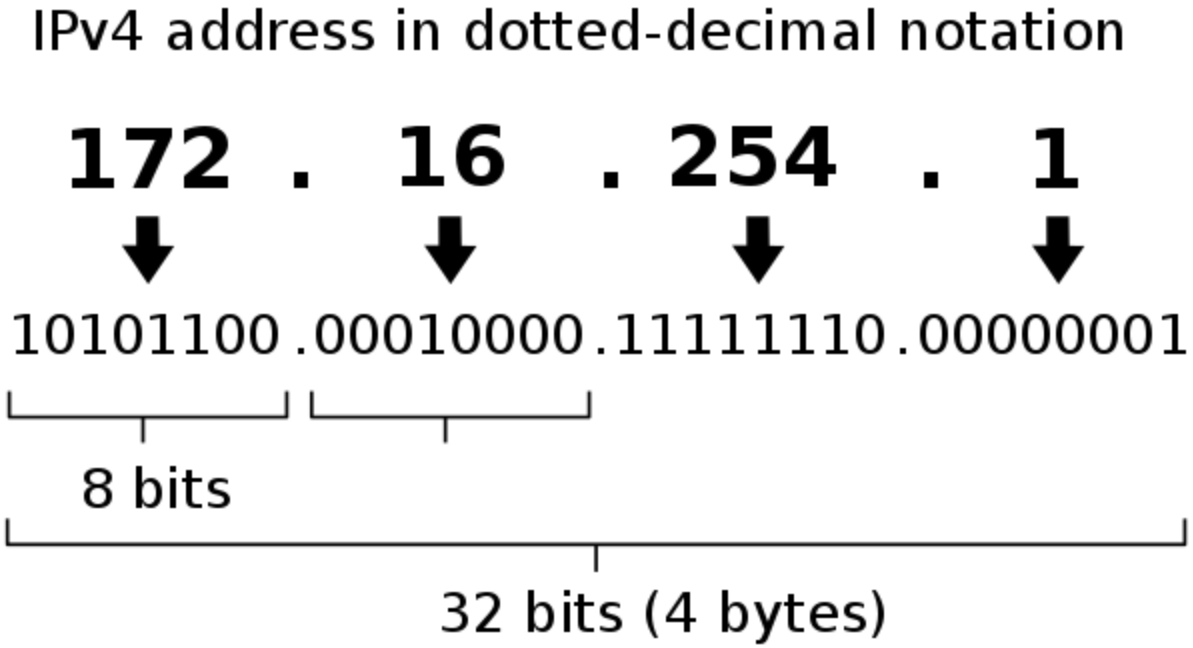
NETWORK LAYER

IP Address

['ī 'pē ə-'dres]

A number used to identify a computer or network of computers.

DOTTED DECIMAL NOTATION OF IP ADDRESS

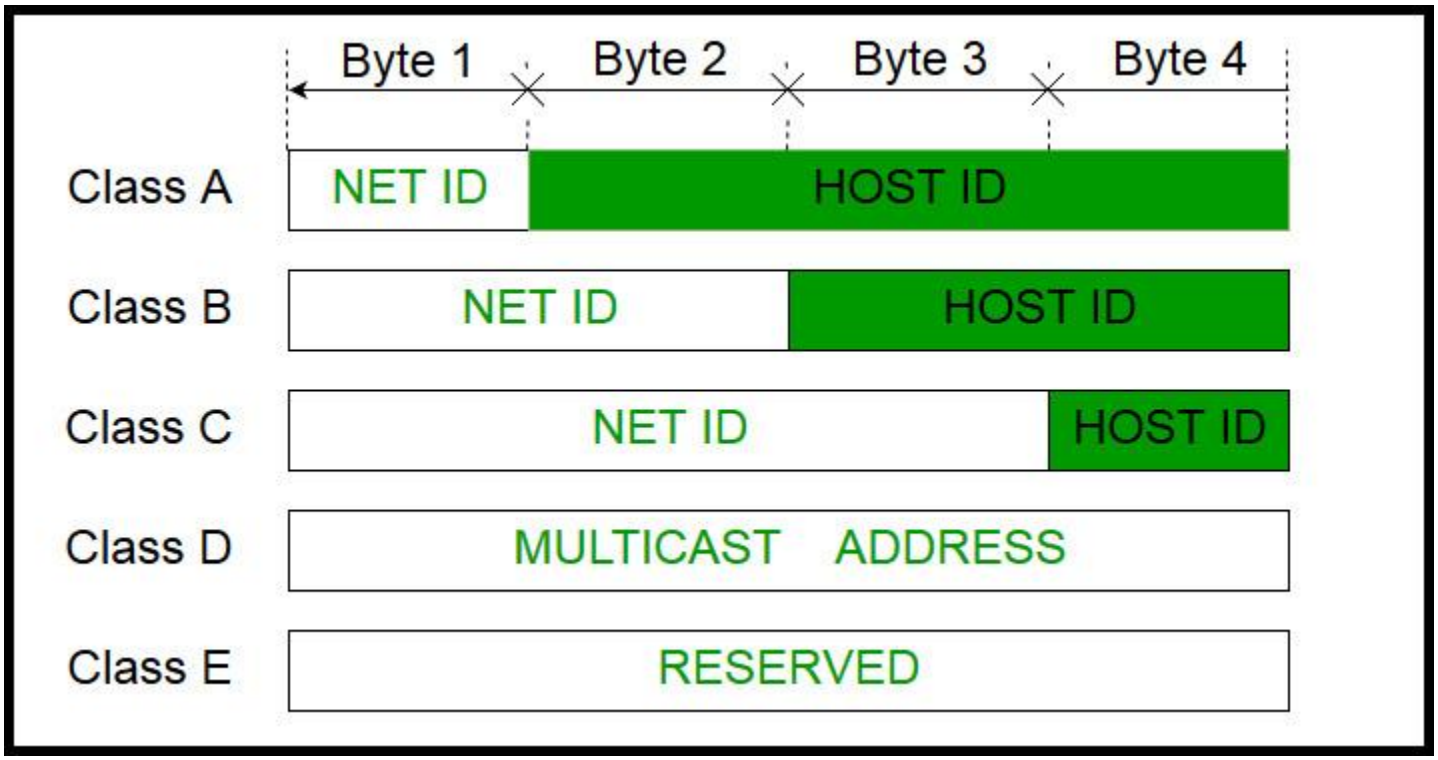


HEXADECIMAL NOTATION OF IP ADDRESS

01110101	00011101	10010101	11101010
75	1D	95	EA

0x751D95EA

IP ADDRESS CLASSES

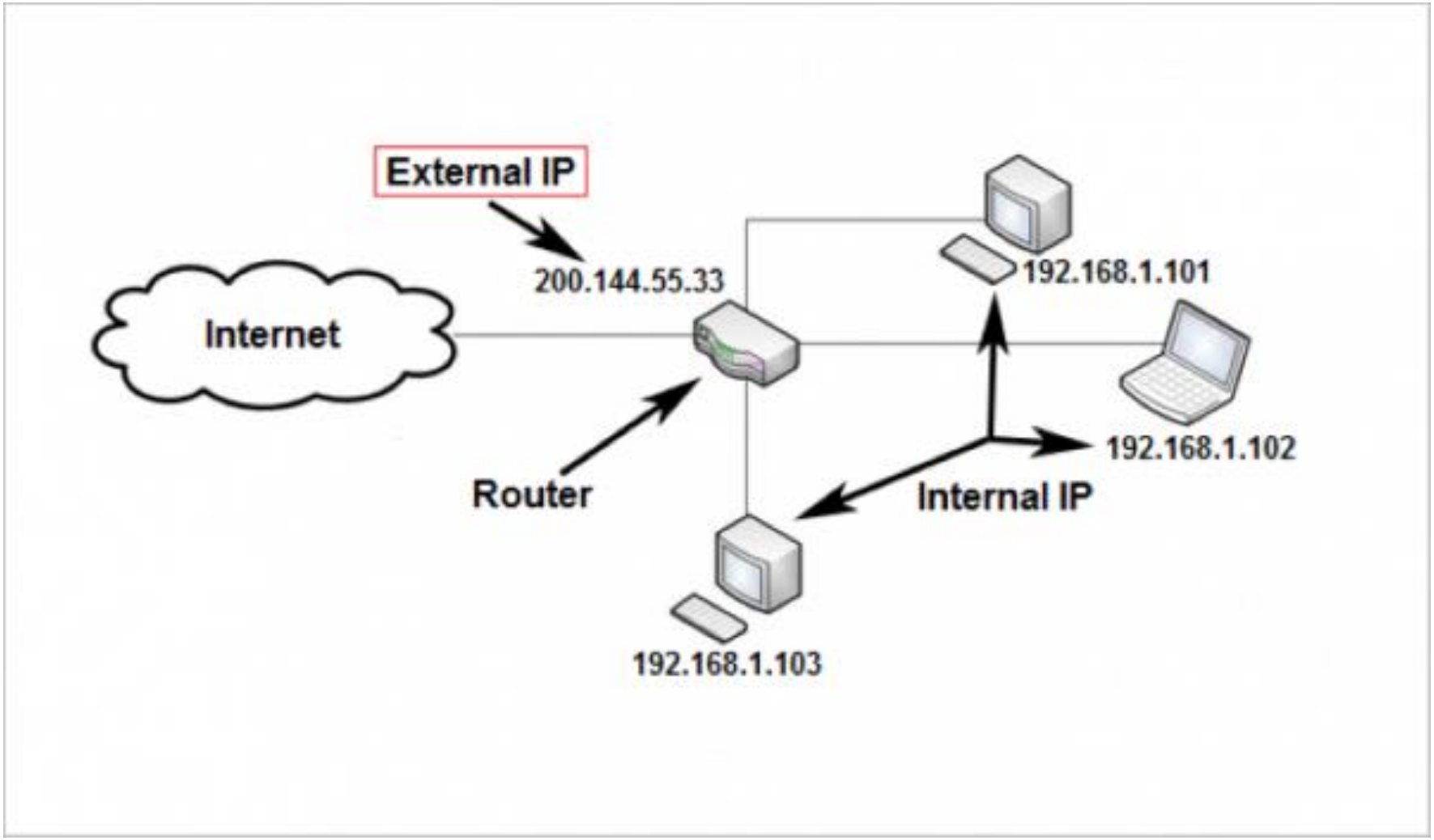


IP ADDRESS CLASSES

Address Class	RANGE	Default Subnet Mask
A	1.0.0.0 to 126.255.255.255	255.0.0.0
B	128.0.0.0 to 191.255.255.255	255.255.0.0
C	192.0.0.0 to 223.255.255.255	255.255.255.0
D	224.0.0.0 to 239.255.255.255	Reserved for Multicasting
E	240.0.0.0 to 254.255.255.255	Experimental

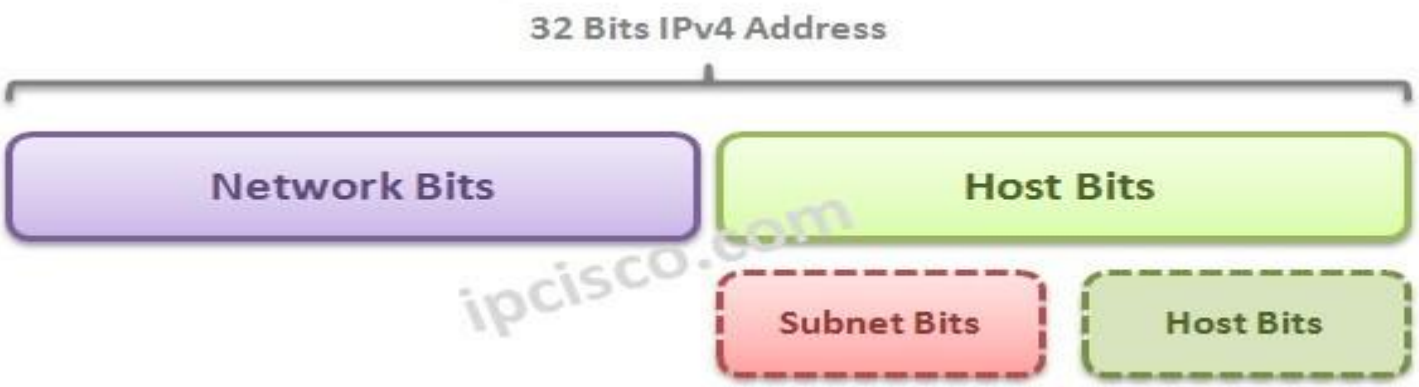
Note: Class A addresses 127.0.0.0 to 127.255.255.255 cannot be used and is reserved for loopback testing.

IP ADDRESS DISTRIBUTION WITHIN A NETWORK



Subnetting

- **What is subnetting**
  - **Process of subdividing a single class of network into multiple subnetworks.**
  - **A subnetted network address contains a network address, subnet address and host address.**



SUBNETTING

DEFINITION

It is a technique in which larger network is divided in to smaller sub networks , so that those sub networks will not communicate with each other.

OR

The process of converting host bits in to network bit is called sub netting.

Types of Sub -netting

There are two types of sub- netting

1. Network based sub -netting
2. Host based sub -netting

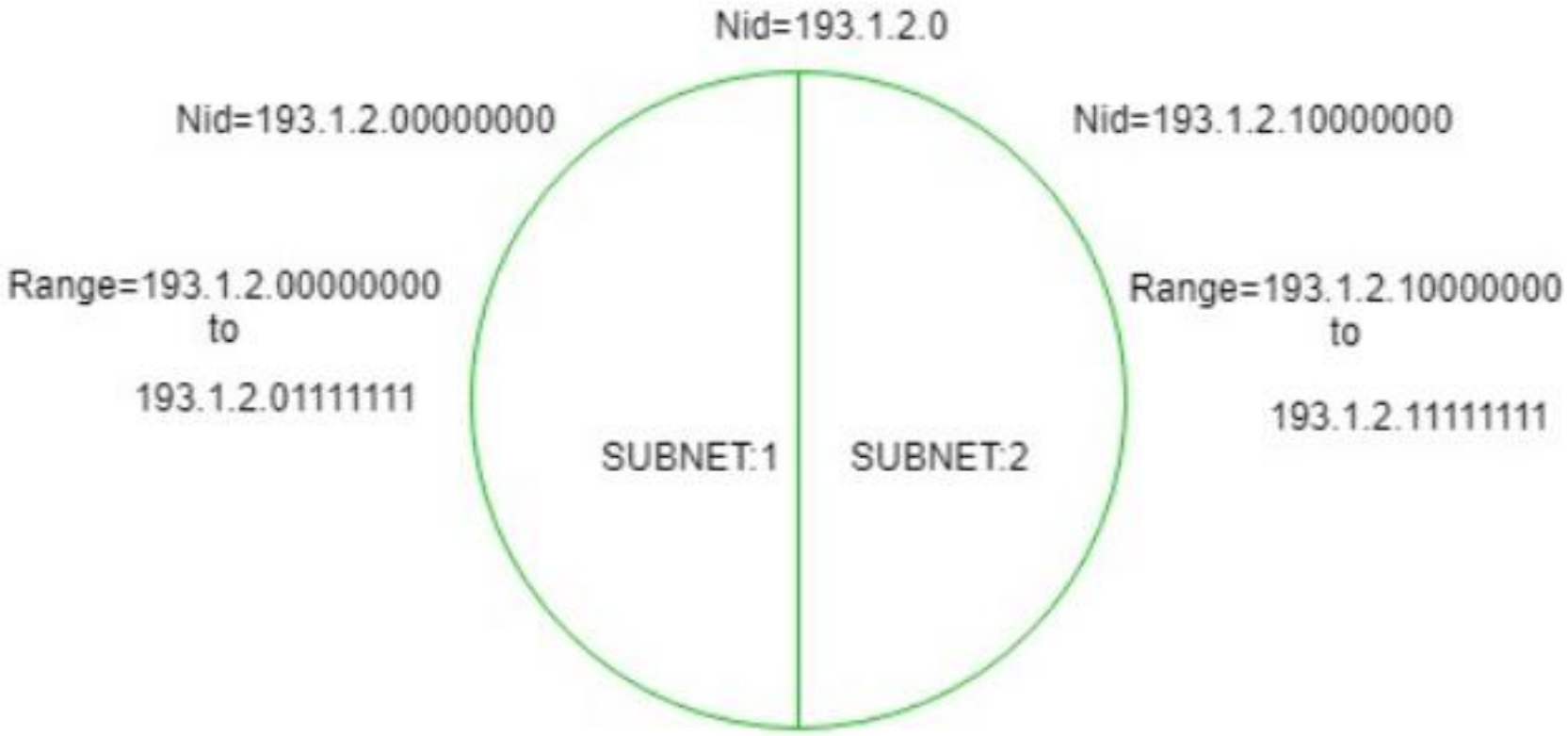
What is a Subnet Mask?

- An Address the accompanies an IP address that indicates which portion of the IP address is the Network ID and which portion of the IP address is the Host ID.
  - 152.107.102.7 (IP Address)
  - 255.255.255.0 (Subnet Mask)
- The IP Address and Subnet Mask (SNM) are interrelated and each only has meaning in the context of the other!

Subnet Masks

- **Tells the device which bits are host address and network address.**

Class	Subnet Mask	Binary
A	255.0.0.0	11111111.00000000.00000000.00000000
B	255.255.0.0	11111111. 11111111. 00000000.00000000
C	255.255.255.0	11111111. 11111111. 1111111.00000000


Nid=193.1.2.0
Nid=193.1.2.00000000
Nid=193.1.2.10000000
Range=193.1.2.00000000
to
193.1.2.01111111
Range=193.1.2.10000000
to
193.1.2.11111111
SUBNET:1
SUBNET:2

- **For Subnet-1:** The first bit which is chosen from the host id part is zero and the range will be from (193.1.2.00000000 till you get all 1's in the host ID part i.e, 193.1.2.01111111) except for the first bit which is chosen zero for subnet id part. Thus, the range of subnet-1:

```
193.1.2.0 to 193.1.2.127
```

Subnet id of Subnet-1 is : 193.1.2.0

Direct Broadcast id of Subnet-1 is : 193.1.2.127

Total number of host possible is : 126 (Out of 128, 2 id's are used for Subnet id & Direct Broadcast id)

- **For Subnet-2:** The first bit chosen from the host id part is one and the range will be from (193.1.2.100000000 till you get all 1's in the host ID part i.e, 193.1.2.11111111). Thus, the range of subnet-2:

```
193.1.2.128 to 193.1.2.255
```

Subnet id of Subnet-2 is : 193.1.2.128

Direct Broadcast id of Subnet-2 is : 193.1.2.255

Total number of host possible is : 126 (Out of 128, 2 id's are used for Subnet id & Direct Broadcast id)

Example1. An organization is assigned a class C network address of 201.35.2.0. It uses a netmask of 255.255.255.192 to divide this into sub-networks. Which of the following is/are valid host IP addresses?

```
A. 201.35.2.129
B. 201.35.2.191
C. 201.35.2.255
D. Both (A) and (C)
```

Example 2. An organization has a class C network address of 201.32.64.0. It uses a subnet mask of 255.255.255.248. Which of the following is NOT a valid broadcast address for any subnetworks?

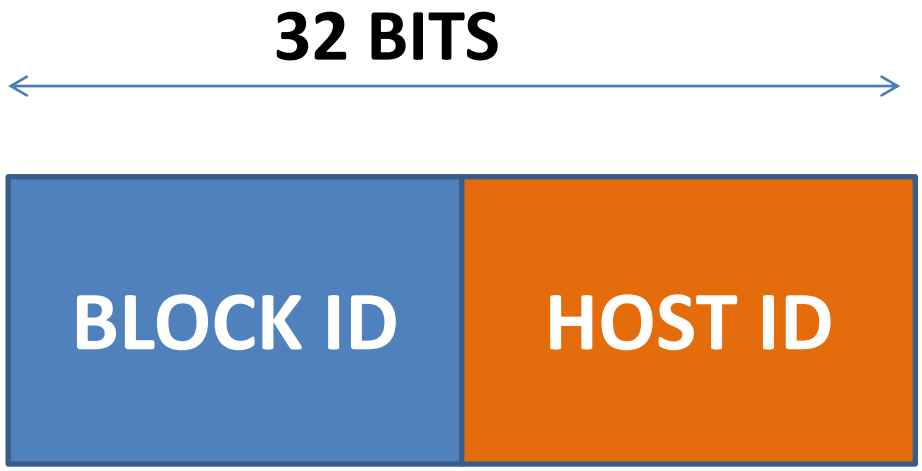
A. 201.32.64.135
B. 201.32.64.240
C. 201.32.64.207
D. 201.32.64.231

Disadvantages of Subnetting

- Subnetting can become confusing as networks grow larger.
- Subnetting itself is not an easy process if the administrator does not practice.

CIDR – Classless InterDomain Routing

- Running out of IP addresses
- class C is too small; class B is too large (more than half of the class B networks have fewer than 50 hosts)
- CIDR (Classless InterDomain Routing) allows to allocate IP address with a variable-sized block (contiguous network numbers to nearby networks), with no regard to the classes.



Notation Used

x.y.z.w/n

Here n shows number of network bits

For eg. 200.10.20.40/28 is a CIDR address

Subnet Mask: 11111111.11111111.11111111.11110000

For Eg. Using CIDR, Find out total no. of

a) **Network bits**
b) **Host bits**
c) **Possible Hosts**
d) **Network/Block ID**

Of following IP 200.10.20.40/28

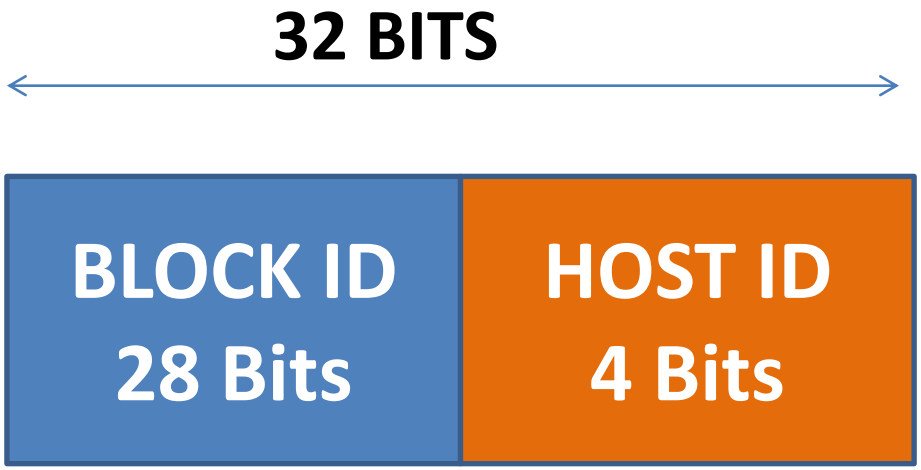
On comparing given IP 200.10.20.40/28 with Notation x.y.z.w/n

Here n=28

So total network bits=28 out of 32

So host bits=32 – 28 = 4

Total no. of Possible Hosts= 2^4 = 16



As network is 28 bits and host is 4 bits

So subnet mask is

11111111.11111111.11111111.11110000

255 . 255 . 255 . 240

Given IP is 200.10.20.40

On doing AND operation over binary of Last Octet of given IP with Subnet Mask we get network ID

240 11110000

40 00101000

---

32 00100000

So Network id/Block id is 200.110.200.32/28

Rules for CIDR Addressing

- **Address should be contiguous.**
- **Number of addresses in a block must be in power of 2**
- **First address of every block must be evenly divisible with size of blocks/total hosts.**

SUBNETTING IN CIDR

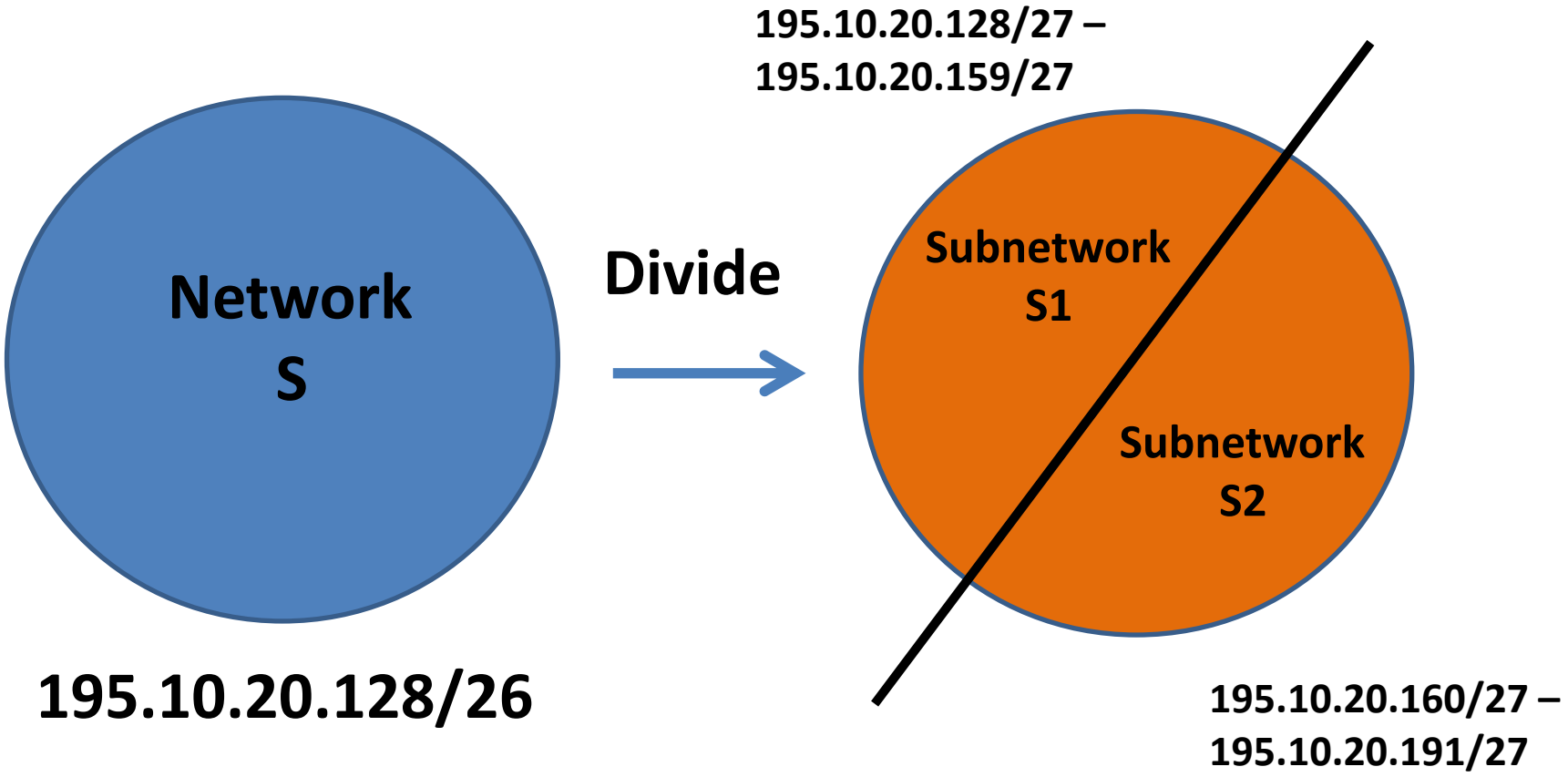
Subnetting in CIDR is same as Classful IP addressing

Bits in host will be used as subnet purpose and to be fixed for a particular network

So number of subnetworks = 2^m

where m= no. of subnet bits

For eg. Network S divides into two subnetworks S1 and S2



Variable Length Subnet Masking(VLSM)

Technique to make subnetworks of different different sizes

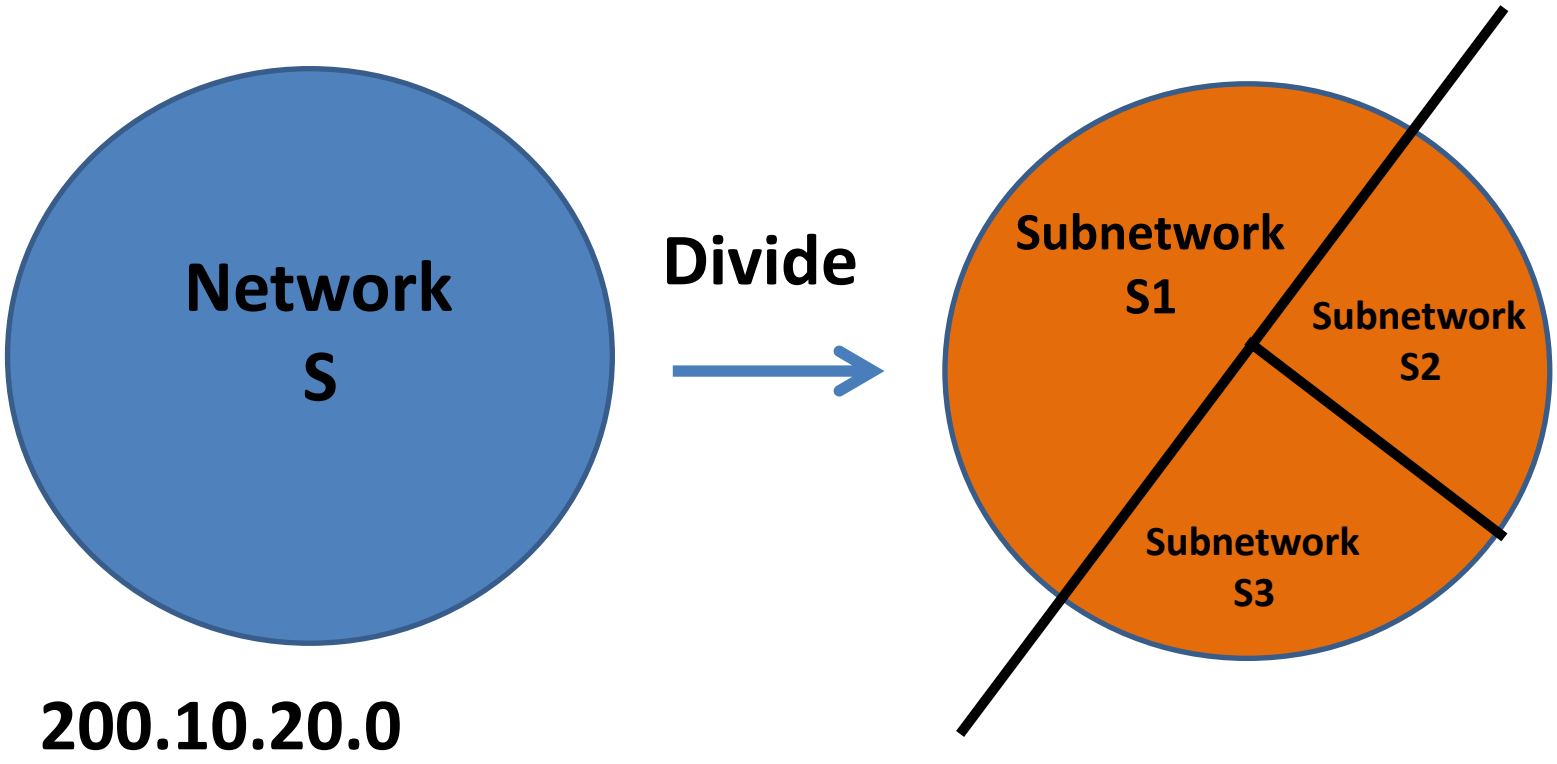
Network allow of different size subnetworks are known as Flexibility

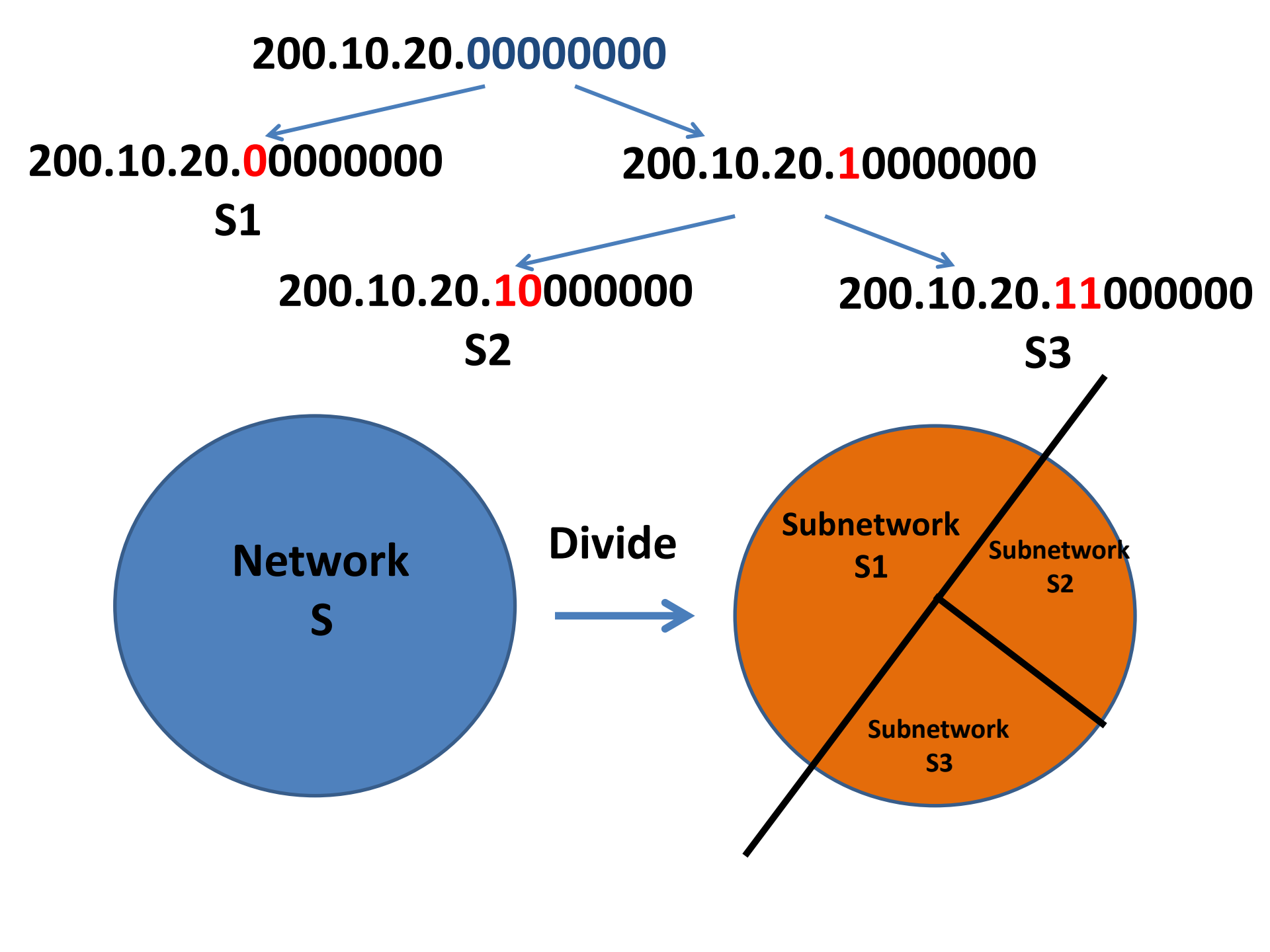
VLSM applied in both Classful and Classless Addressing

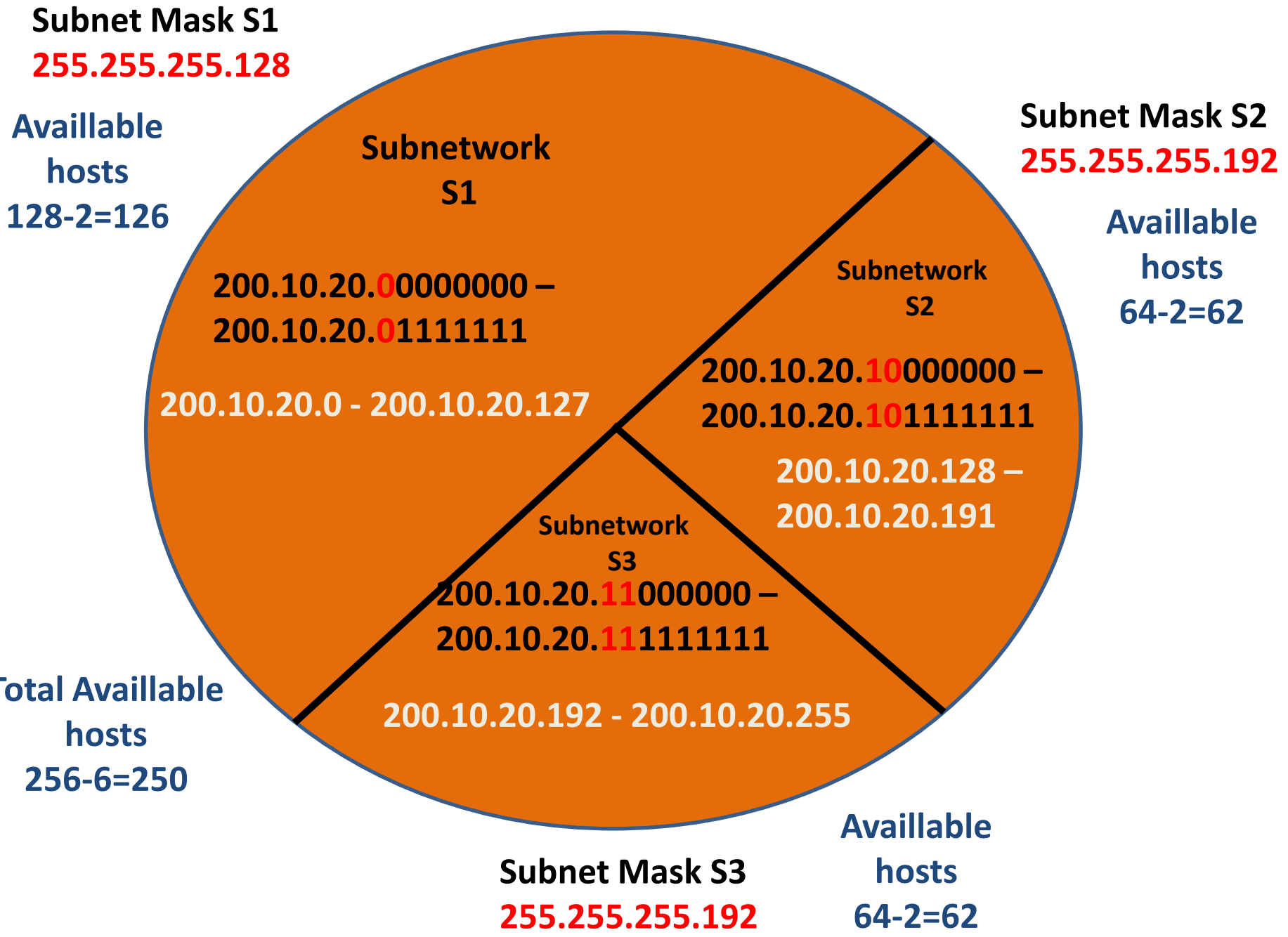
Always Remember: Only host bits used for subnetting

VLSM in Classful Addressing

For eg. Network S divides into three subnetworks S1(50%), S2(25%) and S3(25%) of different sizes




200.10.20.00000000
200.10.20.00000000
S1
200.10.20.10000000
200.10.20.10000000
S2
200.10.20.11000000
S3
Network
S
Divide
Subnetwork
S1
Subnetwork
S2
Subnetwork
S3


Subnet Mask S1
255.255.255.128
Availlable hosts
128-2=126
Subnetwork S1
200.10.20.00000000 –
200.10.20.01111111
200.10.20.0 - 200.10.20.127
Subnet Mask S2
255.255.255.192
Availlable hosts
64-2=62
Subnetwork S2
200.10.20.10000000 –
200.10.20.101111111
200.10.20.128 –
200.10.20.191
Subnetwork S3
200.10.20.11000000 –
200.10.20.111111111
200.10.20.192 - 200.10.20.255
Total Availlable hosts
256-6=250
Subnet Mask S3
255.255.255.192
Availlable hosts
64-2=62

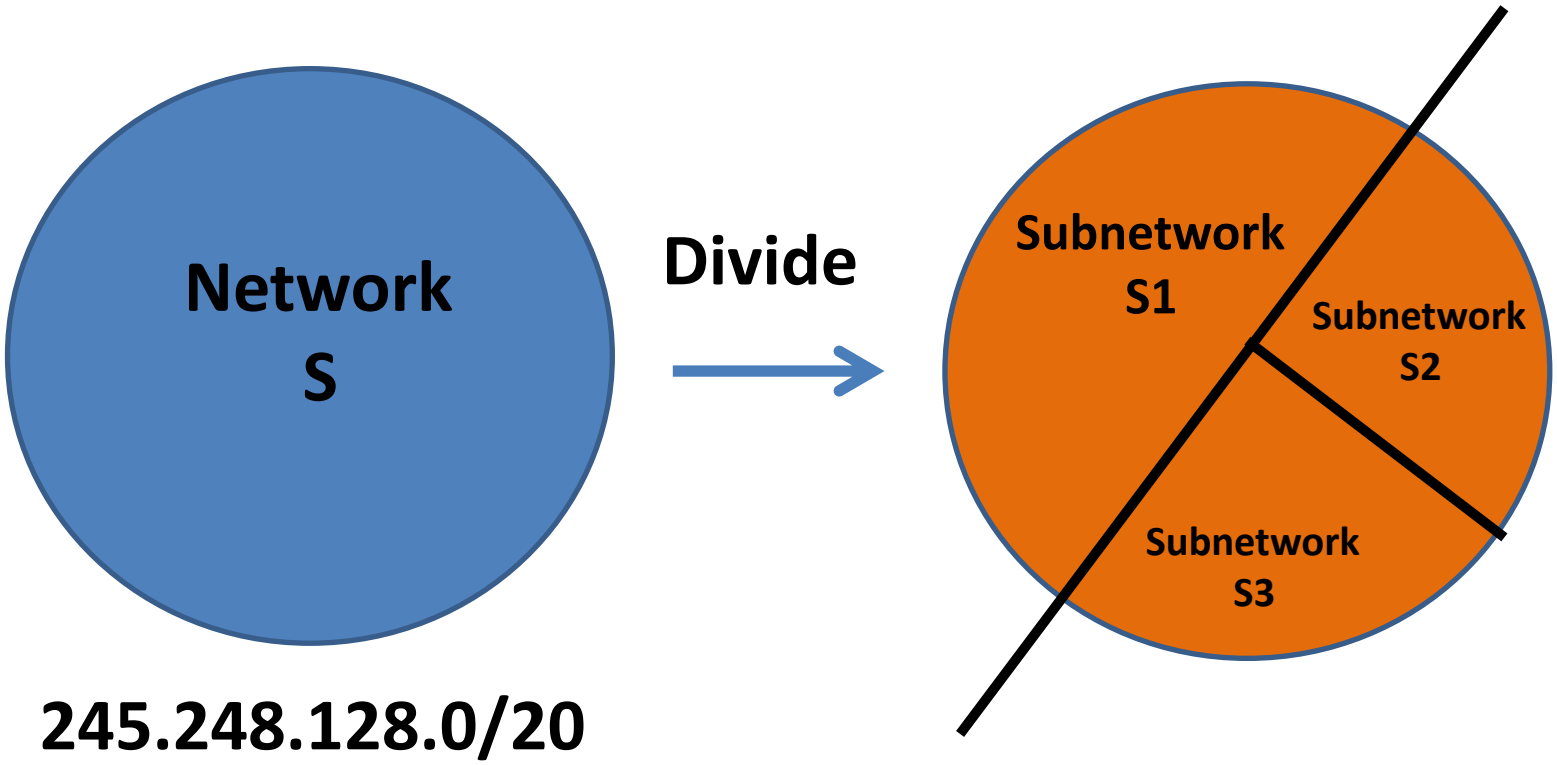
Variable Length Subnet Masking(VLSM)

If n=number of Subnets

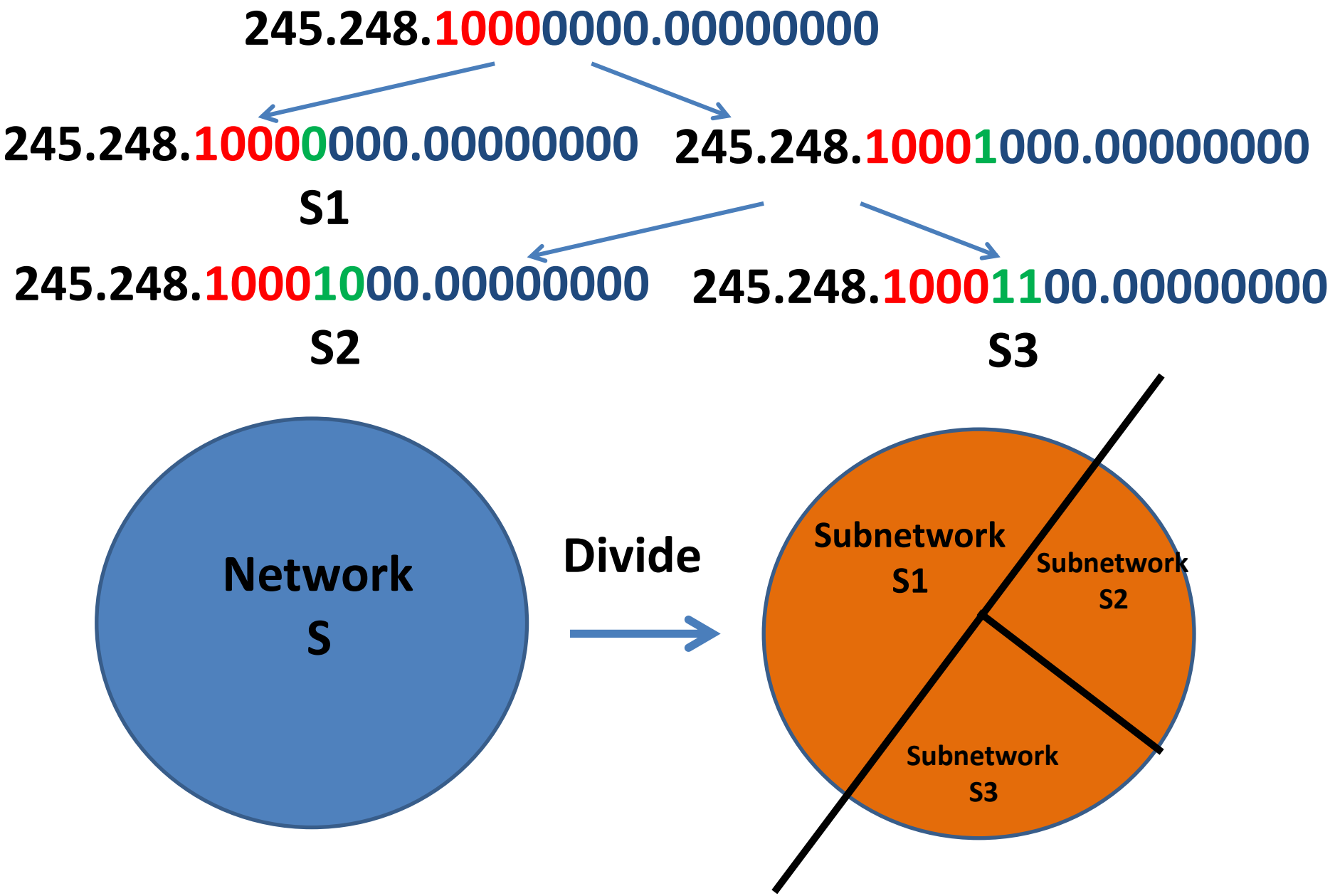
Total Usuable Hosts=Total Hosts – 2n

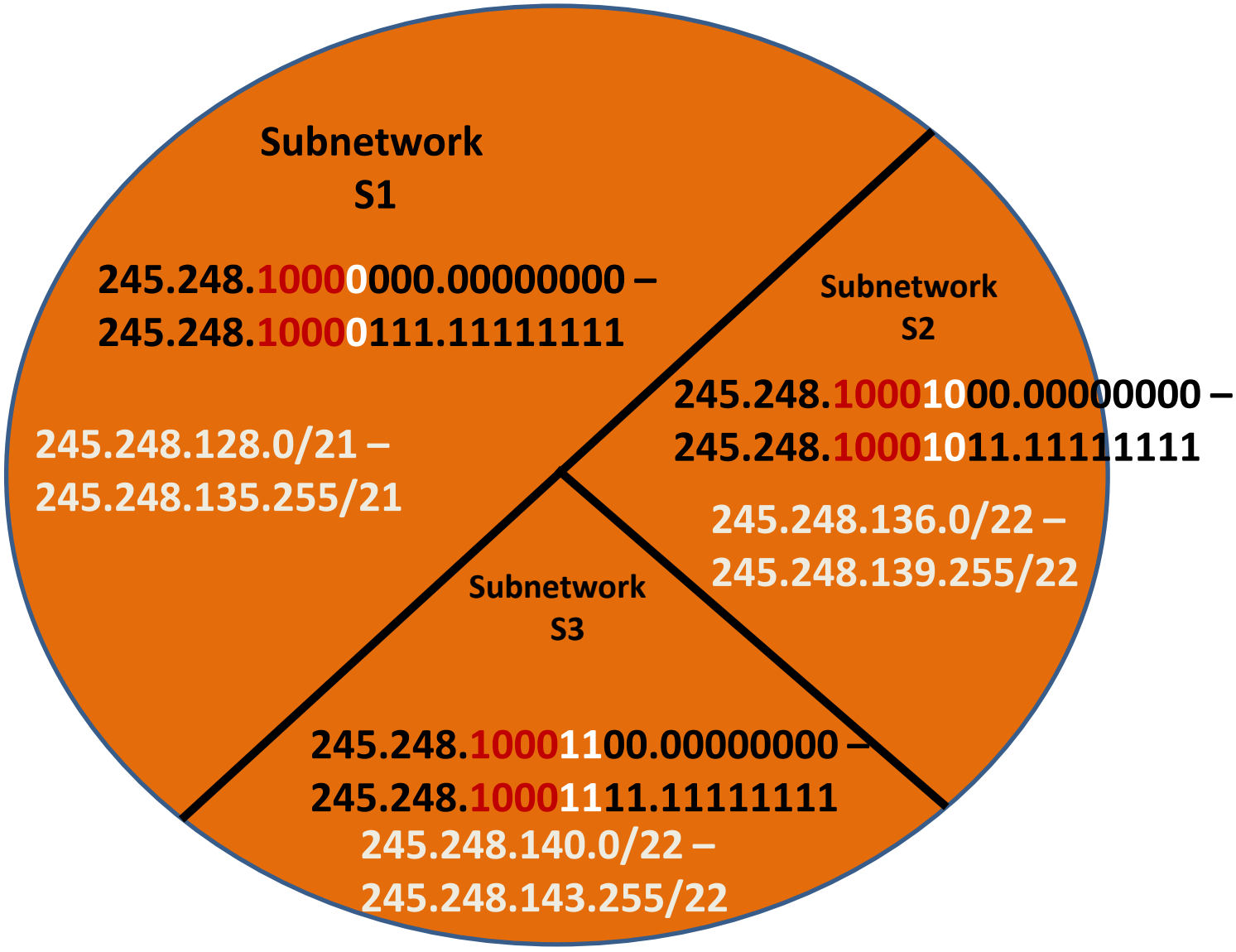
VLSM in CIDR

For eg. Network S divides into three subnetworks S1(50%), S2(25%) and S3(25%) of different sizes



245.248.128.0/20


245.248.10000000.00000000
245.248.10000000.00000000
S1
245.248.10001000.00000000
245.248.10001000.00000000
S2
245.248.10001100.00000000
S3
Network
S
Divide
Subnetwork
S1
Subnetwork
S2
Subnetwork
S3


Subnetwork
S1
245.248.10000000.00000000 –
245.248.10000111.11111111
245.248.128.0/21 –
245.248.135.255/21
Subnetwork
S2
245.248.10001000.00000000 –
245.248.10001011.11111111
245.248.136.0/22 –
245.248.139.255/22
Subnetwork
S3
245.248.10001100.00000000 –
245.248.10001111.11111111
245.248.140.0/22 –
245.248.143.255/22

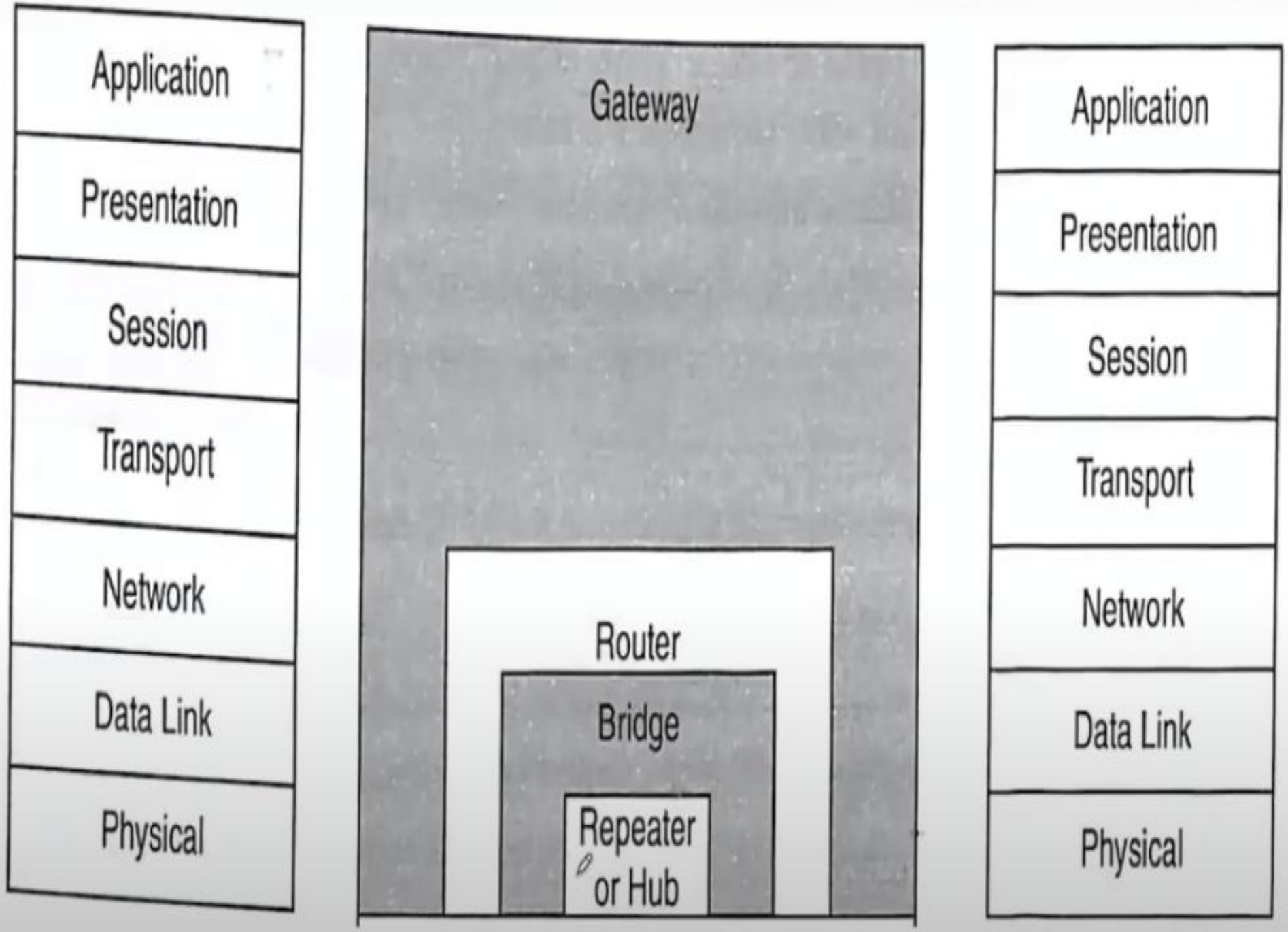
INTERNETWORKING
&
CONNECTING DEVICES

INTERNETWORKING

- When two or more different networks are connected together to form a bigger network, it is known as internetwork.
- These different network may be based on different technologies and may use different protocols like TCP/IP, SNA, DECnet, NCP/IPX, Apple Talk and other specialized protocols for satellites and cellular networks.
- Beside protocols, there are several other parameters that differentiate network. For example, packet size, flow control etc.

NETWORKING AND INTERNETWORKING DEVICES

1. Hub
    - Passive Hub
    - Active Hub
2. Repeater
3. Bridge
4. Router
5. Gateways
6. Switch


Application
Presentation
Session
Transport
Network
Data Link
Physical
Gateway
Router
Bridge
Repeater
or Hub
Application
Presentation
Session
Transport
Network
Data Link
Physical

HUB

- **Passive Hub**

A passive hub is just a connector and connects the wires coming from different sides.

A passive hub is just a point where signal coming from different stations collide. Therefore, passive hub is the collision point.

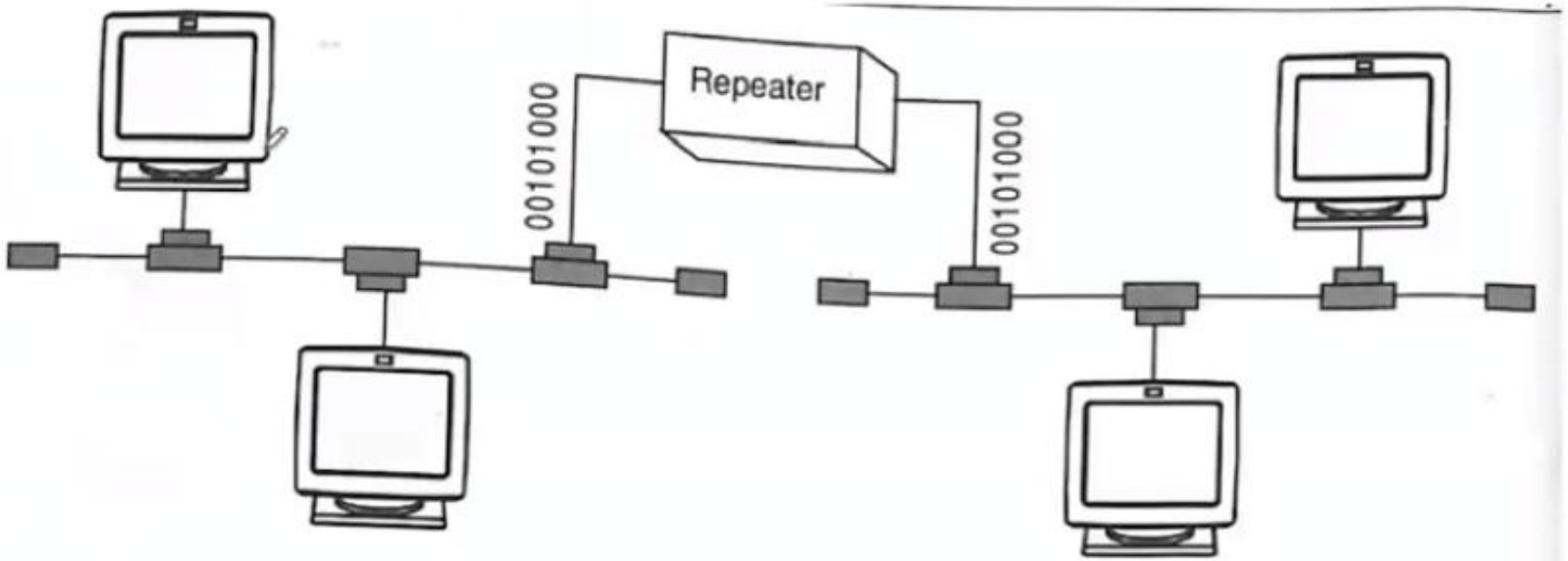
- **Active Hub**

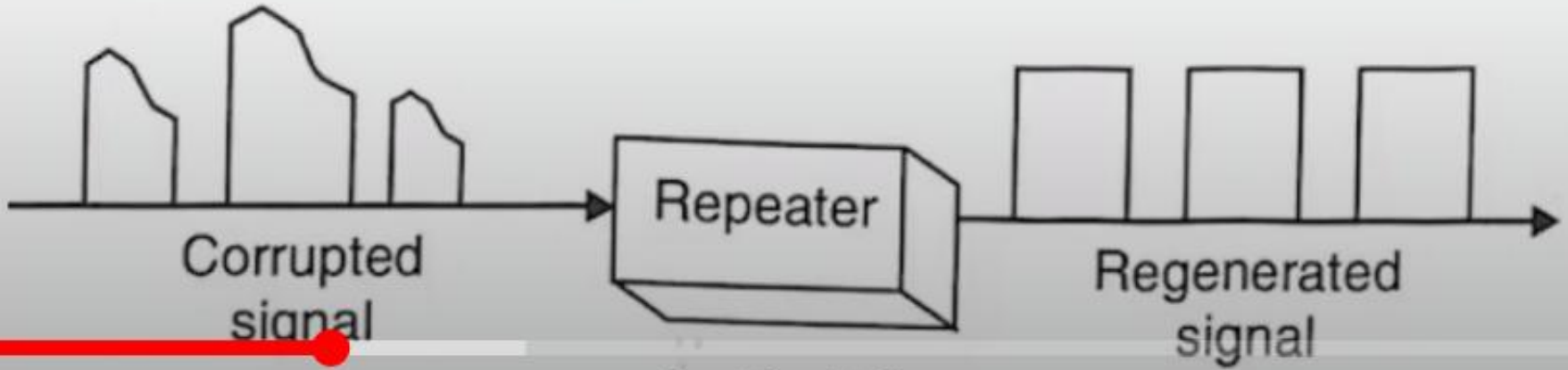
An active hub is a multi-port repeater. It is used in physical star topology to create connections between the stations

2. Repeaters

- A repeater is also known as regenerator.
- It is an electronic device that operates on only physical layer of the OSI model.
- Signal can carry the information within a network only for a fixed distance.
- After this distance attenuation occurs and the signal becomes weak or corrupted.
- The purpose of repeater is to take this signal (before it becomes two weak or corrupted) and regenerate the original bit pattern.
- Thus, it places a fresh copy of signal on the link.

2. Repeaters



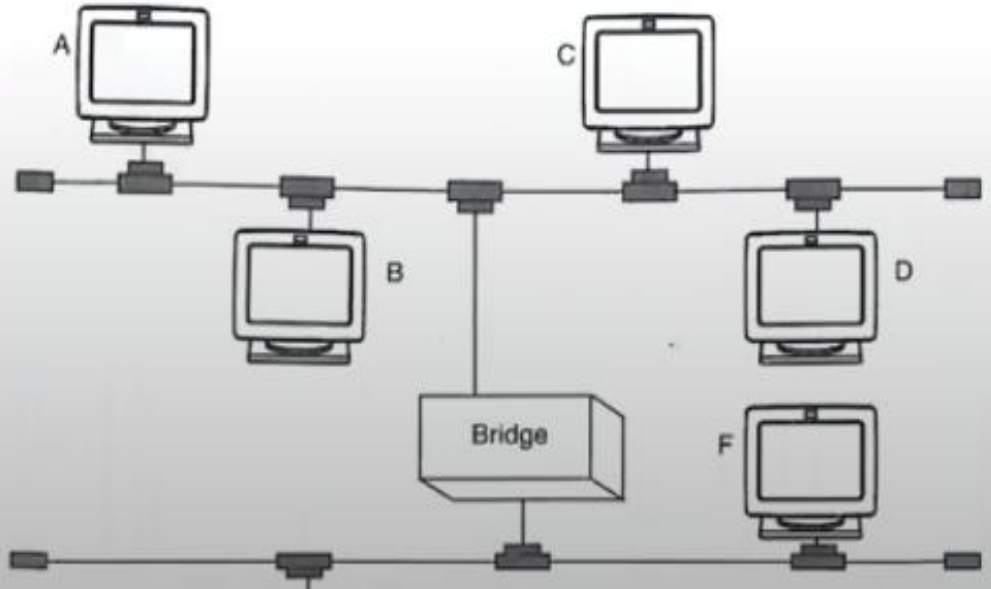


3. BRIDGES

- Bridges can be used to divide a large network into smaller segments and can also pass the frames between two originally separated LANs.
- As a physical layer device, a bridge regenerates the signal.
- As a data link layer device a bridge can access the physical or MAC address of the stations attached to it.
- Unlike repeater, a bridge performs filtering also.
- It has a logic that allows it to keep the traffic for each segment separate.
- It can check the destination address of a frame and decide if the frame should be forwarded or dropped.

3. BRIDGES

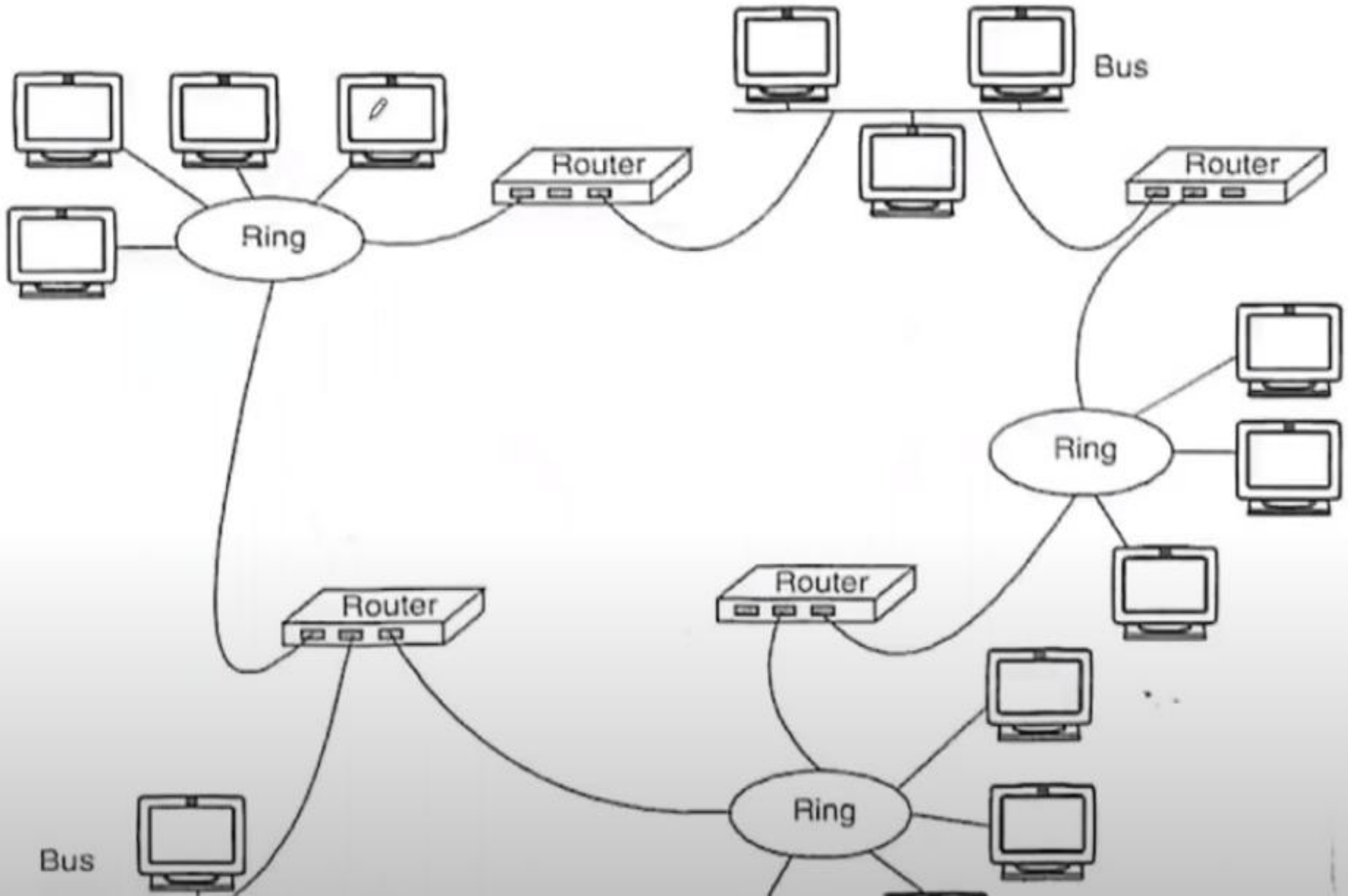
- A bridge maintains a table of MAC address of various stations attached to it.
- When a frame enters a bridge, it checks the address contained in the frame and compares this address with a table of all the stations on both segments.
- When it finds a match, it decides to which segment the destination station belongs and then passes the frame to that segment



4. ROUTERS

- Routers are multiport devices and more sophisticated as compared to repeaters and bridges.
- Routers also support filtering and encapsulation like bridges.
- They operate at physical, data link and network layer of OSI model
- Like bridges, they are self-learning, as they can communicate their existence to other devices and can learn of the existence of new routers, nodes and LAN segments.
- A router may connect LANs and WANs in the Internet and can pass or relay the packets among them. A router has access to the network layer address or logical address (IP address).

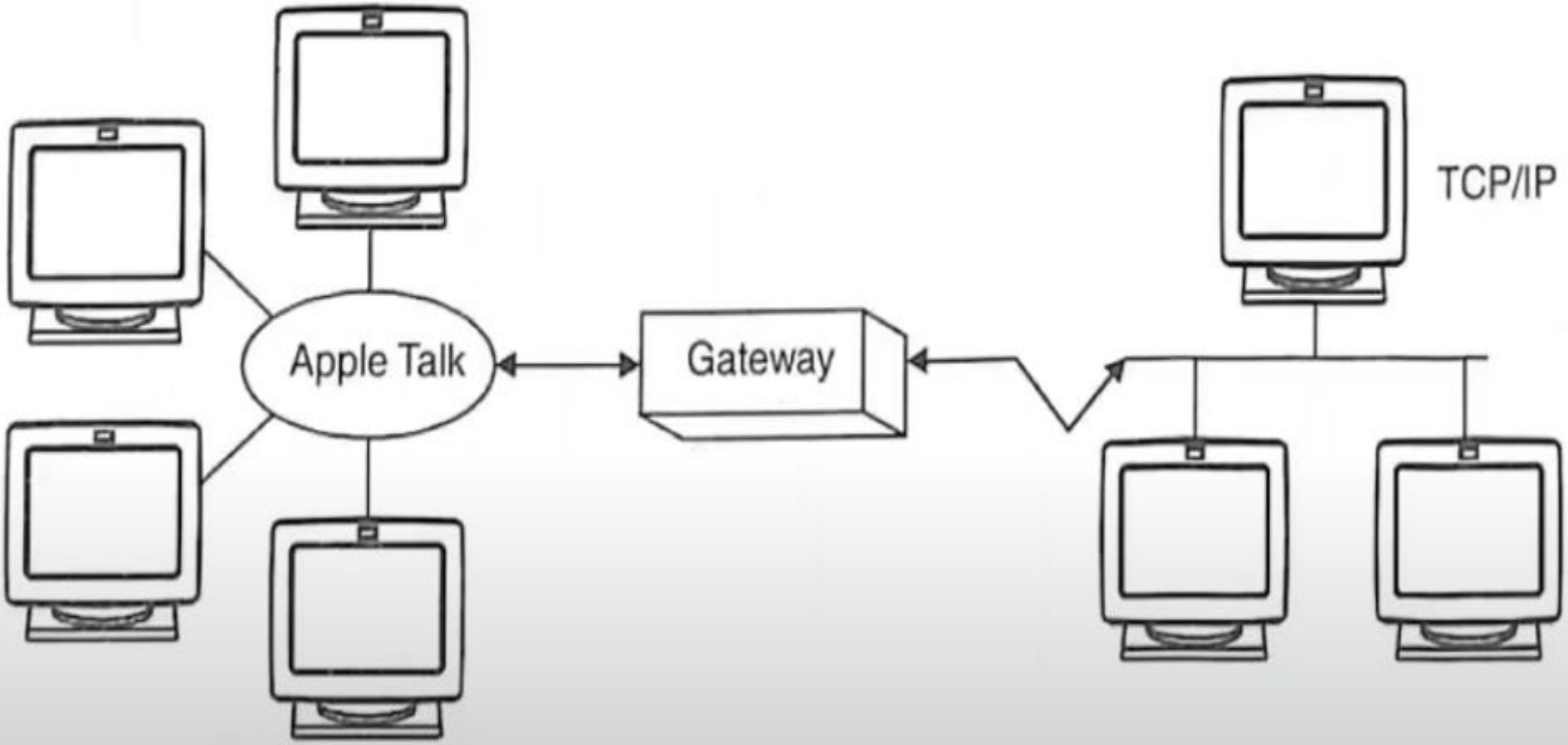
4. ROUTERS



5. GATEWAY

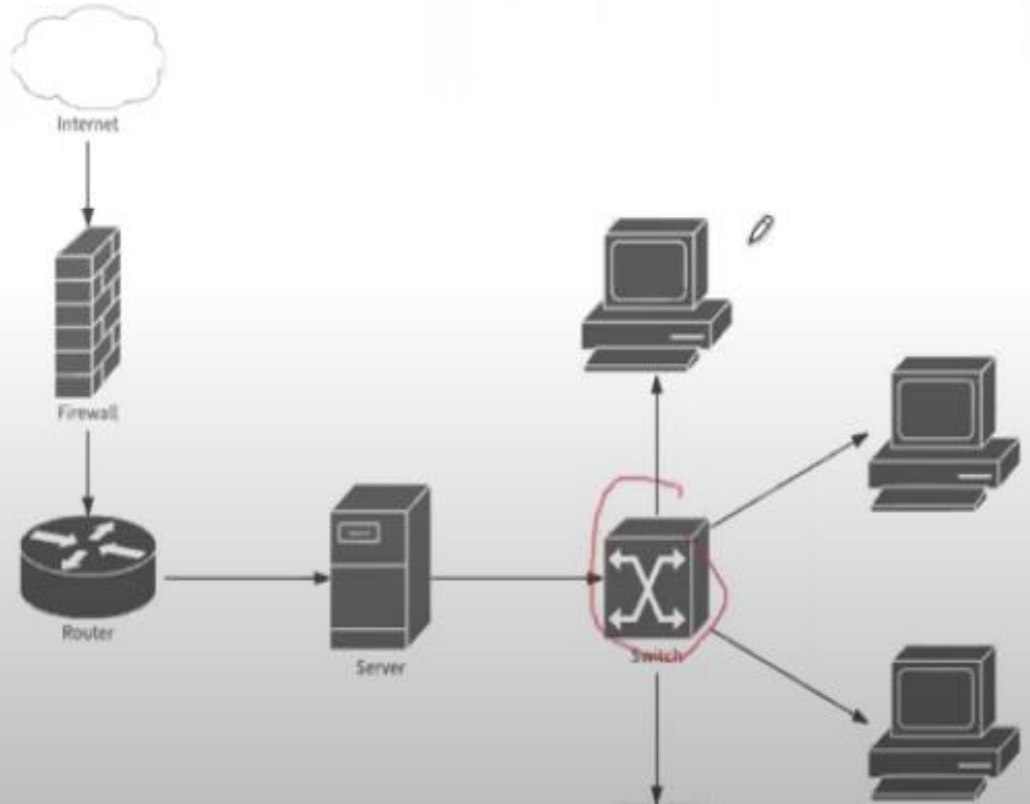
- Gateways are the internetworking devices that operate in all the seven layers of OSI model
- A gateway act as a protocol converter. It is used to connect two different type of network that uses different protocols.
- A gateway can accept a packet formatted for one protocol (e.g. Apple Talk) and convert it to a packet formatted for another protocol (e.g. TCP/IP) before forwarding it
- In addition, gateways can perform all the functions of bridges and routers.
- The gateway links two systems that do not use same communication protocols, data formatting structures, languages and architecture.

5. GATEWAY



6. NETWORK SWITCH

- Switches are used to connect multiple devices together on the same network.
- In a properly designed network, LAN switches are responsible for directing and controlling the data the flow at the access layer to networked resources.



ADDRESS MAPPING

IP Address

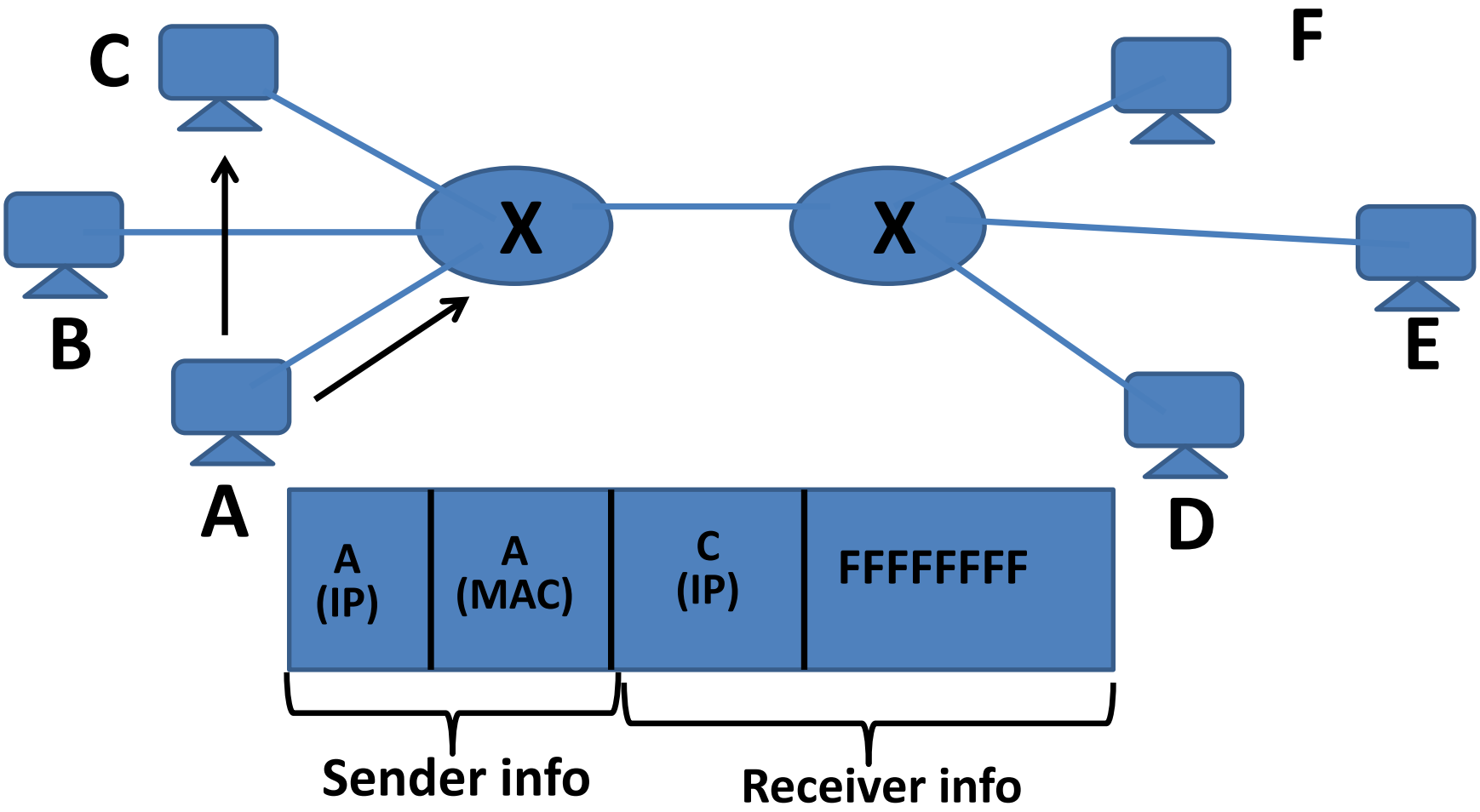
Physical IP
(MAC)
(Private)
48 Bit or 6 Byte

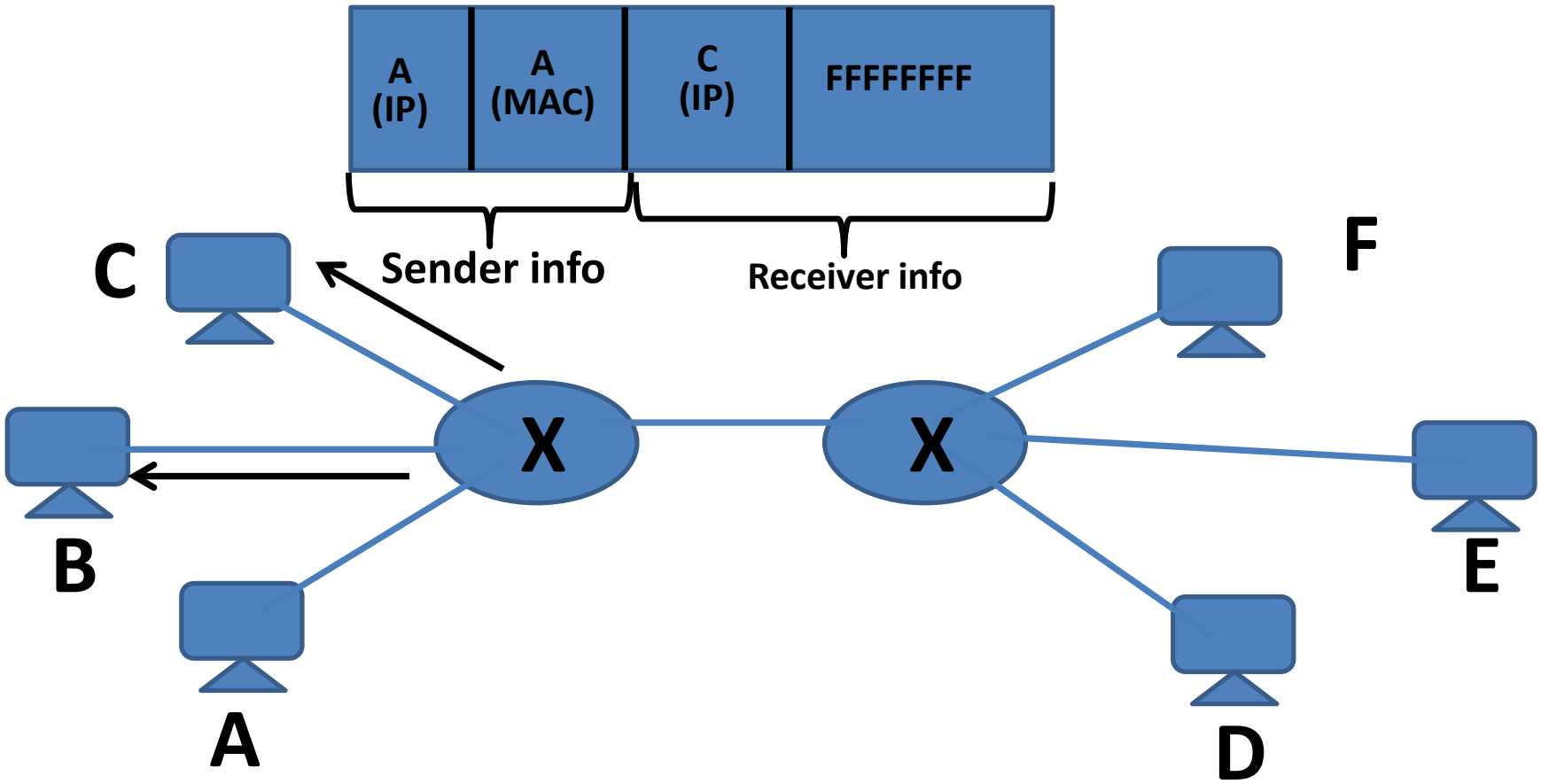
Logical IP
(IP)
(Publically Shared)
32 Bit or 4 Byte

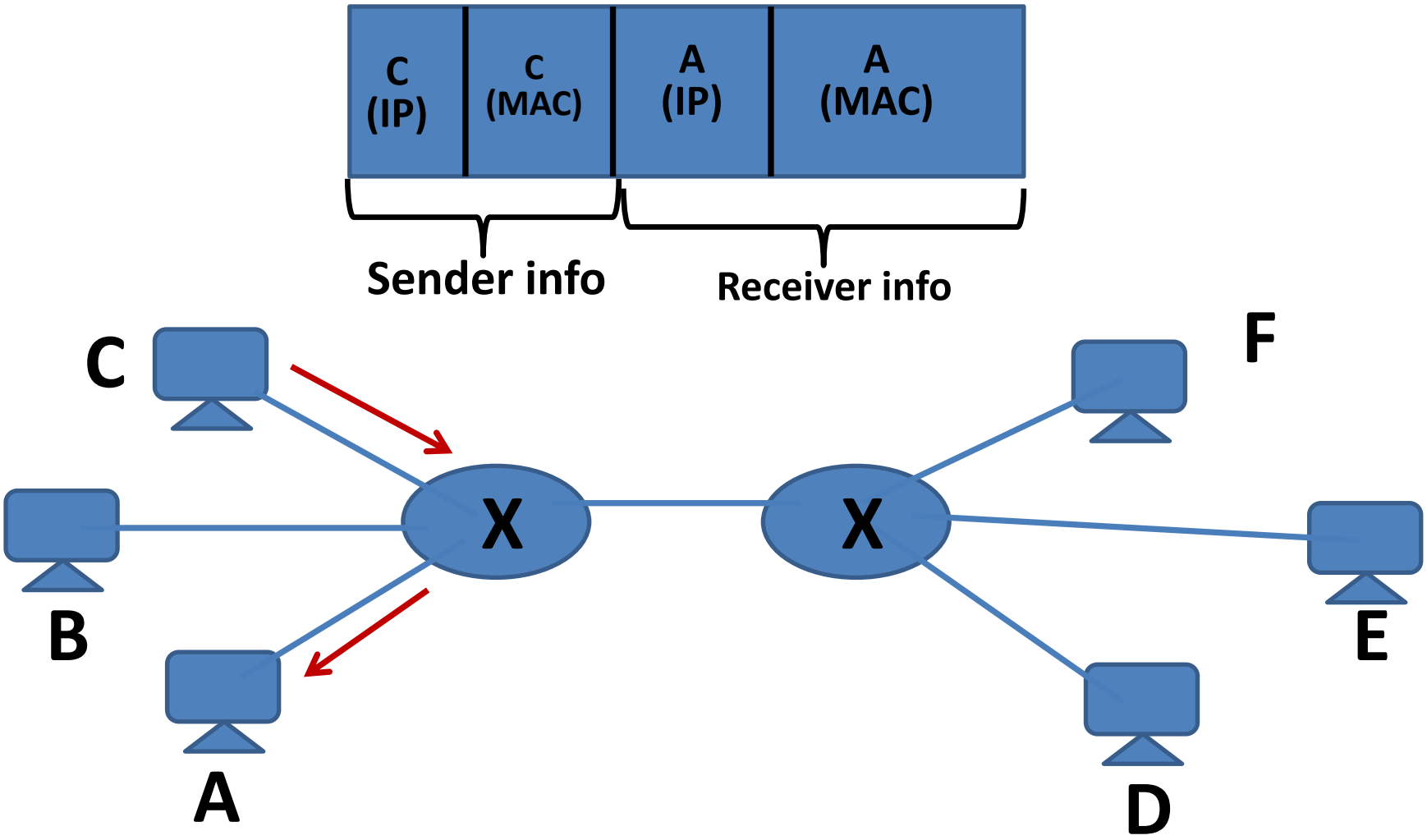
Mapping

Logical IP → **Physical IP**

BroadCasting to establish connection for data transfer
HANDSHAKING

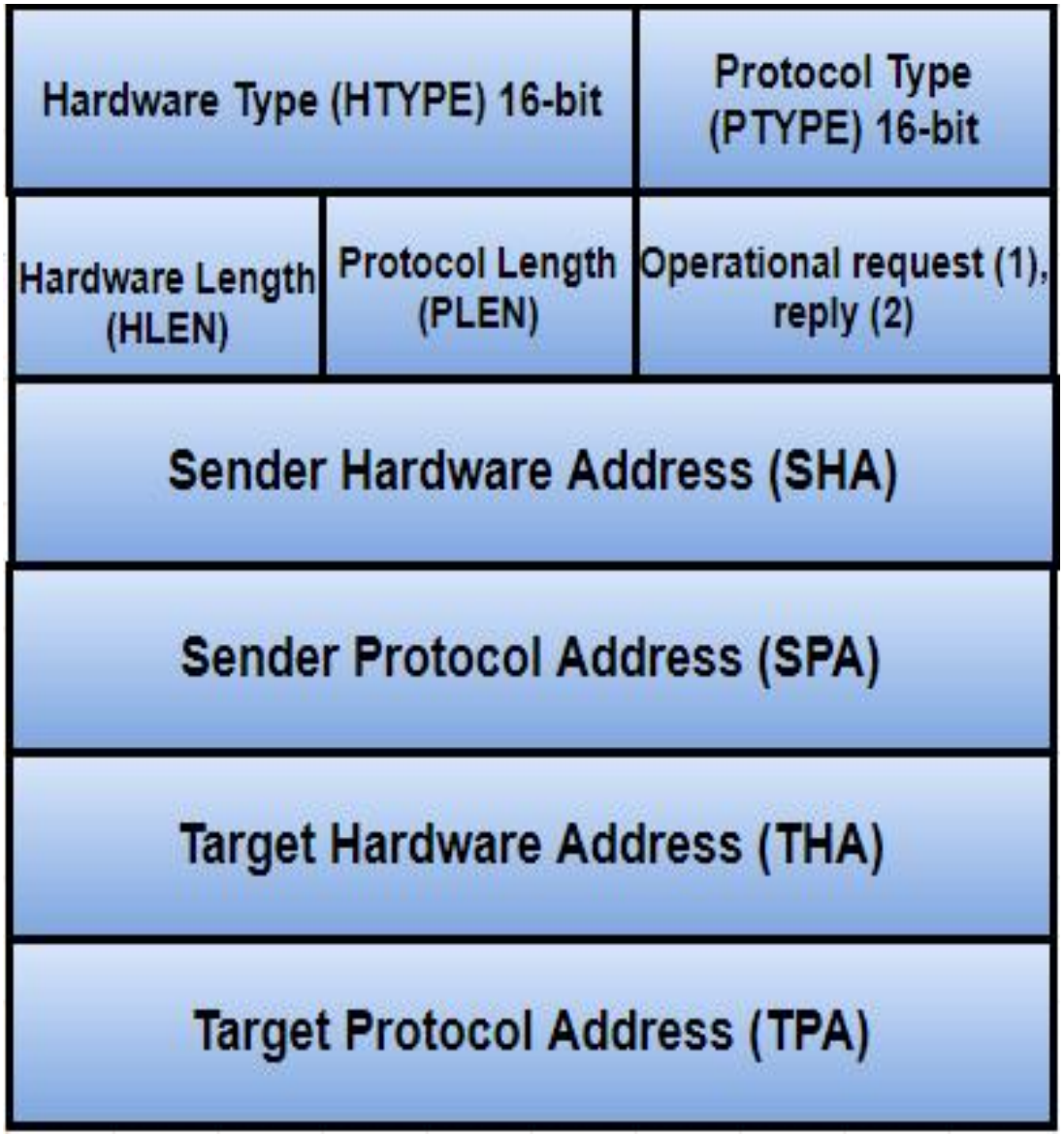

C
F
X
X
B
E
A
D
A (IP)
A (MAC)
C (IP)
FFFFFFFF
Sender info
Receiver info


A (IP)
A (MAC)
C (IP)
FFFFFFFF
Sender info
Receiver info
C
F
X
X
B
E
A
D


C
(IP)
C
(MAC)
A
(IP)
A
(MAC)
Sender info
Receiver info
C
F
X
X
B
E
A
D

Address Mapping

- Host to Host (H to H)
- Host to Router (H to R)
- Router to Host (R to H)
- Router to Router (R to R)

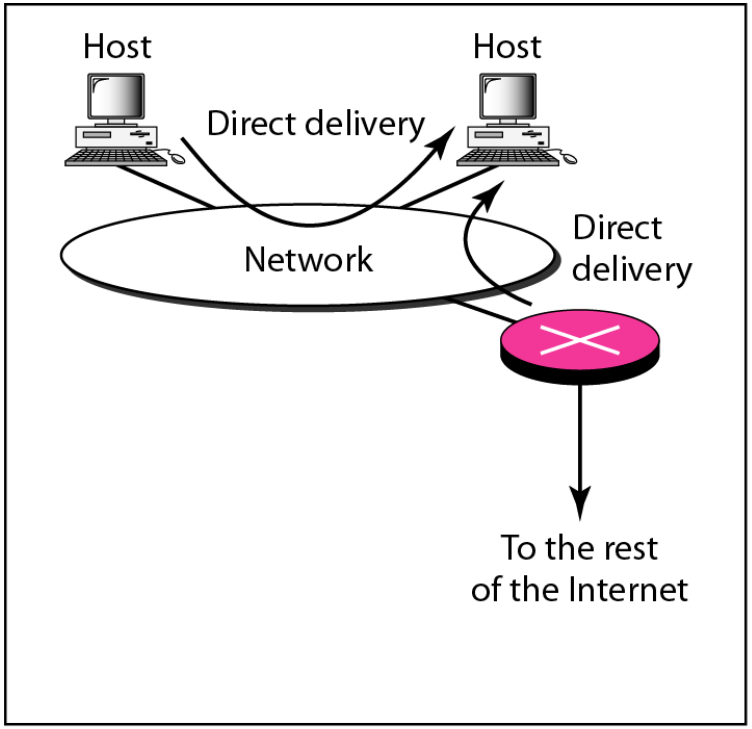


FORWARDING
&
DELIVERY

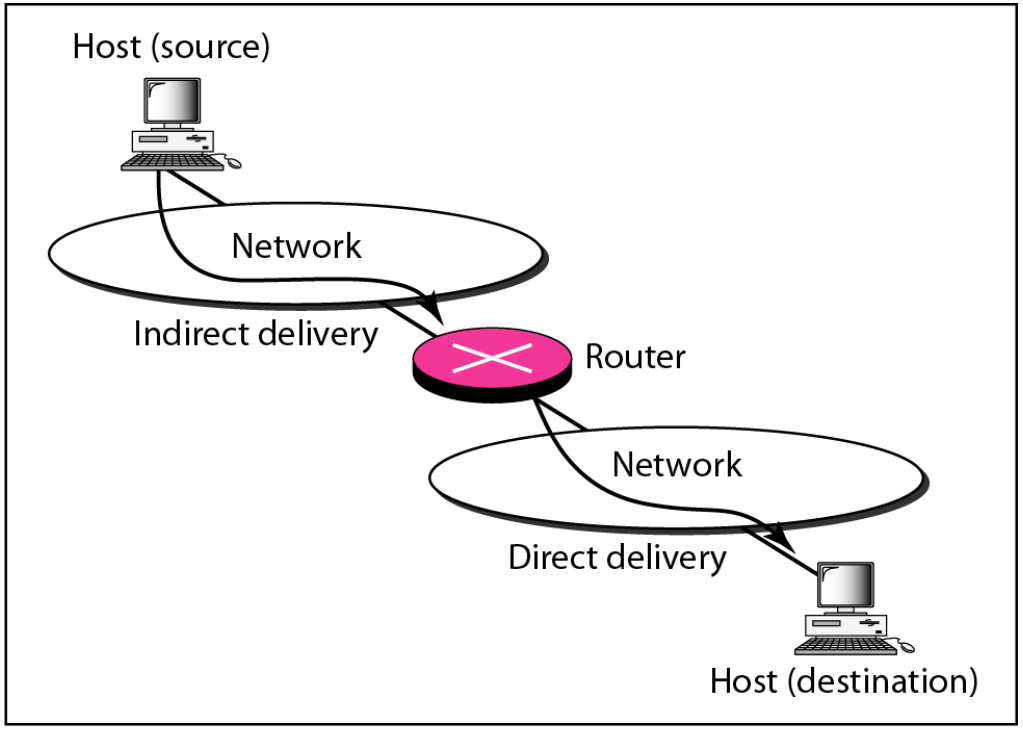
DELIVERY

The network layer supervises the handling of the packets by the underlying physical networks. We define this handling as the delivery of a packet.

Figure 22.1 *Direct and indirect delivery*



a. Direct delivery



b. Indirect and direct delivery

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.

Figure 22.2 *Route method versus next-hop method*

a. Routing tables based on route

Destination	Route
Host B	R1, R2, host B

Routing table for host A

Destination	Route
Host B	R2, host B

Routing table for R1

Destination	Route
Host B	Host B

Routing table for R2

b. Routing tables based on next hop

Destination	Next hop
Host B	R1

Destination	Next hop
Host B	R2

Destination	Next hop
Host B	---

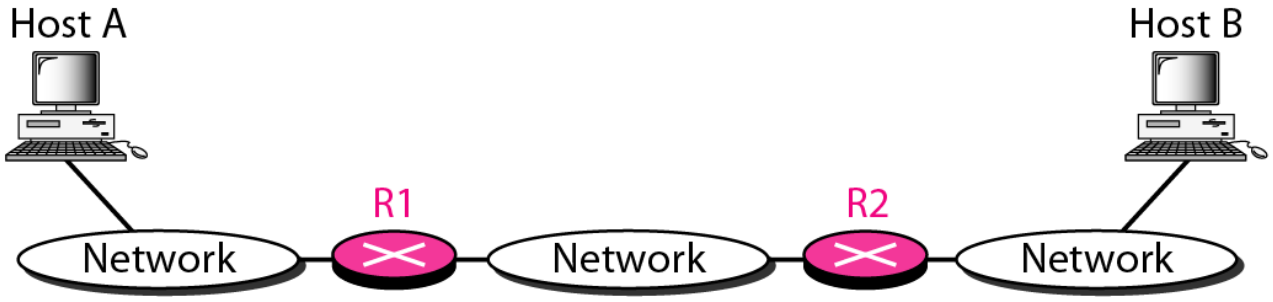


Figure 22.3 *Host-specific versus network-specific method*

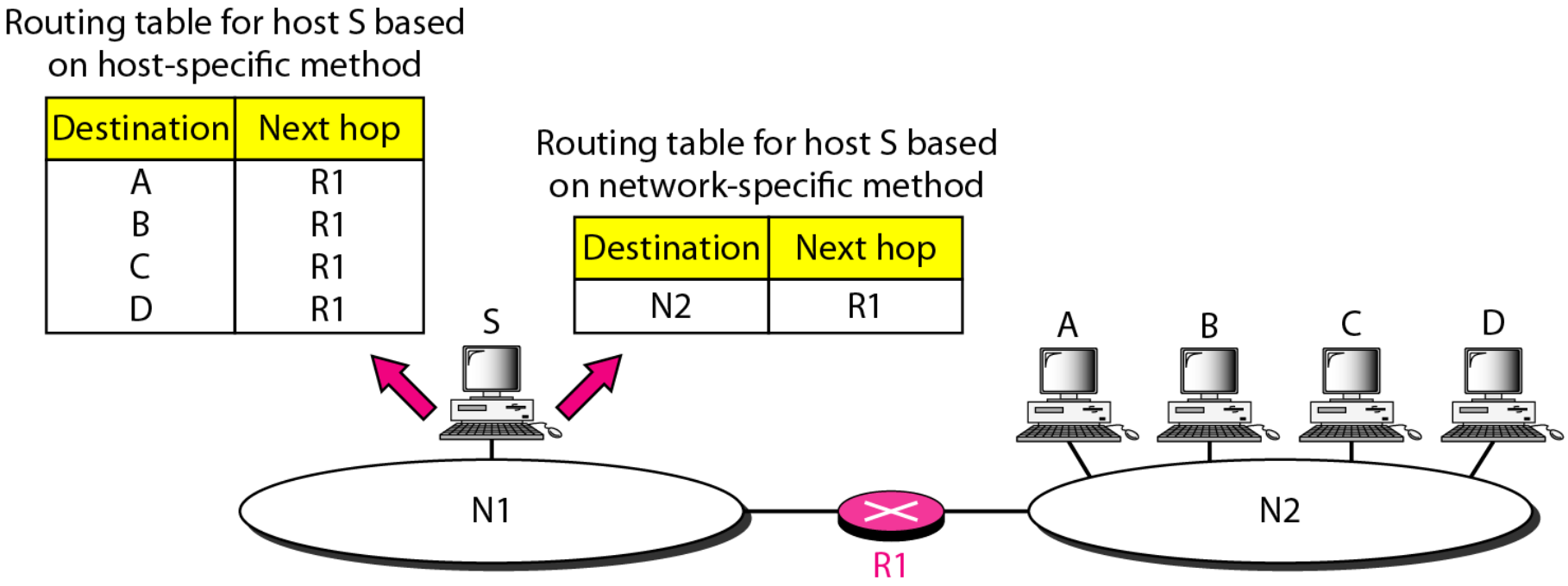


Figure 22.4 *Default method*

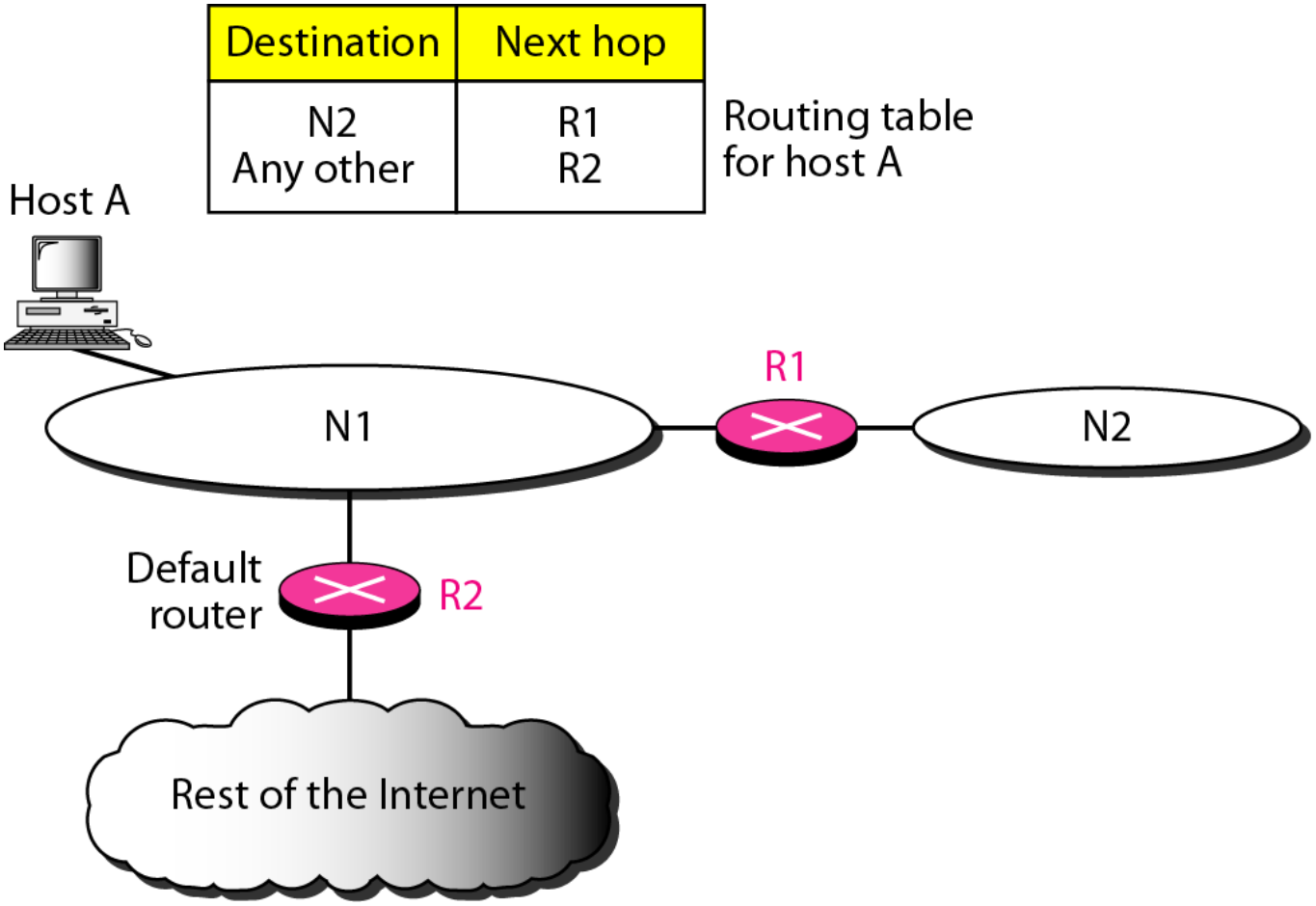
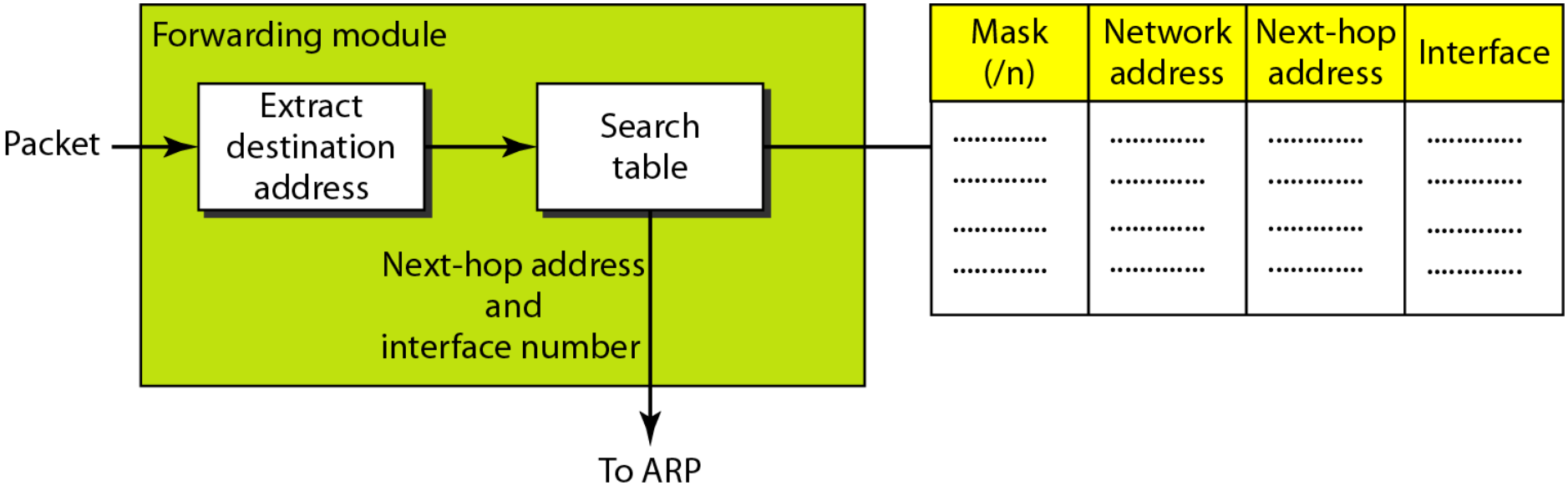


Figure 22.5 *Simplified forwarding module in classless address*



Make a routing table for router R1, using the configuration in Figure 22.6.

Solution

Table 22.1 shows the corresponding table.

Figure 22.6 *Configuration for Example 22.1*

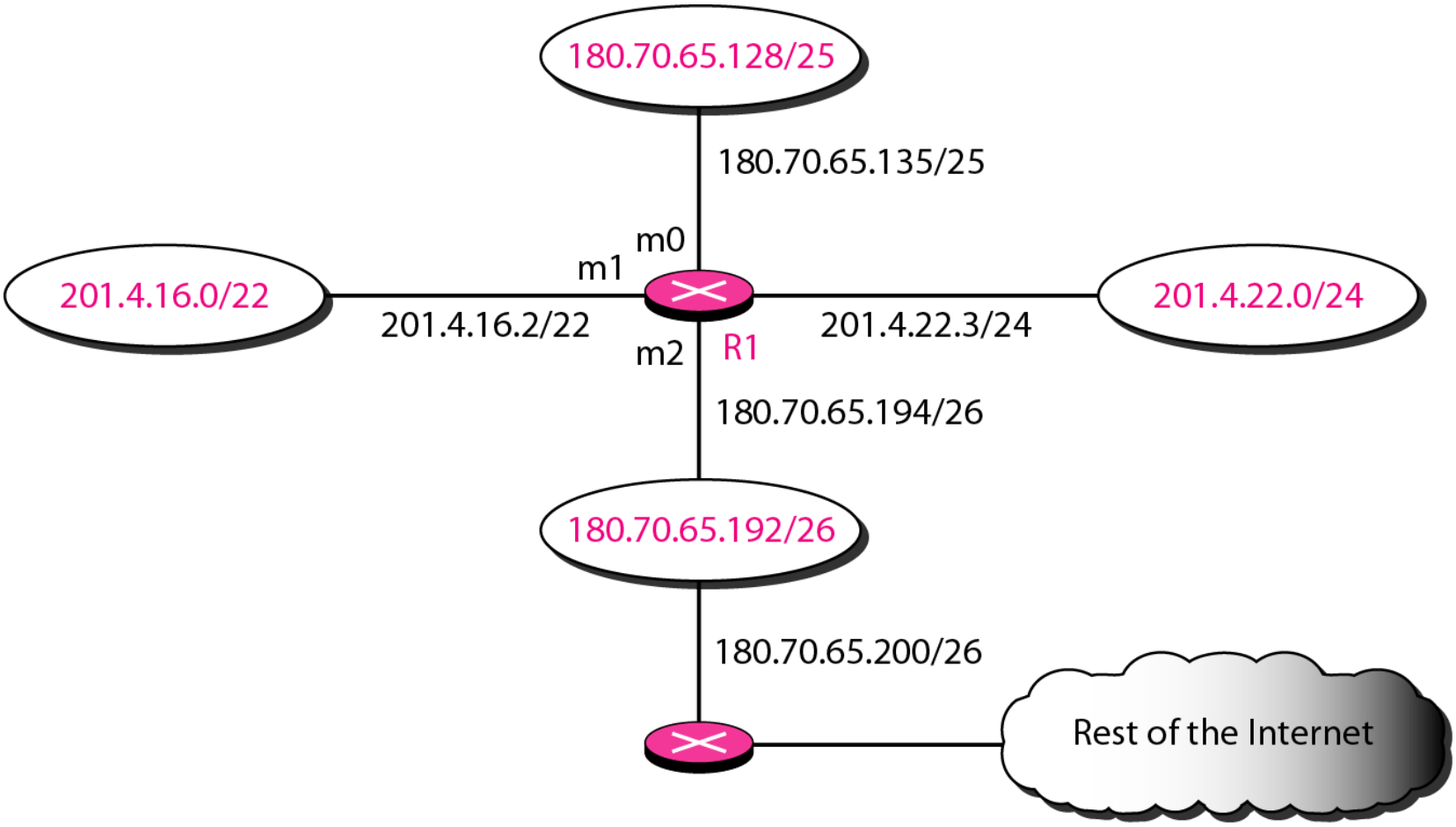


Table 22.1 *Routing table for router R1 in Figure 22.6*

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

Figure 20.5 *IPv4 datagram format*

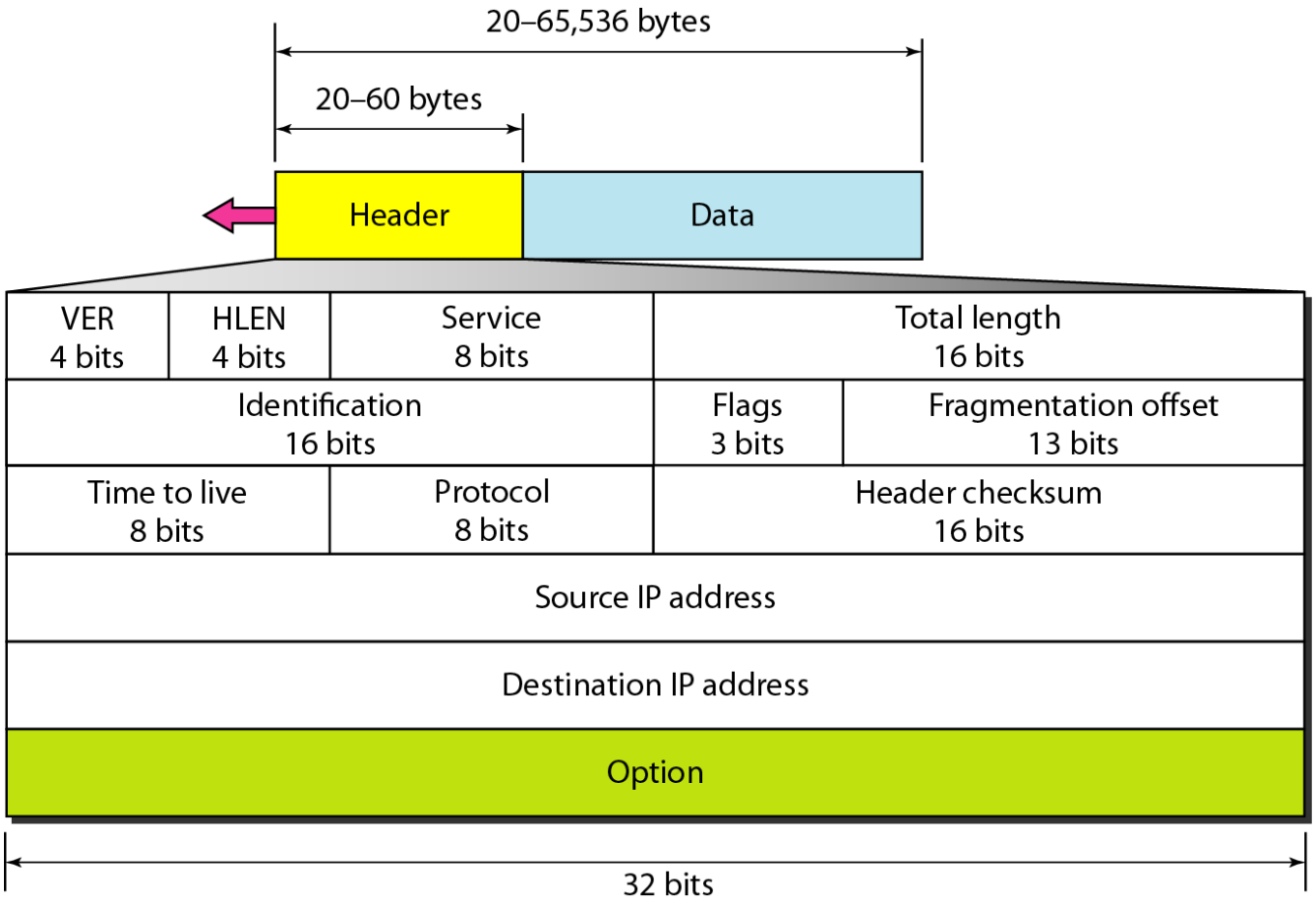
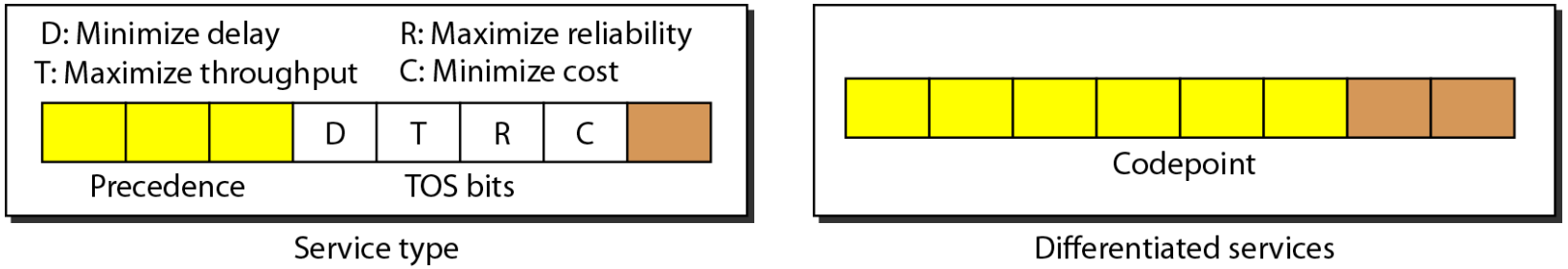


Figure 20.6 *Service type or differentiated services*



Note

The precedence subfield was part of version 4, but never used.

Table 20.1 *Types of service*

TOS Bits	*Description*
0000	Normal (default)
0001	Minimize cost
0010	Maximize reliability
0100	Maximize throughput
1000	Minimize delay

Table 20.2 *Default types of service*

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

Table 20.3 ***Values for codepoints***

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

Note

The total length field defines the total length of the datagram including the header.

Figure 20.7 *Encapsulation of a small datagram in an Ethernet frame*

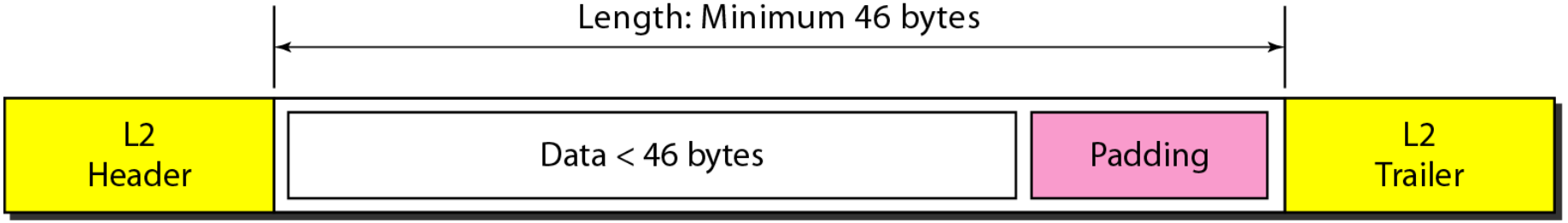


Figure 20.8 *Protocol field and encapsulated data*

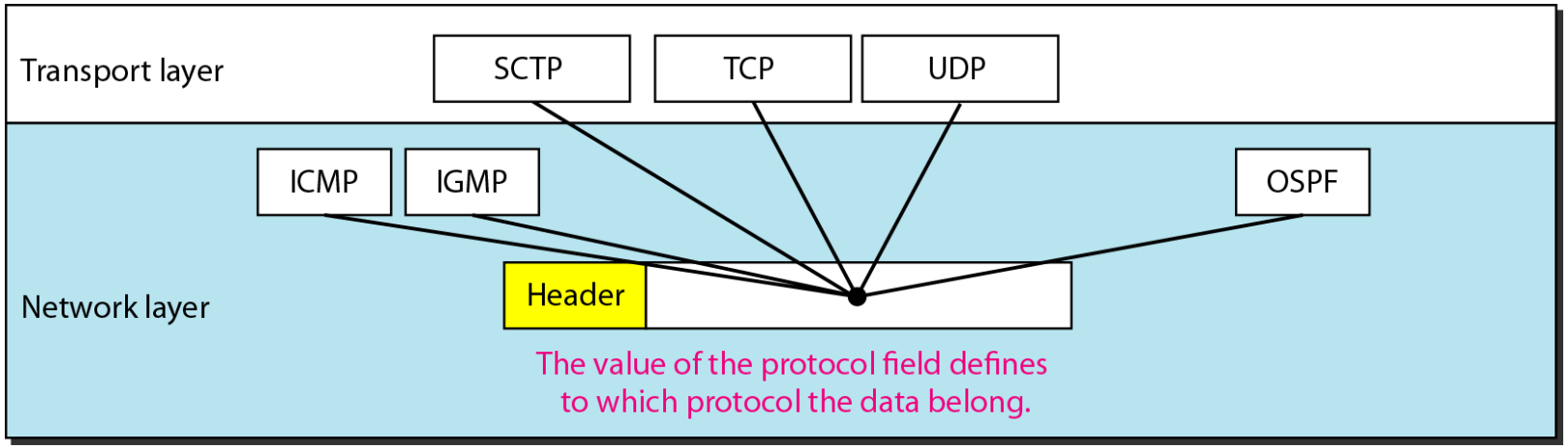


Table 20.4 ***Protocol values***

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

Example 20.1

An IPv4 packet has arrived with the first 8 bits as shown:

01000010

The receiver discards the packet. Why?

Solution

There is an error in this packet. The 4 leftmost bits (0100) show the version, which is correct. The next 4 bits (0010) show an invalid header length ($2 \times 4 = 8$). The minimum number of bytes in the header must be 20. The packet has been corrupted in transmission.

Example 20.2

In an IPv4 packet, the value of HLEN is 1000 in binary. How many bytes of options are being carried by this packet?

Solution

The HLEN value is 8, which means the total number of bytes in the header is 8 × 4, or 32 bytes. The first 20 bytes are the base header, the next 12 bytes are the options.

Example 20.3

In an IPv4 packet, the value of HLEN is 5, and the value of the total length field is 0x0028. How many bytes of data are being carried by this packet?

Solution

The HLEN value is 5, which means the total number of bytes in the header is 5 × 4, or 20 bytes (no options). The total length is 40 bytes, which means the packet is carrying 20 bytes of data (40 − 20).

Example 20.4

An IPv4 packet has arrived with the first few hexadecimal digits as shown.

0x45000028000100000102 . . .

How many hops can this packet travel before being dropped? The data belong to what upper-layer protocol?

Solution

To find the time-to-live field, we skip 8 bytes. The time-to-live field is the ninth byte, which is 01. This means the packet can travel only one hop. The protocol field is the next byte (02), which means that the upper-layer protocol is IGMP.

Figure 20.9 *Maximum transfer unit (MTU)*

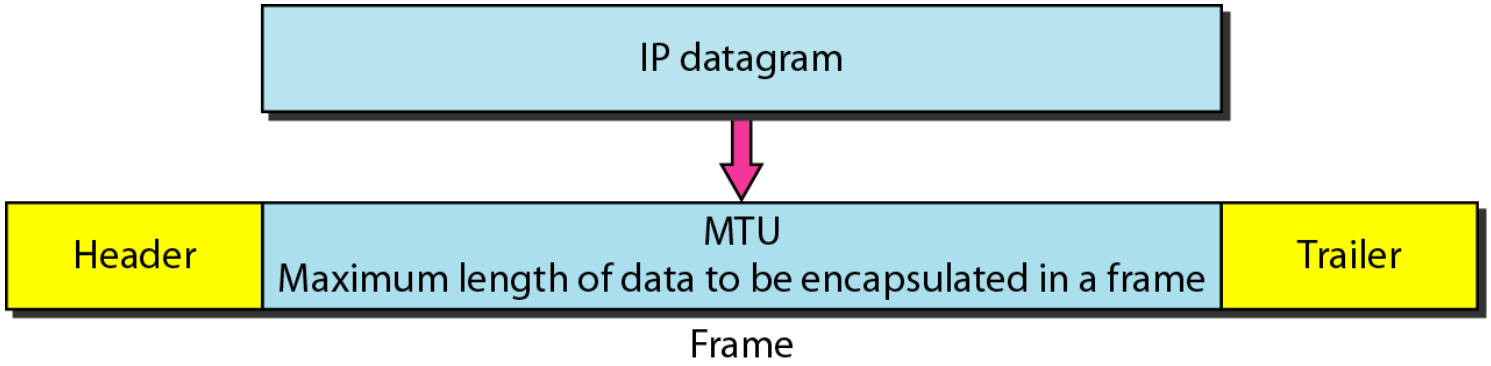
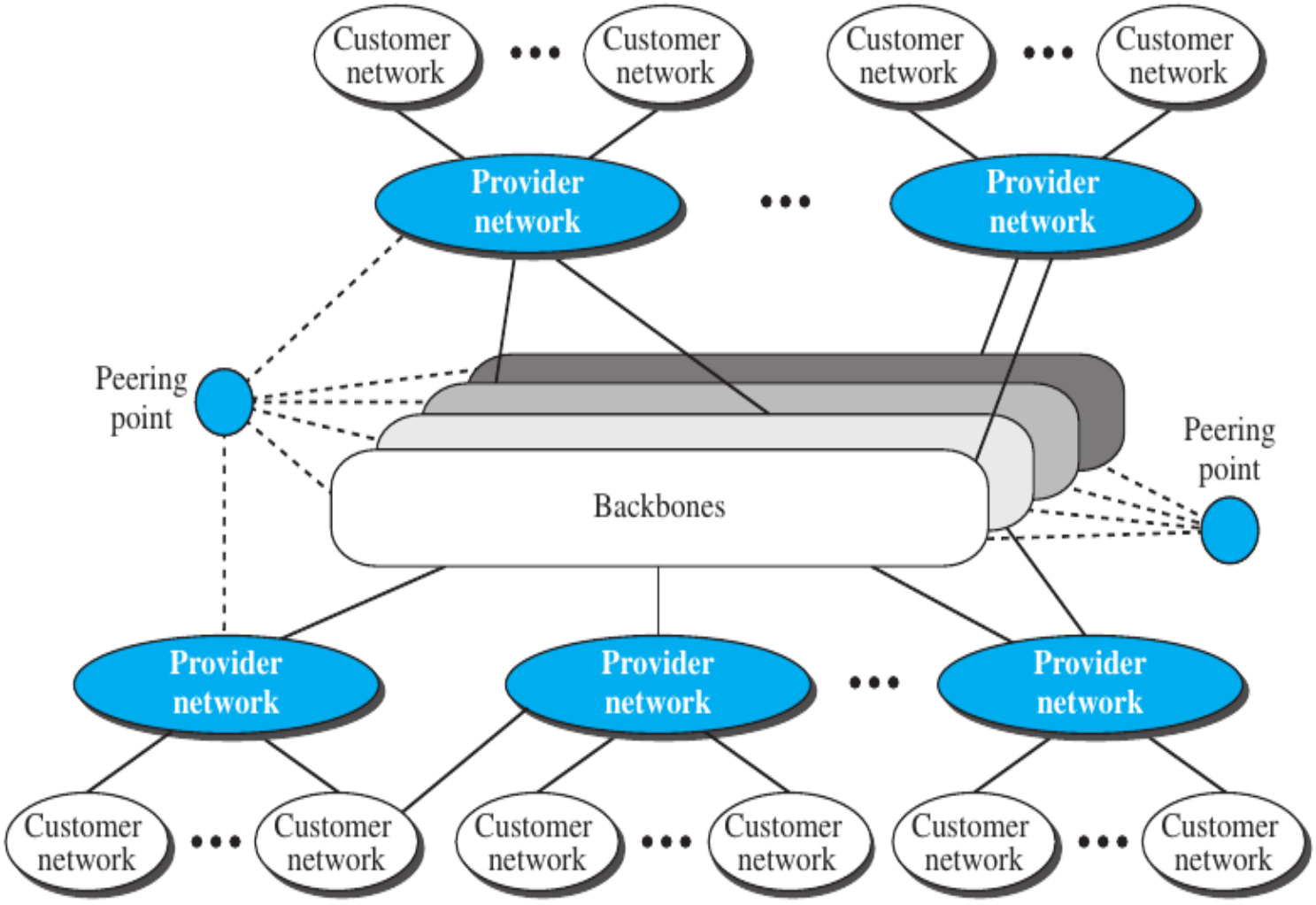
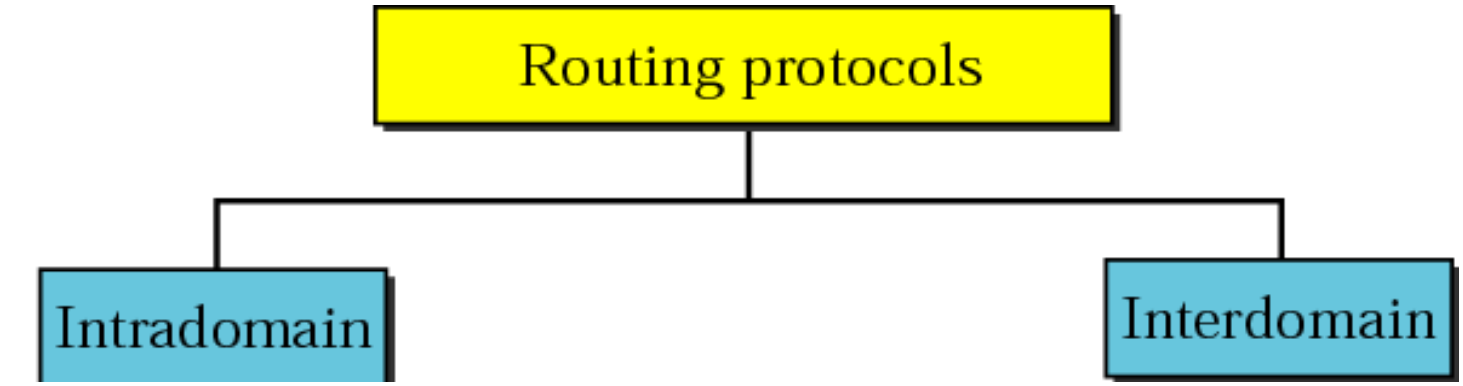


Table 20.5 *MTUs for some networks*

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	296

ROUTING


Customer network
Customer network
Customer network
Customer network
Provider network
Provider network
Peering point
Backbones
Peering point
Provider network
Provider network
Provider network
Customer network
Customer network
Customer network
Customer network
Customer network
Customer network



With in Autonomous System

Between Autonomous System

Figure 14.1 *Autonomous systems*

An *autonomous system* is a set of networks and routers under the control of a single administrative authority.
Routing within an autonomous system is *intra*domain routing.
Routing between autonomous systems is *inter*domain routing.

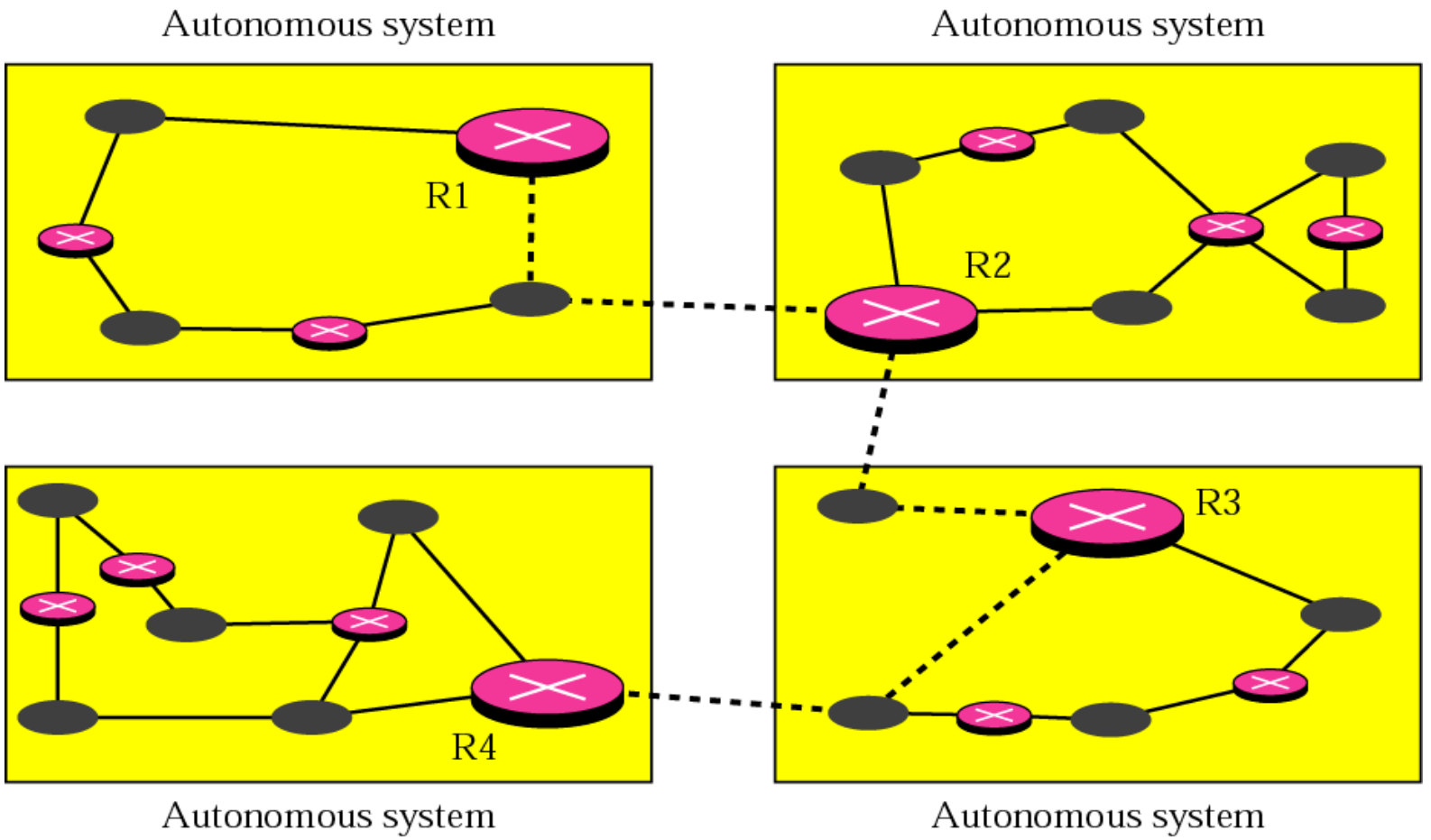
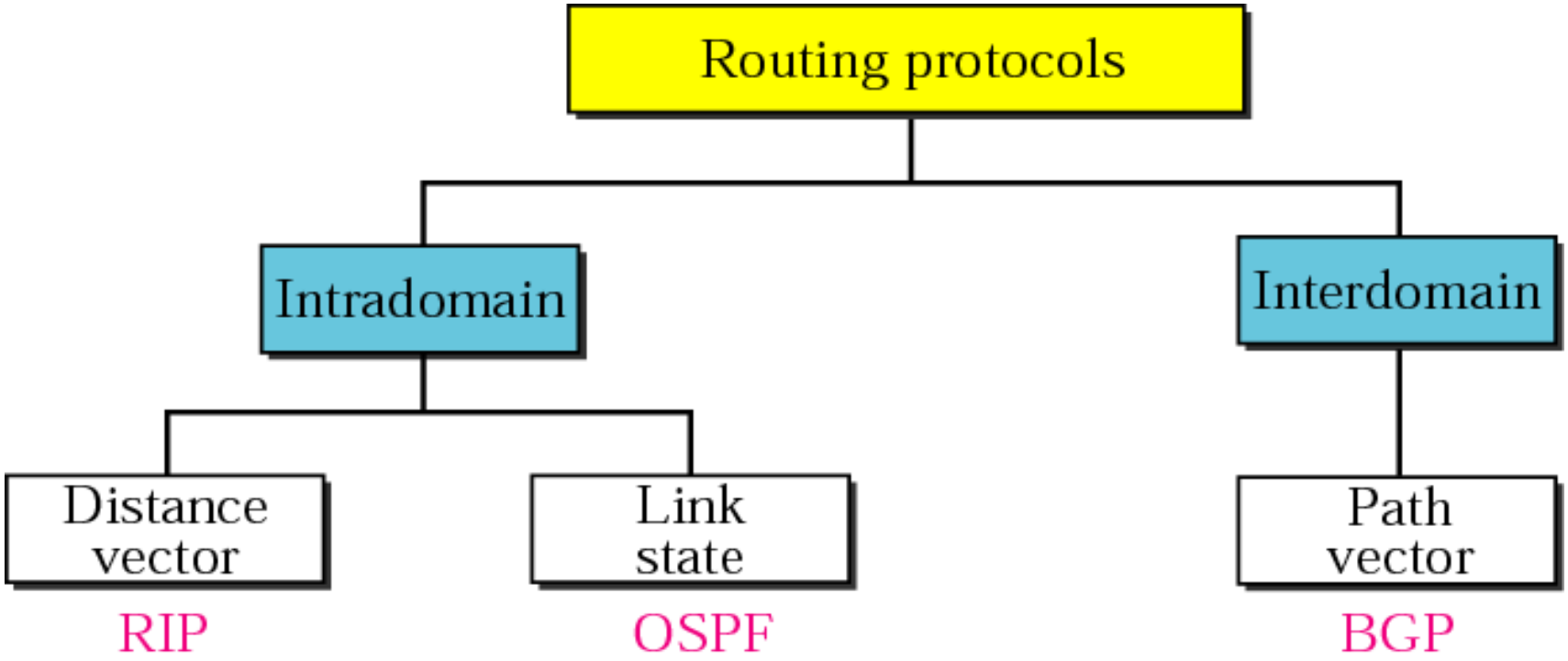


Figure 14.2 *Popular routing protocols*



14.2 DISTANCE VECTOR ROUTING

In distance vector routing, the least cost route between any two nodes is the route with minimum distance.

In this protocol each node maintains a vector (table) of minimum distances to every node

Figure 14.3 *Distance vector routing tables*

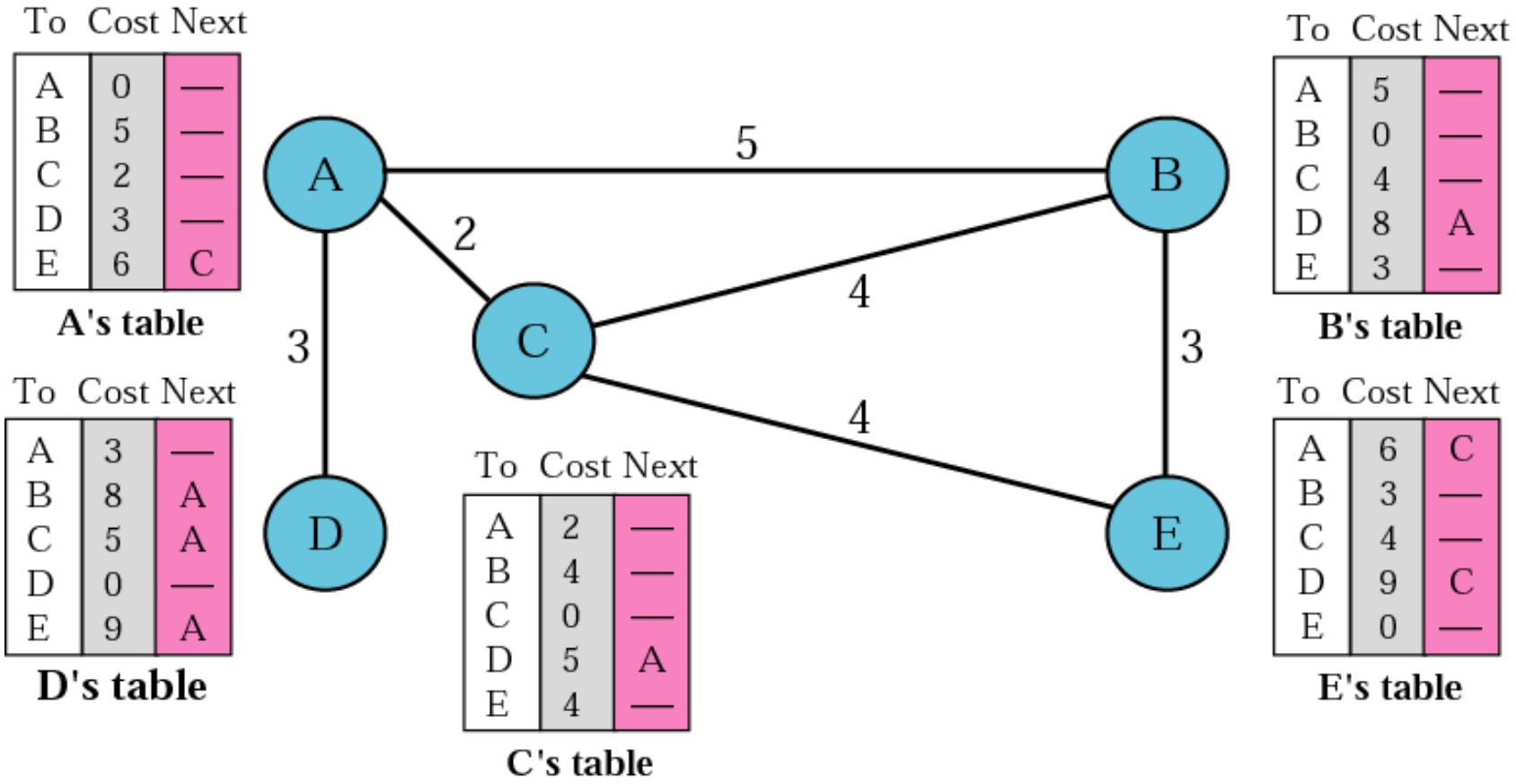
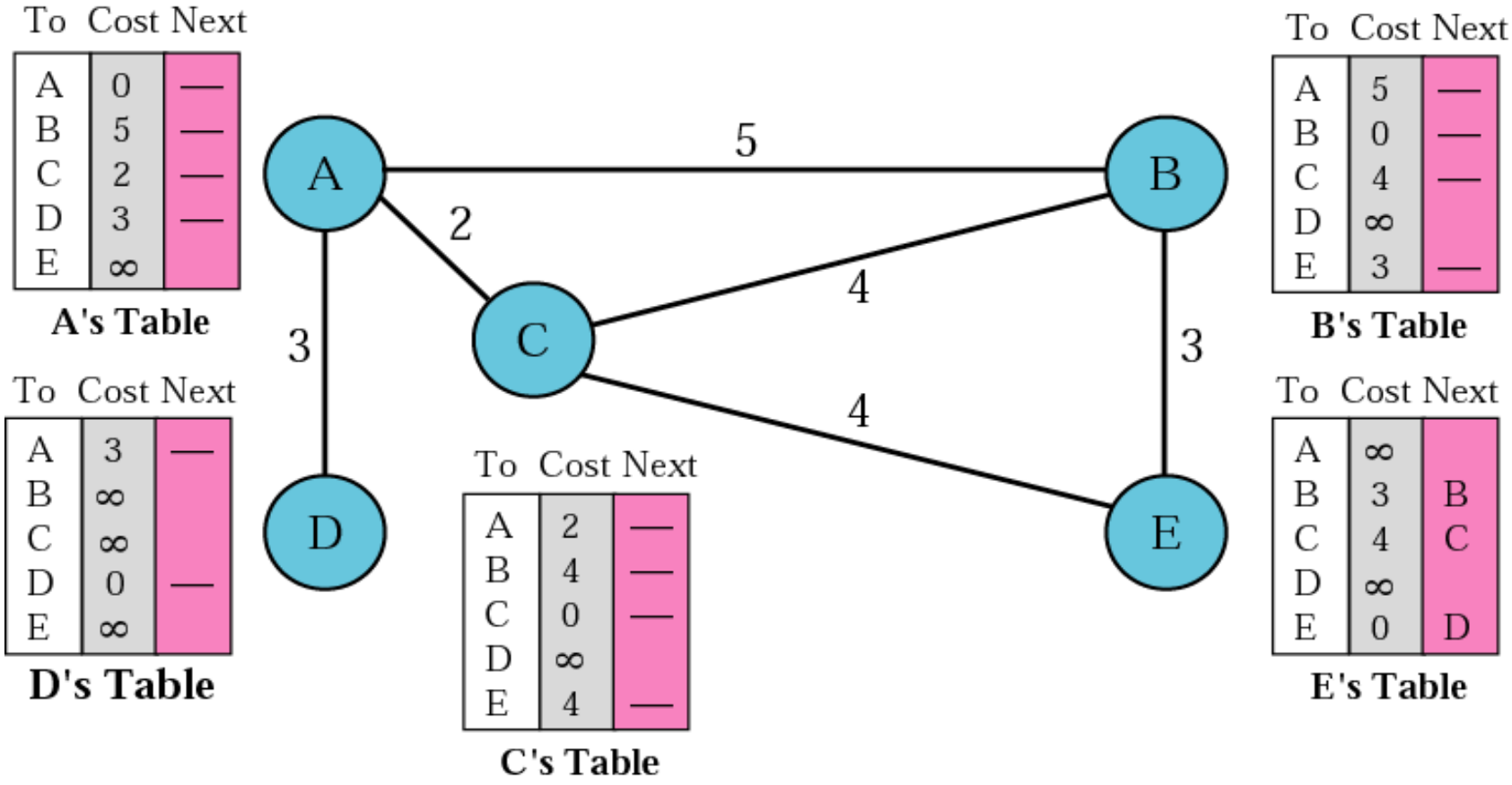


Figure 14.4 *Initialization of tables in distance vector routing*



In distance vector routing, each node shares its table with its immediate neighbor periodically (eg every 30s) and when there is a change.

Figure 14.5 *Updating in distance vector routing*

Step 1: Add cost (2) to table received from neighbor (C).
Step 2: Compare Modified Table with Old Table (row by row).
If Next node entry is different, select the row with the smaller cost. If tie, keep the old one.
If Next node entry the same, select the new row value (regard-less of whether new value is smaller or not).

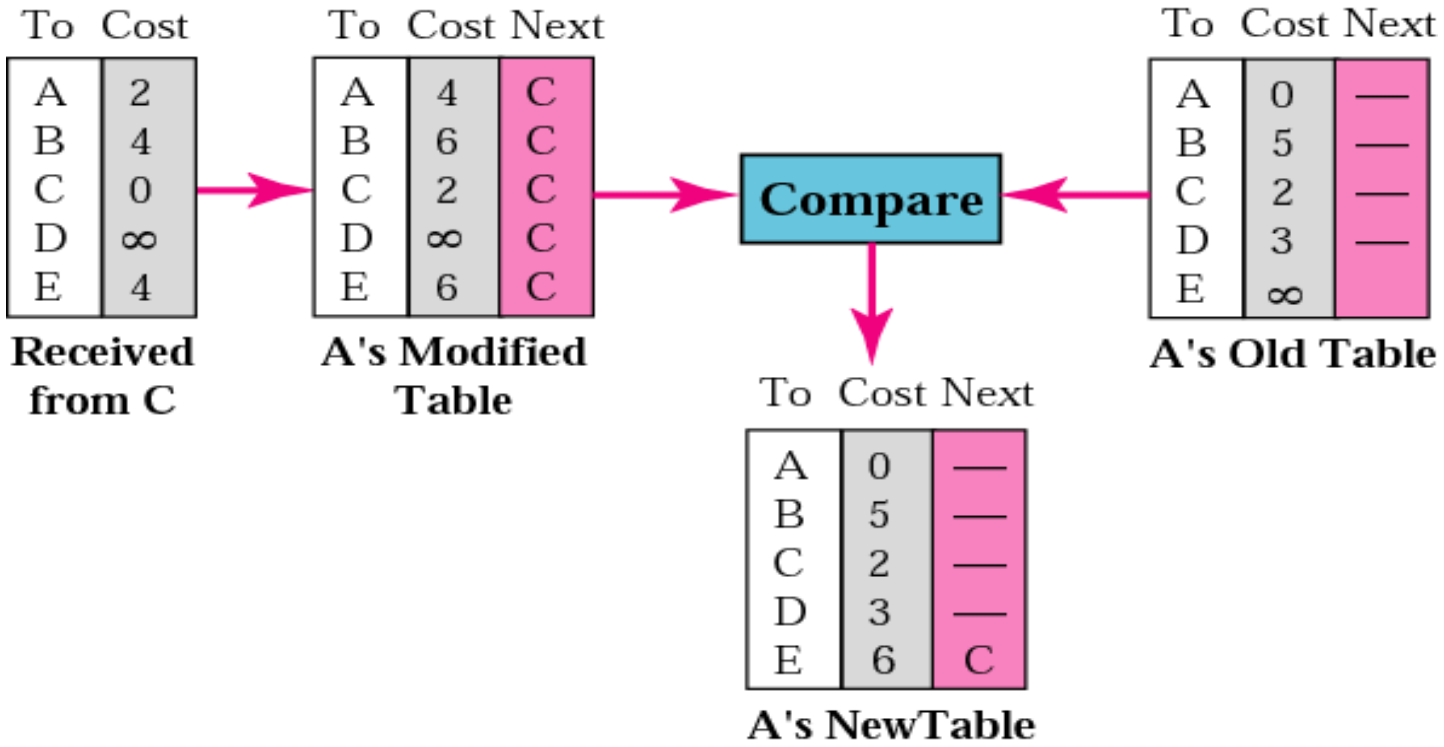
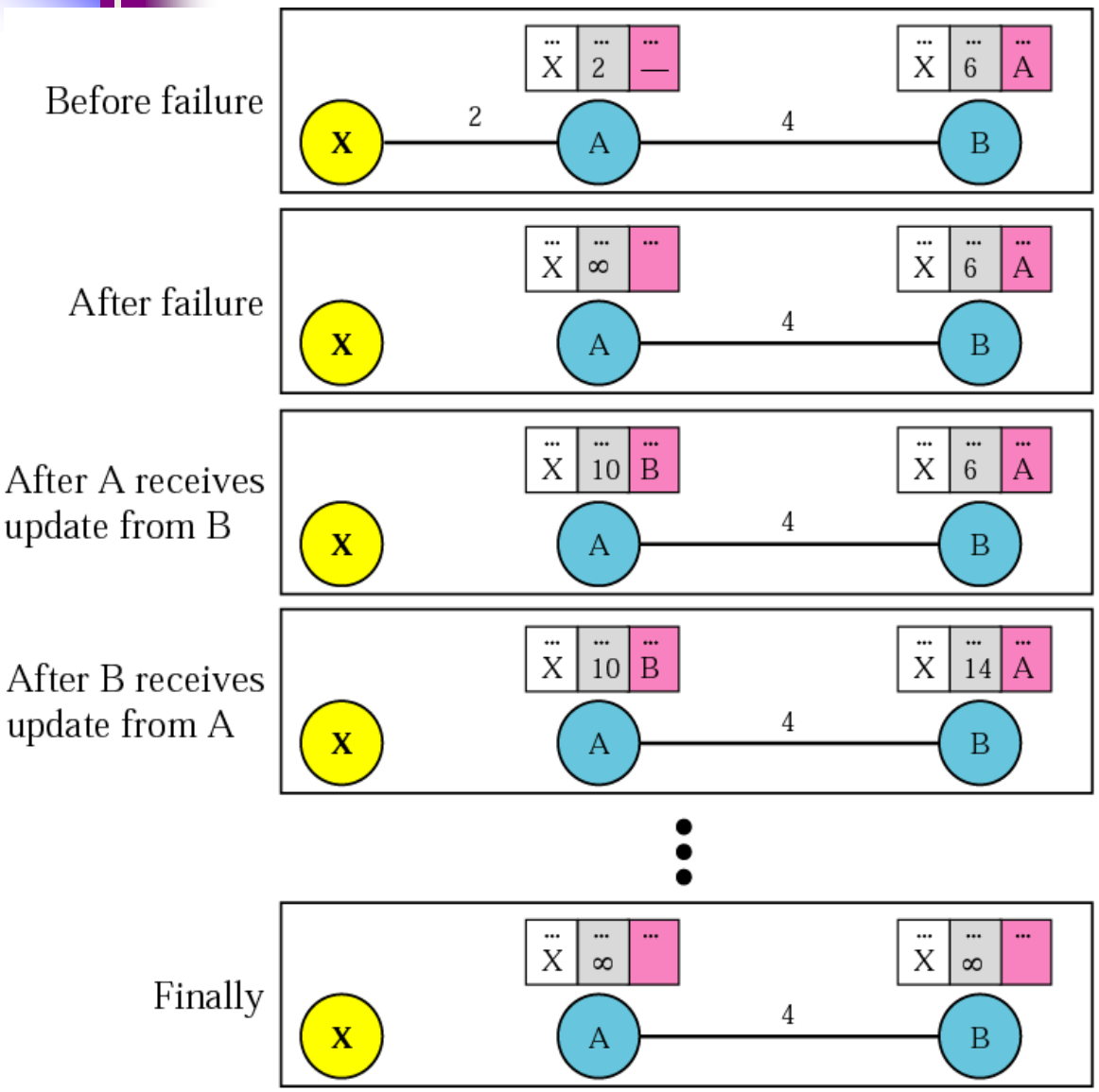


Figure 14.6 *Two-node instability – what can happen with distance vector routing*

Count to Infinity Problem



Both A and B know where X is.

Link between A and X fails. A updates its table immediately. But before A can tell B, B sends its info to A!

A, using B's info, updates its table (error!).

Then A send its table to B and B updates its table (more error). Both routers keep updating tables, eventually hitting infinity. In the meantime, chaos!

Figure 14.6 *Two-node instability – what can happen with distance vector routing*

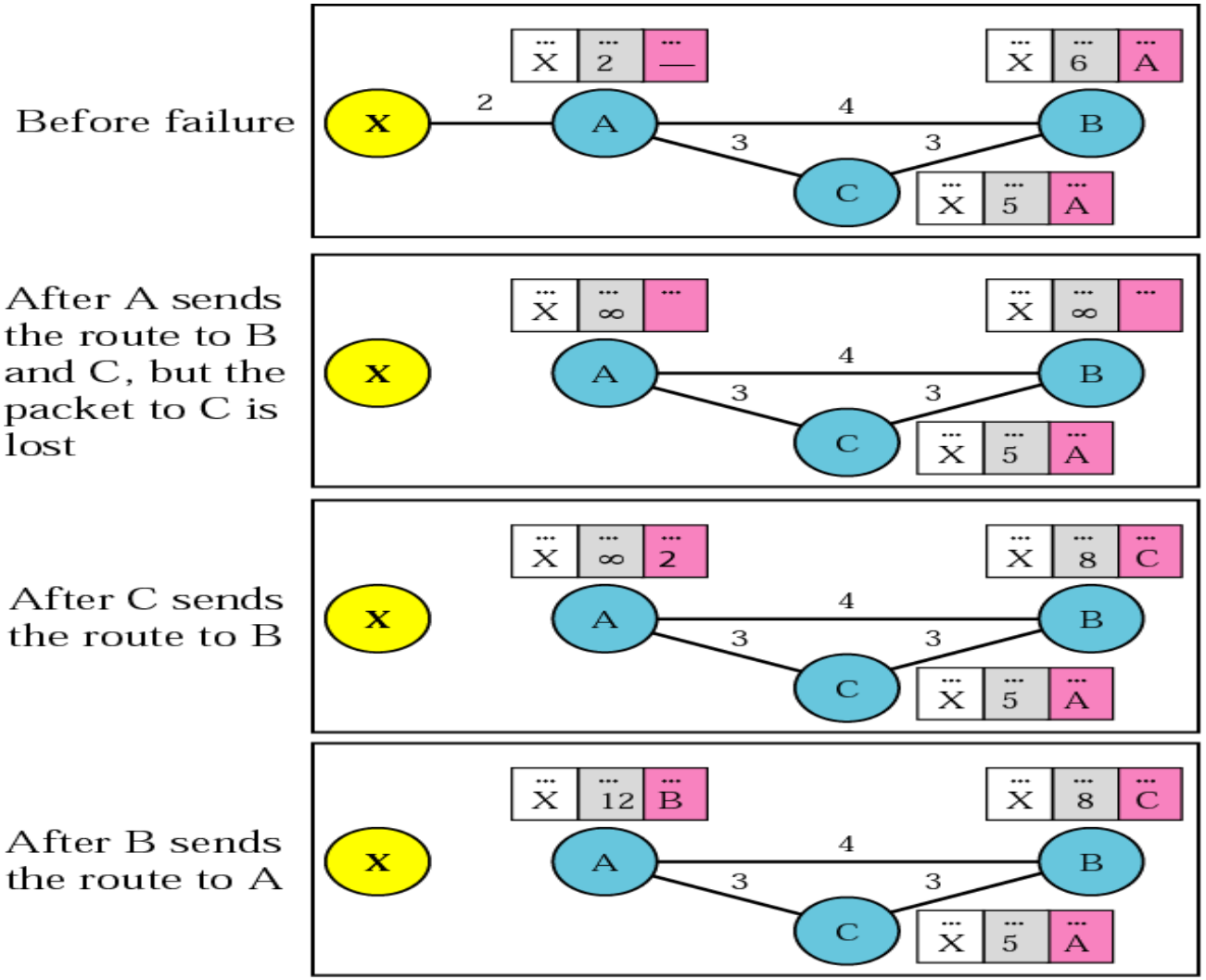
Possible Solutions to two-node instability:

1. Define infinity to be a much smaller value, such as 100. Then it doesn't take too long to become stable. But now you can't use distance vector routing in large networks.

2. **Split Horizon** – instead of flooding entire table to each node, only part of its table is sent. More precisely, if node B thinks that the optimum router to reach X is via A, then B does not need to advertise this piece of info to A – the info has already come from A.

3. **Split Horizon and Poison Reverse** – Normally, the distance vector protocol uses a timer. If there is no news about a route, the node deletes the route from its table. So when A never hears from B about the route to X, it deletes it. Instead, Node B still advertises the value for X, but if the source of info is A, it replaces the distance with infinity, saying "Do not use this value; what I know about this route comes from you."

Three-node instability – no solutions here!



14.3 RIP

The Routing Information Protocol (RIP) is an intradomain routing protocol used inside an autonomous system.

It is a very simple protocol based on distance vector routing.

RIP

The **Routing Information Protocol (RIP)** is an intradomain routing protocol used inside an autonomous system. It is a very simple protocol based on distance vector routing. RIP implements distance vector routing directly with some considerations:

1. In an autonomous system, we are dealing with routers and networks (links). The routers have routing tables; networks do not.
2. The destination in a routing table is a network, which means the first column defines a network address.
3. The metric used by RIP is very simple; the distance is defined as the number of links (networks) to reach the destination. For this reason, the metric in RIP is called a **hop count.**
4. Infinity is defined as 16, which means that any route in an autonomous system using RIP cannot have more than 15 hops.
5. The next-node column defines the address of the router to which the packet is to be sent to reach its destination.

- RIP implements the same algorithm as the distance-vector routing algorithm we discussed in the previous section. However, some changes need to be made to the algorithm to enable a router to update its forwarding table:

❑ Instead of sending only distance vectors, a router needs to send the whole contents of its forwarding table in a response message.

❑ The receiver adds one hop to each cost and changes the next router field to the address of the sending router. We call each route in the modified forwarding table the received route and each route in the old forwarding table the old route.

- The received router selects the old routes as the new ones except in the following

- three cases:
- 1. If the received route does not exist in the old forwarding table, it should be added to the route.
- 2. If the cost of the received route is lower than the cost of the old one, the received route should be selected as the new one.
- 3. If the cost of the received route is higher than the cost of the old one, but the value of the next router is the same in both routes, the received route should be selected as the new one.

Timers in RIP

- RIP uses three timers to support its operation. The periodic timer controls the advertising of regular update messages. Each router has one periodic timer that is randomly to a number between 25 and 35 seconds (to prevent all routers sending their messages at the same time and creating excess traffic). The timer counts down; when zero reached, the update message is sent, and the timer is randomly set once again.
- The expiration timer governs the validity of a route. When a router receives update information for a route, the expiration timer is set to 180 seconds for that particular route. Every time a new update for the route is received, the timer is reset. If there is a problem on internet and no update is received within the allotted 180 seconds, the route is considered expired and the hop count of the route is set to 16, which means the destination unreachable. Every route has its own expiration timer.
- The garbage collection timer used to purge a route from the forwarding table. When the information about a route becomes invalid, the router does not immediately purge that route from its table. Instead, it continues to advertise the route with a metric value of 16. At the same time, a garbage collection timer is set to 120 seconds for that route. When the count reaches zero, the route is purged from the table. This timer allows neighbors to become aware of the invalidity of a route prior to purging.

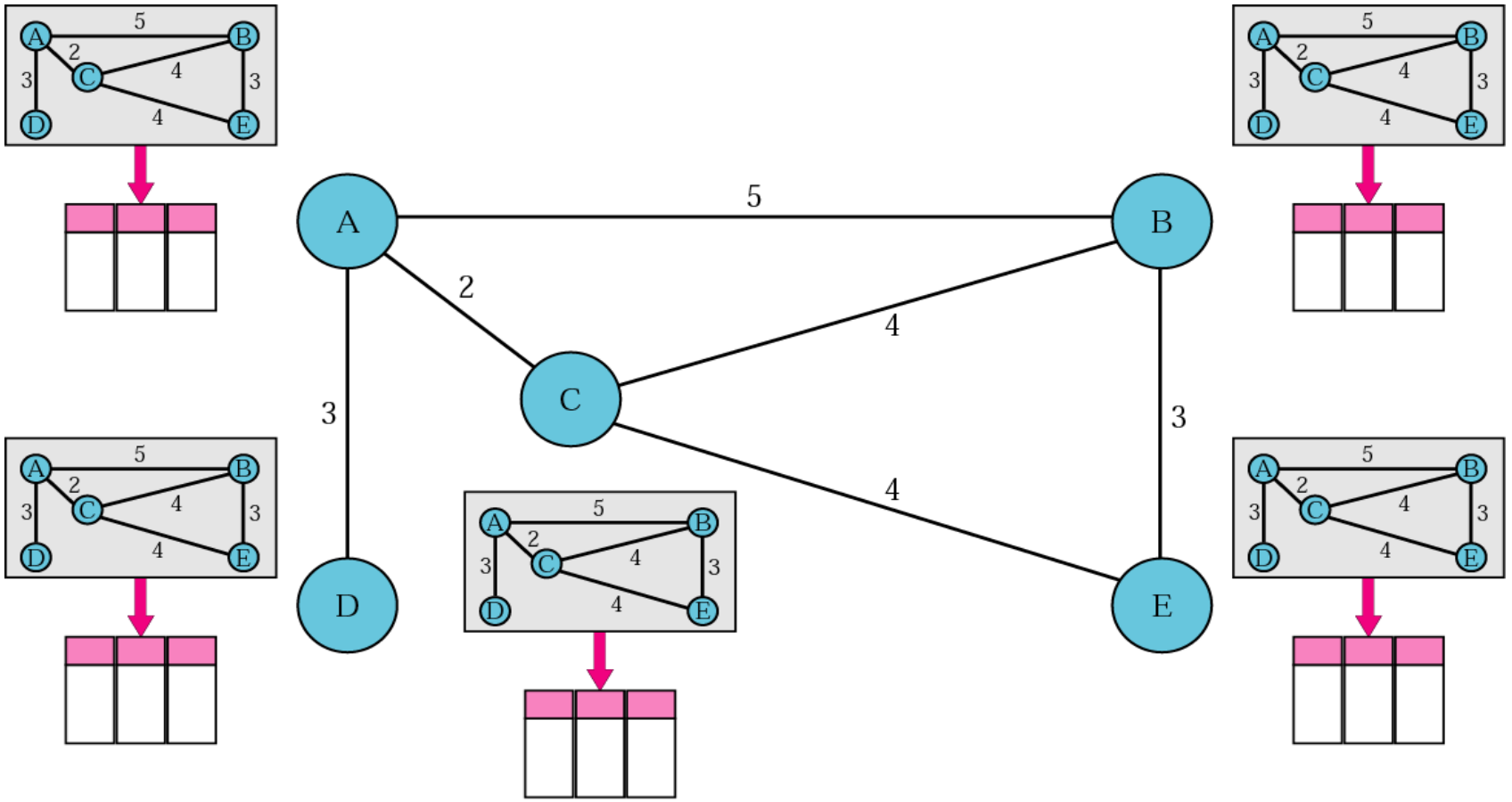
LINK STATE ROUTING

In link state routing, if each node in the domain has the entire topology of the domain,

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

Figure 14.15 *Concept of link state routing*

The figure shows a simple domain with five nodes. Each node uses the same topology to create a routing table, but the routing table for each node is unique because the calculations are based on different interpretations of the topology. This is analogous to a city map. While each person may have the same map, each needs to take a different route to reach her specific destination.



Every router has knowledge about the network, but from its own perspective.

Figure 14.16 *Link state knowledge*

Each router knows (maintains) its states of its links.

Each router floods this info (via a Link State Packet) to other routers periodically (when there is a change in the topology, or every 60 to 120 minutes).

Each router takes in this data and, using Dijkstra's algorithm, creates the shortest path tree and corresponding routing table.

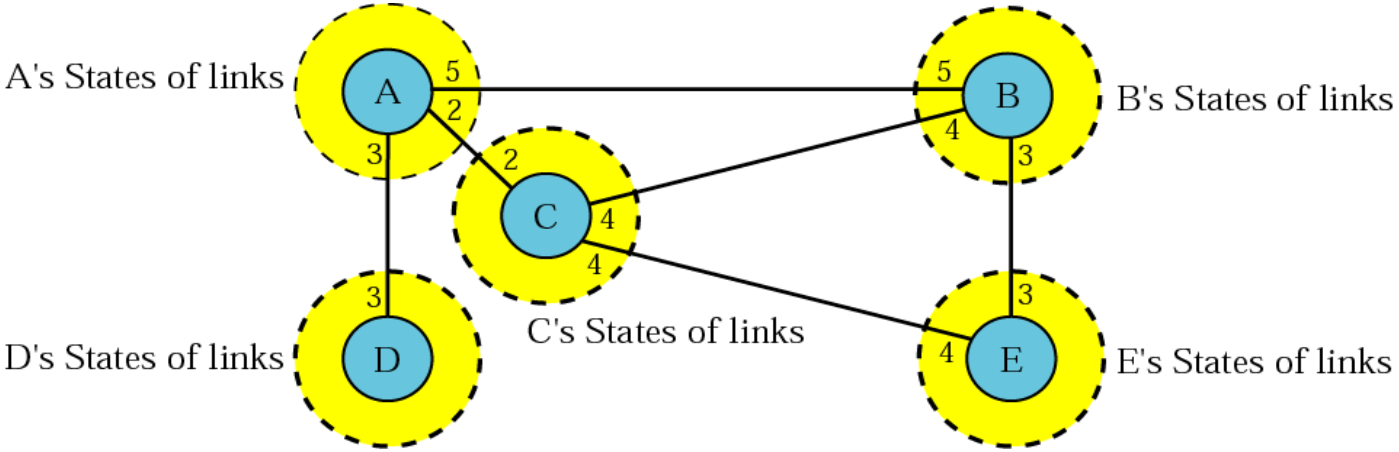


Figure 14.17 *Dijkstra algorithm*

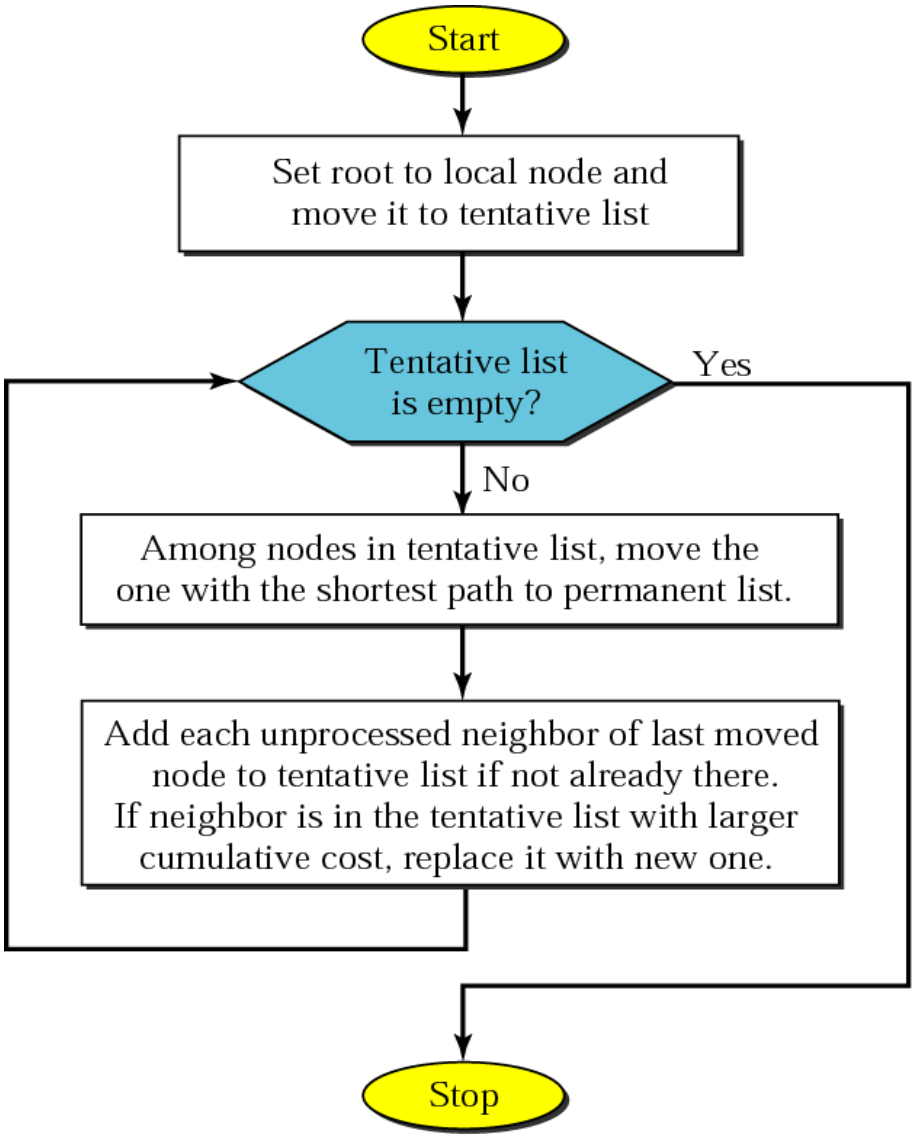
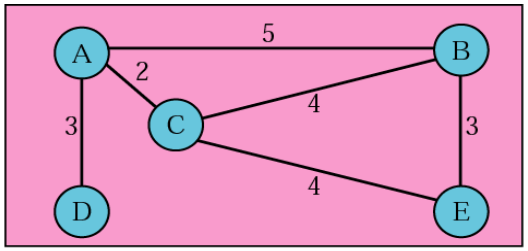
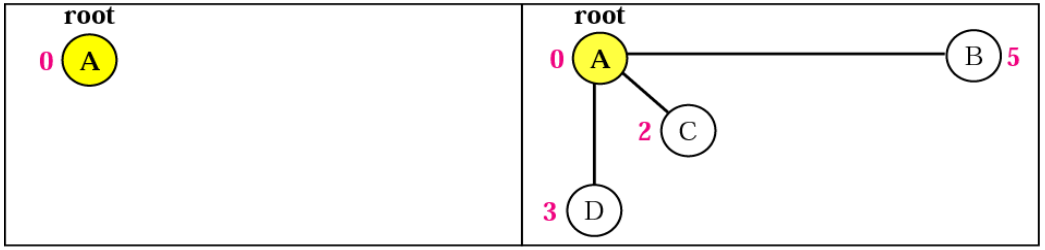


Figure 14.18 *Example of formation of shortest path tree*

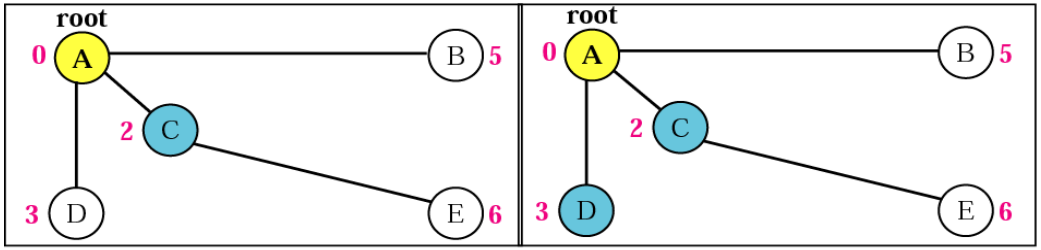


Topology



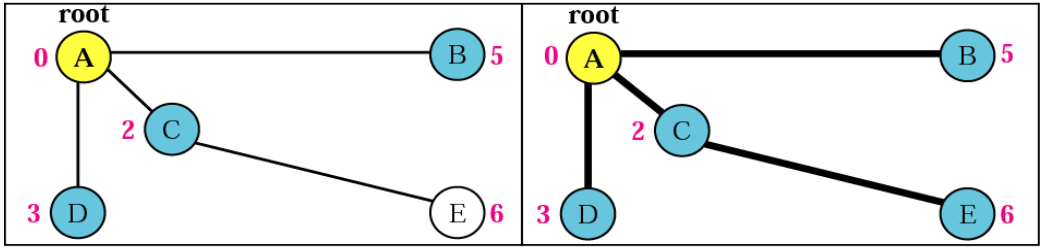
1. Set root to A and move A to tentative list

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



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

4. Move D to permanent list.



5. Move B to permanent list

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

Table 14.1 *Routing table for node A*

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

Now let's try using the Dijkstra's algorithm introduced in TDC 361.

Comparison of Protocols

Link State

- Knowledge of every router's links (entire graph)
- Every router has O(# edges)
- Trust a peer's info, do routing computation yourself
- Use Dijkstra's algorithm
- Send updates on any link-state changes
- Ex: OSPF, IS-IS
- Adv: Fast to react to changes

Distance Vector

- Knowledge of neighbors' distance to destinations
- Every router has O (#neighbors * #nodes)
- Trust a peer's routing computation
- Use Bellman-Ford algorithm
- Send updates periodically or routing decision change
- Ex: RIP, IGRP
- Adv: Less info & lower computational overhead

OSPF

The Open Shortest Path First (OSPF) protocol is an intradomain routing protocol based on link state routing.

Its domain is also an autonomous system.

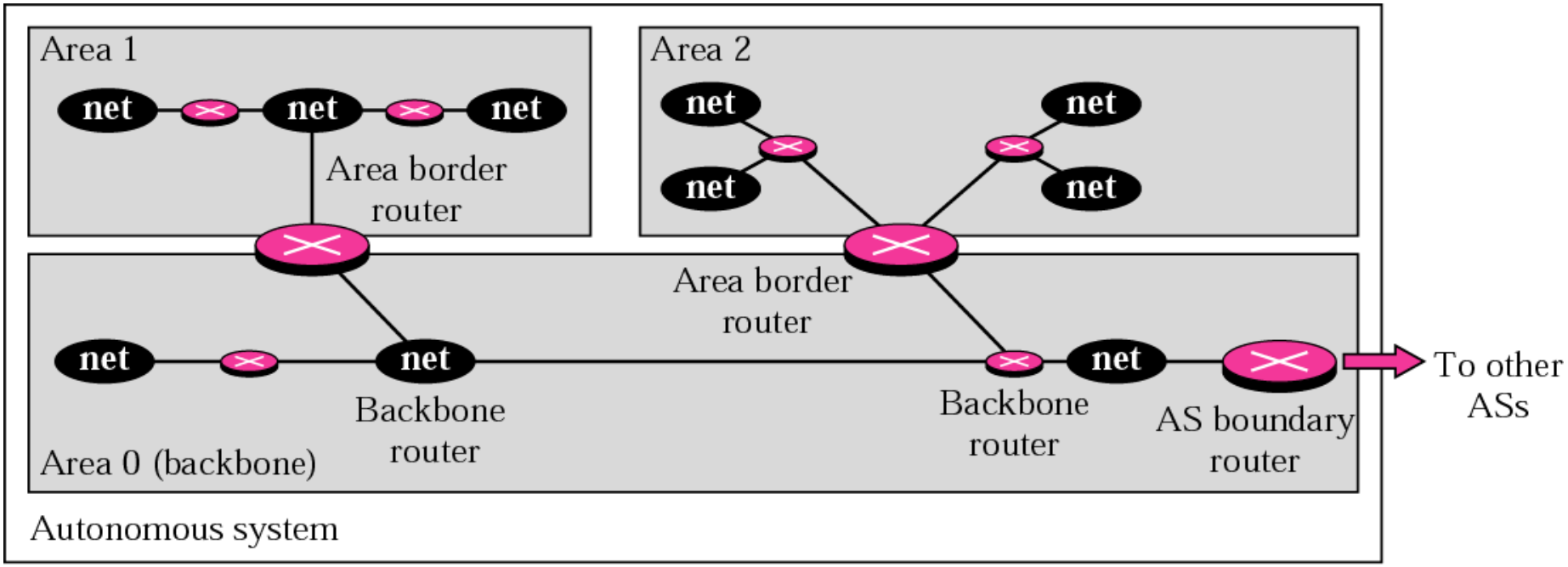
Also known as Dijkistra's Algorithm

Figure 14.19 *Areas in an autonomous system*

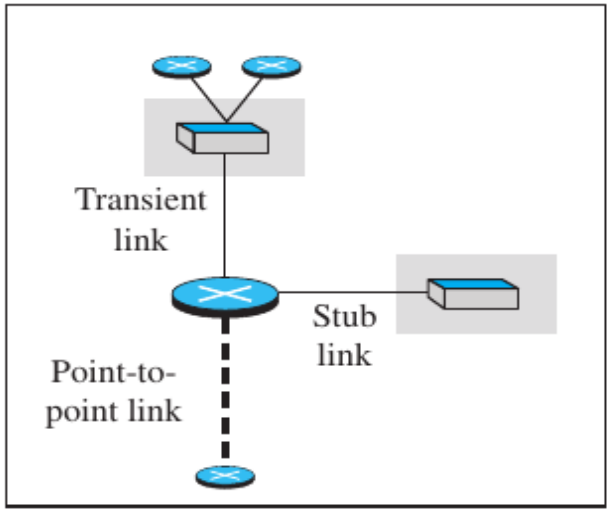
OSPF divides an autonomous system into areas. All networks inside an area must be connected.

area border router; backbones; backbone routers; boundary routers

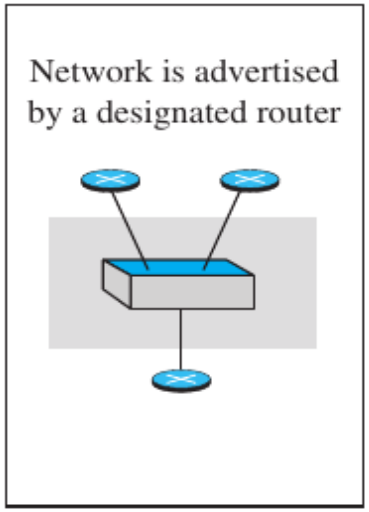
The cost associated with a route is called the metric. Metric could be min delay, max thruput, etc.



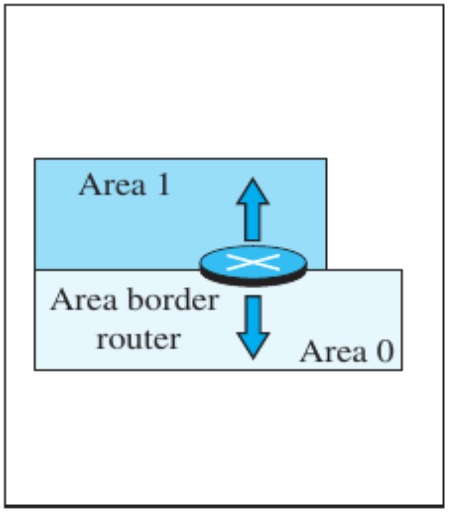
LSPs (Link State Packets)



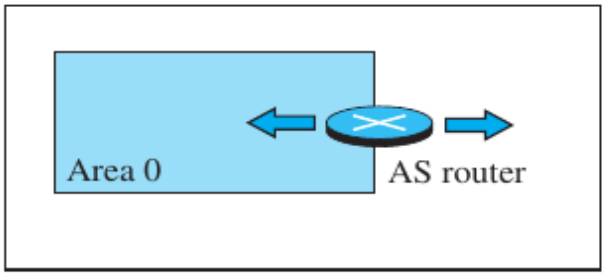
a. Router link



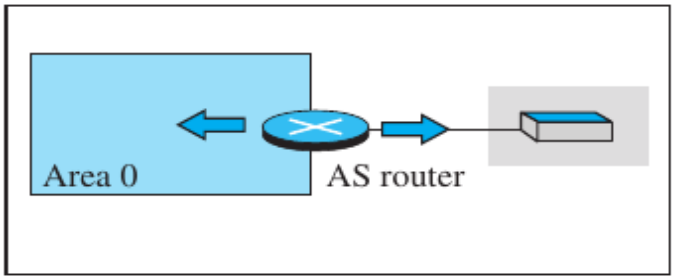
b. Network link



c. Summary link to network



d. Summary link to AS

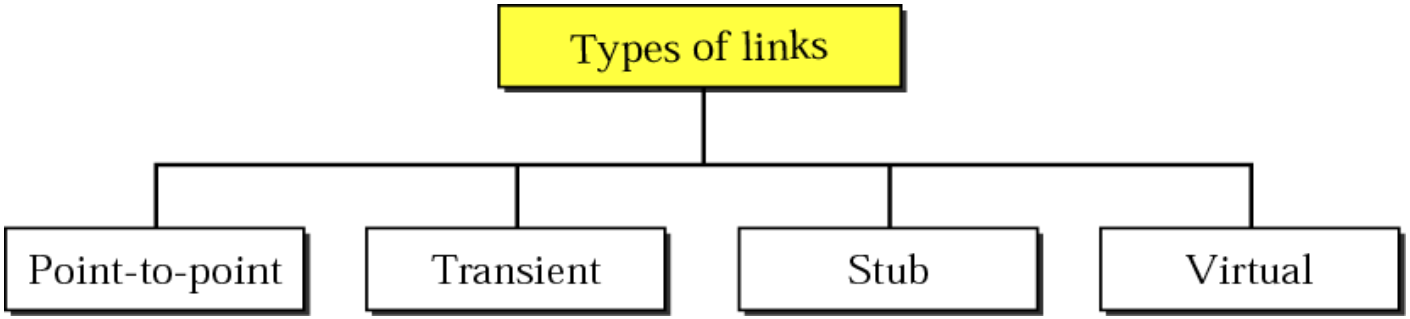


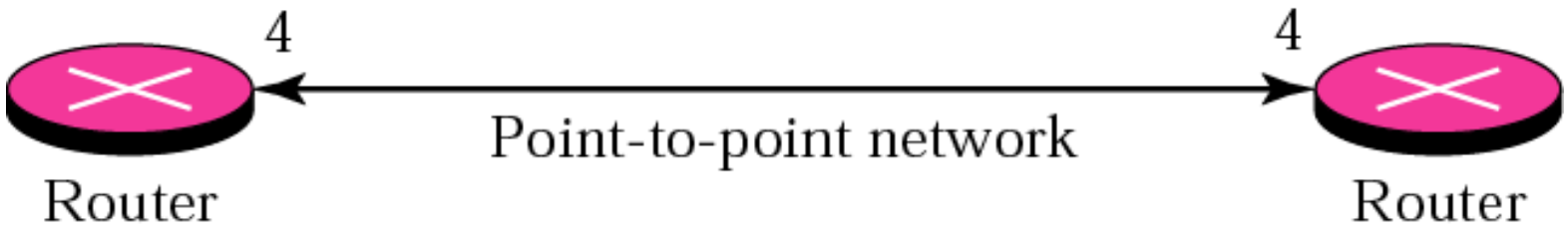
e. External link

LSPs

1. Router Link: A router link advertises the existence of a router as a node. In addition to giving the address of the announcing router, this type of advertisement can define one or more types of links that connect the advertising router to other entities.

Figure 14.21 *Point-to-point link*





No hosts in between; T-1 connection common

Types of link

- A **transient link** announces a link to a transient network, a network that is connected to the rest of the networks by one or more routers. This type of advertisement should define the address of the transient network and the cost of the link.
- A **stub link** advertises a link to a stub network, a network that is not a through network. Again, the advertisement should define the address of the network and the cost.
- A **point-to-point link** should define the address of the router at the end of the point-to-point line and the cost to get there.

2. Network Link: A network link advertises the network as a node. However, since a network cannot do announcements itself (it is a passive entity), one of the routers is assigned as the designated router and does the advertising.

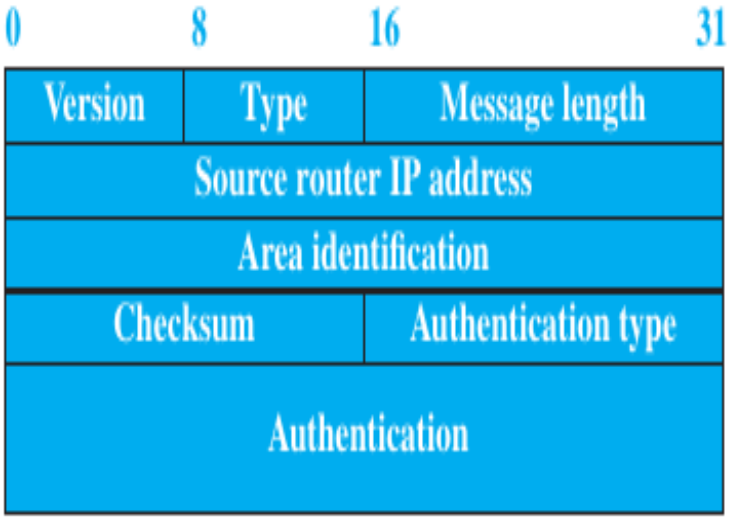
3. Summary link to network: This is done by an area border router; it advertises the summary of links collected by the backbone to an area or the summary of links collected by the area to the backbone.

LSPs

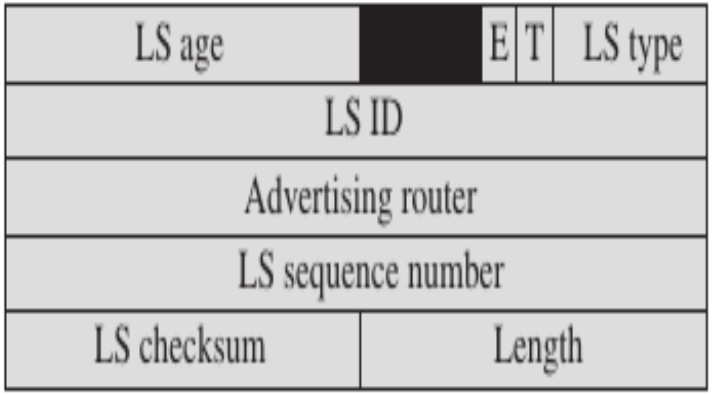
4. Summary link to AS: This is done by an AS router that advertises the summary links from other ASs to the backbone area of the current AS, information which later can be disseminated to the areas so that they will know about the networks in other ASs.

5. External link. This is also done by an AS router to announce the existence of a single network outside the AS to the backbone area to be disseminated into the areas.

OSPF message format



OSPF common header



Link-state general header

14.6 PATH VECTOR ROUTING

Path vector routing is similar to distance vector routing. There is at least one node, called the speaker node, in each AS that creates a routing table and advertises it to other speaker nodes in the neighboring ASs..

The topics discussed in this section include:

Initialization
Sharing
Updating

Figure 14.48 *Initial routing tables in path vector routing*

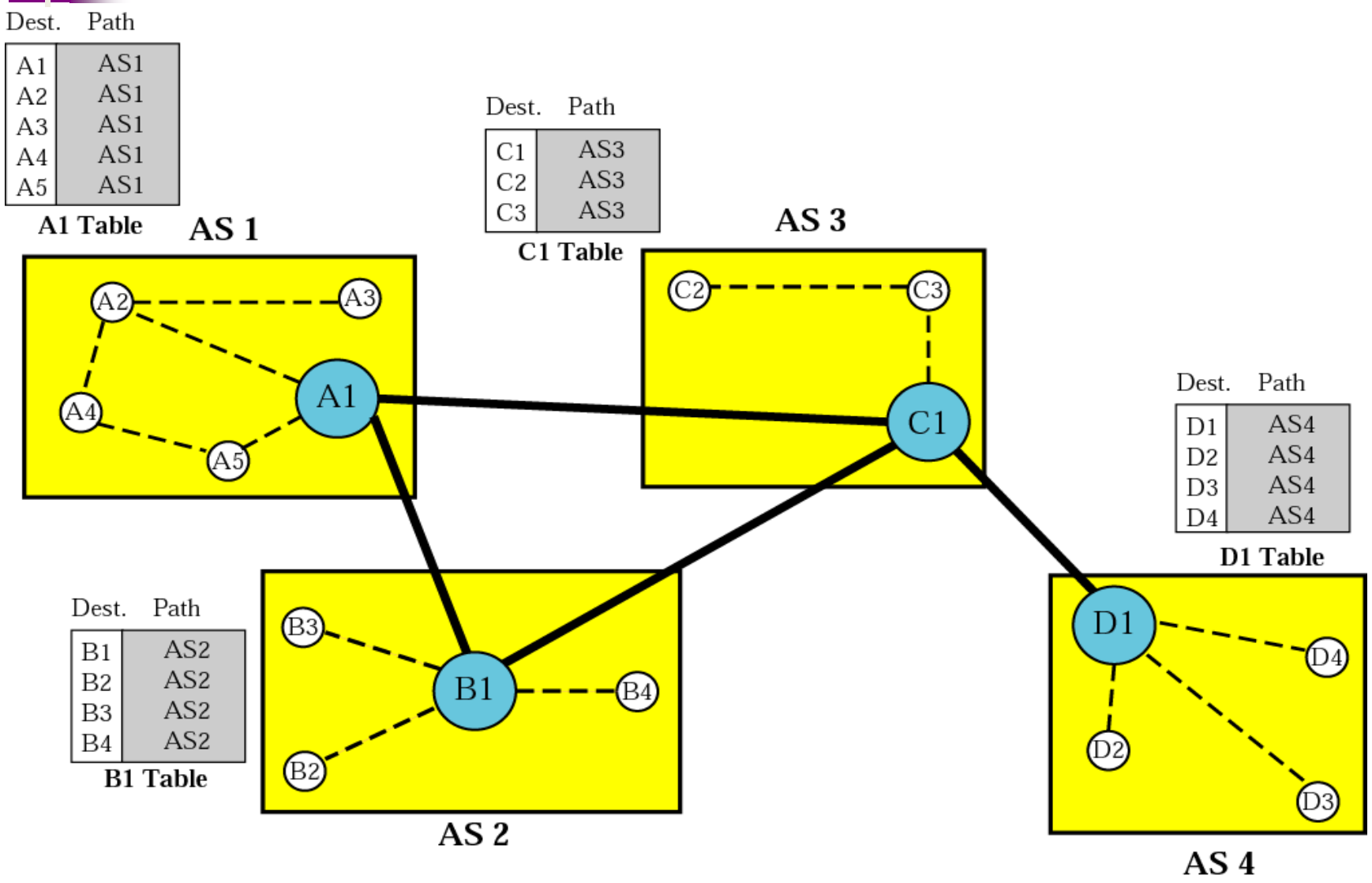


Figure 14.49 *Stabilized tables for four autonomous systems*

Dest.	Path
A1 ... A5	AS1
B1 ... B4	AS1-AS2 ... AS1-AS2
C1 ... C3	AS1-AS3 ... AS1-AS3
D1 ... D4	AS1-AS2-AS4 ... AS1-AS2-AS4

A1 Table

Dest.	Path
A1 ... A5	AS2-AS1
B1 ... B4	AS2 ... AS2
C1 ... C3	AS2-AS3 ... AS2-AS3
D1 ... D4	AS2-AS3-AS4 ... AS2-AS3-AS4

B1 Table

Dest.	Path
A1 ... A5	AS3-AS1
B1 ... B4	AS3-AS2 ... AS3-AS2
C1 ... C3	AS3 ... AS3
D1 ... D4	AS3-AS4 ... AS3-AS4

C1 Table

Dest.	Path
A1 ... A5	AS4-AS3-AS1
B1 ... B4	AS4-AS3-AS2 ... AS4-AS3-AS2
C1 ... C3	AS4-AS3 ... AS4-AS3
D1 ... D4	AS4 ... AS4

D1 Table

14.7 BGP

Border Gateway Protocol (BGP) is an interdomain routing protocol using path vector routing. It first appeared in 1989 and has gone through four versions.

BGP interconnects three different types of AS:

1. Stub AS, e.g. a corporate network

2. Multihomed AS, e.g. a large corporate network with connections to multiple ASs, but does not allow traffic to pass thru (transient)

3. Transit AS - one that allows transient traffic, such as an Internet backbone

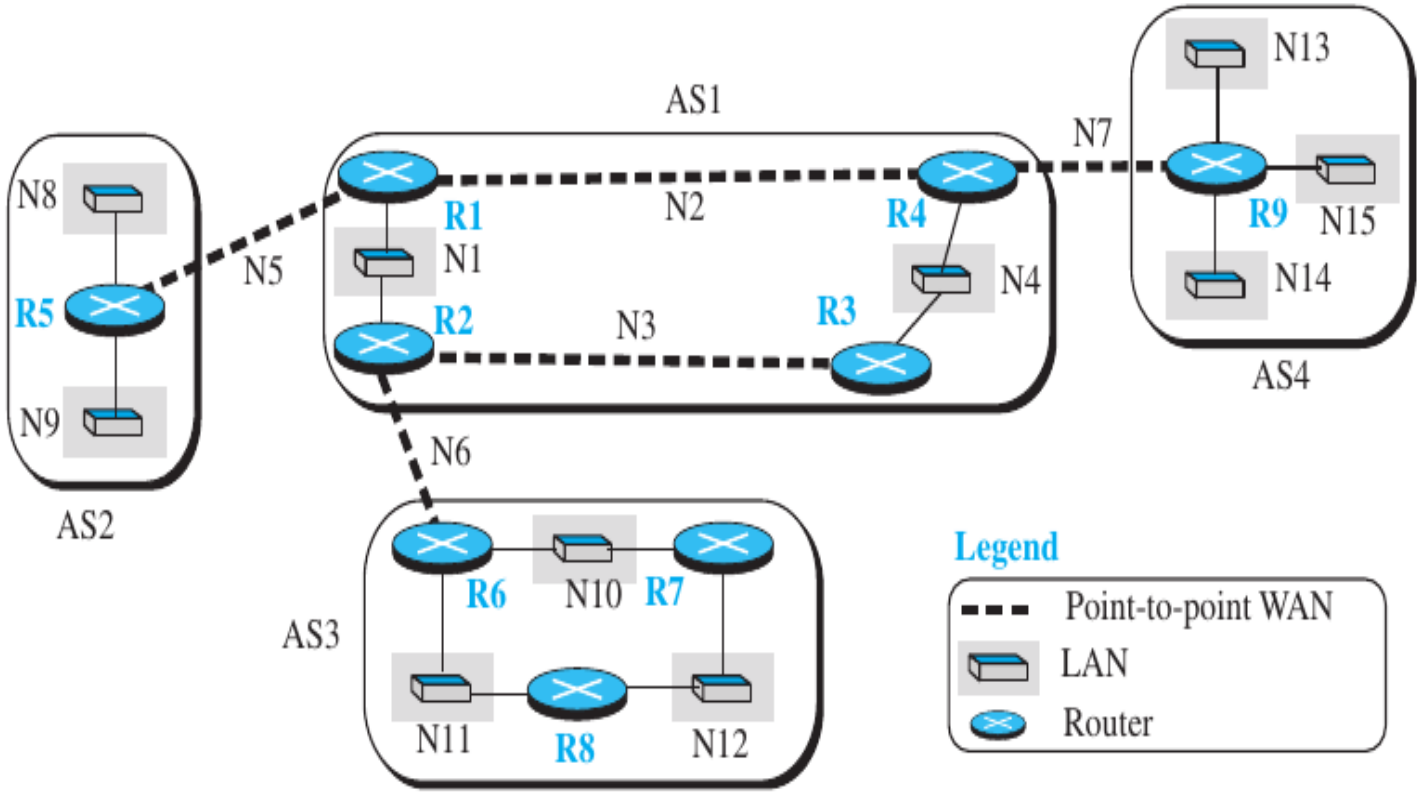

AS1
AS2
AS3
AS4
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
N11
N12
N13
N14
N15
R1
R2
R3
R4
R5
R6
R7
R8
R9
Legend
Point-to-point WAN
LAN
Router

Figure 14.50 *Internal and external BGP sessions*

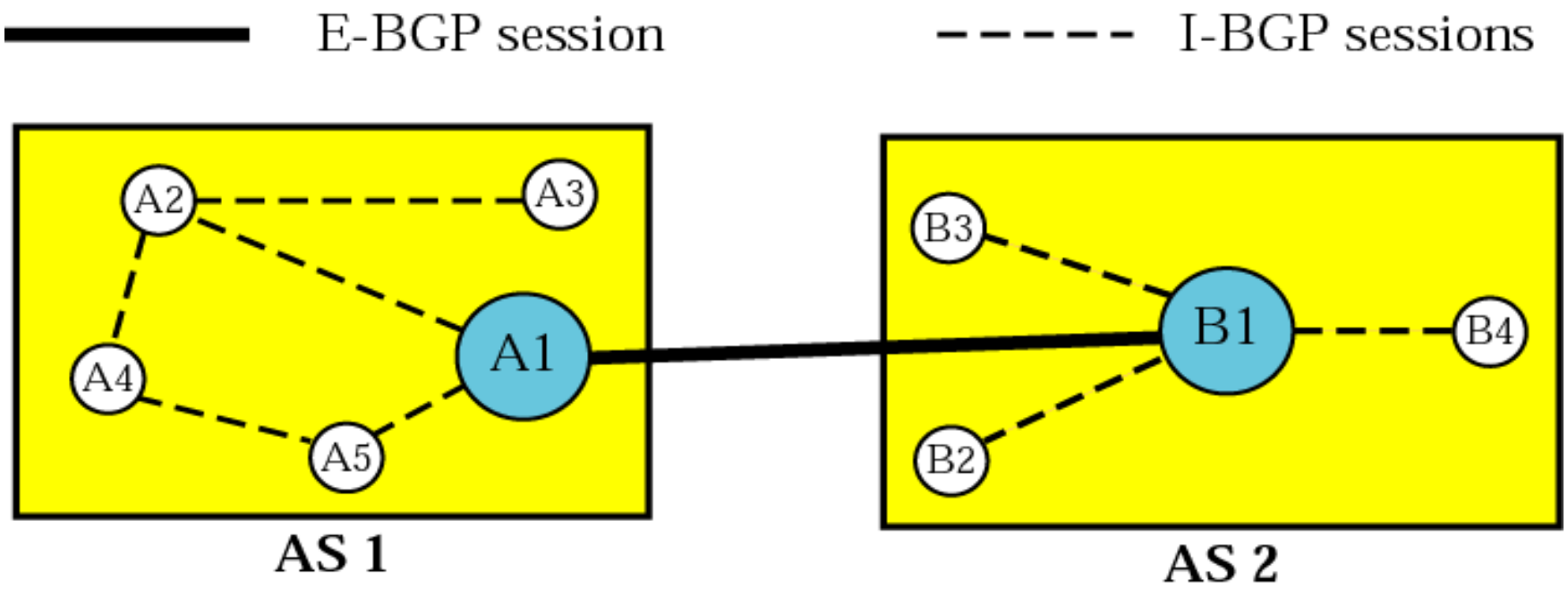
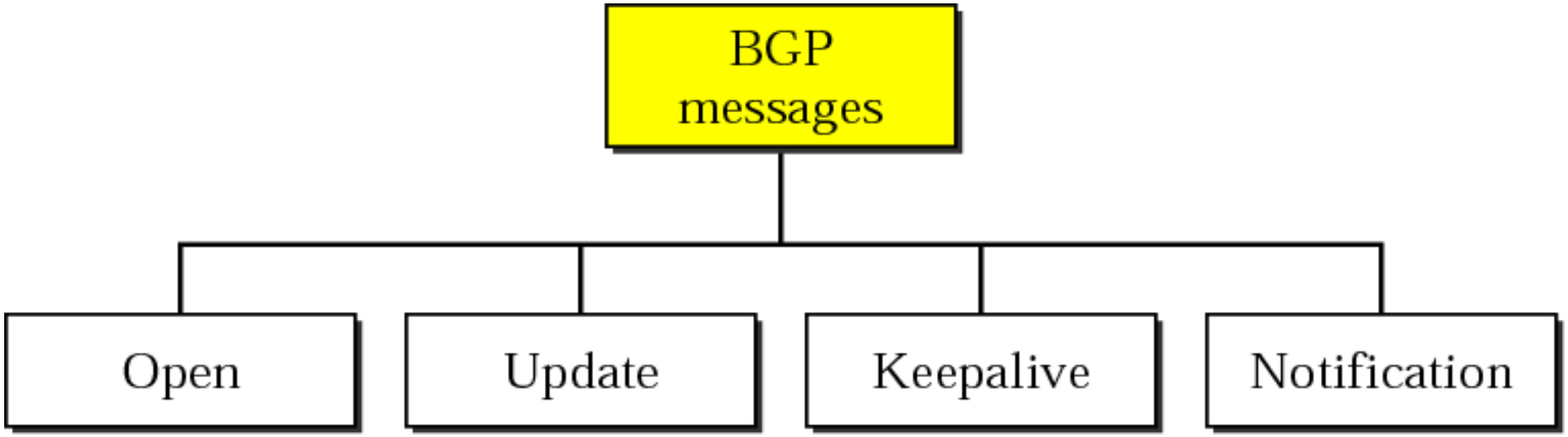


Figure 14.51 *Types of BGP messages*



BGP Messages

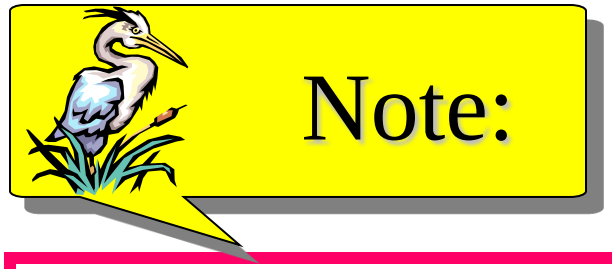
Open Message. To create a neighborhood relationship, a router running BGP opens a TCP connection with a neighbor and sends an open message.

Update Message. The update message is the heart of the BGP protocol. It is used by a router to withdraw destinations that have been advertised previously, to announce a route to a new destination, or both. Note that BGP can withdraw several destinations that were advertised before, but it can only advertise one new destination (or multiple destinations with the same path attributes) in a single update message.

K**eepalive Message.** The BGP peers that are running exchange keepalive messages regularly (before their hold time expires) to tell each other that they are alive.

Notification. A notification message is sent by a router whenever an error condition is detected or a router wants to close the session.

BGP supports classless addressing and CIDR.



BGP uses the services of TCP on port 179.

RIP uses the services of UDP on port 520.