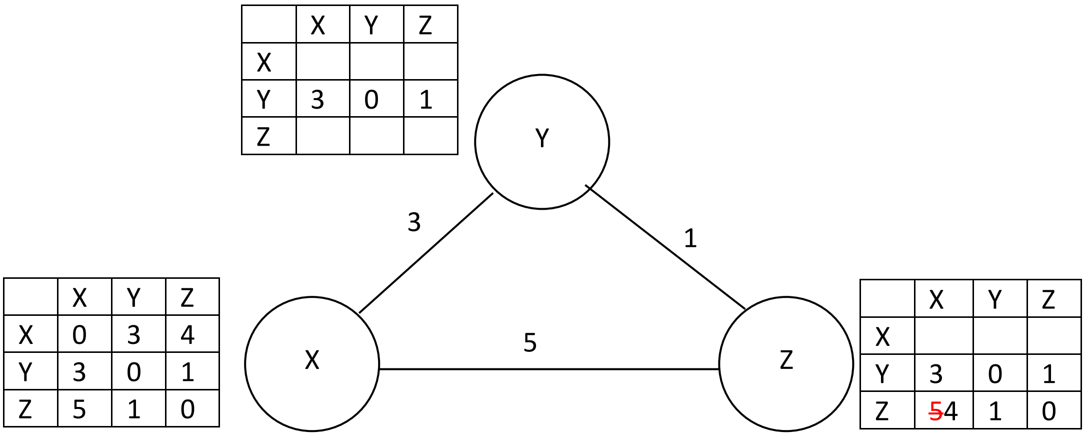
	X	Y	Z
X	0	3	4
Y	3	0	1
Z	5	1	0

	X	Y	Z
X			
Y			
Z	5	1	0

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

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

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

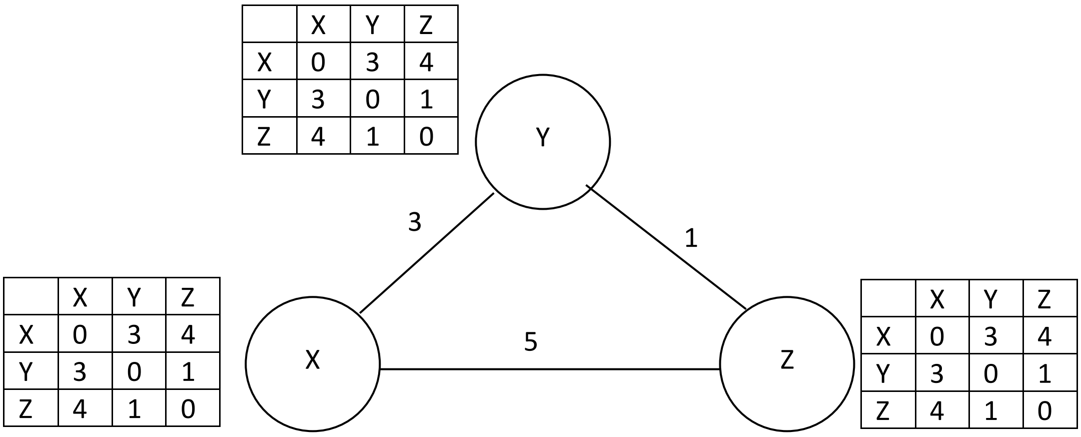


	X	Y	Z
X			
Y	3	0	1
Z			

	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 –



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

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

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

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.

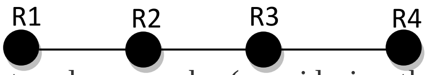
Further, both can be compared on some more points as follows:

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.

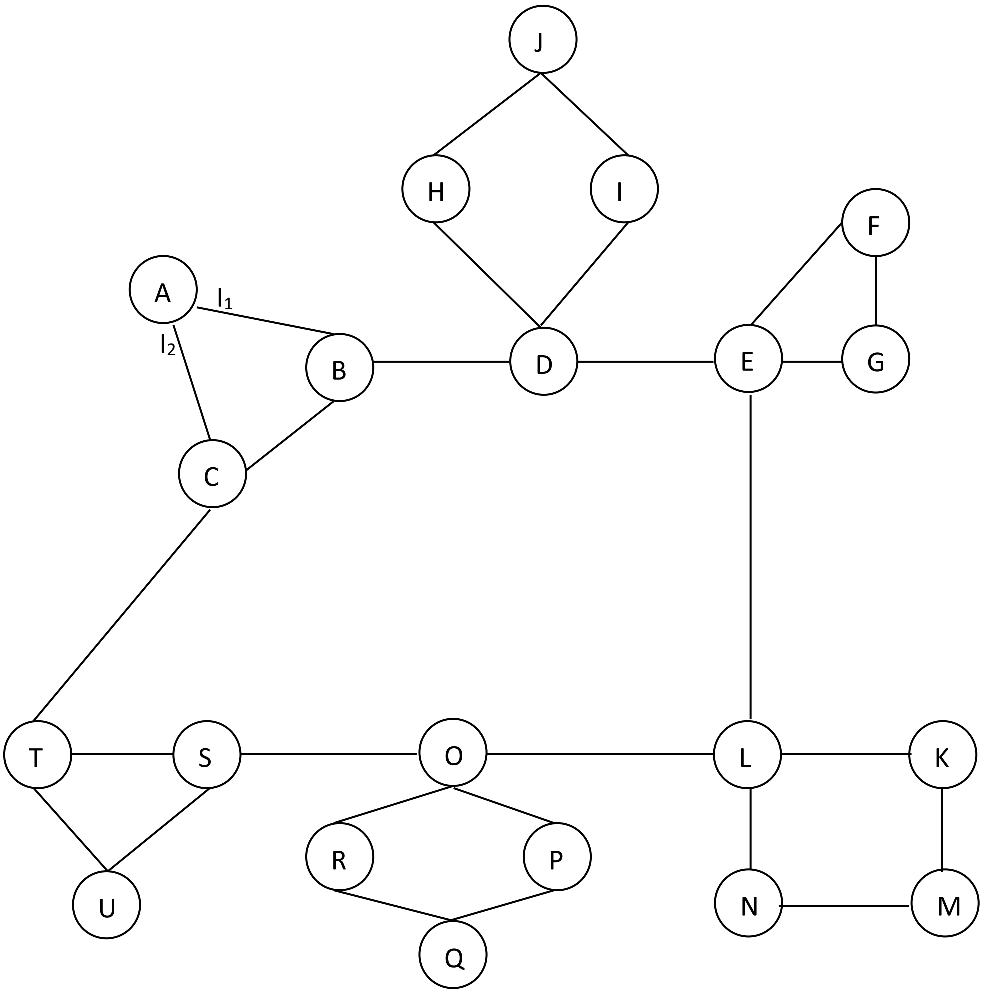
2.6 HIERARCHICAL ROUTING

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) .



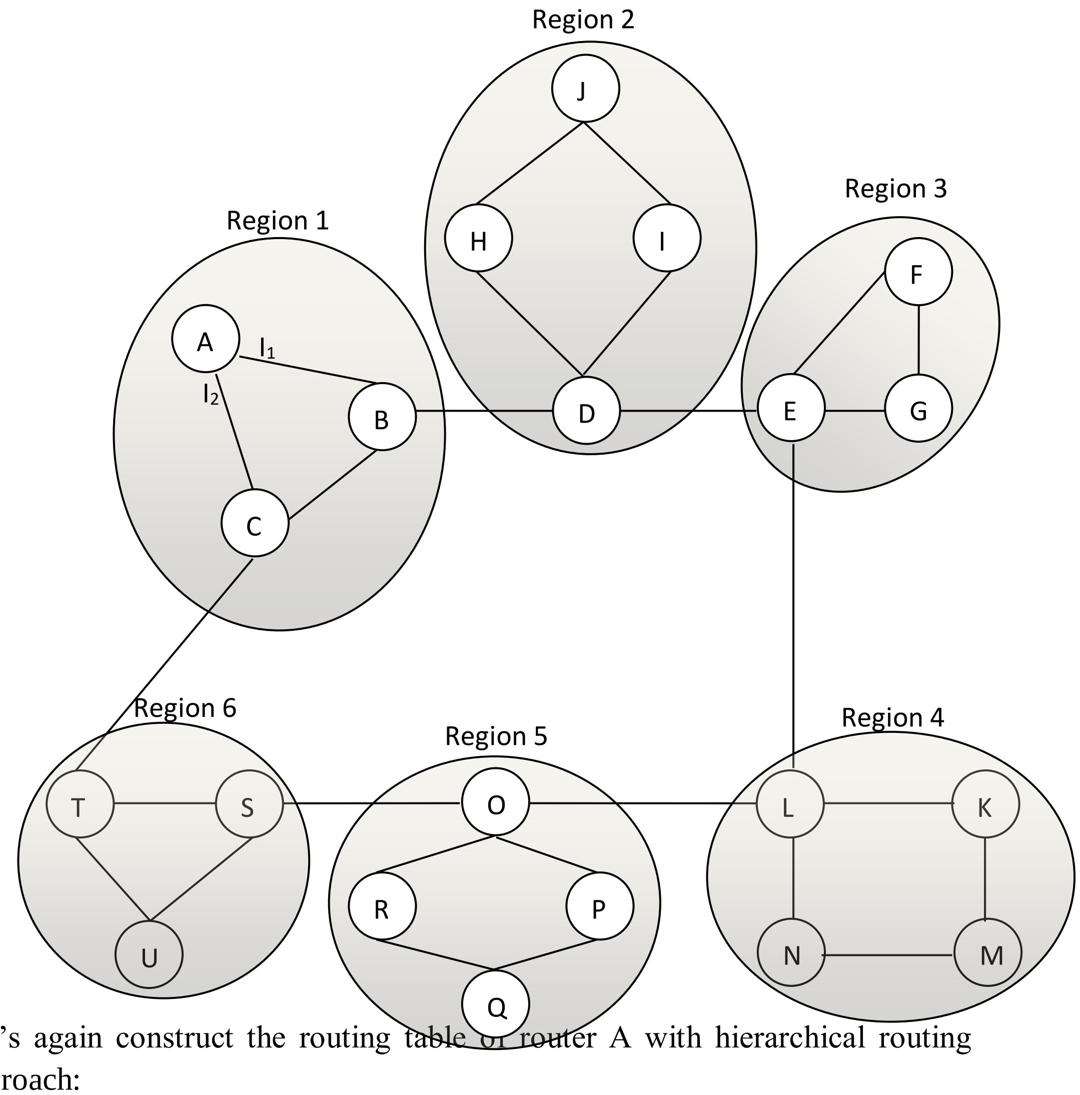
Destination	Exit Interface	Cost
A	--	0
B	I_1	1
C	I_2	1
D	I_1	2
E	I_1	3
F	I_1	4
G	I_1	4
H	I_1	3
I	I_1	3
J	I_1	4
K	I_1	5
L	I_1	4
M	I_1	6

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 4th version 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:

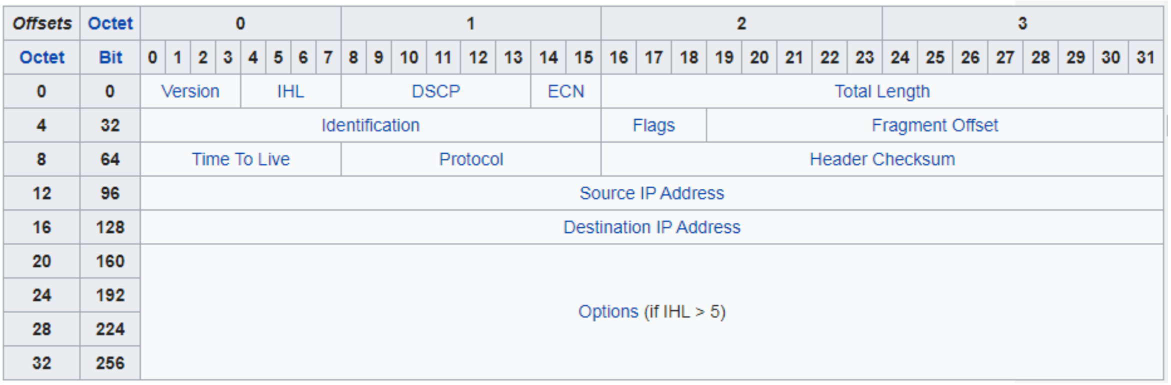
	1st Byte	2nd Byte	3rd Byte	4th Byte	Description
Class A	0 to 127 (in decimal) 00000000 to 01111111 (in binary)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	Open (can take any value between 0 to 255)	The 1st MSB bit of the 1stoctet is always set to 0 (zero) (as 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 $(2^{13} – 1) \times 8 = 65,528$ 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

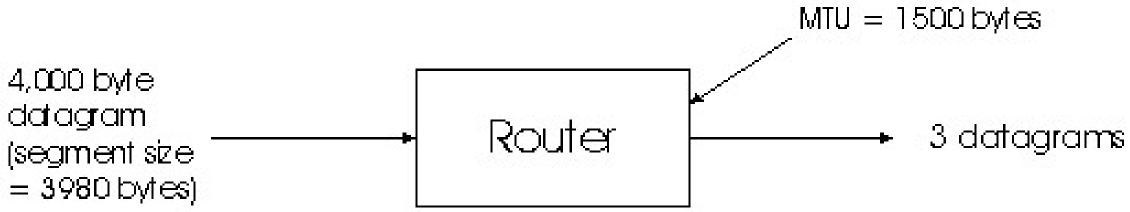
In general this field is not used while forming IP packets.

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 1stfragment:

- 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)

2ndfragment

- 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)

3rdfragment

- 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.

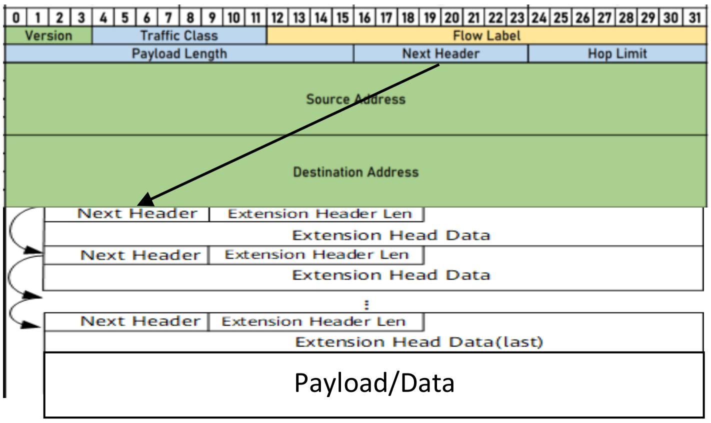
Address Representation

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. <u>Open Shortest Path First (OSPF)</u> :

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.

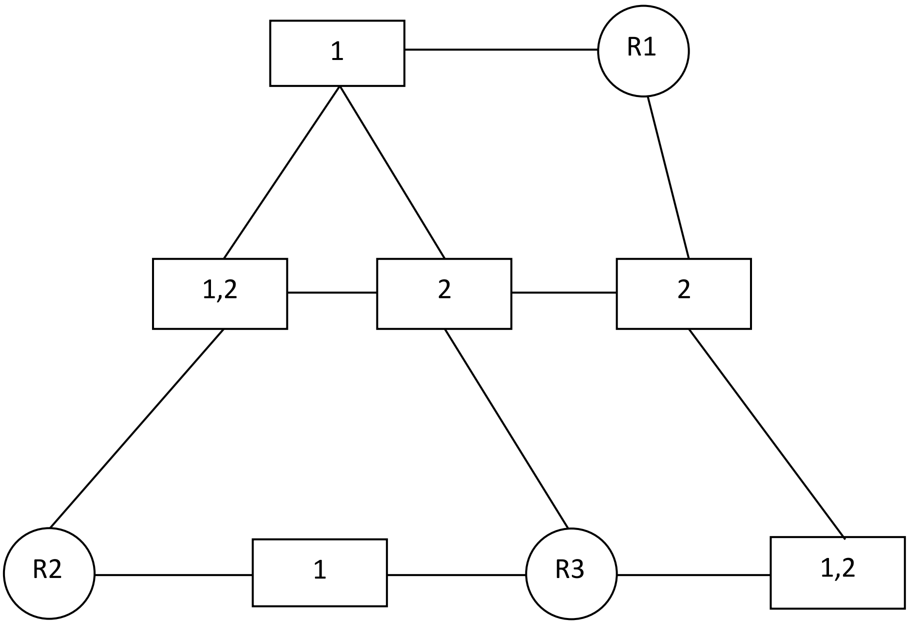
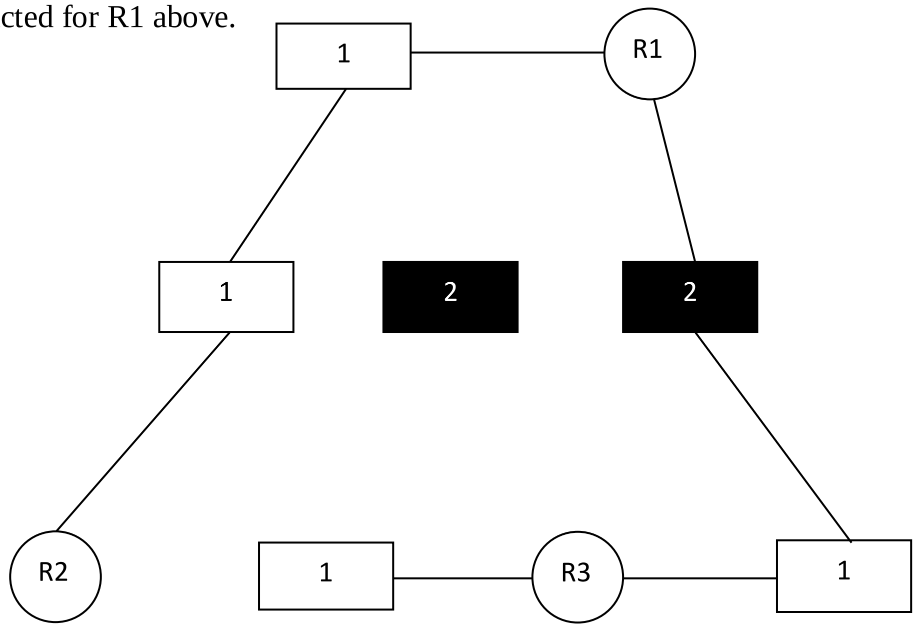
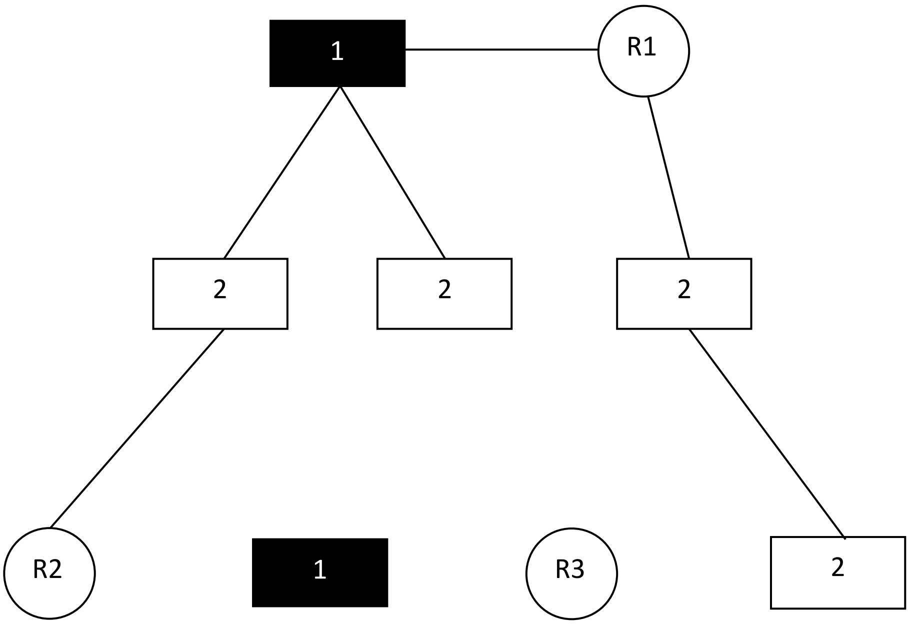


Figure below shows the pruned spanning tree for group 1 for the spanning tree constructed for R1 above.



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.