If you are working on something that you really care about, you don't have to be pushed. The vision pulls you.

Steve Jobs

4

Network Layer

4.1 Internetworking and Devices

Why Interconnect?

- To seperate/connect one corporate division with another.
- To connect two LANs with different protocols.
- To connect a LAN to the internet.
- To break a LAN into segments to relieve traffic congestion.
- To provide as security wall between two different types of users.

Internetworking Devices

Figure **??** depicts the different devices that operates in different layers. The layers matters because different devices use different pieces of information to decide how to switch.

Different networking devices are as follows:

4.1.1 Repeaters

- A repeater operates at the physical layer.

- A repeater connects segments of a LAN.

- It's main function is to regenerate the signal over the same network before the signal becomes too weak or corrupted so as to extend the length to which the signal can be transmitted over the same network.

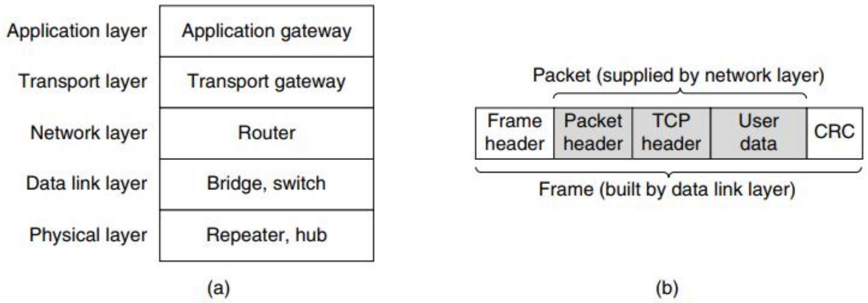


Figure 4.1: a. Which device is in which layer b. Frames, packets, and headers.

- An important point to be noted about repeaters is that they do not amplify the signal. When the signal becomes weak, they copy the signal bit by bit and regenerate it at the original strength.
- A repeater does not understand frames, packets, or headers; it has no filtering capacity. They understand the symbols that encode bits as volts.
- It is a two port device.

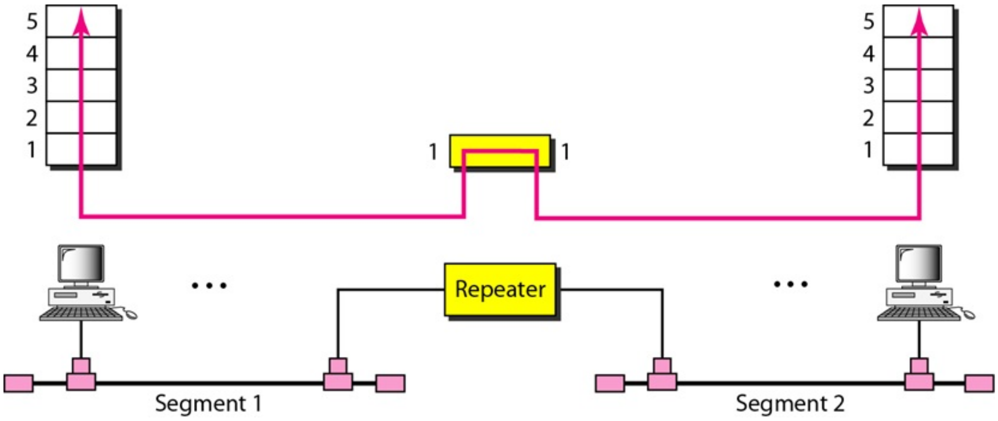


Figure 4.2: A repeater connecting two segments of a LAN.

4.1.2 Hub

- A hub is a multiport repeater that operates on physical layer.
- It has a number of input ports that it joins electrically.
- Frames arriving on any of the lines are sent out on all the others; i.e. the collision domain of all stations connected to the hub are same.
- It has no filtering capability; it does not have the intelligence to find from which port the frame should be sent out.

4.1.3 Bridge

- A bridge operates at the MAC sub-layer of data link layer.
- A bridge is a repeater, with added functionality of filtering content by reading the

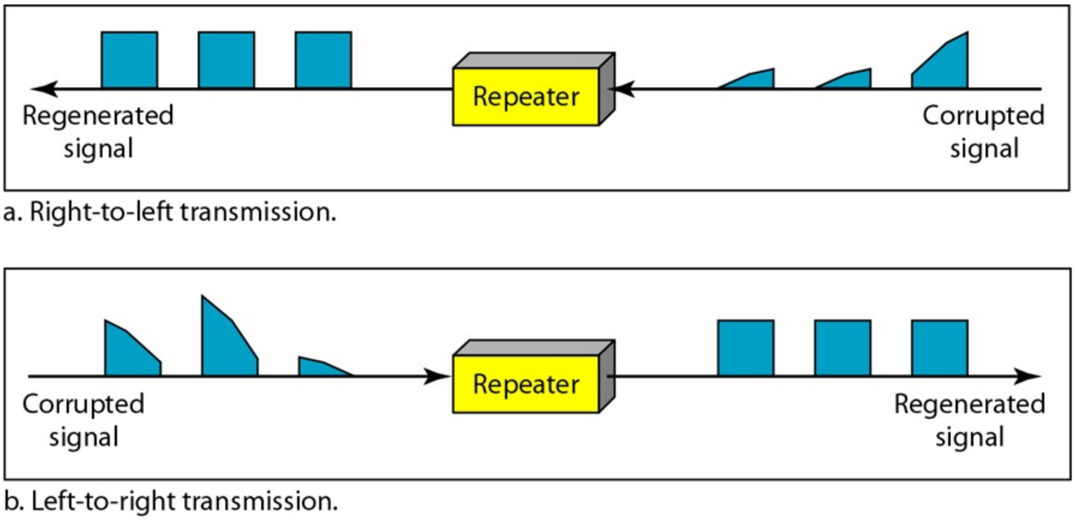


Figure 4.3: Function of a repeater.

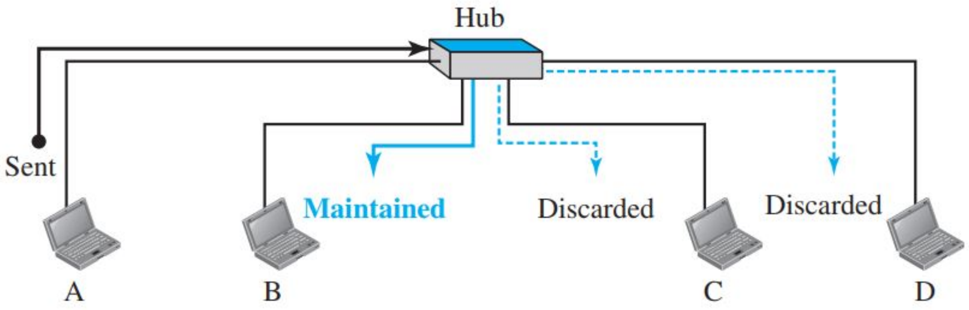


Figure 4.4: A hub.

MAC addresses of source and destination.

- A bridge stores the hardware addresses observed from frames received by each interface.

- It uses this information to learn which frames need to be forwarded by the bridge.

- Each bridge has two connections (ports) and ther is a table associated with each port.

- A bride observes each frame that arrives at a port.
- It extracts the source address from the port.
- and places that address in the port's routing table.

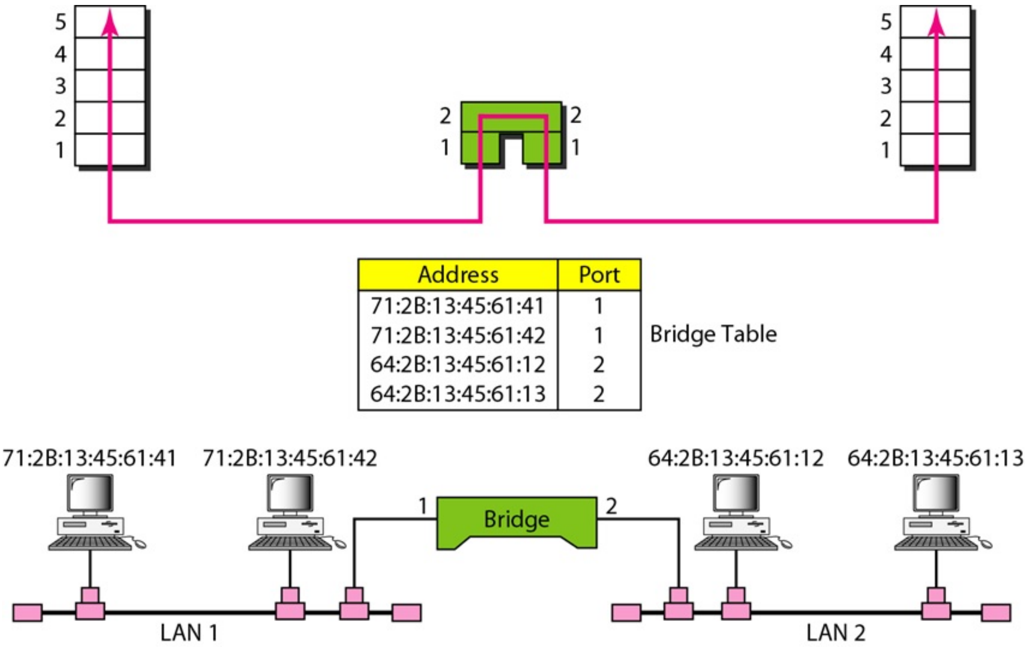
- A bridge is also used for interconnecting two LANs working on the same protocol.

- It has a single input and single output port, thus making it a 2 port device.

4.1.4 Switch/Ethernet Switch

- A switch is essentially essentially a multiport bridge used to connect mostly hosts, as well as LANs.

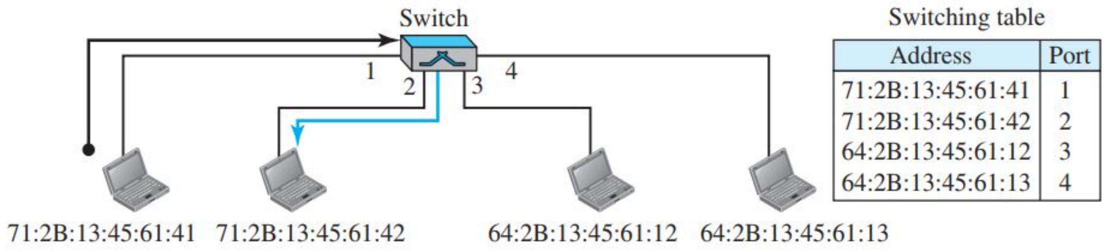
- It stores and forward Ethernet frames.



Address	Port
71:2B:13:45:61:41	1
71:2B:13:45:61:42	1
64:2B:13:45:61:12	2
64:2B:13:45:61:13	2

Figure 4.5: A bridge connecting two LANs.

- A link-layer switch has filtering capability. It can check incoming frame's MAC address, and can decide from which outgoing port(s) the frame should be sent. An illustration is shown in figure **??**.



Address	Port
71:2B:13:45:61:41	1
71:2B:13:45:61:42	2
64:2B:13:45:61:12	3
64:2B:13:45:61:13	4

Figure 4.6: A link layer switch .

- Each connection on a switch creates a separate collision domain, but a switch remains one broadcast domain by default.

Switch: multiple simultaneous transmission

- Each hosts have dedicated, direct connection to switch.
- Switches buffer packets.
- Ethernet protocol used on each incoming link; but no collisions if full duplex is used as each link is its own collision domain.
- In figure **??** switching from A-to-A' and B-to-B' can transmit simultaneously without

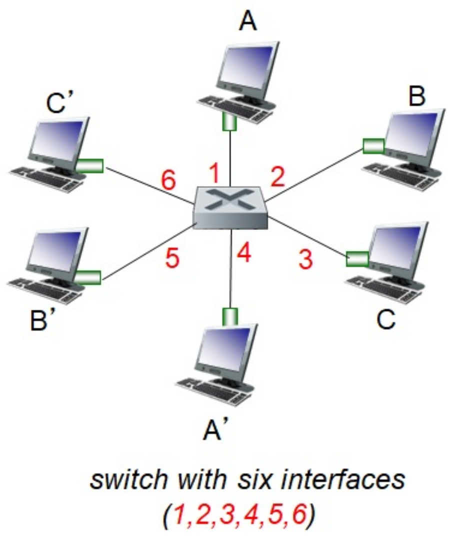


Figure 4.7: A switch with six interfaces.

collisions.

Switch: cut-through switching

- Start transmitting as soon as it reads destination address.

- Transmit the head of the packet via the outgoing link while still receiving the tail via the incoming link.
- Much faster.
- Cannot detect corrupt packets.
- Destination is responsible for detecting corrupt packets.
- Can propagate the corrupt packets to the network.

Switch: self-learning

- Each switch has a switch table; each entry consists of MAC address of host, interfaces to reach host, and time stamp. It looks like a routing table.
- Switch table maps MAC address to outgoing interface.
- Switch table is created as when a frame arrives. It then:

- Inspect the *source* MAC address.
- Associate the address with the *incoming* interface.
- Store the mapping in the switch table.
- Use a time-to-live filed to eventually forget the mapping.

Switch:frame filtering/forwarding

When frame received at switch:

1. record incoming link, MAC address of sending host

```
2. index switch table using MAC destination address
3. if entry found for destination
      then {
     if destination of segment from which frame arrived
       then drop frame
       else forward frame on interface indicated by entry }
     else flood      /* forward on all interfaces except arriving interface */
```

Flooding can lead to Loops

- Switched sometimes need to broadcast frames.
  - Upon receiving a frame with an unfamiliar destination.
  - Upon receiving a frame sent to the braodcast address.
- Broadcasting is implemented by flooding.
  - Transmitting frame out every interface.
  - Except the one where the frame arrived.

Spanning Tree
- Solution to these loop is ***spanning tree***.
- It creates the sub-graph of all nodes but contains no cycles; links not in the spanning tree do not forward frames.
- Spanning tree ensure the topology has no loops. It avoids using some of the links when flooding to avoid forming a loop.

Question: Differentiate between collision domain and broadcast domain.
- *Collision domain* is an Ethernet term that refers to a network scenario wherein one device sends a frame out on a physical network segment forcing every other device on the same segment to pay attention to it. *Broadcast domain* refers to a group of devices on a specific network segment that hear all the broadcasts sent out on that specific network segment.

4.1.5 Router

- A router is a layer-3 switch; that is, it operates on network layer.
- It forwards the packet to next-hop router based on routing table. Routing table is formed based on communication between routers using *routing protocols*.
- Router can select the most appropriate cost effective path to forward the packets.
- It is the router that connects the different networks. In other words, a router is an

internetworking device; it connects independent networks to form an internetwork. According to this definition, two networks connected by a router become an *internetwork* or an *internet*.
- Unlike switch, router breaks up the broadcast domain by default and provide WAN services.

4.1.6 Gateway

- A gateway is a networking device used to connect two dissimilar networks using different protocols.
- Transport gateway connects two computers that use different connection-oriented transport protocols. For example, suppose a computer using the connection-oriented TCP/IP protocol needs to talk to a computer using a different connection-oriented transport protocol called SCTP. The transport gateway can copy the packets from one connection to the other, reformatting them as need be.
- Application gateways understand the format and contents of the data and can translate messages from one format to another. An email gateway could translate Internet messages into SMS messages for mobile phones, for example.

4.2 Network Layer

- The network layer is responsible for the host-to-host delivery of datagrams. That is, it is concerned with getting packets from the source all the way to destination.
- Routers are specified in this layer. It forwards the incoming packets to the next-hop router based on routing table.
- A network layer performs the routing. Routing algorithm determines the best effective path to route the packet from source to destination.
- A network layer provides services to the transport layer and receives services from the data-link layer.
- On sending side, network layer encapsulates segments into *datagrams*; on receiving side, it delivers *segments* to transport layer.
- It can provide connection-less service (datagram network) or connection-oriented service (virtual circuit network).

Forwarding: It is process of moving packets from router's input to appropriate router output.

Routing: It determines route taken by packets from source to destination. Routing algorithms determines the path.

Issue	Datagram Network	Virtual Circuit
Cirucit Setup	No setup is required, but a more complicated lookup procedure is required to locate th entry for the destination.	Setup is required. but takes time and cosumues resources.
Addressing	Each packet contains the full source and destination address.	Each packet contains short VC number.
State Information	Routers do not hold state information about connections.	Each VC requires router table space per connection.
Routing	Each packet is routed independently.	Once setup is finished, all packets follow it.
Effect of router failures	Only those packets queued in the router at that time get lost during crash.	All VCs that passed through the failed router are terminated.
Quality of service	It is difficult to maintain QoS.	Easy if enough resources can be allocated in advance for each VC.
Congestion Control	Congestion avoidance is more difficult.	Easy if enough resources can be allocated in advance for each VC.
Ordering of packet	At destinaion, packets can arrive out of order, since they travel the differetn path.	At destination, packets arrive in order.

4.3 IPv4 Address

- It is a 32-bit long address. They are unique and universal.

- Every Host and router on the internet has an IP Address. This IP address is unique and no two devices on the Internet can have the same address at the same time.

- It is universal in the sense that the addressing scheme 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

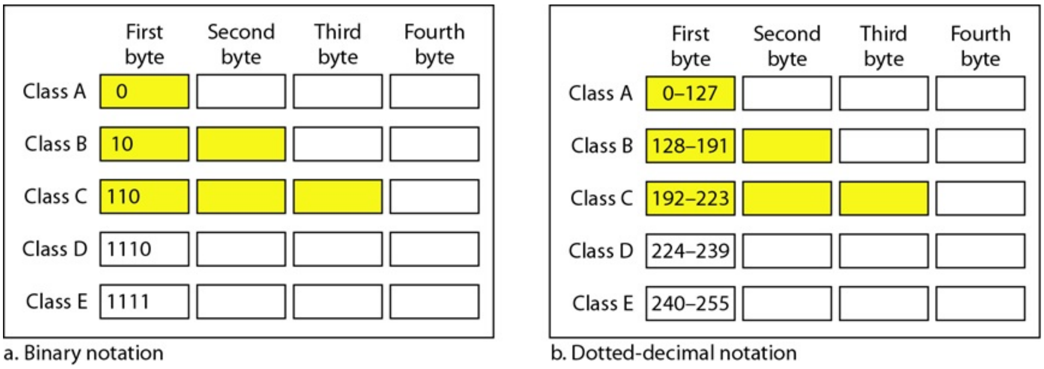
uses N bits to define an address, the address space is 2^N.
- IPv4 uses 32-bit addresses, which means that the address space is 2^32 or 4,294,967,296 (more than 4 billion).

Notations There are two ways to represent IPv4.

i. **Binary notation**
   - The IPv4 address is displayed as 32 bits.
   - ex. 11000001 10000011 00011011 11111111
   - Mostly used by devices for processing
ii. **Dotted-decimal notation**
   - Denoted in decimal format each byte being separated by dot.
   - Each byte (octet) is 8 bits hence each number in dotted-decimal notation is a value ranging from 0 to 255. - eg. 117.25.13.1
   - Mostly used by human configurations

IPv4 Classes (Classful Address)

- IP Addressing uses the concept of class. So this architecture is called Classful addressing. - In Classful addressing, the address space is divided into five classes: A, B, C, D and E.
- Division is based on the first byte in doted decimal format.



	First byte	Second byte	Third byte	Fourth byte
Class A	0			
Class B	10	🟨		
Class C	110	🟨	🟨	
Class D	1110			
Class E	1111			

a. Binary notation

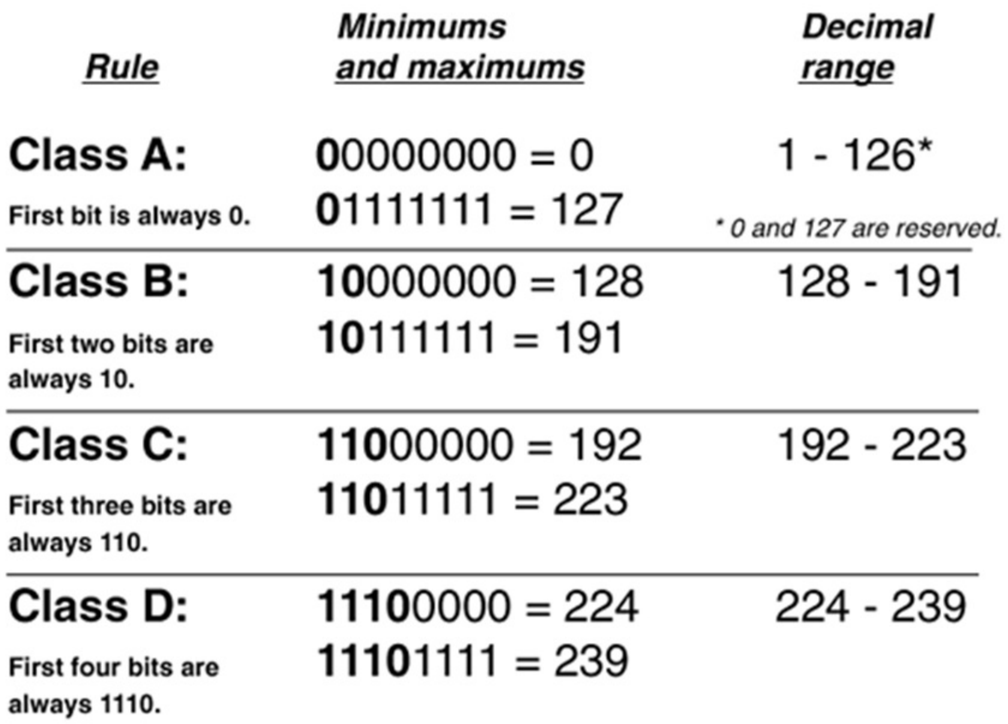
	First byte	Second byte	Third byte	Fourth byte
Class A	0–127			
Class B	128–191	🟨		
Class C	192–223	🟨	🟨	
Class D	224–239			
Class E	240–255			

b. Dotted-decimal notation

Figure 4.8: Finding the classes in binary and dotted-decimal notation.

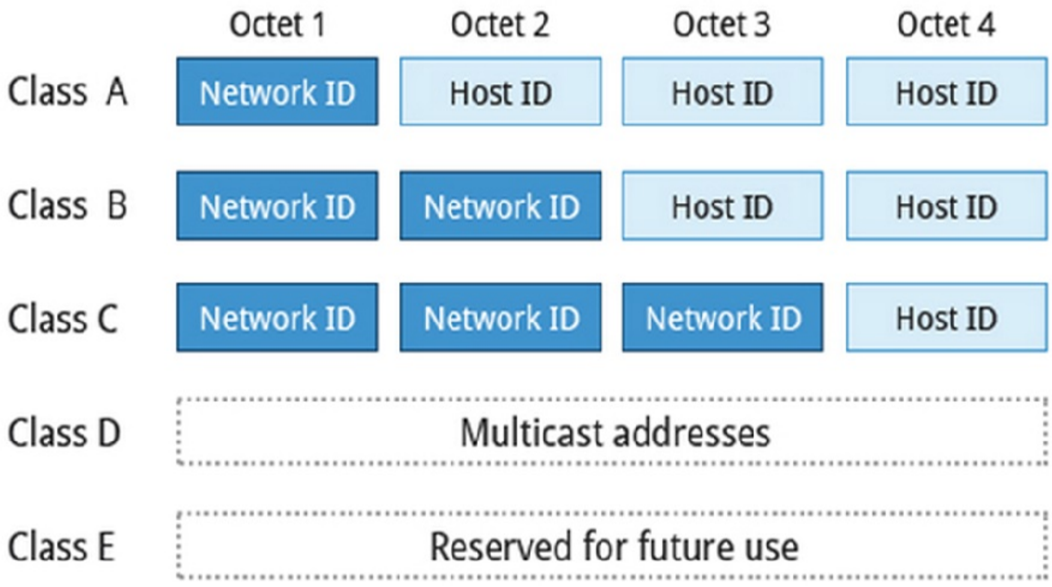
Network ID and Host ID

- Each class of classful IP addressing has network id and host id as shown in figure **??**.
- In class A, the network id is 8 bits, but since the first bit, which is 0, defines the class, we can have only seven bits as the network identifier. This means there are only 2^7 = 128 networks in the world that can have a class A address. The host id has 24 bits, so the number of host we can have with class A is 2^{24} = 16,777,216.
- In this manner, we can find total networks and host for Class B and C IP.



Rule	Minimums and maximums	Decimal range
Class A: First bit is always 0.	00000000 = 0 01111111 = 127	1 - 126* * 0 and 127 are reserved.
Class B: First two bits are always 10.	10000000 = 128 10111111 = 191	128 - 191
Class C: First three bits are always 110.	11000000 = 192 11011111 = 223	192 - 223
Class D: First four bits are always 1110.	11100000 = 224 11101111 = 239	224 - 239

Figure 4.9: Ranges of classes of IP.

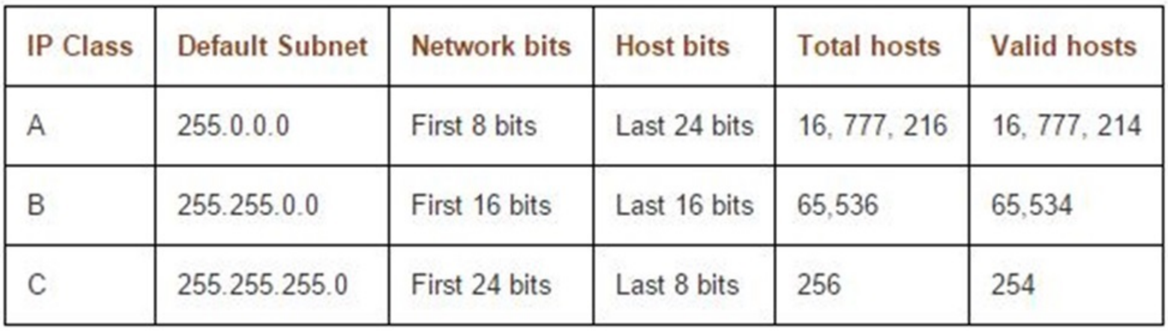


	Octet 1	Octet 2	Octet 3	Octet 4
Class A	Network ID	Host ID	Host ID	Host ID
Class B	Network ID	Network ID	Host ID	Host ID
Class C	Network ID	Network ID	Network ID	Host ID
Class D	Multicast addresses			
Class E	Reserved for future use			

Figure 4.10: Netword ID and Host ID of classess classful IP addressing.

- Class A, B and C IPs are used for unicast address; while class D is used for multicast purpose and Class E IP is used as reserve.

- The reason that classful addressing has become obsolete is *address depletion*.
- To alleviate address depletion, two strategies were proposed and, to some extent, im-



IP Class	Default Subnet	Network bits	Host bits	Total hosts	Valid hosts
A	255.0.0.0	First 8 bits	Last 24 bits	16, 777, 216	16, 777, 214
B	255.255.0.0	First 16 bits	Last 16 bits	65,536	65,534
C	255.255.255.0	First 24 bits	Last 8 bits	256	254

Figure 4.11: Number of hosts for classful IP.

plemented: subnetting and supernetting.

4.4 Subnetting

- Subnetting is the process of dividing a large network up into smaller networks, called subnets or sub networks.
- Each of these subnets has its own specific address.
- To create these additional networks we use a subnet mask. The subnet mask simply determines which portion of the IP address belongs to the host.
- The subnet address is created by dividing the host address into network address and host address. - For example, 172.16.1.0, 172.16.2.0, 172.16.3.0 and 172.16.4.0 are all subnets within network 171.16.0.0.

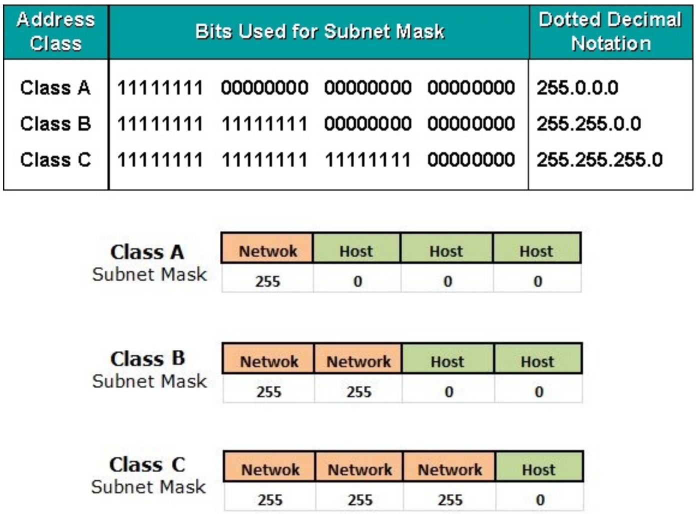
- To create subnetworks, you take (borrow) bits from the host portion of the IP address and reserve them to define the subnet address.
- That means fewer bits for hosts, so the more subnets.

Subnet Masks

- For the subnet address scheme to work, every machine on the network must know which part of the host address will be used as the subnet address.
- This is accomplished by assigning a subnet mask to each machine.
- A subnet is 32 bit value that allows the recipient of IP packets to distinguish the network ID portion of the IP address from the host ID portion of the IP address.
- Subnet mask has binary 1s is in all bits specifying the network field, and binary 0s in all bits specifying the host field.
- A subnet address is created by borrowing the bits from host field of IP address.
- Subnet mask bits should come from the high-order (left most) bits of the host field.

Default Subnet Mask

- Each class has a predefined default subnet mask that tell us the octets, which are already part of the network portion, as well as how many bits we have available to work with.
- Whatever network class is it, we cannot change those bits that are already assigned.
- We cannot assign the network ID and the broadcast address to a host.
- Regardless how many bits are left in the host field, network ID and the broadcast address must be reserved.
- Subnet bits start at the left and go to the right, without skipping bits.



Address Class	Bits Used for Subnet Mask	Dotted Decimal Notation
Class A	11111111 00000000 00000000 00000000	255.0.0.0
Class B	11111111 11111111 00000000 00000000	255.255.0.0
Class C	11111111 11111111 11111111 00000000	255.255.255.0

Class A Subnet Mask

Netwok	Host	Host	Host
255	0	0	0

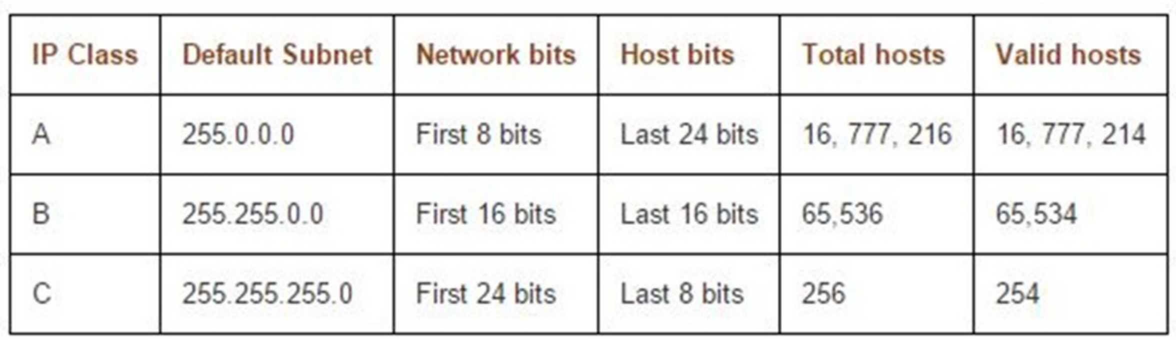
Class B Subnet Mask

Netwok	Network	Host	Host
255	255	0	0

Class C Subnet Mask

Netwok	Network	Network	Host
255	255	255	0

Figure 4.12: Default subnet masks of class A, B and C IPs.



IP Class	Default Subnet	Network bits	Host bits	Total hosts	Valid hosts
A	255.0.0.0	First 8 bits	Last 24 bits	16, 777, 216	16, 777, 214
B	255.255.0.0	First 16 bits	Last 16 bits	65,536	65,534
C	255.255.255.0	First 24 bits	Last 8 bits	256	254

Figure 4.13: Hosts in class A, B and C IP.

Some Terminologies:

Network ID

First address of subnet is called network ID. This address is used to identify one segment or broadcast domain from all the other segments in the network.

Block Size
Block size is the size of subnet including network address, hosts addresses and broadcast address.

Broadcast ID
There are two types of broadcast, direct broadcast and full broadcast.
Direct broadcast or local broadcast is the last address of subnet and can be hear by all hosts in subnet.
Full broadcast is the last address of IP classes and can be hear by all IP hosts in network.
Full broadcast address is 255.255.255.255

Host Addresses
All address between the network address and the directed broadcast address is called host address for the subnet. You can assign host addresses to any IP devices such as PCs, servers, routers, and switches.

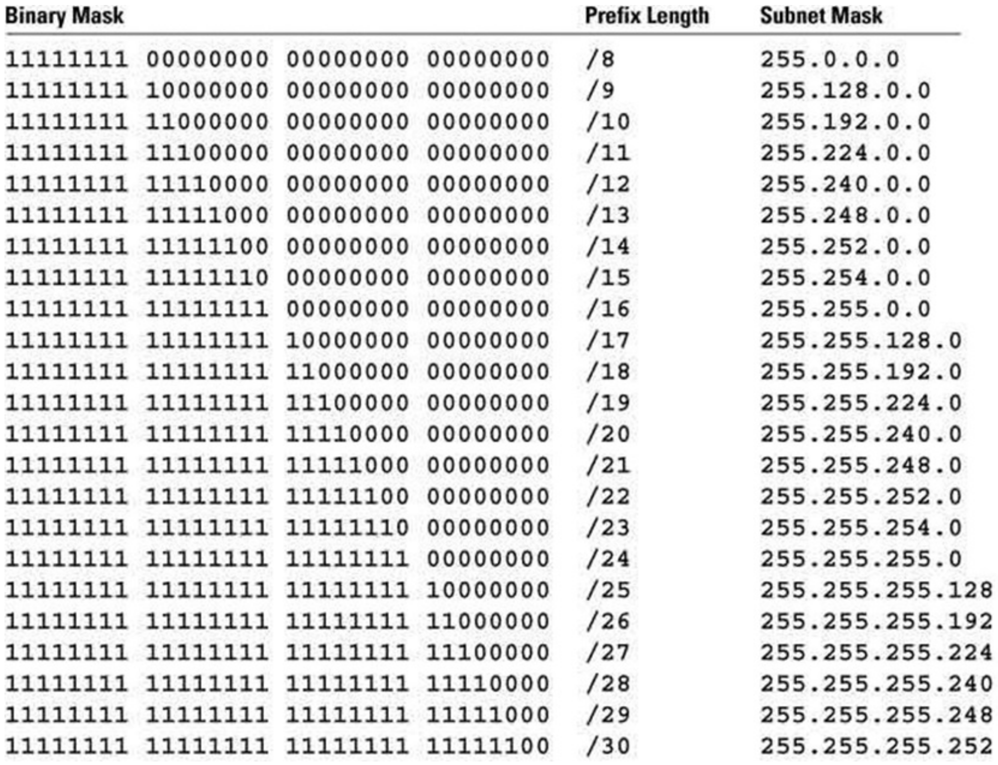
4.5 Classless Inter Domain Routing (CIDR)

- CIDR is a slash notation of subnet mask. CIDR tells us number of on bits(i.e. number of 1s) in a network address.
- It is the basically the methods that ISPs (Internet Service Provider) use to allocate a number of addresses to a company, a home-a customer.
- CIDR provide addresses in a certain block size.

Example:
- When you receive a block of address from an ISP, what you get will look like something like this 192.168.10.32/28.
- It tells what your subnet mask is.
- The slash notation (/) means how many bits are turned on (1s).
- Obviously, the maximum could only be /32 because a byte is 8 bits and there are 4 bytes in an IP address. (4*8=32). Bur, regardless of the class of address, the largest subnet mask can be only /30 because you've got to keep at least 2 bits for host bits.

In subnetting we find the answer of following questions:

- What is subnet mask for given address?
- How many subnets does given subnet mask provide ?
- What is block size for given subnet mask?
- What are the valid subnets?



Binary Mask	Prefix Length	Subnet Mask
11111111 00000000 00000000 00000000	/8	255.0.0.0
11111111 10000000 00000000 00000000	/9	255.128.0.0
11111111 11000000 00000000 00000000	/10	255.192.0.0
11111111 11100000 00000000 00000000	/11	255.224.0.0
11111111 11110000 00000000 00000000	/12	255.240.0.0
11111111 11111000 00000000 00000000	/13	255.248.0.0
11111111 11111100 00000000 00000000	/14	255.252.0.0
11111111 11111110 00000000 00000000	/15	255.254.0.0
11111111 11111111 00000000 00000000	/16	255.255.0.0
11111111 11111111 10000000 00000000	/17	255.255.128.0
11111111 11111111 11000000 00000000	/18	255.255.192.0
11111111 11111111 11100000 00000000	/19	255.255.224.0
11111111 11111111 11110000 00000000	/20	255.255.240.0
11111111 11111111 11111000 00000000	/21	255.255.248.0
11111111 11111111 11111100 00000000	/22	255.255.252.0
11111111 11111111 11111110 00000000	/23	255.255.254.0
11111111 11111111 11111111 00000000	/24	255.255.255.0
11111111 11111111 11111111 10000000	/25	255.255.255.128
11111111 11111111 11111111 11000000	/26	255.255.255.192
11111111 11111111 11111111 11100000	/27	255.255.255.224
11111111 11111111 11111111 11110000	/28	255.255.255.240
11111111 11111111 11111111 11111000	/29	255.255.255.248
11111111 11111111 11111111 11111100	/30	255.255.255.252

Figure 4.14: CIDR values.

- What are the total hosts?
- How many valid hosts are available per subnet?
- What is broadcast address of each subnet?
- What is network address of each subnet?

4.6 Variable Length Subnet Masks (VLSM)

- VLSM is a process of dividing an IP space into the subnets of different sizes without wasting IP addresses.

- VLSM is a process of breaking down subnets into the smaller subnets, according to the need of individual networks. It is subnetting a subnet.

- It can also be called a classless IP addressing When we perform subnetting, all subnets have the same number of hosts, this is known as FLSM (Fixed length subnet mask).

- In FLSM all subnets use same subnet mask, this lead to inefficiencies because of wastage of IP address. In real life scenario, some subnets may require large number of host addresses while other may require only few addresses.

- In VLSM Subnetting, we do subnetting of subnets according the network requirement.

Implementing VLSM Networks

- To create VLSM quickly and efficiently, you need to understand how block sizes and charts work together to create the VLSM.

Table 4.1: Block sizes.

Prefix	Mask	Subnets	Valid Hosts	Block Size
/25	128	2	126	28
/26	192	4	62	64
/27	224	8	30	32
/28	240	16	14	16
/29	248	32	6	8
/30	252	64	2	4

- Table below show the block sizes used when creating VLSMs with Class C networks.

Note: Example problems of subnetting and vlsm will be done in class!!

4.7 Supernetting

Supernetting is the method for combining two or more contiguous network address spaces to simulate a single, larger, address space.

- In Subnetting we are adding the bits from the host part to the network part.

- But in Supernetting we do the reverse. - Here in super netting

- We add bits from the network part to the host part. To supernet two contiguous networks is simple. Just convert the networks in to binaries, compare the bits of the two networks.

- Till where you have the similiar bit pattern, use a subnet mask bit of "1", and after that "0". Use the altered subnet mask for two networks. That's it!

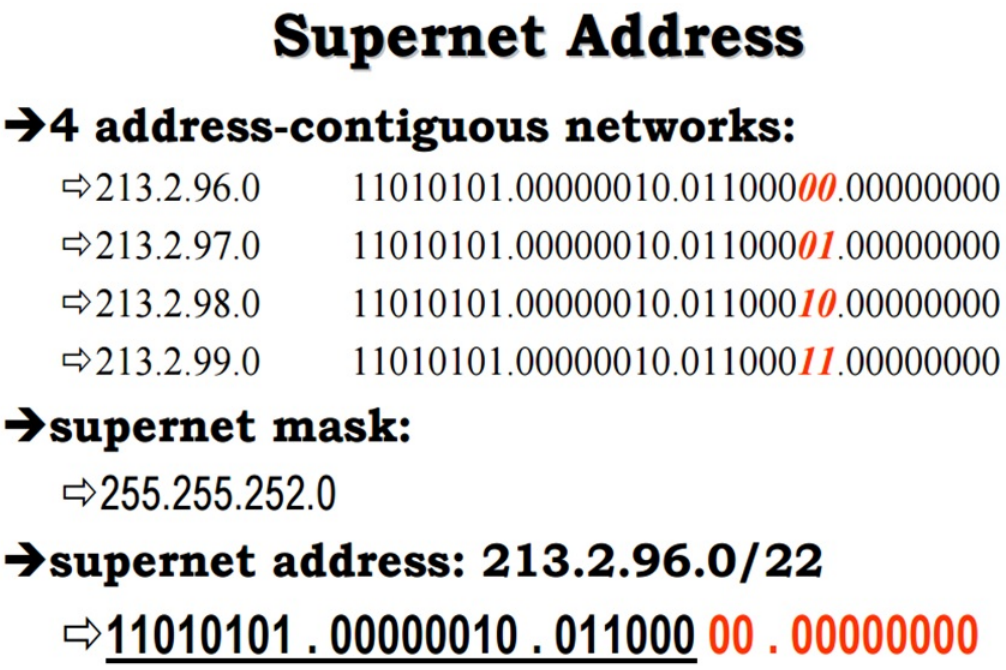
Example:

- For example, you may want to supernet the networks 192.168.10.0 and 192.168.11.0 to make a single, large network. Following two lines are the conversions of the above network addresses to binaries and the last line is the new subnet mask.

11000000.10101000.00001010.00000000
11000000.10101000.00001011.00000000
11111111.11111111.11111110.00000000

- The changed subnet mask is 255.255.254.0 can be used to supernet 192.168.10.0 and 192.168.11.0.

- The concept of supernetting is used in routing protocols for "route summarization".



Supernet Address

➔4 address-contiguous networks:

⇨213.2.96.0 11010101.00000010.01100000.00000000

⇨213.2.97.0 11010101.00000010.01100001.00000000

⇨213.2.98.0 11010101.00000010.01100010.00000000

⇨213.2.99.0 11010101.00000010.01100011.00000000

➔supernet mask:

⇨255.255.252.0

➔supernet address: 213.2.96.0/22

⇨11010101 . 00000010 . 011000 00 . 00000000

4.8 Routing

- Routing refers to taking a packet from one device and sending it through the network to another device on a different network.
- Through routing, packets from the source are delivered to destination network.
- Routing protocols are used to continuously update the routing table that are consulted for forwarding and routing.
- Routing algorithm is that part of network layer software responsible for deciding which output line an incoming packet should be transmitted on.
- Routing algorithm can be grouped into major classes: *non-adaptive* and *adaptive*.

- Routers don't really care about hosts — they only care about networks and the best path to each one of them. The logical network address of the destination host is key to get packets through a routed network. It's the hardware address of the host that's used to deliver the packet from a router and ensure it arrives at the correct destination host.

- Here's an important list of the minimum factors a router must know to be able to effectively route packetes:

- Destination address.
- Neighbor routers from which it can learn about remote networks.
- Possible routes to all remote networks.
- The best route to each remote network.
- How to maintain and verify routing information.

- The router learns about remote networks from neighboring routers or from an administrator. The router then builds a routing table, which is basically a map of the internetwork, and it describes how to find remote networks. If a network is directly connected, then the router already knows how to get to it.

- But if a network isn't directly connected to the router, the router must use one of two ways to learn how to get to the remote network. The static routing method requires someone to hand-type all network locations into the routing table, which can be a pretty daunting task when used on all but the smallest of networks!

- Conversely, when dynamic routing is used, a protocol on one router communicates with the same protocol running on neighboring routers. The routers then update each other about all the networks they know about and place this information into the routing table. If a change occurs in the network, the dynamic routing protocols automatically inform all routers about the event. If static routing is used, the administrator is responsible for updating all changes by hand onto all routers. Most people usually use a combination of dynamic and static routing to administer a large network.

4.8.1 Criteria for Good Routing

- ***Simplicity***: Routing algorithm should be simple.
- ***Correctness***: Each packet should be correctly delivered to the correct destination.
- ***Robustness***: The routing algorithm should be able to cope with changes in the topology and traffic without requiring all jobs in all hosts to be aborted. In addition it should cope with hardware and software failure of all kinds.
- ***Stability***: Routing algorithm should converge quickly. Also, routing algorithm should never converge to a fixed set of paths, no matter how long they run.
- ***Fairness and Efficiency***: Routing algorithm should choose "best" paths that have small delay and high bandwidth. But, there is always trade-off between these two.

4.8.2 Non-adaptive Routing

- It is also called *static routing*.

- In this type of routing, the route to be taken in going from one node to the other is computed in advance, off-line, and downloaded to the routers when the network is booted.

- These algorithms do not base their routing decisions on measurements and estimates of the current traffic and topology.

- Static routing means someone must hand-type all network locations into the routing table. If static routing is used, the administrator is responsible for updating all changes by hand onto all routers.

- Because static routing does not respond to failures, static routing is mostly useful for situations in which the routing is choice is clear.

Advantages of static routing
- Minimal CPU/Memory overhead.
- No bandwidth overhead (updates are not shared between routers).
- Granular control on how traffic is routed.
- secure because the routers are managed statically.

Disadvantages of static routing
- Infrastructure changes must be manually adjusted.
- No "dynamic" fault tolerance if a link goes down.
- Unsuitable for large network.

4.8.3 Adaptive Routing

- It is also called *dynamic routing*.
- These algorithms change their routing decisions to reflect changes in the topology and in traffic as well.
- These get their routing information from adjacent routers or from all routers.
- If a change occurs in the network, the dynamic routing protocols automatically inform all routers about the events e.g. RIP V1, RIP V2, OSPF, EIGRP.

Using Dynamic Routing Protocols
- Dynamic routing protocols help the network administrator manage the time-consuming and exacting process of configuring and maintaining static routes.
- Imagine maintaining the static routing configurations for the seven routers as shown in figure **??**.

- What if the company grew and now has four regions and 28 routers to manage, as shown in Figure above? What happens when a link goes down? How do you ensure that redundant paths are available?
- Dynamic routing is the best choice for large networks like the one shown in Figure **??**.

Advantages of dynamic routing
- Independent of the network size.
- Suitable for all the topologies.
- Will dynamically choose a different (or better) route if a link goes down.
- Ability to load balance between multiple links.

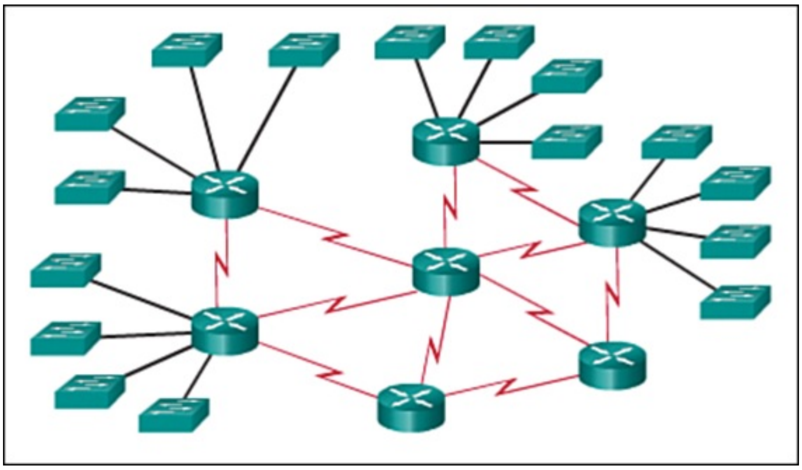


Figure 4.15: A network example.

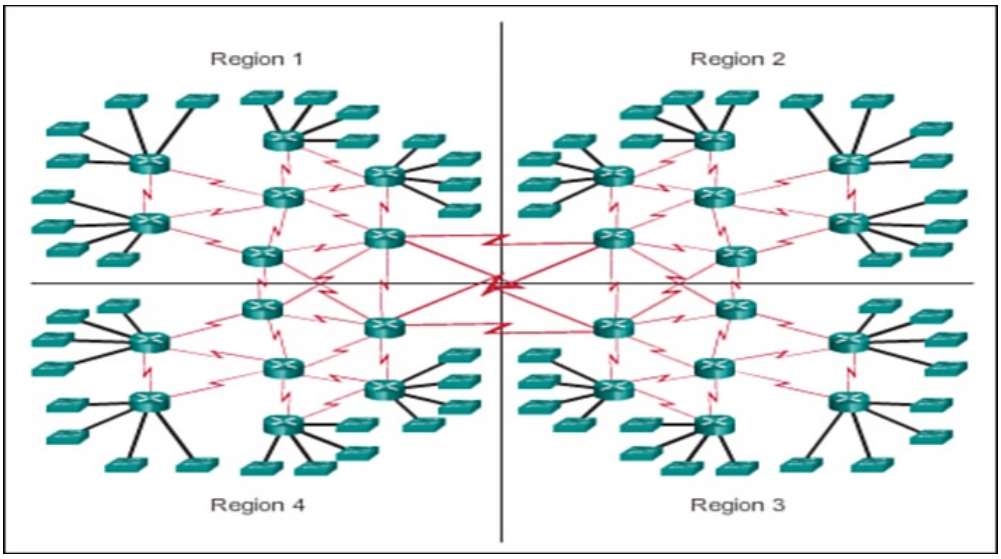


Figure 4.16: A network example.

Disadvantages of dynamic routing

- Can be more complex to initially implement.

- Updates are shared between routers, thus consuming bandwidth.

- Routing protocols put additional load on router CPU/RAM and link bandwidth.

- Route depends on the current topology.

- The broadcast and multicasting of routing updates make it less secure.

4.9 Routing Protocols

- Routing protocols are used to facilitate the exchange of routing information between routers.

- A routing protocol is a set of processes, algorithms, and messages that are used to exchange routing information and populate the routing table with the routing protocol's

Table 4.2: Static vs Dynamic routing

Static Routing	**Dynamic Routing**
Routers are configured manually, and routing table is also created manully.	Router and its routing tables are created automaticaly.
Routers does not reflect change in topology or traffic.	Routers reflect change in topolgy or traffic.
It does not employ complex algorithm.	It involves complex algorithm for calculating best path.
It is useful for small network consisting of few hosts.	It is suitable for large network having number of hosts.
When link fails, routing is discontinued.	When link fails, routing is not disrupted. Alternative route is chosen.
It does not involve protocols.	It involves different protocols like RIP, EIGRP, BGP etc.
It consumes less resources.	It consumes huge resources like bandwidth, memory etc..

choice of best paths.

- The purpose of dynamic routing protocols includes:

- Discovery of remote networks.
- Maintaining up-to-date routing information.
- Choosing the best path to destination networks.
- Ability to find a new best path if the current path is no longer available.

- The main components of dynamic routing protocols include:

- **Data structures**: Routing protocols typically use tables or databases for their operations. This information is kept in RAM.
- **Routing protocol messages**: Routing protocols use various types of messages to discover neighboring routers, exchange routing information, and perform other tasks to learn and maintain accurate information about the network.
- **Algorithm**: An algorithm is a finite list of steps used to accomplish a task. Routing protocols use algorithms for facilitating routing information and for best path determination

4.9.1 The Role of Dynamic Routing Protocols

- Routing protocols allow routers to dynamically share information about remote networks and automatically add this information to their own routing tables.
- Routing protocols determine the best path, or route, to each network. That route is then added to the routing table.
- A primary benefit of dynamic routing protocols is that routers exchange routing information when there is a topology change. This exchange allows routers to automatically learn about new networks and also to find alternate paths when there is a link failure to a current network.

Note:

- Compared to static routing, dynamic routing protocols require less administrative overhead.
- However, the expense of using dynamic routing protocols is dedicating part of a router's resources for protocol operation, including CPU time and network link bandwidth.
- Despite the benefits of dynamic routing, static routing still has its place.
- There are times when static routing is more appropriate and other times when dynamic routing is the better choice. Networks with moderate levels of complexity may have both static and dynamic routing configured.

4.9.2 Types of Dynamic Routing Protocols

- Dynamic Routing protocols can be classified into different groups according to their characteristics. Specifically, routing protocols can be classified by their:

- **Purpose**: Interior Gateway Protocol (IGP) or Exterior Gateway Protocol (EGP)
- **Operation**: Distance vector protocol, link-state protocol, or path-vector protocol
- **Behavior**: Classful (legacy) or classless protocol

- The ***classful routing protocols***, RIPv1 and IGRP, are legacy protocols and are only used in older networks. These routing protocols have evolved into the ***classless routing protocols***, RIPv2 and EIGRP, respectively. Link-state routing protocols are classless by nature.

4.9.3 IGP and EGP Routing Protocols

- An **autonomous system (AS)** is a collection of routers under a common administration such as a company or an organization. An AS is also known as a routing domain, which basically means that all routers sharing the same routing table information are in the same AS. Typical examples of an AS are a company's internal network and an ISP's

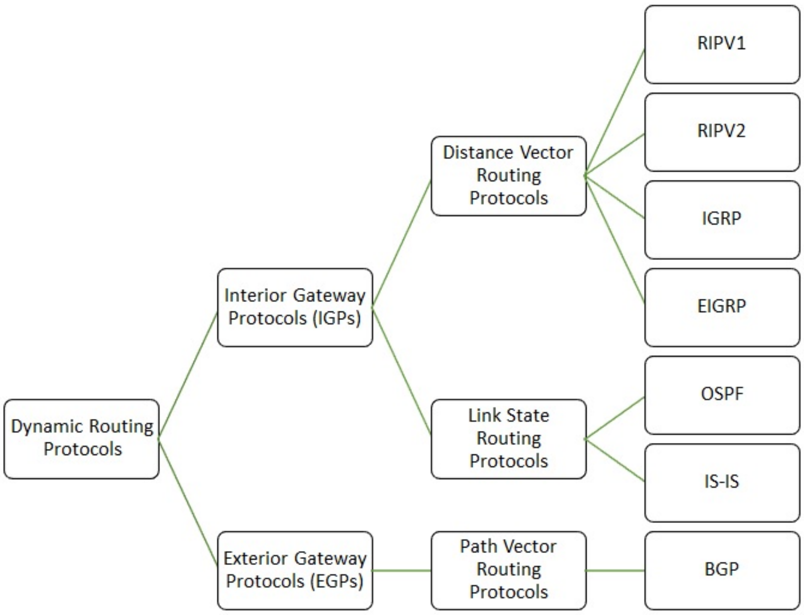

Dynamic Routing Protocols
Interior Gateway Protocols (IGPs)
Exterior Gateway Protocols (EGPs)
Distance Vector Routing Protocols
Link State Routing Protocols
Path Vector Routing Protocols
RIPV1
RIPV2
IGRP
EIGRP
OSPF
IS-IS
BGP

Figure 4.17: Types of routing protocols.

network. - The Internet is based on the AS concept; therefore, two types of routing protocols are required.

1. **Interior Gateway Protocols (IGP)**: Used for routing within an AS. It is also referred to as intra-AS routing. Companies, organizations, and even service providers use an IGP on their internal networks. IGPs include RIP, EIGRP, OSPF, and IS-IS.

2. **Exterior Gateway Protocols (EGP)**: Used for routing between autonomous systems. It is also referred to as inter-AS routing. Service providers and large companies may interconnect using an EGP. The Border Gateway Protocol (BGP) is the only currently viable EGP and is the official routing protocol used by the Internet

Note:

Because BGP is the only EGP available, the term EGP is rarely used; instead, most engineers simply refer to BGP.

4.9.4 Metrics

- Router metrics are metrics used by a router to make routing decisions.

- Metric is the cost assigned for passing through a network.

- The total metric of a particular route is equal to the metrics of networks that comprise the route.

- A router chooses the route with smallest metric.

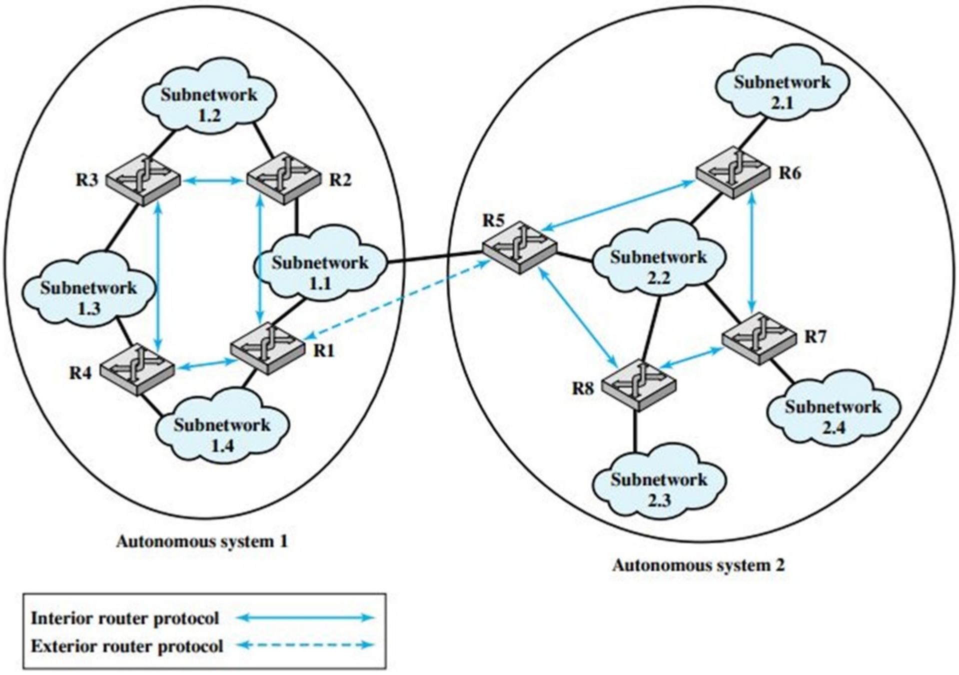


Figure 4.18: Autonomous system.

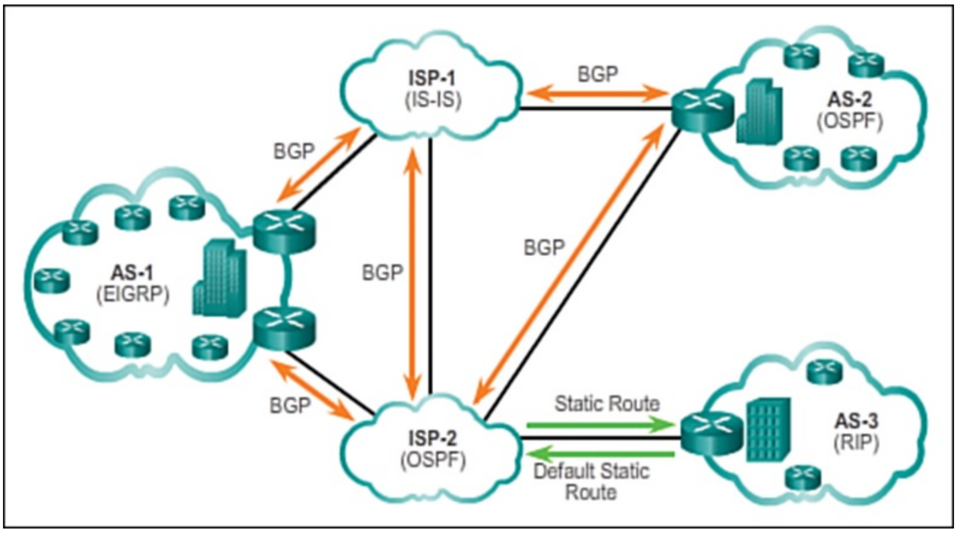


Figure 4.19: Simple scenarios highlighting the deployment of IGPs, BGP, and static routing.

4.9.5 Distance Vector Routing Protocols

- Distance vector means that routes are advertised by providing two characteristics:

i. **Distance**: Identifies how far it is to the destination network and is based on a metric such as the hop count, cost, bandwidth, delay, and more

ii. **Vector**: Specifies the direction of the next-hop router or exit interface to reach the destination

- For example, in Figure **??**, router R1 knows that the distance to reach network 172.16.3.0/24

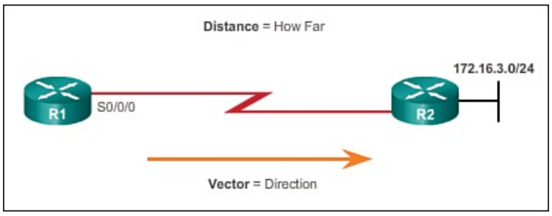


Figure 4.20: The meaning of distance.

is one hop and that the direction is out of the interface Serial 0/0/0 toward router R2.

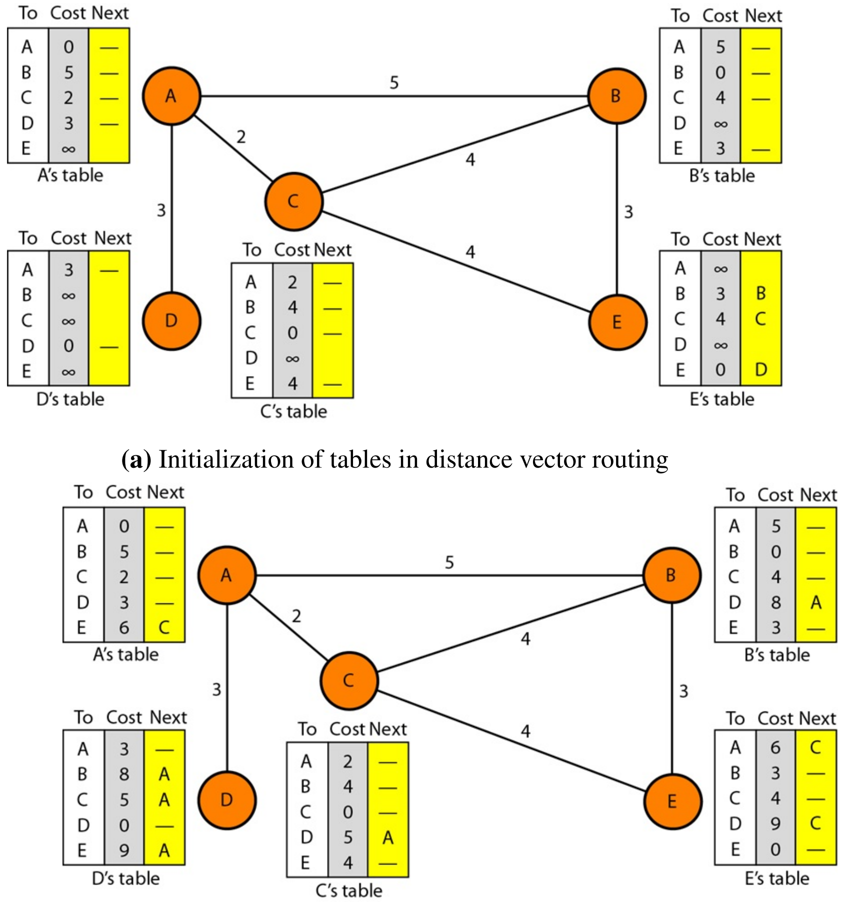
- A router using a distance vector routing protocol does not have the knowledge of the entire path to a destination network.
- Distance vector protocols use routers as sign posts along the path to the final destination.
- The only information a router knows about a remote network is the distance or metric to reach that network and which path or interface to use to get there.
- Each router shares its routing table with its immediate neighbors periodically and when there is a change.
- It used Bellman Ford Algorithm for making routing tables.
- Distance vector routing protocols do not have an actual map of the network topology.
- There are four distance vector IPv4 IGPs:

i. **RIPv1**: First generation legacy protocol
ii. **RIPv2**: Simple distance vector routing protocol
iii. **IGRP**: First generation Cisco proprietary protocol (obsolete and replaced by EIGRP)
iv. **EIGRP**: Advanced version of distance vector routing

Figure **??** shows an example of how metrics for routing table is calculated in distance vector routing.

4.9.5.1 Routing Information Protocol (RIP)

- RIP is true distance vector routing protocol.
- It is an intra AS routing protocol.
- RIP sends the complete routing table out of all active interfaces every 30 seconds.
- It uses hop-count as routing metric. It maximum allowable hop count is 15 by default, so a destination of 16 would be unreachable.
- RIP works okey in very small networks, but inefficient on large network with slow WAN links or on networks with a large number of routers installed and completely useless on networks that have links with variable bandwidth.



A's table

To	Cost	Next
A	0	—
B	5	—
C	2	—
D	3	—
E	∞	

B's table

To	Cost	Next
A	5	—
B	0	—
C	4	—
D	∞	
E	3	—

D's table

To	Cost	Next
A	3	—
B	∞	
C	∞	
D	0	—
E	∞	

C's table

To	Cost	Next
A	2	—
B	4	—
C	0	—
D	∞	
E	4	—

E's table

To	Cost	Next
A	∞	
B	3	B
C	4	C
D	∞	
E	0	D

(a) Initialization of tables in distance vector routing

A's table

To	Cost	Next
A	0	—
B	5	—
C	2	—
D	3	—
E	6	C

B's table

To	Cost	Next
A	5	—
B	0	—
C	4	—
D	8	A
E	3	—

D's table

To	Cost	Next
A	3	—
B	8	A
C	5	A
D	0	—
E	9	A

C's table

To	Cost	Next
A	2	—
B	4	—
C	0	—
D	5	A
E	4	—

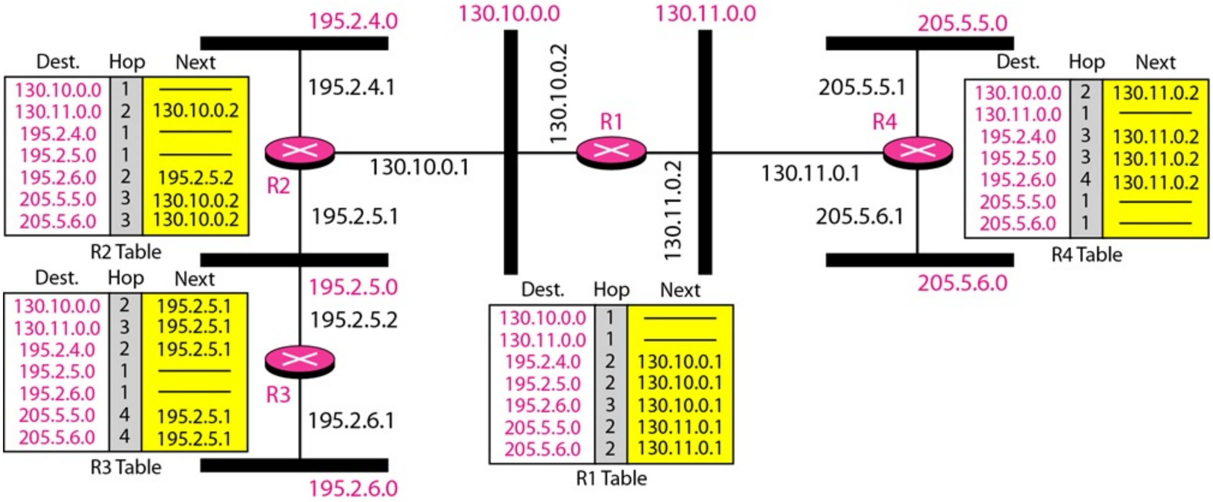
E's table

To	Cost	Next
A	6	C
B	3	—
C	4	—
D	9	C
E	0	—

(b) Updating routing tables.

Figure 4.21: Distance vector routing table calculation example.

- RIP uses the User Datagram Protocol (UDP) as its transport protocol, and is assigned the reserved port number 520.
- It has two version: *RIP version 1* and *RIP version 2*. .
- RIP version 1 is the classful routing protocol, as it doesn't send updates with subnet mask information while RIP version 2 classless routing protocol as it send updates with subnet mask information with its route updates.



Dest.	Hop	Next
130.10.0.0	1	———
130.11.0.0	2	130.10.0.2
195.2.4.0	1	———
195.2.5.0	1	———
195.2.6.0	2	195.2.5.2
205.5.5.0	3	130.10.0.2
205.5.6.0	3	130.10.0.2

R2 Table

Dest.	Hop	Next
130.10.0.0	2	195.2.5.1
130.11.0.0	3	195.2.5.1
195.2.4.0	2	195.2.5.1
195.2.5.0	1	———
195.2.6.0	1	———
205.5.5.0	4	195.2.5.1
205.5.6.0	4	195.2.5.1

R3 Table

Dest.	Hop	Next
130.10.0.0	1	———
130.11.0.0	1	———
195.2.4.0	2	130.10.0.1
195.2.5.0	2	130.10.0.1
195.2.6.0	3	130.10.0.1
205.5.5.0	2	130.11.0.1
205.5.6.0	2	130.11.0.1

R1 Table

Dest.	Hop	Next
130.10.0.0	2	130.11.0.2
130.11.0.0	1	———
195.2.4.0	3	130.11.0.2
195.2.5.0	3	130.11.0.2
195.2.6.0	4	130.11.0.2
205.5.5.0	1	———
205.5.6.0	1	———

R4 Table

Figure 4.22: Example of a domain using RIP.

4.9.6 Link-State Routing Protocols

- Link state protocols are also called shortest-path-first protocols. Link state routing protocols have a complete picture of the network topology. Hence they know more about the whole network than any distance vector protocol.
- Three separate tables are created on each link state routing enabled router. One table is used to hold details about directly connected neighbors, one is used to hold the topology of the entire internetwork and the last one is used to hold the actual routing table.
- Link state protocols send information about directly connected links to all the routers in the network. Examples of Link state routing protocols include OSPF - Open Shortest Path First and IS-IS - Intermediate System to Intermediate System.
- It makes use of Dijkstra's Algorithm for calculating path.
- • It makes use of more memory and large computations.

Operation:

Each link state enabled router must perform the following:

- Discover its neighbors and build its neighbor table.

- Measure link cost.
- Construct and send a routing update telling all it has learned to all routers in the network.
- Apply the Dijkstra algorithm to construct the shortest path to all possible destinations.

So exactly how does a link state protocol work? All routers will complete the following generic link state routing process to reach a state of convergence:

1. **Each router learns about its own links, its own directly connected networks**. This is done by detecting that an interface is in the up state.

2. **Each router is responsible for meeting its neighbors on directly connected networks**. Link state routers do this by exchanging Hello packets with other links-state routers on directly connected networks.

3. **Each router builds a Link-State Packet (LSP) containing the state of each directly connected link.** This is done by recording all the pertinent information about each neighbour, including neighbor ID, link type, and bandwidth.

4. **Each router floods the LSP to all neighbors, who then store all LSPs received in a database**. Neighbors then flood to their neighbors unitl all routers in the area have received the LSPs. Each router stores a copy of each LSP received from its neighbors in a local database.

5. **Each routers uses the database to construct a complete map of the topology and computes the best path to each destination network**. Link having a road map, the router now has a complete map of all destinations in the topology and the routes to reach them. The SPF algorithm is used to construct the map of the topology and to determine the best path to each network.

- RIP-enabled routers send periodic updates of their routing information to their neighbors. Link-state routing protocols do not use periodic updates. After the network has converged, a link-state update is only sent when there is a change in the topology. For example, in Figure below, the link-state update is sent when the 172.16.3.0 network goes down.
- Link-state protocols work best in situations where:

- The network design is hierarchical, usually occurring in large networks
- Fast convergence of the network is crucial
- The administrators have good knowledge of the implemented link-state routing protocol

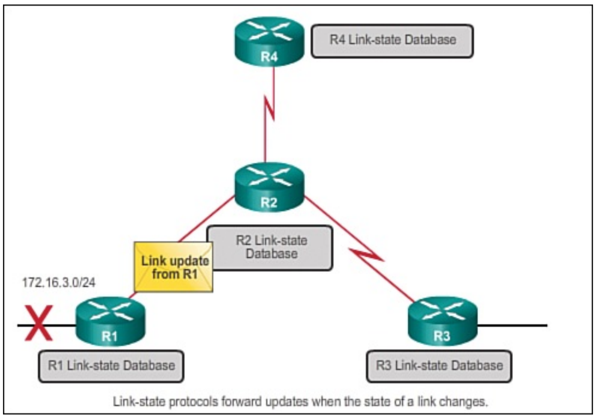


Figure 4.23: Link State protocol forward updates when the state of a link changes.

- There are two link-state IPv4 IGPs:

- **OSPF**:Popular standards-based routing protocol.
- **IS-IS**: Popular in internet service provider networks.

Advantages of Link State Routing Protocol Building topological map:

- Link-state routing protocol creates a topological map, or SPF tree of the network topology. Distance vector routing protocol do not have a topological map of the network.

Faster Convergence:

- When receiving a link-state packet (LSP), link state routing protocols immediately flood the LSP out all interfaces except for the interface from which the LSP was received. This way, it achieve the faster convergence.

Event Driven Updates:

- After the initial flooding of LSPs, link-state routing procols only send out ann LSP when there is a change in the topology. The LSP contains only the information regarding the affected link. Unlike some distance vector routing protocols, link-state routing protocols do not send periodic updates.

Flow based routing:

- A flooding algorithm is an algorithm for distributing material to every part of a conneced network. It implemented by OSPF.

4.9.6.1 Open Shortest Path First (OSEP) Protocol

- OSPF is an open standard routing protocol that's been implemented by a wide variety network vendor's, including CISCO.

- OSPF is an interior as well as classless routing protocol based on the link state routing,

operating within a single autonomous system.
- It uses Dijkstra algorithm to calculate best shortest path and populate the routing table with the resulting best paths.
- To handle routing efficiently, OSPF divides an autonomous system into areas. An area is a collection of networks, hosts, and routers all contained within an autonomous system. Each area has an identification number.

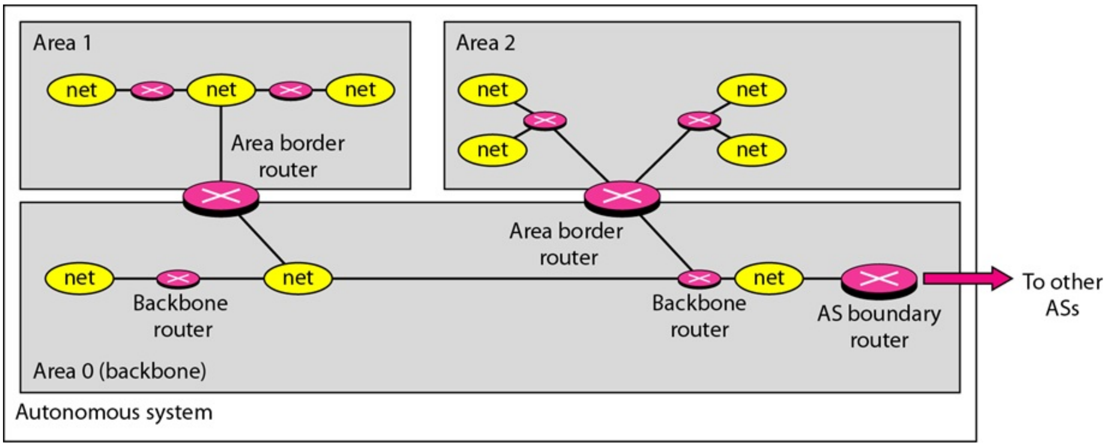


Figure 4.24: Example of a domain using OSPF.

- It uses link bandwidth as path metric.
- It includes the ability to carry subnet information during route updates, thus supporting Classless Inter- Domain Routing (CIDR).
- In addition, it supports authentication.
- It supports both IP and IPv6 routed protocols.

Here's list that summarizes some of OSPF's best features:

- Allows for creations of areas and autonomous systems.
- Minimizes routing update traffic.
- Is highly flexible, versatile, and scalable.
- Supports VLSM/CIDR.
- Offers an unlimited hop count.
- Is open standard and supports multi-vendor deployment.
- Support IPv4 and IPv6 routed protocols.

OSPF operation

OSPF operation is basically divided into these three categories:

- Neighbor and adjacency initialization.

- LSA flooding.
- SPF tree calculation.

Neighbor and adjacency initialization

- When OSPF is initialized on a router, the router allocates memory for it, as well as for the maintenance of both neighbor and topology tables.

- Once the router determines which interfaces have been configured for OSPF, it will then check to see if they're active and begin sending Hello packets.

- The Hello protocol is used to discover neighbors, establish adjacencies, and maintain relationships with other OSPF routers.

- It uses address 244.0.0.5.

- By default, OSPF Hello packets are sent:

- every 10 second on multi-access and point-to-point segments.
- every 30 second on non-broadcast multi-access segment (Frame Relay, X.25, ATM)

LSA Flooding

- A *Link State Advertisement (LSA)* is an OSPF data packet containing link-state and routing information that's shared among OSPF routers.

- An OSPF router will exchange LSA packets only with routers to which it has established adjacencies.

- LSA flooding is the method OSPF uses to share routing information.

- Via Link State Updates (LSU)'s packets, LSA information containing link-state data is shared with all OSPF routers within an area.

- The network topology is created from the LSA updates, and flooding is used so that all OSPF routers have the same topology map to make SPF calculations with.

- Once the LSA updates have been flooded throughout the network, each recipient must acknowledge that the flooded update has been received. It's also important for recipients to validate the LSA update.

SPF Tree Calculation

- Within an area, each router calculates the best/shortest path to every network in that same area. This calculation is based upon the information collected in the topology database and an algorithm called shortest path first (SPF).

- Picture each router in an area constructing a tree—much like a family tree—where the router is the root and all other networks are arranged along the branches and leaves. This is the shortest path tree used by the router to insert OSPF routes into the routing table.

- It's important to understand that this tree contains only networks that exist in the same area as the router itself does. If a router has interfaces in multiple areas, then separate trees will be constructed for each area.

4.9.7 OSPF and RIP comparision

Table 4.3: OSPF and RIP comparision

Characteristics	**OSPF**	**RIPv2**	**RIPv1**
Type of Protocol	Link State	Distance Vector	Distance Vector
Classless support	Yes	Yes	No
VLSM support	Yes	Yes	No
Auto-summarizatio	No	Yes	Yes
Manual summarization	Yes	Yes	No
Noncontiguous support	Yes	Yes	No
Route propagation	Multicast on change	Periodic multicast	Periodic broadcast
Path metric	Bandwidth	Hops	Hops
Hop count limit	None	15	15
Convergence	Fast	Slow	Slow
Peer authentication	Yes	Yes	No
Hierarachical Network requirment	Yes (using areas)	No	No
Updates	Event triggered	Periodic	Periodic
Route computation	Dijkstra	Bellman-ford	Bellman-ford

4.9.8 Flooding

- When a routing algorithm is implemented, each router must make decisions based on local knowledge, not the complete picture of the network. A simple local technique is **flooding**, in which every incoming packet is sent out on every outgoing line except the one it arrived on.
- Flooding obviously generates vast numbers of duplicate packets, in fact, an infinite number unless some measures are taken to damp the process.
- One such measure is to have a hop counter contained in the header of each packet that is decremented at each hop, with the packet being discarded when the counter reaches zero.

Table 4.4: Distance Vector vs. Link-State

Distance Vector	Link State
Sends the entire routing table information to the neigbhbor routers.	Sends only the link state information to neighbor through flooding.
View the network from the perspective of neighbor.	Gets common view of entire network topology.
Has frequent and periodic upadates.	Has even triggered updates.
Slow convergence.	Fast Convergence.
Susceptible of routing loop.	Less susceptible to routing loop.
Easy to configure.	Can be harder to configure.
Requires less memory and processing power of routers.	Requires more processing power and memory.
Example: RIP, IGRP.	Example: OSPF, IS-IS

  - Flooding with a hop count can produce an exponential number of duplicate packets as the hop count grows and routers duplicate packets they have seen before.

- A better technique for damming the flood is to have routers keep track of which packets have been flooded, to avoid sending them out a second time.
  - One way to achieve this goal is to have the source router put a sequence number in each packet it receives from its hosts.
  - Each router then needs a list per source router telling which sequence numbers originating at that source have already been seen.
  - If an incoming packet is on the list, it is not flooded.

Advantage of Flooding
- The main advantage of flooding is increased reliability provided by this routing method. Since the message will be sent at least once to every host it is almost guaranteed to reach its destination.
- In addition, the message will reach the host through the shortest possible path.

Disadvantages of Flooding:
There are disadvantages of flooding.
- It is very wasteful in terms of the network total bandwidth. While a message may only have one destination it has to be sent to every host. This increase the maximum load placed upon the network.

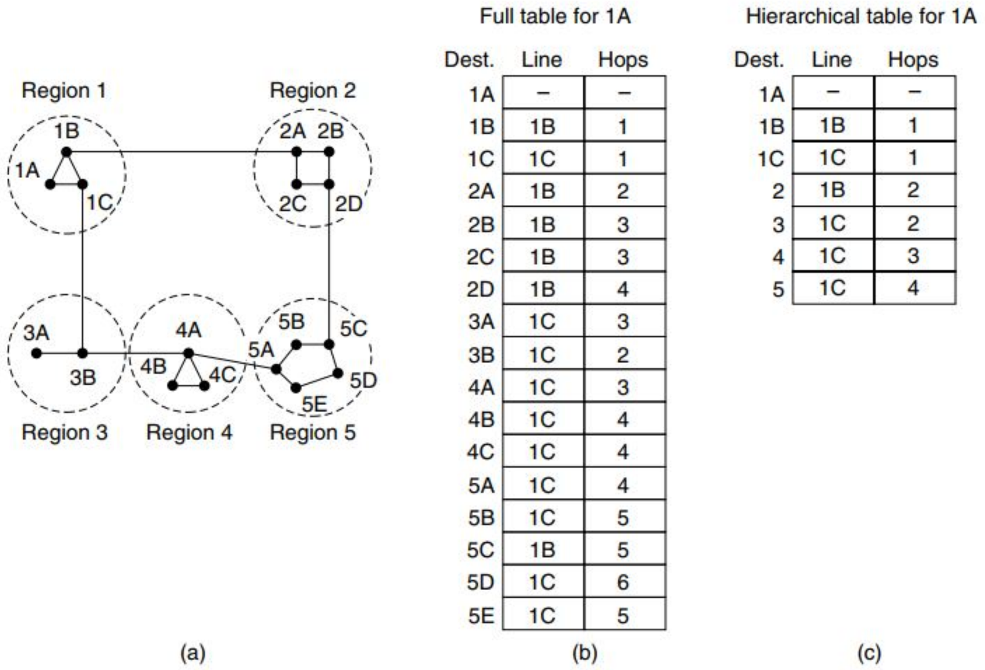
- Message can also become duplicated in the network further increasing the load on the network's bandwidth as well as requiring an increase in processing complexity to disregard duplicate messages.

4.10 Hierarchical Routing

- As network grow in size, the router routing tables grow proportionally.
  - router memory is consumed by ever-increasing tables.
  - more CPU time is needed to scan them, and
  - more bandwidth is needed to send status reports about them.
- To resolve this, the routing will have to be done hierarchically, as it is in the telephone network.
- When hierarchical routing is used, the routers are divided into what we will call regions.
  - Each router knowns all the details about how to route packets to destinations within its own region.
  - Knows nothing about the internal structure of other regions.
- For huge networks, a two-level hierarchy may be insufficient
  - it may be necessary to group the regions into clusters,
  - the clusters into zones,
  - the zones into groups, and so on.

- Figure **??** gives a quantitative example of routing in a two-level hierarchy with five regions.
- The full routing table for router 1A has 17 entries;
- When routing is done hierarchically, there are entries for all the local routers, as before, but all other regions are condensed into a single router, so all traffic for region 2 goes via the 1B-2A line, but the rest of the remote traffic goes via the 1C-3B line.
- Hierarchical routing has reduced the table from 17 to 7 entries. Unfortunately, the penalty of less memory space is paid off in increased in path length.

Example:
Consider a network with 720 routers. If there is no hierarchy, each router needs 720 routing table entries. If the network is partitioned into 24 regions of 30 routers each,



Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

Figure 4.25: A quantitative example of routing in a two-level hierarchy with five regions.

each router needs 30 local entries plus 23 remote entries for a total of 53 entries. If a three-level hierarchy is chosen, with 8 clusters each containing 9 regions of 10 routers, each router needs 10 entries for local routers, 8 entries for routing to other regions within its own cluster, and 7 entries for distant clusters, for a total of 25 entries.

4.11 Broadcast Routing

- In some applications, hosts need to send messages to many or all other hosts. For example, a service distributing weather reports, stock market updates, or live radio programs might work best by sending to all machines and letting those that are interested read the data. Sending a packet to all destinations simultaneously is called **broadcasting**.
- Broadcasting routing requires the router to have a complete list of all destination. This routing is slow and requires much bandwidth.
- It can use reverse path forwarding or **sink tree** (i.e **spanning tree**) to improve the efficiency.
- The principal advantage of reverse path forwarding is that it is efficient while being easy to implement. It sends the broadcast packet over each link only once in each direction, just as in flooding, yet it requires only that routers know how to reach all destinations, without needing to remember sequence numbers (or use other mechanisms to stop the flood) or list all destinations in the packet.

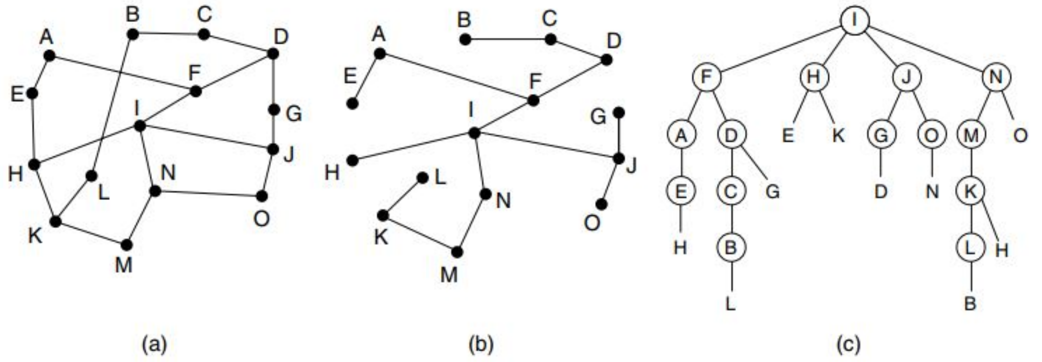


Figure 4.26: Reverse path forwarding. (a) A network (b) A sink tree. (c) The tree built by reverse path forwarding.

4.12 Multicast Routing

- Sending a message to such a group is called multicasting, and the routing algorithm used is called multicast routing.
- All multicasting schemes require some way to create and destroy groups and to identify which routers are members of a group.
- Each group is identified by a multicast address and that routers know the groups to which they belong.
- Broadcast spanning tree is pruned to form multicast spanning tree and there are various way of it. Different multicast groups can have different spanning trees.
- Figure **??** shows an example of network having two groups, 1 and 2, and their spanning trees.

Applications of Multicast

- Video/audio conference
- IP TV, Video on Demand
- Advertisement, Stock, Distance learning
- Distributed interactive gaming or simulations
- Voice-over-IP
- Synchronizing of distributed database, websites

4.13 Anycast Routing

- In anycast, a packet is delivered to the nearest member of a group. Schemes that find these paths are called anycast routing.
- For example, anycast is used in the Internet as part of DNS.

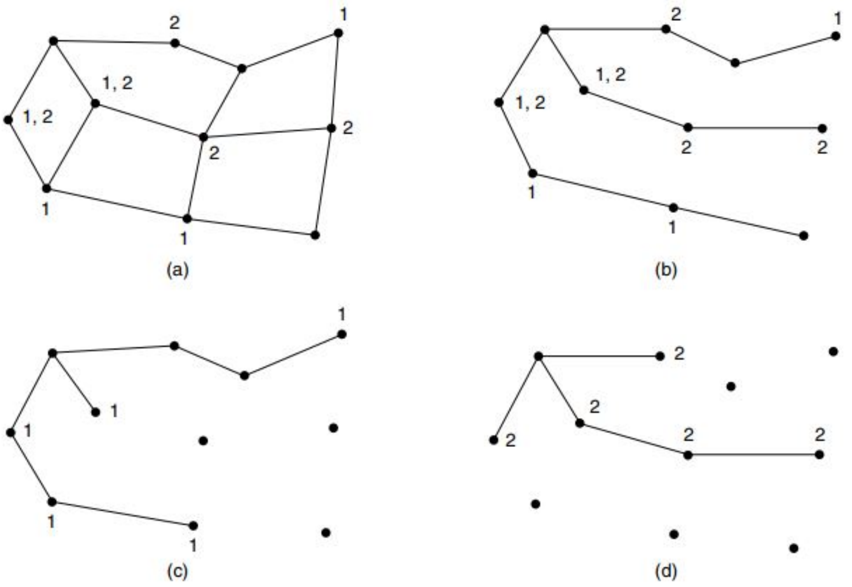


Figure 4.27: (a) A network (b) A spanning tree for the left-most router. (c) A multicast tree for group 1. (d) A multicast tree for group 2.

- Regular distance vector and link state routing can produce anycast routes.

- For example:
- we want to anycast to the members of group 1. They will all be given the address "1" instead of different addresses. Distance vector routing will distribute vectors as usual, and nodes will choose the shortest path to destination 1. This will result in nodes sending to the nearest instance of destination 1. The routes are shown in figure **??** a. This procedure works because the routing protocol does not realize that there are multiple instances of destination 1. That is, it believes that all the instances of node 1 are the same node, as in the topology shown in figure **??** b.

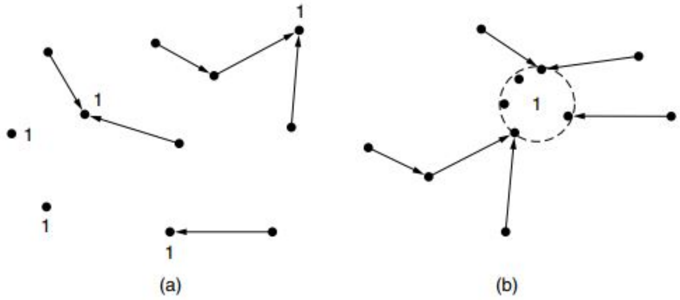


Figure 4.28: (a) Anycast of group 1 (b) Topology seen by routing protocol.

4.14 Hybrid Protocol

There are also routing protocols that are considered to be hybrid in the sense that they use aspects of both distance vector and link state protocols. **EIGRP - Enhanced Interior Gateway Routing Protocol** is one of those hybrid routing protocols.

4.15 Classful/Classless Routing Protocols

- The biggest distinction between classful and classless routing protocols is that classful routing protocols do not send **subnet mask** information in their routing updates.
- Classless routing protocols include subnet mask information in the routing updates.

4.16 Routing Protocol Characteristics

Routing protocols can be compared based on the following characteristics:

- **Speed of convergence**: Speed of convergence defines how quickly the routers in the network topology share routing information and reach a state of consistent knowledge. The faster the convergence, the more preferable the protocol.
- Routing loops can occur when inconsistent routing tables are not updated due to slow convergence in a changing network.
- **Scalability**: Scalability defines how large a network can become, based on the routing protocol that is deployed. The larger the network is, the more scalable the routing protocol needs to be.
- **Classful or classless (use of VLSM)**: Classful routing protocols do not include the subnet mask and cannot support variable-length subnet mask (VLSM). Classless routing protocols include the subnet mask in the updates. Classless routing protocols support VLSM and better route summarization.
- **Resource usage**: Resource usage includes the requirements of a routing protocol such as memory space (RAM), CPU utilization, and link bandwidth utilization. Higher resource requirements necessitate more powerful hardware to support the routing protocol operation, in addition to the packet forwarding processes.
- **Implementation and maintenance**: Implementation and maintenance describes the level of knowledge that is required for a network administrator to implement and maintain the network based on the routing protocol deployed.

4.17 Shortest Path Algorithm

- In this technique, the shortest path is choose to route the packet form source to destination.

- One way of measuring path length is the number of hops. Another metric is the geographical distance in kilometers.
- However, the shortest path is the fastest path rather than the path with the fewer hops. Shortest path is calculated as a function of different factors like distance, bandwidth, communication cost, delay, average traffic etc.
- Two common algorithms are used:

i. Dijkstra's Algorithm
ii. Bellman-Ford Algorithm

4.17.1 Dijkstra's Algorithm

- Dijkstra's algorithm is an algorithm for finding the shortest paths between nodes in a graph.
- This algorithm proceed in stages. After k^{th} stage, the least cost path is found; these nodes are in the set T.
- At stage $(k+1)$, the node not in T that has the shortest path from source node is added to T.
- As each node is added to T, its path from the source is defined.

The algorithm can be formally described as follows. *Define*:
N = set of nodes in the network
s = source node
T = set of nodes so far incorporated by the algorithm
w(i, j) = cost from node *i* to node *j*
w(i ,i) = 0
w(i, j) = ∞, if nodes are not directly connected.
w(i,j) = $\gneq$, if nodes are directly connected.
$L(n)$ = cost of the least-cost from node s to node n that is currently known to the algorithm; at termination, this is the least-cost path in the graph from s to n.

Step 1: Initialization
• T = s
• L(n) = w(s,n) for n$\neq$ s
Step 2: Get Next Node
• Find the neighboring node not in *T* that has the least-cost path from node s and incorporate that node into *T*.
• Also incorporate the edge contributes to the path. Mathematically expressed as:

$$\text{Find x} \notin \text{T such that L(x)} = \min_{j \notin T} L(j).$$

Add *x* to *T*; and find the least cost of edge for nodes in updated *T*.

Step 3: Update Least-Cost Paths

L(n) = min [*L(n), L(x) + w(x,n)*] for all *n* not belongs to *T*.

- Repeat step 2 and 3 until *T* = *N*
- The algorithm terminates when all nodes have been added to *T*.

Example:

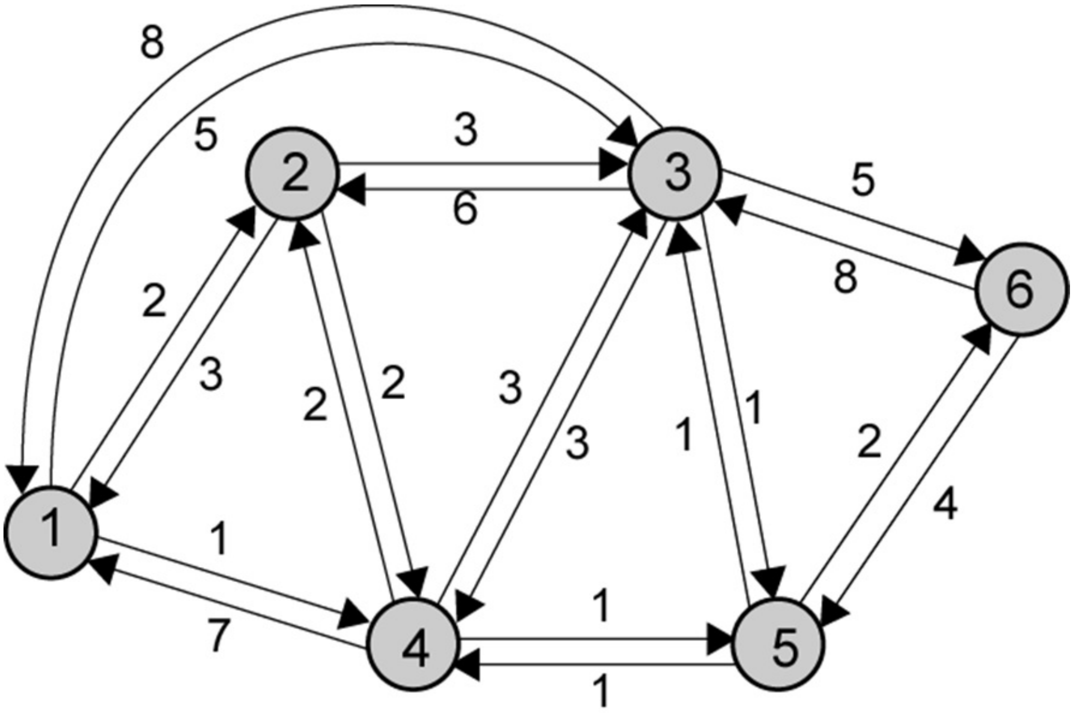


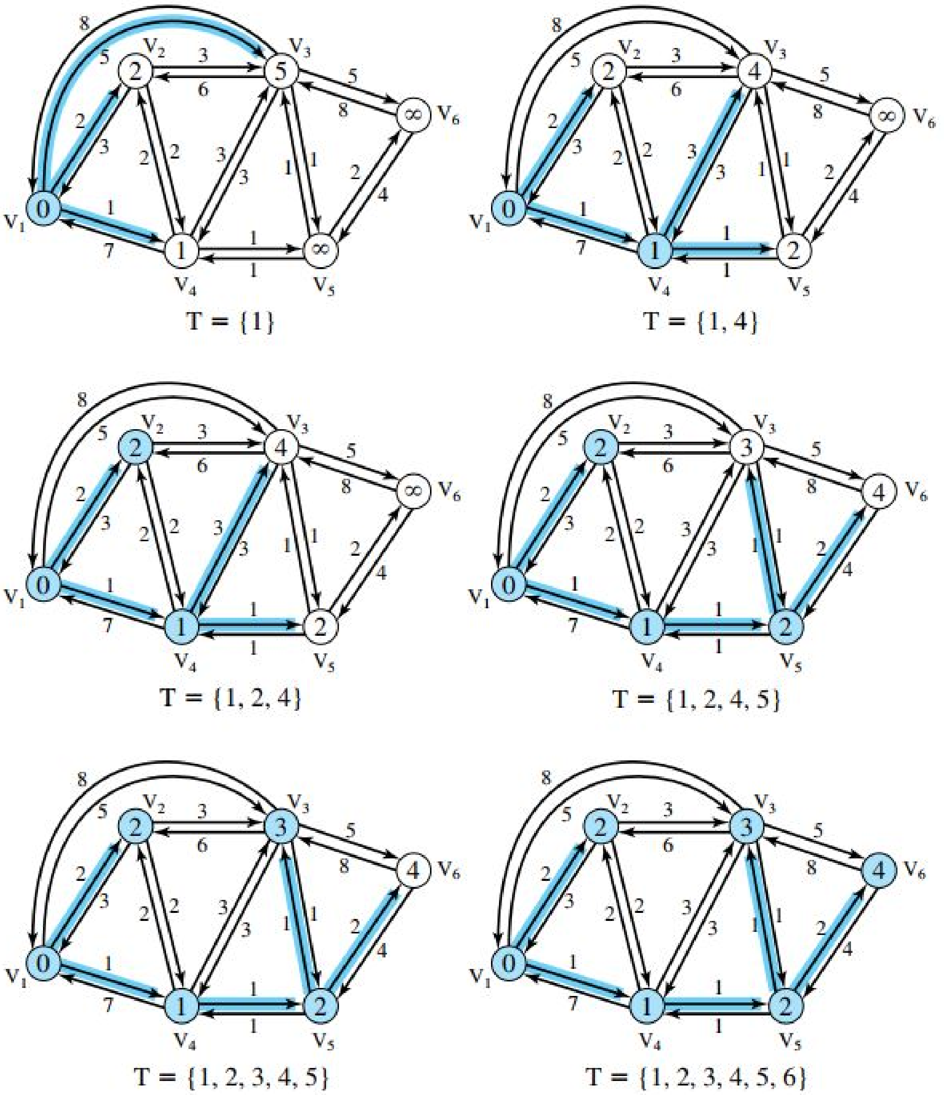
Figure 4.29: A network with cost values.

- In the figure **??**, the values in each circle are the current estimates of *L(x)* for each node *x*.
- A node is shaded when it is added to *T*.
- At each step the path to each node plus the total cost of that path is generated.
- After the final iteration, the least-cost path to each node and the cost of that path have been developed.

4.17.2 Bellman-Ford Algorithm

The algorithm can be formally described as follows. *Define*:

s = source node



T = {1}

T = {1, 4}

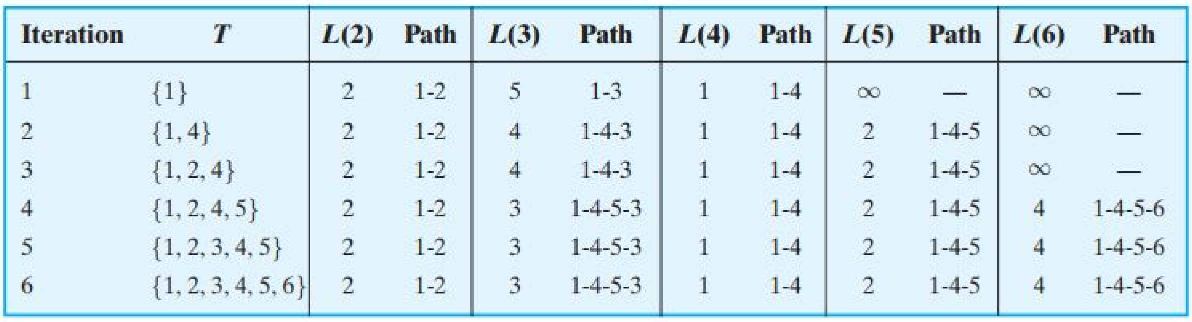
T = {1, 2, 4}

T = {1, 2, 4, 5}

T = {1, 2, 3, 4, 5}

T = {1, 2, 3, 4, 5, 6}

Figure 4.30: Dijkstra's Algorithm Applied to Graph of Figure **??**.



Iteration	*T*	*L*(2)	Path	*L*(3)	Path	*L*(4)	Path	*L*(5)	Path	*L*(6)	Path
1	{1}	2	1-2	5	1-3	1	1-4	∞	—	∞	—
2	{1, 4}	2	1-2	4	1-4-3	1	1-4	2	1-4-5	∞	—
3	{1, 2, 4}	2	1-2	4	1-4-3	1	1-4	2	1-4-5	∞	—
4	{1, 2, 4, 5}	2	1-2	3	1-4-5-3	1	1-4	2	1-4-5	4	1-4-5-6
5	{1, 2, 3, 4, 5}	2	1-2	3	1-4-5-3	1	1-4	2	1-4-5	4	1-4-5-6
6	{1, 2, 3, 4, 5, 6}	2	1-2	3	1-4-5-3	1	1-4	2	1-4-5	4	1-4-5-6

Figure 4.31: Dijkstra's Algorithm (s=1) for figure **??**.

w(i, j) = cost from node*i* to node *j*

w(i ,i) = 0

$w(i, j) = \infty$, if nodes are not directly connected.

$w(i,j) = \geq$, if nodes are directly connected.

h = maximum number of links in a path at the current stage of the algorithm.

$Ln(n)$ = cost of the least-cost path from node s to node n under the constraint of no more than h links.

1. Initialization

$$L_0(n) = \infty, \text{ for all } n \neq s$$

2. Update:

For each successive $h \geq 0$;

For each $n \neq s$, compute,

$$L_{h+1}(n) = \min_j [L_h(j) + w(j,n)]$$

Connect n with the predecessor node j that achieves the minimum, and eliminate any connection of n with a different predecessor node formed during an earlier iteration. The path from s to n terminates with the link from j to n.

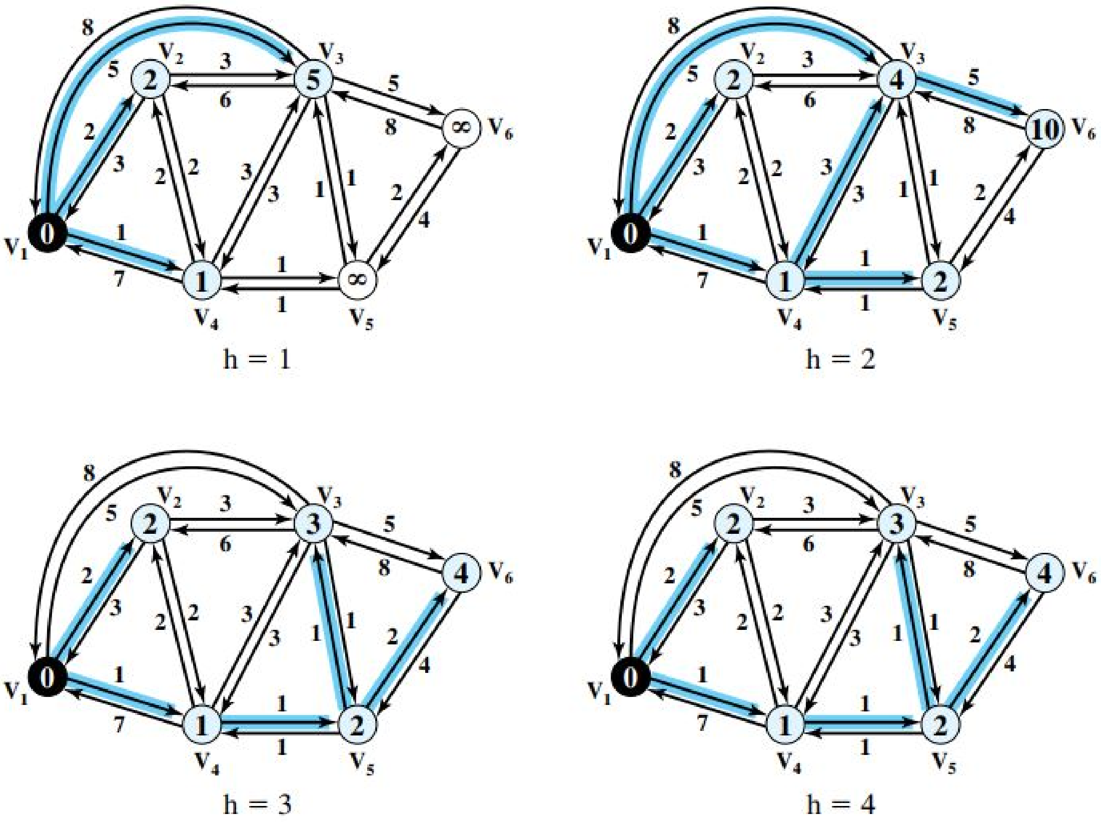
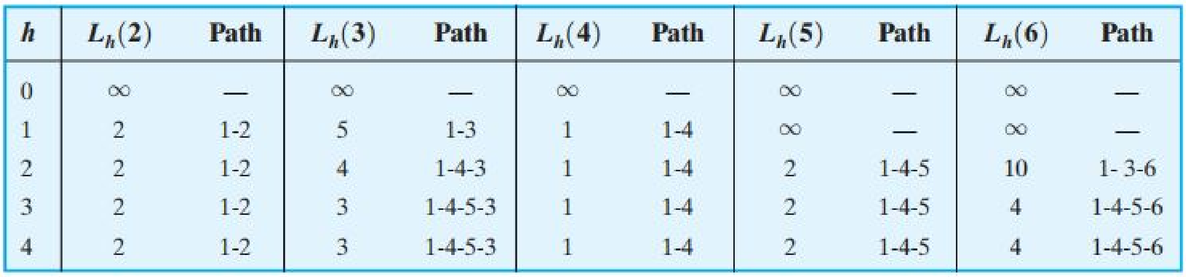


Figure 4.32: Bellman-Ford applied to graph of figure **??**.



h	$L_h(2)$	Path	$L_h(3)$	Path	$L_h(4)$	Path	$L_h(5)$	Path	$L_h(6)$	Path
0	∞	—	∞	—	∞	—	∞	—	∞	—
1	2	1-2	5	1-3	1	1-4	∞	—	∞	—
2	2	1-2	4	1-4-3	1	1-4	2	1-4-5	10	1- 3-6
3	2	1-2	3	1-4-5-3	1	1-4	2	1-4-5	4	1-4-5-6
4	2	1-2	3	1-4-5-3	1	1-4	2	1-4-5	4	1-4-5-6

Figure 4.33: Bellman-Ford Algorithm (s=1) applied to graph of figure **??**.

4.18 IPv4 Datagram

- The Internet Protocol version 4 (IPv4) is the delivery mechanism used by the TCP/IP protocols.

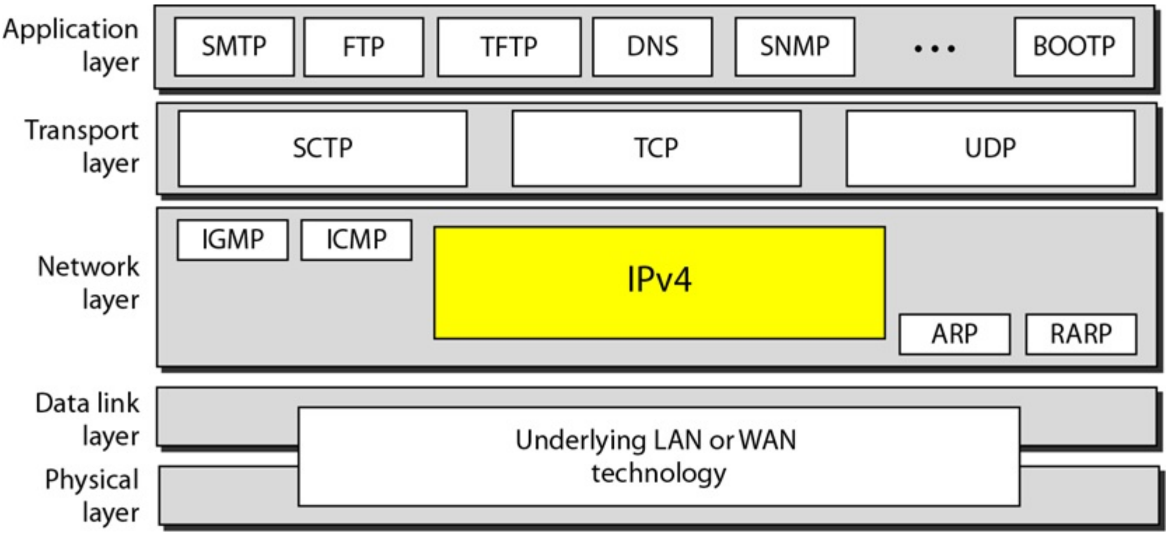


Figure 4.34: Position of IPv4 in TCP/IP protocol suite.

- Packets in IPv4 layer is called datagrams.
- A datagram is a variable length packet consisting of two parts: header and data.
- The header is 20 to 60 bytes in length and contains information essential to routing and delivery.
- Figure **??** shows the format of IPV4 datagram.

Version

- It is a 4-bit field that specifies the version of IP currently being used. It has value 4 for IPV4.

IP Header Length (HLEN)

- Length of header is 32 bit words.

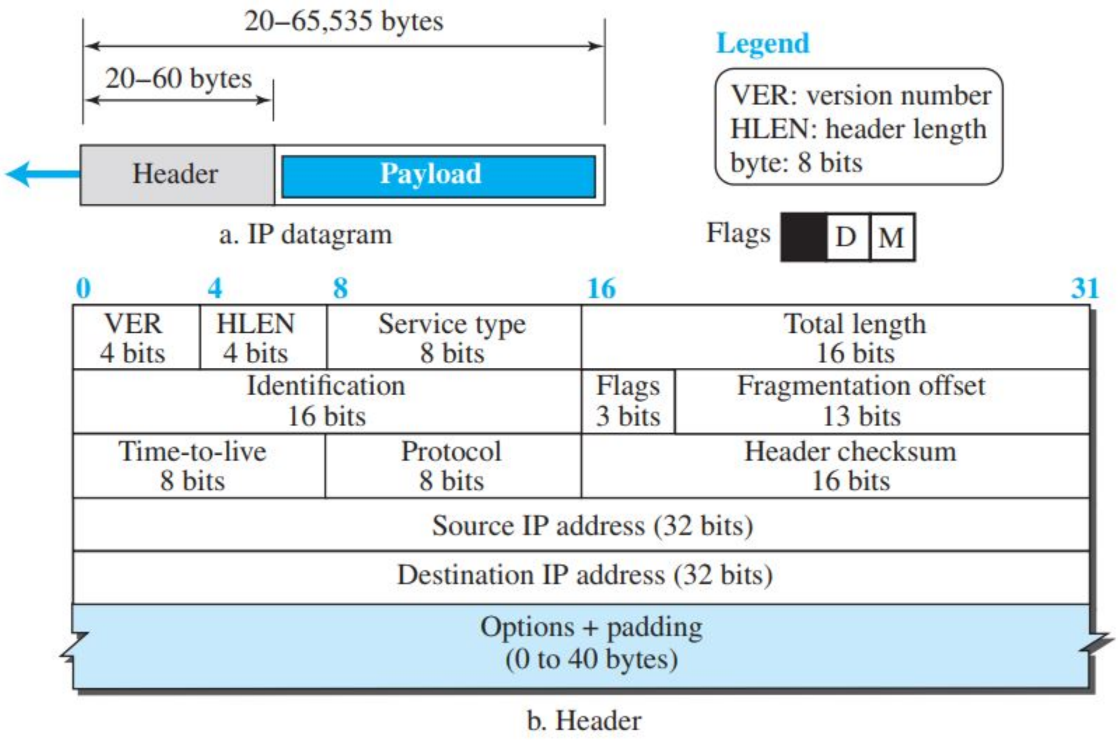


Figure 4.35: IPv4 datagram format.

- The minimum value of HLEN is 5 i.e. 0101, for a minimum header length of 20 bytes.

Service Type

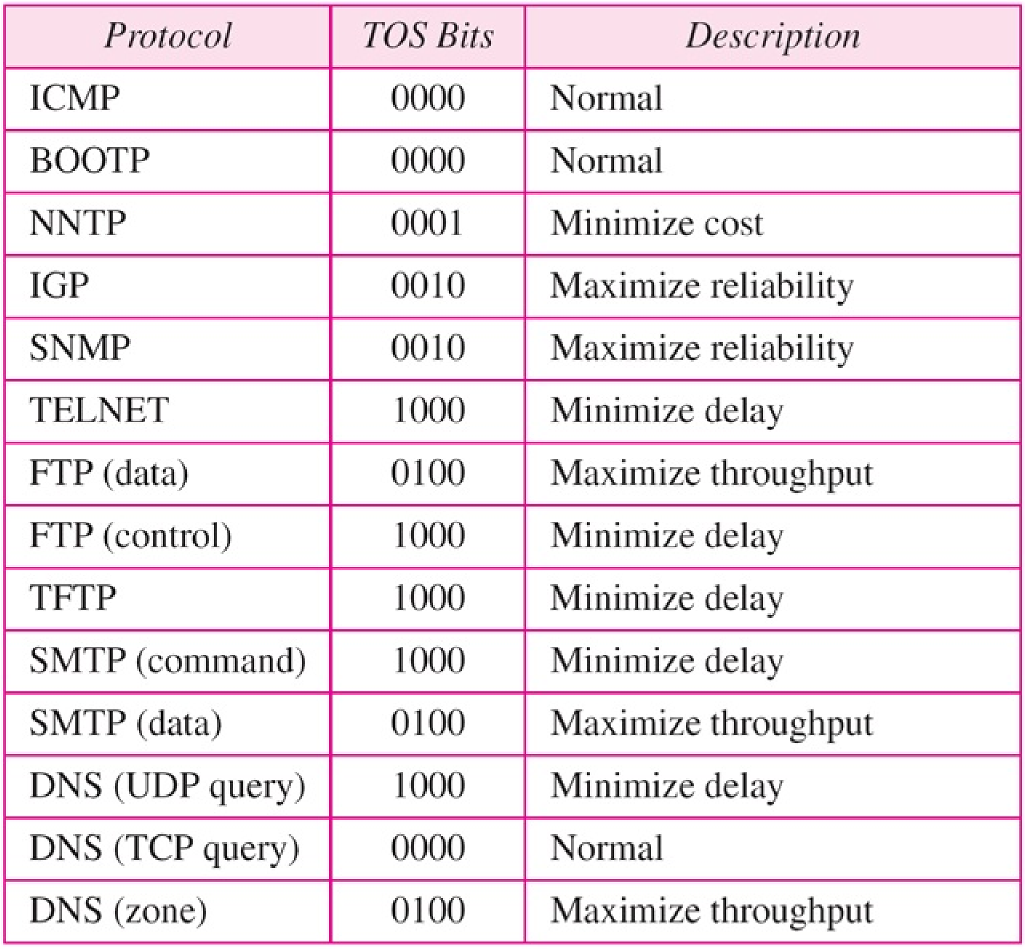
- Originally called *type of service (TOS)*.
- This 8-bit specify how datagram should be handled in its transmission through network.
- For example, datagrams particularly requiring low delay, high throughput, or reliability should be differentiated from each other.
- Now it is called differentiate service.

Total Length

- 16 bit field specifies the total length of entire IP datagram including data and header in bytes.
- Length of data = total length – (HLEN) x 4.

Identification

- This 16 bit number is used combined with the source address, destination address, and user protocol, is intended to identify a datagram uniquely.
- Thus, this number should be unique for the datagram's source address, destination



Protocol	*TOS Bits*	*Description*
ICMP	0000	Normal
BOOTP	0000	Normal
NNTP	0001	Minimize cost
IGP	0010	Maximize reliability
SNMP	0010	Maximize reliability
TELNET	1000	Minimize delay
FTP (data)	0100	Maximize throughput
FTP (control)	1000	Minimize delay
TFTP	1000	Minimize delay
SMTP (command)	1000	Minimize delay
SMTP (data)	0100	Maximize throughput
DNS (UDP query)	1000	Minimize delay
DNS (TCP query)	0000	Normal
DNS (zone)	0100	Maximize throughput

Figure 4.36: Default types of service.

address, and user protocol for the time during which the datagram will remain in the internet.
- This field is used to help piece together datagram fragments.

Flags
- IT consists of 3 bit field.
- Lower 2 bits are defined *D* flag (Do not fragment) and *M* (More fragments) flag.
- If *D* value is 1, the router must not fragment the data.
- *M* specifies whether the packet is the last fragment in a series of fragmented packets. i.e if its value is 1, it means the datagram is not the last fragment; it its value is 0, it means this is the last or only fragment.
- The third or high order bit is not used.

Fragment Offset
- This 13 bit field indicates the position of the fragment's data relative to the beginning

of the data in the original datagram, which allows the destination IP process to properly reconstruct the original datagram.
- The value of the offset is measured in units of 8 bytes.

Time-to-Live

- This 8 bit field Specifies how long, in seconds, a datagram is allowed to remain in the internet.
- Every router that processes a datagram must decrease the TTL by at least one, so the TTL is similar to a hop count.
- If this value, after being decremented, is zero, the router discards the datagram.

Protocol

- This 8 bit field indicates which upper layer protocol the data belongs to;
- so that when the datagram arrives at the destination, the value of this field helps to define to which protocol the payload should be delivered.
- A complete list is maintained at `http://www.iana.org/assignments/protocol-numbers`.

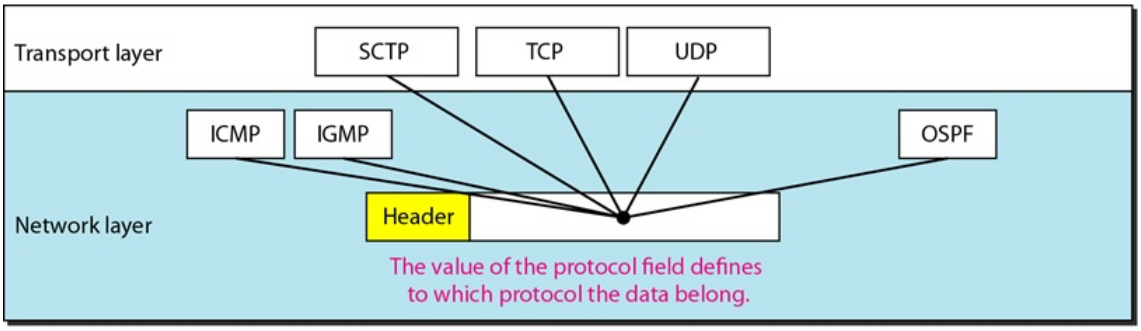

Transport layer
SCTP
TCP
UDP
ICMP
IGMP
OSPF
Network layer
Header
The value of the protocol field defines to which protocol the data belong.

Figure 4.37: Protocol field and encapsulated data.

Table 4.5: Some protocol values

Value	Protocol
1	ICMP
2	IGMP
6	TCP
17	UDP
89	OSPF

Header Checksum

- IP adds a 16-bit header checksum field to check the header, but not the payload.

Because some header field may change during transit (e.g. TTL, destination address, fragmentation-related fields).
- IP is not a reliable protocol; it does not check whether the payload carried by a datagram is corrupted during the transmission. IP puts the burden of error checking of the payload on the protocol that owns the payload, such as UDP or TCP.

Source and Destination Address
- These 32-bit source and destination address fields define the IP address of the source and destination respectively.
- The source host should know its IP address. The destination IP address is either known by the protocol that uses the service of IP or is provided by the DNS.
- The value of these fields must remain unchanged during the time the IP datagram travels from the source host to the destination host.

Options
- A datagram header can have up to 40 bytes of options. It size is variable. It is used to encode the options requested by the sending host.
- It is also used for padding to ensure that the datagram header is a multiple of 32 bits in length.
- Options, as the name implies, are not required for a datagram.

Data
- Payload, or data, is the main reason for creating a datagram.
- Payload is the packet coming from other protocols that use the service of IP.
- The data field must be an integer multiple of 8 bits in length. The maximum length of the datagram (data field plus header) is 65,535 octets.

4.18.1 Maximum Transfer Unit (MTU)

- Each link-layer protocol has its own frame format.
- One of the features of each format is the maximum size of the payload that can be encapsulated.
- In other words, when a datagram is encapsulated in a frame, the total size of the datagram must be less than this maximum size, which is defined by the restrictions imposed by the hardware and software used in the network.

- The value of the MTU differs from one physical network protocol to another. For example, the value for a LAN is normally 1500 byte.

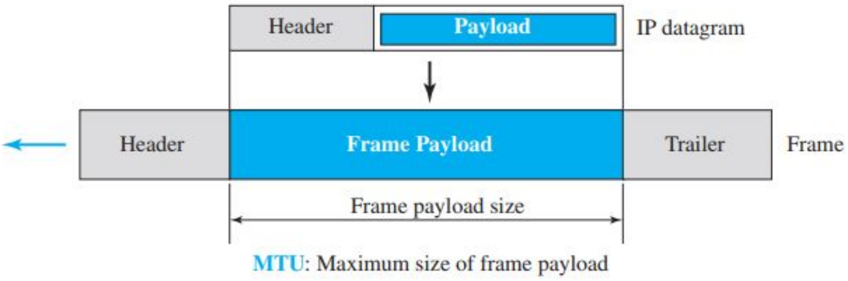


Figure 4.38: Maximum Transfer Unit (MTU).

Table 4.6: MTUs for some networks.

Protocol	MTU
Hyperchannel	65,535
Token Ring (16 Mbps)	17,914
Token Ring (4 Mbps)	4,464
FDDI	4,352
Ethernet	1,500
X.25	576
PPP	269

4.18.2 Fragmentation and Reassembly

- When datagram has to pass through the other physical network, the datagram must be divided into smaller size less that or equal to MTU of that network to make it possible for it to pass through the network. The is called **fragmentation**.

- A datagram can be fragmented by the source host or any router in the path. The reassembly of the datagram, however, is done only by the destination host, because each fragment becomes an independent datagram, which can travel through different path.

- The *identification number* helps the destination in reassembling the datagram. It knows that all fragments having the same identification value should be assembled into one datagram.

- When we talk about fragmentation, we mean that the payload of the IP datagram is fragmented. However, most parts of the header, with the exception of some options, must be copied by all fragments. The host or router that fragments a datagram must change the values of three fields: *flags, fragmentation offset, and total length*. The rest of the fields must be copied. Of course, the value of the checksum must be recalculated regardless of fragmentation.

Fragmentation Example:

- 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.
- The first fragment carries bytes 0 to 1399. The offset for this datagram is 0/8 = 0.
- The second fragment carries bytes 1400 to 2799; the offset value for this fragment is 1400/8 = 175.
- Finally, the third fragment carries bytes 2800 to 3999. The offset value for this fragment is 2800/8 = 350.

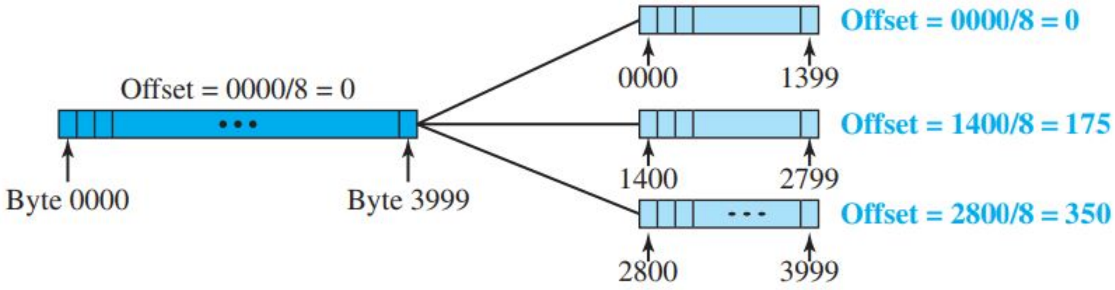


Figure 4.39: Fragmentation example.

4.19 Internet Control Message Protocol (ICMP)

Limitation of Internet Protocol

- The IP protocol has no error-reporting or error-correcting mechanism.
- The IP protocol also lacks a mechanism for host and management queries.
- The Internet Control Message Protocol (ICMP) has been designed to compensate for the above two deficiencies.
- It is a companion to the IP protocol.

- ICMP is a network layer protocol used by network devices (such as routers) to diagnose network communication issues.
- The primary purpose of ICMP is for *error reporting*. When two devices connect over the Internet, the ICMP generates errors to share with the sending device in the event that any of the data did not get to its intended destination. For example, if a packet of data is too large for a router, the router will drop the packet and send an ICMP message

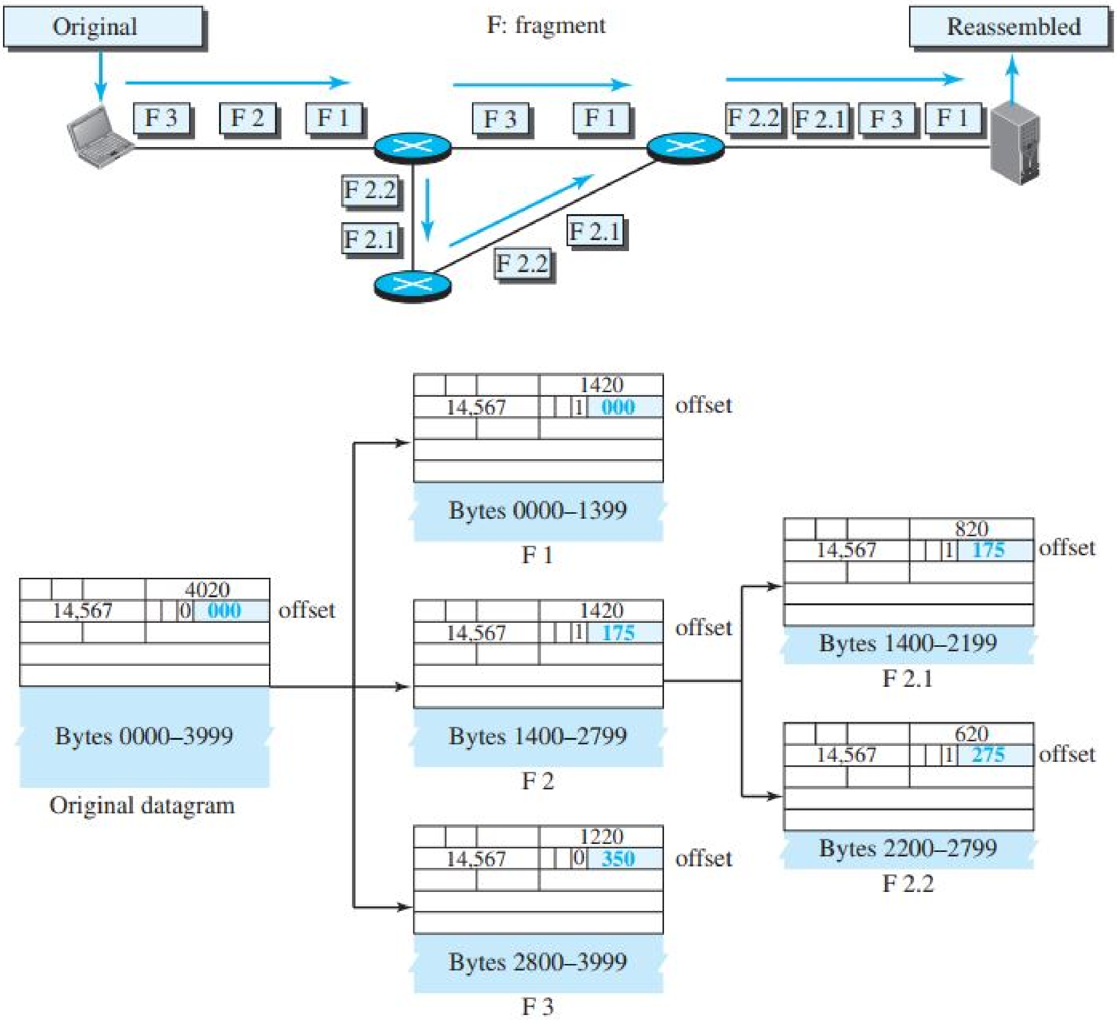


Figure 4.40: Detailed fragmentation example.

back to the original source for the data.
- A secondary use of ICMP protocol is to perform *network diagnostics* to assess the network performance; the commonly used terminal utilities *traceroute* and ping both operate using *ICMP*.
- ICMP is also used to hurt network performance. This is done using an ICMP flood, a Smurf attack, and a ping of death attacks that overwhelms a device on the network and prevent normal functionality.

ICMPv4 Message

- ICMP messages are not directly passed to data link layer; instead it is first encapsulated inside IP datagrams and send to data link layer. The protocol field is set to 1.
- It is of two types:
i. Error- Reporting Messages
ii. Query Messages

- The *error-reporting messages* report problems that a router or a host (destination) may encounter when it processes an IP packet. ICMP always reports error messages to the original source.
- The *query messages*, which occur in pairs, help a host or a network manager gets specific information from a router or another host.

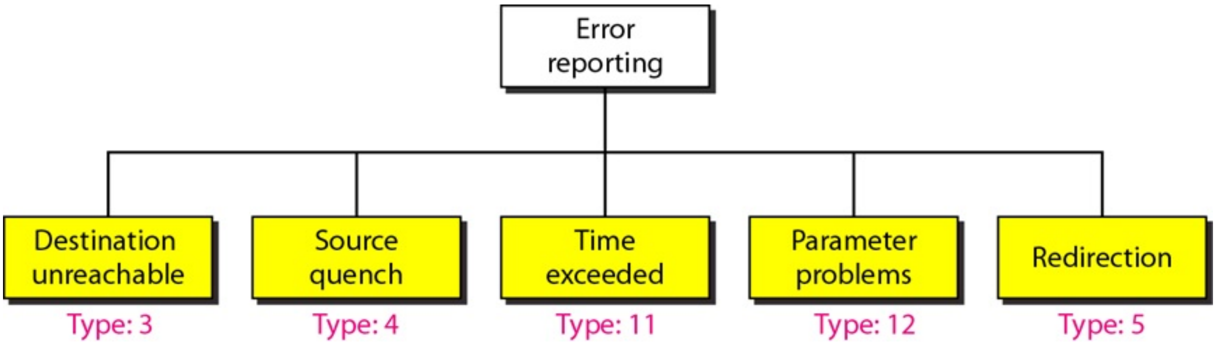


Figure 4.41: Error-reporting messages.

- **Destination Unreachable**: Informs the sender that host is unreachable.
- **Source Quench**: Informs the sender that the network has encountered congestion and the datagram has been dropped.
- **Time exceeded**: If a packet's TTL drops to zero, ICMP Time Exceeded Message is sent to the source host.
- **Redirection Message**: used when the source uses a wrong router to send out its message. The router redirects the message to the appropriate router, but informs the source that it needs to change its default router.
- **Parameter problem**: sent when either there is a problem in the header of a datagram or some options are missing or cannot be interpreted.

Query Messages:
- The *echo request (type 8)* and the *echo reply (type 0)* pair of messages are used by a host or a router to test the liveliness of another host or router. A host or router sends an echo request message to another host or router; if the latter is alive, it responds with an echo reply message. *Ping* and *traceroute* commands are used for it.
- The timestamp request (type 13) and the timestamp reply (type 14) pair of messages are used to find the round-trip time between two devices or to check whether the clocks in two devices are synchronized.

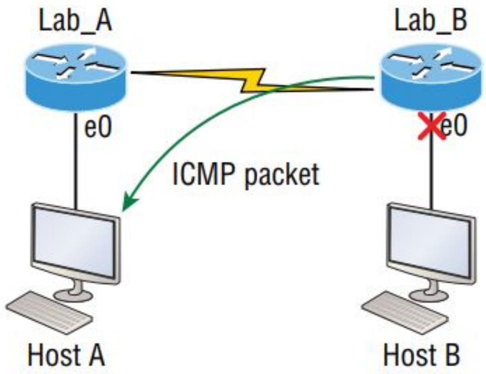


Figure 4.42: ICMP error message is sent to the sending host from the remote router.

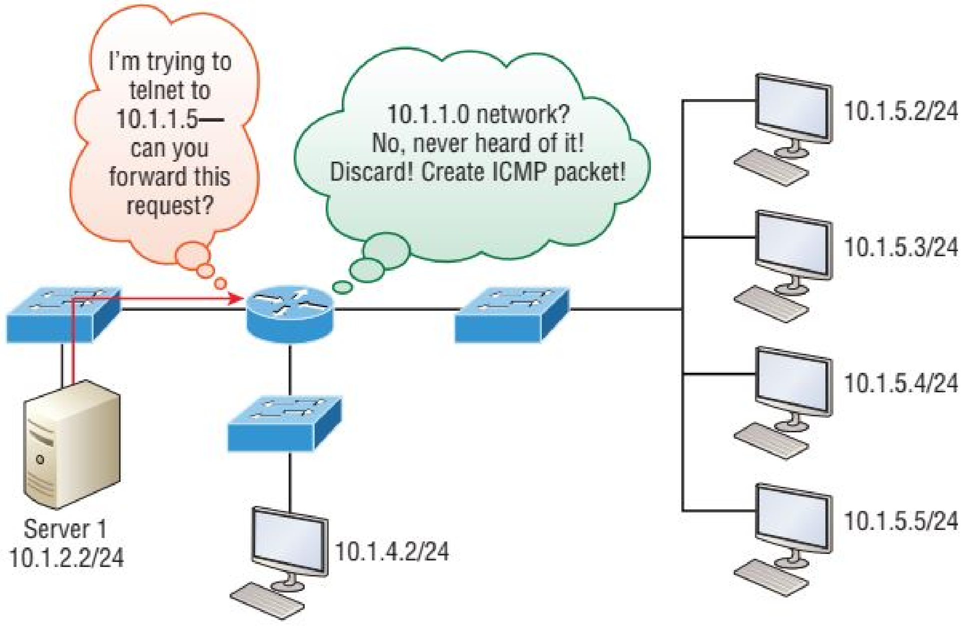

I'm trying to telnet to 10.1.1.5— can you forward this request?
10.1.1.0 network? No, never heard of it! Discard! Create ICMP packet!
10.1.5.2/24
10.1.5.3/24
10.1.5.4/24
10.1.5.5/24
Server 1
10.1.2.2/24
10.1.4.2/24

Figure 4.43: ICMP in action.

4.20 Address Resolution Protocol (ARP)

- Address Resolution Protocol (ARP) is a network layer protocol of the network layer that is used to find the he hardware address, also known as Media Access Control (MAC) address, of a host from its known IP address.
- As IP's detective, ARP interrogates the local network by sending out a broadcast asking the machine with the specified IP address to reply with its hardware address.

- Most of the computer programs/applications use logical address (IP address) to send/receive messages, however the actual communication happens over the physical address (MAC address) i.e from layer 2 of OSI model. So our mission is to get the destination MAC address which helps in communicating with other devices. This is where ARP comes into the picture, its functionality is to translate IP address to physical address.

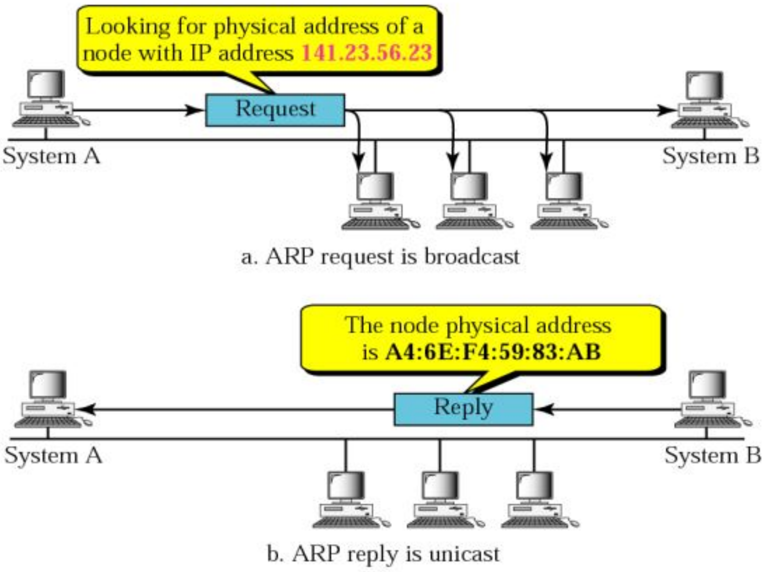

Looking for physical address of a node with IP address 141.23.56.23
Request
System A
System B
a. ARP request is broadcast
The node physical address is A4:6E:F4:59:83:AB
Reply
System A
System B
b. ARP reply is unicast

Figure 4.44: ARP resolves IP address to Ethernet (MAC) address.

4.21 Reverse Address Resolution Protocol (RARP)

- RARP is a network layer protocol used to map a physical address to a logical address.

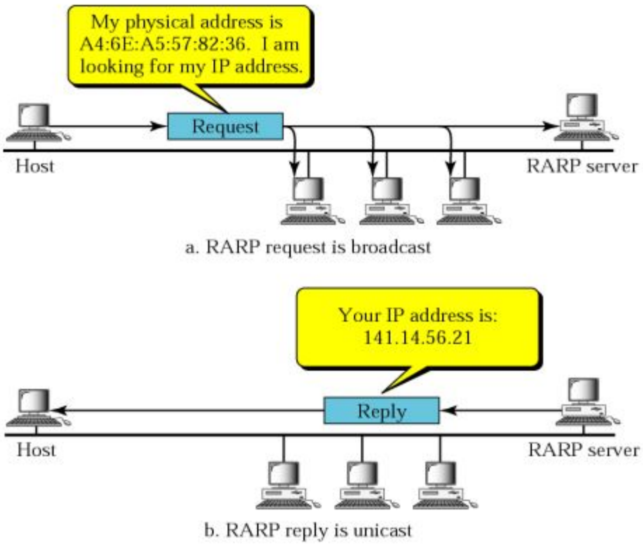

My physical address is A4:6E:A5:57:82:36. I am looking for my IP address.
Request
Host
RARP server
a. RARP request is broadcast
Your IP address is: 141.14.56.21
Reply
Host
RARP server
b. RARP reply is unicast

Figure 4.45: RARP resolves MAC address to IP address.