UNIT-3
NETWORK LAYER

Logical Addressing

Communication at the network layer is host-to-host (computer-to-computer); a computer somewhere in the world needs to communicate with another computer somewhere else in the world. Usually, computers communicate through the Internet. The packet transmitted by the sending computer may pass through several LANs or WANs before reaching the destination computer. For this level of communication, we need a global addressing scheme; we called this logical addressing or IP address.

IPv4 ADDRESSES

An **IPv4** address is a **32-bit** address that uniquely and universally defines the connection of a device (for example, a computer or a router) to the Internet.

- **IPv4 addresses are unique.** They are unique in the sense that each address defines one, and only one, connection to the Internet. Two devices on the Internet can never have the same address at the same time. But by using some strategies, an address may be assigned to a device for a time period and then taken away and assigned to another device.
- On the other hand, if a device operating at the network layer has ***m*** connections to the Internet, it needs to have ***m*** addresses. A router is such a device which needs as many IP addresses as the number of ports are there in it.

Address Space

A protocol such as IPv4 that defines addresses has an 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

This means that, theoretically, if there were no restrictions, more than 4 billion devices could be connected to the Internet. But the actual number is much less because of the restrictions imposed on the addresses.

IPv4 Address Notations

There are two prevalent notations to show an IPv4 address:

a. Binary notation and

b. Dotted decimal notation.

a. Binary Notation

In binary notation, the IPv4 address is displayed as 32 bits. Each octet is often referred to as a byte. So it is common to hear an IPv4 address referred to as a 32-bit address or a 4-byte address. The following is an example of an IPv4 address in binary notation:

01110101 10010101 00011101 00000010

b. Dotted-Decimal Notation

To make the IPv4 address more compact and easier to read, Internet addresses are usually written in decimal form with a decimal point (dot) separating the bytes. The following is the dotted decimal notation of the above address:

117.149.29.2

Figure 19.1 shows an IPv4 address in both binary and dotted-decimal notation. Note that because each byte (octet) is 8 bits, each number in dotted-decimal notation is a value ranging from 0 to 255.

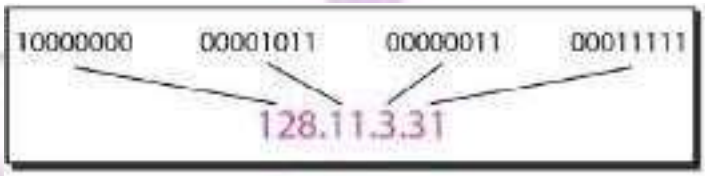


Figure 19.1 Dotted-decimal notation and binary notation for an IPv4 address

Example 19.1

Change the following IPv4 addresses from binary notation to dotted-decimal notation.

a. 10000001 00001011 00001011 11101111

b. 11000001 10000011 00011011 11111111

Solution

We replace each group of 8 bits with its equivalent decimal number (see Appendix B) andadd dots for separation.

a. 129.11.11.239

b. 193.131.27.255

Example 19.2

Change the following IPv4 addresses from dotted-decimal notation to binary notation.

a. 111.56.45.78

b. 221.34.7.82

Solution

We replace each decimal number with its binary equivalent.

a. 01101111 00111000 00101101 01001110

b. 11011101 00100010 00000111 01010010

Types of IPv4 Addressing Schemes

There are two types of IPv4 addressing schemes:

- **Classful Addressing**
- **Classless Addressing**

Classful Addressing

IPv4 addressing, at its inception, used the concept of classes. This architecture is called classful addressing. Although this scheme is becoming **obsolete**, we briefly discuss it here to show the rationale behind classless addressing.

- In classful addressing, the address space is divided into **five classes**: A, B, C, D, and E.
- Each class occupies some part of the address space.
- We can find the class of an address when given the address in binary notation or dotted-decimal notation.
- If the address is given in binary notation, the first few bits can immediately tell us theclass

of the address.

- If the address is given in decimal-dotted notation, the first byte defines the class.Both methods are shown in Figure 19.2.

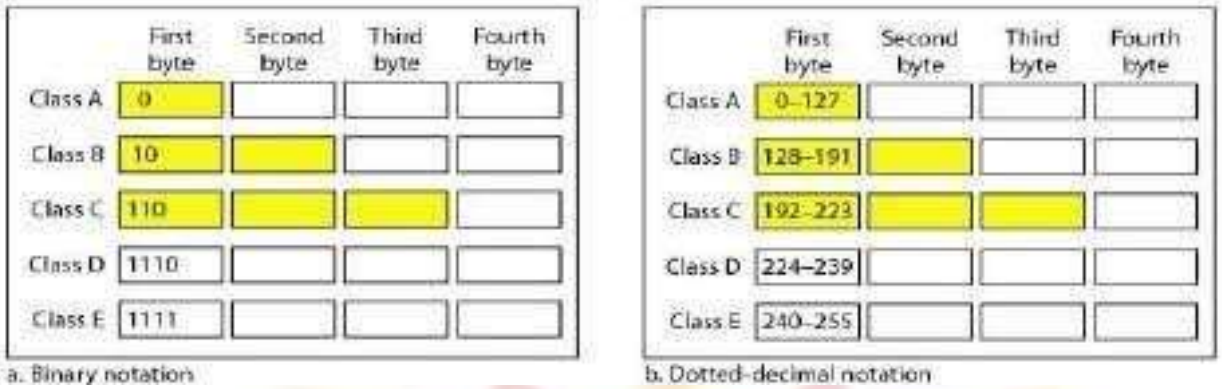


Figure 19.2 Finding the classes in binary and dotted-decimal notation

Example 19.4

Find the class of each address.

a. 00000001 00001011 00001011 11101111

b. 11000001 10000011 00011011 11111111

c. 14.23.120.8

d. 252.5.15.111

Solution

a. The first bit is 0. This is a class A address.

b. The first 2 bits are 1; the third bit is 0. This is a class C address.

c. The first byte is 14 (between 0 and 127); the class is A.

d. The first byte is 252 (between 240 and 255); the class is E.

Classes and Blocks

One problem with classful addressing is that each class is divided into a fixed number of blocks with each block having a fixed size as shown in Table 19.1.

Table 19.1 Number of blocks and block size in Classful IPv4 addressing

Class	Number of Blocks	Block Size	Application
A	2^7=128	2^{24}=16,777,216	Unicast

B	2^{14}=16,384	2^{16}=65,536	Unicast
C	2^{21}=2,097,152	2^{8}=256	Unicast
D	1	2^{28}=268,435,456	Multicast
E	1	2^{28}=268,435,456	Reserved

- **Class A** addresses were designed for **large organizations** with a large number ofattached hosts or routers.
- **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 ofattached hosts or routers.

Limitations of Classful Addressing:

- A block in **class A** address is **too large** for almost any organization. This means mostof the addresses in class A were wasted and were not used.
- A block in **class B** is also **very large**, probably too large for many of the organizations that received a class B block.
- A block in **class C** is probably **too small** for many organizations.
- **Class D** addresses were designed for multicasting. Each address in this class is used to define one group of hosts on the Internet. The Internet authorities wrongly predicted a need for 268,435,456 groups. This never happened and many addresses were wasted here too.
- And lastly, the **class E** addresses were **reserved** for future use; only a few were used, resulting in another waste of addresses.

Netid and Hostid

- In classful addressing, an IP address in class A, B, or C is divided into **netid** and **hostid**.
- These parts are of varying lengths, depending on the class of the address. Figure 19.2shows some netid and hostid bytes.
- The netid is in color, the hostid is in white. Note that the concept does not apply to classes D and E.
- In **class A**, one byte defines the netid and three bytes define the hostid.

- In **class B**, two bytes define the netid and two bytes define the hostid.
- In **class C**, three bytes define the netid and one byte defines the hostid.Table

19.2 Default masks for classful addressing

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

Mask

A mask (also called the default mask) is a 32-bit number made of contiguous 1s followed by contiguous 0s. The masks for classes A, B, and C are shown in Table 19.2. The concept does not apply to classes D and E.

- 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 last column of Table 19.2 shows the mask in the form /*n* where *n* can be 8, 16, or 24 in classful addressing.
- This notation is also called **slash notation** or **Classless Interdomain Routing (CIDR) notation**.

Address Depletion Problem

The fast growth of the Internet led to the near **depletion of the available addresses** in classful addressing scheme. Yet the number of devices on the Internet is much less than the 2^{32} address space. 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.
- **Classful addressing**, which is almost **obsolete**, is **replaced with classless addressing**.

Classless Addressing

To overcome address depletion and give more organizations access to the Internet, classless

addressing was designed and implemented. In this scheme, there are no classes, but the addresses are still granted in blocks.

Address Blocks

- In classless addressing, when an entity, small or large, needs to be connected to the Internet, it is granted a block (range) of addresses.
- The size of the block (the number of addresses) varies based on the nature and size of the entity. For example, a household may be given only two addresses; a large organization may be given thousands of addresses. An ISP, as the Internet service provider, may be given thousands or hundreds of thousands based on the number ofcustomers it may serve.
- The Internet authorities impose **three restrictions on classless address blocks**:

> **1.** The addresses in a *block must be contiguous*, one after another.
>
> **2.** The *number of addresses* in a block must be a *power of 2* (1, 2, 4, 8, ...).
>
> **3.** The *first address* must be *evenly divisible by the number of addresses*.

Example 19.5

Figure 19.3 shows a block of addresses, in both binary and dotted-decimal notation, granted to a small business that needs 16 addresses. We can see that the restrictions are applied to this block. The addresses are contiguous. The number of addresses is a power of 2 ($16 = 2^4$), and 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 in 215,024,210.

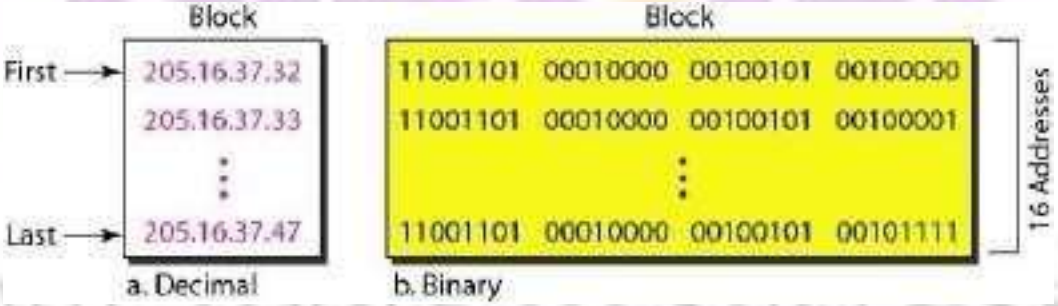


Figure 19.3 A block of16 addresses granted to a small organization

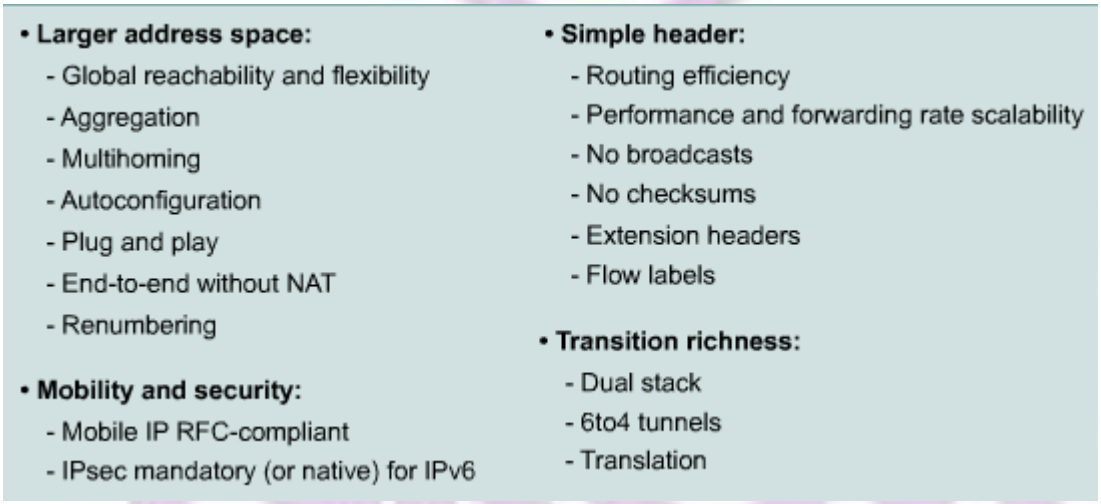
Mask

A better way to define a block of addresses is to select any address in the block and the mask. As we discussed before, a mask is a 32-bit number in which the *n* leftmost bits are 1s and the 32 - *n* rightmost bits are 0s.

- However, in classless addressing the mask for a block can take any value from 0 to 32. It is very convenient to give just the value of *n* preceded by a slash (CIDRnotation).
- 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.
- The address and the /n notation completely define the whole block (the firstaddress, the last address, and the number of addresses).

First Address: The first address in the block can be found by *setting the 32 - n rightmost bits in the binary notation of the address to 0s.*

Features of IPV6:



Larger address space Offers improved global reachability and flexibility; the aggregation of prefixes that are announced in routing tables; multihoming to several Internet service providers (ISPs) auto configuration that can include link-layer addresses in the address space; plug-and-play options; public-toprivate readdressing end to end without address translation; and simplified mechanisms for address renumbering and modification.

Simpler header: Provides better routing efficiency; no broadcasts and thus no potential threat of broadcast storms; no requirement for processing checksums; simpler and more efficient extension header

mechanisms; and flow labels for per-flow processing with no need to open the transport inner packet to identify the various traffic flows.

- **Mobility and security:** Ensures compliance with mobile IP and IPsec standards functionality; mobilityis built in, so any IPv6 node can use it when necessary; and enables people to move around in networks with mobile network devices—with many having wireless connectivity.

  Mobile IP is an Internet Engineering Task Force (IETF) standard available for both IPv4 and IPv6.The standard enables mobile devices to move without breaks in established network connections. Because IPv4 does not automatically provide this kind of mobility, you must add it with additional configurations. IPsec is the IETF standard for IP network security, available for both IPv4 and IPv6. Although the functionalities are essentially identical in both environments, IPsec is mandatory in IPv6. IPsec is enabled on every IPv6 node and is available for use. The availability of IPsec on all nodes makes the IPv6 Internet more secure. IPsec also requires keys for each party, which implies a global key deployment and distribution.

- **Transition richness:** You can incorporate existing IPv4 capabilities in IPv6 in the following ways:
  - Configure a dual stack with both IPv4 and IPv6 on the interface of a network device.
  - Use the technique IPv6 over IPv4 (also called 6to4 tunneling), which uses an IPv4 tunnel to carry IPv6 traffic. This method (RFC 3056) replaces IPv4-compatible tunneling (RFC 2893). Cisco IOS Software Release 12.3(2)T (and later) also allows protocol translation (NAT-PT) between IPv6 and IPv4. This translation allows direct communication between hosts speakingdifferent protocols.

IPv4 VS IPv6

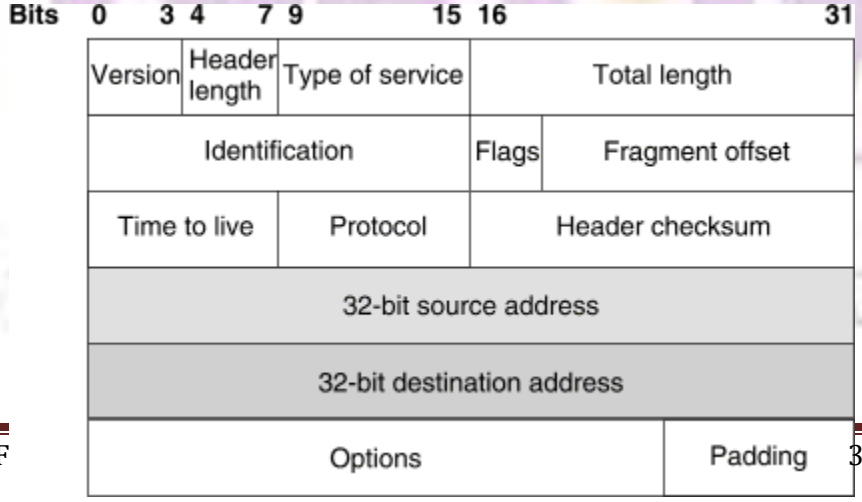


Fig: IPV4 Header

An IPv4 header contains the following fields:
version The IP version number, 4.

lengthThe length of the datagram header in 32-bit words.
type of service Contains five subfields that specify the precedence, delay, throughput, reliability, and cost desiredfor a packet. (The Internet does not guarantee this request.) This field is not widely used on the Internet.
total length The length of the datagram in bytes including the header, options, and the appended transport protocol segment or packet.
Identification An integer that identifies the datagram.

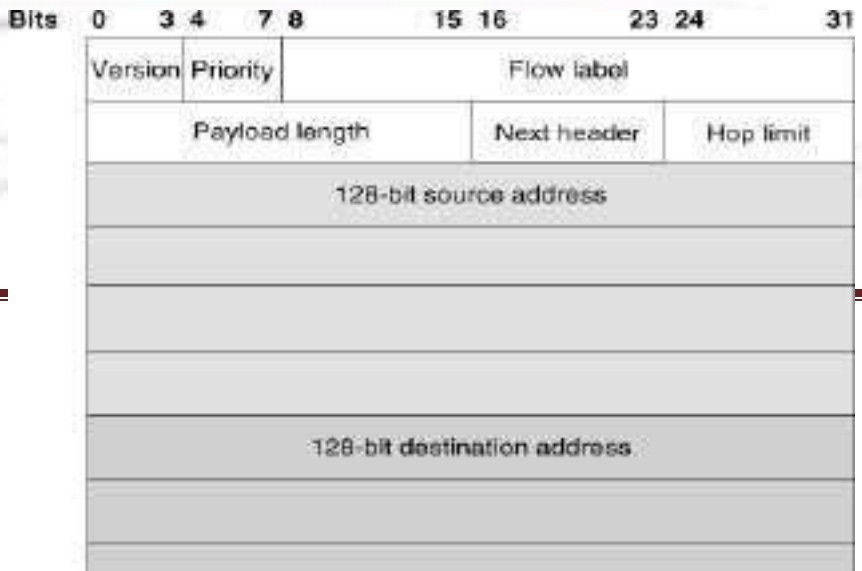
Flags: Controls datagram fragmentation together with the identification field. The flags indicate whether the datagram may be fragmented, whether the datagram is fragmented, and whether the current fragment is the finalone.
fragment offset The relative position of this fragment measured from the beginning of the original datagram inunits of 8 bytes.
time to live How many routers a datagram can pass through. Each router decrements this value by 1 until itreaches 0 when the datagram is discarded. This keeps misrouted datagrams from remaining on the Internet forever.
Protocol The high-level protocol type.

header checksum A number that is computed to ensure the integrity of the header values.
source address The 32-bit IPv4 address of the sending host.

destination address The 32-bit IPv4 address of the receiving host.

Options A list of optional specifications for security restrictions, route recording, and source routing. Not everydatagram specifies an options field.
Padding Null bytes which are added to make the header length an integral multiple of 32 bytes as required bythe header length field.

Ipv6 header:



Specifically, IPv6 omits the following fields in its header.

- header length (the length is constant)
- identification
- flags
- fragment offset (this is moved into fragmentation extension headers)
- header checksum (the upper-layer protocol or security extension header handles data integrity)

IPv6 options improve over IPv4 by being placed in separate extension headers that are located between the IPv6 header and the transport-layer header in a packet. Most extension headers are not examined or processed by any router along a packet's delivery path until it arrives at its final destination. This mechanism improves router performance for packets containing options. In IPv4, the presence of any options requires the router to examine all options.

Another improvement is that IPv6 extension headers, unlike IPv4 options, can be of arbitrary length and the total amount of options that a packet carries is not limited to 40 bytes. This feature, and the manner in which it is processed, permit IPv6 options to be used for functions that were not practical in IPv4, such as the IPv6 Authentication and Security Encapsulation options.

By using extension headers, instead of a protocol specifier and options fields, newly defined extensions can be integrated more easily into IPv6.

IPV6 Addressing:

Address Representation:

Represented by breaking 128 bit into Eight 16-bit segments (Each 4 Hex character

each)Each segment is written in Hexadecimal separated by colons.
Hex digit are not case sensitive.

Rule 1:

Drop leading zeros:
2001:0050:0000:0235:0ab4:3456:456b:e560
2001:050:0:235:ab4:3456:456b:e560

Rule2:

Successive fields of zeros can be represented as "::" , But double colon appear only once in the address.
FF01:0:0:0:0:0:0:1
FF01::1

Note : An address parser identifies the number of missing zeros by separating the two parts and entering 0 until the 128 bits are complete. If two "::" notations are placed in the address, there is no way to identify the size of each block of zeros.

Ipv4 vs ipv6

IPV4	IPV6
1. source and destination addresses are 32 bits.)	1. Source and destination addresses are 128 bits.
2. ipv4 support small address space.	2. Supports a very large address space sufficeint foreach and every people on earth.
3. ipv4 header includes checksum.	3. ipv6 header doesn't includes the checksum. (theupper-layer protocol or security extension header handles data integrity)
4. addresses are represented in dotted decimal format.(Eg. 192.168.5.1)	4. Addresses are represented in 16-bit segments Each segment is written in Hexadecimal separated bycolons. (Eg. 2001:0050:020c:0235:0ab4:3456:456b:e560
5. Header includes options.	All optional data is moved to IPV6 extension header..
6. Broadcast address are used to send traffic to all	6. There is no IPV6 broadcast address. Instead a link
nodes on a subnet.	local scope all-nodes multicast address is used.
7. No identification of packet flow for QOS handlingby router is present within the ipv4 header.	7. Packet flow identification for QOS handling by routers is present within the IPV6 header using theflow label field.

8. uses host address (A) resource records in the Domain name system(DNS) to map host names toipv4 addresses.	8. Uses AAAA records in the DNS to map host namesto ipv6 addresses.
9. Both routers and the sending host fragmentpackets.	9. Only the sending host fragments packets; routersdo not.
10. ICMP Router Discovery is used to determine theIPv4 address of the best default gateway, and it is optional.	10. ICMPv6 Router Solicitation and Router Advertisement messages are used to determine the IPaddress of the best default gateway, and they are required.

ADDRESS MAPPING

ADDRESS MAPPING An 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, they are 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.

Static mapping :Static mapping involves in the creation of a table that associates a logical address with a physical address. This table is stored in each machine on the network. Each machine that knows, for example, the IP address of another machine but not its physical address can look it up in the table. Static mapping has some limitations because physical addresses may change in the following ways:

1. A machine could change its NIC (Network Interface Card), resulting in a new physical address.

2. In some LANs, such as LocalTalk, the physical address changes every time the computer is turned on.

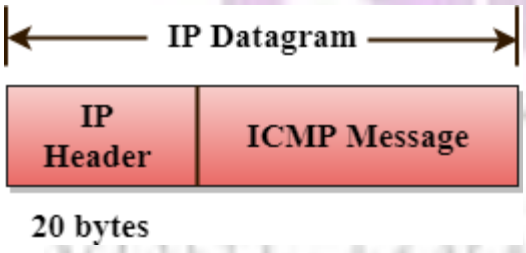
3. A mobile computer can move from one physical network to another, resulting in a change in its physical address.

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

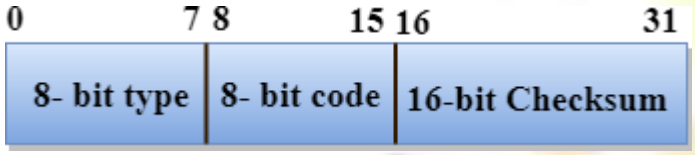
Dynamic mapping: In such mapping each time a machine knows one of the two addresses (logical or physical), it can use a protocol to find the other one.

ICMP

- ICMP stands for Internet Control Message Protocol.
- The ICMP is a network layer protocol used by hosts and routers to send the notifications of IP datagram problems back to the sender.ICMP uses echo test/reply to check whether the destination is reachable and responding.
- ICMP handles both control and error messages, but its main function is to report the error but not to correct them.
- An IP datagram contains the addresses of both source and destination, but it does not know the address of the previous router through which it has been passed. Due to this reason, ICMP can only send the messages to the source, but not to the immediate routers.
- ICMP protocol communicates the error messages to the sender. ICMP messages cause the errors to be returned back to the user processes.
- ICMP messages are transmitted within IP datagram.



The Format of an ICMP message



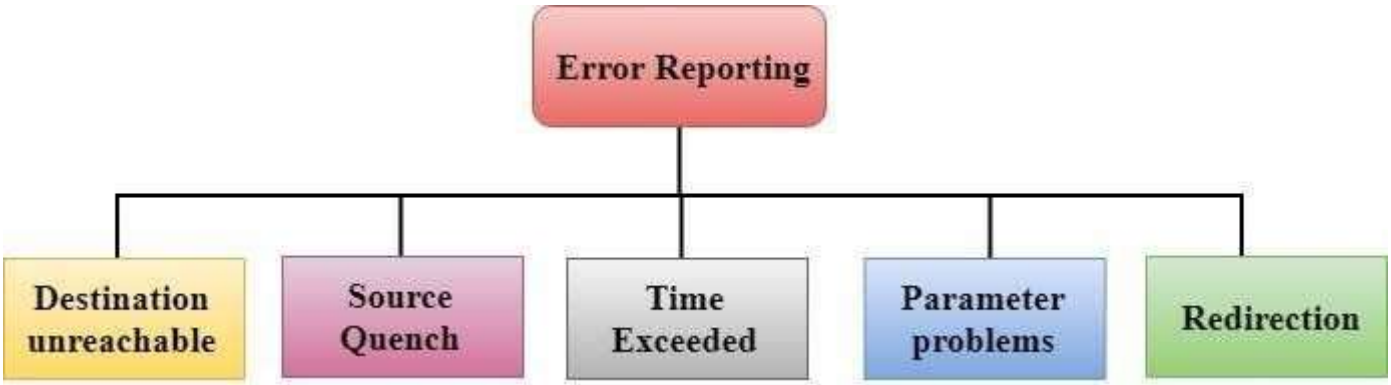
- The first field specifies the type of the message.
- The second field specifies the reason for a particular message type.
- The checksum field covers the entire ICMP message.

Error Reporting

ICMP protocol reports the error messages to the sender.

Five types of errors are handled by the ICMP protocol:

- Destination unreachable
- Source Quench
- Time Exceeded
- Parameter problems
- Redirection



- **Destination unreachable:** The message of "Destination Unreachable" is sent from receiver to the sender when destination cannot be reached, or packet is discarded when the destination is not reachable.
- **Source Quench:** The purpose of the source quench message is congestion control. The message sent from the congested router to the source host to reduce the transmission rate. ICMP will take the IP of the discarded packet and then add the source quench message to the IP datagram to inform the source host to reduce its transmission rate. The source host will reduce the transmission rate so that the router will be free from congestion.
- **Time Exceeded:** Time Exceeded is also known as "Time-To-Live". It is a parameter that defines how long a packet should live before it would be discarded.

There are two ways when Time Exceeded message can be generated:

Sometimes packet discarded due to some bad routing implementation, and this causes the looping issue and network congestion. Due to the looping issue, the value of TTL keeps on decrementing, and when it reaches zero, the router discards the datagram. However, when the datagram is discarded by the router, the time exceeded message will be sent by the router to the source host.

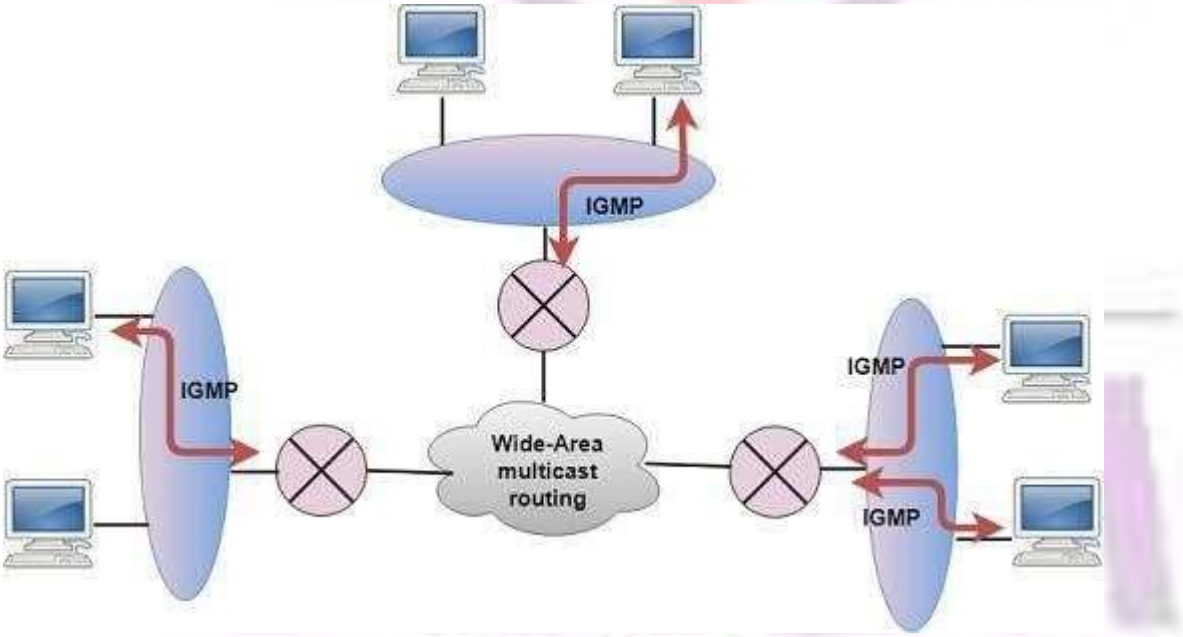
When destination host does not receive all the fragments in a certain time limit, then the received fragments are also discarded, and the destination host sends time Exceeded message to the source host.

- **Parameter problems:** When a router or host discovers any missing value in the IP datagram, the router discards the datagram, and the "parameter problem" message is sent back to the source host.
- **Redirection:** Redirection message is generated when host consists of a small routing table. When the

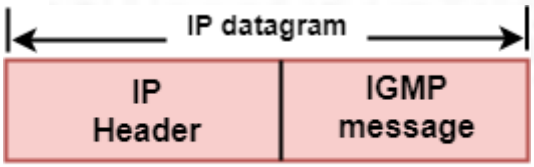
host consists of a limited number of entries due to which it sends the datagram to a wrong router. The router that receives a datagram will forward a datagram to a correct router and also sends the "Redirection message" to the host to update its routing table.

IGMP

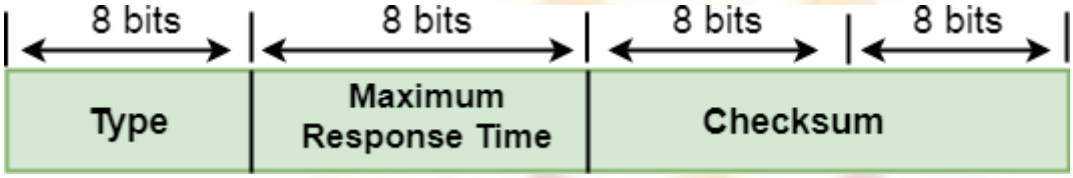
- IGMP stands for **Internet Group Message Protocol**.
- The IP protocol supports two types of communication:
  - **Unicasting:** It is a communication between one sender and one receiver. Therefore, we can say that it is one-to-one communication.
  - **Multicasting:** Sometimes the sender wants to send the same message to a large number of receivers simultaneously. This process is known as multicasting which has one-to-many communication.
- The IGMP protocol is used by the hosts and router to support multicasting.
- The IGMP protocol is used by the hosts and router to identify the hosts in a LAN that are the members of a group.



- IGMP is a part of the IP layer, and IGMP has a fixed-size message.
- The IGMP message is encapsulated within an IP datagram.



The Format of IGMP message



Where,

Type: It determines the type of IGMP message. There are three types of IGMP message: Membership Query, Membership Report and Leave Report.

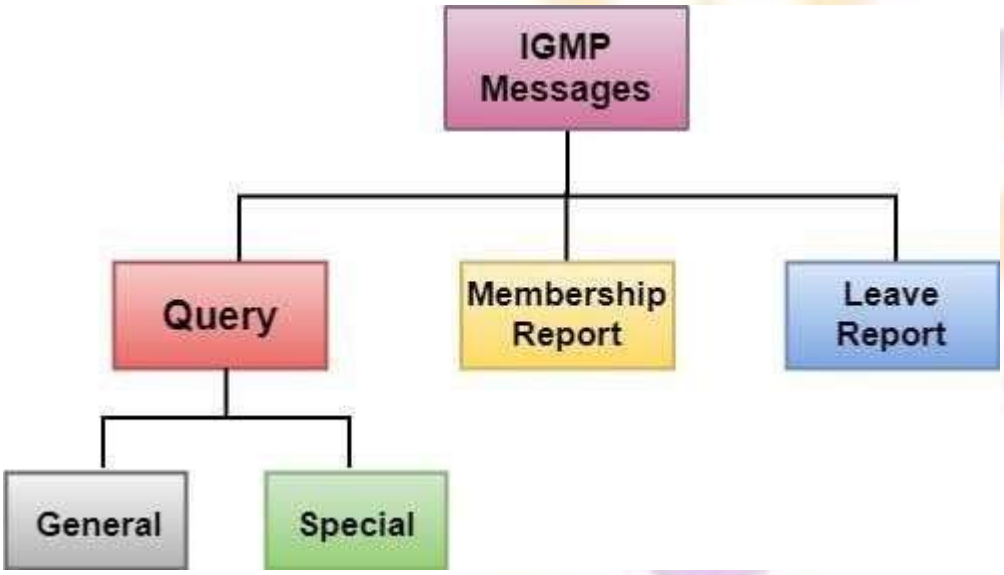
Maximum Response Time: This field is used only by the Membership Query message. It determines the maximum time the host can send the Membership Report message in response to the Membership Query message.

Checksum: It determines the entire payload of the IP datagram in which IGMP message is encapsulated.

Group Address: The behavior of this field depends on the type of the message sent.

- **For Membership Query**, the group address is set to zero for General Query and set to multicast group address for a specific query.
- **For Membership Report**, the group address is set to the multicast group address.
- **For Leave Group**, it is set to the multicast group address.

IGMP Messages



- **Membership Query message**
  - This message is sent by a router to all hosts on a local area network to determine the set of all the multicast groups that have been joined by the host.

- It also determines whether a specific multicast group has been joined by the hosts on a attached interface.
- The group address in the query is zero since the router expects one response from a host for every group that contains one or more members on that host.

- **Membership Report message**
  - The host responds to the membership query message with a membership report message.
  - Membership report messages can also be generated by the host when a host wants to join the multicast group without waiting for a membership query message from the router.
  - Membership report messages are received by a router as well as all the hosts on an attached interface.
  - Each membership report message includes the multicast address of a single group that the host wants to join.
  - IGMP protocol does not care which host has joined the group or how many hosts are present in a single group. It only cares whether one or more attached hosts belong to a single multicast group.
  - The membership Query message sent by a router also includes a "**Maximum Response time**". After receiving a membership query message and before sending the membership report message, the host waits for the random amount of time from 0 to the maximum response time. If a host observes that some other attached host has sent the "**Maximum Report message**", then it discards its "**Maximum Report message**" as it knows that the attached router already knows that one or more hosts have joined a single multicast group. This process is known as feedback suppression. It provides the performance optimization, thus avoiding the unnecessary transmission of a "**Membership Report message**".

Leave Report

When the host does not send the "Membership Report message", it means that the host has left the group. The host knows that there are no members in the group, so even when it receives the next query, it would not report the group.

Forwarding

Forwarding means to place the packet in its route to its destination. Forwarding requires a host or a router to have a routing table. When a host has a packet to send or when a router has received a packet to be forwarded, it looks at this table to find the route to the final destination. However, this simple solution is impossible today in an internetwork such as the Internet because the number of entries needed in the routing table would make table lookups inefficient.

Forwarding Techniques

Several techniques can make the size of the routing table manageable and also handle issues such as security.

a. Next-Hop Method versus Route Method

One technique to reduce the contents of a routing table is called the next-hop method. In this technique, the routing table holds only the address of the next hop instead of information about the complete route (route method). The entries of a routing table must be consistent with one another.

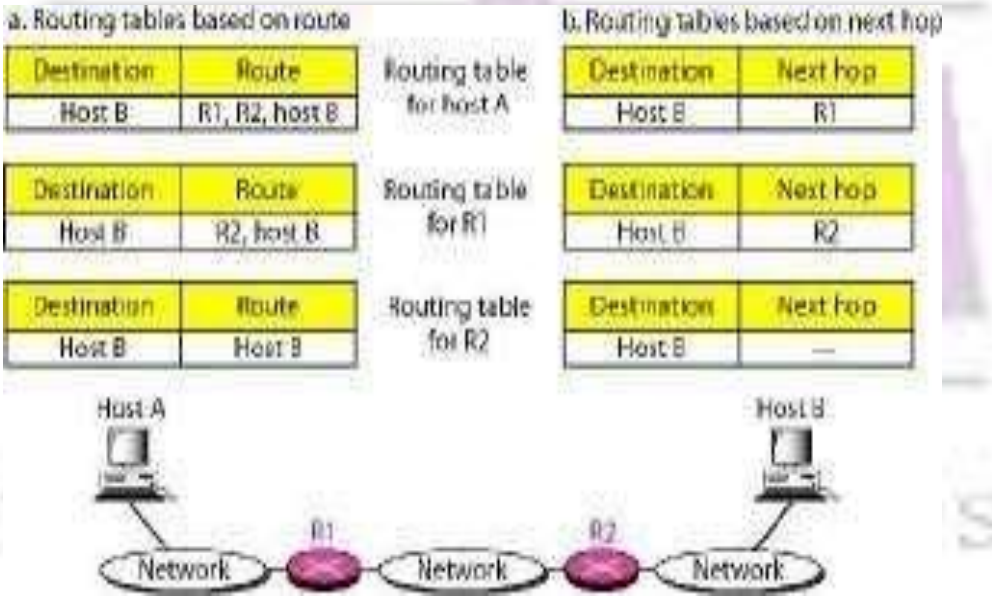


Figure 3.40 Route method versus next-hop method

b. Network-Specific Method versus Host-Specific Method

A second technique to reduce the routing table and simplify the searching process is called the network-specific method. Here, instead of having an entry for every destination host connected to the same physical network (host-specific method), we have only one entry that defines the address of the destination network itself.

Host-specific routing is used for purposes such as checking the route or providing security measures.

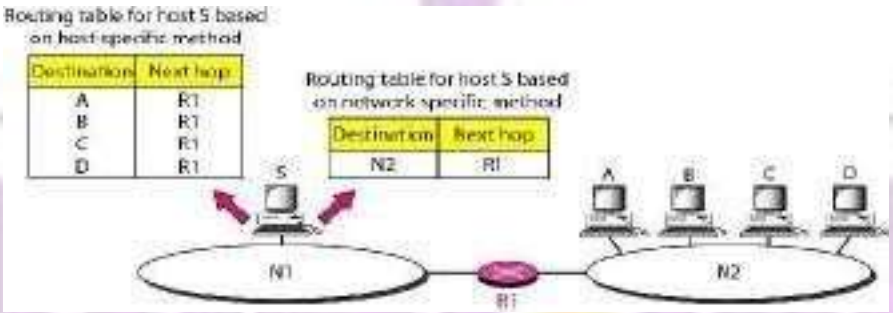


Figure 3.41 Host-specific versus network-specific method

c. Default Method

Another technique to simplify routing is called the default method. Host A is connected to a network with two routers. Router R1 routes the packets to hosts connected to network N2. However, for the rest of the Internet, router R2 is used. So instead of listing all networks in the entire Internet, host A can just have one entry called the default (normally defined as network address 0.0.0.0).

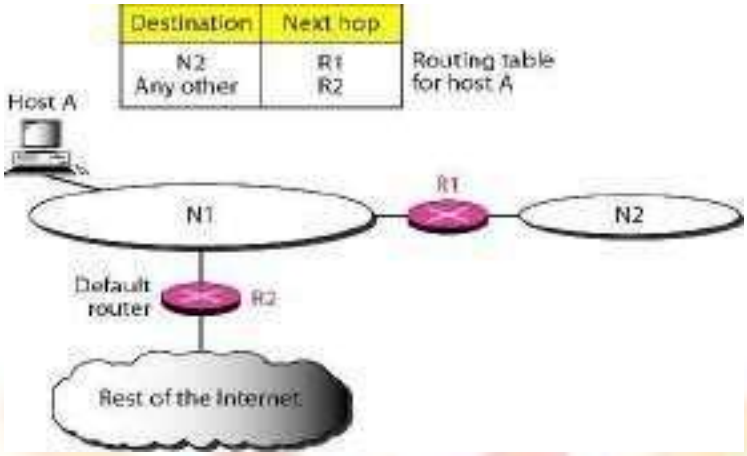


Figure 3.42 Default methodExample 3.18
Make a routing table for router R1, using the configuration in Figure 3.43

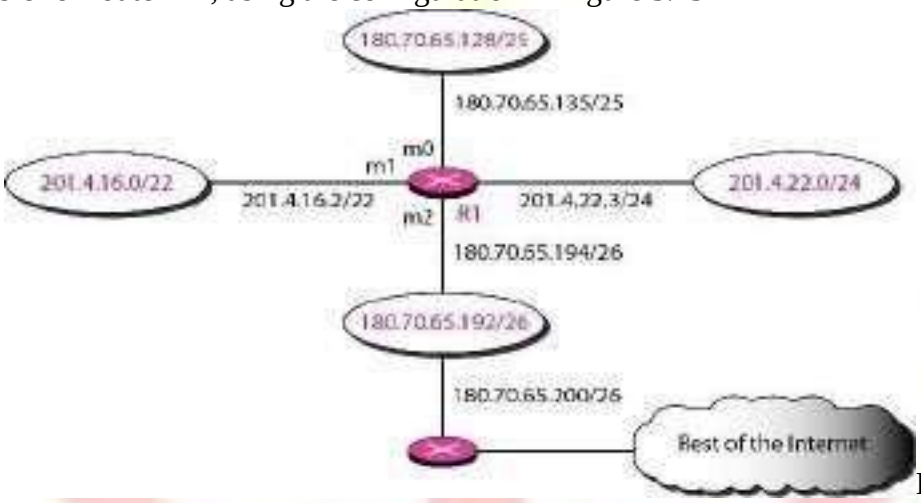


Figure 3.43 Configuration for Example 3.18Solution

Table 3.9 Routing table for router R1 in Figure 3.43

Mask	*Network Address*	*Next Hop*	*Interface*
/26	180.70.65.192	—	m2
/25	180.70.65.128	—	m0
/24	201.4.22.0	—	m3
/22	201.4.16.0	—	m1
Any	Any	180.70.65.200	m2

Unicast Routing Protocols

A routing table can be either static or dynamic. A static table is one with manual entries. A dynamic table, on the other hand, is one that is updated automatically when there is a change somewhere in the internet. Today, an internet needs dynamic routing tables. The tables need to be updated as soon as there is a change in the internet. For instance, they need to be updated when a router is down, and they need to be updated whenever a better route has been found.

Optimization

A router receives a packet from a network and passes it to another network. A router is usually attached to several networks. One approach is to assign a cost for passing through a network. We call this cost a metric. However, the metric assigned to each network depends on the type of protocol. Some simple protocols, such as the Routing Information Protocol (RIP), treat all networks as equals.

The cost of passing through a network is the same; it is one hop count. So if a packet passes through 10 networks to reach the destination, the total cost is 10 hop counts.

Intra- and Inter-domain Routing

An internet can be so large that one routing protocol cannot handle the task of updating the routing tables of all routers. For this reason, an internet is divided into autonomous systems. An autonomous system (AS) is a group of networks and routers under the authority of a single administration. Routing inside an autonomous system is referred to as intradomain routing. Routing between autonomous systems is referred to as interdomain routing.

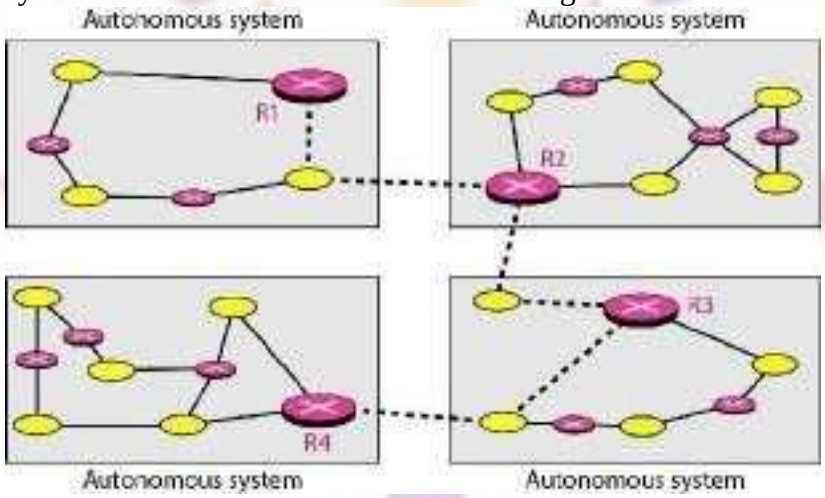


Figure 3.43 Autonomous systems

Several intradomain and interdomain routing protocols are in use.

- Two intradomain routing protocols: Distance vector and link state.
- One interdomain routing protocol: path vector.

Routing Information Protocol (RIP) is an implementation of the distance vector protocol. Open Shortest Path First (OSPF) is an implementation of the link state protocol. Border Gateway Protocol (BGP) is an implementation of the path vector protocol.

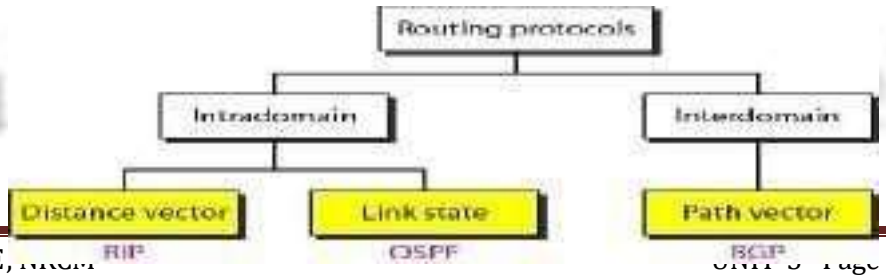


Figure 3.44 Popular routing protocols

Distance Vector Routing

In distance vector routing, the least-cost route between any two nodes is the route with minimum distance. In this protocol, as the name implies, each node maintains a vector (table) of minimum distances to every node. The table at each node also guides the packets to the desired node by showing the next stop in the route (next-hop routing).

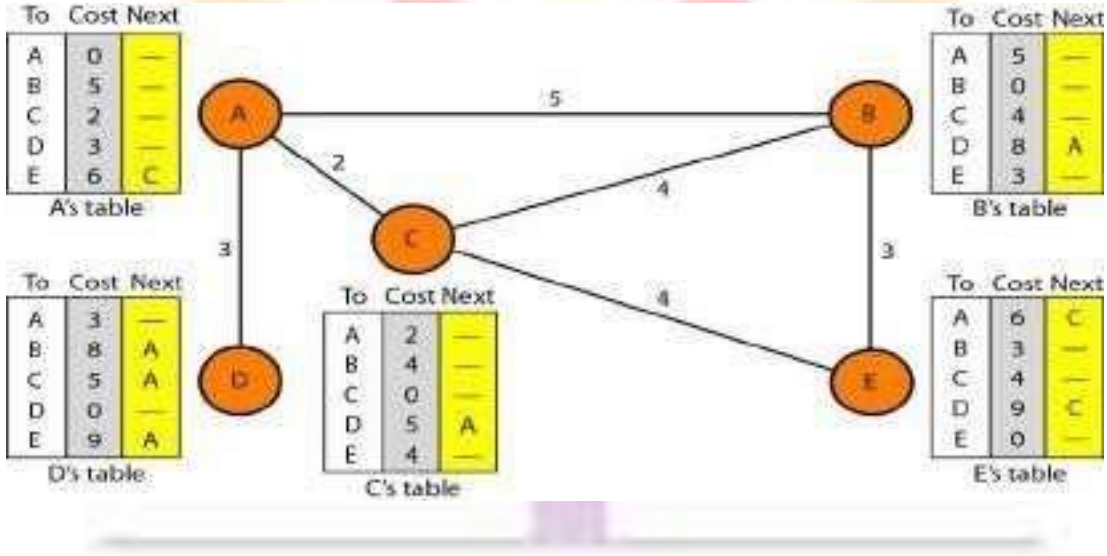


Figure 3.45 Distance vector routing tables

The table for node A shows how we can reach any node from this node. For example, our least cost to reach node E is 6. The route passes through C.

Initialization

The tables in Figure 3.45 are stable; each node knows how to reach any other node and the cost. At the beginning, however, this is not the case. Each node can know only the distance between itself and its immediate neighbors, those directly connected to it. So for the moment, we assume that each node can send a message to the immediate neighbors and find the distance between itself and these neighbors. The distance for any entry that is not a neighbor is marked as infinite (unreachable).

Sharing

The whole idea of distance vector routing is the sharing of information between neighbors. Although node A does not know about node E, node C does. So if node C shares its routing table with A, node A can also know how to reach node E. On the other hand, node C does not know how to reach node D, but node A does. If node A shares its routing table with node C, node C also knows how to reach node D. In other words, nodes A and C, as immediate neighbors, can improve their routing tables if they help each other.

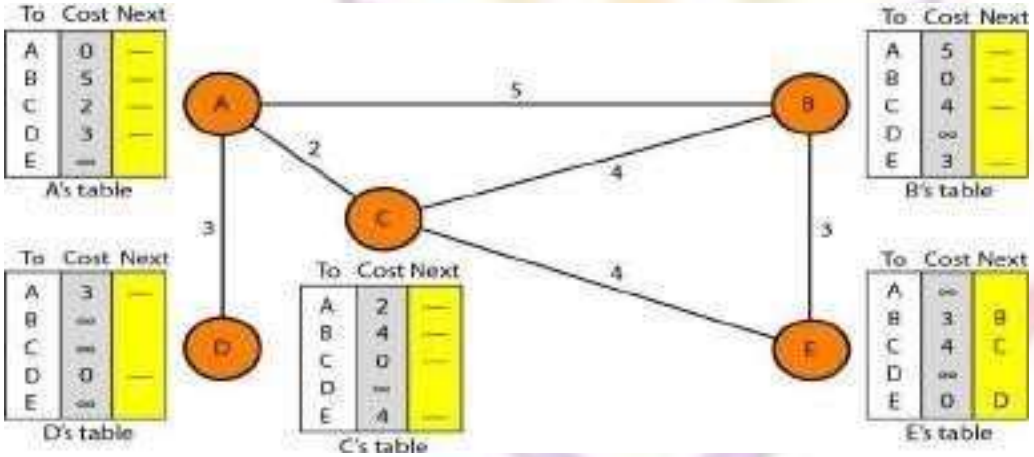


Figure 3.46 Initialization of tables in distance vector routing

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. The logic is clear. If node C claims that its distance to a destination is *x* mi, and the distance between A and C is *y* mi, then the distance between A and that destination, via C, is $x + y$ mi.
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.

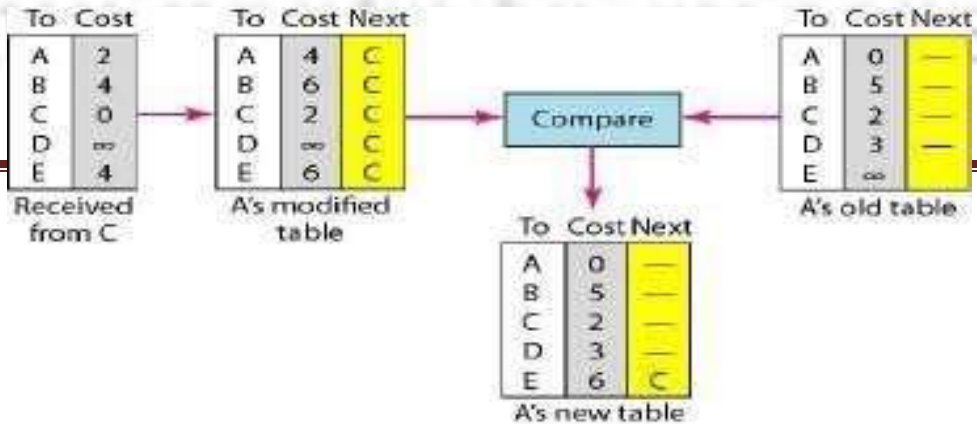


Figure 3.47 Updating in distance vector routingTwo-Node Loop Instability

A problem with distance vector routing is instability, which means that a network using this protocol can become unstable. To understand the problem, let us look at the scenario depicted.

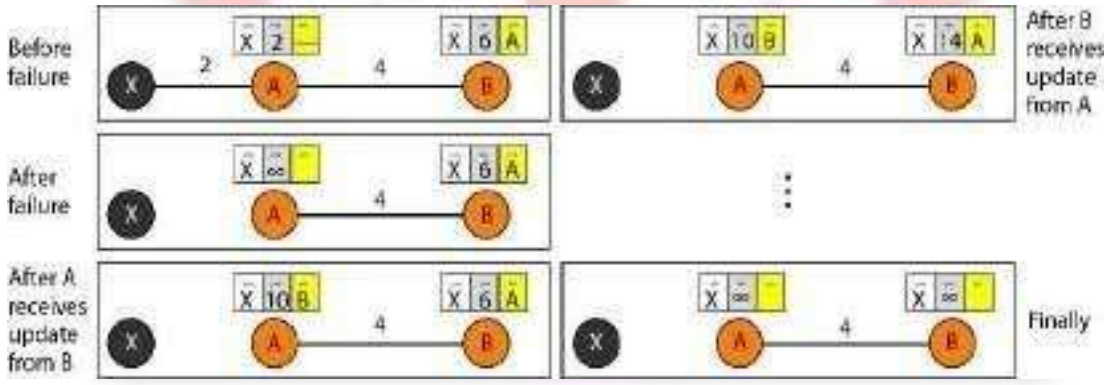


Figure 3.48 Two-node instability

Defining Infinity The first obvious solution is to redefine infinity to a smaller number, such as 100. For our previous scenario, the system will be stable in less than 20 update s. As a matter of fact, most implementations of the distance vector protocol define the distance between each node to be I and define 16 as infinity. However, this means that the distance vector routing cannot be used in large systems. The size of the network, in each direction, cannot exceed 15 hops.

Split Horizon Another solution is called split horizon. In this strategy, instead of flooding the table through each interface, each node sends only part of its table through each interface. If, according to its table, node B thinks that the optimum route to reach X is via A, it does not need to advertise this piece of information to A; the information has corne from A (A already knows). Taking information from node A, modifying it, and sending it back to node A creates the confusion. In our scenario, node B eliminates the last line of its routing table before it sends it to

A. In this case, node A keeps the value of infinity as the distance to X.

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 of LSPs to every other router, called **flooding,** 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.

3.7 Multicast Routing Protocols

Unicast, Multicast, and Broadcast:

A message can be unicast, multicast, or broadcast.

Unicasting

In unicast communication, there is one source and one destination. The relationship between the source and the destination is one-to-one. In this type of communication, both the source and destination addresses, in the IP datagram, are the unicast addresses assigned to the hosts (or host interfaces, to be more exact).

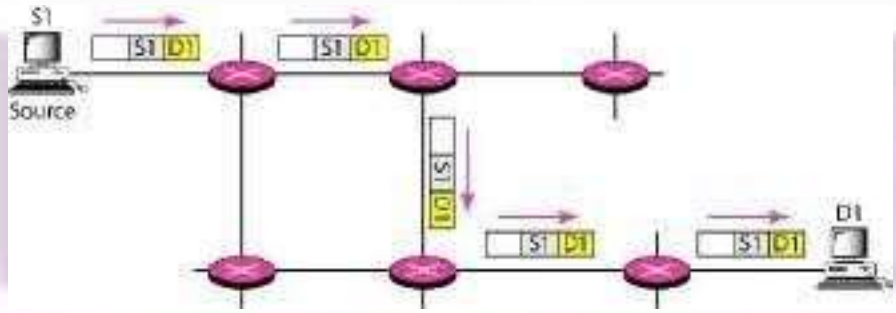


Figure 3.54 Unicasting

Multicasting

In multicast communication, there is one source and a group of destinations. The relationship is one-to-many. In this type of communication, the source address is a unicast address, but the

destination address is a group address, which defines one or more destinations. The group address identifies the members of the group.

A multicast packet starts from the source S1 and goes to all destinations that belong to group G1. In multicasting, when a router receives a packet, it may forward it through several of its interfaces.

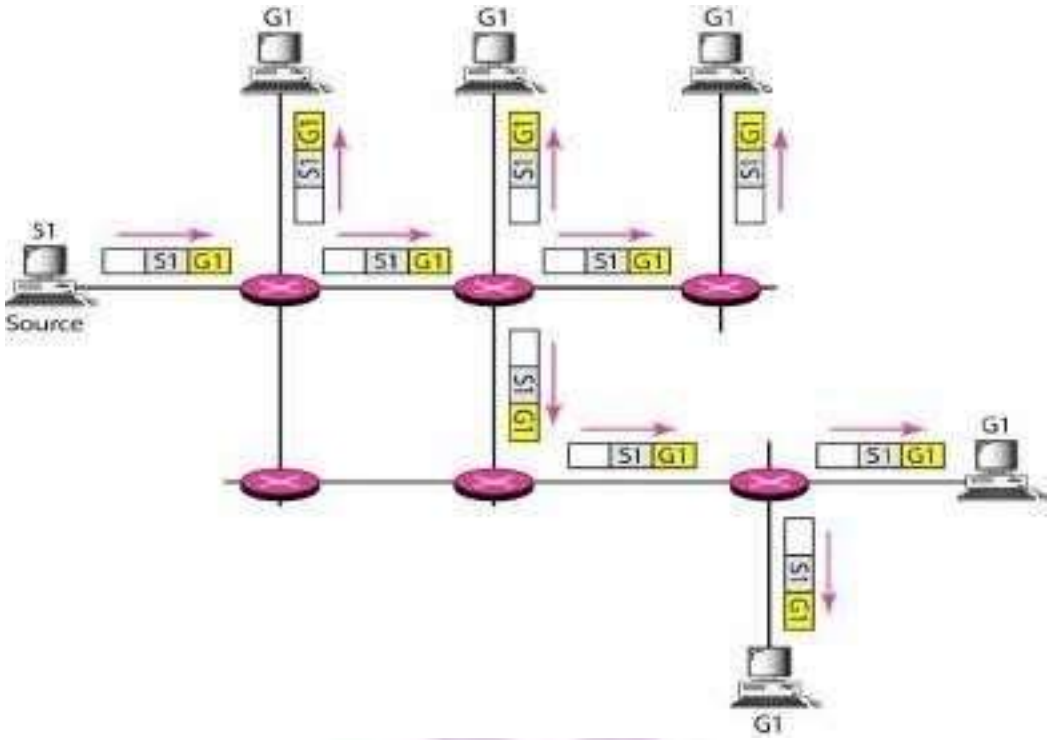


Figure 3.55 Multicasting

Broadcasting

In broadcast communication, the relationship between the source and the destination is one-to-all. There is only one source, but all the other hosts are the destinations. The Internet does not explicitly support broadcasting because of the huge amount of traffic it would create and because of the bandwidth it would need. Imagine the traffic generated in the Internet if one person wanted to send a message to everyone else connected to the Internet.

Multicasting versus Multiple Unicasting

Multicasting starts with one single packet from the source that is duplicated by the routers. The destination address in each packet is the same for all duplicates. Note that only one single copy of the packet travels between any two routers.

In multiple unicasting, several packets start from the source. If there are five destinations, for example, the source sends five packets, each with a different unicast destination address. Note that there may be multiple copies traveling between two routers.

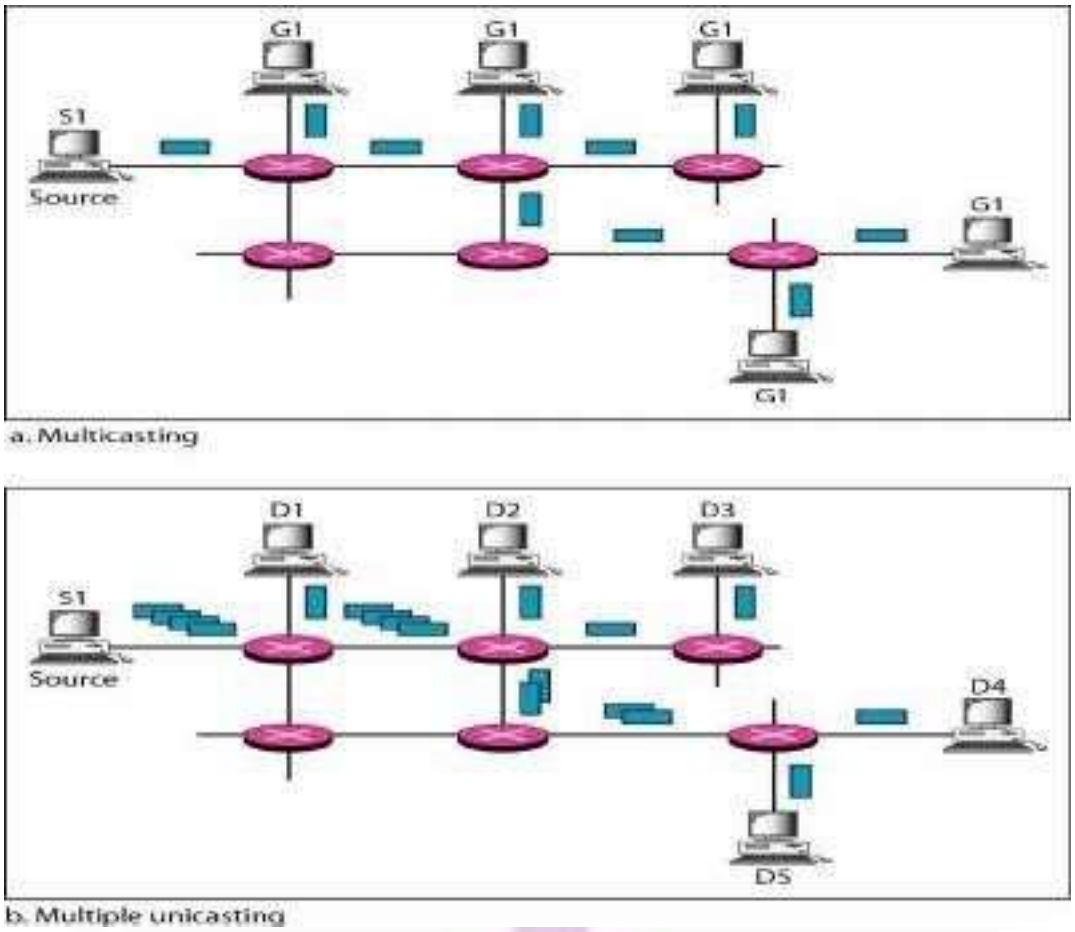


Figure 3.56 Multicasting versus multiple unicasting

Applications

Multicasting has many applications today such as access to distributed databases, information dissemination, teleconferencing, and distance learning.

Unicast, Multicast, and Broadcast:

A message can be unicast, multicast, or broadcast.

Unicasting

In unicast communication, there is one source and one destination. The relationship between

the source and the destination is one-to-one. In this type of communication, both the source and destination addresses, in the IP datagram, are the unicast addresses assigned to the hosts (or host interfaces, to be more exact).

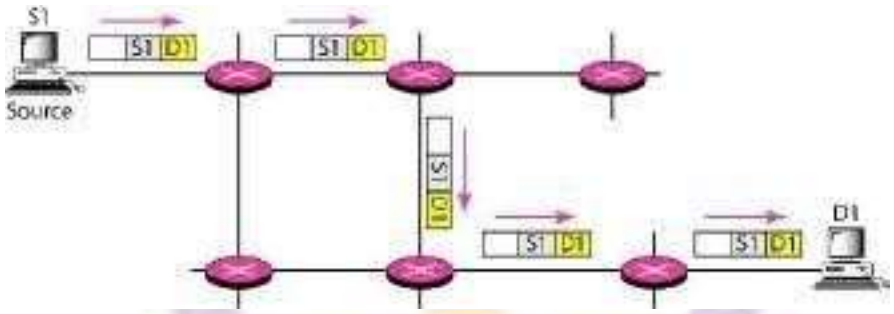


Figure 3.54 Unicasting

Multicasting

In multicast communication, there is one source and a group of destinations. The relationship is one-to-many. In this type of communication, the source address is a unicast address, but the destination address is a group address, which defines one or more destinations. The group address identifies the members of the group.

A multicast packet starts from the source S1 and goes to all destinations that belong to group G1. In multicasting, when a router receives a packet, it may forward it through several of its interfaces.

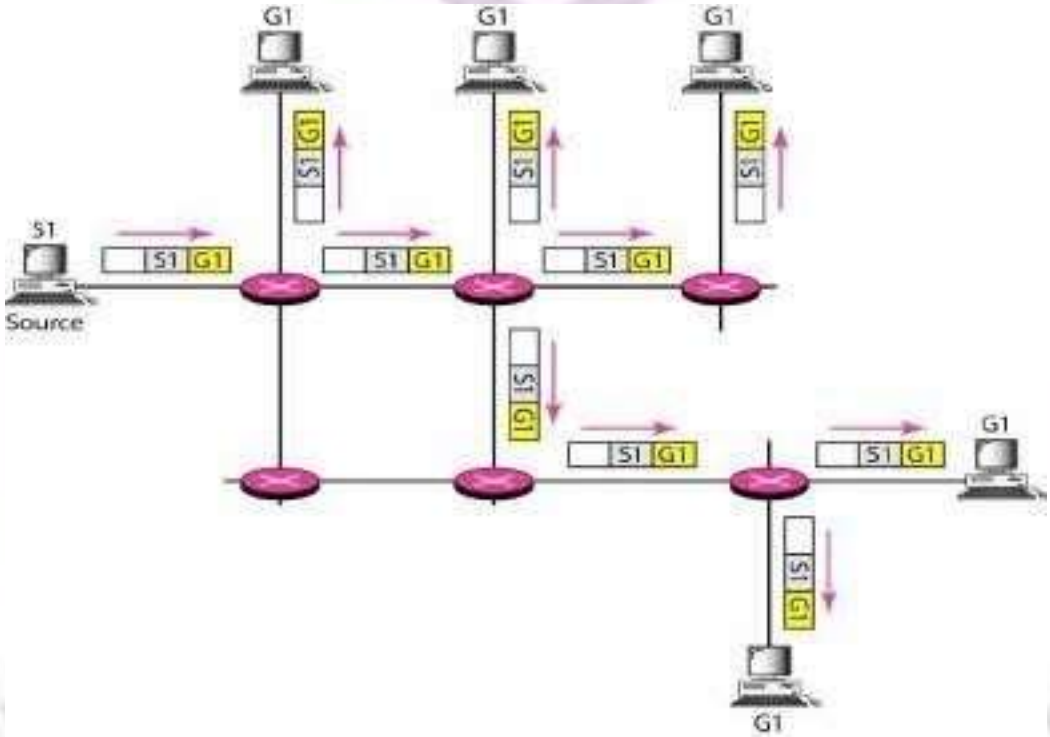


Figure 3.55 Multicasting

Broadcasting

In broadcast communication, the relationship between the source and the destination is one-to-all. There is only one source, but all the other hosts are the destinations. The Internet does not explicitly support broadcasting because of the huge amount of traffic it would create and because of the bandwidth it would need. Imagine the traffic generated in the Internet if one person wanted to send a message to everyone else connected to the Internet.

Multicasting versus Multiple Unicasting

Multicasting starts with one single packet from the source that is duplicated by the routers. The destination address in each packet is the same for all duplicates. Note that only one single copy of the packet travels between any two routers.

In multiple unicasting, several packets start from the source. If there are five destinations, for example, the source sends five packets, each with a different unicast destination address. Note that there may be multiple copies traveling between two routers.

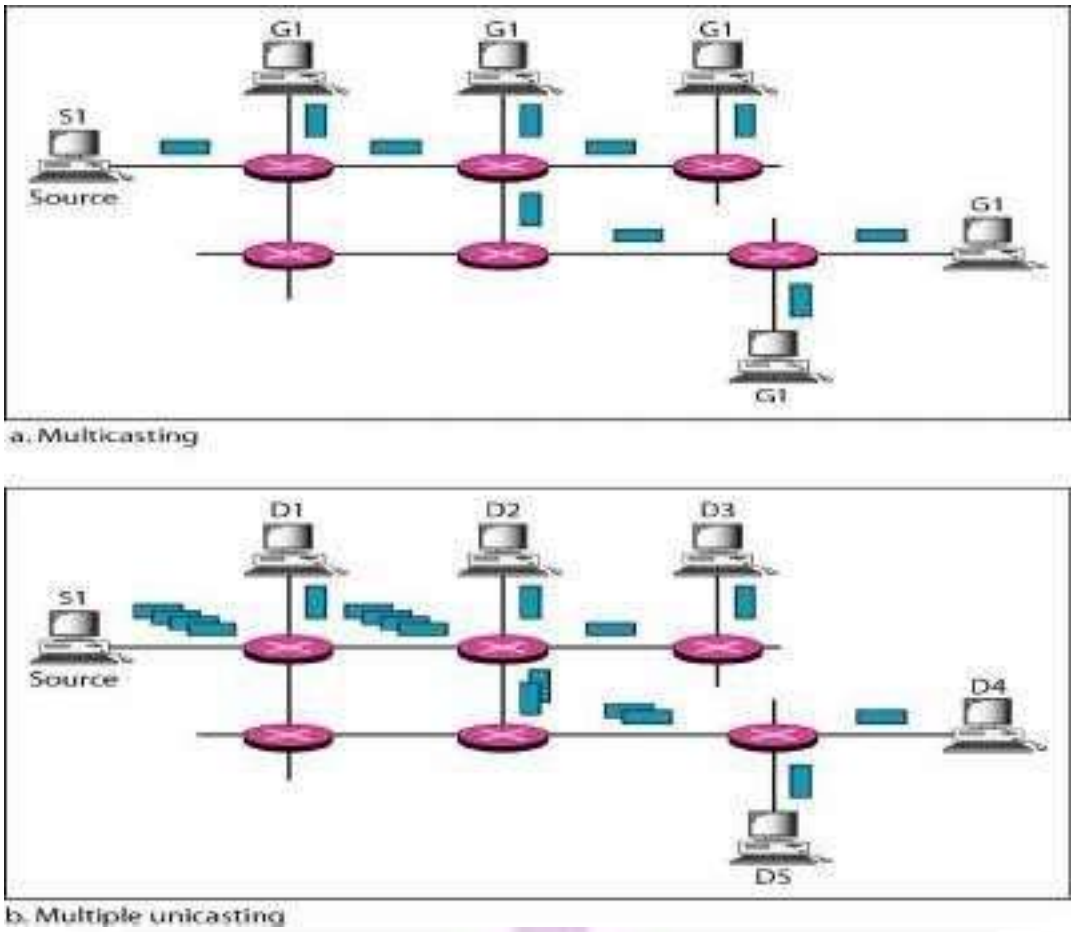


Figure 3.56 Multicasting versus multiple unicasting

Applications

Multicasting has many applications today such as access to distributed databases, information dissemination, teleconferencing, and distance learning.

Access to Distributed Databases

Most of the large databases today are distributed. That is, the information is stored in more than one location, usually at the time of production. The user who needs to access the database does not know the location of the information. A user's request is multicast to all the database locations, and the location that has the information responds.

Information Dissemination

Businesses often need to send information to their customers. If the nature of the information

is the same for each customer, it can be multicast. In this way a business can send one message that can reach many customers. For example, a software update can be sent to all purchasers of a particular software package.

Optimal Routing: Shortest Path Trees

The process of optimal interdomain routing eventually results in the finding of the shortest path tree. The root of the tree is the source, and the leaves are the potential destinations. The path from the root to each destination is the shortest path. However, the number of trees and the formation of the trees in unicast and multicast routing are different.

Unicast Routing: In unicast routing, when a router receives a packet to forward, it needs to find the shortest path to the destination of the packet. The router consults its routing table for that particular destination. The next-hop entry corresponding to the destination is the start of the shortest path. The router knows the shortest path for each destination, which means that the router has a shortest path tree to optimally reach all destinations. In other words, each line of the routing table is a shortest path; the whole routing table is a shortest path tree. In unicast routing, each router needs only one shortest path tree to forward a packet; however, each router has its own shortest path tree.

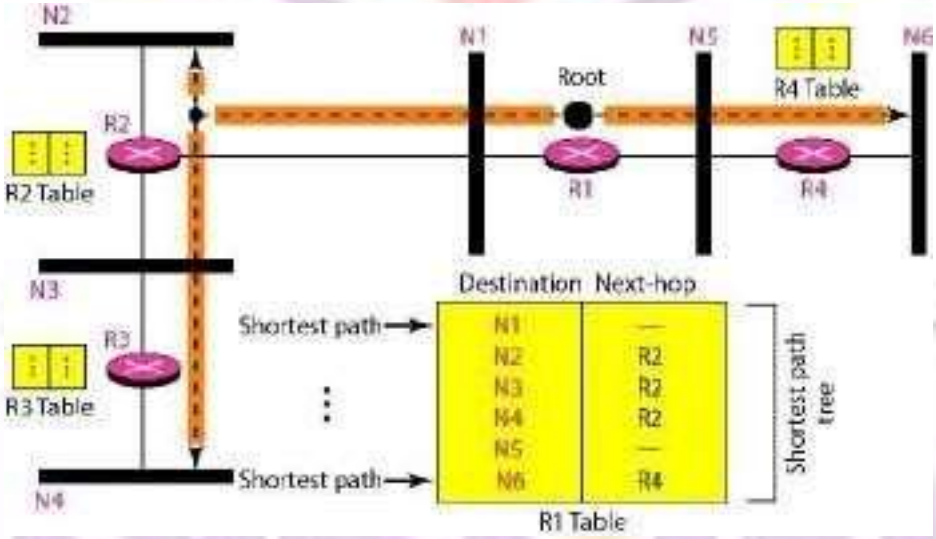


Figure 3.57 Shortest path tree in unicast routing

Multicast Routing: When a router receives a multicast packet, the situation is different from when it receives a unicast packet. A multicast packet may have destinations in more than one network. Forwarding of a single packet to members of a group requires a shortest path tree. If we have n groups, we may need *n* shortest path trees. We can imagine the complexity of multicast routing. Two

approaches have been used to solve the problem: source-based trees and group- shared trees.

a. Source-Based Tree: In the source-based tree approach, each router needs to have one shortest path tree for each group. The shortest path tree for a group defines the next hop for each network that has loyal member(s) for that group. Five groups in the domain: GI, G2, G3, G4, and G5.

At the moment G1 has loyal members in four networks, G2 in three, G3 in two, G4 in two, and G5 in two. We have shown the names of the groups with loyal members on each network. There is one shortest path tree for each group; therefore there are five shortest path trees for five groups.

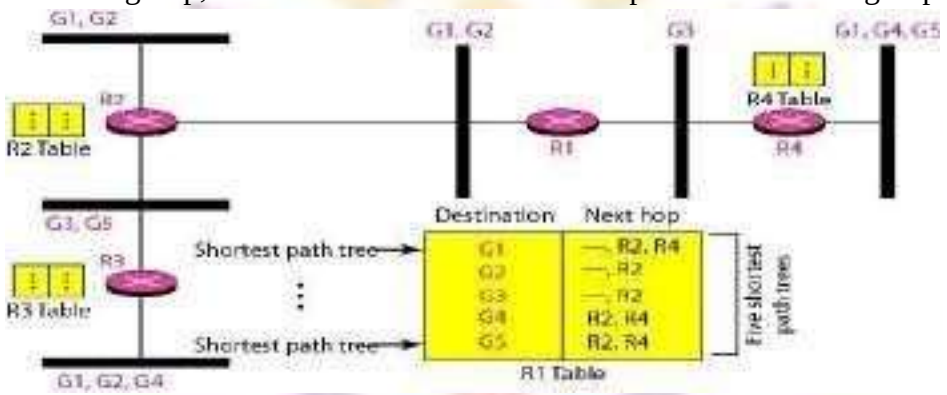


Figure 3.58 Source-based tree approach

b. Group-Shared Tree:. In the group-shared tree approach, instead of each router having *m* shortest path trees, only one designated router, called the center core, or rendezvous router, takes the responsibility of distributing multicast traffic. The core has *m* shortest path trees in its routing table. The rest of the routers in the domain have none. If a router receives a multicast packet, it encapsulates the packet in a unicast packet and sends it to the core router. The core router removes the multicast packet from its capsule, and consults its routing table to route the packet.

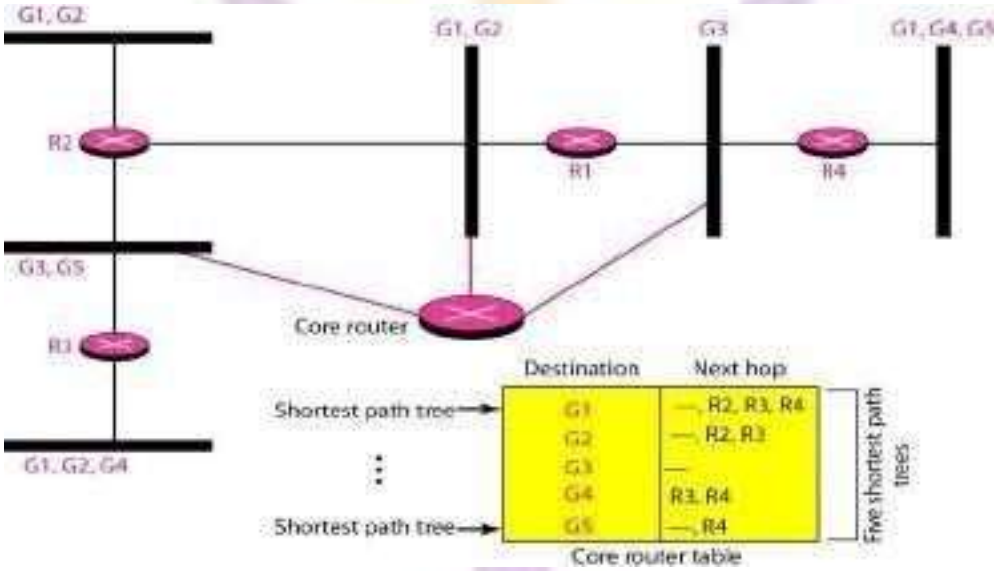


Figure 3.59 Group-shared tree approach

Routing Protocols

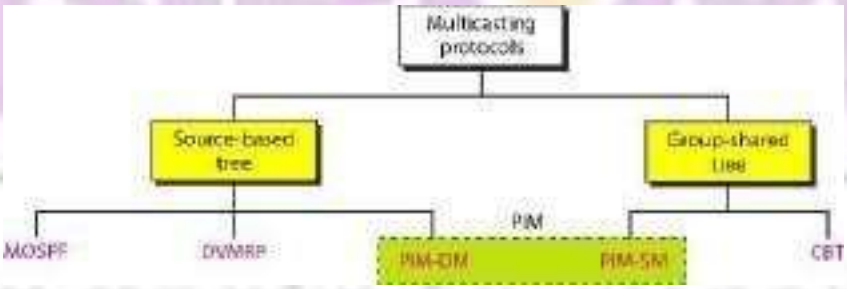


Figure 3.60 Taxonomy of common multicast protocols

a. Multicast Link State Routing: MOSPF

Multicast link state routing uses the source-based tree approach. links. For multicast routing, a node needs to revise the interpretation of *state.* A node advertises every group which has any loyal member on the link. Here the meaning of state is "what groups are active on this link." The information about the group comes from IGMP. Each router running IGMP solicits the hosts on the link to find out the membership status.

MOSPF Multicast Open Shortest Path First (MOSPF) protocol is an extension of the OSPF protocol that uses multicast link state routing to create source-based trees. The protocol requires a new link state update packet to associate the unicast address of a host with the group address or addresses the host is sponsoring. This packet is called the group-membership LSA.

b. Multicast Distance Vector: DVMRP

Multicast Distance Vector Routing Unicast distance vector routing is very simple; extending it to support multicast routing is complicated. Multicast routing does not allow a router to send its routing table to its neighbors. The idea is to create a table from scratch by using the information from the unicast distance vector tables.

Multicast distance vector routing uses source-based trees, but the router never actually makes a routing table. When a router receives a multicast packet, it forwards the packet as though it is consulting a routing table.

1Flooding

1. **Reverse Path Forwarding (RPF)**
2. **Reverse Path Broadcasting (RPB)**
3. **Reverse Path Multicasting (RPM)**

DVMRP The Distance Vector Multicast Routing Protocol (DVMRP) is an implementation of multicast distance vector routing. It is a source-based routing protocol, based on RIP.

c. CBT

The Core-Based Tree (CBT) protocol is a group-shared protocol that uses a core as the root of the tree. The autonomous system is divided into regions, and a core (center router or rendezvous router) is chosen for each region.

The Core-Based Tree (CBT) is a group-shared tree, center-based protocol using one tree per group. One of the routers in the tree is called the core. A packet is sent from the source to members of the group following this procedure:

1. The source, which may or may not be part of the tree, encapsulates the multicast packet inside a unicast packet with the unicast destination address of the

core and sends it to the core. This part of delivery is done using a unicast address; the only recipient is the core router.

2. The core decapsulates the unicast packet and forwards it to all interested intetfaces.
3. Each router that receives the multicast packet, in tum, forwards it to all interested interfaces.

d. PIM

Protocol Independent Multicast (PIM) is the name given to two independent multicastrouting protocols: Protocol Independent Multicast, Dense Mode (PIM-DM) and Protocol Independent Multicast, Sparse Mode (PIM-SM). Both protocols are unicast protocol-dependent, but the similarity ends here.

PIM-DM

PIM-DM is used when there is a possibility that each router is involved in multicasting (dense mode). In this environment, the use of a protocol that broadcasts the packet is justified because almost all routers are involved in the process. PIM-DM is a source-based tree routing protocol that uses RPF and pruning and grafting strategies for multicasting. Its operation is like that of DVMRP.

PIM-SM

PIM-SM is used when there is a slight possibility that each router is involved in multicasting (sparse mode). In this environment, the use of a protocol that broadcasts the packet is not justified; a protocol such as CBT that uses a group-shared tree is more appropriate. PIM- SM is used in a sparse multicast environment such as a WAN. PIM-SM is a group-shared tree routing protocol that has a rendezvous point (RP) as the source of the tree.

Tunneling

Tunneling is a method of transmitting data that is intended for use only within a private network through a public network in such a way that the routing nodes in the public network are unaware that the transmission is a part of private network. Handling the general case of making two different networks interwork is exceedingly difficult. However, there is a common special case that is manageable. This case is where the source and destination hosts are on the same type of

network, but there is a different network in between. As an example, think of an international bank with a TCP/IP-based Ethernet in Paris, a TCP/IP-based Ethernet in London, and a non-IP wide area network (e.g., ATM) in between, as shown in Fig. 10.1.

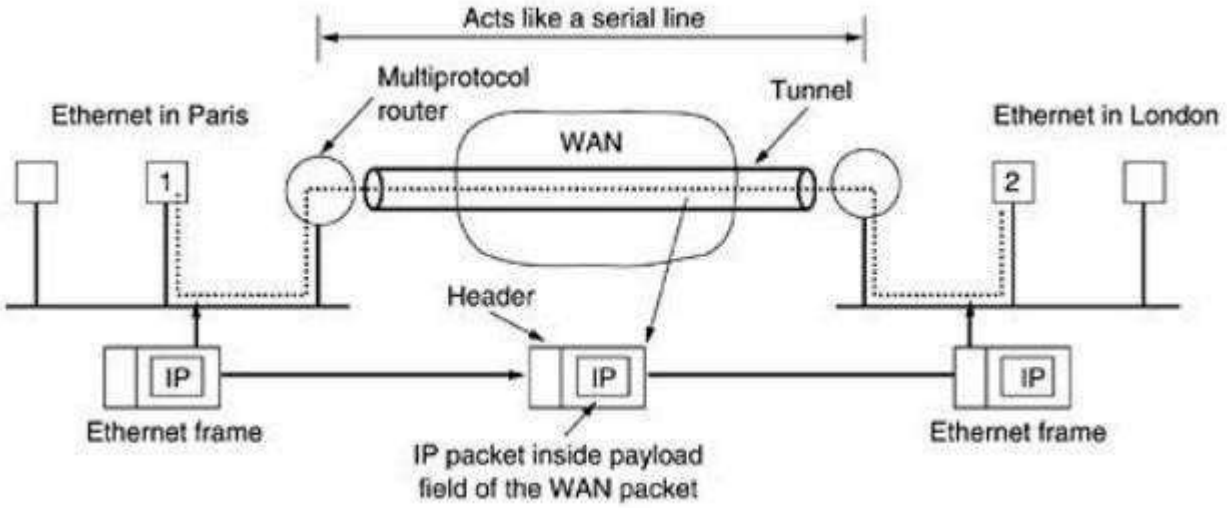


Fig 10.1. Tunneling a packet from Paris to London.

The solution to this problem is a technique called tunneling. To send an IP packet to host 2, host 1 constructs the packet containing the IP address of host 2, inserts it into an Ethernet frame addressed to the Paris multiprotocol router, and puts it on the Ethernet. When the multiprotocol router gets the frame, it removes the IP packet, inserts it in the payload field of the WAN network layer packet, and addresses the latter to the WAN address of the London multiprotocol router. When it gets there, the London router removes the IP packet and sends it to host 2 inside an Ethernet frame.

The WAN can be seen as a big tunnel extending from one multiprotocol router to the other. The IP packet just travels from one end of the tunnel to the other, snug in its nice box. It does not have to worry about dealing with the WAN at all. Neither do the hosts on either Ethernet. Only the multiprotocol router has to understand IP and WAN packets. In effect, the entire distance from the middle of one multiprotocol router to the middle of the other acts like a serial line. An analogy may make tunneling clearer. Consider a person driving her car from Paris to London. Within France, the car moves under its own power, but when it hits the English Channel, it is loaded into a high-speed train and transported to England through the Chunnel (cars are not permitted to drive through the Chunnel). Effectively, the car is being carried as freight, as depicted in Fig. 10.2. At the far end, the car is let loose on the English roads and once again continues to move under its own power. Tunneling of packets through a foreign network works the same way.

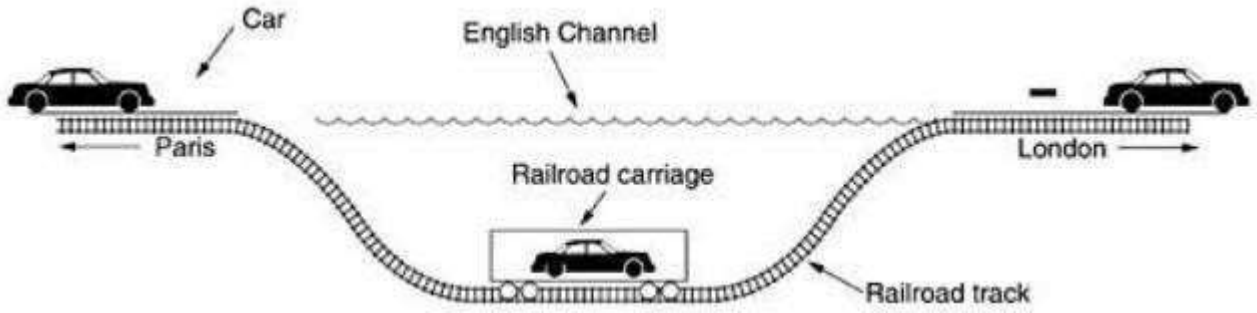


Fig 10.2. Tunneling a car from France to England.

Internetwork routing

Routing through an internetwork is similar to routing within a single subnet, but with some added complications. Consider, for example, the internetwork of Fig. 11(a) in which five networks are connected by six (possibly multiprotocol) routers. Making a graph model of this situation is complicated by the fact that every router can directly access (i.e., send packets to) every other router connected to any network to which it is connected. For example, B in Fig. 11(a) can directly access A and C via network 2 and also D via network 3. This leads to the graph of Fig. 11(b).

Once the graph has been constructed, known routing algorithms, such as the distance vector and link state algorithms, can be applied to the set of multiprotocol routers. This gives a two-level routing algorithm: within each network an interior gateway protocol is used, but between the networks, an exterior gateway protocol is used ("gateway" is an older term for "router"). In fact, since each network is independent, they may all use different algorithms. Because each network in an internetwork is independent of all the others, it is often referredto as an Autonomous System (AS). A typical internet packet starts out on its LAN addressed to the local multiprotocol router (in the MAC layer header). After it gets there, the network layer code decides which multiprotocol router to forward the packet to, using its own routing tables. If that router can be reached using the packet's native network protocol, the packet is forwarded there directly. Otherwise it is tunneled there, encapsulated in the protocol required by the intervening network. This process is repeated until the packet reaches the destination network.

One of the differences between internetwork routing and intranetwork routing is that internetwork routing may require crossing international boundaries. Various laws suddenly come into play, such as Sweden's strict privacy laws about exporting personal data about Swedish citizens from Sweden. Another example is the Canadian law saying that data traffic originating in Canada and ending in Canada may not leave the country. This law means that traffic from

Windsor, Ontario to Vancouver may not be routed via nearby Detroit, even if that route is the fastest and cheapest

Another difference between interior and exterior routing is the cost. Within a single network, asingle charging algorithm normally applies. However, different networks may be under different managements, and one route may be less expensive than another. Similarly, the quality of service offered by different networks may be different, and this may be a reason to choose one route over another.

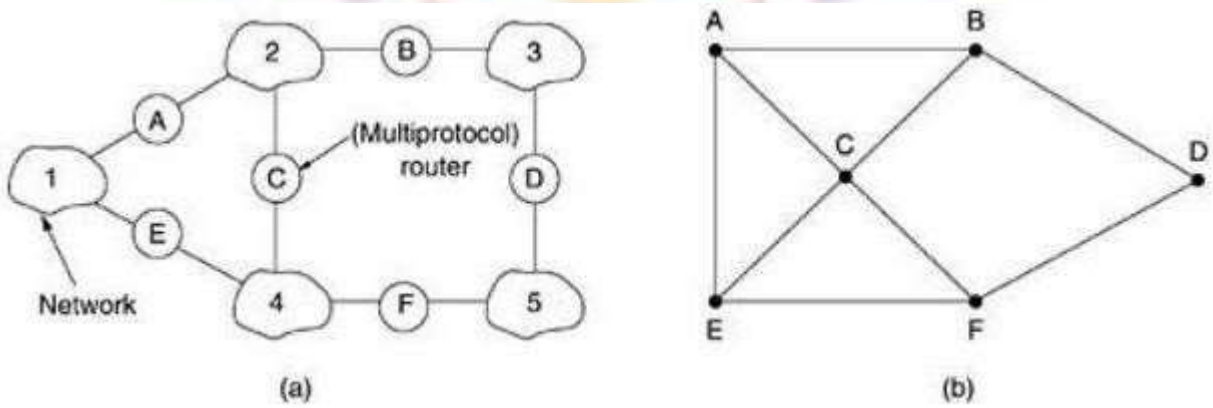


Fig.11 (a) An internetwork. (b) A graph of the internetwork