LOGICAL ADDRESSING

Logical addressing is implemented by network layer.

Logical addressing is a Global Addressing scheme.

Data-link layer handles the addressing problem locally, but if packets passes the network boundary there is a need for logical addressing system to help distinguish source and destination systems.

The network layer adds a header to the packet coming from the upper layer that includes the logical addresses of the sender and receiver.

There are 2 types of addressing mechanisms are present:

1. IPv4 (IPversion4)
2. IPv6 (IP version6)

IPv4 ADDRESSES

An **IPv4** address is a **32-bit** address that **uniquely** and **universally** defines the connection of a device to the Internet.

Unique: Two devices on the Internet can never have the same address at the same time.

Universal: The addressing system must be accepted by any host that wants to be connected to the Internet.

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 (0 or 1) 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** (more than **4 billion**).

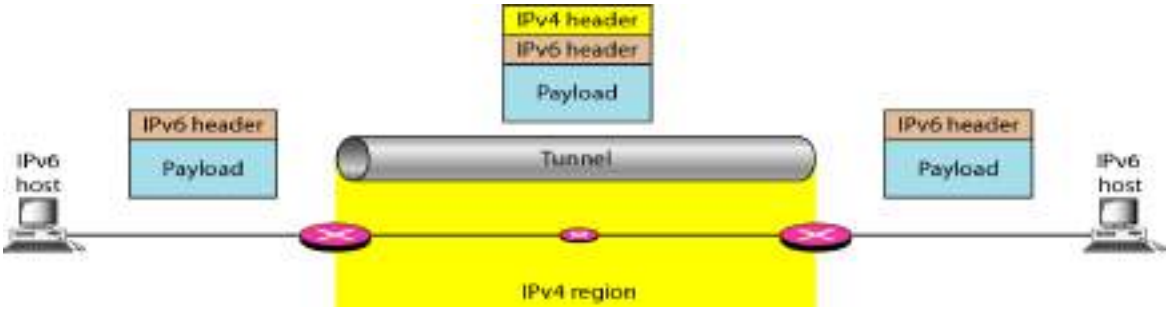
Notations

There are two notations to show an IPv4 address: Binary and Dotted-Decimal Notation.

Binary	Dotted-Decimal
• IPv4 address is displayed as 32 bits. Each octet is often referred to as a byte. • It is a 4 byte address Ex: 10000000 00001011 00000011 00011111	• Internet addresses are written in decimal form with a dot separating the bytes. • Each number in dotted-decimal notation is a value ranging from 0 to255. Ex: 128.11.3.31

Tunneling

- Tunneling is a strategy used when two computers using IPv6 want to communicate with each other and the packet must pass through a region that usesIPv4.
- To pass through this region, the packet must have an IPv4address.
- So the IPv6 packet is encapsulated in an IPv4 packet when it enters the region, and it leaves its capsule when it exits the region.
- It seems as if the IPv6 packet goes through a tunnel at one end and emerges at the other end. The IPv4 packet is carrying an IPv6 packet as data, the protocol value is set to41.



IP (Internet Protocol):

IP (Internet Protocol) was designed as a **best-effort delivery protocol**, but it **lacks some features** such as **flow control** and **error control**. To make IP more responsive it takes the help of other protocols as depicted in Figure 21.1:

- Protocols to create a mapping between physical and logical addresses: **ARP** (Address Resolution Protocol).
- Protocols to create a reverse mapping i.e. mapping a physical address to a logical address: **RARP, BOOTP**, and **DHCP**.
- Lack of flow and error control in the Internet Protocol has resulted in another protocol, **ICMP**. It reports congestion and some types of errors in the network or destination host.

Figure 21.1 Position of ARP and RARP Protocols

2. ADDRESS MAPPING

The internet is made of a combination of physical networks connected by internetworking devices such as routers. A packet starting from a source host may pass through several different physical networks before finally reaching the destination host. The hosts and routers are recognized at the **network level** by their **logical (IP) addresses**, while at the **physical level**, theyare recognized by their **physical (MAC) addresses.**

Thus delivery of a packet to a host or a router requires **two levels** of **addressing**: **logical**

(IP) and **physical (MAC)**.

We need to be able to map a logical address to its corresponding physical address and vice-versa. These can be done by using either static or dynamic mapping.

2.1. Static mapping

Static mapping involves the creation of a table that associates a logical address with a physical address.

Limitations:

- A machine could change its NIC (Network Interface Card), resulting in a new physical address.
- In some LANs, such as LocalTalk, the physical address changes every time the computer is turned on.
- A mobile computer can move from one physical network to another, resulting in a change in its physical address.

To implement these changes, a static mapping table must be updated periodically. This overhead could affect network performance.

2.2. Dynamic mapping

In such mapping each time a machine knows one of the two addresses (logical or physical), it can use a protocol to find the other one.

3. Mapping Logical to Physical Address: ARP

ARP stands for **Address Resolution Protocol** which is one of the most important protocols of the Network layer in the OSI model. ARP finds the physical address, also known as Media Access Control (MAC) address, of a host from its known IP address Figure 21.2.

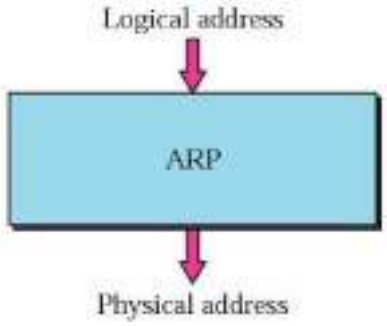


Figure 21.2: ARP Mapping

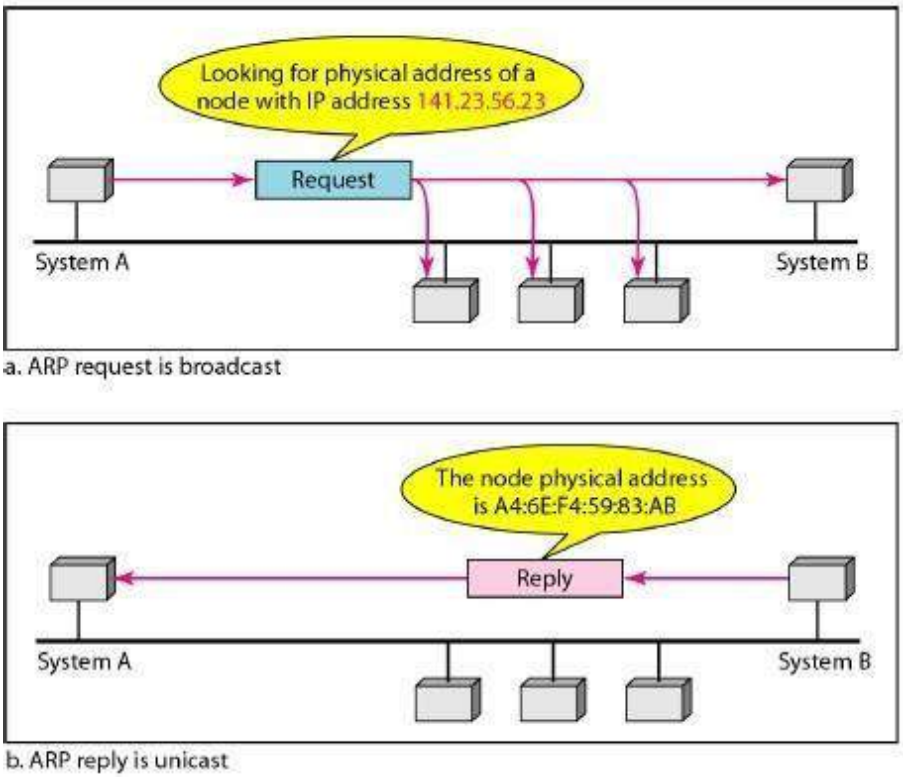


Figure 21.3 ARP operation

Following **steps** are involved in logical to physical address mapping:

a. The host or the router sends an **ARP query packet**. The ARP query packet includes the physical and IP addresses of the sender and the IP address of the receiver. As the sender does not know the physical address of the receiver, the **ARP query is broadcast over the network** (see Figure 21.3).

b. Every host or router on the network receives and processes the ARP query packet, but only the intended recipient recognizes its IP address and sends back an **ARP response packet**.

c. The ARP response packet contains the recipient's IP and physical addresses. The ARP response packet is **unicast directly to the inquirer** (host/router) by using the physical address received in the query packet.

Example: (Figure 21.3)The system on the left (A) has a packet that needsto be delivered to another system (B) with IP address 141.23.56.23.
System A needs to pass the packet to its data link layer for the actual delivery, but it does not know the physical address of the recipient. It uses the services of ARP by asking the ARP protocol to send a broadcast ARP request packet to ask for the physical address of a system with an IP address of 141.23.56.23. This packet is received by every system on the physical network, but only system B will answer it, as shown in Figure 21.3b.
System B sends an ARP reply packet that includes its physical address.

Now system A can send all the packets it has for this destination by using the physical address it received.

3.1. Cache Memory

Using ARP is inefficient if system A needs to broadcast an ARP request for each IP packet it needs to send to system B. ARP can be useful if the ARP reply is cached (kept in cache memory for a while) because a system normally sends several packets to the same destination. A system that receives an ARP reply stores the mapping in the cache memory and keeps it for 20 to 30 minutes unless the space in the cache is exhausted. Before sending an ARP request, the system first checks its cache to see if it can find the mapping.

3.2. ARP Packet Format

Figure 21.4 shows the format of an ARP packet.

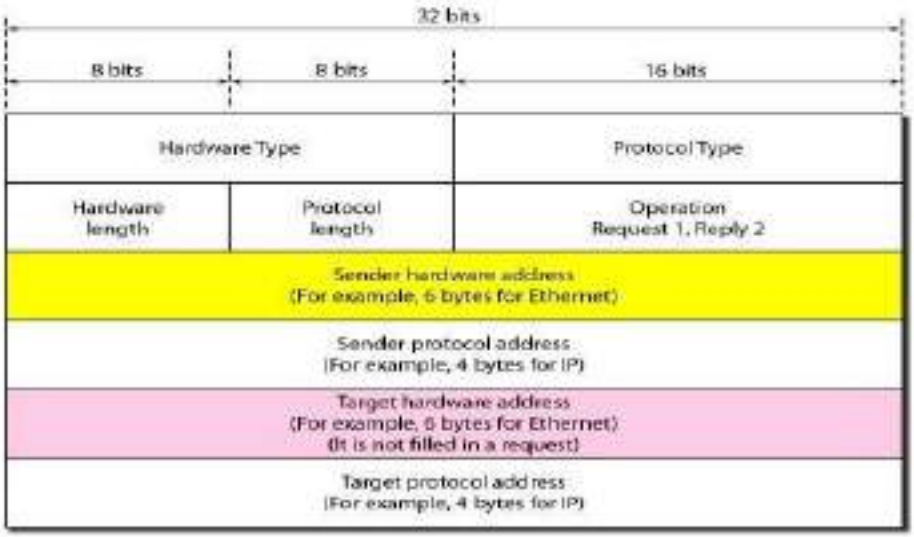


Figure 21.4 ARP packet

The fields are as follows:

a. **Hardware type.** This is a 16-bit field defining the type of the network on which ARP is running. Each LAN has been assigned an integer based on its type. For **example**, Ethernet is given type 1. ARP can be used on any physical network.

b. **Protocol type.** This is a 16-bit field defining the protocol. For example, the value of this field for the IPv4 protocol is 080016, ARP can be used with any higher-level protocol.

c. **Hardware length**. This is an 8-bit field defining the length of the physical address in bytes. For example, for Ethernet the value is 6.

d. **Protocol length**. This is an 8-bit field defining the length of the logical address in bytes. For example, for the IPv4 protocol the value is 4.

e. **Operation**. This is a 16-bit field defining the type of packet. Two packet types are defined: ARP request (1) and ARP reply (2).

f. **Sender hardware address**. This is a variable-length field defining the physical address of the sender. For example, for Ethernet this field is 6 bytes long.

g. **Sender protocol address**. This is a variable-length field defining the logical (for example, IP) address of the sender. For the IP protocol, this field is 4 bytes long.

h. **Target hardware address**. This is a variable-length field defining the physical address of the target. For example, for Ethernet this field is 6 bytes long. For an ARP request message, this field is alI 0s because the sender does not know the physical address of the target.

i. **Target protocol address**. This is a variable-length field defining the logical (for example, IP) address of the target. For the IPv4 protocol, this field is 4 bytes long.

3.3. Encapsulation

An ARP packet is encapsulated directly into a data link frame. For example, in Figure 21.5 an ARP packet is encapsulated in an Ethernet frame. Note that the type field indicates that the data carried by the frame are an ARP packet.

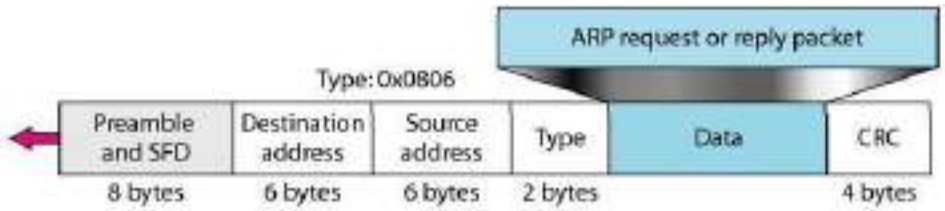


Figure 21.5 Encapsulation of ARP packet

3.4. ARP Operation

Let us see how ARP functions on a typical internet. First we describe the steps involved. Then we discuss the four cases in which a host or router needs to use ARP. These are the steps involved in an ARP process:

1. The sender knows the IP address of the target. We will see how the sender obtains thisshortly.
2. IP asks ARP to create an ARP request message, filling in the sender physical

address, thesender IP address, and the target IP address. The target physical address field is filled with 0s.

3. The message is passed to the data link layer where it is encapsulated in a frame by using the physical address of the sender as the source address and the physical broadcast address as the destination address.

4. Every host or router receives the frame. Because the frame contains a broadcast destination address, all stations remove the message and pass it to ARP. All machines except the one targeted drop the packet. The target machine recognizes its IP address.

5. The target machine replies with an ARP reply message that contains its physical address. The message is unicast.

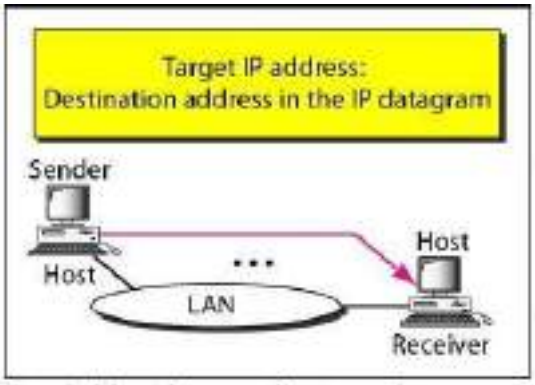
6. The sender receives the reply message. It now knows the physical address of the target machine.

7. The IP datagram, which carries data for the target machine, is now encapsulated in a frameand is unicast to the destination.

3.5. Four Different Cases of ARP Operation

The following are **four different cases** in which the services of ARP can be used (see Figure 21.6(a)to 21.6(d)).

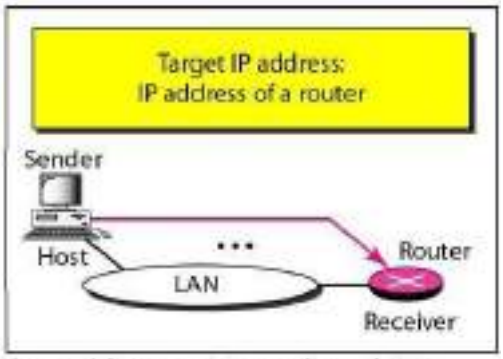
Case 1: *The sender is a host and wants to send a packet to another host on the same network.*In this case, the logical address that must be mapped to a physical address is the destination IP address in the datagram header.



Case 1. A host has a packet to send to another host on the same network.

Figure 21.6(a): Case 1

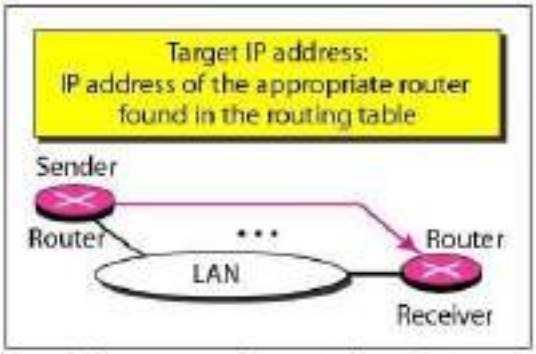
Case 2: *The sender is a host and wants to send a packet to another host on another network.* In this case, the host looks at its routing table and finds the IP address of the next hop (router) for this destination. If it does not have a routing table, it looks for the IP address of the default router. The IP address of the router becomes the logical address that must be mapped to a physical address.



Case 2. A host wants to send a packet to another host on another network. It must first be delivered to a router.

Figure 21.6(b): Case 2

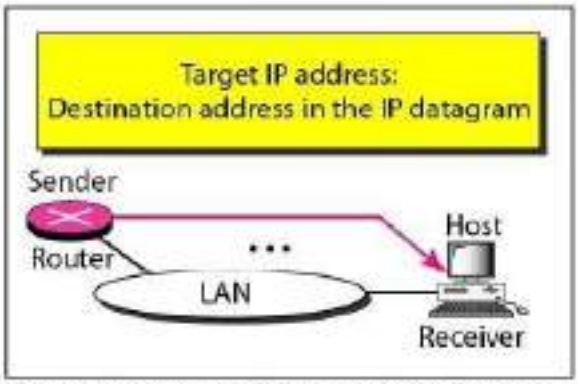
Case 3: *The sender is a router that has received a datagram destined for a host on another network.* It checks its routing table and finds the IP address of the next router. The IP address of the next router becomes the logical address that must be mapped to a physical address.



Case 3. A router receives a packet to be sent to a host on another network. It must first be delivered to the appropriate router.

Figure 21.6(c): Case 3

Case 4: *The sender is a router that has received a datagram destined for a host on the same network.* The destination IP address of the datagram becomes the logical address that must be mapped to a physical address.



Case 4. A router receives a packet to be sent to a host on the same network.

Figure 21.6(d): Case 4

Proxy ARP

A proxy ARP is an ARP that acts on behalf of a set of hosts. Whenever a router running a proxy ARP receives an ARP request looking for the IP address of one of these hosts, the router sends an ARP reply announcing its own hardware (physical) address. After the router receives the actual IP packet, it sends the packet to the appropriate host or router. Let us give an example.In Figure 21.8 the ARP installed on the right-hand host will answer only to an ARP request witha target IP address of 141.23.56.23.

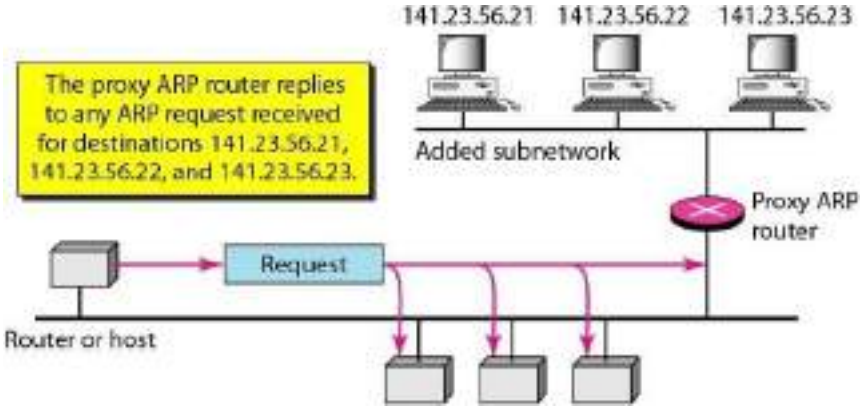


Figure 21.8 Proxy ARP

However, the administrator may need to create a subnet without changing the whole system to recognize subnetted addresses. One solution is to add a router running a proxy ARP. In this case, the router acts on behalf of all the hosts installed on the subnet. When it receives an ARP request with a target IP address that matches the address of one of its host (141.23.56.21, 141.23.56.22, or 141.23.56.23), it sends an ARP reply and announces its hardware address as the target hardware address. When the router receives the IP packet, it sends the packet to the appropriate host.

4. Mapping Physical to Logical Address: RARP, BOOTP, and DHCP

There are occasions in which a **host knows its physical address, but needs to know its logical address** Figure 21.9. This may happen in **two cases**:

Case 1: *A diskless station is just booted*. The station can find its physical address by checking its interface, but it does not know its IP address.

Case 2: *An organization does not have enough IP addresses to assign to each station*; it needs to assign IP addresses on demand. The station can send its physical address and ask for a short time lease.

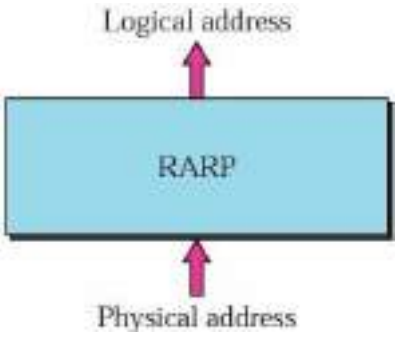


Figure 21.9: RARP Mapping

4.1. RARP

Reverse Address Resolution Protocol (RARP) finds the logical address for a machine that knows only its physical address. RARP Operation

RARP operation is displayed in Figure 21.10

a. A **RARP request** is created and **broadcast** on the local network.

b. Another machine on the local network that knows all the IP addresses will respond with a **RARP reply**.

c. The requesting machine must be running a **RARP client** program; the responding machine must be running a **RARP server** program.

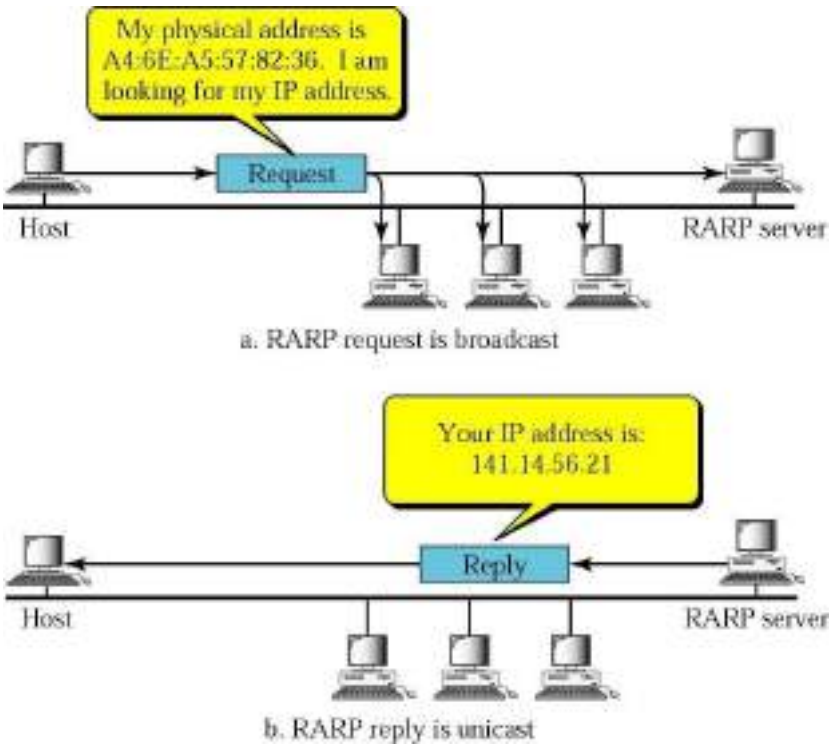


Figure 21.10 RARP Operation

4.1.2. RARP Packet Format & Encapsulation

The format of the RARP packet is the same as the ARP packet format as

displayed in Figure 21.4, except that the Operation field. It's value is 3 for RARP request message and 4 for RARP reply message.
An RARP packet is also encapsulated directly into a data link frame just like ARP packet as displayed in Figure 21.5.

4.1.3. Limitations of RARP:

- As broadcasting is done at the data link layer. The physical broadcast address, all 1's inthe case of Ethernet, does not pass the boundaries of a network.
- This means that if an administrator has several networks or several subnets, it needs toassign a RARP server for each network or subnet.
- This is the reason that **RARP is almost obsolete**.
- Two protocols, BOOTP and DHCP, are replacing RARP.

4.2. BOOTP

The Bootstrap Protocol (BOOTP) is a client/server based protocol at application layer, designed to **provide physical address** to **logical address mapping**. The administrator may put the client and the server on the same network or on different networks, as shown in Figure 21.11a and Figure 21.11b respectively. **BOOTP** messages are **encapsulated** in a **UDP packet**, and the UDP packet itself is encapsulated in an **IP packet**, as shown in Figure 21.12.

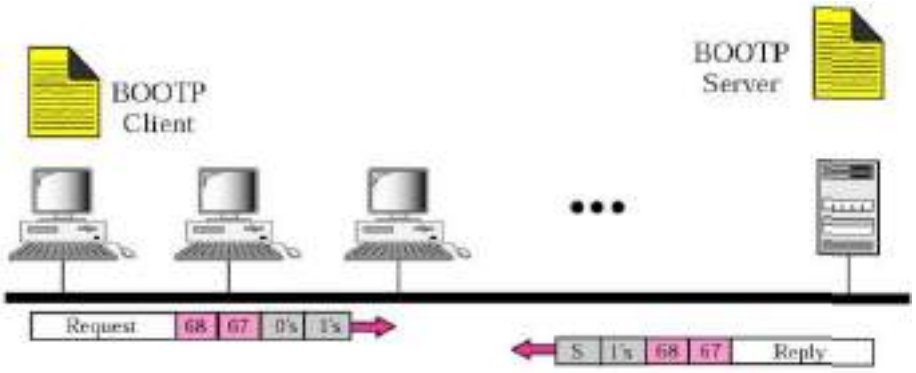


Figure 21.11a BOOTP client and server on the same network

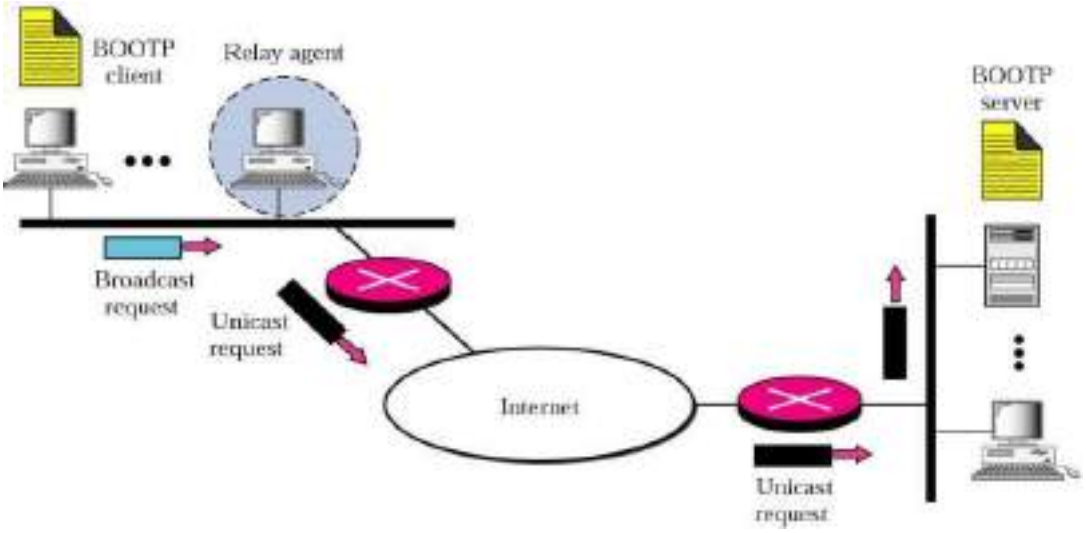


Figure 21.11b BOOTP client and server on the same and different network

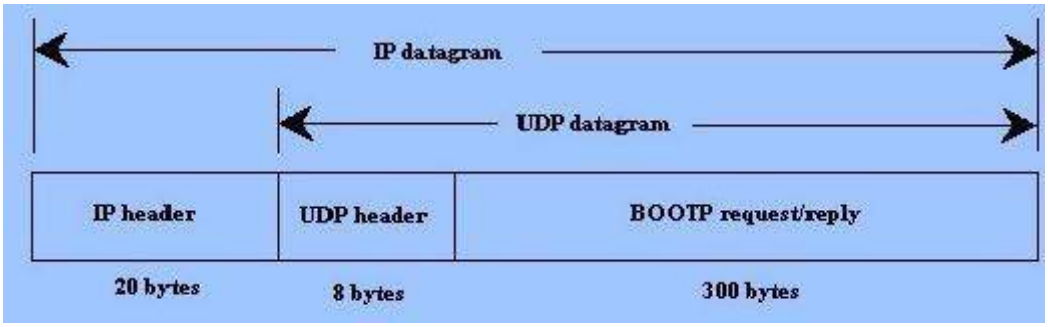


Figure 21.12 BOOTP data Encapsulation

4.2.1. BOOTP Operation

There are two cases of BOOTP operation described below:

Case 1: Client and server on same network (*Figure 21.11a)*

1. When a BOOTP client is started, it has no IP address, so it broadcasts a message containing its MAC address onto the network. This message is called a "BOOTP request," and it is picked up by the BOOTP server, which replies to the client with the following information that the client needs:
    a. The client's IP address, subnet mask, and default gateway address.
    b. The IP address and host name of the BOOTP server.
    c. The IP address of the server that has the boot image, which the client needs to load its operating system.
2. When the client receives this information from the BOOTP server, it configures and initializes its TCP/IP protocol stack, and then connects to the server on which the boot image is shared.

Case 2 : Client and server on different networks(Figure 21.11b**)**

1. If the server exists on some distant network the BOOTP request is broadcast because the client does not know the IP address of the server.
2. The client simply uses all as 0's the source address and all 1's as the destination address.
3. But a broadcast IP datagram cannot pass through any router. To solve the problem, there is a need for an intermediary.
4. One of the hosts in local network (or a router that can be configured to operate at the application layer) can be used as a relay . The host in this case is called a **relay agent**.
5. The relay agent knows the unicast address of a BOOTP server. When it receives this typeof packet, it encapsulates the message in a unicast datagram and sends the request to the BOOTP server.
6. The packet, carrying a unicast destination address, is routed by any router and reaches the BOOTP server.
7. The BOOTP server knows the message comes from a relay agent because one of thefields in the request message defines the IP address of the relay agent.
8. BOOTP server sends a BOOTP reply message to the relay agent.
9. The relay agent, after receiving the reply, sends it to the BOOTP client.

4.2.2. Limitations of BOOTP

- BOOTP is **not a dynamic configuration** protocol.
- BOOTP cannot handle these situations because the binding between the physical and IPaddresses is static and fixed in a table until changed by the administrator.

4.3. DHCP

The **Dynamic Host Configuration Protocol** (DHCP) has been devised to provide static anddynamic address allocation that can be manual or automatic as required.

- **Static Address Allocation** In this capacity DHCP acts as BOOTP does. It is backward compatible with BOOTP, which means a host running the BOOTP client can request a static address from a DHCP server. A DHCP server has a database that statically binds physical addresses to IP addresses.
- **Dynamic Address Allocation** DHCP has a second database with a pool of available IPaddresses. This second database makes DHCP dynamic. When a DHCP client requests a temporary IP address, the DHCP server goes to the pool of available (unused) IP addresses and assigns an IP address for a negotiable period of time.

- When a DHCP client sends a DHCP request to a DHCP server, the server first checks its static database. If an entry with the requested physical address exists in the static database, the permanent IP address of the client is returned.
- On the other hand, if the entry does not exist in the static database, the server selectsan IP address from the available pool, assigns the address to the client, and adds the entry to the dynamic database.

- The dynamic aspect of DHCP is needed when a host moves from network to network oris connected and disconnected from a network (as is a subscriber to a s rvice provider).
- DHCP provides temporary IP addresses for a limited time. The addresses assigned from the pool are temporary addresses.
- The DHCP server issues a lease for a specific time. When the lease expires, the clientmust either stop using the IP address or renew the lease.
- The server has the option to agree or disagree with the renewal. If the server disagrees,the client stops using the address.

4.3.1. DHCP Operation

DHCP provides an automated way to distribute and update IP addresses and other configuration information on a network. A DHCP server provides this information to a DHCP client through the exchange of a series of messages, known as the DHCP conversation or the DHCP transaction displayed in Figure 21.13.

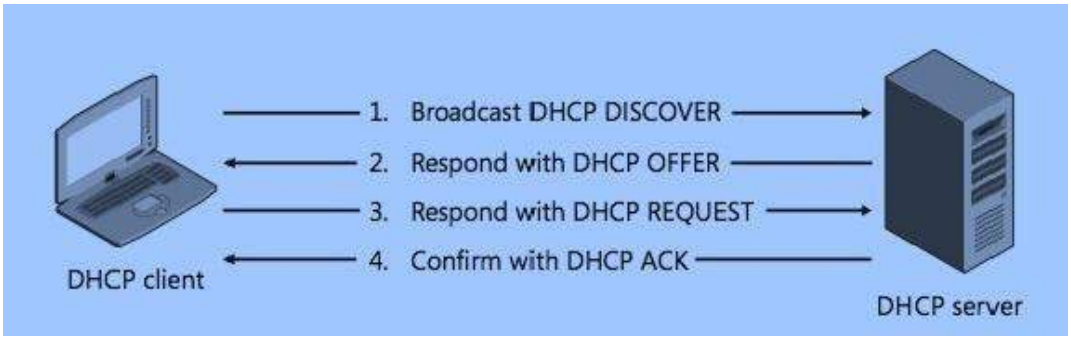


Figure 21.13: DHCP

c

Operation DHCP client goes through the **f ur step**

process:

1. A DHCP client sends a broadcast packet (**DHCP Discover**) to discover DHCP servers on the LAN segment.

2. The DHCP servers receive the **DHCP Discover** packet and respond with **DHCP Offer** packets, offering IP addressing information.

3. If the client receives the **DHCP Offer** packets from multiple DHCP servers, the first **DHCP Offer** packet is accepted. The client responds by broadcasting a **DHCP Request** packet, requesting network parameters from a single server.

4. The DHCP server approves the lease with a **DHCP Acknowledgement (DHCP Ack)** packet. The packet includes the lease duration and other configuration information.

5. ICMP

IP provides unreliable and connectionless datagram delivery. It was designed this way to make efficient use of network resources. The IP protocol is a best-effort delivery service that delivers a datagram from its original source to its final destination. However, **IP protocol has two deficiencies:** lack of error control and lack of assistance mechanisms.

- The IP protocol has no error-reporting or error-correcting mechanism.
- What happens if something goes wrong?
- What happens if a router must discard a datagram because it cannot find a router to the final destination, or because the time-to-live field has a zero value?
- What happens if the final destination host must discard all fragments of a datagram because it has not received all fragments within a predetermined time limit?

These are examples of situations where an error has occurred and the IP protocol has no built-in mechanism to notify the original host.

- The IP protocol also lacks a mechanism for host and management queries.
- A host sometimes needs to determine if a router or another host is alive.
- And sometimes a network administrator needs information from another host or router.

The Internet Control Message Protocol (ICMP) has been designed to compensate for the above two deficiencies. It is a companion to the IP protocol.

5.1. Types of Messages

ICMP messages are divided into two broad categories: **Error-reporting messages and Query messages**

The **error-reporting messages** report problems that a router or a host (destination) may encounter when it processes an IP packet.

The **query messages**, which occur in pairs, help a host or a network manager get specific information from a router or another host.

5.2. Message Format

An ICMP message has an **8-byte header** and a **variable-size data section**. Although the general format of the header is different for each message type, the first 4 bytes are common to all. As Figure 21.14 shows:

The first field, **ICMP type**, defines the type of the message.

The **code field** specifies the reason for the particular message type.

The last common field is the **checksum field** used for securing ICMP

header. The rest of the header is specific for each message type.

The **data section** in error messages carries information for finding the original packet that had the error.

In ICMP query messages, the data section carries extra information based on the type of the query.

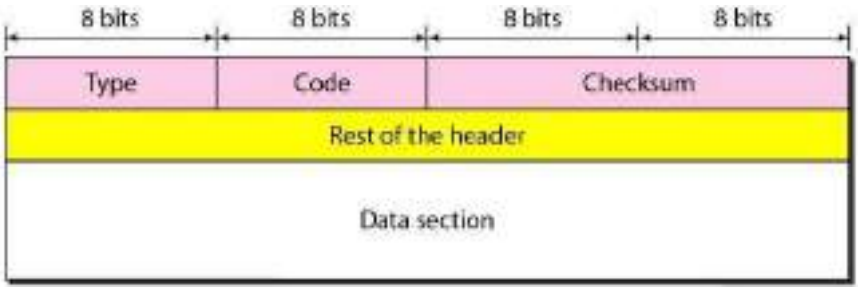


Figure 21.14 General format of ICMP messages

5.3. ICMP Encapsulation:

ICMP itself is a network layer protocol. However it's messages are not passed directly to datalink layer. Instead the messages are first encapsulated inside IP datagrams before going to the lower layer (see Figure 21.15).

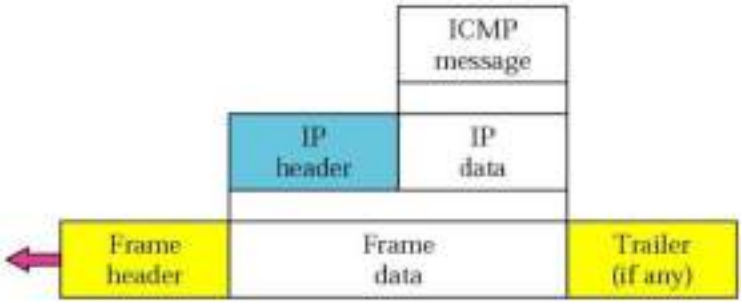


Figure 21.15 Contents of data field for the error messages

5.4. Error Reporting Messages

One of the main responsibilities of ICMP is to report errors.

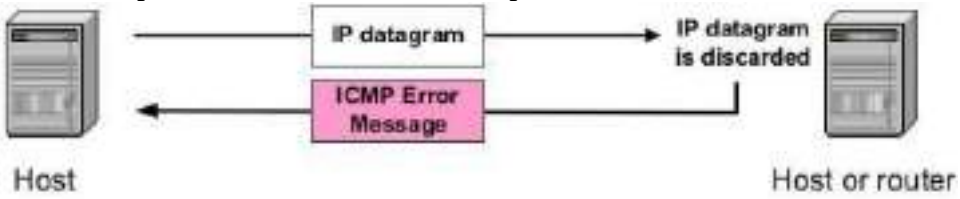


Figure 21.16 ICMP Error Reporting Message

Error messages are typically sent when a datagram is discarded due to some error asdisplayed in Figure 12.16.

Error messages are always sent to the original source because the only information available in the datagram about the route is the source and destination IP addresses.

Five types of errors are handled: *destination unreachable, source quench, timeexceeded,parameter problems, and redirection* (see Figure 21.17).

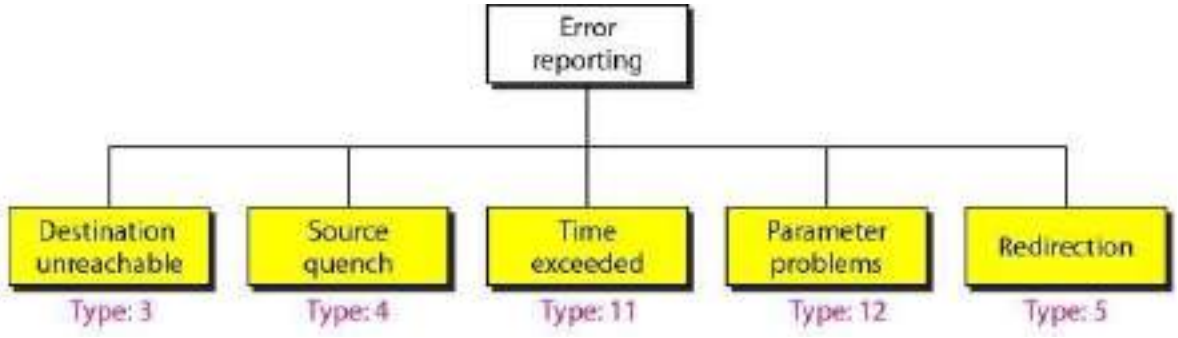


Figure 21.17 Error-reporting messages

a. Destination Unreachable

When a router cannot route a datagram or a host cannot deliver a datagram, the datagram is discarded and the router or the host sends a destination-unreachable message back to the source host that initiated the datagram.

b. Source Quench

- **The source-quench message in ICMP was designed to add a kind of flow control to the IP.**
- When a router or host discards a datagram due to congestion**, it sends a source- quench message to the sender of the datagram**. This message has **two purposes**.
- **First**, it informs the source that the datagram has been discarded.
- **Second**, it warns the source that there is congestion somewhere in the path and that the source should slow down (quench) the sending process.

c. Time Exceeded

The time-exceeded message is generated in **two cases**:

Case1: As routers use routing tables to find the next hop (next router) that must receive the packet. If there are errors in one or more routing tables, a packet can travel in a loopor a cycle, going from one router to the next or visiting a series of routers endlessly. Each datagram contains a field called *time to live* that controls this situation. When a datagram visits a router, the value of this field is decremented by 1. When the time-to-live value reaches 0, after decrementing, the router discards the datagram. However, when the datagram is discarded, a time-exceeded message must be sent by the router to the original source.

Case2: A time-exceeded message is also generated when not all fragments that make upa message arrive at the destination host within a certain time limit.

d. Parameter Problem

Any **ambiguity in the header** part of a datagram can create serious problems as the datagram travels through the Internet. If a router or the destination host discovers an ambiguous or missing value in any field of the datagram, it discards the datagram and sends a parameter-problem message back to the source.

e. Redirection

- This concept of redirection is shown in Figure 21.18. Host A wants to send a datagram to host B.

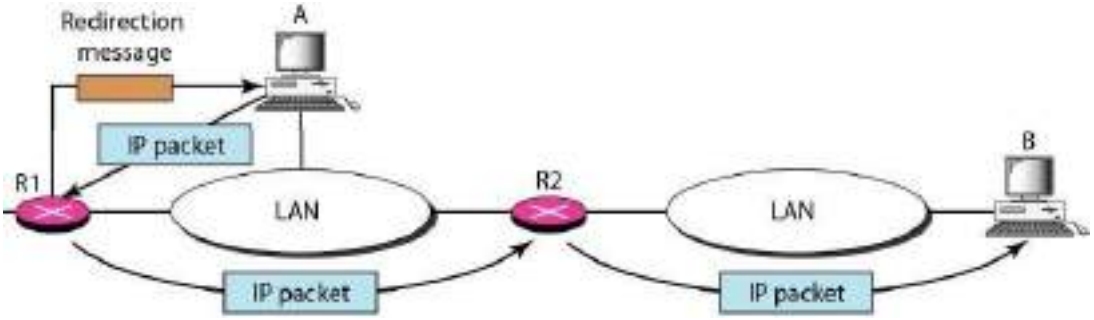


Figure 21.18 Redirection concept

- Router R2 is obviously the most efficient routing choice, but host A did notchoose router R2. The datagram goes to R1 instead.
- Router R1, after consulting its table, finds that the packet should have gone toR2.
- It sends the packet to R2 and, at the same time, sends a redirection message tohost A.
- Host A's routing table can now be updated.

5.5. ICMP Query Messages

In addition to error reporting, ICMP can diagnose some network problems. This is accomplished through the query messages, a group of **four different pairs of messages**, as shown in Figure 21.19.

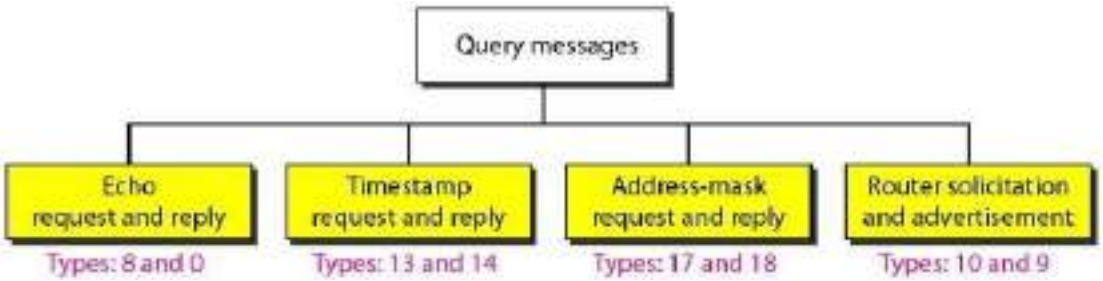


Figure 21.19 Query messages

In this type of ICMP message, a node sends a ICMP request message that is answered ina specific format as ICMP reply by the destination node, depicted in Figure 21.20.

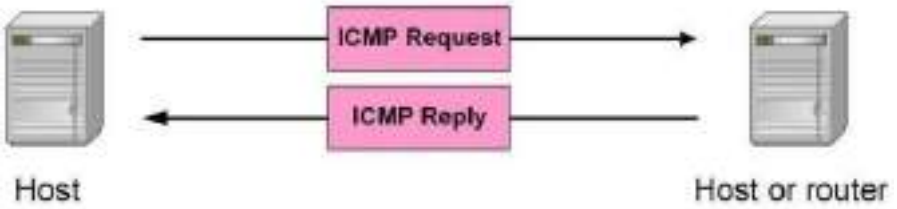


Figure 21.20 ICMP Query Message

A query message is encapsulated in an IP packet, which in turn is encapsulated in a data link layer frame.

However, in this case, no bytes of the original IP are included in the message, as shownin Figure 21.21.

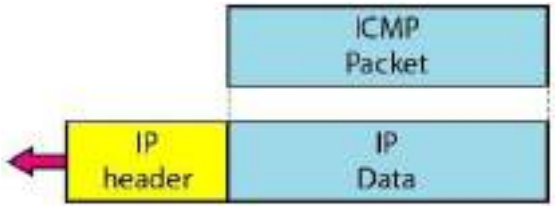


Figure 21.21 Encapsulation of ICMP query messages

a. Echo Request and Echo Reply

The echo-request and echo-reply messages are designed for diagnostic purposes. The combination of echo-request and echo-reply messages determines whether two systems (hosts or routers) can communicate with each other Figure 21.22. It also confirms that the intermediate routers are receiving, processing, and forwarding IP datagrams.

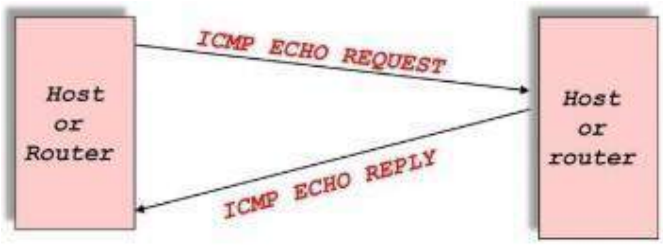


Figure 21.22 ICMP Echo Request and Echo Reply

Today, most systems provide a version of the ***ping* command** that can create a series (instead of just one) of echo-request and echo-reply messages, providing statistical information. We can use the *ping* program to find if a host is alive and responding.

b. Timestamp Request and Timestamp Reply

Two machines (hosts or routers) can use the timestamp request and timestamp reply messages to determine the round-trip time needed for an IP datagram to travel between them. It can also be used to synchronize the clocks in two machines.

c. **Address-Mask Request and Address-Mask Reply**

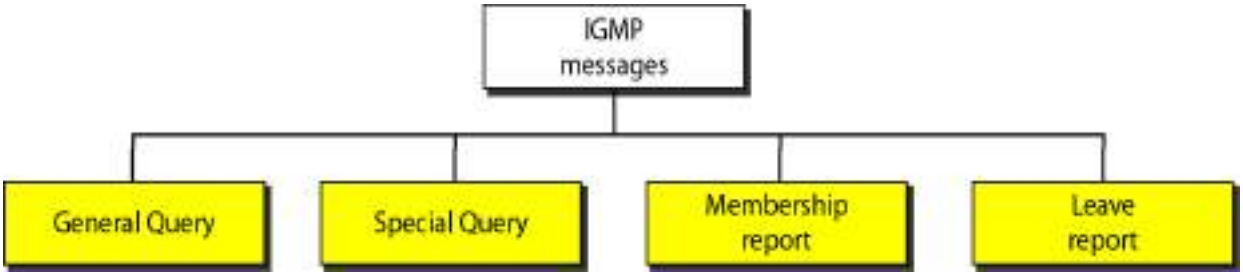
A host may know its IP address, but it may not know the corresponding mask. For example, a host may know its IP address as 159.31.17.24, but it may not know that the corresponding mask is /24. To obtain its mask, a host sends an address-mask-request message to a router on the LAN. If the host knows the address of the router, it sends the request directly to the router. If it does not know, it broadcasts the message. The router receiving the address-mask-request message responds with an address-mask-reply message, providing the necessary mask for the host. This can be applied to its full IP address to get its subnet address.

d. **Router Solicitation and Router Advertisement**

The router-solicitation and router-advertisement messages can help a host to check whether the neighboring routers are alive and functioning. A host can broadcast (or multicast) a router-solicitation message. The router or routers that receive the solicitation message broadcast their routing information using the router-advertisement message. A router can also periodically send router-advertisement messages even if no host has solicited.

IGMP:

The IP protocol can be involved in two types of communication: unicasting and multicasting. The Internet Group Management Protocol (IGMP) is one of the necessary, but not sufficient, protocols that is involved in multicasting.. IGMP is a companion to the IP protocol.



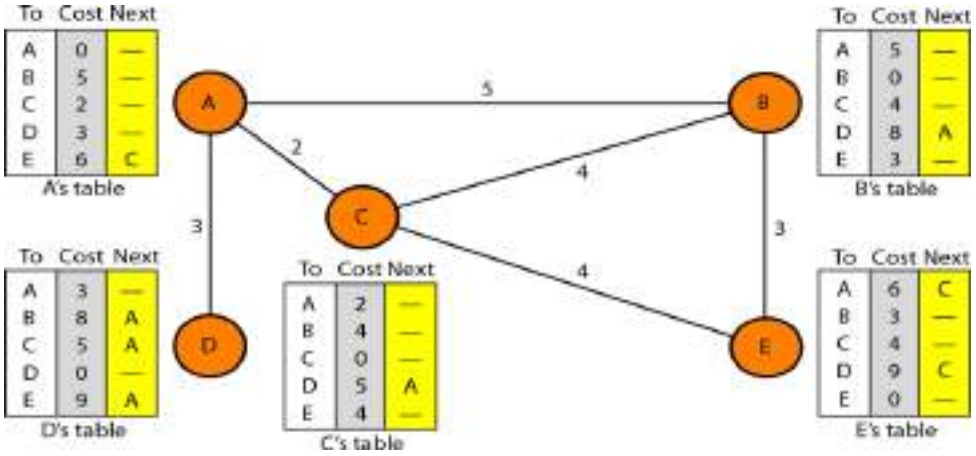
ROUTING:

DISTANCE VECTOR ROUTING

In distance vector routing, the least-cost route between any two nodes is the route with minimum distance. The term **Vector** means a **Table**.

- In this protocol each node maintains a table of minimum distances to every node.

- The table at each node also guides the packets to the desired node by showing the next hop in the route.



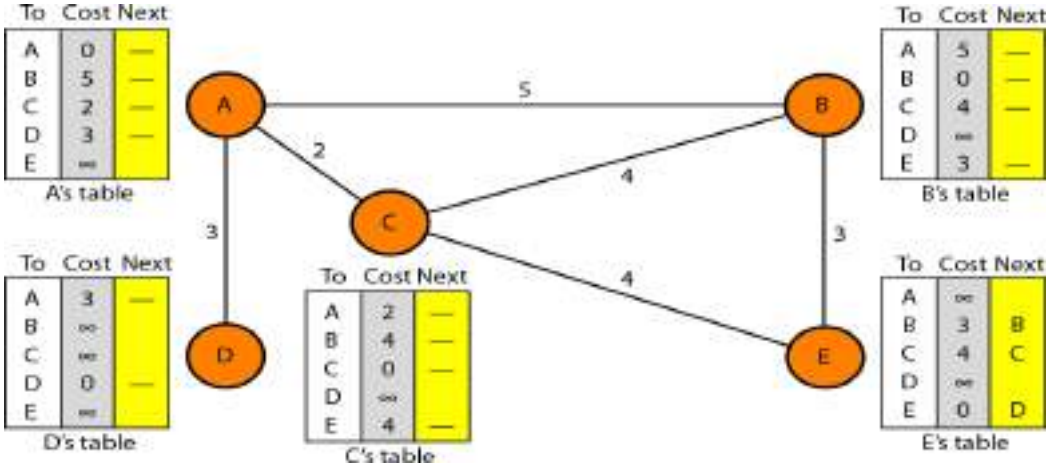
The table for node A shows how we can reach to any node from node A.

Ex: From node A the least cost to reach node E is 6. The route passes through C.

Initialization

- The above table is the final step of Distance vector routing each node knows about each and every node in the network.
- But at the initial stage each node can only know the distance between itself and its immediate neighbors. Neighbors are the nodes which are directly connected to the node.

The below figure shows the initialization of the table:



- Each node will take the immediate neighbor distance into the table.
- The distance for any entry that is not a neighbor is marked as infinite.
- Infinite means Unreachable.

Sharing

The idea behind Distance Vector Routing is the **sharing** of information between **Neighbors**.

By observing the above figure:

1. Node A does not know about node E but node C knows how to reach node E.

If Node C shares its routing table with node A then node A can also know how to reach E.

2. Node C does not know how to reach Node D but Node A knows how to reach Node D.

If Node A shares its routing table with node C then node C can also know how to reach node A.

(i.e.) Immediate Neighbor's nodes A and C can improve their routing tables if they share each other's routing tables.

Problem: How many columns of the table must be shared with each neighbor?

- A node is not aware of a neighbor's table.
- A node can send only the first two columns of its table to any neighbor.
- Because the third column of a table next hop is not useful for the neighbor.
- When the neighbor receives a table, third column needs to be replaced with the sender's name.

Updating

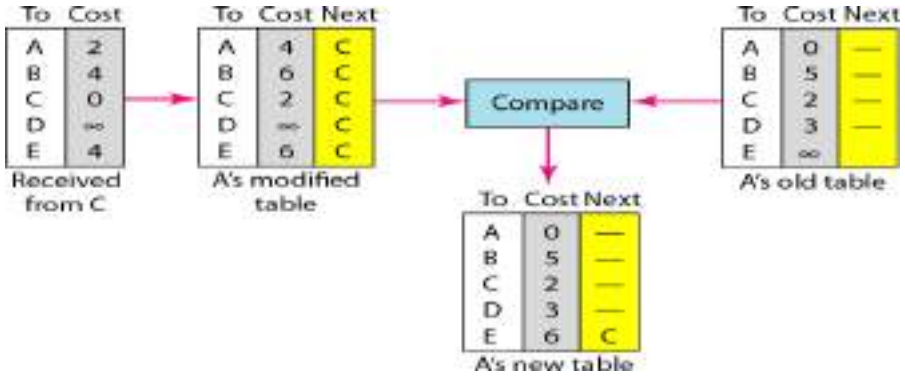
When a node receives a **Two-Column Table** from a neighbor, it needs to update its routing table. Updating takes three steps:

1. The receiving node needs to add the cost between itself and the sending node to each value in the second column.

Example:

- If node C claims that its distance to a destination is **x** km
- The distance between A and C is **y** km then,
- The distance between A and that destination via C is **(x + y)** km.

2. The receiving node needs to add the name of the sending node to each row as the third column if the receiving node uses information from any row. The sending node is the next node in the route.
3. The receiving node needs to compare each row of its old table with the corresponding row of the modified version of the received table.
   a. If the next-node entry is different, the receiving node chooses the row with the smaller cost. If there is a tie, the old one is kept.
   b. If the next-node entry is the same, the receiving node chooses the new row. For example, suppose node C has previously advertised a route to node X with distance 3. Suppose that now there is no path between C and X; node C now advertises this route with a distance of infinity. Node A must not ignore this value even though its old entry is smaller. The old route does not exist anymore. The new route has a distance of infinity.



Note:

1. The modified table shows how to reach A from A via C. If A needs to reach itself via C, it needs to go to C and come back, so the distance will be 4 (A□C=2 and C□A=2).

2. The only benefit from this updating of node A is that A now knows how to reach E with cost=6 via C.

Each node can update its table by using the tables received from other nodes. If there is no change in the network itself, such as a failure in a link, each node reaches a stable condition in which the contents of its table remain the same.

When to Share

When does a node send its partial routing table (only two columns) to all its immediate neighbors?

1. **Periodic Update**: A node sends its routing table, normally every 30 s, in a periodic update. The period depends on the protocol that is using distance vector routing.
2. **Triggered Update** A node sends its two-column routing table to its neighbors anytime there is a change in its routing table.

The change can result from the following:

- A node receives a table from a neighbor, resulting in changes in its own table after updating.
- A node detects some failure in the neighboring links which results in a distance change to infinity.

Count to Infinity Problems

Two-Node Loop Instability

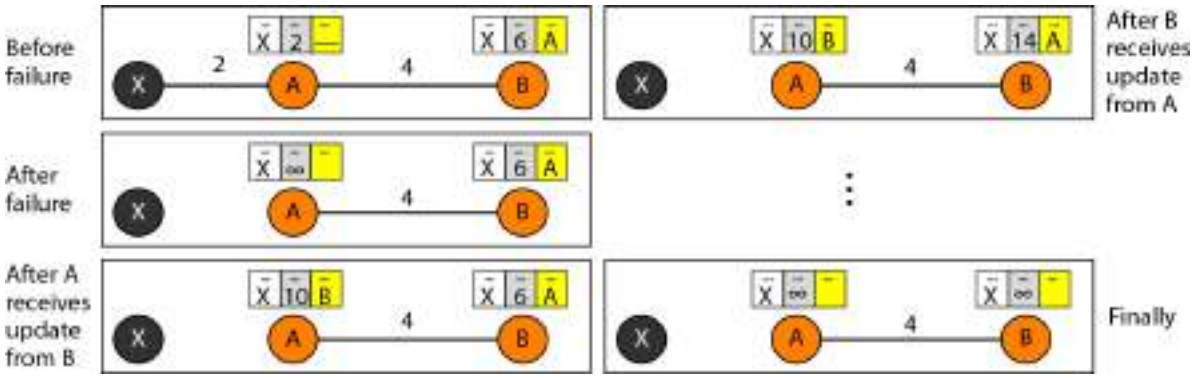
A problem with distance vector routing is instability, which means that a network using this protocol can become unstable.

Consider the above figure:

It shows a system with 3 nodes: X, A and B.

- At the beginning, both nodes A and B know how to reach node X.
- But suddenly, the link between A and X fails.
- Node A changes its table.

If A can send its table to B immediately there will be no problem because B can identify by looking at the table value ∞**.**



Problem is: The system becomes unstable if B sends its routing table to A before receiving A's routing table.

- Node A receives the update and, assuming that B has found a way to reach X and immediately updates its routing table.
- Based on the triggered update strategy, A sends its new update to B.
- Now B thinks that something has been changed around A and updates its routing table.
- The cost of reaching X increases gradually until it reaches infinity.
- At this moment, both A and B know that X cannot be reached.

During this counting to infinity the system is not stable:

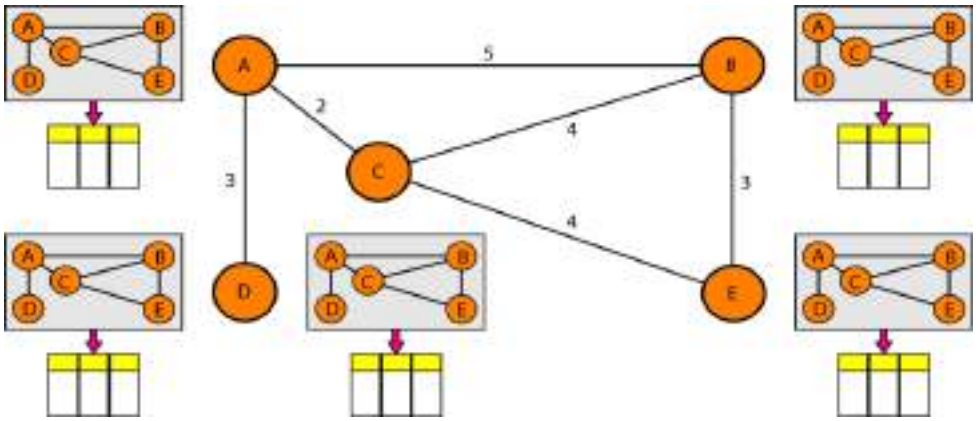
- Node A thinks that the route to X is via B.
- Node B thinks that the route to X is via A.
- If A receives a packet destined for X, it goes to B and then comes back to A.
- If B receives a packet destined for X, it goes to A and comes back to B.
- Packets bounce between A and B, creating a two-node loop problem.

Link State Routing (LSR):

In Link State Routing, Each node in the domain has the entire topology of the domain –

- List of nodes and links
- How they are connected including the type
- Cost (metric)
- Condition of the links (up or down)

The node can use Dijkstra's algorithm to build a routing table.



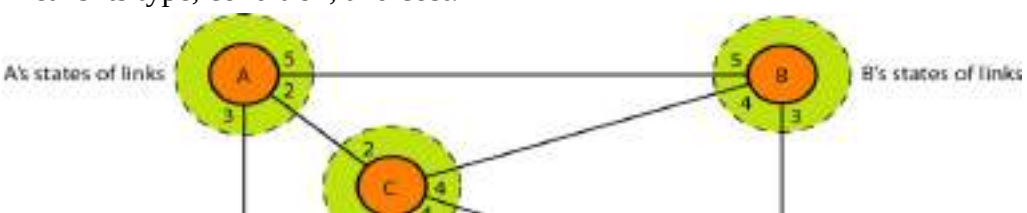
The above figure shows a Simple domain with Five Nodes.

- Each node uses the same topology to create a routing table, but the routing table for each node is unique because the calculations are based on different interpretations of the topology.
- The topology must be dynamic, representing the latest state of each node and each link.
- If there are changes in any point in the network the topology must be updated for each node.

For Example: a link is down then each and every node in the domain should update this change.

How can a common topology be dynamic and stored in each node?

In LSR each node in the domain has the partial knowledge about the state of its links. The state means its type, condition, and cost.



Consider the above figure that shows List of nodes and their partial knowledge.

Node A knows that it is connected

- To Node B with metric5
- To Node C with metric2
- To Node D with metric3

Node C knows that it is connected

- To Node A with metric2
- To Node B with metric4
- To Node E with metric4

Although there is an overlap in the knowledge, the overlap guarantees the creation of a common topology-a picture of the whole domain for each node.

Building Routing Tables

In link state routing, four sets of actions are required to ensure that each node has the routing table showing the least-cost node to every other node.

1. Creation of the states of the links by each node, called the Link State Packet(LSP).
2. Dissemination (Distribution) of LSPs to every other router called **Flooding**. The flooding can be done in an efficient and reliable way.
3. Formation of a shortest path tree for each node.
4. Calculation of a routing table based on the shortest path tree.

Creation of Link State Packet (LSP)

A link state packet can carry a large amount of information such as the node identity, the list of links, a sequence number, and age etc.

- Node identity and the List of links are needed to make the topology.
- Sequence number facilitates flooding and distinguishes new LSPs from old ones.
- Age prevents old LSP's from remaining LSP's in the domain for a longtime.

LSPs are generated on two occasions:

1. **When there is a change in the topology of the domain:**

Triggering of LSP dissemination is the main way of quickly informing any node in the domain to update its topology.

2. **On a periodic basis:**
    - It is done to ensure that old information is removed from the domain.
    - The timer set for periodic dissemination is normally in the range of 60 min or 2 hours based on the implementation.

- A longer period ensures that flooding does not create too much traffic on the network.

Note: As a matter of fact, there is no actual need for this type of LSP dissemination.

Flooding of LSP's

After a node has prepared an LSP, it must be disseminated to all other nodes, not only to its neighbors. The process is called flooding.

Flooding will be done based on the following:

1. The creating node sends a copy of the LSP out of each interface.
2. A node that receives an LSP compares it with the copy it may already have.

Each and every LSP will be given a Sequence number at the time of their creation. Comparison of sequence numbers determines which LSP is older and which LSP is latest. If the newly arrived LSP is older than the one it already has, then the node discards the LSP.

If LSP arrived is newer, the node does the following:

- It discards the old LSP and keeps the new one.
- It sends a copy of it out of each interface except the one from which the packet arrived. This guarantees that flooding stops somewhere in the domain where a node has only one interface.

Formation of Shortest Path Tree: Dijkstra Algorithm

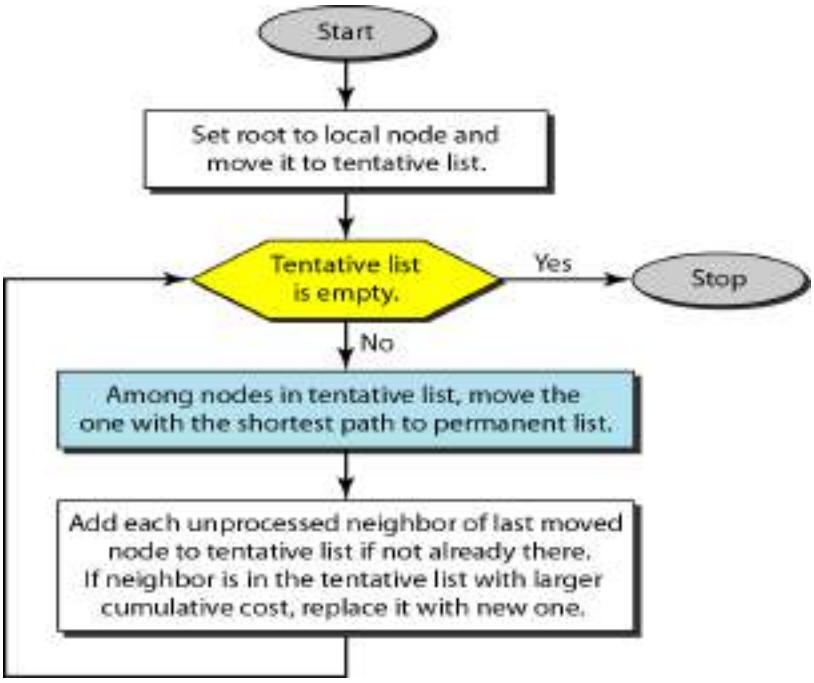
- After receiving all LSPs, each node will have a copy of the whole topology.
- The topology is not sufficient to find the shortest path to every other node; a shortest path tree is needed.
- A tree is a graph of nodes and links, where one node is called Root.
- A shortest path tree is a tree in which the path between the root and every other node is the shortest.
- The Dijkstra's algorithm creates a shortest path tree from a graph.

The algorithm divides the nodes into two sets:

1. Tentative nodes
2. Permanent nodes

Dijkstra's algorithm finds the neighbors of a current node, makes them tentative, examines them, and if they pass the criteria, makes them permanent.

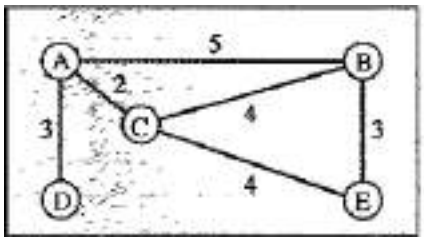
Flow Chart of Dijksta's Algorithm



Example:

Consider the below graph with five nodes: A,B,C,D,E.

- Apply the Dijkstra's algorithm to node A.
- To find the shortest path in each step, we need the cumulative cost from the root to each node, which is shown next to the node.



Topology

At the end of each step, we show the permanent (filled circles) and the tentative (open circles) nodes and lists with the cumulative costs.

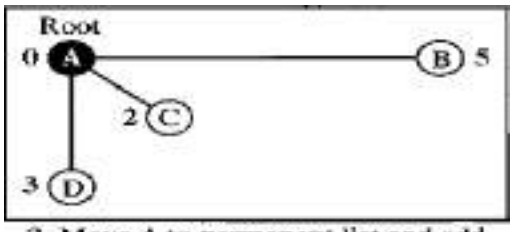
Step 1: We make node A the root of the tree and move it to the tentative list. Our two lists are

Permanent list: **Empty** Tentative list: **A(0)**

1. Set root to A and move A to tentative list.

Step 2: Node A has the shortest cumulative cost from all nodes in the tentative list.
We move A to the permanent list and add all neighbors of A to the tentative list. Our new lists are

Permanent list: A(0) Tentative list: B(5), C(2), D(3)

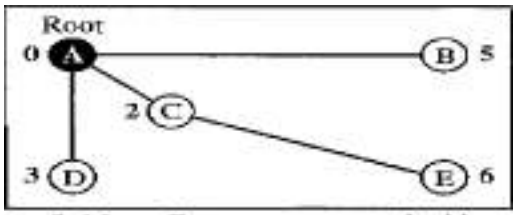


2. Move A to permanent list and add B, C, and D to tentative list.

Step 3: Node C has the shortest cumulative cost from all nodes in the tentative list.

- We move C to the permanent list.
- Node C has three neighbors, but node A is already processed, which makes the unprocessed neighbors just B and E.
- However, B is already in the tentative list with a cumulative cost of 5.
- Node A could also reach node B through C with a cumulative cost of 6.
- Since 5 is less than 6, we keep node B with a cumulative cost of 5 in the tentative list and do not replace it.

Our new lists are: Permanent list: A(0),C(2) Tentative list: B(5), D(3),E(6).

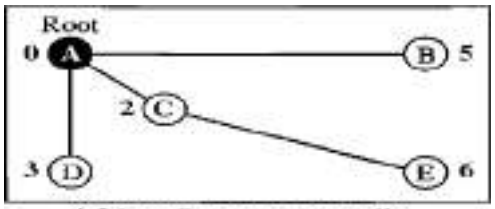


3. Move C to permanent and add E to tentative list.

Step 4: Node D has the shortest cumulative cost of all the nodes in the tentative list.

- We move D to the permanent list. Node D has no unprocessed neighbor to be added to the tentative list.

Our new lists are: Permanent List: A(0),C(2),D(3) Tentative List: B(5),E(6).

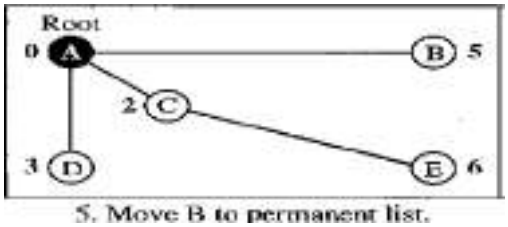


4. Move D to permanent list.

Step 5: Node B has the shortest cumulative cost of all the nodes in the tentative list.

- We move B to the permanent list. We need to add all unprocessed neighbors of B to the tentative list (i.e. just node E).
- E(6) is already in the list with a smaller cumulative cost.
- The cumulative cost to node E, as the neighbor of B, is 8. We keep node E(6) in the tentative list.

Our new lists are: Permanent list: A(0), B(5),C(2),D(3) Tentative list: E(6)

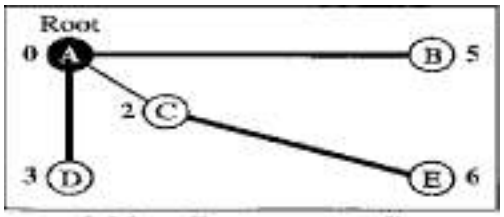


5. Move B to permanent list.

Step 6: Node E has the shortest cumulative cost from all nodes in the tentative list.

- Move E to the permanent list. Node E has no neighbor. Now the tentative list is empty.
- We stop the process here. The shortest path tree is ready for graph ABCDE.

The finalists are: Permanent list: A(0), B(5), C(2), D(3), E(6) Tentative list: Empty



6. Move E to permanent list (tentative list is empty).

Calculation of Routing Table from Shortest Path Tree

- Each node uses the shortest path tree protocol to construct its routing table.
- The routing table shows the cost of reaching each node from the root.

The below table shows routing table for Node A.

Node	Cost	Next Router
A	0	-
B	5	-
C	2	-
D	3	-
E	6	C

IPv4 Delivery Mechanism

- IPv4 delivery mechanism is used in TCP/IP protocols.
- IPv4 is an unreliable and connectionless datagram protocol.
- If reliability is important, IPv4 must be paired with a reliable protocol such as TCP.

Datagram

Packets in the IPv4 layer are called datagrams.

A datagram is a variable-length packet consisting of two parts:

1. Header
2. Data

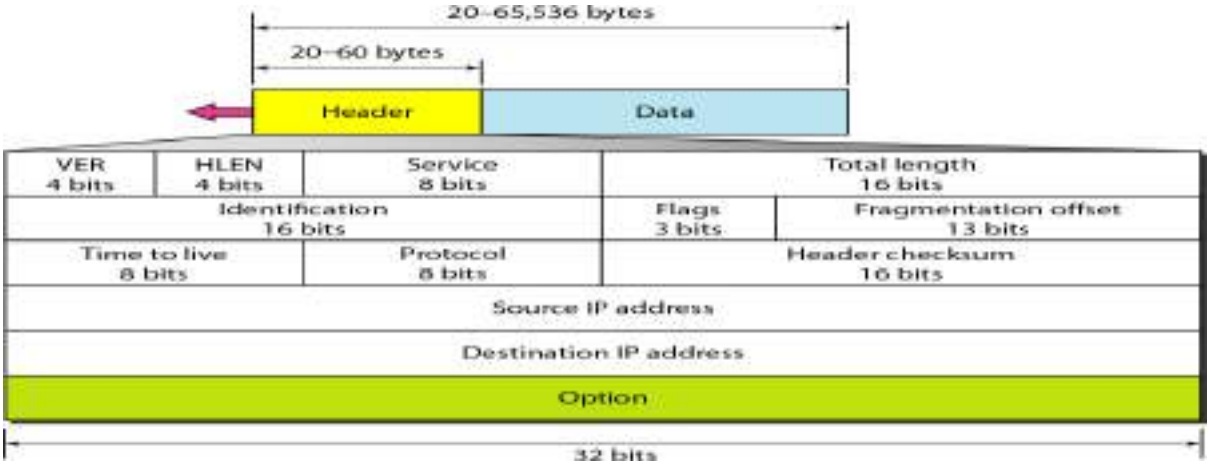
The header is 20 to 60 bytes in length and contains information essential to routing and delivery.

1. Version (VER) – 4bits

- It defines the version of the IPv4 protocol. Currently 4th version of IPv4 is using.

2. Header length (HLEN) – 4bits

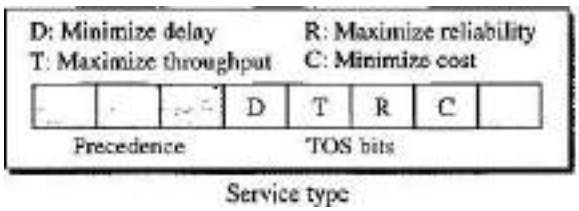
- It defines the total length of the datagram header in 4-bytewords.
- The length of the header is variable between 20 and 60bytes.
- When there are no options, the header length is 20 bytes, and the value of this field is 5 (5x 4 =20).
- When the option field is at its maximum size, the value of this field is 15 (15 x 4 = 60).



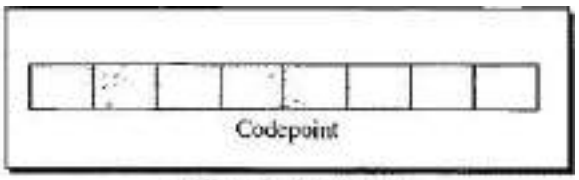
3. Services

IETF has changed the interpretation and name of this 8-bit field.

Previously this field is called **Service Type** but now the name changed to **Differentiated Services**.



Service type



Differentiated services

Service Type

In this interpretation, the first 3 bits are called precedence bits. The next 4 bits are called type of service (TOS) bits, and the last bit is not used.

i. Precedence

- It is a 3-bit subfield ranging from 0 to 7 (000 to111).
- The precedence defines the priority of the datagram in issues such as congestion.
- If a router is congested and needs to discard some datagrams, those datagrams with lowest precedence are discarded first.

ii. Type of Service(TOS)

It is a 4-bit subfield with each bit having a special meaning. Out of 4 bits only one bit will have the value of 1.

TOS Bits	Description
0000	Normal (default)
0001	Minimize Cost
0010	Maximize Reliability
0100	Maximize Throughput
1000	Minimize Delay

Differentiated Services

In this interpretation, the first 6 bits make up the code-point subfield, and the last 2 bits are not used. The code-point subfield can be used in two different ways:

a. When the 3 rightmost bits are 0's, the 3 leftmost bits are interpreted the same as the precedence bits in the service type interpretation.

b. When the 3 rightmost bits are not all Os, the 6 bits define 64 services based on the priority assignment by the Internet or local authorities.

Category	**Code-point**	**Assigning Authority**	**No of service types**	**Numbers**
1	XXXXX0	Internet	32	0,2,4,6,8,……60,62
2	XXXX11	Local	16	3,7,11,15,……59,63
3	XXXX01	Temporary or Experiment	16	1,5,9,13,17...……,61

4. Total length

This is a 16-bit field that defines the total length (**header plus data**) of the IPv4 datagram in bytes. Total length of IPv4 is 65,535 (2^{16}-1).

Length of data = Total length - header length

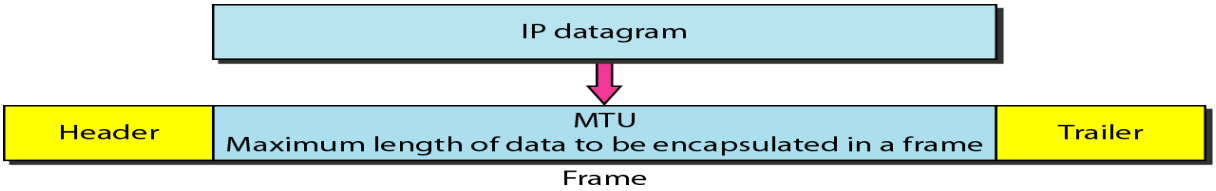
FRAGMENTATION

- A datagram is fragmented if it is too large for a network to carry it.

Maximum Transfer Unit (MTU)

- the Total size of the datagram must be less than this maximum size.

- The maximum length of IPV4 datagram is 65,535bytes.



- If the length of the datagram exceeds the MTU then the datagram must be fragmented to make it possible to pass through the networks.
- When a datagram is fragmented, each fragment has its own header with only few fields are changed. Remaining fields are copied by all fragments.

Fields Related to Fragmentation: Identification, Flag, Offset.

Identification (16 bits)

- When a datagram is fragmented, all fragments have the same identification number the same as the original datagram. All fragments having the same identification value must be assembled into one datagram.
- The identification number helps the destination in reassembling the datagram.

Flags (3 bits)

The first bit is **Reserved**.

The second bit is called the **Do Not Fragment** bit.

- If its value is 1, the machine must not fragment the datagram.
- If its value is 0, the datagram can be fragmented if necessary.

The third bit is called the **More Fragment** bit.

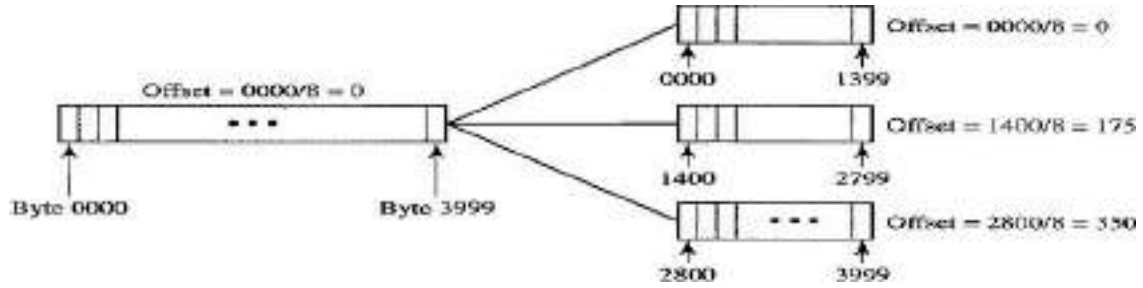
- If its value is 1, it means the datagram is not the last fragment.
- If its value is 0, it means this is the last or only fragment.

Fragmentation offset (13 bits)

It shows the relative position of this fragment with respect to the whole datagram.

It is the offset of the data in the original datagram measured in units of 8 bytes.

Example: Consider the below figure shows a datagram with a data size of 4000 bytes fragmented into three fragments.



The bytes in the original datagram are numbered 0 to 3999.

Fragment Number	Range	Offset Value
First	0-1399	0/8=0
Second	1400-2799	1400/8=175
Third	2800-3999	2800/8=350

Time To Live - TTL (8 bits)

A datagram has a limited lifetime in its travel through an internet.

This field can be used in two ways:

i. This field was originally designed to hold a timestamp, which was decremented by each visited router. The datagram was discarded when the value became zero.

ii. This field is used mostly to control the maximum number of hops (routers) visited by the datagram. Each router that processes the datagram decrements this number by 1. The router discards the datagram, if **TTL=0**.
When a source host sends the datagram, it stores a number in TTL field. This value is approximately 2 times the maximum number of routes between any two hosts.

Protocol (8 bits)

- This field defines the higher-level protocol that uses the services of the IPv4layer.
- An IPv4 datagram can encapsulate data from several higher-level protocols such as TCP, UDP, ICMP, and IGMP.
- This field specifies the final destination protocol to which the IPv4 datagram is delivered.

Checksum (16 bits)

The checksum in the IPv4 packet covers only the header, not the data. There are two reasons:

i. All higher-level protocols that encapsulate data in the IPv4 datagram have a checksum field that covers the whole packet. The checksum for the IPv4 datagram does not have to check the encapsulated data.

ii. The header of the IPv4 packet changes with each visited router, but the data do not changes. So the checksum includes only the part that has changed.

Options

Options are not required for a datagram. They can be used for network testing and debugging.

End of Option: An end-of-option option is a 1-byte option used for padding at the end of the option

field.

Record Route: A record route option is used to record the Internet routers that handle the datagram. It can list up to nine router addresses. It can be used for debugging and management purposes.

Strict Source Route: A strict source route option is used by the source to predetermine a route for the datagram as it travels through the Internet. The sender can choose a route with a specific type of service, such as minimum delay or maximum throughput.
If the datagram visits a router that is not on the list, the datagram is discarded and an error message is issued.

Loose Source Route: A 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.

Timestamp: A timestamp option is used to record the time of datagram processing by a router.

Source Address (32 bits) & Destination Address (32 bits)

- These two fields define the IPv4 address of the Source and Destination respectively.
- These fields must remain unchanged during the time the IPv4 datagram travels from the source host to the destination host.

Disadvantages of IPv4

1. Despite all short-term solutions, such as subnetting, classless addressing, and NAT, address depletion is still a long-term problem in the Internet.
2. The Internet must accommodate real-time audio and video transmission. This type of transmission requires minimum delay strategies and reservation of resources not provided in the IPv4 design.
3. The Internet must accommodate encryption and authentication of data for some applications. No encryption or authentication is provided byIPv4.

IPV6 DELIVERY MECHANISM

- IPV6 is introduced to overcome the deficiencies ofIPv4.
- IPv6 is also called as IPng (Internetworking Protocol next generation).
- In IPv6, the Internet protocol was extensively modified to accommodate the growth of the Internet. Packet format, Length of IP address, ICMP, IGMP, ARP, RARP, RIP routing protocol are also modified inIPv6.

Advantages of IPv6

- **Larger address space** An IPv6 address is 128 bits long whereas IPv4 is 32-bitaddress.
- **Better header format** IPv6 uses a new header format in which options are separated from the base header. when options are needed it is inserted between the base header and the upper-layer data.
- **New options** IPv6 has new options to allow for additional functionalities.
- **Allowance for extension** IPv6 is designed to allow the extension of the protocol if required by new technologies or applications.
- **Resource Allocation** In IPv6 the type-of-service field has been removed, but a mechanism called ***flow label*** has been added to enable the source to request special handling of the packet. This mechanism can be used to support traffic such as real-time audio and video.
- **More Security** The encryption and authentication options in IPv6 provide confidentiality and integrity of the packet.

Packet Format

In IPv6 each packet is composed of a mandatory base header followed by the payload.
The **Payload** consists of two parts:

- Optional extension headers
- Data from an upper layer

The **Base Header** occupies 40 bytes, whereas the extension headers and data from the upper layer contain up to 65,535 bytes of information.

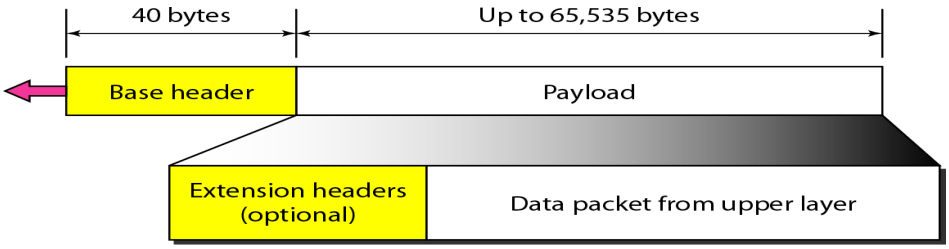
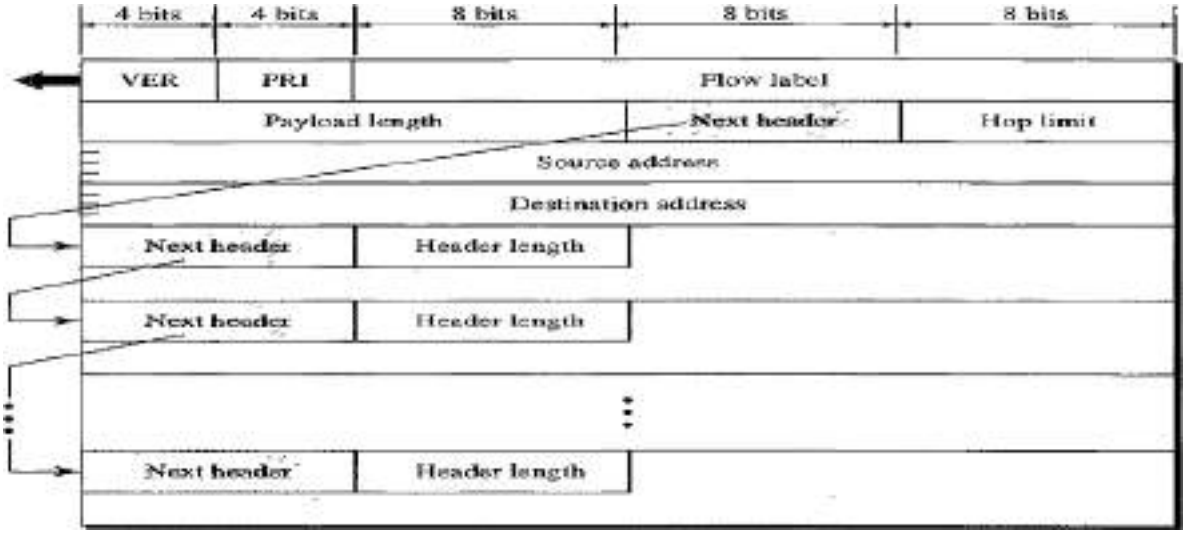


Fig: IPv6 Datagram header and payload

Base Header

Fields in IPv6 datagram are:



- **Version (4-bit)**

  This field defines the version number of the IP. For IPv6, the value is 6.

- **Priority(4-bit)**

  The priority field defines the priority of the packet with respect to traffic congestion.

- **Flow label (24-bit or 3Byte)**

  Flow label field that is designed to provide special handling for a particular flow of data.

- **Payload length (16 bit or 2Byte)**

  Payload length field defines the length of the IP datagram excluding the base header.

- **Next header(8-bit)**

  The next header is an 8-bit field defining the header that follows the base header in the datagram. The next header is either optional extension headers used by IP or the header of TCP or UDP encapsulated packet.

  Note: This field in IPv4 is called the *protocol.*

- **Hop limit (8 bit)**

  Hop limit field serves the same purpose as the TTL field in IPv4

- **Source address (128-bit or 16 Byte) and Destination Address (128 bit or 16Byte)**

  The source address field is a 16-byte (128-bit) Internet address that identifies the original source of the datagram.

  The destination address field is a 16-byte (128-bit) Internet address that usually identifies the final destination of the datagram. If source routing is used, this field contains the address of the next router.

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 (No Next Header)
60	Destination Option

Priority

The priority field of the IPv6 packet defines the priority of each packet with respect to other packets from the same source.

Example: If one of two consecutive datagrams must be discarded due to congestion, the datagram with the lower **packet priority** will be discarded.

IPv6 divides traffic into two broad categories:

i. Congestion-Controlled
ii. Non congestion-controlled

Congestion-Controlled Traffic

When there is congestion a source adapts itself to slowdown the traffic.

Example: TCP uses sliding window protocol can easily respond to the traffic.

Congestion-controlled data are assigned priorities from 0 to 7

Priority	Meaning	Description
0	No specific traffic	Priority 0 is assigned to a packet when the process does not define a priority.
1	Background data	defines data that are usually delivered in the background. Ex: Delivery of the news.
2	Unattended data traffic	If the user is not waiting (attending) for the data to be received, the packet will be given a priority of 2. Ex: Email
3	Reserved	
4	Attended bulk data traffic	A protocol that transfers data while the user is waiting to receive the data is given a priority of 4 Ex: FTP and HTTP
5	Reserved	
6	Interactive traffic	Protocols that need user interaction are assigned 6. Ex: TELNET
7	Controlled traffic	Routing Protocols are given Highest Priority 7. Ex: OSPF, RIP, SNMP

Non congestion-Controlled Traffic

- The source does not adapt itself to congestion. It is a type of traffic that expects minimum delay. Priority numbers from 8 to 15 are assigned to Non congestion-controlled traffic.
- In this traffic Discarding of packets is not desirable and Retransmission in most cases is impossible.

Examples: Real-time audio and video.

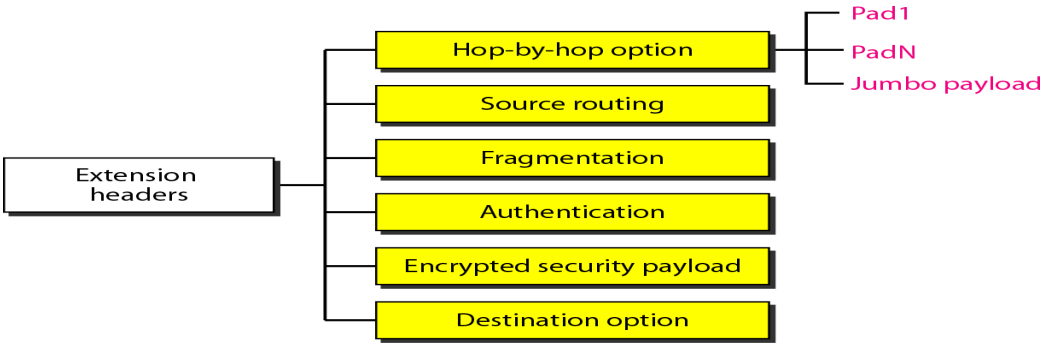
- **Priority 15:** It is given to data containing **Less Redundancy** (low-fidelity audio or video)
- **Priority 8:** It is given to data containing M**ore Redundancy** (high-fidelity audio or video)

Flow Label

- A sequence of packets, sent from a particular source to destination that needs special handling by routers is called a F**low of packets.**
- The combination of the source address and the value of the **Flow Label** uniquely define a flow of packets.
- To a router, a flow is a sequence of packets that share the same characteristics such as traveling the same path, using the same resources, having the same kind of security etc.
- A router that supports the handling of flow labels has a flow label table. The table has an entry for each active flow label.
- Each entry defines the services required by the corresponding flow label.
- When a router receives a packet it consults the flow label table instead of consulting the routing table and going through a routing algorithm to define the address of the next hop, it can easily look in a flow label table for the next hop.
- This mechanism speed up the processing of a packet by a router.

Extension Headers

To give greater functionality to the IP datagram, the base header can be followed by up to six types of extension headers.



Hop-by-Hop Option

The hop-by-hop option is used when the source needs to pass information to all routers visited by the datagram.

Only three options have been defined: Pad l, Pad N, and jumbo payload.

- The Pad l option is 1 byte long and is designed for 1 byte alignment purposes.
- Pad N is used when 2 or more bytes is needed for alignment.
- The jumbo payload option is used to define a payload longer than 65,535bytes.

Source Routing

- The source routing extension header combines the concepts of the strict source route and the loose source route options ofIPv4.

Fragmentation

- In IPv4, the source or a router is required to fragment if the size of the datagram is larger than the MTU of the network over which the datagram travels.
- In IPv6, only the original source can fragment. A source must use a path MTU discovery technique to find the smallest MTU supported by any network on the path.
- The source then fragments using this knowledge.

Authentication

- The authentication extension header has a dual purpose: it validates the message sender and ensures the integrity of data.

Encrypted Security Payload (ESP)

- ESP is an extension that provides confidentiality and guards against eavesdropping.

Destination Option

- It is used when the source needs to pass information to the destination only.
- Intermediate routers are not permitted access to this information.

Comparison between IPv4 Options and IPv6 Extension Headers

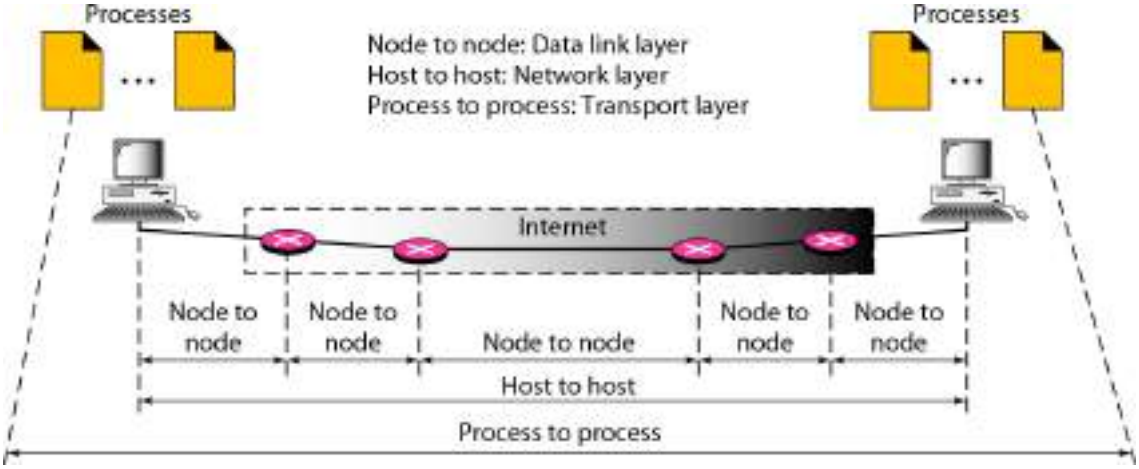
1. The no-operation and end-of-option options in IPv4 are replaced by Pad1 and Pad N options in IPv6.
2. The record route option is not implemented in IPv6 because it was not used.
3. The timestamp option is not implemented because it was not used.
4. The source route option is called the source route extension header inIPv6.
5. The fragmentation fields in the base header section of IPv4 have moved to the fragmentation extension header inIPv6.
6. The authentication and Encrypted Security Payload extension headers are new inIPv6.

PROCESS TO PROCESS DELIVERY

A Process is an Application program that runs on the host. At any moment several processes may be running on the source host and destination host.

Transport Layer is responsible for delivery of a packet from one process on source host to another process on destination host.

Two processes communicate in a client/server relationship.



Client/Server Paradigm

A Client is a process on a local host, whereas Server is a process on remote host. A Client needs services from Server. Both Client and Server processes have the same name.

Example: To get the day and time from a remote machine, we need a Daytime client process running on the local host and a Daytime server process running on a remote machine.

For communication to be done between two processes we must have to define following:

1. Local host
2. Local process
3. Remote host
4. Remote process

Addressing

- Transport layer needs an address called a **Port Number** or **Port Address** to choose among multiple processes running on the destination host.
- The port numbers are 16-bit integers between 0 and 65,535.
- Destination Port number is needed for Delivery. Source Port number is needed for the Reply.
- At the client side the port numbers are defined randomly whereas at the server side it uses Well-Known Port numbers.

Socket address

- The combination of an IP address and a port number is called a **Socket address**. UDP or TCP header contains the Port numbers.

- A transport layer protocol needs a pair of socket addresses: the client socket address and the server socket address.

TRANSPORT LAYER PROTOCOL

There are 3 transport layer protocols are implemented:

1. UDP
2. TCP
3. SCTP

USER DATAGRAM PROTOCOL (UDP)

- UDP is called a connectionless, unreliable transport protocol.
- UDP is a very simple protocol using a minimum of overhead. UDP performs very limited error checking.
- UDP is used when a process wants to send a small message and does not need reliability.
- Sending a small message by using UDP takes much less interaction between the sender and receiver than using TCP or SCTP.

User Datagram

UDP packets are called User Datagrams. Datagrams have fixed size header of length 8 bytes.

Fields of user datagram are:

Source port number (16 bits)

- This is the port number used by the process running on the source host.

Destination port number

- This is the port number used by the process running on the destination host.

Note: Source and Destination port number can range from 0 to 65,535.

Total Length (16 bits)

The total length includes UDP header plus Data, total length ranges from 0-65535.

Checksum (16 bits)

This field is used to detect errors over the entire user datagram (header plus data).

UDP Operation

UDP uses the following four concepts:

1. Connectionless Service
2. No Flow and Error control
3. Encapsulation and Decapsulation
4. Queuing

Connectionless Services

- UDP provides a connectionless service. There is no connection establishment and no connection termination. Each user datagram sent by UDP is an independent datagram.
- The user datagrams are not numbered. Each user datagram can travel on a different path.
- There is no relationship between the different user datagrams even if they are coming from the same source process and going to the same destination program.
- UDP processes are capable of sending short messages only.

No Flow and Error Control

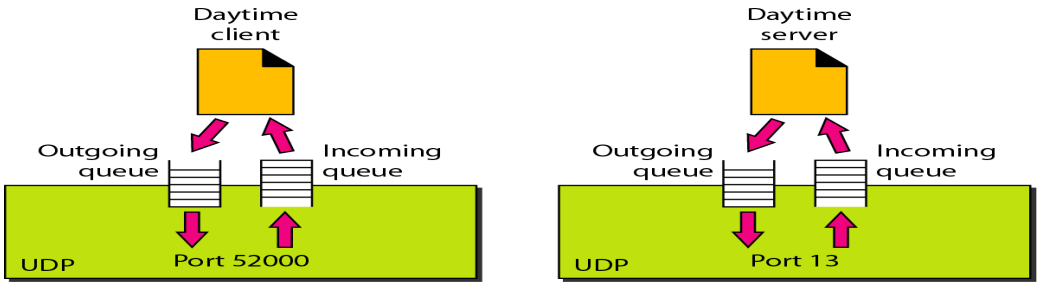
- There is no error control mechanism in UDP except for the checksum.

Encapsulation and Decapsulation

To send a message from one process to another, the UDP protocol encapsulates and decapsulates messages in an IP datagram.

Queuing

- Queues are associated with ports in UDP. Queues are associated with each process.
- Queues are created when a process is started. There are two types of queues are created: Incoming queue and Outgoing queue.
- A process wants to communicate with multiple processes; it obtains only one port number and one outgoing queue and one incoming queue.
- The queues function as long as the process is running. When the process terminates, the queues aredestroyed.



Uses of UDP

The following lists some uses of the UDP protocol:

- UDP is suitable for a process that requires simple request-response communication with little concern for flow and error control.
- UDP is suitable for a process with internal flow and error control mechanisms.
- UDP is a suitable transport protocol for multicasting.
- UDP is used for management processes such as SNMP.
- UDP is used for Route Updating Protocols such as Routing Information Protocol(RIP).

TRANSMISSION CONTROL PROTOCOL (TCP)

TCP is called a connection-oriented, reliable, process-to-process transport protocol. It adds connection-oriented and reliability features to the services of IP.

TCP Services

The following five services offered by TCP to the processes at the application layer

1. Process-to-Process Communication
2. Full-Duplex Communication
3. Connection-Oriented Service
4. Reliable Service
5. Stream Delivery Service

Process-to-Process Communication

TCP provides process-to-process communication using Well-known port numbers.

Full-Duplex Communication

TCP offers full-duplex service, in which data can flow in both directions at the same time. Each TCP then has a sending and receiving buffer and segments move in both directions.

Connection-Oriented Service

- When a process at site A wants to send and receive data from another process at site B, the following steps will occur:
  1. The two TCPs establish a virtual connection between them.
  2. Data are exchanged in both directions.
  3. The connection is terminated.
- The TCP segment is encapsulated in an IP datagram and can be sent out of order or lost or corrupted must be resent.
- TCP creates a stream-oriented environment in which it accepts the responsibility of delivering the bytes in order to the other site.

Reliable Service

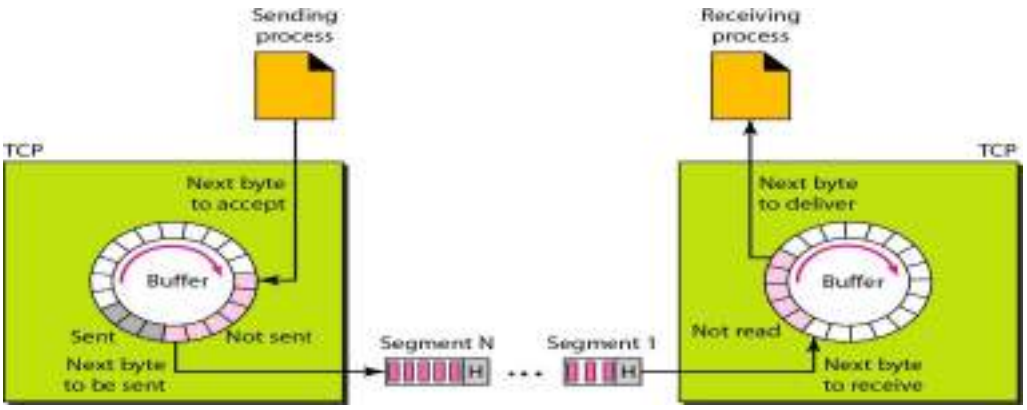
TCP is a reliable transport protocol. It uses an acknowledgment mechanism to check the safe and sound arrival of data.

Stream Delivery Service

- TCP allows the sending process to deliver data as a stream of bytes and allows the receiving process to obtain data as a stream of bytes.
- The sending process produces the stream of bytes, and the receiving process consumes them.

Sending and Receiving Buffers

- TCP needs buffers for storage, because the sending and the receiving processes may not write or read data at the same speed.
- The buffers are implemented by using Circular Array where each location carries1-byte.
- There are two types of buffers are implemented: **Sending buffer** and **Receiving buffer**.



- At the sending site TCP keeps bytes in the buffer that have been sent but not yet acknowledged until it receives an acknowledgment.
- After the bytes in the buffer locations are acknowledged, the locations are recycled and they are available for use by the sending process.
- At the receiver site the circular buffer is divided into two areas:
- One area contains empty chambers to be filled by bytes received from the network.
- Other are contain received bytes that can be read by the receiving process.
- When a byte is read by the receiving process, the chamber is recycled and added to the pool of empty chambers.

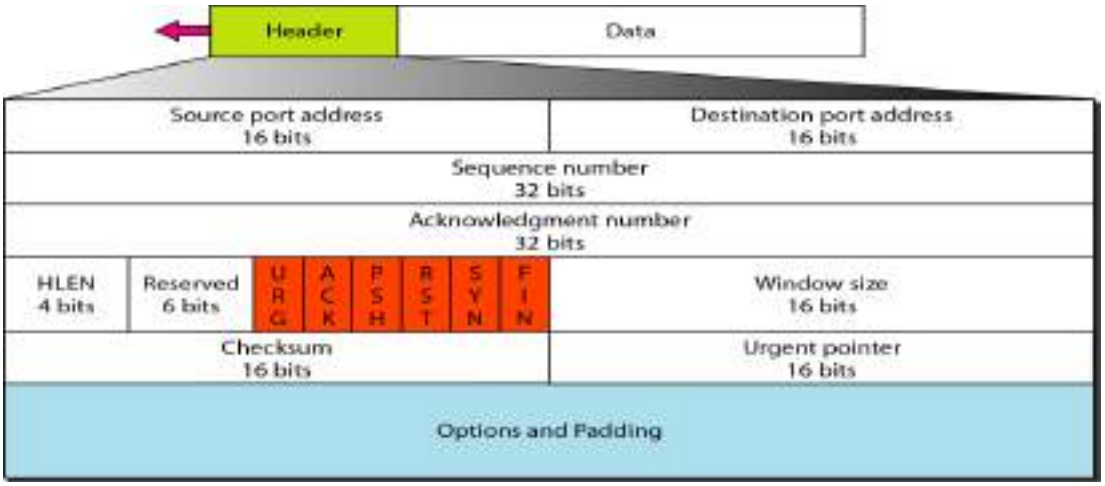
Segments

- TCP groups a number of bytes together into a packet called a segment.
- TCP adds a header to each segment and delivers them to the IP layer for transmission.
- The segments are encapsulated in IP datagrams and transmitted.
- The segments are not necessarily the same size.

TCP Segment

- A packet in TCP is called a segment.
- A segment consists of Segment Header and Data.
- The segment Header consists of 20-60 Byte. Header is 20 bytes if there are no options. If there are any options the length of the header varied upto 60bytes.

The segment format can be given as:



- **Source port address (16bit)**

It defines the port number of the application program in the host that is sending the segment.

- **Destination port address(16-bit)**

It defines the port number of the application program in the host that is receiving the segment.

- **Sequence number (32bit)**

This field defines the number assigned to the first byte of data contained in this segment. The sequence number tells the destination which byte in this sequence comprises the first byte in the segment.

- **Acknowledgment number (32 bit)**

This field defines the byte number that the receiver of the segment is expecting to receive from the other party. Acknowledgment and data can be piggybacked together.

- **Header length (4bit)**

This field indicates the number of 4-byte words in the TCP header. The
length of the header can be between 20 and 60 bytes.
The value of this field can be between 5 (5 x 4 =20) and 15 (15 x 4 =60).

- **Reserved (6 bit)**

This field is reserved for future use.

- **Window size (16bit)**

This field defines the size of the window in bytes that the other party must maintain. The maximum size of the window is 65,535 bytes.
This value is normally referred to as the receiving window (rwnd) and is determined by the receiver.

- **Checksum (16bit)**

The checksum field in TCP is mandatory. For the TCP pseudo header, the value for the protocol field is 6.

- **Urgent pointer(l6-bit)**

This field is valid only if the urgent flag is set. It is used when the segment contains
urgent data.
It defines the number that must be added to the sequence number to obtain the number of the last urgent byte in the data section of the segment.

- **Options (up to 40 bytes)** It defines the optional information in the TCP header.

- **Control flags (6bit)**

This field defines 6 different control bits or flags. One or more of these bits can be set at a time. These bits enable flow control, connection establishment and termination, connection abortion, and the mode of data transfer in TCP.

Flag	Description
URG	The value of the urgent pointer field is valid
ACK	The value of the acknowledgement field is valid.
PSH	Push the data
RST	Reset the connection
SYN	Synchronize sequence numbers during connection
FIN	Terminate the connection

TCP CONNECTION CONTROL:

Connection-oriented TCP transmission requires three phases:

1. Connection establishment
2. Data transfer
3. Connection termination

Connection Establishment Phase

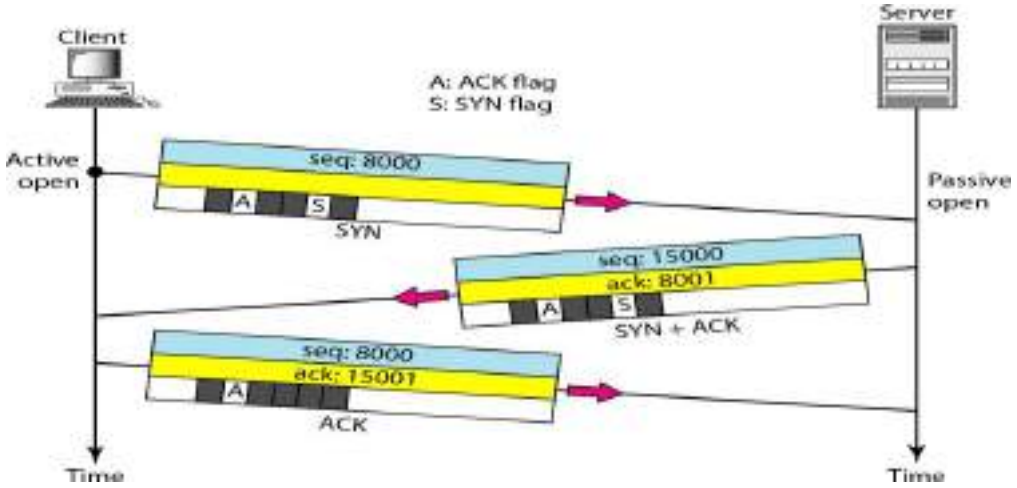
TCP uses Three way handshaking to establish a connection. The three way handshaking uses the sequence number, the acknowledgment number, the control flags and the window size.

- The process starts with the server. The server program tells its TCP that it is ready to accept a connection. This is called a request for a **Passive open**.
- A client that wishes to connect to an open server tells its TCP that it needs to be connected to that particular server. The client program issues a request for an **Active open.**
- Now TCP starts the Three way handshaking protocol process.

The three steps in this phase are as follows:

1. The client sends the first segment a SYN segment. The SYN flag in control field is set. This segment is for synchronization of sequence numbers. The SYN segment does not carry real data. SYN consumes one sequence number. When the data transfer starts the sequence number is incremented by1.
2. The server sends the second segment a SYN + ACK segment with 2 flag bits set: SYN and ACK. This segment has a dual purpose. It is a SYN segment for communication in the other direction and serves as the acknowledgment for the SYN segment. It consumes one sequence number.

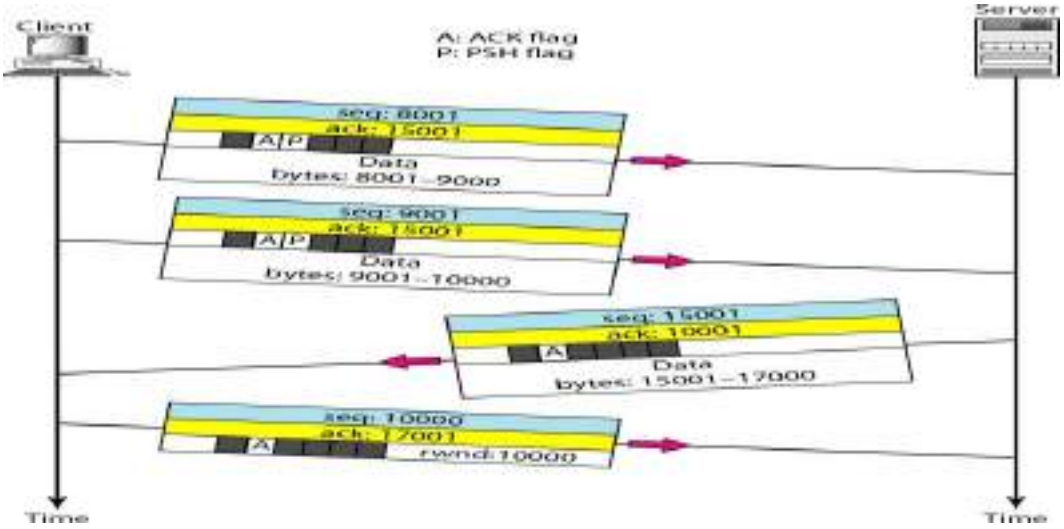
3. The client sends the third segment an ACK segment. It acknowledges the receipt of the second segment with the ACK flag and acknowledgment number field. The sequence number in this segment is the same as the one in the SYN segment.



Data Transfer Phase

- After connection is established, bidirectional data transfer can take place.
- The client and server can both send data and acknowledgments. The acknowledgment is piggybacked with the data.

Consider the below figure that shows the data transfer after connection is established.

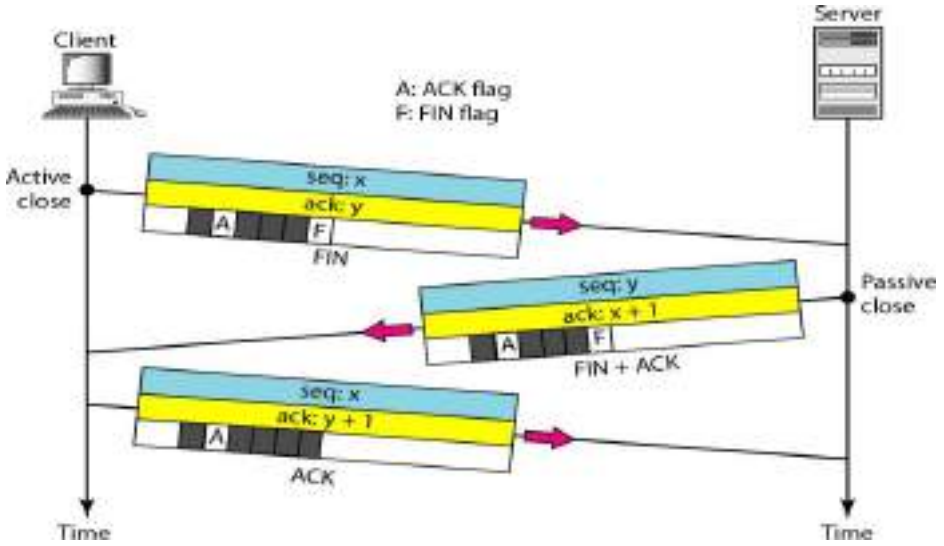


- The client sends 2000 bytes of data in two segments.
- The server then sends 2000 bytes in one segment.
- The client sends one more segment.
- The first three segments carry both data and acknowledgment, but the last segment carries only an acknowledgment because there are no more data to be sent.
- The data segments sent by the client have the PSH (push) flag set so that the server TCP knows to deliver data to the server process as soon as they are received.
- The segment from the server does not set the push flag.

Connection Termination

Client and Server involved in exchanging data can close the connection is called Connection Termination. It is usually initiated by the client.

Three way Handshaking is also used to terminate the connection by following steps:



1. The client TCP after receiving a close command from the client process sends the first segment a **FIN** segment in which the FIN flag is set.

A FIN segment can include the last chunk of data sent by the client or it can be just a control segment. If it is only a control segment, it consumes only one sequence number.

2. The server TCP after receiving the FIN segment informs server process of the situation and sends the second segment a **FIN + ACK** segment to confirm the receipt of the FIN segment from the client and at the same time to announce the closing of the connection in the other direction.

This segment can also contain the last chunk of data from the server. If it does not carry data, it consumes only one sequence number.

3. The TCP client sends the last segment an ACK segment to confirm the receipt of the FIN segment from the TCP server. This segment contains the acknowledgment number, which is **(sequence number + 1)** received in the FIN segment from the server. This segment cannot carry data and consumes no sequence numbers.

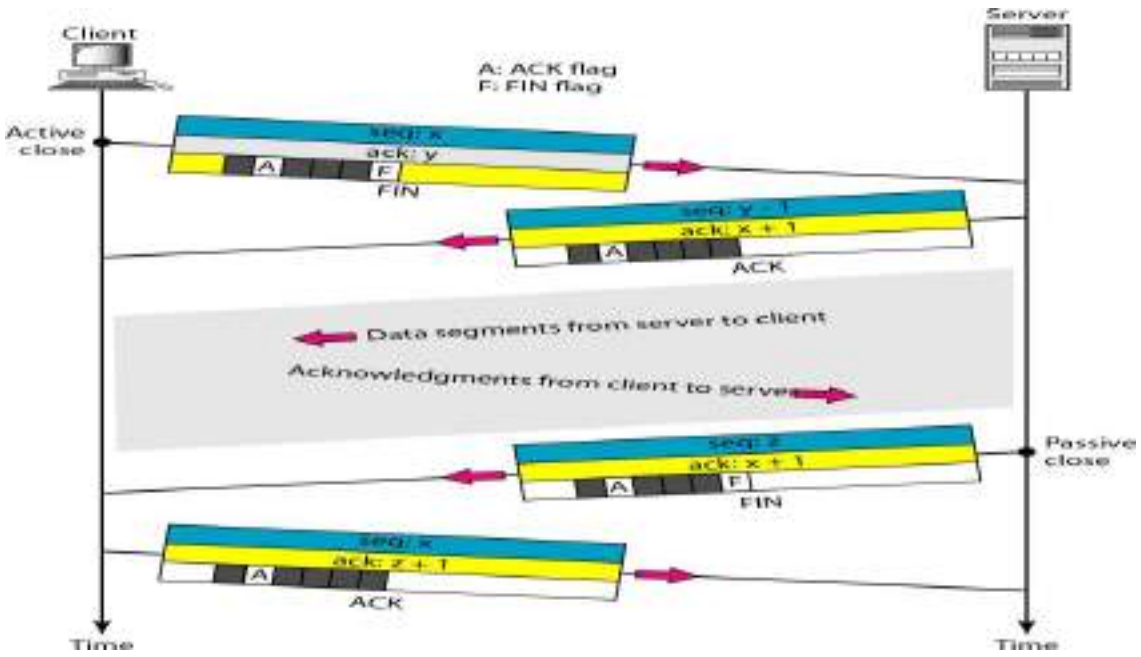
Note: In some rare situations Connection Termination can be done in Four way handshaking by using Half-close option.

Four way Handshaking using Half-Close

- In TCP one end can stop sending data while still receiving data is called a **Half-Close**. It is normally initiated by the client.
- It can occur when the server needs all the data before processing can begin.

Example: Sorting.

- When the client sends data to the server to be sorted, the server needs to receive all the data before sorting can start.
- This means the client after sending all the data can close the connection in the outbound direction.
- The inbound direction must remain open to receive the sorted data.
- The server after receiving the data still needs time for sorting. Its outbound direction must remain open.



- The client half-closes the connection by sending a FIN segment. The server accepts the half-close by sending the ACK segment.
- The data transfer from the client to the server stops. The server can still send data.
- When the server has sent all the processed data, it sends a FIN segment which is acknowledged by an ACK from the client.
- After half-closing of the connection, the data can travel from the server to the client and acknowledgments can travel from the client to the server but the client cannot send any more data segments to the server.
- The second segment (ACK) consumes no sequence number. Although client has received sequence number *y* - 1 and is expecting *y*, the server sequence number is still *y* -1.
- When the connection finally closes, the sequence number of the last ACK segment is still *x*, because no sequence numbers are consumed during data transfer in that direction.

CONGESTION

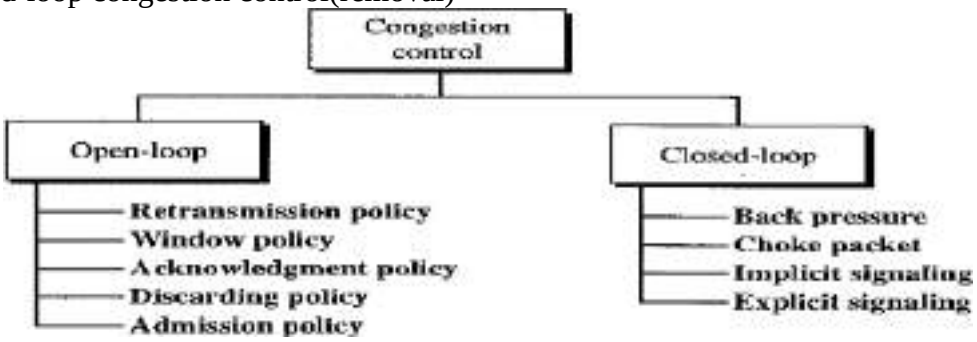
Congestion in a network may occur if the **load** on the network is greater than the *capacity* of the network.

CONGESTION CONTROL

Congestion control refers to techniques and mechanisms that can either prevent congestion before it happens or remove congestion after it has happened.

Congestion control mechanisms can be divided into two categories:

1. Open-loop congestion control(prevention)
2. Closed-loop congestion control(removal)



Open-Loop Congestion Control

In open-loop congestion control, policies are applied to prevent congestion before it happens. Here congestion control is handled by either the source or the destination.

Retransmission Policy

- The packet needs to be retransmitted by sender, when a packet is lost or corrupted.
- Retransmission is sometimes unavoidable. It may increase congestion in the network.
- The retransmission policy and the retransmission timers must be designed to optimize efficiency and at the same time prevent congestion.

Example: Retransmission policy used by TCP is designed to prevent or alleviate congestion.

Window Policy

- The Selective Repeat window is better than the Go-Back-N window for congestion control.
- In the Go-Back-N window, when the timer for a packet is expired several packets will be resent, although some may have arrived safe and sound at the receiver. This duplication may make the congestion worse.
- The Selective Repeat window tries to send the specific packets that have been lost or corrupted.

Acknowledgment Policy

- The acknowledgments are also part of the load in a network. Sending fewer acknowledgments means imposing less load on the network.
- If the receiver does not acknowledge every packet it receives, it may slow down the sender and help prevent congestion.
- A receiver may send an acknowledgment only if it has a packet to be sent or a special timer expires.

Discarding Policy

- A good discarding policy by routers may prevent congestion.
- Example: In audio transmission if the policy is to discard less sensitive packets when congestion happens, the quality of sound is still preserved and congestion is prevented.

Admission Policy

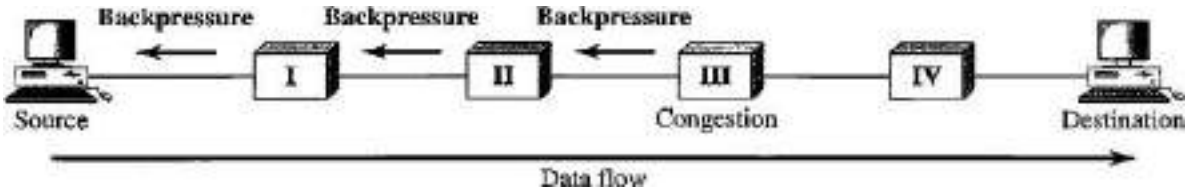
- An admission policy can prevent congestion in virtual-circuit networks.
- Switches first check the resource requirement of a data flow before admitting it to the network.
- A router can deny establishing a virtual-circuit connection if there is congestion in the network or if there is a possibility of future congestion.

Closed-Loop Congestion Control

Closed-loop congestion control mechanisms try to alleviate congestion after it happens. Several mechanisms have been used by different protocols are: Back pressure, Choke packet, Implicit signaling, Explicit signaling.

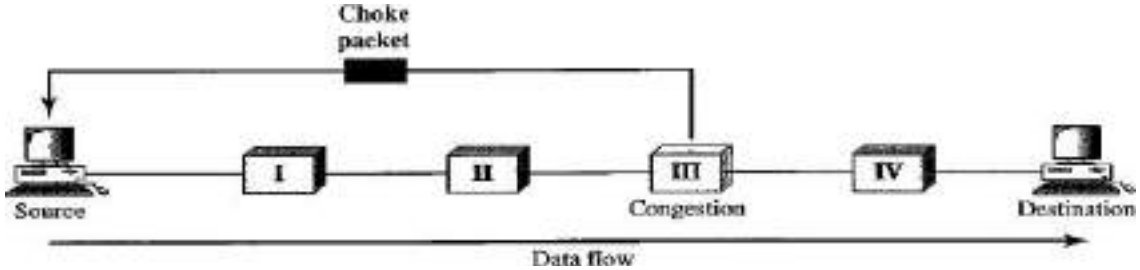
Backpressure

- In Backpressure mechanism, a congested node stops receiving data from the immediate upstream node.
- This may cause the upstream nodes to become congested and they reject data from their upstream nodes.
- Backpressure is a node-to-node congestion control that starts with a node and propagates in the opposite direction of data flow to the source.
- The backpressure technique can be applied only to virtual circuit networks, in which each node knows the upstream node from which a data flow is coming.



Choke Packet

- A choke packet is a packet sent by a node to the source to inform that congestion has occurred.
- In the choke packet method, the warning is sent from the router, which has encountered congestion to the source station directly. The intermediate nodes through which the packet has traveled are not warned.



Implicit Signaling

- In implicit signaling, there is no communication between the congested nodes and the source.
- Source guesses that there is congestion somewhere in the network from other symptoms.

Example: when a source sends several packets and there is no acknowledgment for a while,

one assumption is that the network is congested. The delay in receiving an acknowledgment is interpreted as congestion in the network and the source should slow down sending speed.

Explicit Signaling

- The node that experiences congestion can explicitly send a signal to the source or destination.
- In explicit signaling method, the signal is included in the packets that carry data.
- Explicit signaling can occur in either the forward or the backward direction.
- **Backward Signaling** A bit can be set in a packet moving in the direction opposite to the congestion. This bit can warn the source that there is congestion and that it needs to slow down to avoid the discarding of packets.
- **Forward Signaling** A bit can be set in a packet moving in the direction of the congestion. This bit can warn the destination that there is congestion. The receiver in this case can use policies, such as slowing down the acknowledgments to get rid of the congestion.

CONGESTION CONTROL in TCP

TCP uses congestion control to avoid congestion or alleviate congestion in the network.

Congestion Window

- In TCP the sender window size is determined by the available buffer space in the receiver (rwnd). That means receiver dictates the sender about the size of the sender's window.
- If the network cannot deliver the data as fast as they are created by the sender then the network must tell the sender to slow down. So the network can also determine the size of the sender's window.
- Hence the Sender window size is determined by the receiver and by congestion in the network.
- The actual size of the window is the minimum of (receiver window, congestion window).

Actual Window Size = minimum (rwnd,cwnd)

Congestion Policy

There are three phases in TCP's Congestion policy: Slow start, Congestion Avoidance and Congestion Detection.

Slow Start: Exponential Increase

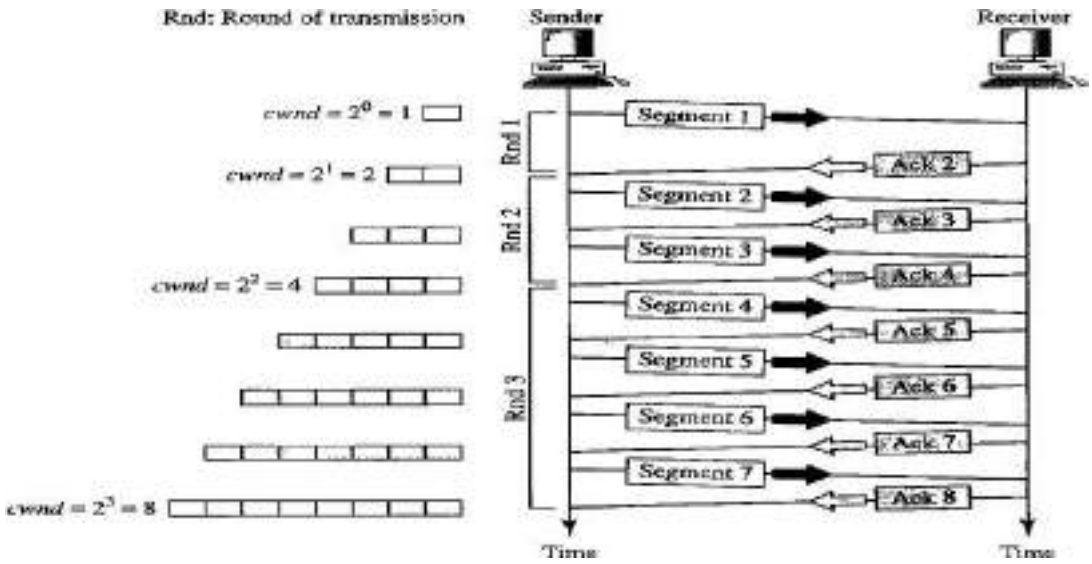
- In this algorithm is based on the size of the congestion window (cwnd) starts with one maximum segment size (i.e. cwnd=1MSS).
- The MSS is determined during connection establishment by using an option of the same name.
- The size of the window increases one MSS each time an acknowledgment is received.
- As the name implies, the window starts slowly but grows exponentially.

Example: Consider the below figure that shows the communication between client and server by using Slow start mechanism.

- If Receiver window (rwnd) is much higher than Congestion window (cwnd) then the sender window size always equals Congestion window size (cwnd). Each segment is acknowledged individually.
- The sender starts with **cwnd=1**MSS (i.e.) the sender can send only one segment.
- After receipt of the acknowledgment for segment 1, the size of the congestion window is

increased by 1 (i.e.) **cwnd =2.**

- Now two more segments can be sent. When each acknowledgment is received, the size of the window is increased by 1MSS.
- When all seven segments are acknowledged **cwnd =8.**



Start □ **cwnd=1**

After round 1 □ **cwnd=2^1=2**

After round 2 □ **cwnd=2^2=4**

After round 3 □ **cwnd=2^3=8**

- If there is delayed ACKs, the increase in the size of the window is less than power of2.
- Slow start cannot continue indefinitely. There must be a threshold to stop this phase.
- The sender keeps track of a variable called **ssthresh** (slow-start threshold). In general the value of **ssthresh** is 65,535bytes.
- When the size of window in bytes reaches threshold value then the slow start stops and the next phase starts.

QUALITY OF SERVICE (QoS)

It is an internetworking issue. There are four Flow characteristics are related to Qos.
They are: Reliability, Delay, Jitter and Bandwidth.

Reliability: Lack of reliability means that losing a packet or acknowledgment, which entails retransmission. Electronic mail, file transfer, and Internet access have reliable transmissions.

Delay: Applications can tolerate delay in different degrees. Telephony, audio conferencing, video conferencing, and remote log-in need minimum delay.

Jitter: It is the variation in delay for packets belonging to the same flow. If four packets depart at times 0, 1, 2,3 and arrive at 20, 21, 22, 23 all have the same delay = 20 units of time. Audio and video applications accept the delay of packets as long as if the delay is same for all the packets.

Bandwidth: Different applications need different bandwidths. In video conferencing, it needs to send millions of bits per second to refresh a color screen while the total number of bits in an e-mail may not reach even a million.

TECHNIQUES TO IMPROVE QoS

There are four techniques that will improve the QoS:

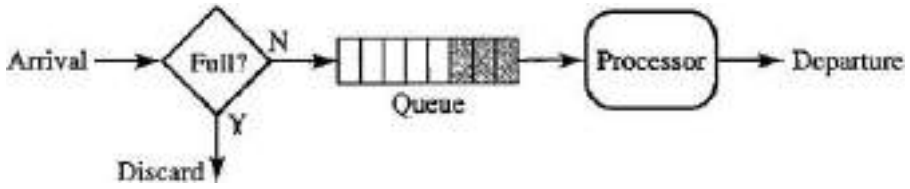
1. Scheduling
2. Traffic shaping
3. Admission control
4. Resource reservation

Scheduling

Packets from different flows arrive at a switch or router for processing.
Several scheduling techniques are designed to improve the quality of service such as: FIFO Queuing, Priority Queuing and Weighted fair queuing.

First-In-First-Out Queuing (FIFO)

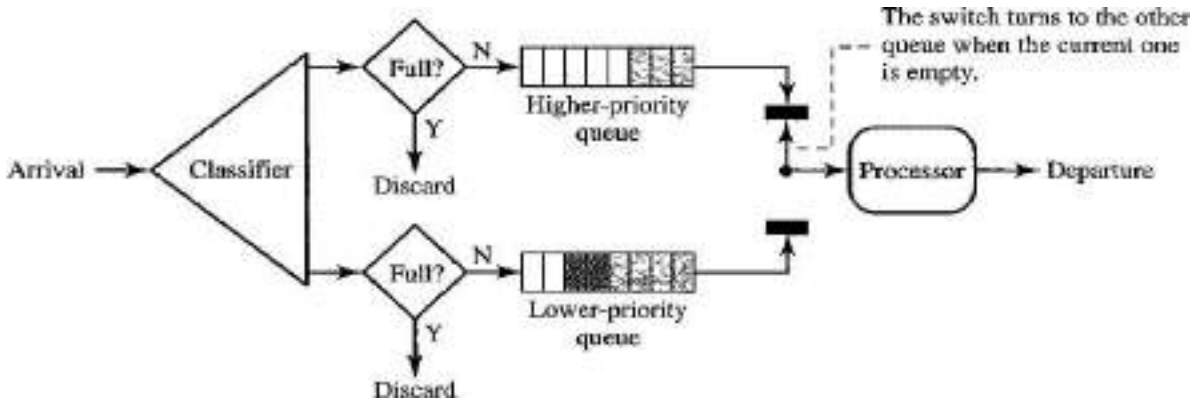
- In FIFO queuing, the packets wait in a buffer (queue) until the node (router or switch) is ready to process them.
- If the average arrival rate is higher than the average processing rate then the queue will fill up and new packets will be discarded.
- That means the speed of receiving the packets in to buffer is more than the speed of processing the packets by processor then the buffer will be filled completely and then there is no place for newly arrived packets hence these packets will be discarded.



Priority Queuing

- In priority queuing, packets are first assigned to a priority class. Each priority class has its own queue.
- The packets in the highest-priority queue are processed first. Packets in the lowest-priority queue are processed last.
- The system does not stop serving a queue until it is empty.

- There is drawback with priority queues called **Starvation** (i.e.) if there is a continuous data flow in a high-priority queue, the packets in the low-priority queues will never have a chance to be processed.

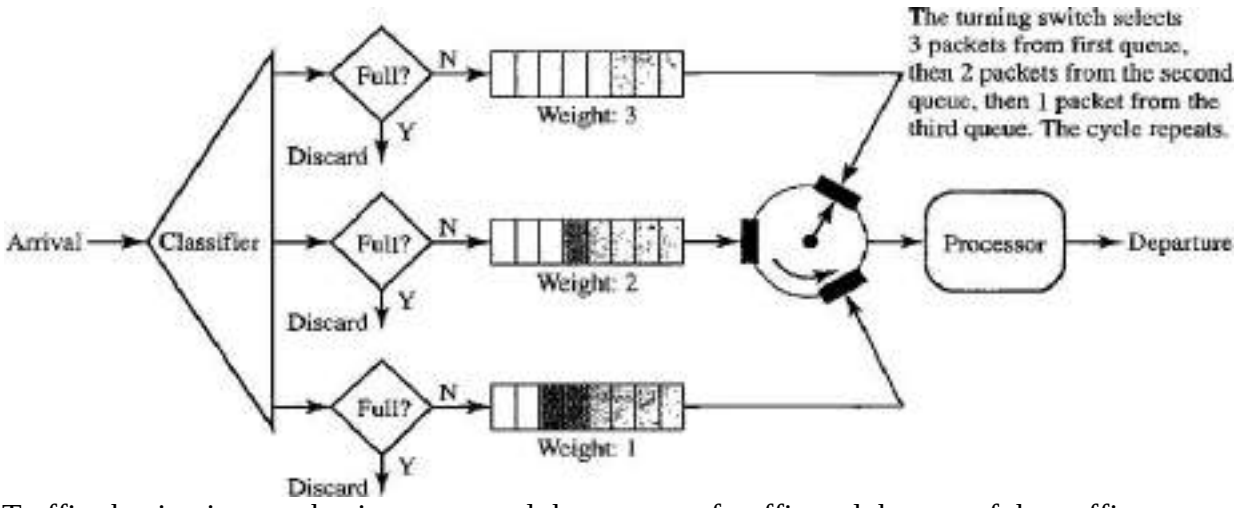


Weighted Fair Queuing

- In this technique, the packets are still assigned to different classes and admitted to different queues.
- The queues are weighted based on the priority of the queues; higher priority means a higher weight.
- The system processes packets in each queue in a **Round-Robin** fashion with the number of packets selected from each queue based on the corresponding weight.

Example: If the weights are 3, 2, and 1, three packets are processed from the first queue, two from the second queue, and one from the third queue. If the system does not impose priority on the classes, all weights can be equal. In this way, we will achieve fair queuing with priority.

Traffic Shaping

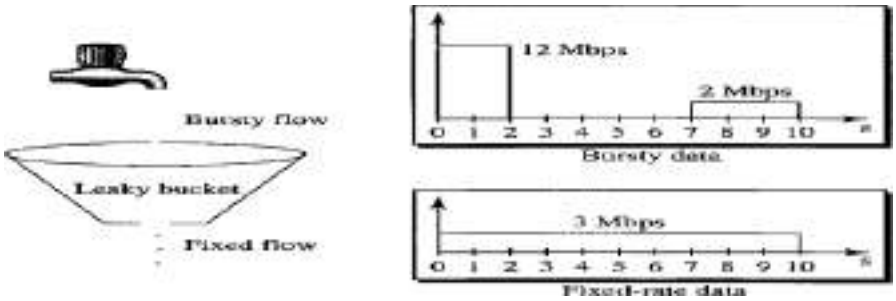


Traffic shaping is a mechanism to control the amount of traffic and the rate of the traffic sent to the network. Two techniques can shape traffic: **Leaky bucket** and **Token bucket.**

Leaky Bucket

- If a bucket has a small hole at the bottom, the water leaks from the bucket at a constant rate as long as there is water in the bucket. The rate at which the water leaks does not depend on the rate at which the water is input to the bucket unless the bucket is empty. The input rate can vary, but the output rate remains constant.

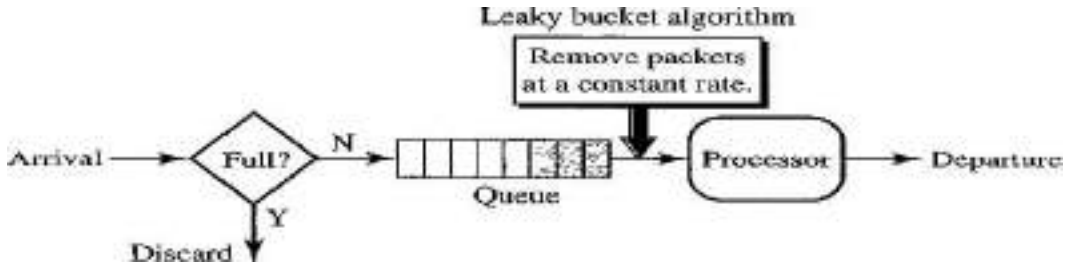
- In networking, a technique called leaky bucket can smooth out Bursty traffic. Bursty chunks are stored in the bucket and sent out at an average rate.



- In the above figure, the network has committed a bandwidth of 3 Mbps for a host. The use of the leaky bucket shapes the input traffic to make it conform to this commitment.
- The host sends a burst of data at a rate of 12 Mbps for 2 sec, for a total of 24 M bits of data.
- The host is silent for 5 sec and then sends data at a rate of 2 Mbps for 3 sec, for a total of 6 M bits of data.
- In total the host has sent 30 M bits of data inl0s.
- The leaky bucket smoothen the traffic by sending out data at a rate of 3 Mbps during the same 10sec.
- Without the leaky bucket, the beginning burst may have hurt the network by consuming more bandwidth than is set aside for this host.

This way the leaky bucket may prevent congestion.

Consider the below figure that shows implementation of Leaky Bucket:



- A FIFO queue holds the packets. If the traffic consists of fixed-size packets the process removes a fixed number of packets from the queue at each tick of the clock.
- If the traffic consists of variable-length packets, the fixed output rate must be based on the number of bytes orbits.

The following is an algorithm for variable-length packets:

1. Initialize a counter to *n* at the tick of the clock
2. If *n* is greater than the size of the packet, send the packet and decrement the counter by the packet size. Repeat this step until *n* is smaller than the packet size.
3. Reset the counter and go to step1.

Problems with Leaky Bucket

1. The leaky bucket is very restrictive. If a host is not sending for a while, its bucket becomes empty.
2. After some time if the host has bursty data, the leaky bucket allows only an average rate. The time when the host was idle is not taken into account.

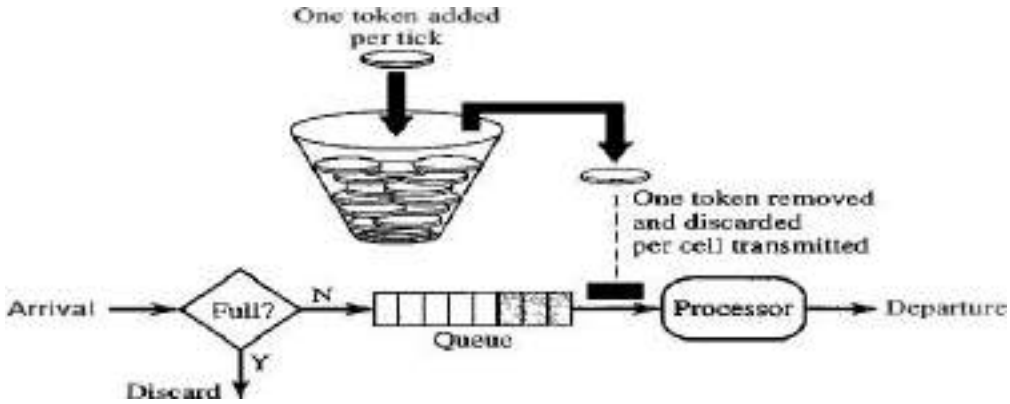
These problems can be overcome by Token bucket algorithm.

Token Bucket

- The token bucket algorithm allows idle hosts to accumulate credit for the future in the form of tokens.
- For each tick of the clock, the system sends *n* tokens to the bucket.
- The system removes one token for every cell (or byte) of data sent.

Example: If *n* is 100 and the host is idle for 100 ticks, the bucket collects 10,000 tokens.

- Now the host can consume all these tokens in one tick with 10,000 cells, or the host takes 1000 ticks with 10 cells per tick.
- The host can send bursty data as long as the bucket is not empty.



The token bucket can easily be implemented with a counter.

- The token is initialized to zero.
- Each time a token is added, the counter is incremented by1.
- Each time a unit of data is sent, the counter is decremented by1.
- When the counter is zero, the host cannot send data.

Resource Reservation

- A flow of data needs resources such as a buffer, bandwidth, CPU time, and soon.
- The quality of service is improved if these resources are reserved before data transfer.

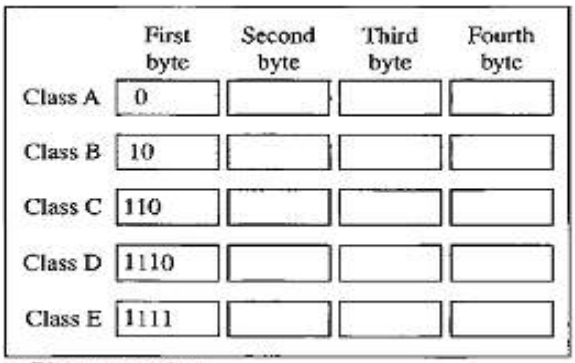
Admission Control

- Admission control refers to the mechanism used by a router or a switch to accept or reject a flow based on predefined parameters called flow specifications.
- Before a router accepts a flow for processing, it checks the flow specifications to see if its capacity and its previous commitments to other flows can handle the new flow.

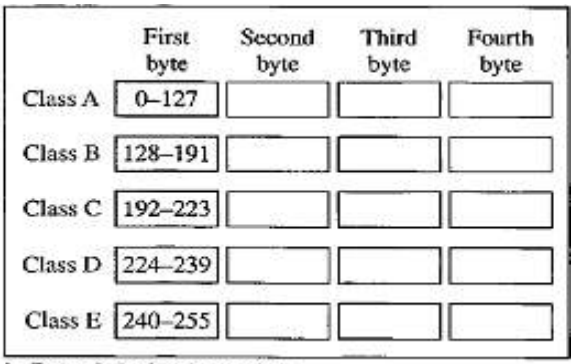
Note: Capacity is in terms of bandwidth, buffer size, CPU speed, etc.

CLASSFUL ADDRESSING

- Initially IPv4 used the concept of Classful addressing.
- In classful addressing, the address space is divided into five classes: A, B, C, D, and E.
- If the address is given in binary notation, the first few bits can immediately tell us the class of the address.
- If the address is given in decimal-dotted notation, the first byte defines the class.



a. Binary notation



b. Dotted-decimal notation

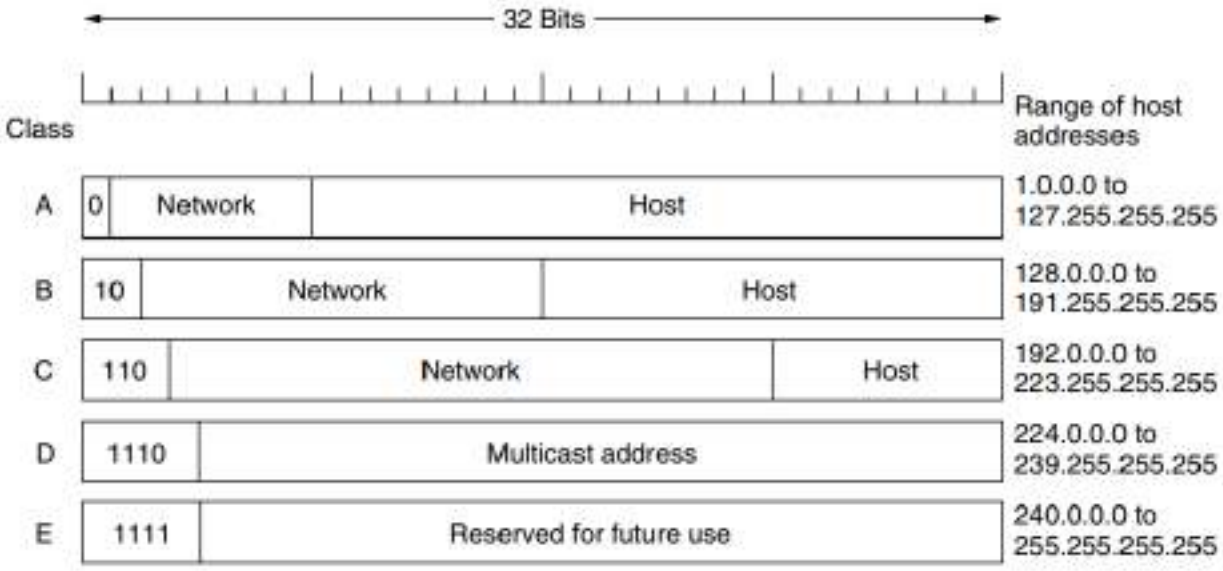
Classes and Blocks

- Each class is divided into a fixed number of blocks.
- Size of the each block is also fixed.

Class	No of Blocks	Block Size	Application
A	128	16,777,216	Unicast
B	16,384	65,536	Unicast
C	2,097,152	256	Unicast
D	1	268,435,456	Multicast
E	1	268,435,456	Reserved

Purpose of classes:

- **Class A** addresses were designed for large organizations with a large number of attached hosts or routers but most of the addresses in class A were wasted and were not used.
- **Class B** addresses were designed for midsize organizations with tens of thousands of attached hosts or routers.
- **Class C** addresses were designed for small organizations with a small number of attached hosts or routers.
- **Class D** addresses were designed for multicasting, each address in class D is used to define one group of hosts on the Internet.
- **Class E** addresses were reserved for future use.



Netid and Hostid

- In classful addressing, an IP address in class A, B, or C is divided into Network id and Host id.

Mask or Default Mask

- A default mask is a 32-bit number made of contiguous 1's followed by contiguous0's.
- The mask can help us to find the Netid and the Hostid.
- For example, the mask for a class A address has eight 1s, which means the first 8 bits of any address in class A define the Netid; the next 24 bits define the Hostid.

The masks for classes A, B, and C are:

Class	Binary	Dotted-Decimal	CIDR
A	**11111111** 00000000 00000000 00000000	**255**.0.0.0	**/8**
B	**11111111 11111111** 00000000 00000000	**255.255**.0.0	**/16**
C	**11111111 11111111 11111111** 00000000	**255.255.255**.0	**/24**

Subnetting

- Subnetting is a process of dividing a large block into smaller contiguous groups and assigns each group to smaller networks (subnets) or share a part of the addresses with neighbors.
- Subnetting increases the number of 1's in the mask.

Supernetting

- In supernetting, an organization can combine several blocks to create a larger range of addresses. Supernetting decreases the number of 1's in the mask.

Example: an organization that needs 1000 addresses can be granted four contiguous class C blocks. The organization can then use these addresses to create one super

network.

Address Depletion

- The number of available IPv4 addresses is decreasing as the number of internet users are increasing.
- We have run out of class A and B addresses, and a class C block is too small for most midsize organizations.
- One solution that has alleviated the problem is the idea of **Classless Addressing**.

CLASSLESS ADDRESSING

Purpose

- Classless addressing was designed and implemented to overcome address depletion and give more organizations access to the Internet.
- In this scheme, there are no classes, but the addresses are still granted in blocks.

Address Blocks

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

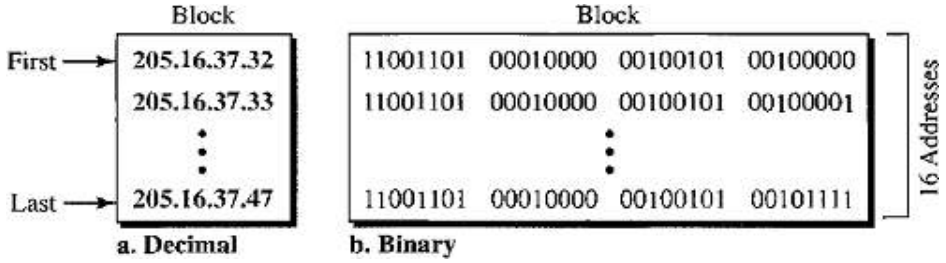
Example:

- For a large organization may be given thousands of addresses
- For a house two addresses are sufficient
- An Internet service provider may be given hundreds of thousands based on the number of customers it may serve.

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 (1, 2, 4, 8, ...).
3. The first address must be evenly divisible by the number of addresses.

Consider the below figure for classless addressing that shows a block of addresses, in both binary and dotted-decimal notation, granted to a small business that needs 16 addresses.



It satisfies all 3 restrictions:

- The addresses are contiguous.
- The number of addresses is a power of 2 (16 =2^4).
- The first address is divisible by 16. The first address, when converted to a decimal number, is 3,440,387,360, which when divided by 16 results in215,024,210.

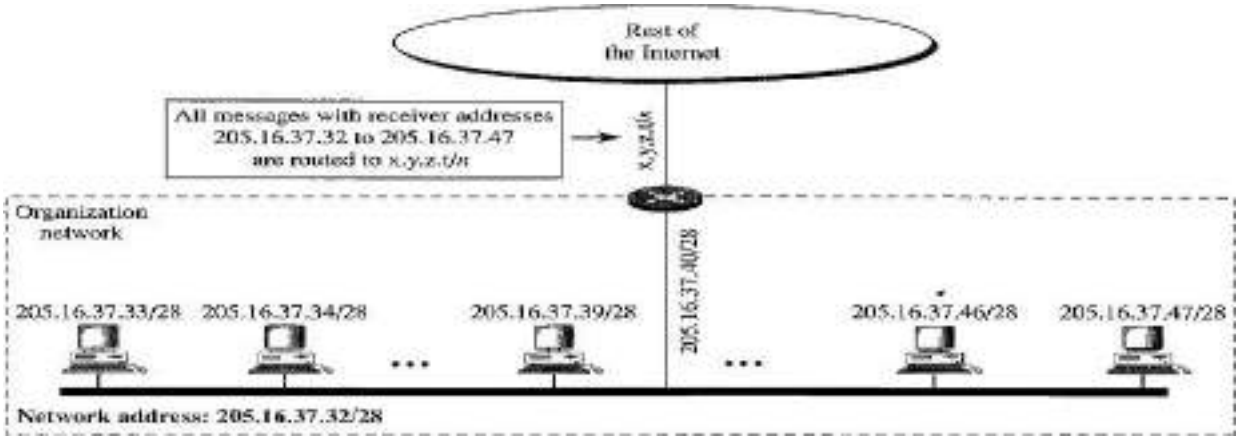
Mask

A mask is a 32-bit number in which the ***n*** leftmost bits are 1's and the **32 - *n*** rightmost bits are 0's, where ***n= 0 to32.***

In 1Pv4 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*** defines the mask. ***/n*** is called as CIDR notation.

- **First Address** in the block can be found by setting the rightmost **32 - *n*** bits to0's.
- **Last Address** in the block can be found by setting the rightmost **32 - *n*** bits to1's.
- **Number of Addresses** in the block can be found by using the formula 2^{32-n}.

- **Network Address** is the first address in the block and defines the organization network. Usually the first address is used by routers to direct the message sent to the organization from the outside.



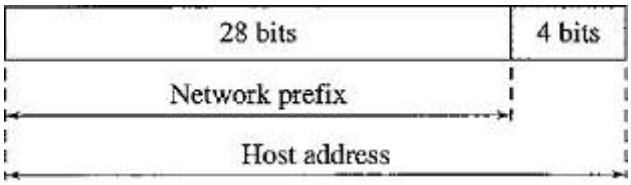
Example: **205.16.37.39/28** or **11001101 00010000 00100101 00100111**

- First address: 11001101 000100000100101 0010000 or 205.16.37.**32**
- Last address: 11001101 00010000 001001010010 1111 or205.16.37.**47**
- Number of Addresses: $2^{32-28} = 2^4$**=16.**
- Network Address (First Address) 11001101 000100000100101 0010000 or205.16.37.32

Netid and Hostid

- The ***n*** leftmost bits of the address **x.y.z.t/n** define the **network address** or **prefix**.
- The **(32 – *n)*** rightmost bits define the particular **suffix** or **host address** (computer or router) connected to the network.

205.16.37.39/28 or **11001101 00010000 00100101 0010 0111**



Network Address Translation (NAT)

- As the number of home users and small business users are increasing day by day

it is not possible to give each and every user to one IPv4 address due to shortage of IPv4 addresses.

- In order to overcome this problem the developers designed the concepts of private IP address and Network Address Translation(NAT).
- **NAT** enables a user to have a large set of addresses internally (private IP addresses) and a small set of addresses externally (public IP addresses).

The Internet authorities have reserved three sets of addresses as private addresses:

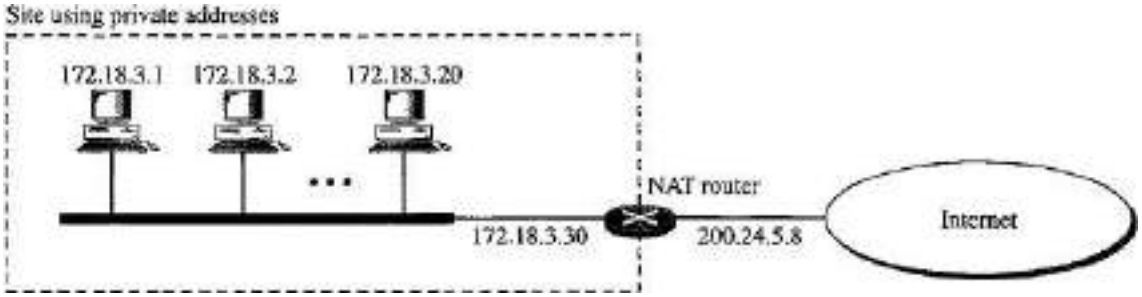
Range	Total
10.0.0.0 to 10.255.255.255	2^{24}
172.16.0.0 to 172.31.255.255	2^{20}
192.168.0.0 to 192.168.255.255	2^{16}

Any organization can use an address out of this set without permission from the Internet authorities.

- They are unique inside the organization, but they are not unique globally. No router will forward a packet that has one of these addresses as the destination address.

Example:

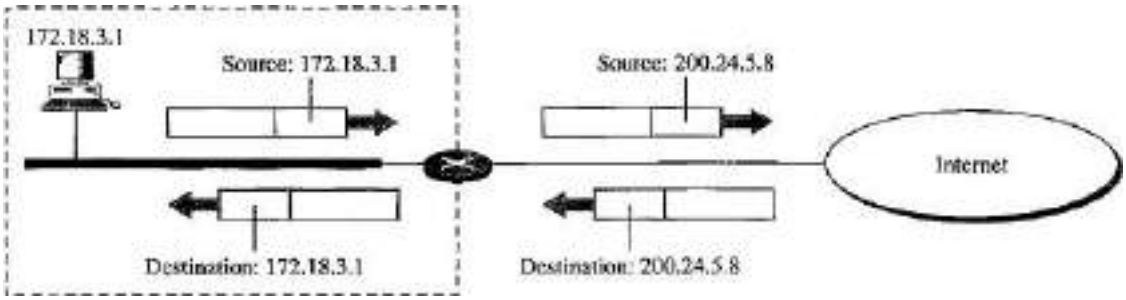
Consider the below figure describes the private network with private addresses:



- The NAT router has one public address **200.24.5.8,** it is a global address.
- The internal devices having addresses from **172.18.3.1** to **172.18.3.30** these are local addresses.

Address Translation

- All the outgoing packets go through the NAT router, which replaces the **source address** in the packet with the **global** NAT address.
- All incoming packets also pass through the NAT router, which replaces the **destination address** in the packet (the NAT router global address) with the appropriate **private address.**



Translation Table

There are two types of translation tables:

1. Two Column translation table (Using one IP address)
2. Five column translation table (Using IP addresses and Port Numbers)

Two Column Translation Table

- It contains two columns: Private Address and External Address.
- In this strategy, communication must always be initiated by the private network.
- When the router translates the source address of the outgoing packet, it also makes note of the destination address-where the packet is going.
- When the response comes back from the destination, the router uses the source address of the packet (as the external address) to find the private address of the packet.

Translation Table

Private	External
172.18.3.1	25.8.2.10
…	….
….	….

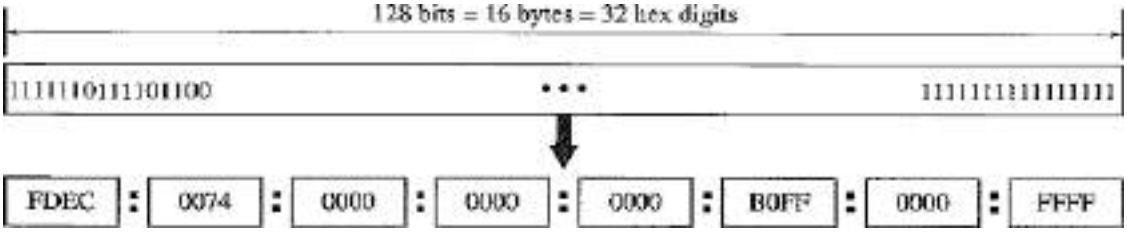
IPv6 ADDRESSES (Internetworking Protocol version6)

Why IPv6?

- In order to overcome the problems of address depletion.
- It eliminates the concept of NAT and Private Addresses.
- There is no need for classless addressing and DHCP.

Structure

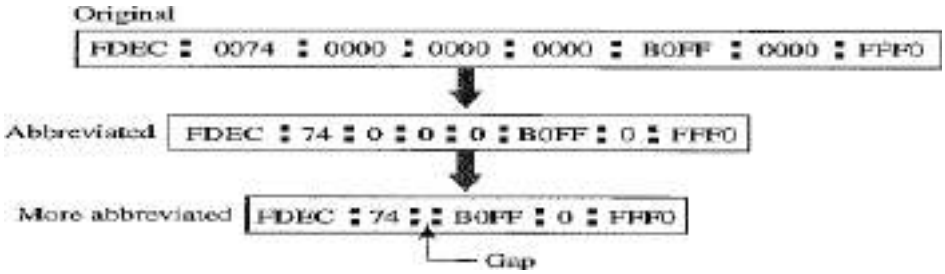
IPv6 specifies hexadecimal colon notation (0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F).
An IPv6 address consists of 16 bytes (octets); it is 128 bits long. 128 bits is divided into eight sections. Each of 4 hex digits separated by a colon.



Abbreviation

- Hexa decimal format of IPv6 is very long and many of the digits are zeros.
- We can abbreviate this address as the leading zeros of a section (four digits between two colons) can be omitted.

Note: Only leading zeros are omitted not
trailing zeros. Example:



- In the above example: 0074 written as 74, 0000 written as 0.
- If there are consecutive sections consisting of zeros only. We can remove the zeros altogether and replace them with a double semicolon.

Address Space

- IPv6 has 2^{128} addresses are available. It is a much larger address space than IPv4.