small amount of buffer space to temporarily hold packets. If the outgoing line is busy, the packet stay in queue until the line becomes available. Packet switching handles burstly traffic well.

- Packet switching method uses two routing approaches :
  1. **Datagram** and 2. **Virtual circuit.**

1) Datagram Packet Switching

- In **datagram** each packet is routed independently through the network. Header is attached to each packet. It provides all of the information required to route the packet to its destination. While routing the packet, the destination address in the header are examined to determine the next hop in the path to the destination. If the required line is busy then the packet is placed in the queue until the line becomes free. Packet share the transmission line with other packets. Then it deliver to the destination. Datagram approach is also called **connectionless**.
- Disadvantage of datagram approach is a lot of overhead because of independent routing. Another disadvantage is that packet may not arrive in the order at destination in which they were sent.
- Since each packet is routed independently, packets from the same source to the same destination may traverse through different paths. This is shown in Fig. 4.3.1.

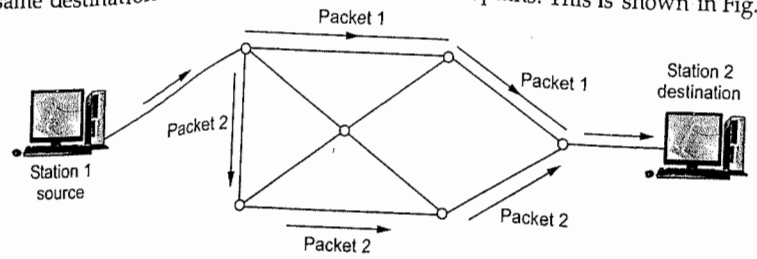


Fig. 4.3.1 Connectionless packet switching

- The packets at station 2 or destination may arrive out of order, and resequencing may be required at the destination. At each node a routing table is maintained which specifies the next hops that is to be taken by packets for the given destination.

2) Virtual Circuit Packet Switching

- In **virtual circuit packet switching** a fixed path between a source and a destination is established prior to transfer of packets.
- Connection-oriented network is also known as **virtual circuit**. Virtual circuit is similar to telephone system. A route, which consists of a logical connection is first established between two users. The connection that is established is not a dedicated path between stations. The path is generally shared by many other virtual connections.

- The process is completed in three main phases -
  i) Establishment phase.
  ii) Data transfer phase.
  iii) Connection release phase.

i) Establishment phase :

- During setting up of logical connection, the two users not only agree to setup a connection between them but also decide upon the quality of service associated with the connection. After this the sequences of packetized information are transmitted bidirectionally between the nodes. The information is delivered to the receiver in the same order as transmitted by sender.

ii) Data transfer phase :

- During this phase it performs flow control and error control services.
- The error control service ensures correct sequencing of packets and correct arrival of packets.
- Flow control service ensures a slow receiver from being overwhelming with data from a faster transmitter.

iii) Connection release :

- When the station wish to close down the virtual circuit, one station can terminate the connection with a clear request packet. Fig. 4.3.2 shows the virtual circuit packet switching.

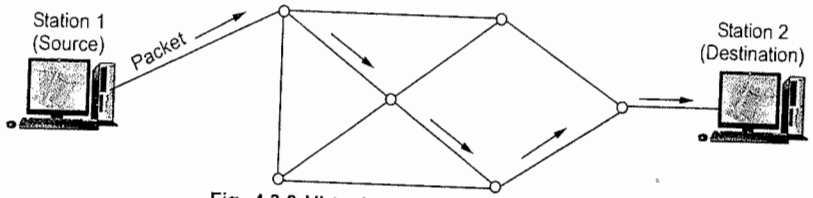


Fig. 4.3.2 Virtual circuit packet switching

4.3.1 Switching Fabric

- Component of packet switching are as follows :
  1. Input ports
  2. Number of output ports
  3. Switching fabric
  4. Routing processor
- Capacity of switch is the maximum rate at which it can move information, assuming all data paths are simultaneously active. Circuit switch must reject call if cannot find a path for samples from input to output. Packet switch must reject a packet if it can find a buffer to store it awaiting access to output trunk.

- Fig. 4.3.3 shows packet switch components.
- In packet switch, physical and data link function are performed by input ports. It constructs the packets from bits and perform error correction and detection.

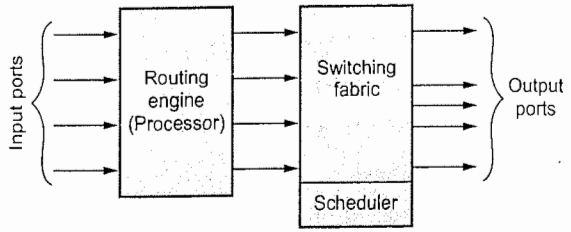


Fig. 4.3.3 Packet switch components

- It transfer data from input to output. It usually consists of links and switching elements.
- The routing engine looks-up the packet address in routing table and determines which output port to send the packet. It performs functions of network layer. Each packet is tagged with port number. The switch uses the tag to send the packet to the proper output port.
- Simplest switch fabric is a shared bus. Switch fabrics are created from certain building blocks of smaller switches arranged in stages.
- The simplest switch is a 2×2 switch, which can be either in the through or crossed position.

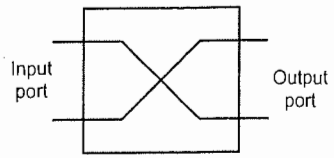


Fig. 4.3.4 Crossed position

4.3.2 Advantages of Packet Switching

1. Uses resources more efficiently.
2. Very little setup or tear down time.
3. It is more flexible. i.e. packets can be routed through any switching node.
4. Improved bandwidth.
5. Small sized packet reduces transmission delay.

4.3.3 Disadvantages of Packet Switching

1. Complex protocol for packet switching.
2. Algorithms are more complicated.
3. Difficults to bill customers.
4. Switching processor must be powerful.
5. Packets may lost during switching.

4.3.4 Circuit Switching

- The telephone system as it historically developed was designed for voice and analog signals. Sending data requires bandwidth. The amount of bandwidth needed is directly related to the data rate that is desired. An analog voice signal contains its data in a relatively narrow bandwidth, in proportion to the amount of data it carries.
- For voice signals, a relatively large amount of distortion is acceptable, since the human ear can understand voice even with distortion that looks severe to the eye. For digital signals, these distortions may cause the receiver to misinterpret the signal that is sent and so produce an error. The regular telephone loop from the local office to the phone is guaranteed by the phone company to have some specific characteristics. This type of line is the lowest performance line, called voice grade conditioning.
- Similar line characteristics are offered by telephone companies on the lines that go between phone company offices. These interoffice lines are called trunks. Any phone line can connect one user to another user through the phone system, the user has a line assigned randomly, through the phone offices. This is called the dial-up or switched network.
- Telephone networks are connection oriented because they require the setting up of connection before the actual transfer of information can take place.
- An end-to-end path setup beginning of a session, dedicated to the application, and then released at the end of session. This is called **circuit switching**. Circuit switching is effective for application which make comparatively steady use of channel. Fig. 4.3.5 shows the circuit switching.

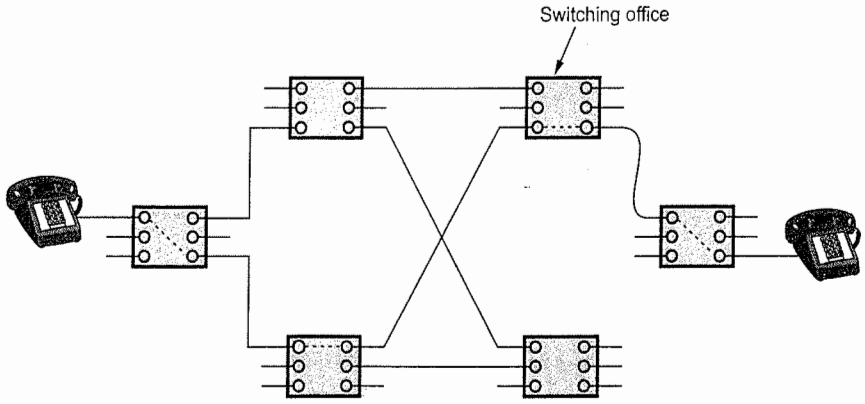


Fig. 4.3.5 Circuit switching

- For application which need greater performance than these dial up lines can offer, telephone companies offer specially conditional lines. These lines both from the phone to the office and between phone offices, provide better frequency response and time delay characteristics. This kind of conditioned line is leased by the user. The term dedicated and leased are used when the phone company has set a side a conditional line for a communications link.

Advantages of Circuit Switching

1. Fixed bandwidth, guaranteed capacity.
2. Low variance end to end delay.

Disadvantages

1. Connection setup and tear down introduces extra overhead.
2. User pay for circuit, even when not sending data.
3. Other user cannot use circuit even if it is free of traffic.

4.3.5 Message Switching

- Message switching is used to describe the telegraph network. When this form of switching is used, no physical copper path is established in advance between sender and receiver. When the sender has a block of data to be sent, it is stored in the first switching office i.e. router and then forwarded later, one hope at a time. Each block is received in its entirely, inspected for errors, and then transmitted. A network using this technique is called a **store and forward network**.
- The message was punched on paper tape off line at the sending office and then read in and transmitted over a communication line to the next office along the way, where it was punched out on paper tape. An operator tore the tape off and read it in on one of the many tape readers, one per outgoing trunk. Such a switching office was called a torn tape office.
- With message switching, there is no limit on block size, which means that routers must have disks to buffer long blocks. It also means that a single block may tie up a router, router line for minutes, rendering message switching uses for interactive traffic.
- Message switching does not involve a call setup. It can achieve a high utilization of the transmission line. Message switching is not suitable for interactive applications. Fig. 4.3.6 shows the message switching.

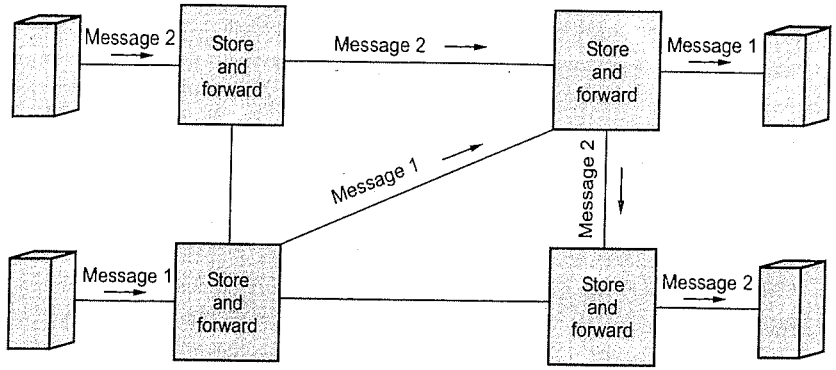


Fig. 4.3.6 Message switching

Advantages of Message Switching

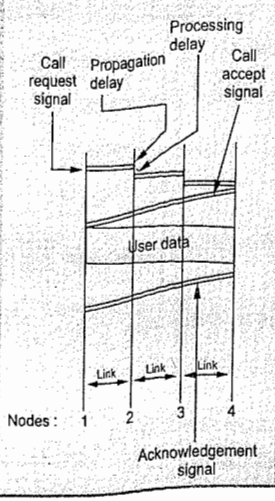
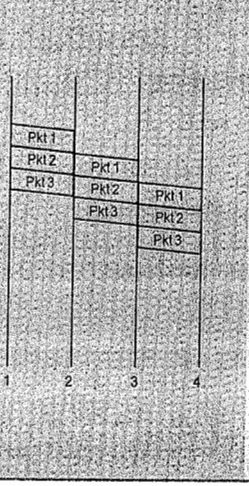
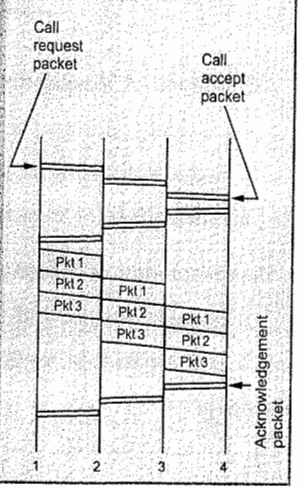
1. Efficient traffic management.
2. Reduces network traffic congenstion.
3. Efficient use of transmission channel.

Disadvantages of Message Switching

1. Because of store and forward, transmission delay is in deduced.
2. Each node requires large capacity for storing.

4.3.6 Comparison of Circuit, Packet and Message Switching

Sr. No.	Circuit switching	Packet switching	Message switching
1.	There is physical connection between transmitter and receiver.	No physical path is established between transmitter and receiver.	No physical path is set in advance between transmitter and receiver.
2.	All the packet uses same path.	Packet travels independently.	Packets are stored and forward.
3.	Needs an end to end path before the data transmission.	No needs of end to end path before data transmission.	Same as packet switching.
4.	Reverses the entire bandwidth in advance.	Does not reserve the bandwidth in advance.	Same as packet switching.
5.	Charge is based on distance and time, but not on traffic.	Charge is based on both number of bytes and connect time.	Charge is based on number of bytes and distance.
6.	Waste of bandwidth is possible.	No waste of bandwidth.	No waste of bandwidth.

7.	Congestion occur for per minute.	Congestion occurs for per packet.	No congestion or very less congestion.
8.	It cannot support store and forward transmission.	It support store and forward transmission.	It also support store and forward transmission.
9.	Not suitable for handling interactive traffic.	Suitable for handling interactive traffic.	Same as circuit switching.
10.	Recording of packet can never happen with circuit switching.	Recording of packet is possible.	Same as packet switching.
11.	Timing diagram 	Timing diagram 	Timing diagram 

University Questions

1. *Compare between circuit switching and packet switching.* **SPPU : May-12, Marks 6**
2. *Compare : i) Circuit and packet switching ii) Datagram and virtual circuits.* **SPPU : Dec.-12, Marks 8**
3. *Compare circuit switching and packet switching.* **SPPU : May-13, Marks 9**
4. *Compare circuit switched networks and datagram networks.* **SPPU : May-14, Marks 4**
5. *Explain virtual circuit networks in detail.* **SPPU : May-14, Marks 4**

4.4 IPv4

SPPU : May-12, 13, Dec.-12, 13, 14

- IP corresponds to the network layer in the OSI reference model and provides a connectionless best effort delivery service to the transport layer. An Internet Protocol (IP) address has a fixed length of 32 bits.

- IPv4 addresses are unique. Two devices on the internet can never have the same address at the same time.
- The address structure was originally defined to have a two level hierarchy : Network ID and host ID. The **network ID** identifies the network the host is connected to. The **host ID** identifies the network connection to the host rather than the actual host.

Address space

- An address space is the total number of addresses used by the protocol. If a protocol uses N bits to define an address, the address space is 2^N because each bit can have two different values and N bits can have 2^N values.
- IPv4 uses 32-bit addresses, which means that the address space is 2^{32} or 4, 294, 967, 296.
- IP addresses are usually written in **dotted decimal** notation so that they can be communicated conveniently by people. The address is broken into four bytes with each byte being represented by a decimal number and separated by a dot.
- For example, an IP address of
  10000000 10000111 01000100 00000100 is written as 128.135.68.4 in dotted-decimal notation.

  The address 193.32.216.9 in binary notation is

  11000001 00100000 11011000 00001001
- An IP address is a numeric identifier assigned to each machine on an IP network. IP address is a software address, not a hardware address, which is hard-coded in the machine or NIC. An IP address is made up of 32 bits of information. These bits are divided into four parts containing 8 bit each.
- There are three method for depicting an IP address.
  1. Dotted-decimal as in 131.57.30.57
  2. Binary, as, 10000010.00111001.00011110.00111000
  3. Hexadecimal, as in 8B.39.C2.43
- The 32-bit IP address is a structured or hierachical address. The network address uniquely identifies each network. Every machine on the same network shares that network address as part of its IP address. The IP address 131.57.30.57, the 131.57 is the network address and 30.57 is the node address. The node address is assigned to and uniquely identifies, each machine on a network.
- The router might able to speed a packet on its way after reading only the first bits of address. The format used for IP address are shown in Fig. 4.4.1 (b).

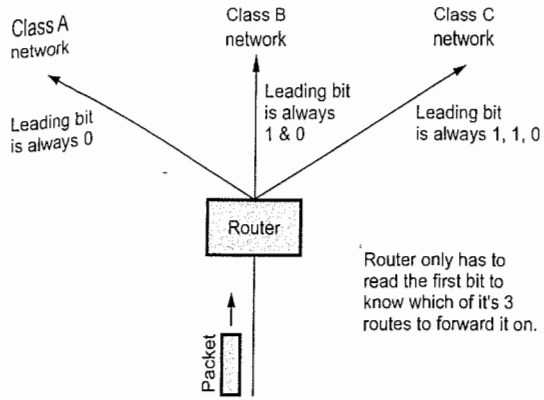


Fig. 4.4.1 (a) Leading bits of a network address

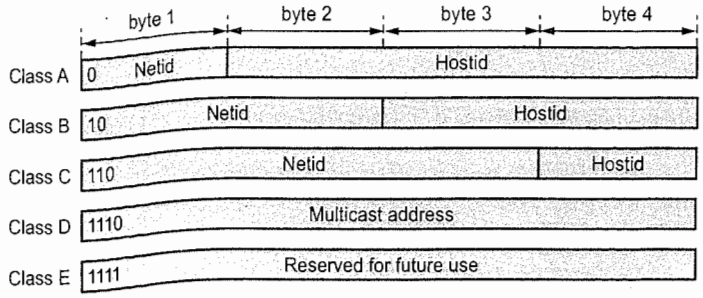


Fig. 4.4.1 (b) Internet classes (IP addresses)

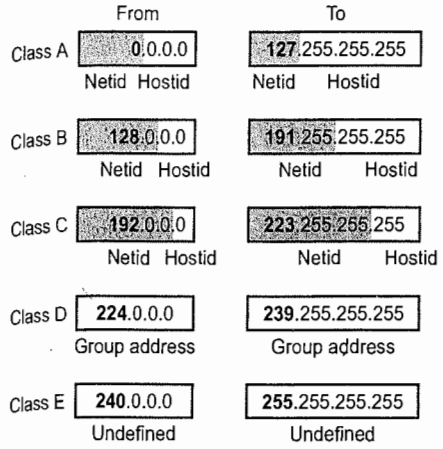


Fig. 4.4.1 (c) Classes range of IP

4.4.1 Classful Addressing

- The IP address structure is divided into five address classes : Class A, Class B, Class C, Class D and Class E, identified by the most significant bits of the addresses.
- Fig. 4.4.2 shows the five classes of IP addresses.

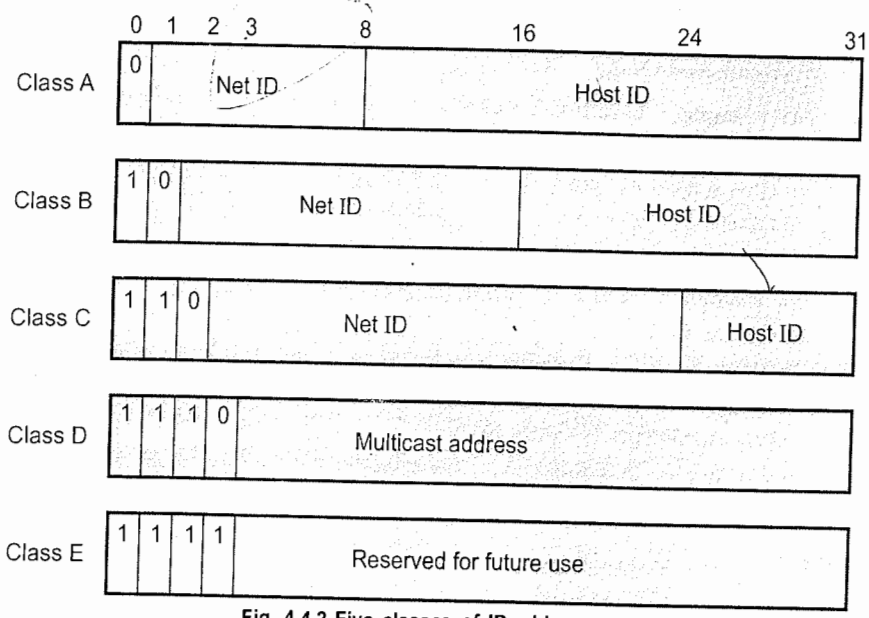


Fig. 4.4.2 Five classes of IP addresses

- Class D addresses are used for multicast services that allow a host to send information to a group of hosts simultaneously. Class E addresses are reserved for future use.
- Class A addresses were designed for large organizations with a large number of attached hosts or routers.
- Class B addresses were designed for midsize organizations with tens of thousands of attached hosts or routers.
- One problem with classful addressing is that each class is divided into a fixed number of blocks with each block having a fixed size.

Class	Number of blocks	Block size
A	128	16777216
B	16384	65536

C	2097152	256
D	1	268435456
E	1	268435456

- In a class A network, the first byte is assigned to the network address and the remaining three bytes used for the node addresses. The class A format is **Network.Node.Node.Node**

  For example : 14.28.101.120 in this IP address 14 is the network address and 28.101.120 is the node address.
- In class B network, the first two bytes are assigned to the network address and the remaining two bytes are used for node addresses. The format is **Network.Network.Node.Node**

  For example : 150.51.30.40 in this IP address network address is 150.51 and node address is 30.40.
- In class C network, the first three bytes are assigned to network address and only one byte is used for node address. The format is **Network.Network.Network.Node**
- **For example :** 200.20.42.120 in this example 200.20.42 is the network address and 120 is the node address.

4.4.2 Special IP Addresses

Some IP addresses are reserved for special purposes.

Sr. No.	Special address	Net ID	Host ID
1.	Network address	Specific	All 0
2.	Direct broadcast address	Specific	All 1
3.	Limited broadcast address	All 1s	All 1
4.	This host on this network	All 0s	All 0
5.	Loopback address	127	Any
6.	Specific host on this network	All 0s	Specific

4.4.3 Classless Addressing

- In classless addressing variable length blocks are assigned that belong to no class. In this, the entire address space is divided into blocks of different sizes. An organization is granted a block suitable for its purposes.
- Fig. 4.4.3 shows the architecture of classless addressing.

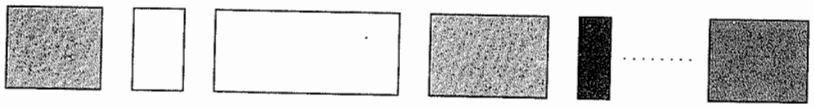


Fig. 4.4.3 Architecture of classless addressing

In classless addressing, when an entity, small or large, needs to be connected to the internet it is granted a block of addresses. The size of the block varies based on the nature and size of the entity.

Restriction

- To simplify the handling of addresses, the internet authorities impose three restrictions on classless address blocks.
  1. The addresses in a block must be contiguous, one after another.
  2. The number of addresses in a block must be a power of 2.
  3. The first address must be evenly divisible by the number of addresses.

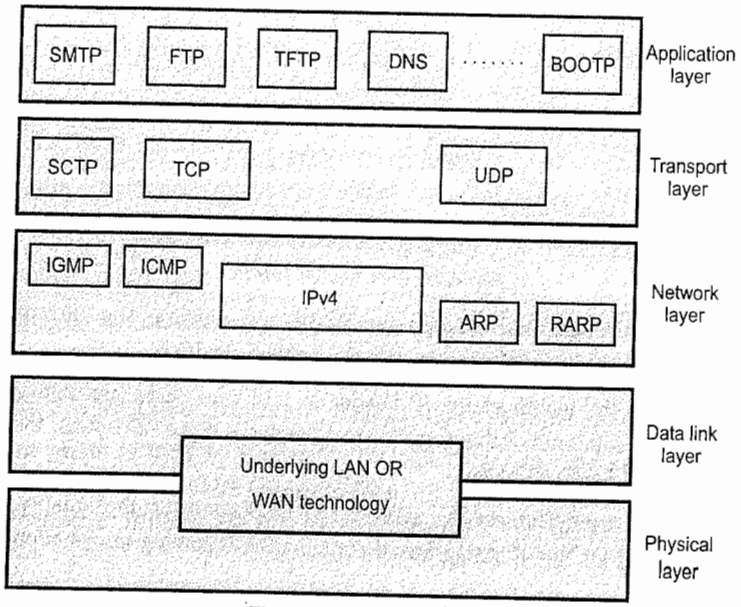


Fig. 4.4.4 IPv4 in TCP/IP

- In IPv4 addressing, a block of addresses can be defined as x.y.z.t/n in which x.y.z.t defines one of the addresses and the /n define the mask. The address and the /n notation completely define the whole block.
- IPv4 is the delivery mechanism used by the TCP/IP protocols. IPv4 is an unreliable and connectionless datagram protocol.
- IPv4 is also a connectionless protocol for a packet switching network that uses the datagram approach.
- Fig. 4.4.4 shows the positions of IPv4 in TCP/IP protocol suite.

4.4.4 Header Format

- Packets in the IPv4 layer are called datagrams. A datagram is a variable length packet consisting of two parts : Header and data.
- Fig. 4.4.5 shows IPv4 header format

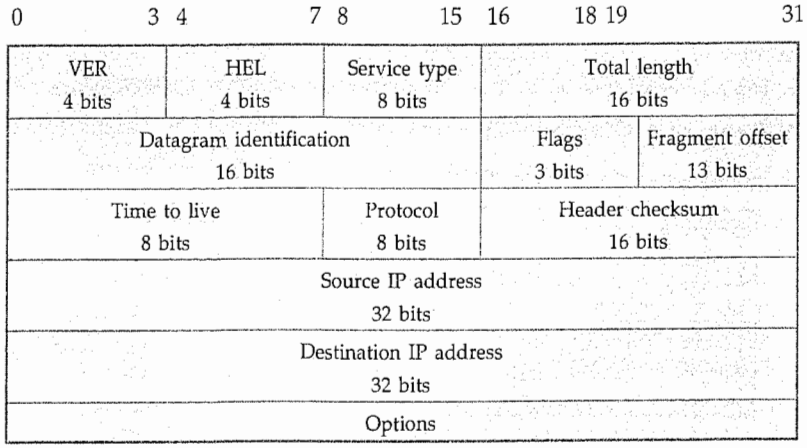


Fig. 4.4.5 IPv4 header format

1. **VER** is the field that contains the IP protocol version. The current version is 4.5 is an experimental version. 6 is the version for IPv6.
2. **HLEN** is the length of the IP header in multiples of 32 bits without the data field.The minimum value for a correct header is 5 (i.e. 20 bytes), the maximum value is 15 (i.e., 60 bytes).
3. **Service type** The service type is an indication of the quality of service requested for this IP datagram. It contains the following information.

Precedence	Types of service	R

Precedence specifies the nature / priority :

000	Routine
001	Priority
010	Immediate
011	Flash
100	Flash override
101	Critical
110	Internetwork control
111	Internetwork control

TOS specifies the type of service value :

TOS bits	Description
1000	Minimize delay
0100	Maximum throughout
0010	Maximize reliability
0001	Minimize monetary cose
0000	Normal service

The last bit is reserved for future use.

4. **Total length** specifies the total length of the datagram, header and data, in octets.
5. **Identification** is a unique number assigned by the sender used with fragmentation.
6. **Flags** contain control flags :
   a. The first bit is reserved and must be zero;
   b. The 2nd bit is DF (Do not Fragment), 0 means allow fragmentation;
   c. The third is MF (More Fragments), 0 means that this is the last fragment.
7. **Fragment offset** is used to reassemble the full datagram. The value in this field contains the number of 64-bit segments (header bytes are not counted) contained in earlier fragments. If this is the first (or only) fragment, this field contains a value of zero.
8. **TTL** (Time To Live) specifies the time (in seconds) the datagram is allowed to travel. In practice, this is used as a hop counter to detect routing loops.

9. **Protocol number** indicates the higher level protocol to which IP should deliver the data in this datagram. E.g., ICMP = 1; TCP = 6; UDP = 17.
10. **Header checksum** is a checksum for the information contained in the header. If the header checksum does not match the contents, the datagram is discarded.
11. **Source/Destination IP addresses** are the 32-bit source/destination IP addresses.
12. **IP options** is a variable-length field (there may be zero or more options) used for control or debugging and measurement. For instance :
    a. The **loose source routing** option provide a means for the source of an IP datagram to supply explicit routing information;
    b. The **timestamp** option tell the routers along the route to put timestamps in the option data.
13. **Padding** is used to ensure that the IP header ends on a 32 bit boundary. The padding is zero.

4.4.5 IP Fragmentation

- IP provides fragmentation/reassembly of datagrams. The maximum length of an IP datagram is 65,535 octets. When an IP datagram travels from one host to another, it may pass through different physical networks. Each physical network has a maximum frame size, called Maximum Transmission Unit (MTU), which limits the datagram length.
- A fragment is treated as a normal IP datagram while being transported to their destination. Thus, fragments of a datagram each have a header. If one of the fragments gets lost, the complete datagram is considered lost. It is possible that fragments of the same IP datagram reach the destination host via multiple routes. Finally, since they may pass through networks with a smaller MTU than the sender's one, they are subject to further fragmentation. Fig. 4.4.6 shows the MTU.

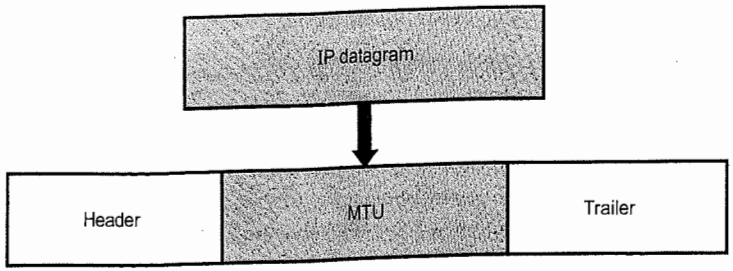


Fig. 4.4.6 MTU

Fragmentation process

- The DF flag bit is checked to see if fragmentation is allowed. If the bit is set, the datagram will be discarded and an ICMP error returned to the originator.
- Based on the MTU value, the data field is split into two or more parts. All newly created data portions must have a length that is a multiple of 8 octets, with the exception of the last data portion. Each data portion is placed in an IP datagram.
  Fig. 4.4.7 shows the examples of fragmentation.

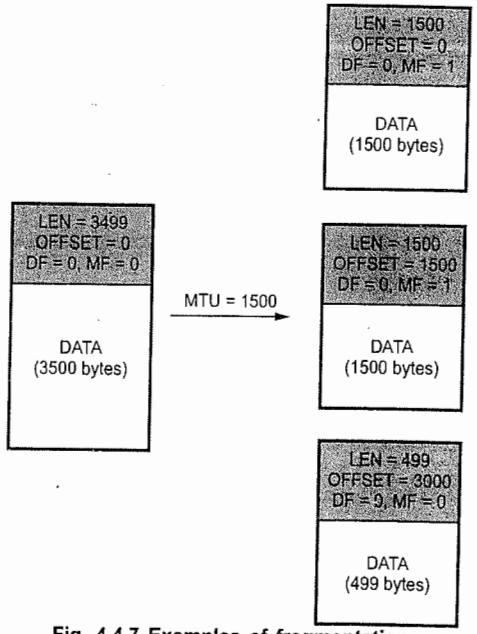


Fig. 4.4.7 Examples of fragmentation

Modification to the headers of fragments :

a. The MF flag is set in all fragments except the last;
b. The fragment offset field is updated;
c. If options were included in the original datagram, they may be copied to all fragment datagram's or only the first datagram (depends on the option);
d. The header length field is set;
e. The total length field is set;
f. The header checksum is re-calculated.

- At the destination host, data are reassembled into the original datagram. The identification field set by the sending host is used together with the source and destination IP addresses in the datagram. Fragmentation does not alter this field.
- In order to reassemble the fragments, the receiving host allocates a storage buffer when the first fragment arrives. The host also starts a timer. If the timer is exceeded and fragments remain outstanding, the datagram is discarded. When subsequent fragments of the datagram arrive, data are copied into the buffer storage at the location indicated by the fragment offset field. When all fragments have arrived, the original unfragmented datagram is restored and passed to upper layers, if needed.

Problem in fragmentation

1. The end node has no way of knowing how many fragments there be. The end node has to manage enough buffer space to handle reassembly process.
2. If any fragments lost, all datagram must be discarded.
3. End node starts a timer when received the first fragment, if any fragments fails to arrive (usually 30 secs), all datagram's must be discarded.
4. Since the IP service is connectionless. No attempt is made by IP to recover these situations, through ICMP error message may be generated.

4.4.6 Options

The header of the IPv4 datagram is made of two parts : A fixed part and a variable part. Options used in IPv4 are as follows

1. **No operation** option is 1-byte option used as a filter between options.
2. **End of option** is a 1-byte option used for padding at the end of the option field.
3. Record route option is used to record the internet routers that handle the datagram. It can list upto nine router addresses. It can be used for debugging and management purposes.
4. Strict source route option is used by the source to predetermine a route for the datagram as it travels through the Internet.
5. Loose source route option is similar to the strict source route, but it is less rigid. Each router in the list must be visited, but the datagram can visit other routers as well.
6. Timestamp option is used to record the time of datagram processing by a router.

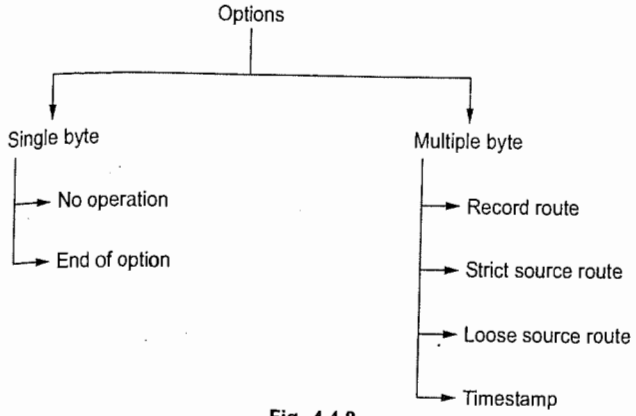


Fig. 4.4.8

4.4.7 Subnetting a Network

- If a organization is large or if its computers are geographically dispersed, it makes good sense to divide network into smaller ones, connected together by routers. The benefits for doing things this way include.
  1. Reduced network traffic
  2. Optimized network performance
  3. Simplified network management
  4. Facilities spanning large geographical distances.
- If Network Information Center (NIC) assign only one network address to an organization which having multiple network, that organization has a problem. A single network address can be used to refer to multiple physical networks. An organization can request individual network **address** for each one of its physical networks. If these were granted, there wouldn't be enough to go around for everyone.
- Another problem is, if each router on the internet needed to know about each existing physical network, routing tables would be impossibly huge. This is physical overhead on the router. To solve this type of problem, the subnet addressing method is used.

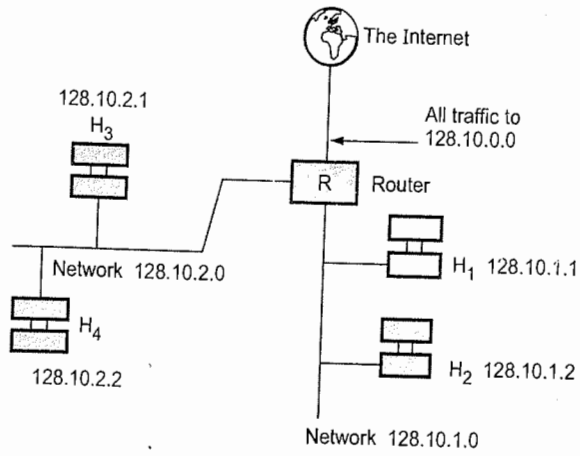


Fig. 4.4.9 Multiple network

- To allow a single network address to span multiple physical networks is called **subnet addressing** or **subnet routing** or **subnetting**. Subnetting is a required part of IP addressing.
- To understand subnet addressing, consider the next example. Consider the site has a single class B IP network address assigned to it, but the organization has two or more physical networks. Only local routers know that there are multiple physical networks and how to route traffic among them.

- In the example, the organization is using the single class B network address for two networks. For the subnet address scheme to work, every machine on the network must know which part of the host address will be used as the subnet address. This is accomplished by assigning each machine a subnet mask.
- The network administrator creates a 32-bit subnet mask comprised of ones and zeros. The ones in the subnet mask represent the positions that refer to the network or subnet addresses. The zeros represent the positions that refer to the host part of the address. Class B address format is Net.Net.Node.Node. The third byte, normally assigned as part of the host address is now used to represent the subnet address. Hence, these bit positions are represented with ones in the subnet mask. The fourth byte is the only part in example that represents the unique host address.

Subnet mask code

1 = Positions representing network or subnet addresses.

0 = Positions representing the host address.

Subnet mask format

1111 1111. 1111 1111 1111 1111. 0000 0000

Network address positions *Subnet positions* *Host positions*

The subnet mask can also be denoted using the decimal equivalents of the binary patterns. The default subnet masks for the different classes of networks are as below in Table 4.4.1

Class	Format	Defult subnet mask
A	Net.Node.Node.Node	255.0.0.0
B	Net.Net.Node.Node	255.255.0.0
C	Net.Net.Net.Node	255.255.255.0

Table 4.4.1 Default subnet mask of IP address

Masking

- A process that extracts the address of the physical network from an IP address is called Masking. If we done the subnetting, then masking extracts the subnetwork address from an IP address.
- To find the subnetwork address, two method are used. There are boundary level masking and non-boundary level masking, we take one by one.

- In boundary level masking, two masking numbers are consider (i.e. 0 or 255). In non-boundary level masking other value of masking is used Apart from 0 and 255.

A. Rules for boundary level masking

1. In this mask number is either 0 or 255.
2. If the mask number is 255 in the mask IP address, then the IP address is repeated in subnetwork address.
3. If the mask number is 0(zero) in the mask IP address, then the 0 is repeated in subnetwork address.

B. Rules for non-boundary level masking

1. In this mask numbers are not 0 or 255 mask number is greater than 0 or less than 255.
2. If the mask number is 255 in the mask IP address, then the original IP address (byte) is repeated in subnetwork address.
3. If the mask number is 0 in the mask IP address, then the 0 is repeated in the subnetwork address.
4. For any other mask numbers, bit-wise AND operator is used. Bit-wise ANDing is done in between mask number (byte) and IP address (byte).

- The first address in the block is used to identify the organization to rest of the Internet. This address is called the **network address.**

1. How many subnets ?

- Number of subnet is calculated as follows :

Number of subnet = 2^x

where x is the number of masked bits or the 1s (ones).

For example 11100000, the number of 1s gives us 2^3 subnets. In this example there are 8 subnets.

2. How many host per subnet ?

Number of host per subnet = $2^y - 2$

Where y is the number of unmasked bits or the 0s (zeros).

For example 11100000, the number of 0s gives us $2^5 - 2$ hosts. In this example there are 30 hosts per subnet. You need to subtract 2 for subnet address and the broadcast address.

3. What are the valid subnets ?

For valid subnet = 256 – Subnet mask = Block size. An example would be

256 – 224 = 32. The block size of a 224 mask is always 32.

Start counting at zero in block of 32 until you reach the subnet mask value and these are your subnets. 0, 32, 64, 96, 128, 160, 192, 224.

4. What is the broadcast address for each subnet ?

Our subnets are 0, 32, 64, 96, 128, 160, 192, 224, the broadcast address is always the number right before the next subnet. For example, the subnet 0 has a broadcast address of 31 because next subnet is 32. the subnet 32 has a broadcast address of 63 because next subnet is 64.

5. What are the valid hosts ?

Valid hosts are the numbers between the subnets, omitting the all 0s and all 1s. For example, if 32 is the subnet number and 63 is the broadcast address, then 32 to 63 is the valid host range. It is always between the subnet address and the broadcast address.

Example 4.4.1 *What is the sub-network address if the destination address is 200.45.34.56 and the subnet mask is 255.255.240.0 ?*

Solution : Using AND operation, we can find sub-network address.

1. Convert the given destination address into binary format :

   200.45.34.56 => 11001000 00101101 00100010 00111000

2. Convert the given subnet mask address into binary format :

   255.255.240.0 => 11111111 11111111 11110000 00000000

3. Do the AND operation using destination address and subnet mask address.

   200.45.34.56 => 11001000 00101101 00100010 00111000

   255.255.240.0=> 11111111 11111111 11110000 00000000

   11001000 00101101 00100000 00000000

Subnetwork address is 200.45.32.0

Example 4.4.2 *For a network address 192.168.10.0 and subnet mask 255.255.255.224 then calculate :*

i) Number of subnet and number of host

ii) Valid subnet

Solution : Given network address 192.168.10.0 is class C address. Subnet mask address is 255.255.255.224. Here three bits is browse for subnet.

i) Number of subnet and number of host :

255.255.255.224 convert into binary => 11111111 11111111 11111111 11100000

Number of subnet = 2^x

= 2^3

= 8

So there are 8 subnet.

Number of host per subnet = $2^y - 2$

= $2^5 - 2$

= 30

ii) Valid subnets :

For valid subnet = 256 – Subnet mask = Block size. An example would be 256 – 224 = 32. The block size of a 224 mask is always 32.

Start counting at zero in block of 32 until you reach the subnet mask value and these are your subnets. 0, 32, 64, 96, 128, 160, 192, 224.

Example 4.4.3 *Find the sub-network address for the following :*

Sr. No.	IP address	Mask
a)	140.11.36.22	255.255.255.0
b)	120.14.22.16	255.255.128.0

Solution :

a) IP address 140.11.36.22 Mask 255.255.255.0

The values of mask (i.e. 255.255.255.0) is boundary-level. So

IP address 140.11.36.22

Mask 255.255.255.0

140.11.36.0

b) IP address 120.14.22.16

Mask 255.255.128.0

The byte 1, byte 2 and byte 4 is boundary values and byte 3 is non-boundary value.

Example 4.4.4 *Find the sub-network address for the following.*

Sr. No.	IP address	Mask
a)	141.181.14.16	255.255.224.0
b)	200.34.22.156	255.255.255.240
c)	125.35.12.57	255.255.0.0

Solution :

a)

141.181.14.16	IP address
255.255.224.0	Mask
141.181.0.0	Sub-network address

b)

200.34.22.156	IP address
255.255.255.240	Mask
200.34.22.144	Sub-network address

c)

125.35.12.57	IP address
255.255.0.0	Mask
125.35.0.0	Sub-network address

(i.e. 128) So for byte-3 value use bit-wise AND operator. It is shown below.

120.14.22.16	IP address
255.255.128.0	Mask
120.14.0.0	Sub-network address

In the above example, the bit wise ANDing is done in between 22 and 128. it is as follows

22	Binary representation	0 0 0 1 0 1 1 0
128	Binary representation	1 0 0 0 0 0 0 0
0		0 0 0 0 0 0 0 0

Thus the sub-network address for this is 120.14.0.0.

Example 4.4.5 *Find the class of the following address.*

a) 1.22.200.10 b) 241.240.200.2 c) 227.3.6.8 d) 180.170.0.2

Solution :

a) 1.22.200.10 Class A IP address

b) 241.240.200.2 Class E IP address

c) 227.3.6.8 Class D IP address

d) 180.170.0.2 Class B IP address

Example 4.4.6 *Find the netid and Hostid for the following.*

a) 19.34.21.5 b) 190.13.70.10 c) 246.3.4.10 d) 201.2.4.2

Solution :

a) netid ⇒ 19 Hostid ⇒ 13.70.10

b) netid ⇒ 190.13 Hostid ⇒ 70.10

c) No netid and No Hostid because 246.3.4.10 is the class E address.

d) netid ⇒ 201.2.4 Hostid ⇒ 2

Example 4.4.7 *An organization is granted the block 211.17.180.0/24. The administrator wants to create 32 subnets.*

a) Find the subnet mask. b) Find the number of addresses in each subnets.

c) Find the first and last addresses in subnet 1.

d) Find the first and last addresses in subnet 32. **Dec.-10**

Solution : a) Find the subnet mask :

log 232 = 5 Extra 1s = **5** Possible subnets : **32** Mask : **/29** (24 + 5)

Subnet mask is 255.255.255.248

b) Find the number of addresses in each subnets :

2^{32-29} = **8** Addresses per subnet.

c) Find the first and last addresses in subnet 1 :

Subnet 1 : The first address is the begining address of the block or **211.17.180.0. First address in subnet 1 : 211. 17. 180 . 0**

Number of addresses : 0. 0. 0. 7

Last address in subnet 1 : 211. 17. 180. 7

d) Find the first and last addresses in subnet 32 :

Subnet 32 : To find the first address in subnet 32, we need to add 248 (31× 8) in base 256 (0.0.0.248) to the first address in subnet 1. So that 211.17.180.0 + 0.0.0.248.

OR

211.17.180.248. Now we can calculate the last address in subnet 32.

First address in subnet 32 : 211. 17. 180. 248.

Number of addresses : 0.0.0.7.

Last address in subnet 32 : 211.17.180.255.

4.4.8 Network Address Translation (NAT)

- Within the company, every machine has a unique address of the form 10 X.Y.Z. when a packet leaves the company premises, it passes through the NAT box that convert the internal IP source address 10.0.0.1. NAT box is often combined in a single device with a firewall. It is also possible to integrate the NAT box into the company router.
- Whenever an outgoing packet enters the NAT box, the 10.X.Y.Z. SA is replaced by the company true IP address. In, addition, TCP source port field is replaced by an index into the NAT box 65536 entry translation table. This table entry contains the original IP address and original source port. Finally both the IP and TCP header checksums are recomputed and inserted into the packet.
- Fig. 4.4.10 shows the placement of NAT box.
- When process want to establish a TCP connection with a remote process, it

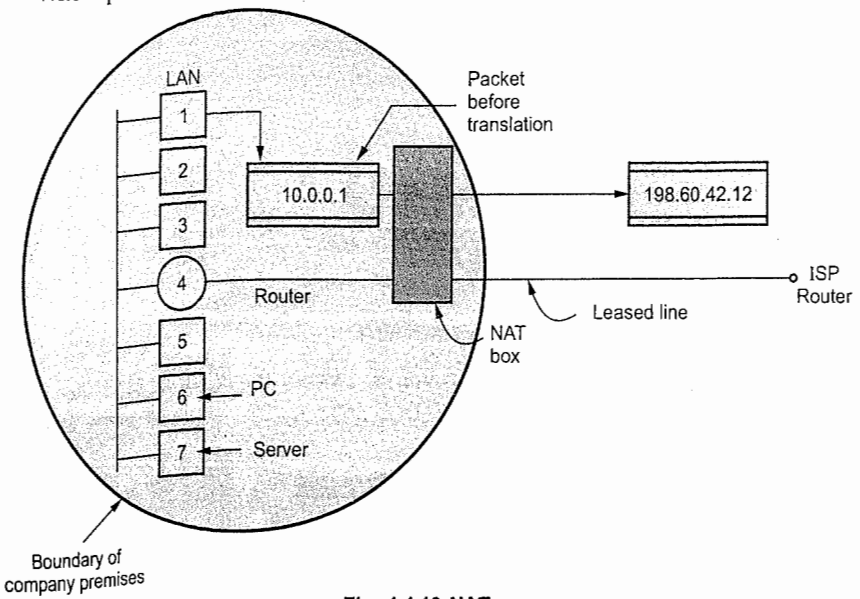


Fig. 4.4.10 NAT

attached itself to an unused TCP port on its own machine. This is called a source port and tells the TCP code where to send incoming packets belonging to this connection. The process also supplies a destination port to tell who to give the packet to on the remote side.

4.4.9 Classless InterDomain Routing (CIDR)

- Dividing the IP address space into A, B and C classes turned out to be inflexible. IP is rapidly becoming a victim of its own popularity, it is running out of addresses. In 1993 the classful address space restriction was lifted. An arbitrary prefix length to indicate the network number, known as Classless InterDomain Routing (CIDR), was adopted in place of the classful scheme.
- Using a CIDR notation, a prefix 205.100.0.0 of length 22 is written as 205.100.0.0/22. The corresponding prefix range runs from 205.100.0.0 through 205.100.3.0. The /22 notation indicates that the network mask is 22 bits, or 255.255.252.0. CIDR routes packets according to the higher order bits of the IP address.
- The entries in a CIDR routing table contain a 32-bit IP address and a 32-bit mask. CIDR uses a technique called supernetting so that a single routing entry covers a block of classful addresses.
- For example, address of class C i.e. 205.100.0.0, 205.100.2.0, 205.100.2.0 and 205.100.3.0, CIDR allows a single entry 205.100.16.0/22. The use of variable length prefixed requires that the routing tables be searched to find the longest prefix match. For example, a routing table may contain entries for the above supernet 205.100.0.0/22 as well as for 205.100.0.0/20. This situation may arise when a large number of destinations have been aggregated into the block 205.100.0.0/20, but packets destined to 205.100.16.0/22 are to be routed differently. A packet with destination address 205.100.1.1 will match both of these entries, so the algorithm must select the match with the longest prefix.

University Questions

1. *For a given classless IP address, how will you extract network address and host address ? Explain with suitable example.* **SPPU : May-12, Marks 8**
2. *Draw IPv4 headers and explain briefly.* **SPPU : Dec.-12, Marks 8**
3. *Explain various classes of IP-addressing.* **SPPU : May-13, Marks 9**
4. *How can we distinguish a multicast address in IPv4 ? How can we do so in IPv6 ?* **SPPU : Dec.-13, Marks 6**
5. *For a given classful IP address, how will you extract network address and host address ? Explain with suitable example ?* **SPPU : Dec.-14, Marks 8**

4.5 IPv6

SPPU : May-12, 13, Dec.-12

- IPv4 provides the host to host communication between systems in the Internet. IPv4 has played a central role in the internetworking environment for many years. It has proved flexible enough to work on many different networking technologies.

- In the early 1990 the IETF began to work on the successor of IPv4 that would solve the address exhaustion problem and other scalability problems.

Advantages of IPv6

1. Larger address space
2. Better header format
3. Security capabilities
4. Support for resource allocation
5. New options
6. Allowance for extension

- IPv6 addresses are 128 bits in length. Addresses are assigned to individual interface on nodes, not to the node themselves. A single interface may have multiple unique unicast addresses. The first field of any IPv6 address is the variable length format prefix, which identifies various categories of addresses.

IPv6 addresses

- A new notation has been devised for writing 16-byte addresses. They are written as eight groups of four hexadecimal digits with colons beween the groups, like this
  8000 : 0000 : 0000 : 0000 : 0123 : 4567 : 89AB : CDEF

Optimization

1. Leading zeros within a group can be omitted so 0123 can be written as 123.
2. One or more groups of 16 zero bits can be repalced by a pair of colons. The address new becomes
   8000 : : 123 : 4567 : 89AB : CDEF

4.5.1 Address Types

- IPv6 allows three types of addresses.
  1. Unicast 2. Anycast 3. Multicast

1. Unicast : An identifier for a single interface. A packet sent to a unicast address is delivered to the interface identified by that address.

2. Anycast : An identifier for a set of interfaces. A packet sent to an anycast address is delivered to one of the interfaces identified by the address.

3. Multicast : An identifier for a set of interfaces. A packet sent to a multicast address is delivered to all interfaces identified by that address.

- Following Table 4.5.1 shows the current allocation of addresses based on the format prefix.
- The first field of any IPv6 address is the variable-length fromat prefix, which identifies various categories of addresses.

Allocation space	Prefix (binary)	Fraction of address space
Reserved	0000 0000	1/256
Unassigned	0000 0001	1/256
Reversed for NSAP allocation	0000 001	1/128
Reversed for IPX allocation	0000 010	1/128
Unassigned	0000 011	1/128
Unassigned	0000 1	1/32
Unassigned	0001	1/16
Unassigned	001	1/8
Provider-based unicast address	010	1/8
Unassigned	011	1/8
Reserved for geographic-based unicast addresses	100	1/8
Unassigned	101	1/8
Unassigned	110	1/8
Unassigned	1110	1/16
Unassigned	1111 0	1/32
Unassigned	1111 10	1/64
Unassigned	1111 110	1/128
Unassigned	1111 1110 0	1/512
Link local use addresses	1111 1110 10	1/1024
Site local use addresses	1111 1110 11	1/1024
Multicast addresses	1111 1111	1/256

Table 4.5.1 Address allocation

4.5.2 Packet Format

- The IPv6 packet is shown in Fig. 4.5.1. Each packet is composed of a mandatory base header followed by the payload. The payload consists of two parts : Optional and data.

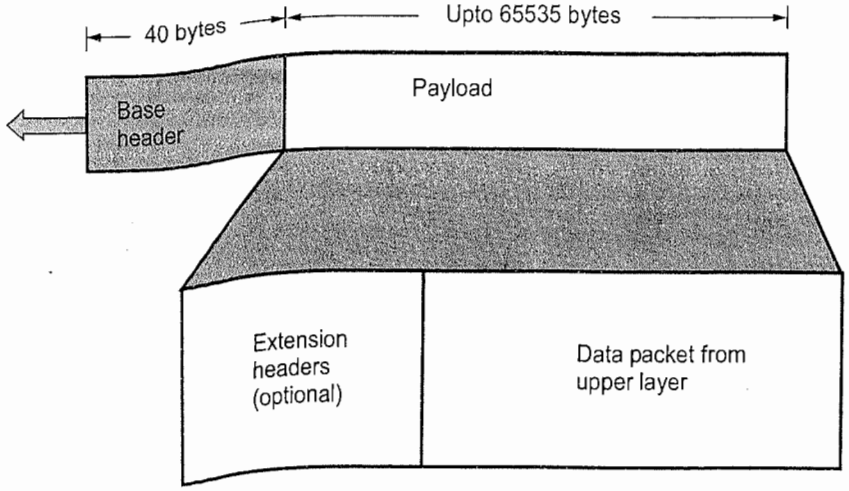


Fig. 4.5.1 IPv6 datagram header of payload

- Fig. 4.5.2 shows the IPv6 datagram header format.

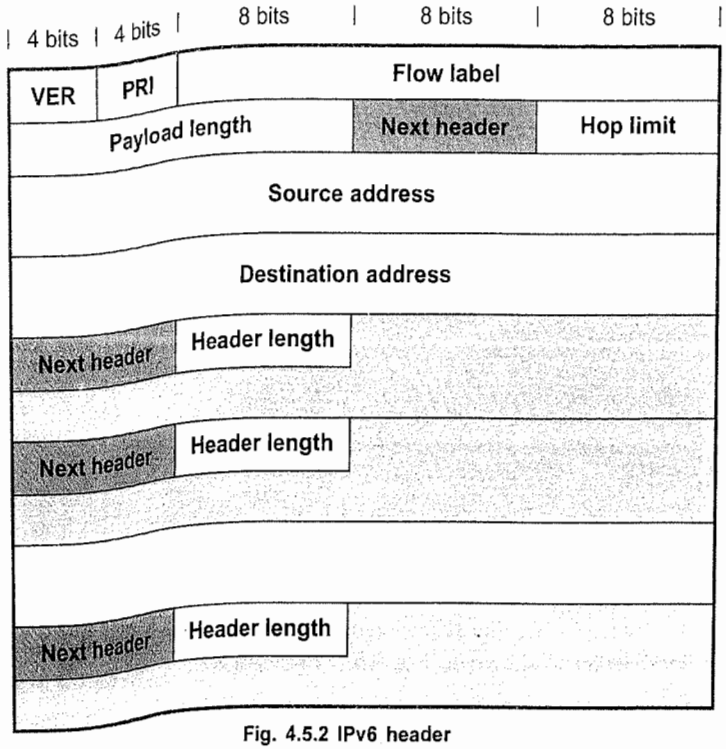


Fig. 4.5.2 IPv6 header

1. **Versions :** This 4 bits field defines the version number of the IP. The value is 6 for IPv6.
2. **Priority :** The 4 bits priority field defines the priority of the packet with respect to traffic congestion.
3. **Flow label :** It is 24 bits field that is designed to provide special handling for a particular flow of data.
4. **Payload length :** The 16 bits payload length field defines the length of the IP datagram excluding the base header.
5. **Next header :** It is an 8 bits field defining the header that follows the base header in the datagram.
6. **Hop limit :** This 8 bits hop limit field serves the same purpose as the TTL field in IPv4.
7. **Source address :** The source address field is a 128 bits internet address that identifies the original.
8. **Destination address :** It is 128 bits Internet address that usually identifies the final destination of the datagram.

- Next header codes for IPv6

Code	Next header
0	Hop by hop option
2	ICMP
6	TCP
17	UDP
43	Source routing
44	Fragmentation
50	Encrypted security payload
51	Authentication
59	Null
60	Destination option

Priority

- The priority field defines the priority of each packet with respect to other packets from the same source. IPv6 divides traffic into two broad categories
  1. Congestion controlled 2. Noncongestion controlled

- If a source adapts itself to traffic slowdown when there is congestion, the traffic is referred to as congestion controlled traffic. congestion controlled data are assigned priorities from 0 to 7.

Priority	Meaning
0	No specific traffic
1	Background data
2	Unattended data traffic
3	Reserved
4	Attended bulk data traffic
5	Reserved
6	Interactive traffic
7	Control traffic

- A priority of 0 is the lowest; a priority of 7 is the highest.
- Noncongestion controlled traffic refers to a type of traffic that excepts minimum delay. Discarding of packets is not desirable. Retransmission in most cases is impossible. Real time audio and video are examples of this type of traffic.
- Priority numbers from 8 to 15 are assigned to noncongestion controlled traffic.

4.5.3 Extension Headers

- The length of the base header is fixed at 40 bytes. Types of extension headers are
  1. Hop by hop option
  2. Source routing
  3. Fragmentation
  4. Authentication
  5. Encrypted security payload
  6. Destination option
- **Hop by hop option** is used when the source needs to pass information to all routers visited by the datagram.
- **Source routing** extension header combines the concepts of the strict source route and the loose source route options of IPv4.
- The concept of **fragmentation** is the same as that in IPv4. In IPv6, only the original source can fragment.

- The **authentication** header has a dual purpose : It validates the message sender and ensures the integrity of data.
- The **encrypted security payload** is an extension that provides confidentiality and guards against eavesdropping.
- The **destination option** is used when the source needs to pass information to the destination only. Intermediate routers are not permitted access to this information.

4.5.4 Comparison between IPv4 and IPv6

Sr. No.	IPv4	IPv6
1.	Header size is 32 bits.	Header size is 128 bits.
2.	It cannot support autoconfiguration.	Supports autoconfiguration
3.	Cannot support real time application.	Supports real time application.
4.	No security at network layer.	Provides security at network layer.
5.	Throughput and delay is more.	Throughput and delay is less.

University Questions

1. *Compare between IPv4 and IPv6.* **SPPU : May-12, Marks 8**
2. *Compare IPv4 and IPv6.* **SPPU : Dec.-12, May-13, Marks 10**

4.6 Transition from IPv4 to IPv6

- Three strategies have been devised by the IETF to help the transition.
  1. Dual stack
  2. Tunneling
  3. Header translation

4.6.1 Dual Stack

- All the host must run IPv4 and IPv6 simultaneously until all the Internet uses IPv6.
- Fig. 4.6.1 shows the dual stack.
- To determine which version to use when sending a packet to destination, the source host queries the DNS. If the DNS returns an IPv4 address, the source host sends an IPv4 packet. If the DNS returns an IPv6 address, the source host sends an IPv6 packet.

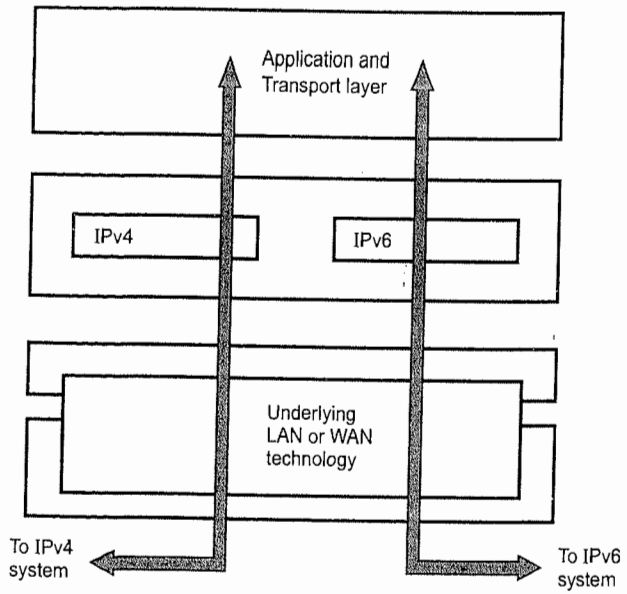


Fig. 4.6.1 Dual stack

Tunneling

- When two computers using IPv6 want to communicate with each other and the packet must pass through a region that uses IPv4. The IPv6 packet is encapsulated in an IPv4 packet when it enters the region, and it leaves its capsule when it exits the region.
- Fig. 4.6.2 shows the tunneling.

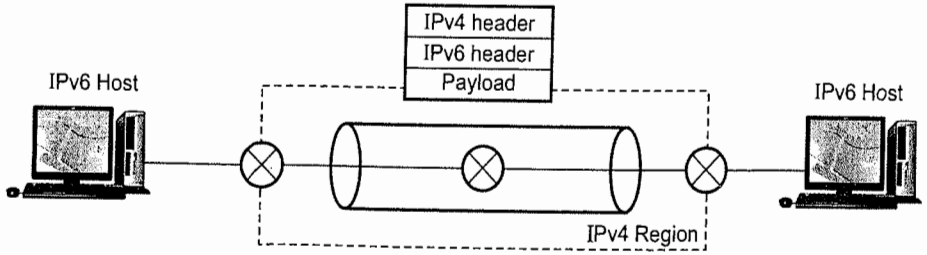


Fig. 4.6.2 Tunneling

Header Translation

- Header translation is used when some of the system uses IPv4. The sender wants to use IPv6, but the receiver does not understand IPv6.

- Fig. 4.6.3 shows the header translation.

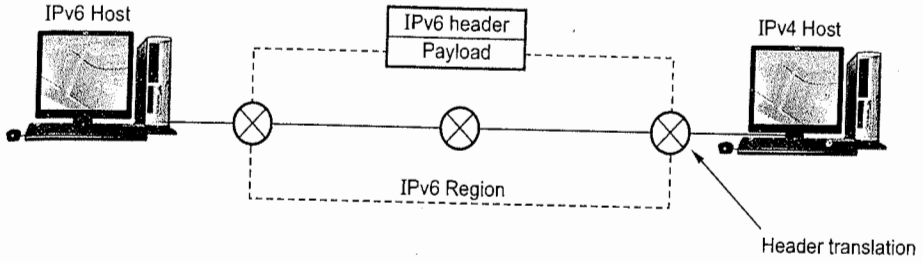


Fig. 4.6.3 Header translation

- The header format must be totally changed through header translation. The header of the IPv6 packet is converted to an IPv4 header.

4.6.2 Comparison between IPv4 and IPv6

Sr. No.	IPv4	IPv6
1	Header size is 32 bits.	Header size is 128 bits.
2	It cannot support autoconfiguration	Supprots autoconfiguration
3	Cannot support real time application.	Supports real time application.
4	No security at network layer.	Provides security at network layer.
5	Throughput and delay is more.	Throughput and delay is less.

4.7 Forwarding

SPPU : May-12

- Forwarding refers to the way a packet is delivered to the next node. It requires a host or router to have a routing table.
- Forwarding refers to the router local action of transferring a datagram from an input link interface to the appropriate output link interface.
- When host has a packet to send, it looks at routing table to find the route to the final destination.

Types of forwarding techniques

1. Next hop versus route method.
2. Network specific versus host specific method.
3. Default method.

1) Next hop versus route method

- Fig. 4.7.1 shows network with routing table for this method. This method reduce the content of rounting table. Routing table stores only the address of the next hop.

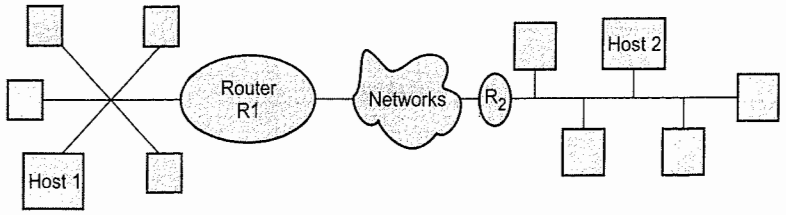


Fig. 4.7.1

Routing table for next hop

For host 1

Destination address	Next hop
Host 2	R1

For router R2

Destination Address	Next hop
Host 2	R2

For Router R2

Destination Address	Next hop
Host 2	-

2) Network specific versus host specific method

- It simplify the searching process and also reduce the routing table size.

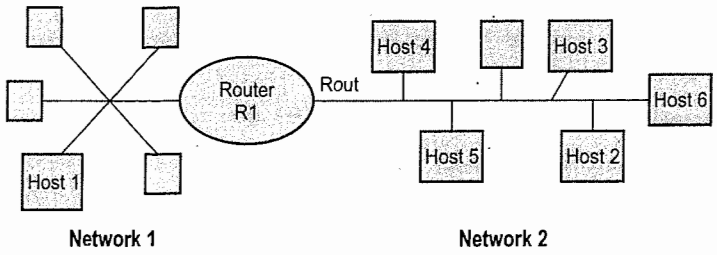


Fig. 4.7.2

- Routing table contains only the address of the destination network.
- It provides good security.

Routing table for host 1

Destination address	Next hop
Network 2	R1

3) Default method

- Host is connected with more than one routers.
- A router is assigned to receive all packets with no match in the routing table.
- Default router is used for communication with outside world.

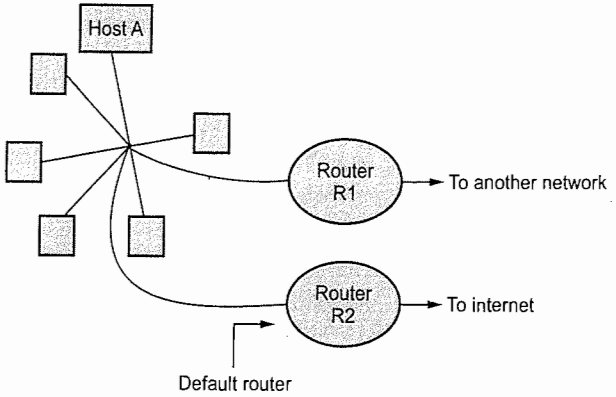


Fig. 4.7.3

University Question

1. *Explain different forwarding techniques used in computer network.* **SPPU : May-12, Marks 7**

4.8 Routing

SPPU : May-14

- A host or a router has a routing table with an entry for each destination, or a combination of destinations, to route IP packets. Routing table can be either static or dynamic.
- A static routing table contains information entered manually. The administrator enters the route for each destination into the table.
- Dynamic routing table is updated periodically by using one of the dynamic routing protocols such as RIP, OSPE or BGP.
- The main function of the network layer is to route packets from source to destination. To accomplish this a route through the network must be selected, generally more than one route is possible. The selection of route is generally based on some performance criteria. The simplest criteria is to choose shortest root through the network.

- The shortest root means a route that passes through the least number of nodes. This shortest root selection results in least number of hops per packet. A routing algorithm is designed to perform this task. The routing algorithm is a part of network layer software.

Properties of routing algorithm

Certain properties which are desirable in a routing algorithm are -

Correctness, simplicity, robustness, stability, fairness, optimality and efficiency.

1. Correctness and simplicity are self-explanatory.
2. Robustness means the ability to cope with changes in the topology and traffic without requiring all jobs in hosts to be aborted and network to be rebooted everytime.
3. Stability refers to equilibrium state of algorithm. It is the technique that react to changing conditions such as congestions. Under any conditions the network must not react too slow or experience unstable swings from one extreme to another.
4. Some performance criteria may favour the exchange of data packets between nearby stations and discourage the exchange between distant stations. Some compromise is needed between fairness and optimality.

Routing algorithm classification

Routing algorithm can be classified in several ways. Based on their responsiveness it can be classified into two types -

1. Static (non-adaptive) routing algorithms.
2. Dynamic (adaptive) routing algorithms.

1. Static (non-adaptive) routing algorithms

In static routing the network topology determines the initial paths. The precalculated paths are then loaded to the routing table and are fixed for a longer period. Static routing is suitable for small networks. Static routing becomes **cumber some** for bigger networks.

The disadvantage of static routing is its inability to respond quickly to network failure.

2. Dynamic (Adaptive) routing algorithms

Dynamic routing algorithms change their routing decision if there is change in topology, traffic. Each router continuously checks the network status by communicating with neighbours. Thus a change in network topology is eventually propagated to all the routers. Based on this information gathered, each router computes the suitable path to the destination.

The disadvantage of dynamic routing is its complexity in the router.

Routing tables

Once the routing decision is made, this information is to be stored in routing table so that the router knows how to forward a packet. In virtual circuit packet switching, the routing table contains each incoming packet number and outgoing packet number and output port to which the packet is to forward. In datagram networks, routing table contains the next hop to which to forward the packet, based on the destination address.

4.8.1 Advantages and Disadvantages of Static Routing

Advantages

1. Minimal CPU/Memory overhead.
2. Granular control on how traffic is routed.
3. Simple to configure and maintain.
4. Secure as only defined routes can be accessed.
5. Bandwidth is not used for sending routing updates.

Disadvantages

1. Manual update of routes after changes
2. Explicit addition of routes for all networks
3. Impractical on large network.

4.8.2 Advantages and Disadvantages of Dynamic Routing

Advantages

1. Simpler to configure on larger networks.
2. Will dynamically choose a different (or better) route if a link goes down.
3. Ability to load balance between multiple links.

Disadvantages

1. Updates are shared between routers, thus consuming bandwidth.
2. Routing protocols put additional load on router CPU/RAM.
3. The choice of the "best route" is in the hands of the routing protocol, and not the network administrator.

4.8.3 Difference between Static and Dynamic Routing

Sr. No.	Static routing (Non adaptive)	Dynamic routing (Adaptive)
1.	Static routing manually sets up the optimal paths between the source and the destination computers.	Dynamic routing uses dynamic protocols to update the routing table and to find the optimal path between the source and the destination computers.
2.	The routers that use the static routing algorithm do not have any controlling mechanism if any faults in the routing paths.	The dynamic routing algorithms are used in the dynamic routers and these routers can sense a faulty router in the network.
3.	These routers do not sense the faulty computers encountered while finding the path between two computers or routers in a network.	the dynamic router eliminates the faulty router and finds out another possible optimal path from the source to the destination.
4.	The static routing is suitable for very small networks and they cannot be used in large networks	Dynamic routing is used for larger networks.
5.	The static routing is the simplest way of routing the data packets from a source to a destination in a network.	The dynamic routing uses complex algorithms for routing the data packets.
6.	The static routing has the advantage that it requires minimal memory.	Dynamic routers have quite a few memory overheads, depending on the routing algorithms used.
7.	The network administrator finds out the optimal path and makes the changes in the routing table in the case of static routing.	In the dynamic routing algorithm, the algorithm and the protocol is responsible for routing the packets and making the changes accordingly in the routing table.

4.8.4 Design Goals

Routing algorithms often have one or more of the following design goals :

1. Optimality
2. Simplicity and low overhead
3. Robustness and stability
4. Rapid convergence
5. Flexibility.

1. Optimality : Optimality refers to the ability of the routing algorithm to select the best route. The best route depends on the metrics and metric weightings used to make the calculation. For example, one routing algorithm might use number of hops and delay, but might weight delay more heavily in the calculation. Naturally, routing protocols must strictly define their metric calculation algorithms.

2. Simplicity : Routing algorithms are also designed to be as simple as possible. In other words, the routing algorithm must offer its functionality efficiently, with a minimum of software and utilization overhead. Efficiency is particularly important when the software implementing the routing algorithm must run on a computer with limited physical resources.

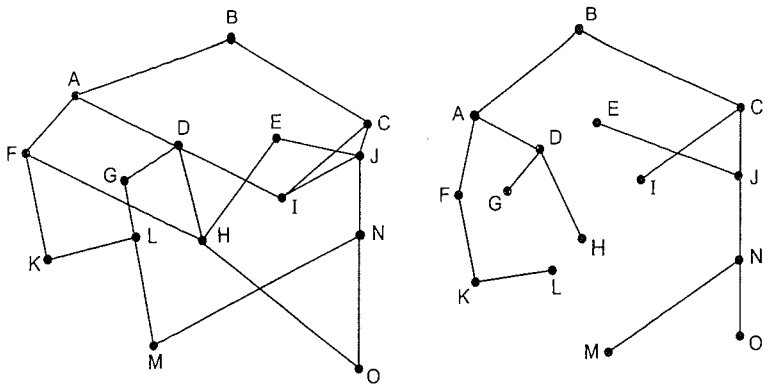
3. Robustness : Routing algorithms must be robust. In other words, they should perform correctly in the face of unusual or unforeseen circumstances such as hardware failures, high load conditions and incorrect implementations. Because routers are located at network junction points, they can cause considerable problems when they fail. The best routing algorithms are often those that have withstood the test of time and proven stable under a variety of network conditions.

4. Rapid convergence : Routing algorithms must converge rapidly. Convergence is the process of agreement, by all routers, on optimal routes. When a network event causes routes to either go down or become available, routers distribute routing update messages. Routing update messages permeate networks, simulating recalculation of optimal routes and eventually causing all routers to agree on these routes. Routing algorithms that converge slowly can cause routing loops or network outages.

5. Flexibility : Routing algorithms should also be flexible. In other words, routing algorithms should quickly and accurately adapt to a variety of network circumstances. For example, assume that a network segment has gone down. Many routing algorithms, on becoming aware of this problem, will quickly select the next-best path for all routes normally using that segment. Routing algorithms can be programmed to adapt to changes in network bandwidth, router queue size, network delay, and other variables.

4.8.5 Optimally Principle

- Optimally principle states that if router J is on the optimal path from router I to router K, then the optimal path from J to K also falls along the same route.
- Suppose route from I to J is r_1 and rest of the route is called r_2. If a route better than r_2 is existed from J to K, it could be concatenated with r_1 to improve the route from I to K, so that r_1r_2 is optimal.
- Fig. 4.5.1 shows the subnet and sink tree with distance metric is measured as the number of hops.
- Sink tree is not necessarily unique, other trees with the same path lengths may exist.
- Sink tree does not contain any loops, so each packet will be delivered within a finite and bounded number of hops.



(a) Subnet (b) Sink tree for router B

Fig. 4.8.1 Subnet and sink tree

University Questions

1. *Explain routing and routed protocol.* **SPPU : May-14, Marks 4**
2. *Compare static and dynamic routing algorithm with suitable example.* **SPPU : May-14, Marks 6**

4.9 Unicast Routing Protocol

SPPU : May-13, 14, Dec.-13

- Routing table can be static or dynamic. Manual entries are done in static table.
- Dynamic table is updated automatically when there is a change somewhere in the internet.
- Now a day, dynamic table is used because of sudden changes in the internet. One of the routers in the internet may fail or link between any two routers is down. So because of these reasons dynamic table is required.
- Routing protocol is a combination of rules and procedures that let routers in the internet inform each other of changes.

4.9.1 Intra and Inter-domain Routing

- An internet is divided into autonomous systems. An **autonomous system** is a group of networks and routers under the authority of a single administration.
- Routing inside an autonomous system is referred to as **intradomain** routing.
- Routing between autonomous system is referred to as **interdomain** routing.
- Distance vector and link state routing is the example of intradomain routing protocols.
- Path vector is an example of interdomain routing protocol.

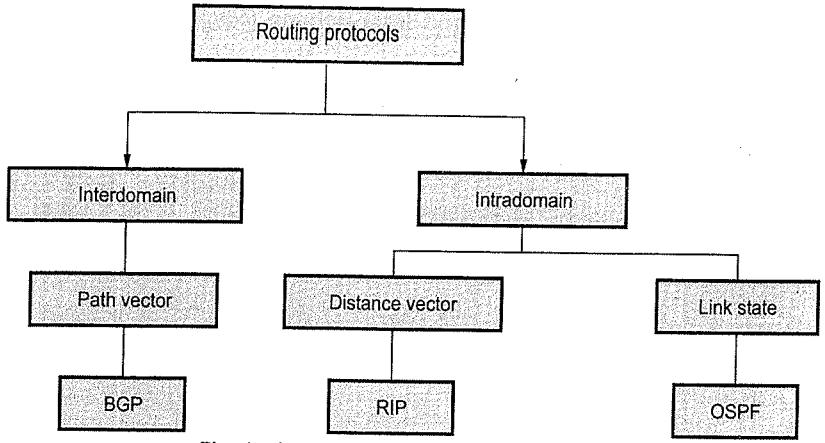


Fig. 4.9.1 Classification of routing protocols

- Fig. 4.9.2 shows an autonomous system.

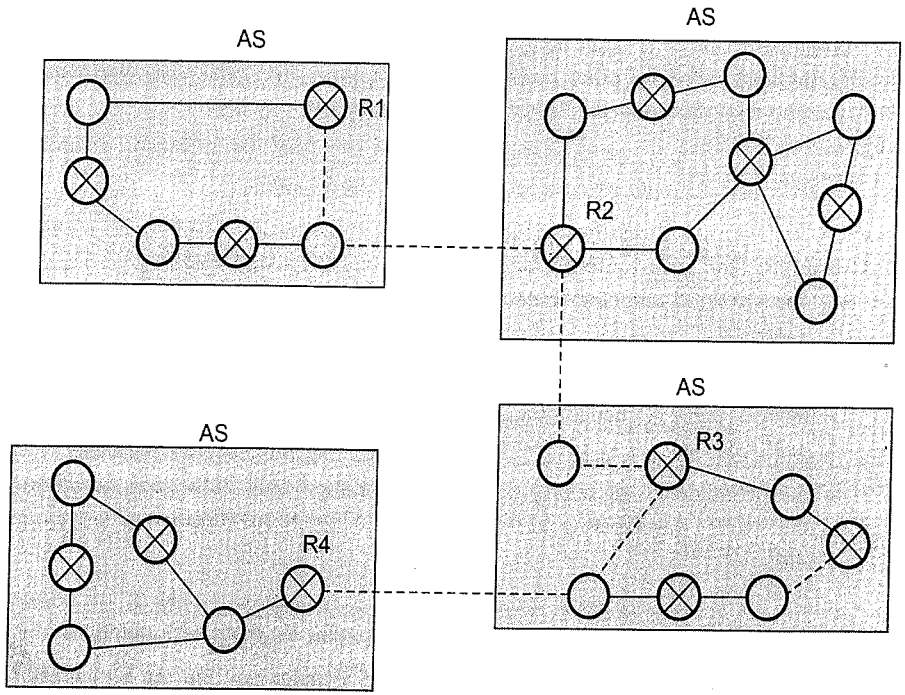


Fig. 4.9.2 Autonomous system

- Only one interdomain routing protocol handles routing between autonomous systems.

4.9.2 Comparison between Intra and Inter-domain Routing

Sr. No.	Intra-domain routing	Inter-domain routing
1.	Routing within an AS.	Routing between AS's.
2.	Ignores the Internet outside the autonomous system.	Assumes that the Internet consists of a collection of interconnected AS's.
3.	Protocols for Intra-domain routing are also called Interior Gateway Protocols.	Protocols for inter-domain routing are also called Exterior Gateway Protocols.
4.	Popular protocols are RIP and OSPF.	Routing protocols are BGP.

4.9.3 Distance Vector Routing

- Distance vector routing algorithm is the dynamic routing algorithm. It was designed mainly for small network topologies. Distance vector routing algorithm is sometimes called by other names, most commonly the distributed **Bellman-Ford** routing algorithm and the **Ford-Fulkerson** algorithm.
- The term distance vector derives from the fact that the protocol includes its routing updates with a vector of distances, or hop counts.
- In this algorithm, each router maintains a routing table indexed by, and containing one entry for, each router in the subnet. This entry contains two parts :
  a. The preferred outgoing line to use for that destination.
  b. An estimate of the time or distance to that destination.
- The metric used might be number of hops, time delay in milliseconds, total number of packets queued along the path, etc.
- Assume that delay is used as a metric and that the router knows the delay to each of its neighbours. All nodes exchange information only with their neighbouring nodes. Nodes participating in the same local network are considered neighbouring nodes.
- Once every 'T' msec each router sends to each neighbor a list of its estimated delays to each destination. It also receives a similar list from each neighbor.
- By performing calculation for each neighbour, a router can find out which estimate seems the best and use that estimate and the corresponding line in its new routing table. Old routing table is not used in the calculation.
- Fig. 4.9.3 shows the subnet with 12 routers.

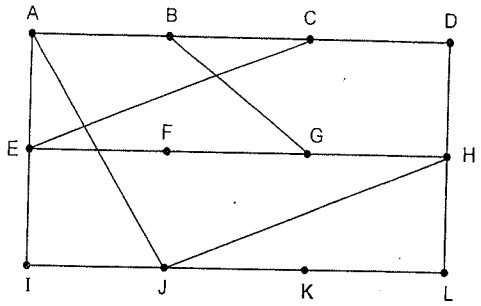


Fig. 4.9.3 Subnet

Routing table is shown below.

To	A	I	H	K
A	0	24	20	21
B	12	36	31	28
C	25	18	19	36
D	40	27	8	24
E	14	7	30	22
F	23	20	19	40
G	18	31	6	31
H	17	20	0	19
I	21	0	14	22
J	9	11	7	10
K	24	22	22	0
L	29	33	9	9

JA delay is 8 | JI delay is 10 | JH delay is 12 | JK delay is 6

Vectors received from J's four neighbors

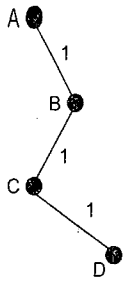
New estimated delay from J

Line
A
A
I
H
I
I
H
H
I
–
K
K

New routing table for J

4.9.3.1 Count-to-Infinity Problem

- Fig. 4.9.4 shows an imagined network and denotes the distances from router A to every other router. Until now everything works fine.

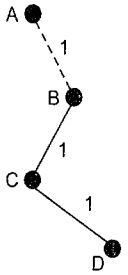


Routing table for A

B	C	D
1	2	3

Fig. 4.9.4

- Suppose that link (A, B) is broken. Router B observed it, but in his routing table he sees, that router C has a route to A with 2 hops. The problem is, that B does not know that C has router B as successor in his routing table on the route to A. That followed **count-to-infinity** problem. Router B actualizes his routing table and takes the router to A over router C.



B	C	D
3	2	3

B	C	D
3	4	3

B	C	D
5	4	5

Fig. 4.9.5

- In Fig. 4.9.5, we can see the new distances to A. In router C's routing the route to A contains router B as next hop router, so if B has increased his costs to A, C is forced to do so. Router C increases his cost to A about B + 1 = 4.
- Now we see the conseuqence of the distributed Bellman-Ford protocal : Because router B takes the path over C to A, he reactualizes his routing table and so on.
- There are several partial solutions to the count-to-infinity problem. The first one is to use some relatively small number as an approximation of infinity. For example, we might decide that the maximum number of hops to get across a certain network is never going to be more than 16 and so we could pick 16 as the value that represents infinity.
- This at least bounds the amount of time that it takes to count to infinity. Of course, it could also present a problem if our network grew to a point where some nodes were separated by more than 16 hops.
- One technique to improve the time to stabilize routing is called **split horizon**. Split horizon technique implies that routing information about some network stored in the routing table of a specific router is never sent to the router from which it was received.

Issues with the Distance Vector Routing

1. The primary drawback of this algorithm is its vulnerability to the 'Count-to-Infinity' problem. There have been proposed many partial solutions but none works under all circumstances.
2. Another drawback of this scheme is that it does not take into account Link Bandwidth.
3. Yet another problem with this algorithm is that it takes appreciably long time for convergence as the network-size grows.
4. A fallout of the Count-to-Infinity issue and slow convergence has been to limit the maximum number of hops to 15 which means more than 16-router subnets, it may not be appropriate routing algorithm.

4.9.3.2 Routing Information Protocol

- In RIP, routing updates are exchanged between neighbours approximately every 30 seconds using a so-called **RIP response message**. The response message sent by a router or host contains a list of upto 25 destination networks within an autonomous system (AS). Response messages are also known as **RIP advertisements**.
- Fig. 4.9.6 shows a portion of an autonomous system.

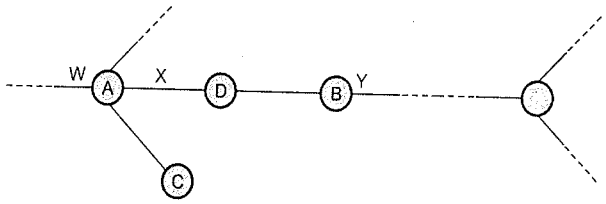


Fig. 4.9.6 Portion of AS

- Forwarding table in router D before receiving advertisement from router A. For this example, the table indicates that to send a datagram from router D to destination network W, the datagram should first be forwarded to neighbouring router A; the table also indicates that destination network W is two hops away along the shortest path.
- The Table 4.9.1 also indicates that network Z is seven hops away via router B.

Destination network	Next router	Number of hops to destination
W	A	2
Y	B	2

Z	B	7
X	–	1
.....	.....	.....

Table 4.9.1 Forwarding table

- Now suppose that 30 seconds later, router D receives from router A the advertisement shown in Table 4.9.1

Destination network	Next router	Number of hops to destination
Z	C	4
W	–	1
X	–	1
.....	.....	.....

Table 4.9.2 Advertisement from router A

- Note that the advertisement is nothing other than the forwarding table information from router A. This information indicates, in particular, that network Z is only four hops away from router A. Router D, upon receiving this advertisement, merges the advertisement with the old routing table.
- Router D learns that there is now a path through router A to network Z that is shorter than the path through router B. Thus, router D updates its forwarding table to account for the shorter shortest path, as shown in Table 4.9.3

Destination network	Next router	Number of hops to destination
W	A	2
Y	B	2
Z	A	3
.....	.....	.....

Table 4.9.3

- RIP routers exchange advertisements approximately every 30 seconds. If a router does not hear from its neighbour atleast once every 180 seconds, that neighbour is considered to be no longer reachable; i.e. either the neighbour has died or the connecting link has gone down. When this happens, RIP modifies the local

forwarding table and then propagates this information by sending advertisements to its neighbouring routers.

- A router can also request information about its neighbour's cost to a given destination using RIP's request message. Routers send RIP request and response messages to each other over UDP using port number 520.

RIP Message Format

- Fig. 4.9.7 shows the RIP message format.

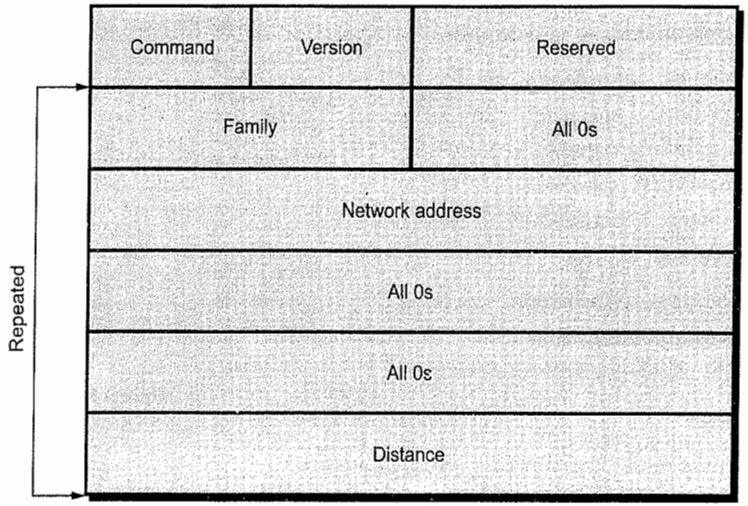


Fig. 4.9.7 RPI message format

1. **Command :** This is 8 bits field specifies the type of message: 1 for request and 2 for response.
2. **Version :** This is 8 bits field define the version.
3. **Family :** This 16 bits field defines the family of the protocol used. For TCP/IP the value is 2.
4. **Network address :** The address field defines the address of the destination network.
5. **Distance :** This 32 bits field defines the hop count from the advertising router to the destination network.

Request and Response

- RIP support two types of messages : Request and Response.

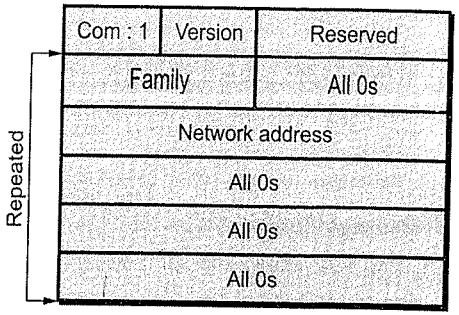
Request

- A request message is sent by a router that has just comp up or by a router that has some time out entries.

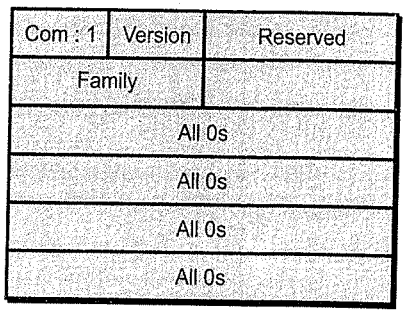
Response

- A response message can be either solicited or unsolicited.
  1. Solicited response
- Is sent only in answer to a request.
  - Containing information about the destination specified in the corresponding request.
  2. Unsolicited response
  - Is sent periodically, every 30 seconds.
  - **Containing** information covering the whole routing table

Fig. 4.9.8 shows the request message.



(a) Request for some



(b) Request for all

Fig. 4.9.8 Request message format

Timers in RIP

- RIP uses three timers to support its operation.
  1. Periodic timer (25 - 35 sec)
  2. Expiration (180 sec)
  3. Garbage collection (120 sec).

1. **Periodic timer** : This type of timer controls the advertising of regular update messages. Each router has one periodic timer that is randomly set to a number between 25 to 35 seconds.
2. **Expiration timer :** The expiration timer governs the validity of a route. In normal situation, the new update for the route occurs every 30 seconds. But, if there is a problem on an internet and no update is received within the allotted 180 seconds, the route is considered expired and the hop count of the route is set to 16. Each router has its own expiration timer.
3. **Garbage collection timer :** When the information about a route becomes invalid, the router continues to advertise the route with a metric value of 16 and the garbage collection timer is set to 120 sec for that route. When the count reaches zero, the route is purged from the table.

RIPv2

- RIP version 2 was designed to overcome some of the shortcomings of version 1. Replaced fields in version 1 that were filled with 0s for the TCP/IP protocols with some new fields.
- **Advantages**
  1. An AS can include several hundred routers with RIP-2 protocol.
  2. Compatible upgrade of RIPv1 including subnet routing, authentication, CIDR aggregation, route tags and multicast transmission.
  3. Subnet support : Uses more convenient partitioning using variable-length subnets
  4. An end system can run RIP in passive mode to listen for routing information without supplying any.
  5. Low requirement in memory and processing at the node .
  6. RIP and RIP2 are for the IPv4 network while the RIPng is designed for the IPv6 network.

Fig. 4.9.9 shows the message format.

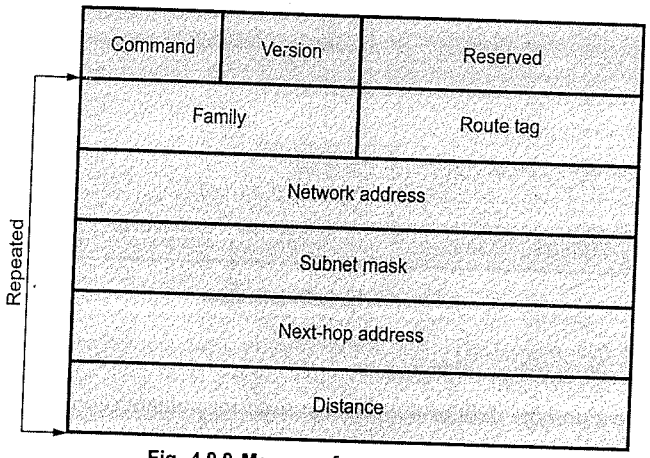


Fig. 4.9.9 Message format of RIPv2

1. Command - The command field is used to specify the purpose of the datagram.
2. Version - The RIP version number. The current version is 2.
3. Identifier - Indicates what type of address is specified in this particular entry.
4. Route tag - Attribute assigned to a route which must be preserved and readvertised with a route. The route tag provides a method of separating internal RIP routes from external RIP routes, which may have been imported from an EGP or another IGP.

5. IP address - The destination IP address.
6. Subnet mask - Value applied to the IP address to yield the non-host portion of the address. If zero, then no subnet mask has been included for this entry.
7. Next hop - Immediate next hop IP address to which packets to the destination specified by this route entry should be forwarded.
8. Distance - Represents the total cost of getting a datagram from the host to that destination.

Authentication

- Authentication is added to protect the message against unauthorized advertisement. No new field is added to the packet.
- To indicate that the entry is authentication information and not routing information, the value of FFFFH is entered in the family field.

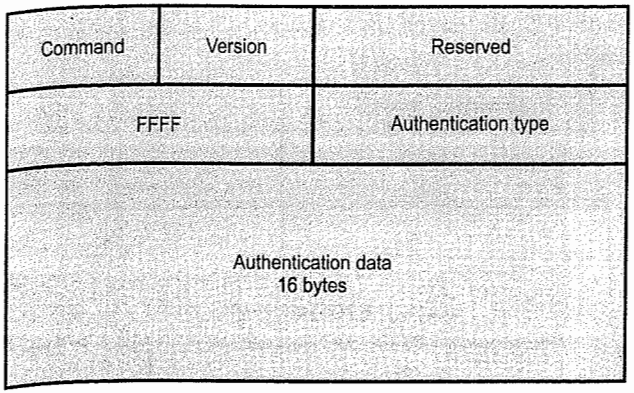


Fig. 4.9.10 Authentication

- Fig. 4.9.10 shows the authentication.
- Authentication type defines the protocol used for authentication.
- Authentication data is the actual data.

RIP2 - Disadvantages

1. RIP2 supports generic notion of authentication, but only "password" is defined so far. Still not very secure.
2. RIP2 packet size increases as the number of networks increases hence it is not suitable for large networks.

3. RIP2 generates more protocol traffic than OSPF, because it propagates routing information by periodically transmitting the entire routing table to neighbour routers.
4. RIP2 may be slow to adjust for link failures.

Advantages of RIP and Disadvantages of RIP version 1

Advantages of RIP

1. RIP is very useful in a small network, where it has very little overhead in terms of bandwidth used and configuration and management time.
2. Easy to implement than newer IGP's.
3. Many implementations are available in the RIP field.

Disadvantages of RIP1

1. Minimal amount of information for router to route the packet and also very large amount of unused space.
2. Subnet support : Supports subnet routes only within the subnet network.
3. Not secure; anyone can act as a router just by sending RIP1 messages.

RIP1 was developed for an AS that originally included less than a 100 routers.

4.9.4 Link State Routing

- Link state routing is the second major class of intradomain routing protocol. It is **dynamic** type routing algorithm.
- The idea behind link state routing is simple and can be stated as five parts. Each router must do the following :
  1. **Learning about the neighbors :** When a router is booted, it sends a special HELLO packet on each point-to-point line. The router on the other end is expected to send back a reply telling who it is. When two or more routers are connected by a LAN, the LAN can be modeled as a node.
  2. **Measuring line cost :** To determine the cost for a line, a router sends a special ECHO packet over the line that the other side is required to send back immediately. By measuring the round-trip time and dividing it by two, the sending router can get a reasonable estimate of the delay. Should the load be taken into account when measuring the delay ?
  3. **Building link state packets :** State packets may be built periodically, or when some significant event occurs, such as a line or neighbor going down or coming back up again.

4. **Distributing the link state packets :** The basic algorithm

- Each state packet contains a sequence number that is incremented for each new packet sent.
- Routers keep track of all the (source router, sequence) pairs they see.
- When a new link state packet comes in, it is checked against the list of packets already seen. If it is new, it is forwarded on all lines except the one it arrived on (i.e., flooding). If it is a duplicate, it is discarded. If a packet with a sequence number lower than the highest one seen so far ever arrives, it is rejected as being obsolete.

Problems with the basic algorithm :

1. The sequence numbers may wrap around, causing confusion. Solution : using a 32-bit sequence number. With one packet per second, it would take 137 years to wrap around.
2. If a router ever crashes, it will lose track of its own sequence number. If it starts again at the sequence number 0, new packets will be rejected as obsolete/duplicate by other routers.
3. If a sequence number is ever corrupted and 65,540 is received instead of 4 (a 1-bit error), packets 5-65540 will be rejected as obsolete.

The solution to router crashes and sequence number corruption is to associate an age with each state packet from any router and decrement the age once per second. When the age hits zero, the information from that router is discarded. Normally a new packet comes in every 10 seconds, so router information only times out when a router is down.

Some refinements to the basic algorithm make it more robust

When a state packet comes into a router for flooding, it is put in a holding area to wait a short while first. If another state packet from the same source comes in before it is transferred, their sequence numbers are compared. If they are equal, the duplicate is discarded. If they are different, the older one is thrown out. To guard against errors on the lines, all state packets are acknowledged. When a line goes idle, the holding area is scanned in round robin to select a packet or acknowledgement to send.

5. **Computing the new routes :** Once a router has accumulated a full set of link state packets, it can construct the entire subnet graph. Then Dijkstra's algorithm can be run locally to construct the shortest path to all possible destinations.

- Link state routing protocols use event driven updates rather than periodic updates. Link state routing is widely used in actual networks. OSPF protocol uses in a link state algorithm.

- Link state routing protocols are as follows :
  a. Open Shortest Path First (OSPF)
  b. Netware Link Services Protocol (NLSP).
  c. Apple's AURP.
  d. ISO's Intermediate System-Intemediate System (IS-IS).

4.9.4.1 Shortest Path Routing

- In shortest path routine the path length between each node is measured as a function of distance, bandwidth, average traffic, communication cost, mean queue length, measured delay etc.
- By changing the weighing function, the algorithm then computes the shortest path measured according to any one of a number of criteria or a combination of criteria. For this a graph of subnet is drawn. With each node of graph representing a router and each arc of the graph representing a communication link. Each link has a cost associated with it. To choose a route between a given pair of routers, the algorithm just finds the shortest path between them on the graph.

Two algorithms for computing the shortest path between two nodes of a graph are known.

i) Dijkstra's algorithm ii) Bellman-Ford algorithm.

i) Dijkstra's algorithm :

Each node is labelled with its distance from the source node along the best known path. Initially no paths are known, so all nodes are labelled with infinity. The algorithm proceed in stages. As the algorithm proceeds, the paths are found, the labels are changed, reflecting better paths. Stepwise proceeding of algorithm is as follows.

Step-I : Source node is initialized and can be indicated as a filled circle.

Step-II : Initial path cost to neighbouring nodes (adjacent nodes) or link cost is computed and these nodes are relabelled considering source node.

Step-III : Examine the all adjacent nodes and find the smallest label, make it permanent.

Step-IV : The smallest label node is now working node, then step-II and step-III are repeated till the destination node reaches.

Following example illustrates Dijkstra's algorithm.

Example 4.9.1 *Find the shortest path between node A and node H for the following Fig. 4.9.11 by applying Dijkstra's algorithm. Show each steps output.*

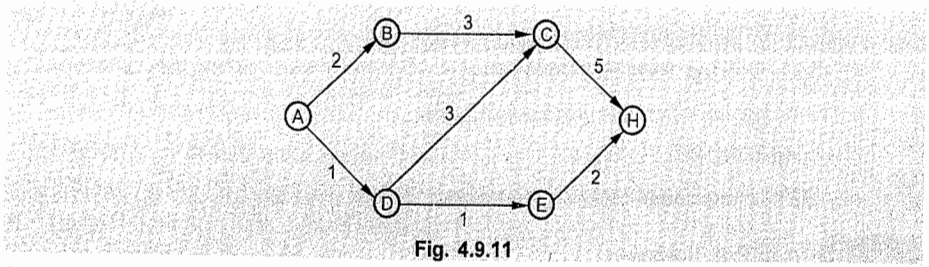


Fig. 4.9.11

Solution :

Step-I : Node A is initialized as source node.

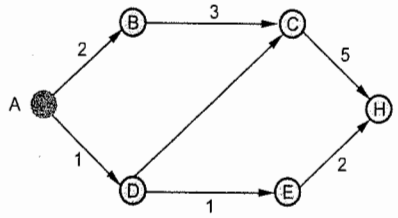


Fig. 4.9.11 (a)

Step-II : Link cost is computed for the adjacent node.

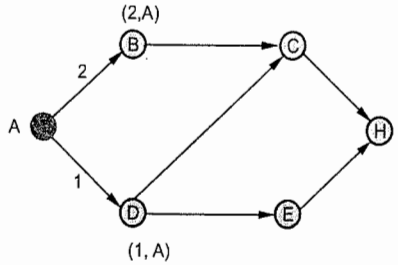


Fig. 4.9.11 (b)

Step-III : Since AD is smallest path, now D is working node.

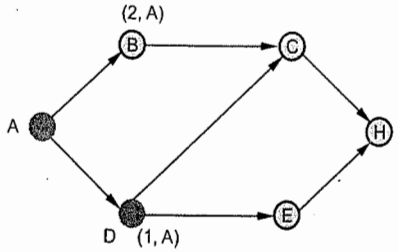


Fig. 4.9.11 (c)

Step-IV : Adjacent nodes to D are C and E.

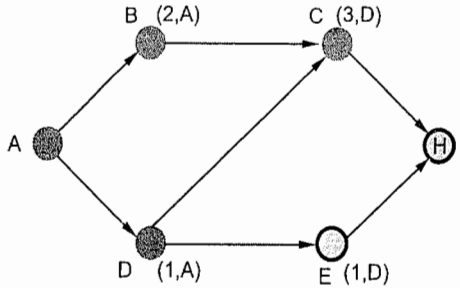


Fig. 4.9.11 (d)

Step-V : Since shortest is E, now E is working node.

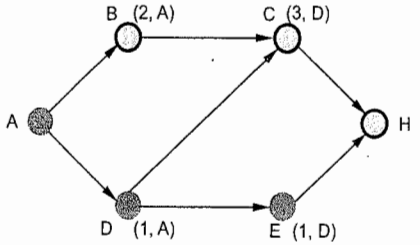


Fig. 4.9.11 (e)

Step-VI :

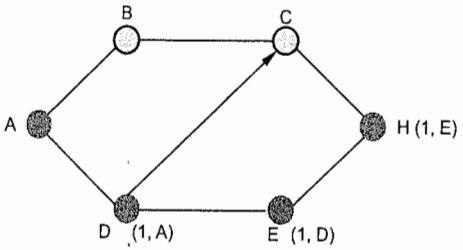


Fig. 4.9.11 (f)

Hence the shortest path between node A and node H is ADEH.

ii) Bellman-Ford algorithm :

Bellman-Ford algorithm is somewhat similar to Dijkstra's algorithm but here the shortest paths from a given source node is computed subject to the constraint that the path contain at most one link, i.e. from source node, at each step least-cost path with maximum number of links are found. Finally the least-cost path to each node and the cost of that path is computed. Bellman-Ford algorithm is illustrated in the following example.

4. An Autonomous System Boundary Router (ASBR) is a router that has its links connected to another autonomous system.

- As shown in the Fig. 4.9.13 routers R_1, R_2 and R_7 are internal routers. Routers R_3, R_6, R_8 are area border routers. Routers R_3, R_4, R_5, R_6, R_8 are backbone routers. Router R_4 is an ASBR.

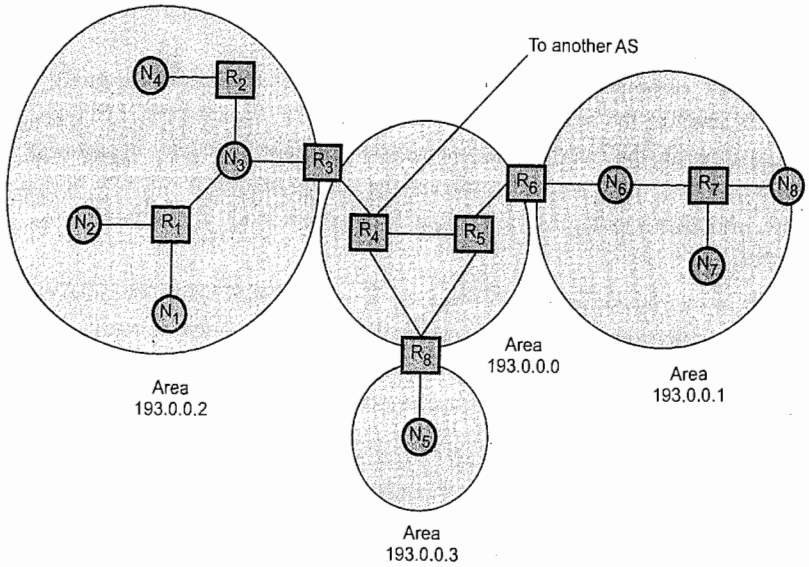


Fig. 4.9.13 OSPF areas

- A hello protocol allows neighbours to be discovered automatically. Two routers are said to be neighbours if they have an interface to a common network. The OSPF protocol runs directly over IP, using IP protocol 89. The header format for OSPF is shown in the Fig. 4.9.14

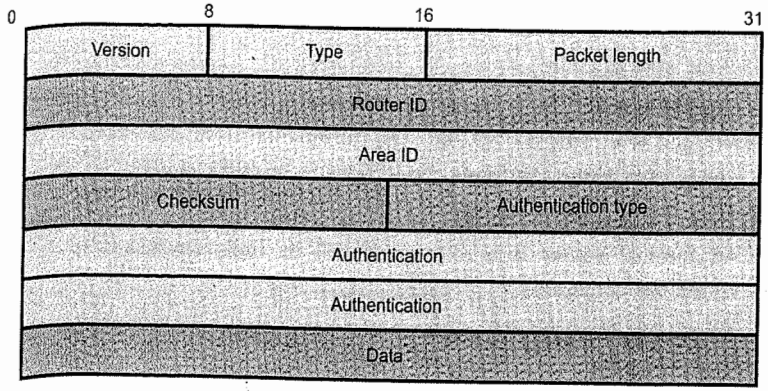


Fig. 4.9.14 OSPF common header

- OSPF header analysis is given below :

1. **Version :** This field specifies the protocol version.
2. **Type :** This field indicates messages as one of the following type.
   a. Hello
   b. Database description
   c. Link status
   d. Link status update
   e. Link status acknowledgement.
3. **Packet length :** This field specifies the length of OSPF packet in bytes, including the OSPF header.
4. **Router ID :** It identifies the sending router. This field is typically set to the IP address of one of its interfaces.
5. **Area ID :** This field identifies the area this packet belongs to (Transmitted).
6. **Checksum :** The checksum field is used to detect errors in the packet. The checksum is performed on the entire packet.
7. **Authentication type :** It identifies the authentication type that is used.
8. **Authentication :** This field includes a value from the authentication type.

The OSPF operation consists of the following stages.

1. OSPF send the Hello messages for discovering the neighbours and designated routers are elected in multiaccess networks.
2. Adjacencies are established and link state databases are synchronized.
3. Link state advertisement are exchanged by adjacent routers to allow topological databases to be maintained and to advertise inter area and inter AS routes. The routers use the information in the database to generate routing tables.

OSPF Advantages

1. Low traffic overhead. OSPF is enconomical of network bandwidth on links between routers.
2. Fast convergence. OSPF routers flood updates to changes in the network around the internet, so that all routers quickly agree on the new topology after a failure.
3. Larger network metrics. This allows a network planner the freedom to assign costs for each path around the network, to give fine control over routing paths.

4. Area based topology. Large OSPF networks are organized as a set of areas linked by a backbone. Routing within each area is isolated to minimize cross area discovery traffic.
5. Route summaries. OSPF can minimize the routes propagated across an area boundary by collapsing several related sub-net routes into one. This reduces routing table sizes, and increases the practical size of a network.
6. Support for complex address structures. OSPF allows variable size sub-netting within a network number, and sub-nets of a network number to be physically disconnected. This reduces waste of address space, and makes changing a network incrementally much easier.
7. Authentication. OSPF supports the use of passwords for dynamic discovery traffic, and checks that paths are operational in both directions. The main use for this is to prevent misconfigured routers from "poisoning" the routing tables throughout the internet.

OSPF Disadvantages

1. Memory overhead. OSPF uses a link state database to keep track of all routers and networks within each attached area. With a complex topology, this database can be much larger than the corresponding routing pool, and may limit the maximum size of an area.
2. Processor overhead. During steady state operation the OSPF CPU usage is low, mainly due to the traffic between routers. However, when a topology change is detected, there is a large amount of processing required to support flodding of changes, and recalculation of the routing table.
3. Configuration. OSPF can be complex to configure.

4.9.5 Path Vector Routing

- RIP and OSPF are not suitable for inter-domain routing protocols. Both require homogenous metrics that may be the case within an AS, but we cannot assume then same for several AS systems.
- Flooding the link state information across multiple AS systems is not scalable.
- It is different from the distance vector routing and link state routing. Each entry in the routing table contains the destination network, the next router and the path to reach the destination.
- Distance vector routing is subject to instability if there are more than a few hops in the domain of operation.

- Link state routing needs a huge amount of resources to calculate routing tables. It also creates heavy traffic because of flooding.
- Path vector routing protocol provides information about how to reach a network given a certain router and identifies which autonomous system should be visited.
- The path vector routing is different from distance vector algorithm, in which each path has information about cost and distance.
- BGP is an example of a path vector protocol. In BGP the routing table maintains the autonomous systems that are traversed in order to reach the destination system. Exterior Gateway Protocol (EGP) does not use path vectors.
- In path vector routing we assume there is one node in each autonomous system which acts on behalf of the entire autonomous system. This node is called the **speaker node.**
- The speaker node creates a routing table and sends information to its neighboring speaker nodes in neighboring autonomous systems. The idea is the same as Distance vector routing except that only speaker nodes in each autonomous system can communicate with each other.
- The speaker node sends information of the path, not the metric of the nodes, in its autonomous system or other autonomous systems. The path vector routing algorithm is somewhat similar to the distance vector algorithm in the sense that each border router advertises the destinations it can reach to its neighboring router
- A route is defined as a pairing between a destination and the attributes of the path to that destination, thus the name, path vector routing, where the routers receive a vector that contains paths to a set of destinations.
- The main advantage of a path vector protocol is its flexibility.

4.9.5.1 BGP

- The purpose of an exterior gateway protocol is to enable two different Autonomous System (AS) to exchange routing information so that IP traffic can flow across the autonomous system border. BGP was developed for use in conjunction with internets that employ the TCP/IP protocol suite. The BGP is an interdomain routing protocol that is used to exchange network reachability information among BGP routers (Also called BGP speakers). Each BGP speaker establishes a TCP connection with one or more BGP speakers (routers). Two routers are considered to be neighbours if they are attached to the same subnetwork. If the two routers are in different autonomous systems, they may wish to exchange routing information.

- BGP performs three functional procedures.
  1. Neighbour acquisition
  2. Neighbour reachability
  3. Network reachability.
- Neighbour acquisition procedures used for exchanging the routing information between two routers in different Autonomous Systems (AS). To perform neighbour acquisition, one router sends an open message to another. If the target router accepts the request, it returns a keepalive message in response.
- Once a neighbour relationship is established, the neighbour reachability procedure is used to maintain the relationship. Both sides needs to be assured that the other side still exists and is still engaged in the neighbour relationship. For this purpose, both routers send keepalive messages to each other. Both sides router maintains a database of the subnetworks that it can reach and the perferred route for reaching that subnetwork.
- If the database changes, router issues an update message that is broadcast to all other routers implementing BGP. By the broadcasting of these updata message, all the BGP routers can build up and maintain routing information. BGP connections inside an autonomous system are called internal BGP (iBGP) and BGP connections between different autonomous systems are called external BGP (eBGP).
- Fig. 4.9.15 shows the internal and external BGP.

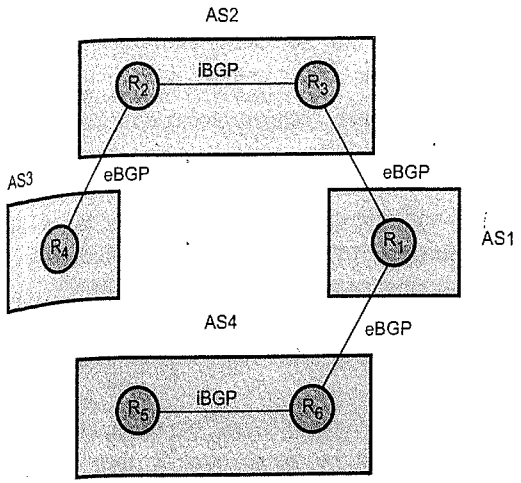


Fig. 4.9.15 Internal and external BGP

BGP messages : Header of the all BGP messages is fixed size that identifies the message type. Fig. 4.9.16 shows the BGP message header format.

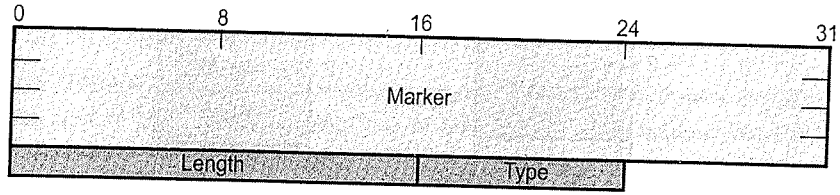
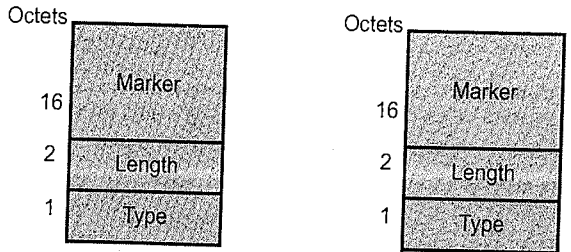


Fig. 4.9.16 BGP header format

1. **Marker :** Marker field is used for authentication. The sender may insert value in this field that would be used as part of an authentication mechanism to enable the recepient to verify the identity of the sender.
2. **Length :** This field indicates the total length of the message in octets, including the BGP header. Value of the length must be between 19 and 4096.
3. **Type :** Type field indicates type of message. BGP defines four message type.
   a) OPEN b) UPDATE c) NOTIFICATION d) KEEPALIVE

- Following Fig. 4.9.17 shows the four types of BGP message formats.



(a) Keepalive

(b) Notification

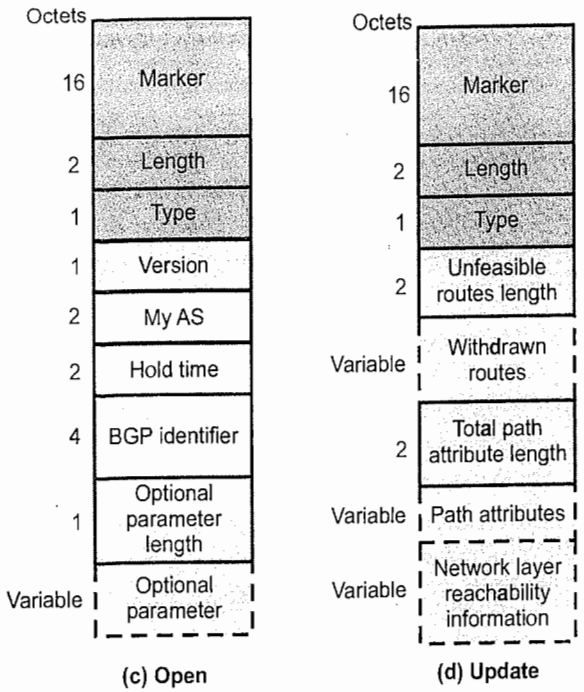


Fig. 4.9.17 BGP message format

- To acquire a neighbour, a router first opens a TCP connection to the neighbour router of interest. It then sends the open message. This message identifies the AS (autonomous system) to which the sender belongs and provides the IP address of the router. It also includes a Hold time parameter. If the recipient is prepared to open a neighbour relationship, it calculate a value of Hold Timer that is the minimum of its Hold Time in the open message. This calculated value is the maximum number of seconds that may elapse between the receipt of successive keepalive and update message by the sender.
- The KEEPALIVE message is just the BGP header with the type field set to 4. The KEEPALIVE messages are exchanged often enough as to not cause the hold timer to expire. A recommended time between successive KEEPALIVE messages is one-third of the hold time interval. This value ensures that KEEPALIVE messages arrive at the receiving router almost always before the hold timer expires even if the transmission delay of a TCP is variable. If the hold time is zero, then KEEPALIVE messages will not be sent.

- When a BGP router detects an error, the router sends a NOTIFICATION message and then close the TCP connection. After the connection is established, BGP peers exchange routing information by using the UPDATE messages.
- The UPDATE messages may contain three pieces of information. Unfeasible routes, path attributes and network layer reachability information
- An UPDATE message can advertise a single route and withdraw a list of route. An update message may contain one or both types of information. The UPDATE messages are used to construct a graph of Autonomous System (AS) connectivity. The withdrawn routes field provides a list of IP address prefixes for the routes that need to be withdrawn from BGP routing tables. The unfeasible routes length field indicates the total length of the withdrawn routes field in actets.
- An UPDATE message can withdraw multiple unfeasible routes from service. A BGP router uses Network Layer Reachability Information (NLRI), the total path attributes length and the path attributes to advertise a route. The NLRI field contains a list of IP address perfixed that can be reached by the route.

Advantages of BGP

1. BGP is a very robust and scalable routing protocol.
2. CIDR is used by BGP to reduce the size of the Internet routing tables.
3. BGP easily solves the count-to-infinity problem.

Disadvantages of BGP

1. BGP is complex.
2. BGP routes to destination networks, rather than to specific hosts or routers.

4.9.6 Comparison between RIP and OSPF

Sr. No.	RIP	OSPF
1.	RIP is easy to configure.	OSPF is complicated to configure and requires network design and planning.
2.	An end system (a system with only one network interface) can run RIP in passive mode to listen for routing information.	OSPF does not have a passive mode.
3.	RIP may be slow to adjust for link failures.	OSPF is quick to adjust for link failures.

4.	RIP generates more protocol traffic than OSPF.	OSPF generates less protocol traffic than RIP.
5.	RIP is not well suited to large networks, because RIP packet size increases as the number of networks increases.	OSPF works well in large networks.
6.	RIP is distance vector routing protocol.	OSPF is link state routing protocol.

4.9.7 Difference between Distance Vector and Link State Routing

Sr. No.	Distance vector	Link state
1.	Bellman-ford algorithm used to calculate the shortest cost path.	Dijkstra's algorithm used to calculate link state cost.
2.	Sends message to their neighbors.	Sends message to every other node in the network.
3.	It is decentralized routing algorithm.	It is centralized global routing algorithm.
4.	Sends larger updates only to neighbouring routers.	Send small updates every where.
5.	Protocol example - RIP	Protocol example - OSPF and BGP
6.	Require less CPU power and less memory space.	Require more CPU power and more memory space.
7.	Simple to implement and support.	Expensive to implement and support.

4.9.8 Hierarchical Routing

- At a certain point the network may grow to the point where it is no longer feasible for every router to have an entry for other router, so the routing will have to be done hierarchically. When this routing is used, the routers are divided into regions. It contains all the details about how to route packets to destinations within its own region.
- Some time, two-level hierarchy may be insufficient. It is necessary to group the regions into clusters, the clusters into zones and zones into groups and so on. Fig.4.9.18 shows routing in a two-level hierarchy.

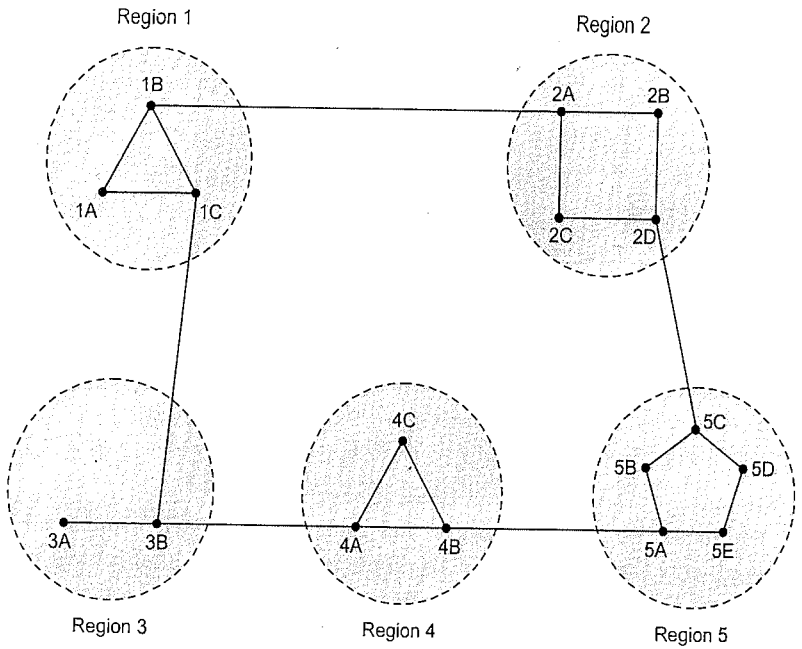


Fig. 4.9.18 Subnet

Hierarchical table for 1A

Dest	Line	Hops
1A	-	-
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	5

Full table for 1A

Dest	Line	Hops
1A	-	-
1B	1B	1
1C	1C	1
2A	1B	2