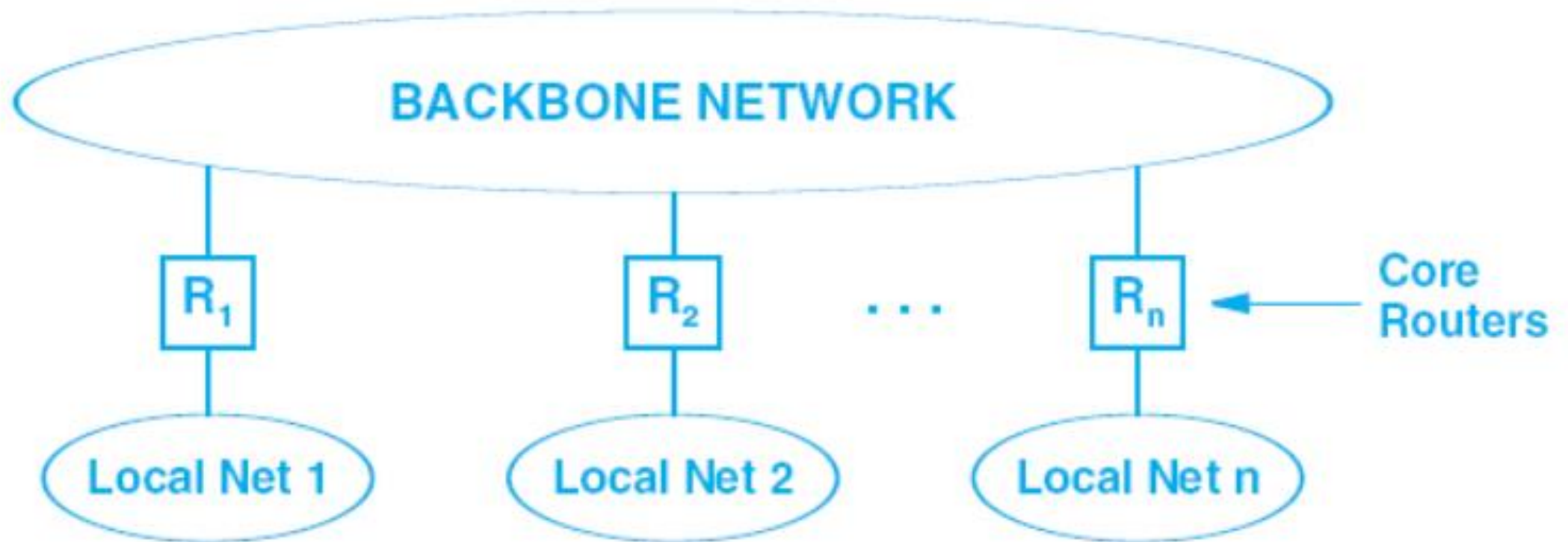
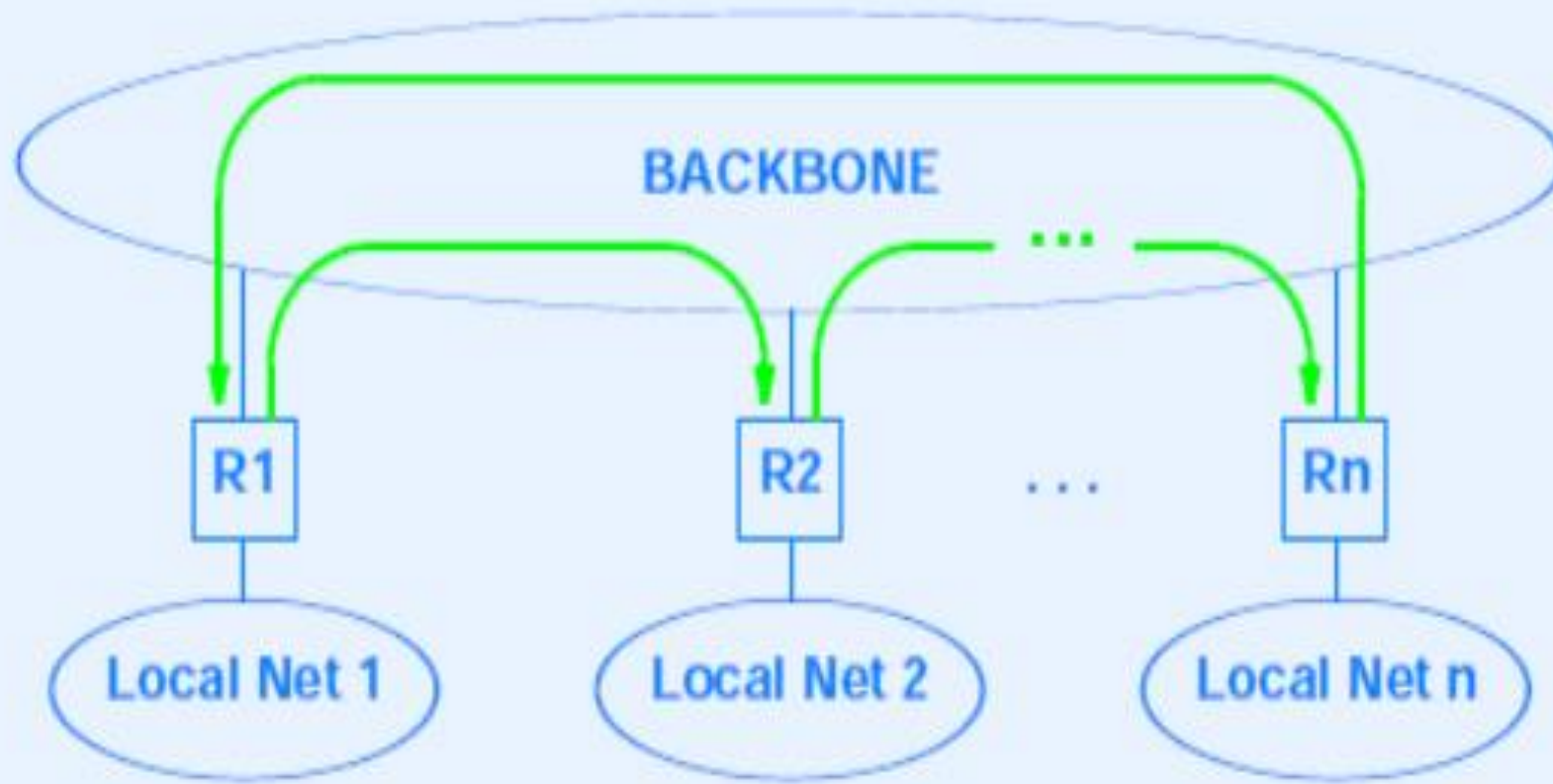






# **MODULE 3 - Routing**

# ORIGINAL INTERNET ARCHITECTURE & CORES

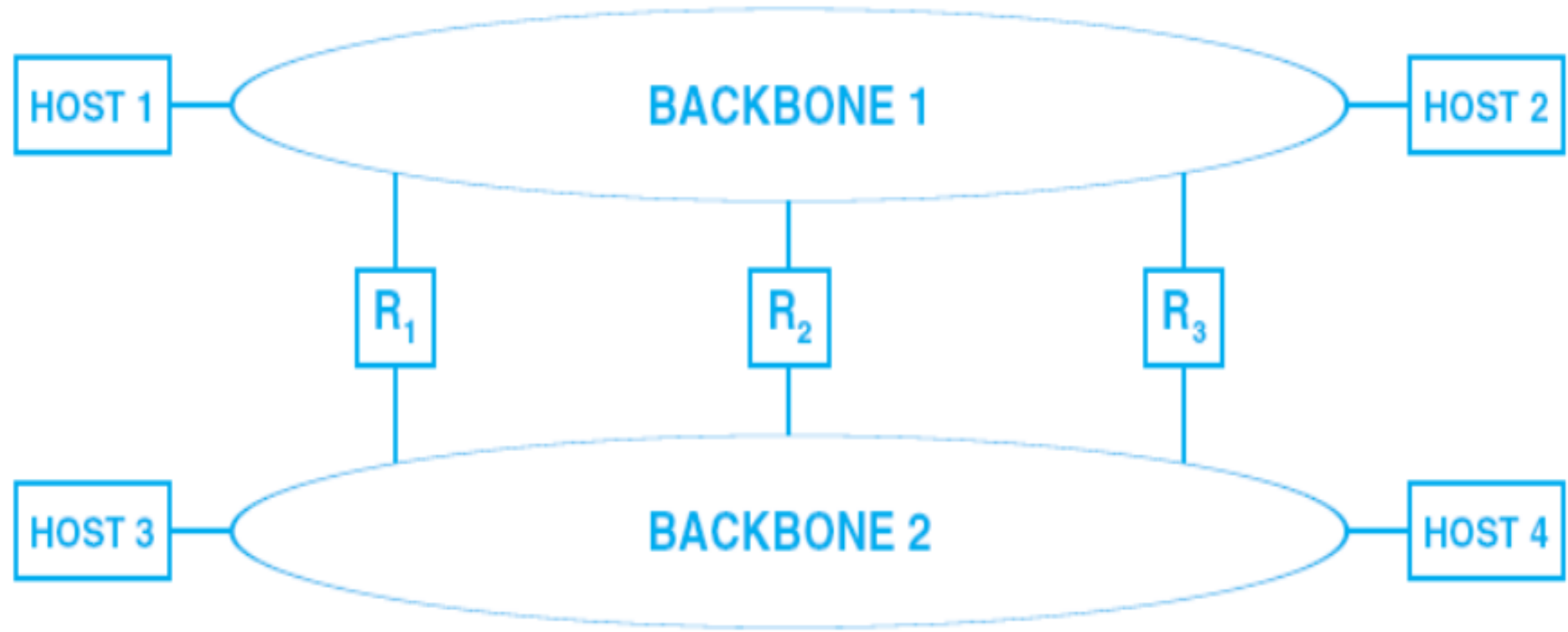




Datagram sent to nonexistent destination loops until TTL expires

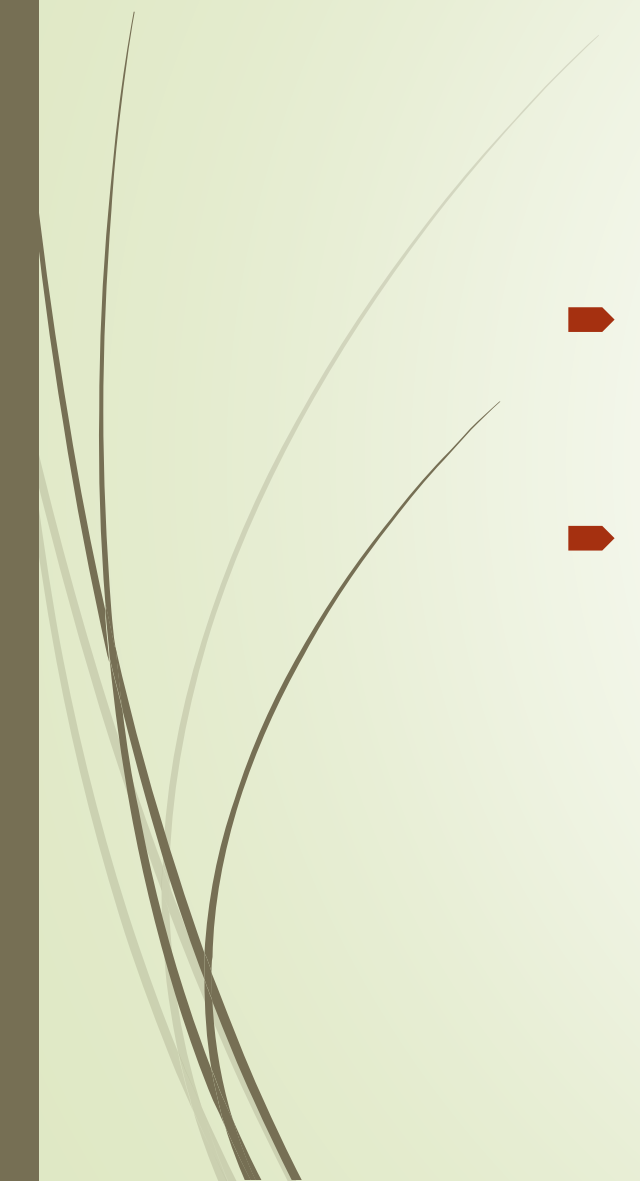
- 
- 
- At the source site, the local router checks to see if it has an explicit route to the destination, and if not, sends the datagram along the path specified by its **default route**.
  - **All datagrams for which the router has no explicit route follow the same default path** regardless of their ultimate destination.
  - The next router along the path diverts datagrams for which it has an explicit route, and sends the rest along its default route.

# PEER BACKBONES





# ALGORITHMS

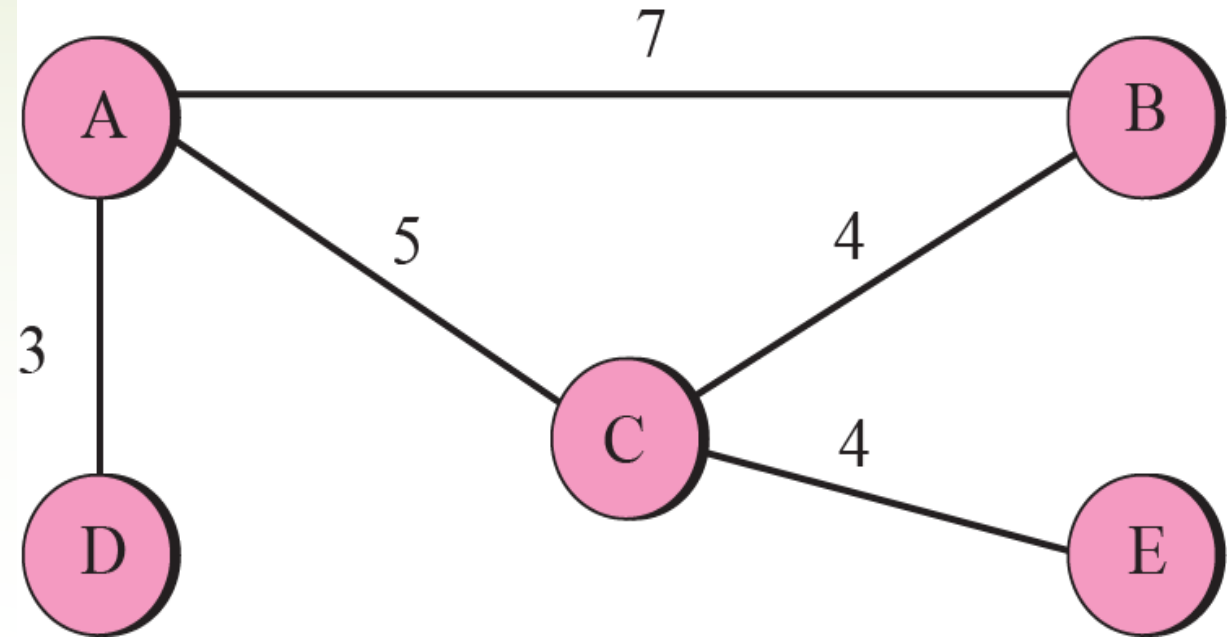
- Bellman-Ford Algorithm
  - Dijkstra's Shortest Path Algorithm
- 

# Distance Vector Routing

- Network is considered as **Graph** - router represented by node and network represented by links connecting two nodes
- Graph theory : **Bellman-Ford Algorithm** – to find the shortest path between nodes given the distance between the nodes.
- This can be modified to be used for **updating routing tables** in a distance vector routing



# Bellman- Ford Algorithm



- Find the least cost route between any two nodes
- It is the route with minimum distance
- Each node maintains a table of minimum distance to every node – shortest distance table – distance vector – (routing table)
- The table at each node guides the packets to the desired node by showing the next stop in the route.





# Example

Class – chalk and board



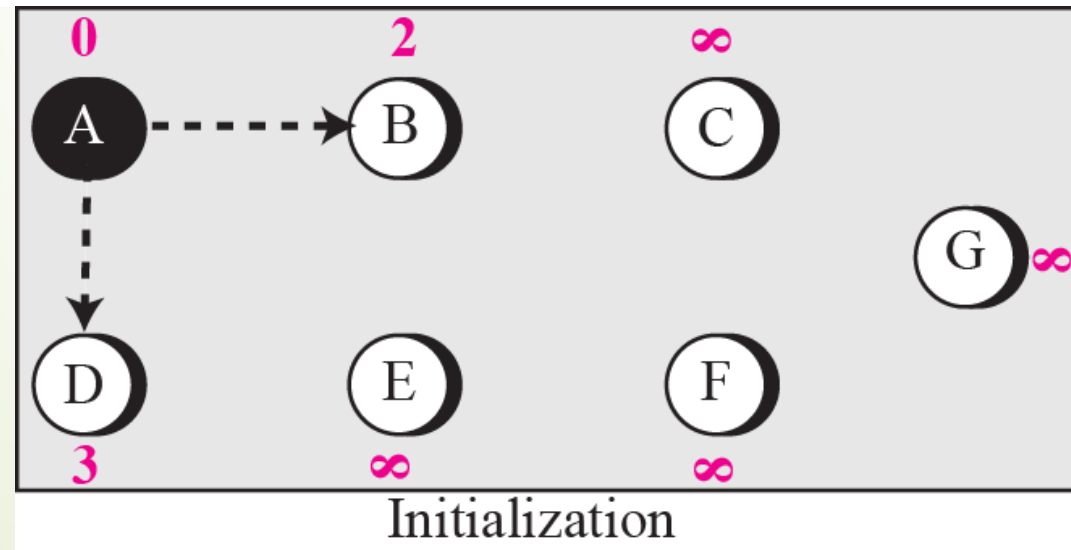
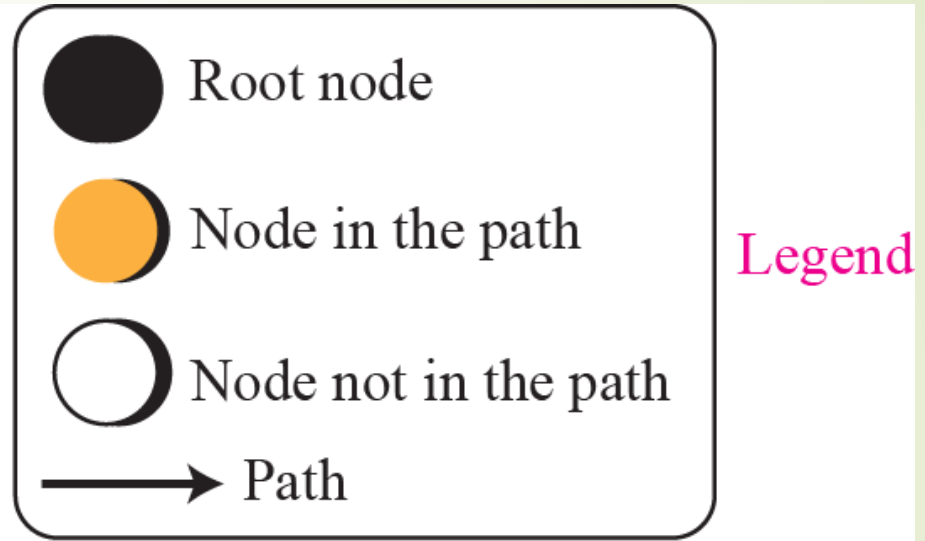
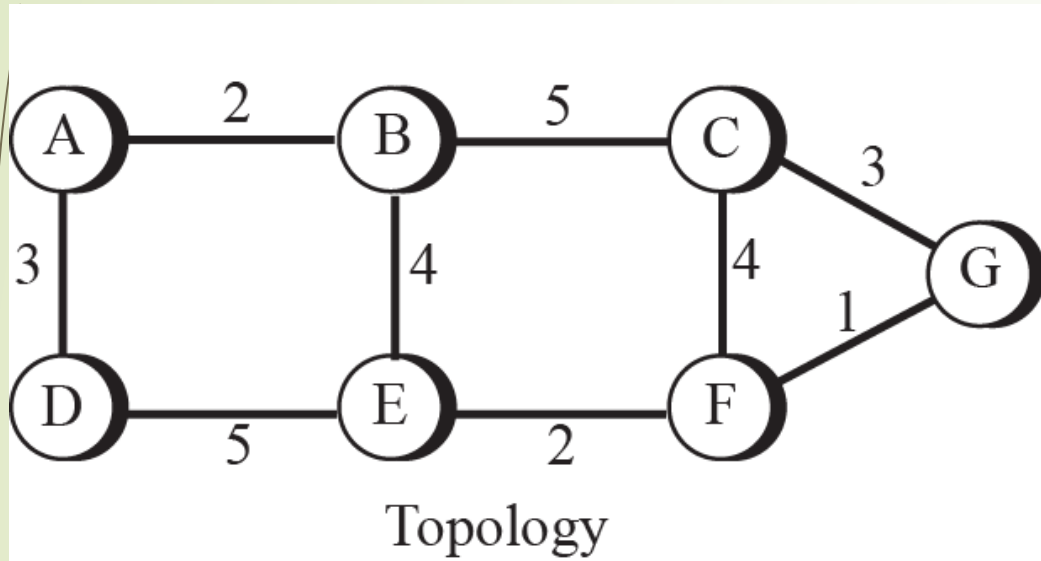
Link State Routing :

**Dijkstras algorithm** is used to create a shortest path tree from a given graph

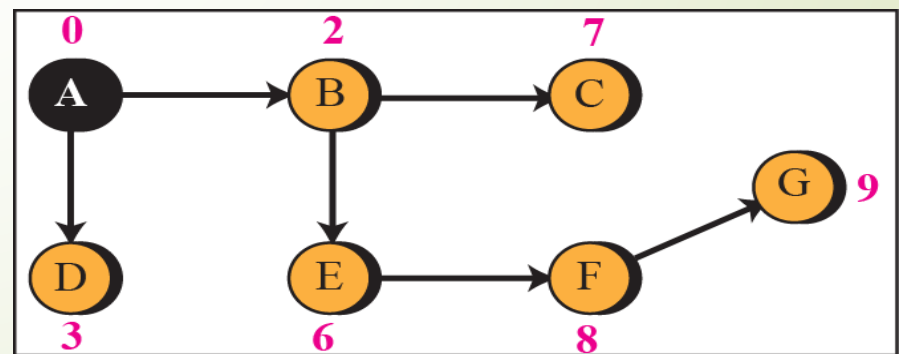
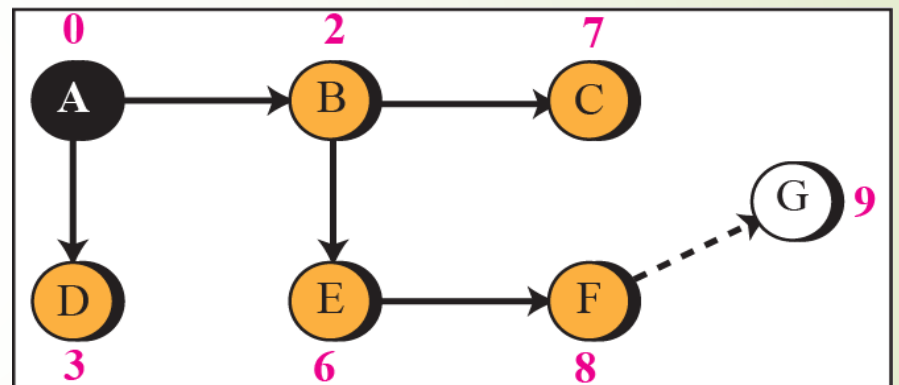
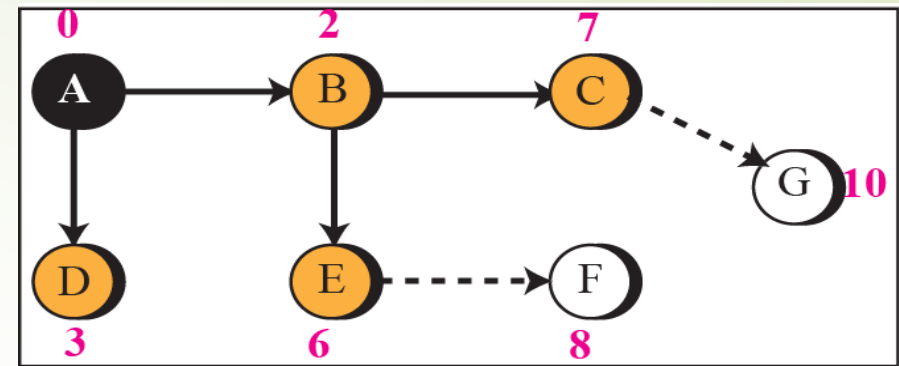
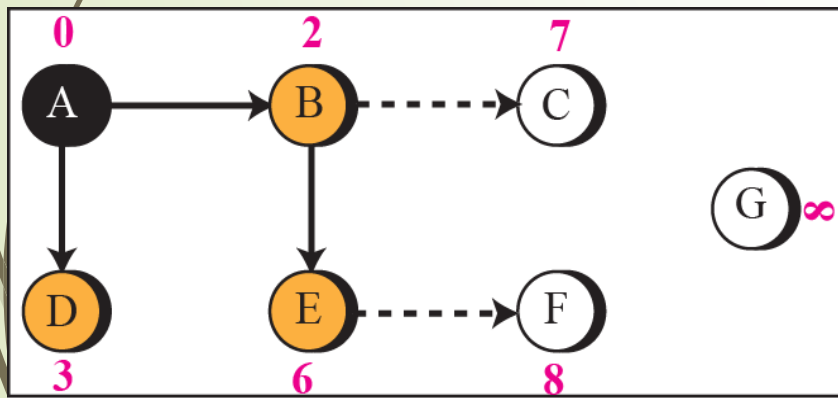
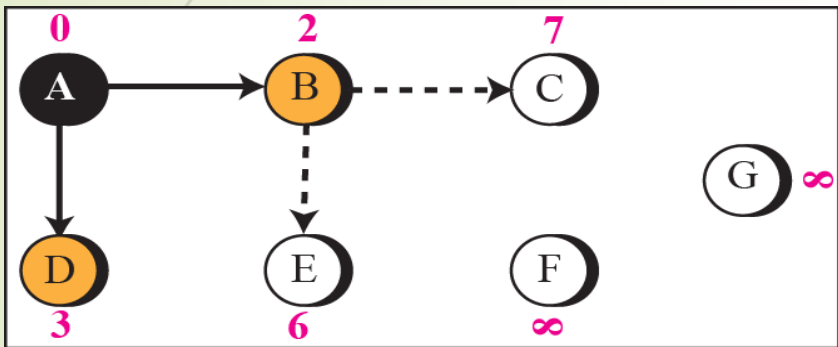
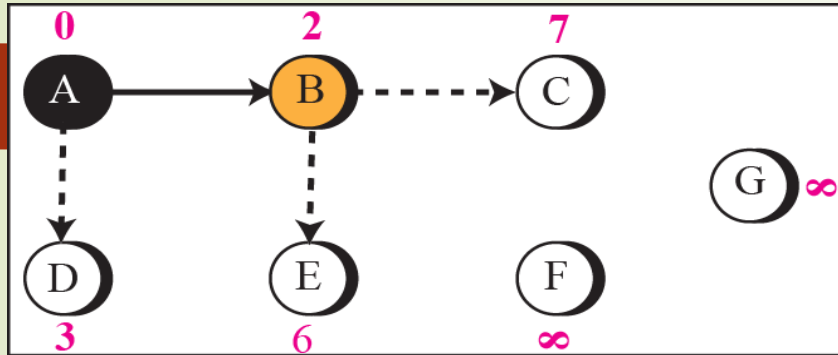
1. **Initialization:** Select the node as the root of the tree and add it to the ***path***. Set the shortest distance for all the root's neighbors to the cost between the root and those neighbors. Others as infinity. Set the shortest distance of the root to zero
2. **Iteration:** Repeat the following two steps until all nodes are added to the ***path***:
  - a) **Adding next node to the path:** Search the nodes not in the ***path***. Select the one with minimum shortest distance and add it to the ***path***.
  - a) **Updating:** Update the shortest distance for all remaining nodes using the shortest distance of the node just moved to the path in step 2.

$$D_j = \text{minimum } (D_j, D_i + c_{ij}) \text{ for all remaining nodes}$$

## Example : *Forming shortest path Tree for router A in a graph*

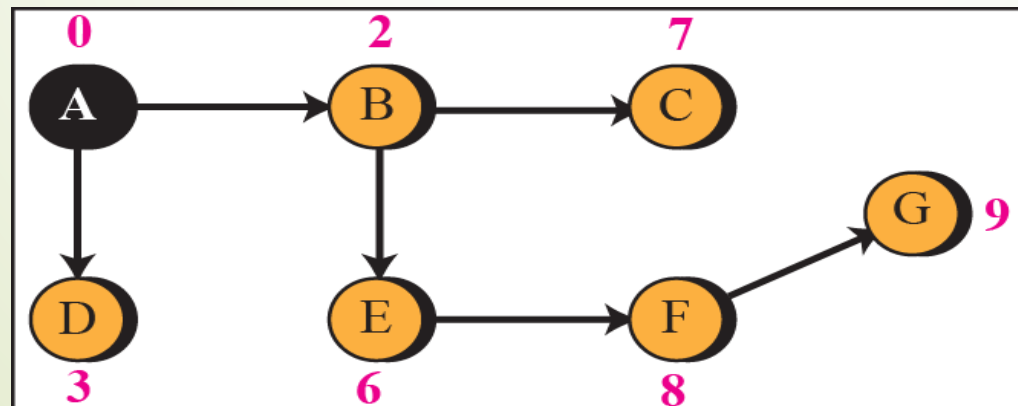


# EXAMPLE : Forming shortest path Tree for router A in a graph



# Calculation of Routing Table from Shortest Path Tree

- Each node uses the shortest path tree found in the previous discussion to construct its routing table.
- The **routing table shows the cost of reaching each node from the root.**
- Table shows the routing table for node A using the shortest path tree found



**Table 11.4** *Routing Table for Node A*

<i>Destination</i>	<i>Cost</i>	<i>Next Router</i>
A	0	—
B	2	—
C	7	B
D	3	—
E	6	B
F	8	B
G	9	B



# Delivery of Packets



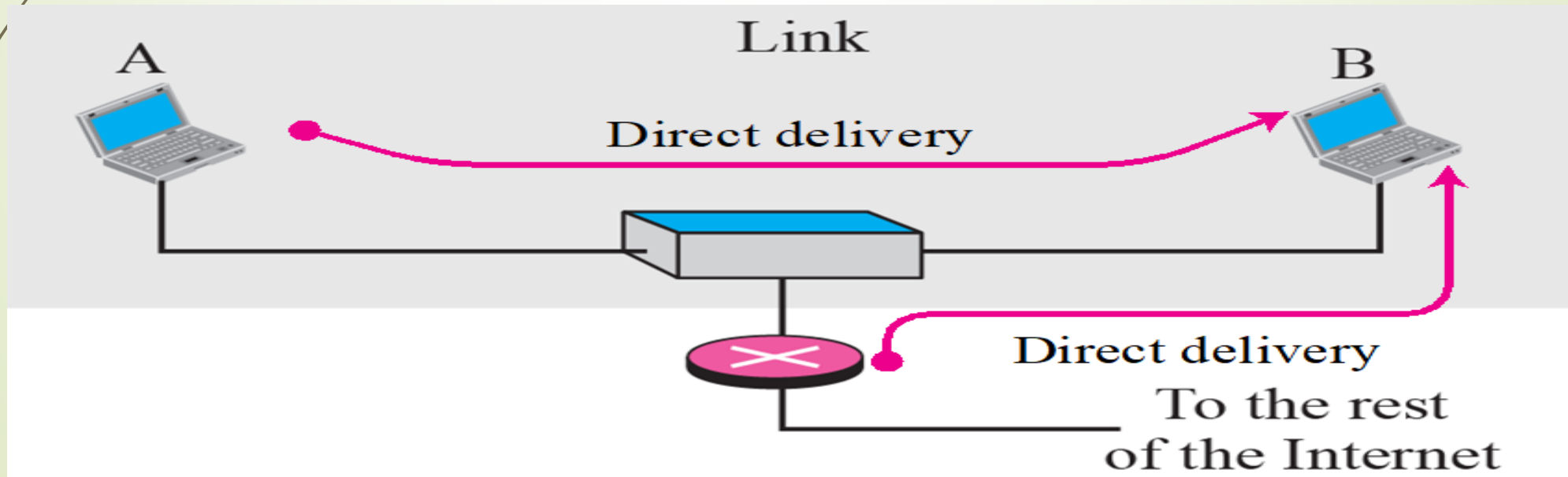
➤ Two types:

**Direct** and **Indirect**



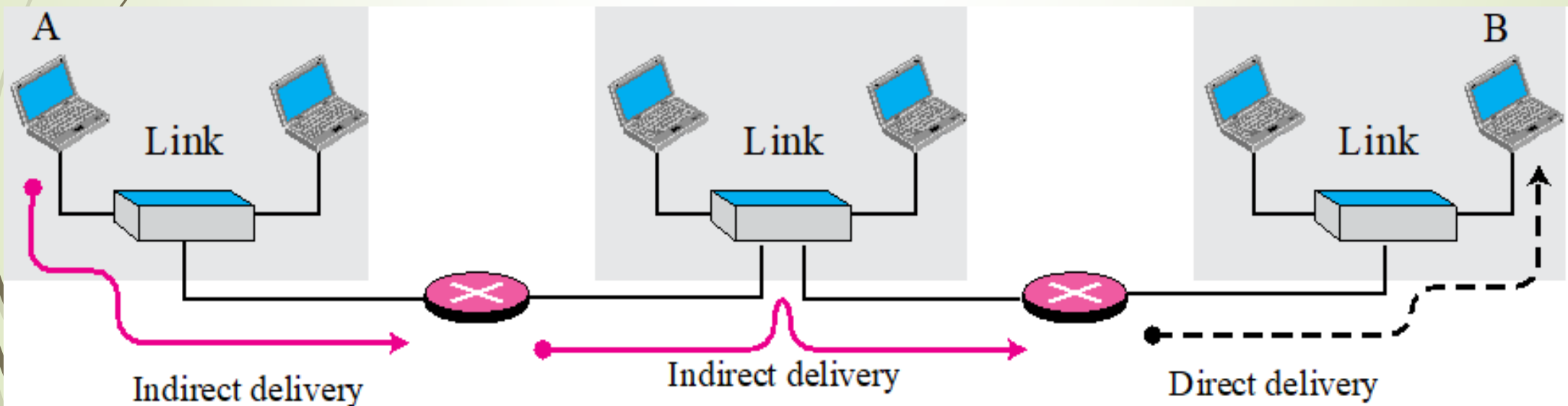
# Direct Delivery

- The final destination of the packet is a host connected to the same network
- **Source and destination located on the same network**
- Sender easily determine the destination
- Sender uses the destination address



# Indirect Delivery

- **Destination is not in the same network**
- Packet goes from router to router until it reaches the destination
- Sender uses the IP address and routing table to find the address of the next router
- **Last delivery is always direct delivery**



# Routing

- Important issue in Internet Layer - **how a router creates its routing table** to help in forwarding a datagram from one network to another network
- Process of establishing communication between two or more networks
- Choose the best path from the routing table.
- This can be done by **Routing Protocols**, that help routers to make their routing table, maintain them, and update them.

## Types of Routing :

- **Static Routing (Basic)**
- **Dynamic Routing (Common)**

## ➤ Static Routing

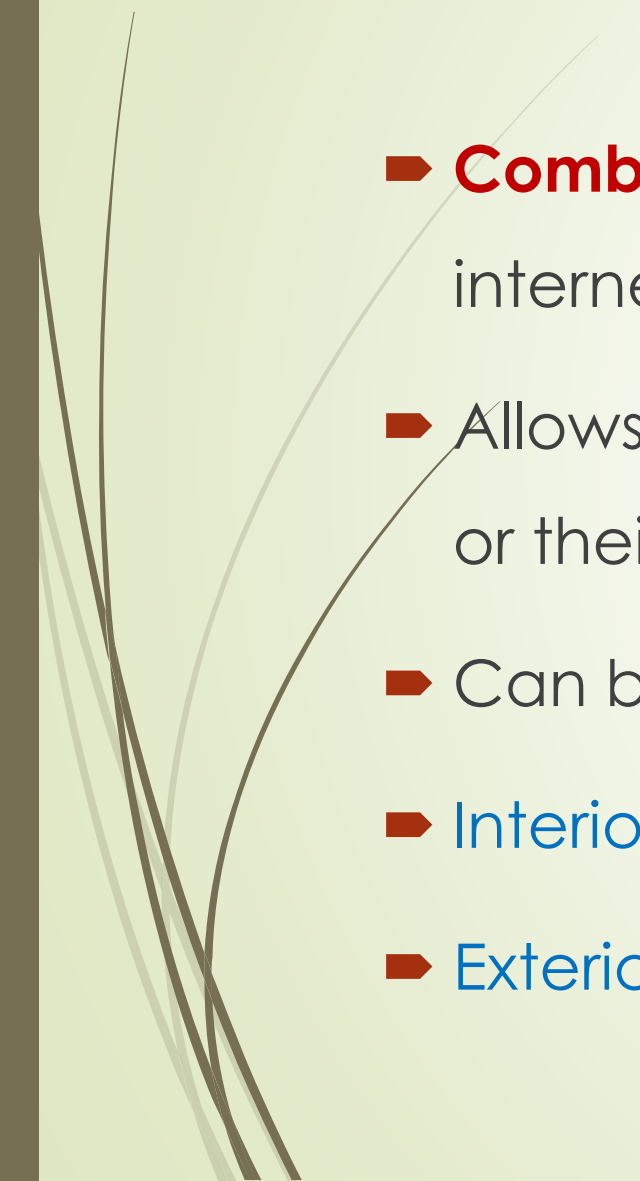
- Manual Routing
- Administrator decide the best route
- **Manual entries – Routing Table**
- For simple static networks or static routing , a manually configured routing table may be adequate.

## ➤ Dynamic Routing (Common)

- Automatic routing
- Automatically learning about the networks
- **Routing Table : updated based on the networks**
- When the network is more complex, need an automated and reactive way to distribute information about the connectivity within the network.
- Best route is decided by the router itself with the help of routing protocols

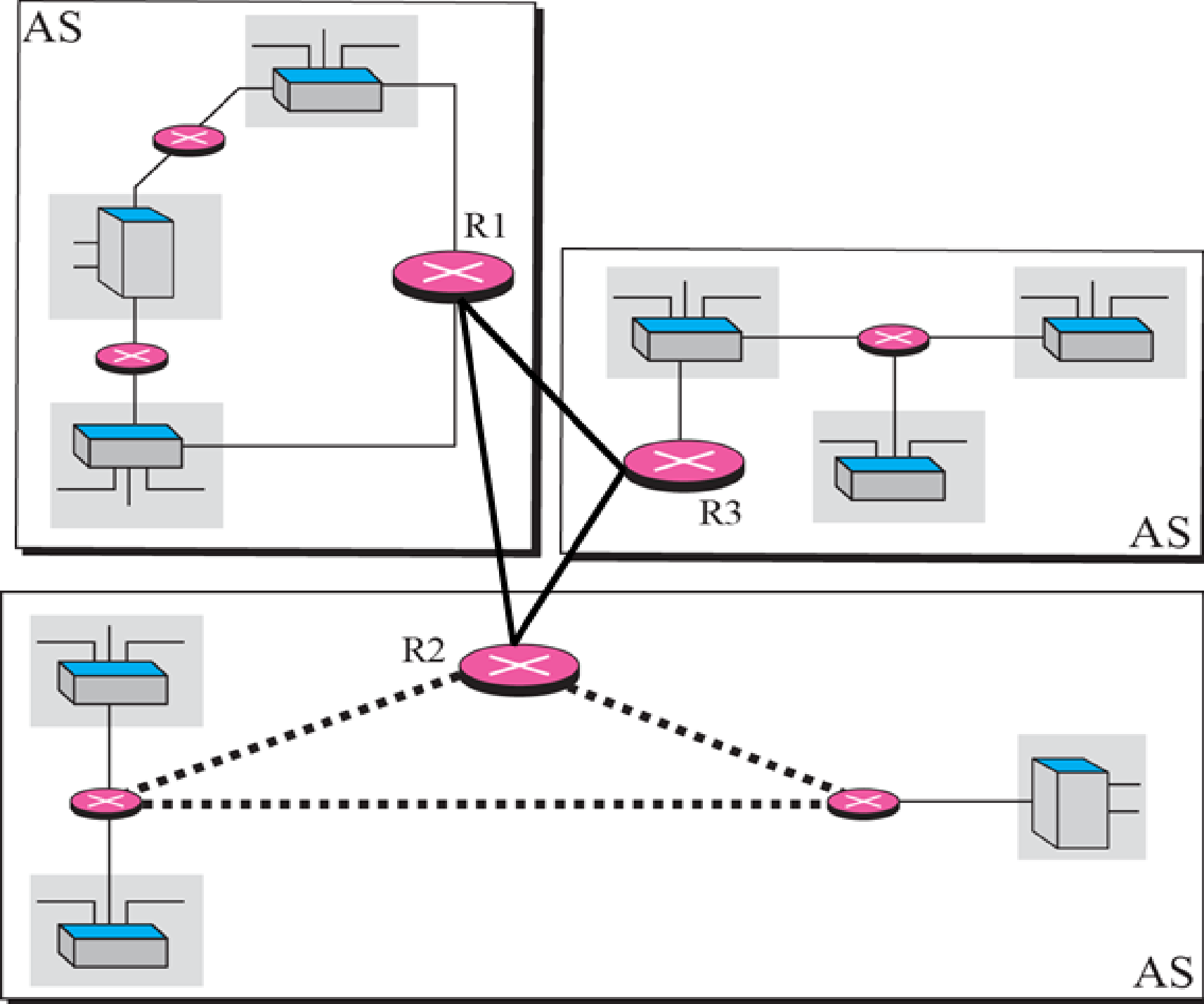


# Routing Protocols

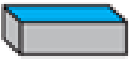


- **Combination of rules and procedures** that lets routers in the internet inform each other of changes.
  - Allows **routers to share whatever they know about the internet** or their neighbourhood
  - Can be either an **interior protocol** or an **exterior protocol**
  - **Interior protocol** – used for **Intra-domain routing**
  - **Exterior protocol** – used for **Inter-domain routing**
- 

# Intra and inter domain routing – Autonomous System Concept

- An internet can be so large that one routing protocol cannot handle the task of updating the routing tables of all routers
- An internet is divided into **autonomous systems (AS)**
- An **AS** is a group of networks and routers under the authority of a single administration
- Routing inside an autonomous system is called **intra-domain routing** (interior protocol)(**Interior Gateway Protocol**)(**IGP**)
- Routing between autonomous systems is called **inter-domain routing** (exterior protocol)(**Exterior Gateway Protocol**)(**EGP**)

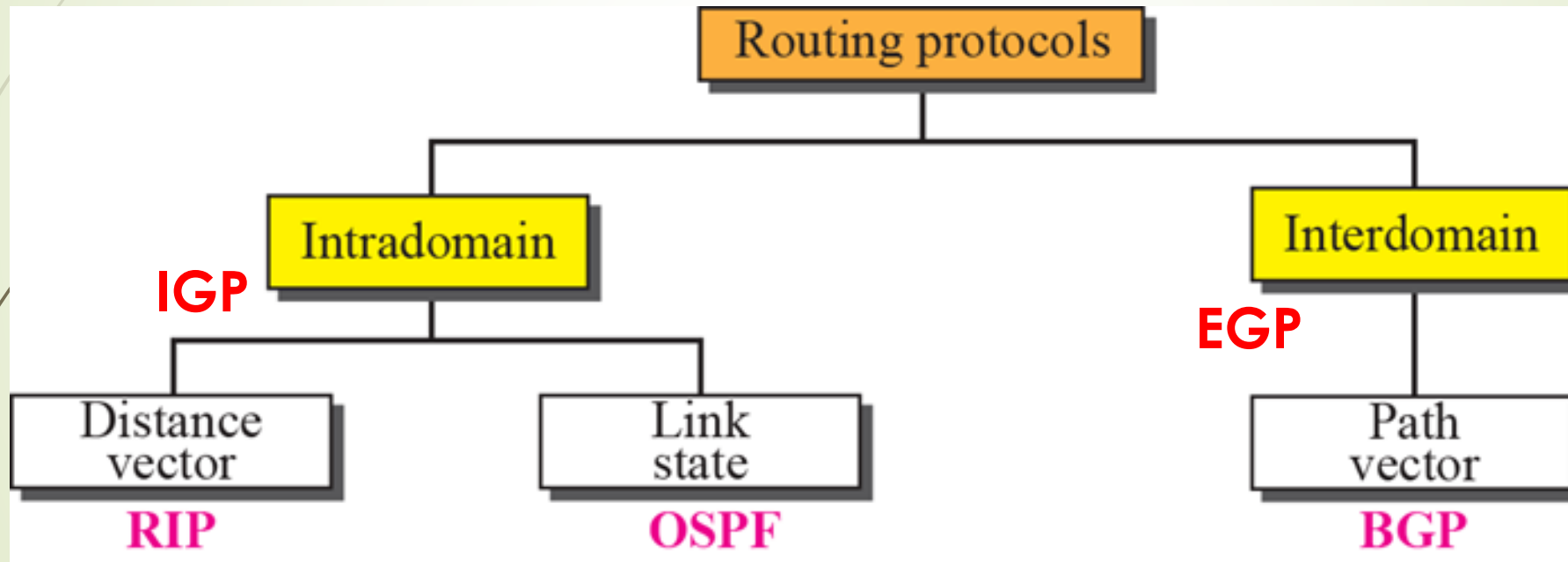


### Legend

AS	Autonomous System
	Ethernet switch
	Point-to-point WAN
	Inter-system connection



# Popular Routing Protocols



# Interior Gateway Protocol (IGP)

- Routing Protocol used for exchanging routing table information between gateways (commonly routers) **within an autonomous system**
- Some of the common interior gateway protocols are:

## 1. Routing Information Protocol (RIP)

- RIP uses a **distance vector routing algorithm** to calculate the best path to a destination based on the number of hops in the path.

## 2. Open Shortest Path First (OSPF)

- OSPF uses a **link state routing** algorithm.
- It computes the shortest-path tree for each router using a method based on Dijkstra's algorithm.
- OSPF divides an AS into routing **areas** to simplify administration and optimize traffic

# Interior Gateway Protocol (IGP)

## 3. Intermediate system to intermediate system (IS-IS)

- IS-IS is a link-state routing protocol, uses Dijkstra's algorithm for computing the best path through the network.
- Packets (datagrams) are then forwarded, based on the computed ideal path, through the network to the destination.

## 4. Enhanced Interior Gateway Routing Protocol (EIGRP)

- This have **both the features** of distance vector routing protocols and link-state routing protocols.

# Exterior Gateway Protocol (EGP)

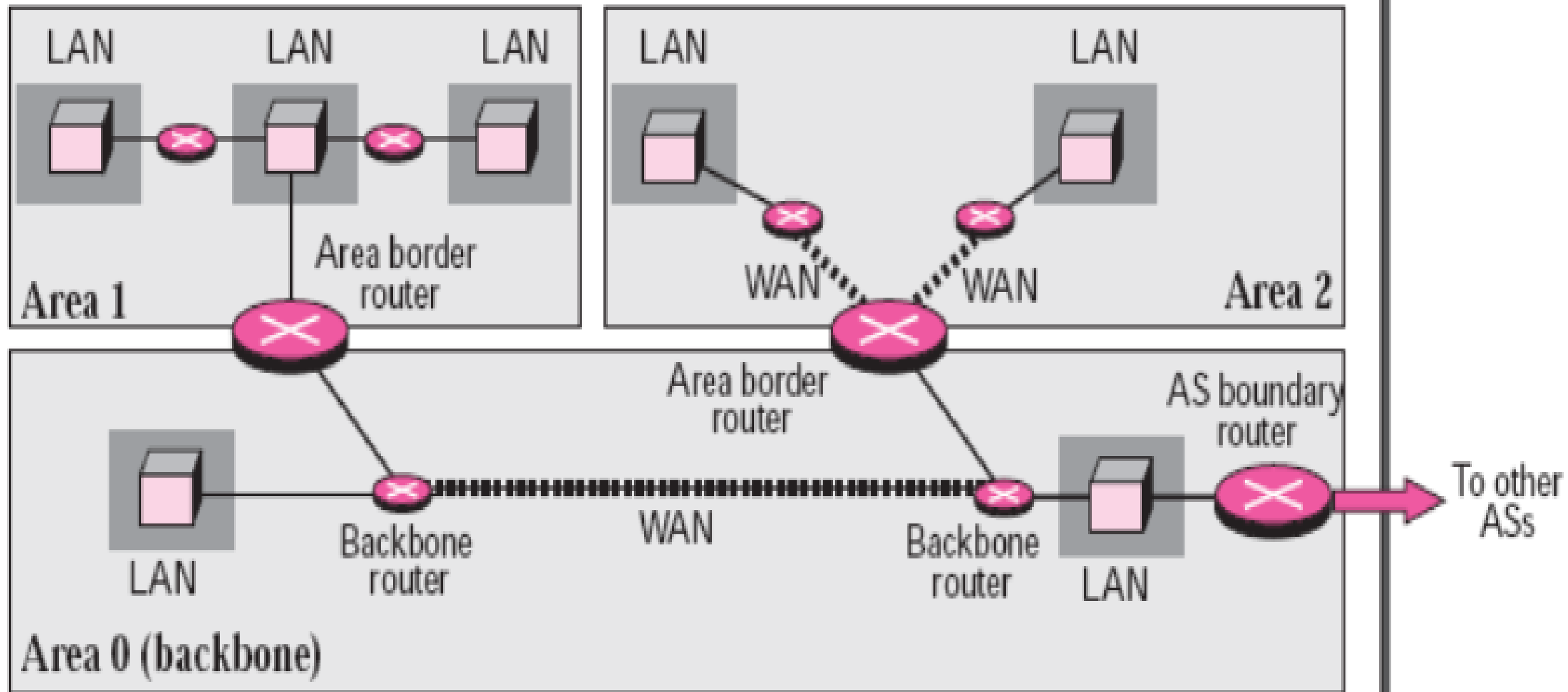
- Routing protocol used to exchange **network reachability information** (routing table information) **between two neighbor gateway(boundary) routers** belonging to the **different autonomous systems** on the large internetworks based on the TCP/IP protocol - from the mid-1980s.
- **Border Gateway Protocol (BGP)** is the only Exterior Gateway Protocol (EGP) from mid-1990s
- BGP is classified as a **path vector routing protocol**, and it makes **routing decisions based on paths**, network policies, or rule-sets configured by a network administrator.





# Areas (in OSPF)

- To handle routing efficiently and in a timely manner, can divide an autonomous system into areas.
- An **area** is a collection of networks, hosts, and routers all contained within an autonomous system.
- An autonomous system can be divided into many different areas.
- All networks inside an area must be connected.

# Autonomous System (AS)

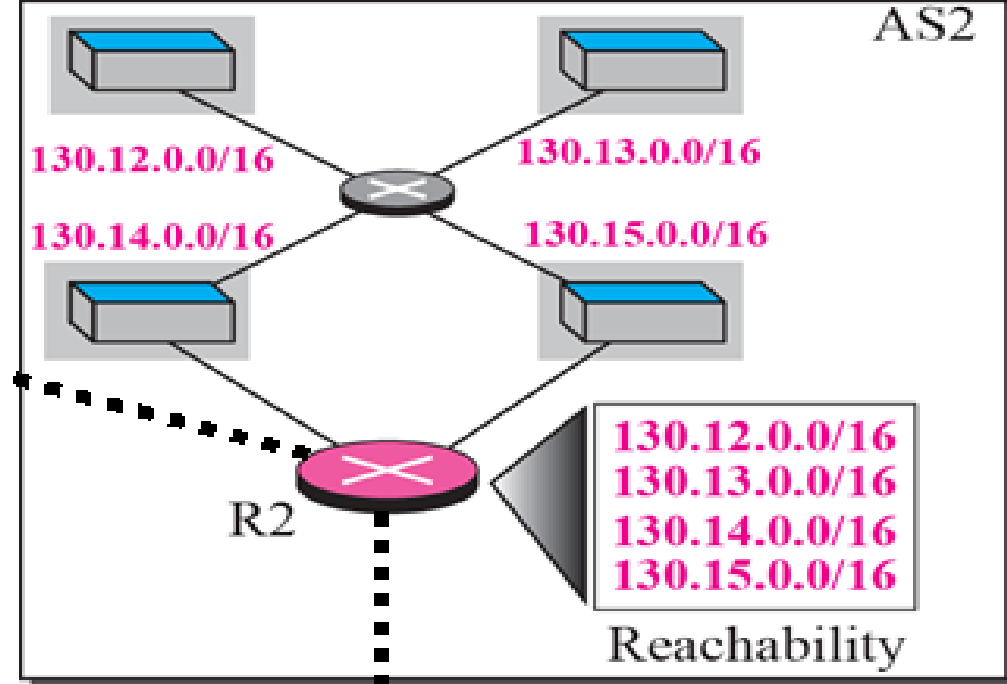
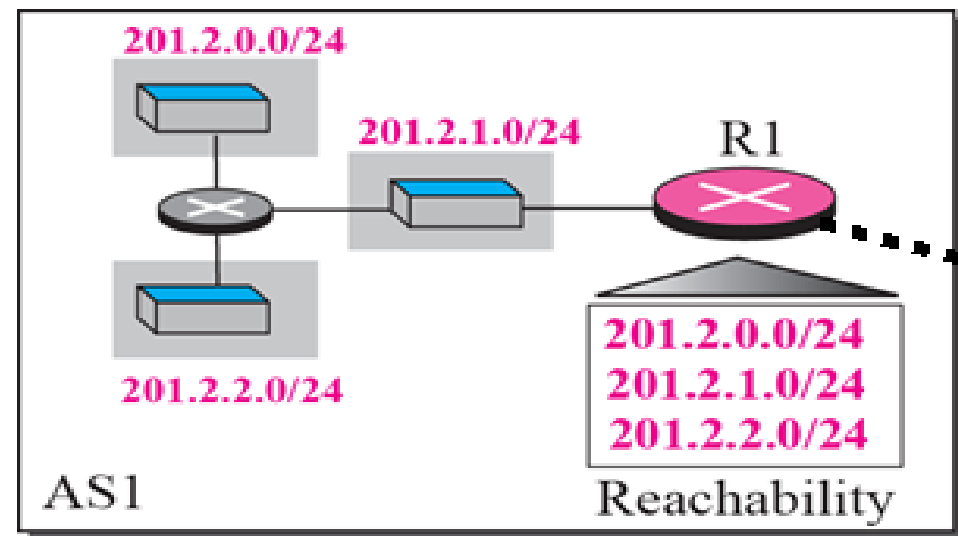


- 
- 
- Routers inside an area flood the area with routing information.
  - At the border of an area, special routers called **area border routers** summarize the information about the area and send it to other areas.
  - Among the areas inside an autonomous system is a special area called the **backbone**; all of the areas inside an autonomous system must be connected to the backbone.
  - In other words, the backbone serves as a primary area and the other areas as secondary areas.
  - The routers inside the backbone are called the **backbone routers**.
  - Each area has an **area identification**.
  - The area identification of the backbone is zero.
  - **AS Boundary Router** : communicates to other AS's



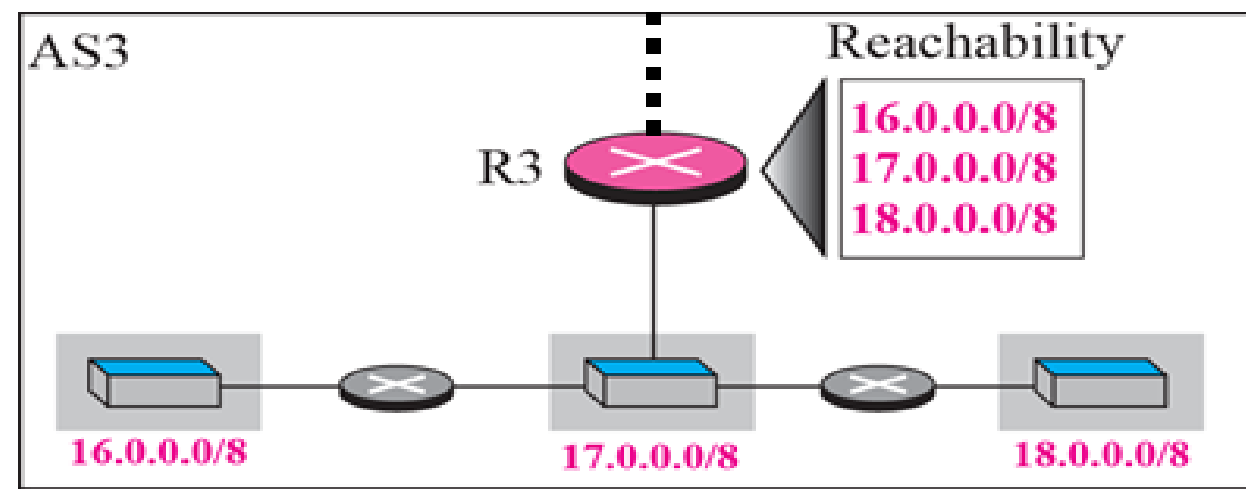
# Reachability

- The routing information passed between autonomous systems is called *reachability information*.
- Reachability information is simply information about which networks can be reached through a specific autonomous system.
- Each router created a list which shows which network is *reachable in that AS* with its network address (CIDR prefix)
- AS needs to have list of existing networks in its territory

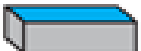




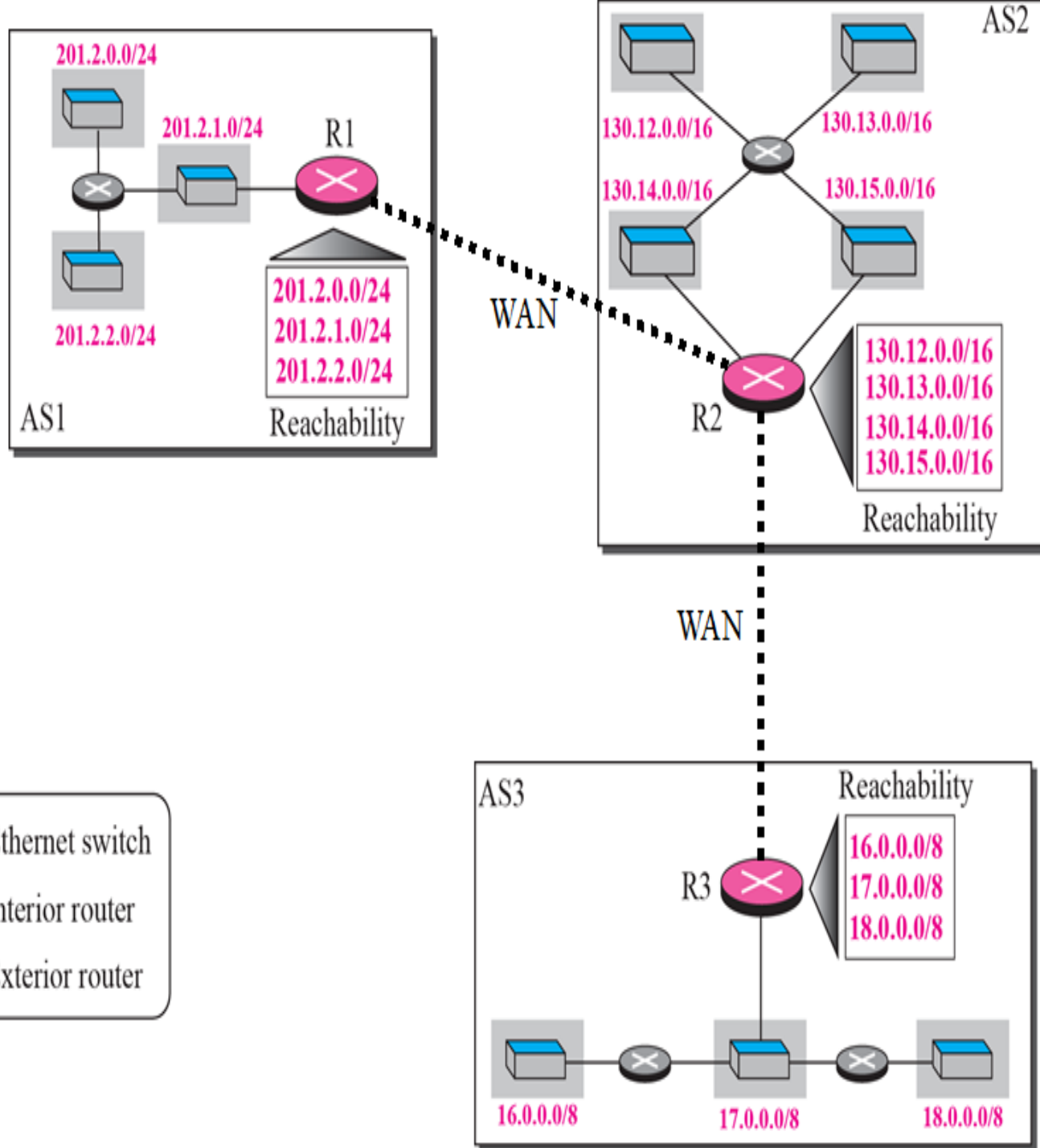
WAN

WAN



### Legend

-  Ethernet switch
-  Interior router
-  Exterior router



R1


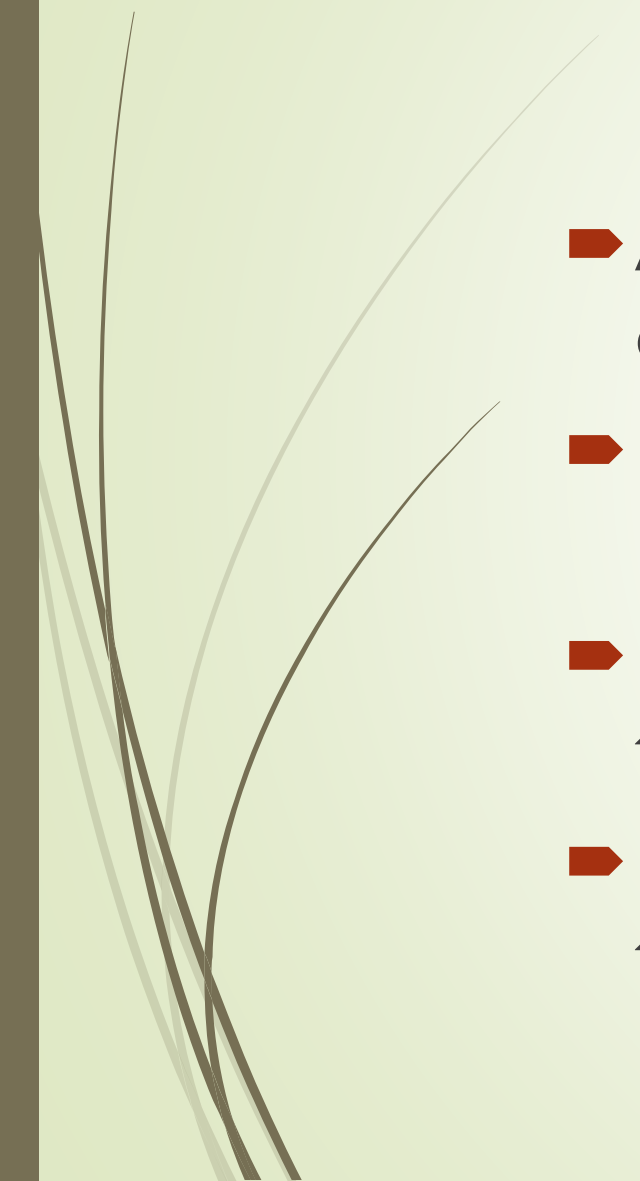
Network	Path
201.2.0.0/24	AS1 (This AS)
201.2.1.0/24	AS1 (This AS)
201.2.2.0/24	AS1 (This AS)
130.12.0.0/16	AS1, AS2
130.13.0.0/16	AS1, AS2
130.14.0.0/16	AS1, AS2
130.15.0.0/16	AS1, AS2
16.0.0.0/8	AS1, AS2, AS3
17.0.0.0/8	AS1, AS2, AS3
18.0.0.0/8	AS1, AS2, AS3

Path-Vector Routing Table

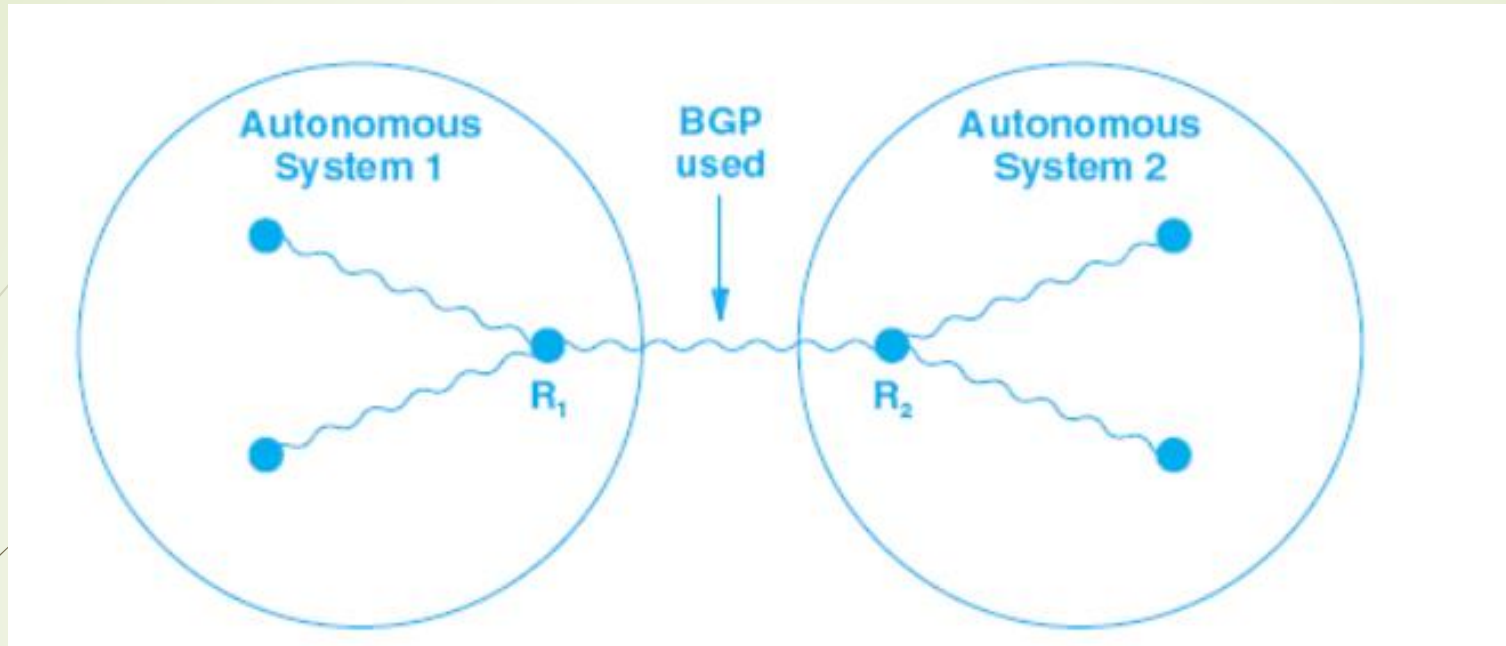
R3

Network	Path	Path
201.2.0.0/24	AS3, AS2, AS1	AS1
201.2.1.0/24	AS3, AS2, AS1	AS1
201.2.2.0/24	AS3, AS2, AS1	AS1
130.12.0.0/16	AS3, AS2	This AS)
130.13.0.0/16	AS3, AS2	This AS)
130.14.0.0/16	AS3, AS2	This AS)
130.15.0.0/16	AS3, AS2	This AS)
16.0.0.0/8	AS3 (This AS)	AS3
17.0.0.0/8	AS3 (This AS)	AS3
18.0.0.0/8	AS3 (This AS)	AS3

Path-Vector Routing Table

- 
- 
- A path vector routing table for each router can be created if **ASs share their reachability list with each other.**
  - Router R1 in AS1 can send its reachability list to router R2.
  - Router R2, after combining its reachability list, can send the result to both R1 and R3.
  - Router R3 can send its reachability list to R2, which in turn improves its routing table, and so on.

# About BGP



- R1 and R2 - *BGP peers*
- Router *R1* gathers information about networks in autonomous system 1
- Uses BGP to report the information to router *R2*
- Router *R2* gathers information about networks in autonomous system 2 and uses BGP to report the information to router *R1*.



# About BGP

- ➡ **BGP** advertises **reachability** instead of routing information
- ➡ BGP does not use either the distance-vector algorithm or the link-state algorithm.
- ➡ BGP uses a modification known as a ***path-vector* algorithm** - gives the path to autonomous systems



# BGP CHARACTERISTICS

## **1. *Inter-Autonomous System Communication :***

BGP is as an **exterior gateway protocol**, its primary role is to **allow one autonomous system to communicate with another.**

## **2. *Coordination Among Multiple BGP Speakers :***

If an **autonomous system** has **multiple routers** each communicating with a peer in an outside autonomous system, a form of BGP known as **iBGP** can be used to **coordinate among routers inside the system** to guarantee that **they all propagate consistent information.**



# BGP CHARACTERISTICS

## 3. *Propagation Of Reachability Information:*

BGP allows an autonomous system to **advertise destinations that are reachable** either in or through it, and to **learn** such information from another autonomous system.

## 4. *Next-Hop Paradigm :*

BGP supplies *next hop(path)* information for each destination.

## 5. *Policy Support:*

A router running BGP can be configured to distinguish between the **set of destinations reachable** by computers inside its autonomous system and the **set of destinations advertised** to other autonomous systems.

## 6. *Reliable Transport:*      **BGP** uses **TCP** for all communication

# BGP CHARACTERISTICS

## 7. *Path Information:*

BGP uses a **path-vector paradigm** in which advertisements **specify path information** that allows a receiver to learn a series of autonomous systems along a path to the destination.

## 8. *Incremental Updates:*

BGP does **not pass full information** in each update message.

Instead, full information is exchanged once, and then successive messages carry **incremental changes** called **deltas**.

## 9. *Support For IPv4 and IPv6 :*

BGP supports **IPv4 classless addresses** and **IPv6 addresses**.

That is, **BGP** sends a **prefix length** along with each address.



# BGP CHARACTERISTICS

## ***10. Route Aggregation:***

BGP allowing a **sender to aggregate route information** and **send a single entry** to represent multiple, related destinations

## ***11. Authentication:***

BGP allows a receiver to authenticate messages (i.e., verify the identity of a sender).

# BGP Functionality

## ➤ BGP peers perform three basic functions.

### 1. Initial peer acquisition and authentication.

The two peers **establish a TCP connection** and perform a message exchange that guarantees both sides have agreed to communicate.

### 2. Forms the **primary focus of the protocol** — each side **sends positive or negative reachability information**.

That is, a sender can advertise that one or more destinations are reachable by giving a next hop for each, or the sender can declare that one or more previously advertised destinations are no longer reachable.

### 3. Provides **ongoing verification** that the peers and the network connections between them are **functioning correctly**.

# BGP Message Types

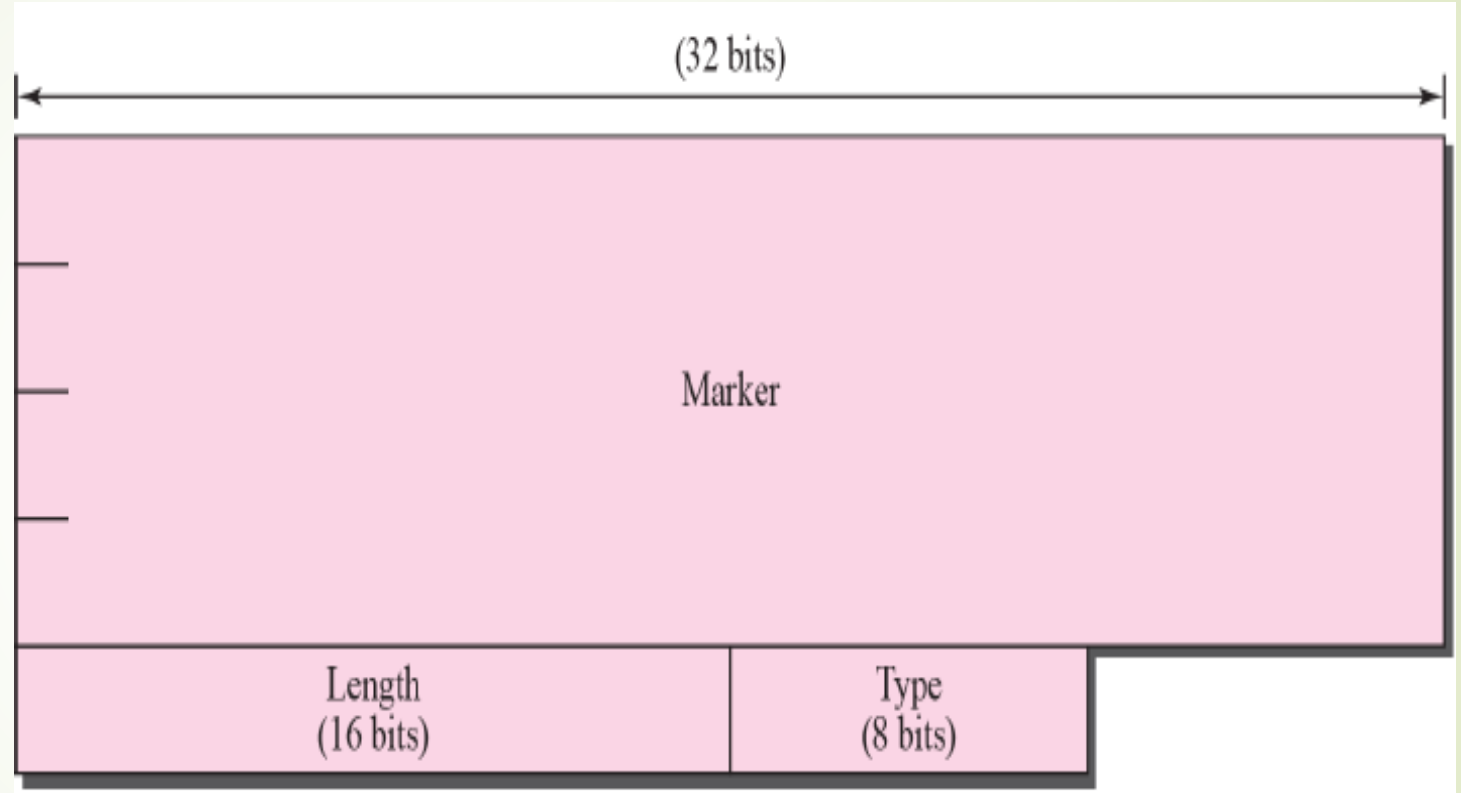
- To handle these three functions, BGP defines **five basic message types**.

Type Code	Message Type	Description
1	OPEN	Initialize communication
2	UPDATE	Advertise or withdraw routes
3	NOTIFICATION	Response to an incorrect message
4	KEEPALIVE	Actively test peer connectivity
5	REFRESH	Request readvertisement from peer



# BGP Message Header

- Each BGP message begins with a **fixed header** that **identifies the message type**.



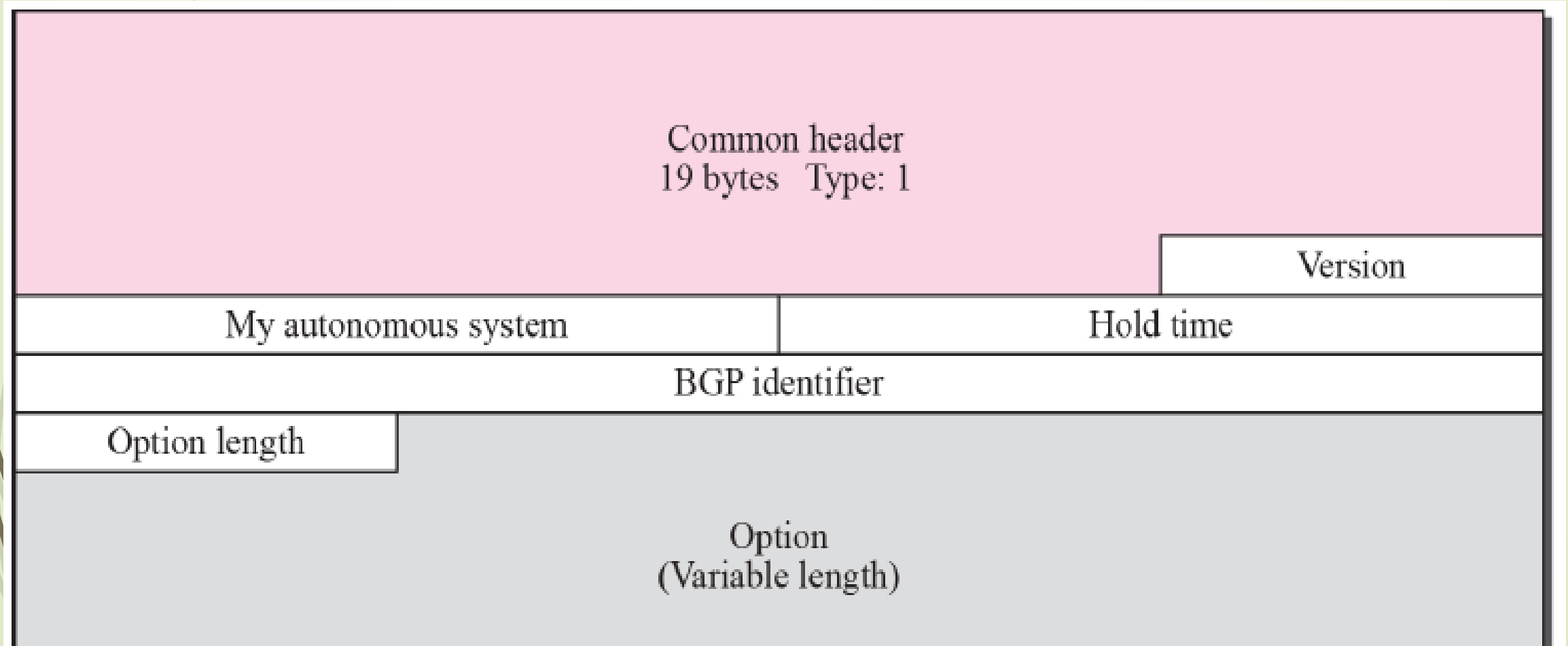
- **Marker:** 16-byte – reserved **for authentication**
- **Length:** 2-byte - **length of the total message** including the header
- **Type:** 1-byte - defines the **type of the packet**(1-5)

# I. Open Message

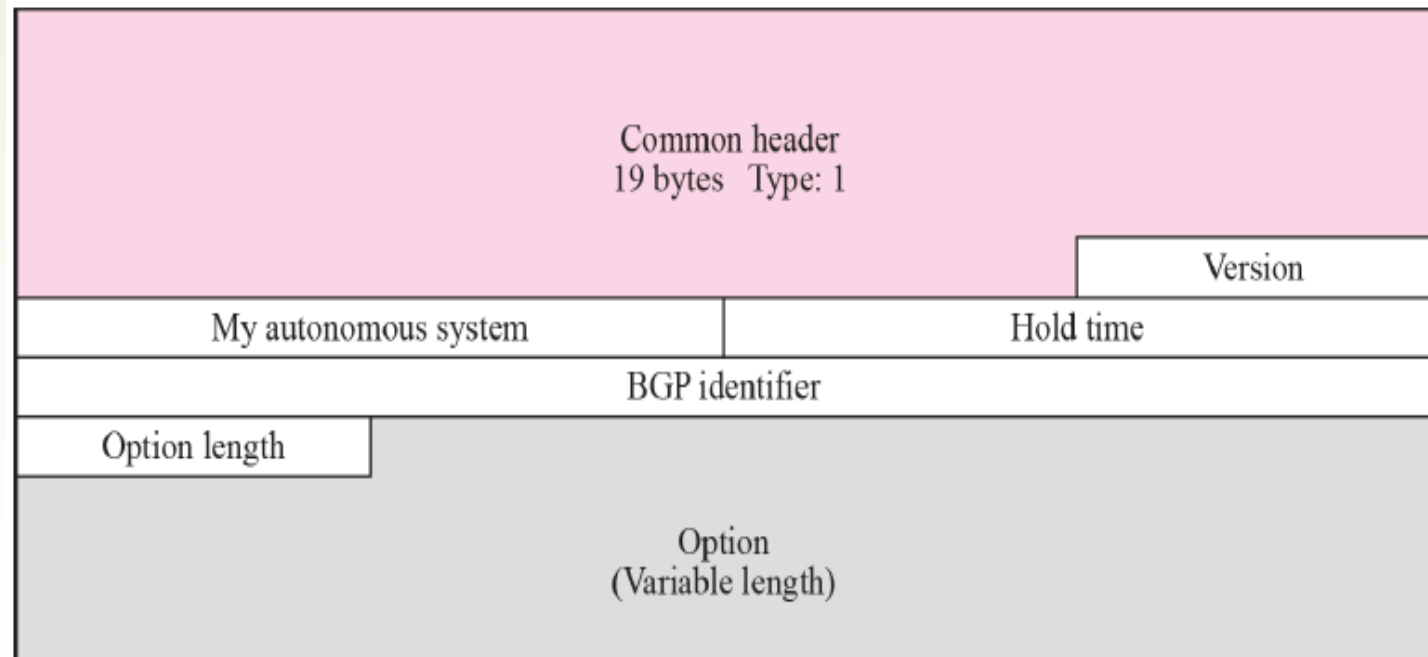
- ➡ To create a neighborhood relationship, **a router running BGP opens a TCP connection with a neighbor** and sends an ***open message***.
- ➡ If the neighbor accepts the neighborhood relationship, it responds with a ***keepalive message***, which means that a **relationship has been established between the two routers**.



# I. Open Message



# Open Message Format

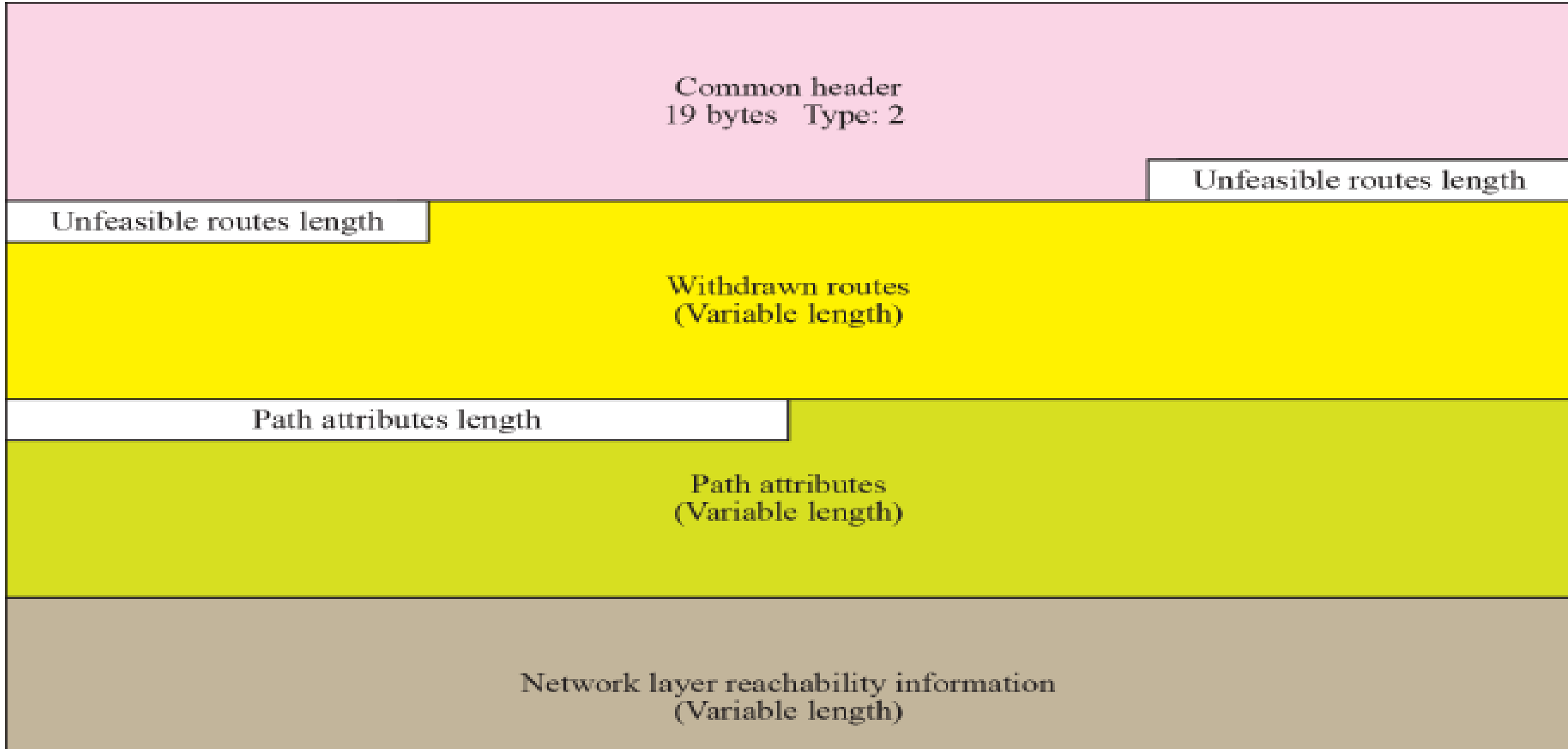


- ➔ **Version:** 1-byte - current version is **4**
- ➔ **My autonomous system:** 2-byte - defines **AS number**
- ➔ **Hold time:** 2-byte - **maximum number of seconds** that can elapse until receive a keepalive or update message
- ➔ **BGP identifier:** 4-byte - **defines the router(IP Address)** that send open message
- ➔ **Option length:** 1-byte - length of the total **option parameters**
- ➔ **Option parameters:** **authentication** parameter

## 2. Update Message

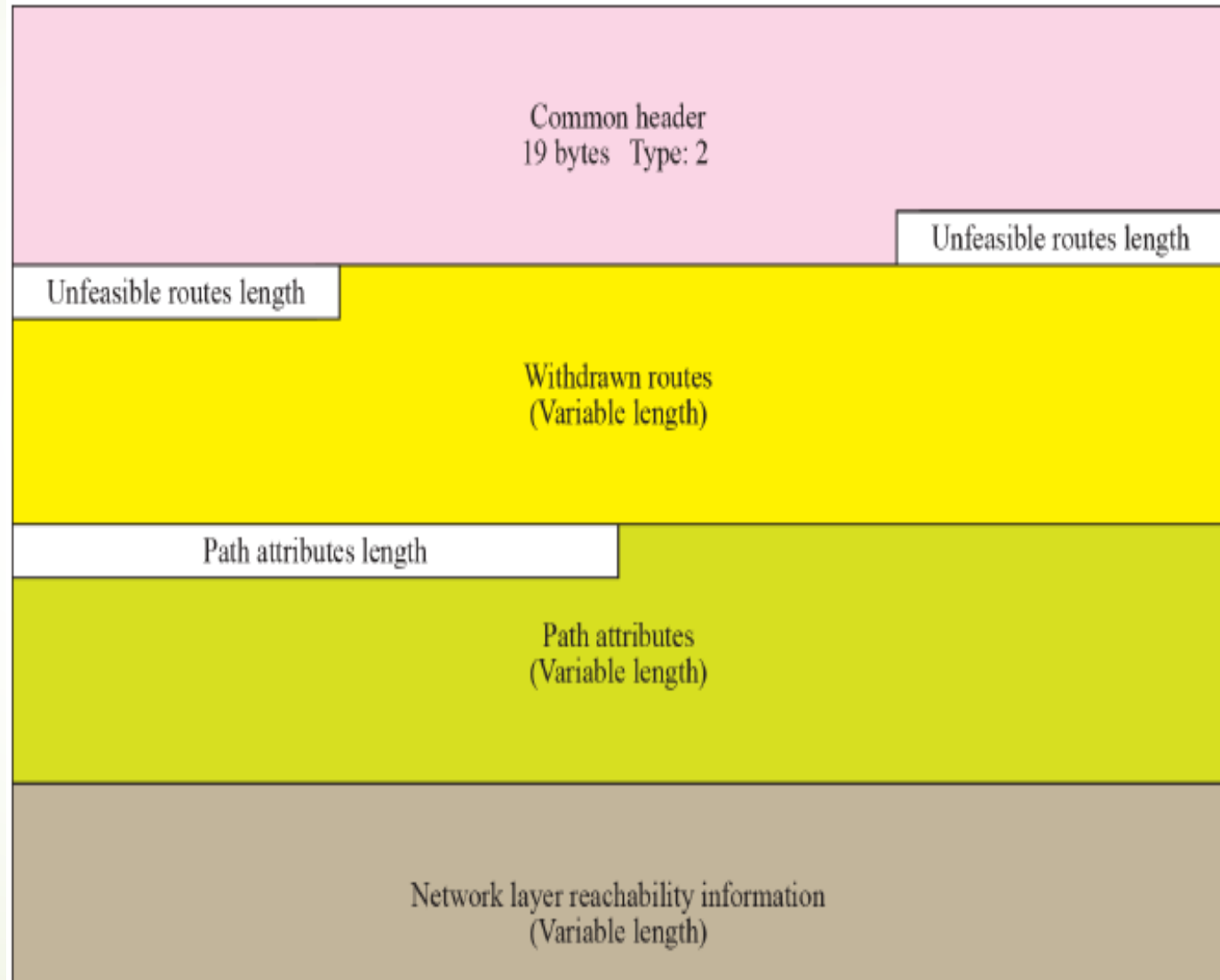
- ➡ It is used by a BGP router
  - ➡ to **withdraw destinations that have been advertised previously**
  - ➡ **announce a route to a new destination**

# Update Message Format



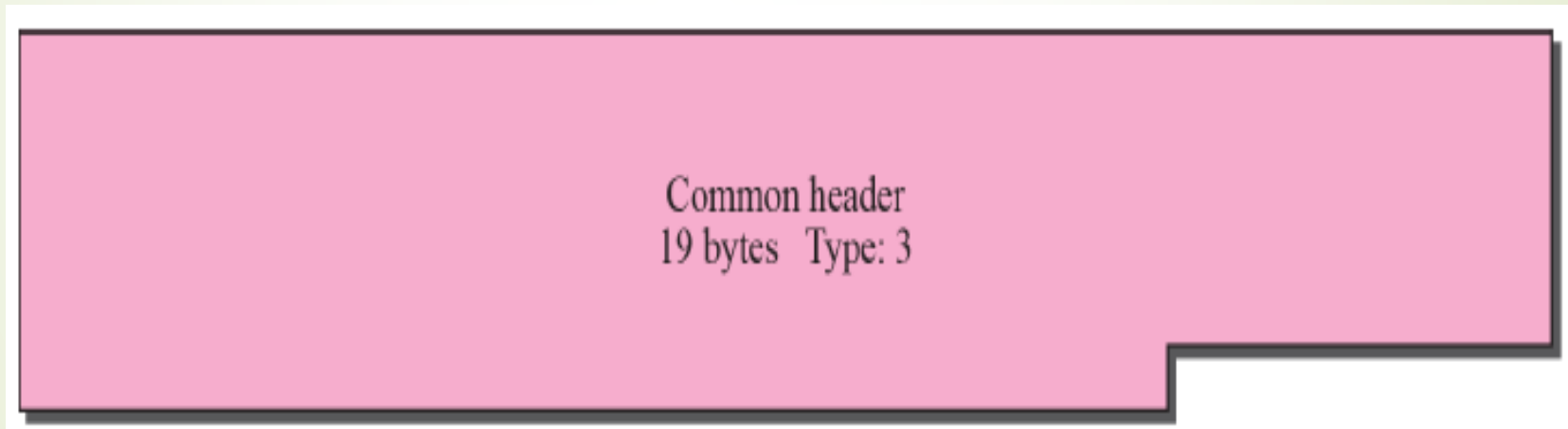
# Update Message Format

- **Unfeasible routes length:** 2-byte - length of the next field
- **Withdrawn routes:** **list the routes deleted** from previous advertised list
- **Path attribute length:** 2-byte - length of the next field
- **Path attributes:** **attributes of the path**
- **Network layer reachability information(NLRI):** **defines the network actually advertise this message**



### 3. Keepalive Message

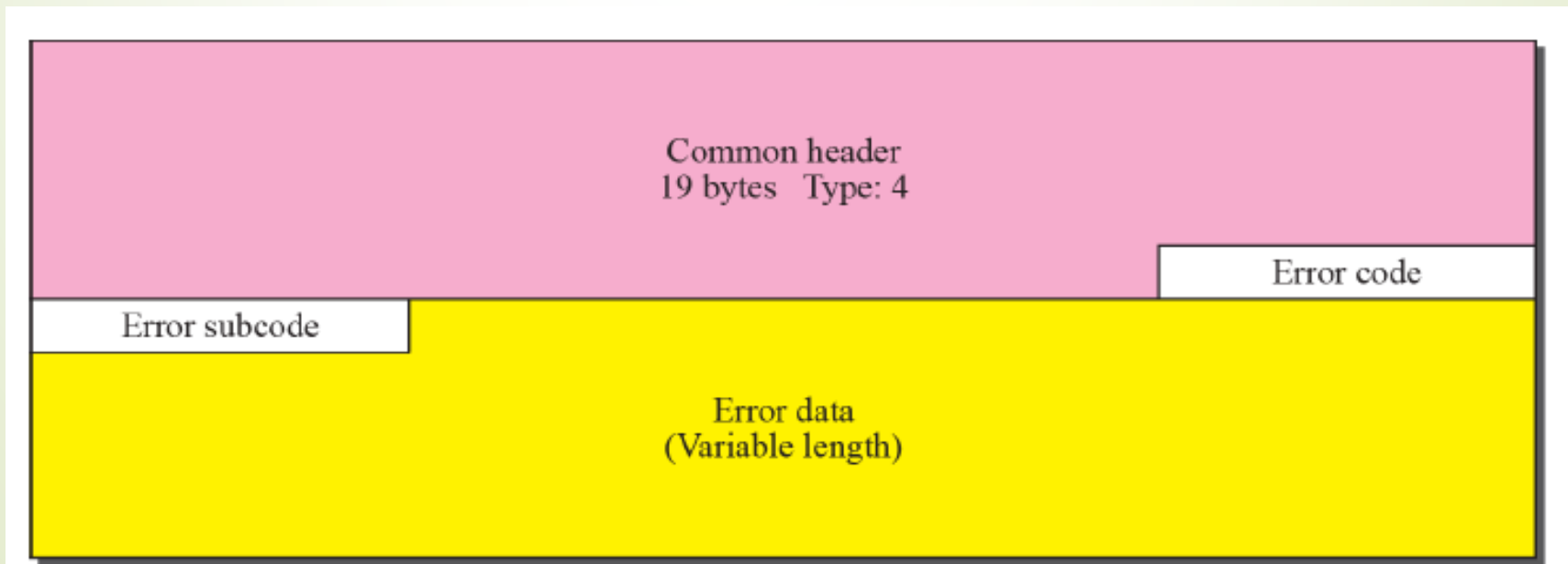
- Two BGP peers regularly send this message to tell that are alive and to test network connectivity



- They will be informed if the TCP connection fails.


## 4. Notification Message

- Notification send by router **whenever an error condition detected** and then closes the TCP connection
- Error code: 1-byte - **category of the error**
- Error subcode: 1-byte - **type of error in each category**
- Error data: **more information about error**





<i>Error Code</i>	<i>Error Code Description</i>	<i>Error Subcode Description</i>
1	Message header error	Three different subcodes are defined for this type of error: synchronization problem (1), bad message length (2), and bad message type (3).
2	Open message error	Six different subcodes are defined for this type of error: unsupported version number (1), bad peer AS (2), bad BGP identifier (3), unsupported optional parameter (4), authentication failure (5), and unacceptable hold time (6).
3	Update message error	Eleven different subcodes are defined for this type of error: malformed attribute list (1), unrecognized well-known attribute (2), missing well-known attribute (3), attribute flag error (4), attribute length error (5), invalid origin attribute (6), AS routing loop (7), invalid next hop attribute (8), optional attribute error (9), invalid network field (10), malformed AS_PATH (11).
4	Hold timer expired	No subcode defined.
5	Finite state machine error	This defines the procedural error. No subcode defined.
6	Cease	No subcode defined.



## 5. Refresh Message

- For requesting re-advertisement from the BGP router

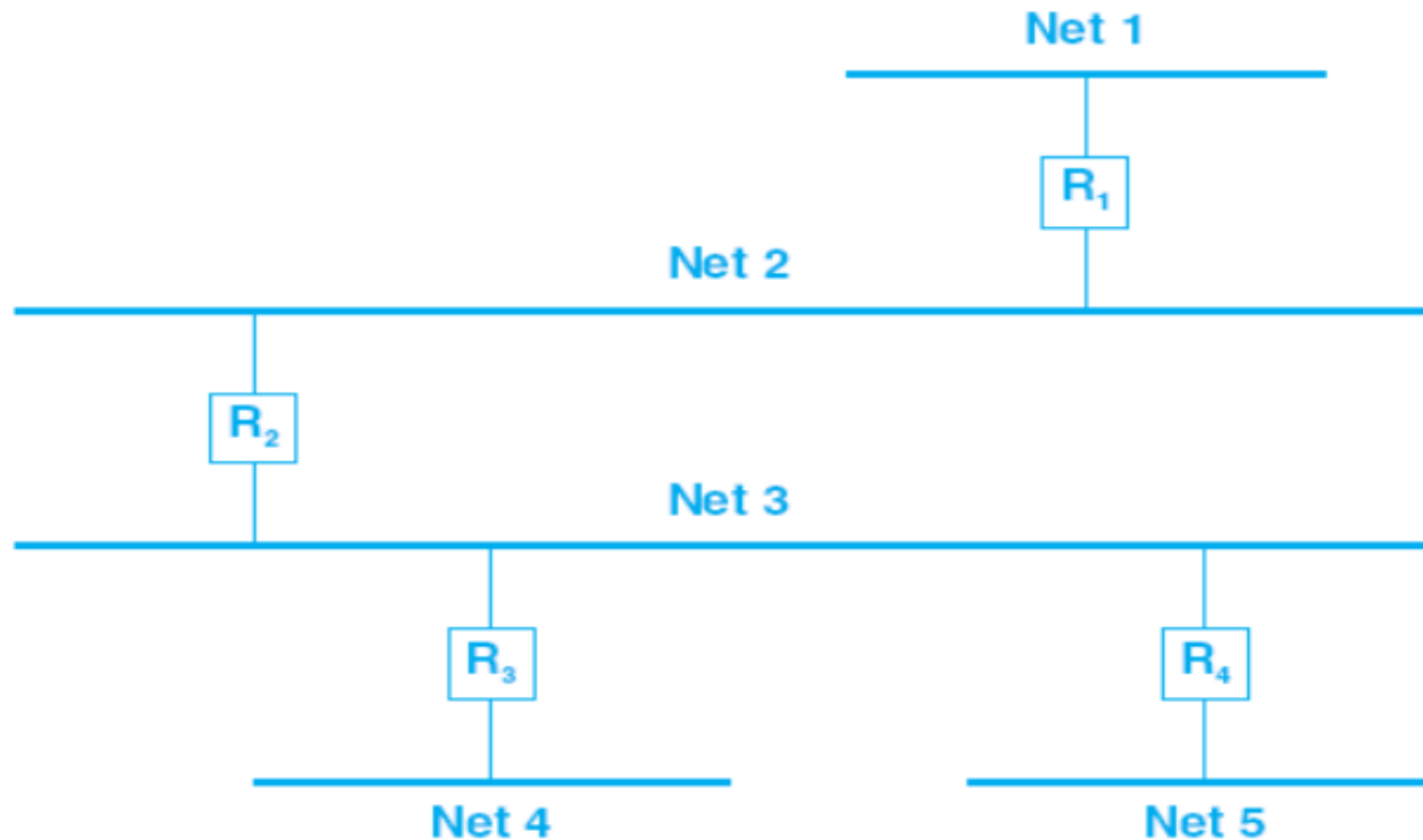


# **Routing Within An Autonomous System (RIP, RIPng, OSPF, IS-IS)**


Chapter 14 - Comer




# Static Vs. Dynamic Interior Routes

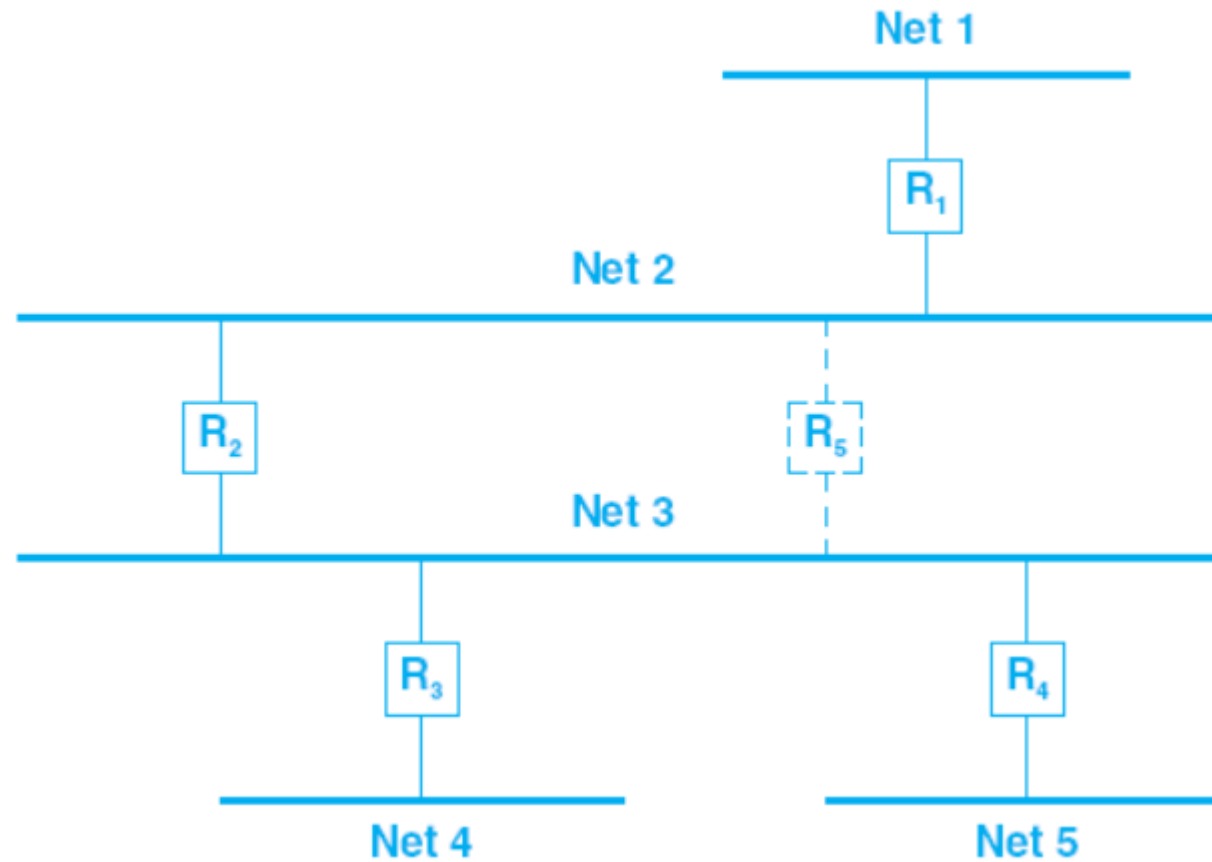


- Routing for the intranet is trivial because **only one path exists between any two points**.
- If a router fails, the intranet will be **disconnected** because there are no redundant paths.
- A **manager can configure** routes in all hosts and routers manually, and never needs to change the routes.
- But, if the intranet changes (e.g., **a new network is added**), the manager must **reconfigure** the routes accordingly

- 
- Routers within an autonomous system are said to be ***interior routers***
  - How can routers in an autonomous system learn about networks within the autonomous system?
  - In the smallest intranets, **network managers** can establish and modify routes manually.
  - The manager keeps a list of networks and updates the tables whenever a new network is added to, or deleted from, the autonomous system.

## ➤ **Static Routes**

- Disadvantages:
  - Manual systems cannot accommodate rapid growth and rely on humans to change routes whenever a network failure occurs.
- 



- **Multiple paths** exist between some hosts.
- In such cases, a manager usually chooses one path to be a **primary path**
- If a router or network along the primary path fails, routes must be changed to send traffic along an alternate path.



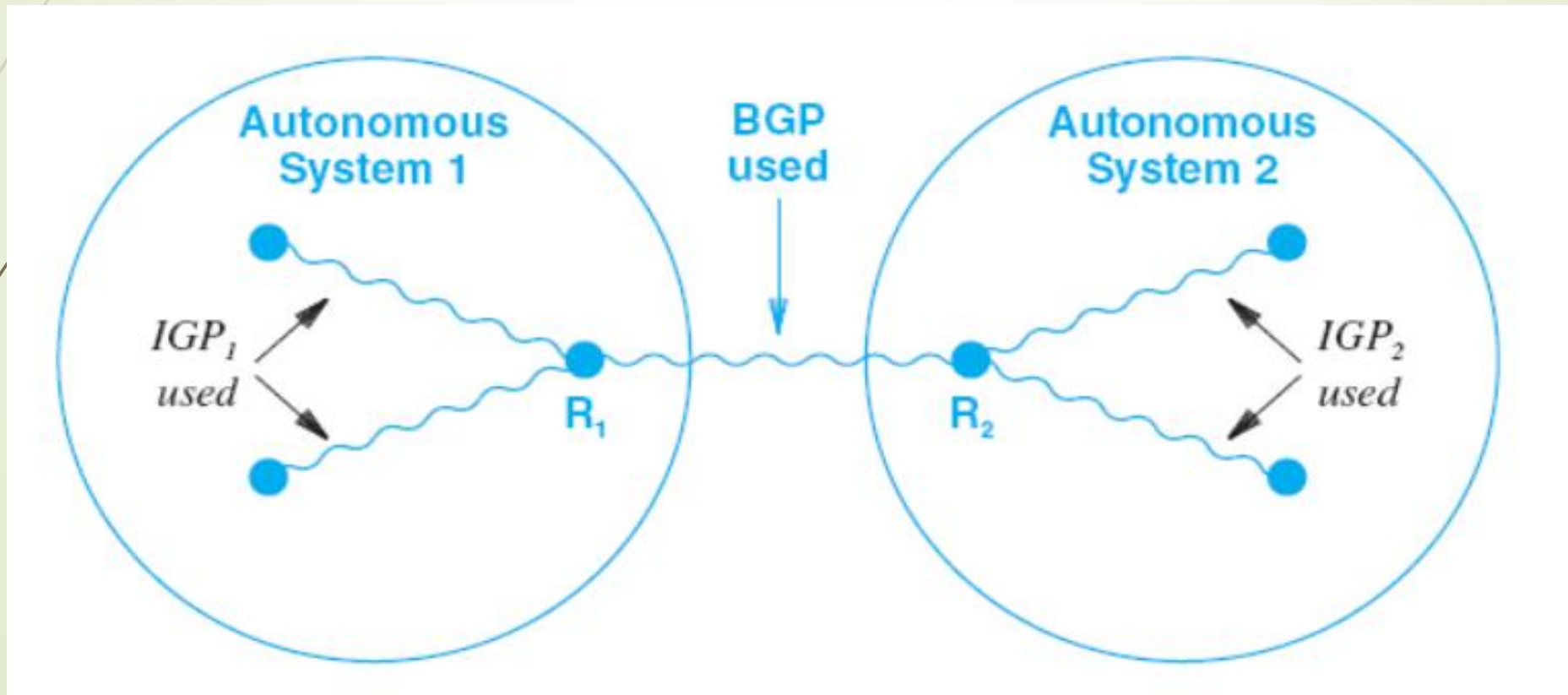



## ➤ **Advantages:**

- First, because computers can respond to failures much faster than humans, automated route changes are **less time consuming**.
- Second, because humans can make small errors when entering network addresses, automated routing is **less error-prone**.
- Thus, even in small internets, an **automated system is used to change routes quickly and reliably**.

## ➤ **Dynamic Routes**

Any protocol that interior routers use when they exchange routing information- ***Interior Gateway Protocol (IGP)***.





In the figure, *IGP1* refers to the interior routing protocol used within autonomous system 1, and *IGP2* refers to the protocol used within autonomous system 2.

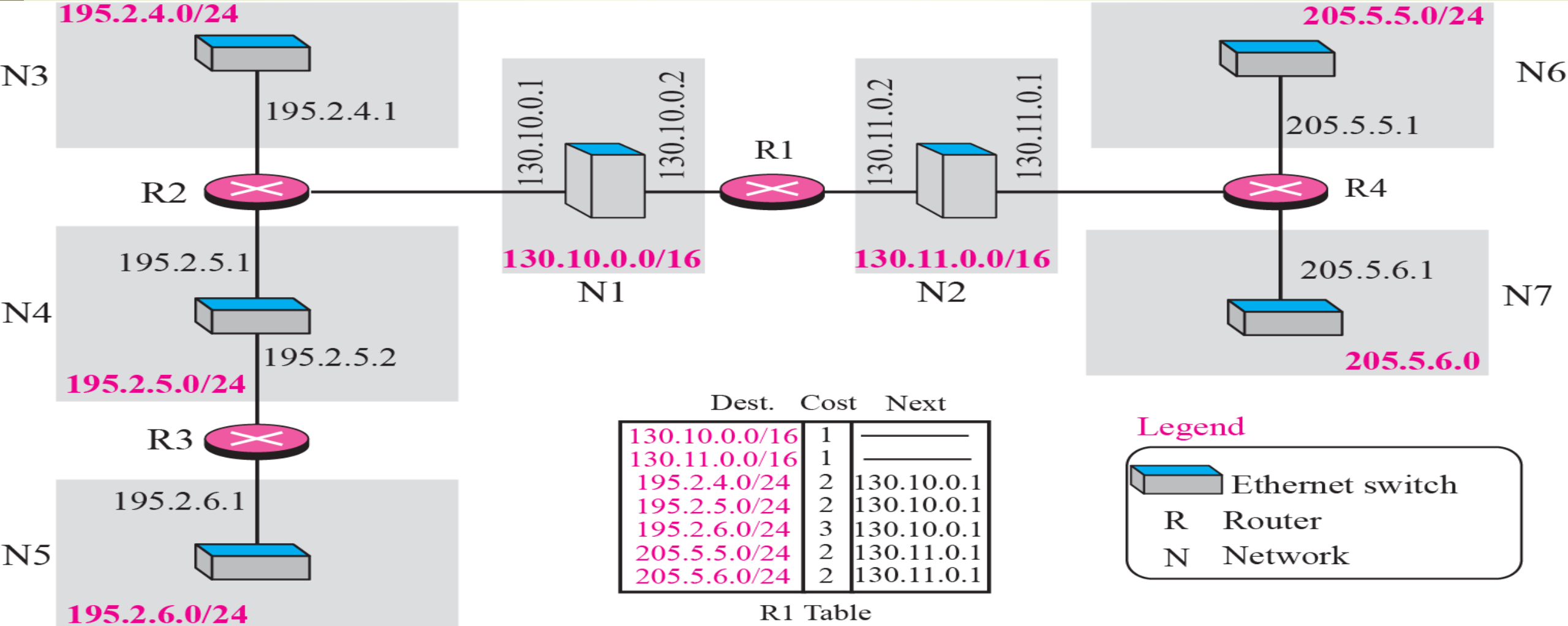
- Router *R1* will use *IGP1* to obtain routes internally, summarize the information, apply policies, and then use BGP to export the resulting information
- Similarly, router *R2* will use *IGP2* to obtain information that it exports.
- **Routers that run BGP to advertise reachability usually also need to run an IGP to obtain information from within their autonomous system.**

# Routing Information Protocol (RIP)

- An **intra-domain (interior) routing protocol** used inside an autonomous system
- It is a very simple protocol based on **distance vector routing**.
- RIP supports two type of participants: **active** and **passive**.
- **Active participants** **advertise their routes to others**;
- **Passive participants** **listen to RIP messages** and use them to update their forwarding table, but do not advertise.

# Routing Table

- RIP implements DVR with some considerations :
  - In an AS there are routers and network – **The routers have routing table.**
  - The **destination** in a table is a network, which means the **first column defines a network address.**
  - The **cost** is defined as the **number of links that have to be used to reach the destination** (hop count)
  - **Infinity** is defined as **16**. An AS using RIP **can not have more than 15 hops**
  - The **next node column** defines the **address of the router to which the packet needs to be sent to reach its destination**



Dest.	Cost	Next
130.10.0.0/16	1	_____
130.11.0.0/16	1	_____
195.2.4.0/24	2	130.10.0.1
195.2.5.0/24	2	130.10.0.1
195.2.6.0/24	3	130.10.0.1
205.5.5.0/24	2	130.11.0.1
205.5.6.0/24	2	130.11.0.1

R1 Table

Dest.	Cost	Next
130.10.0.0/16	1	_____
130.11.0.0/16	2	130.10.0.2
195.2.4.0/24	1	_____
195.2.5.0/24	1	_____
195.2.6.0/24	2	195.2.5.2
205.5.5.0/24	3	130.10.0.2
205.5.6.0/24	3	130.10.0.2

R2 Table

Dest.	Cost	Next
130.10.0.0/16	2	195.2.5.1
130.11.0.0/16	3	195.2.5.1
195.2.4.0/24	2	195.2.5.1
195.2.5.0/24	1	_____
195.2.6.0/24	1	_____
205.5.5.0/24	4	195.2.5.1
205.5.6.0/24	4	195.2.5.1

R3 Table

Dest.	Cost	Next
130.10.0.0/16	2	130.11.0.2
130.11.0.0/16	1	_____
195.2.4.0/24	3	130.11.0.2
195.2.5.0/24	3	130.11.0.2
195.2.6.0/24	4	130.11.0.2
205.5.5.0/24	1	_____
205.5.6.0/24	1	_____

R4 Table



# Slow Convergence Problem/Count to Infinity

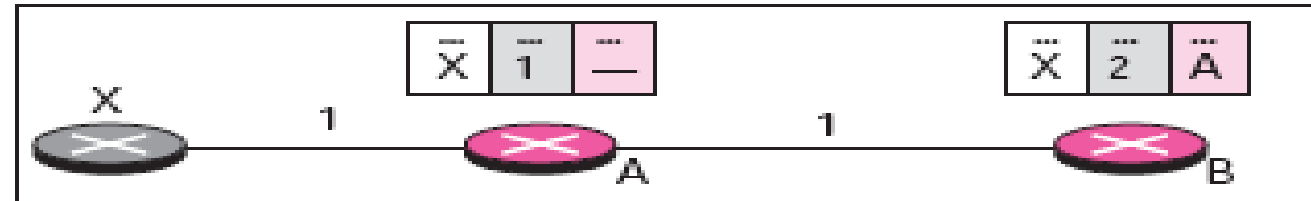
- The distance vector algorithm used by RIP can create a problem.
- **Routing update messages propagates slowly across the network.**
- Inconsistencies arises.
- Any decrease in cost propagates quickly, but any increase in cost propagates slowly.
- For a routing protocol to work properly, **if a link is broken, every other router should be aware of it immediately , but in DVR, this takes some time.**
- It takes several updates before the cost for a broken link is recorded as infinity by all routers.



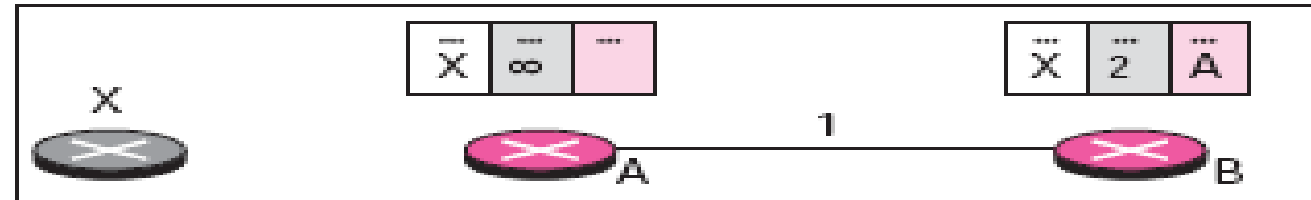
# Slow Convergence Problem/Count to infinity - Two node loop

## Example

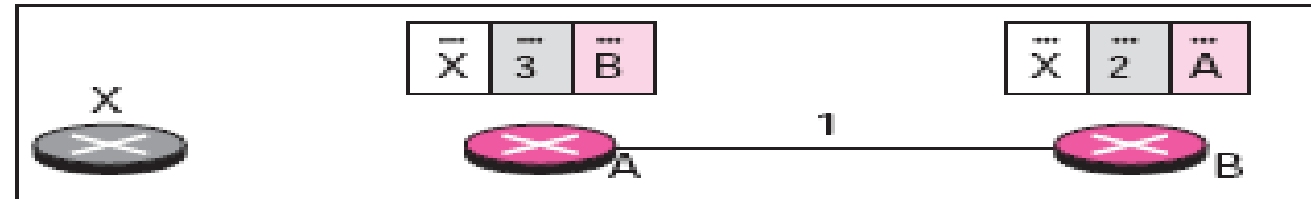
Before failure



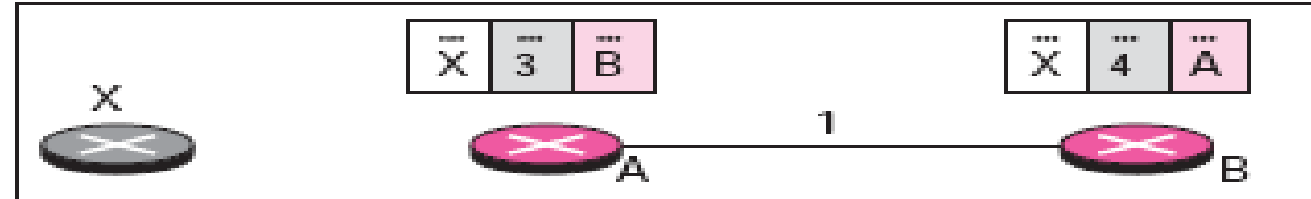
After failure



After A receives update from B

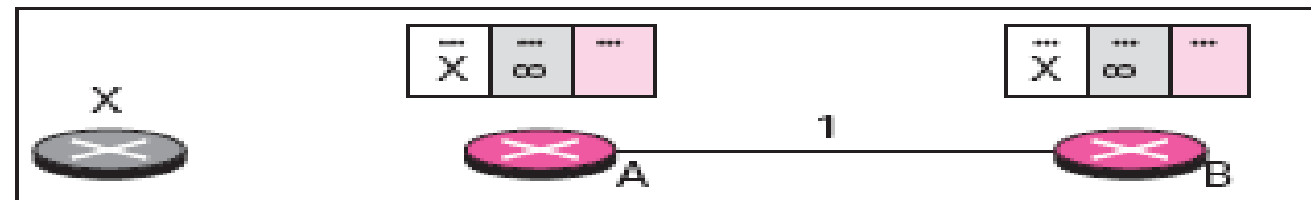


After B receives update from A



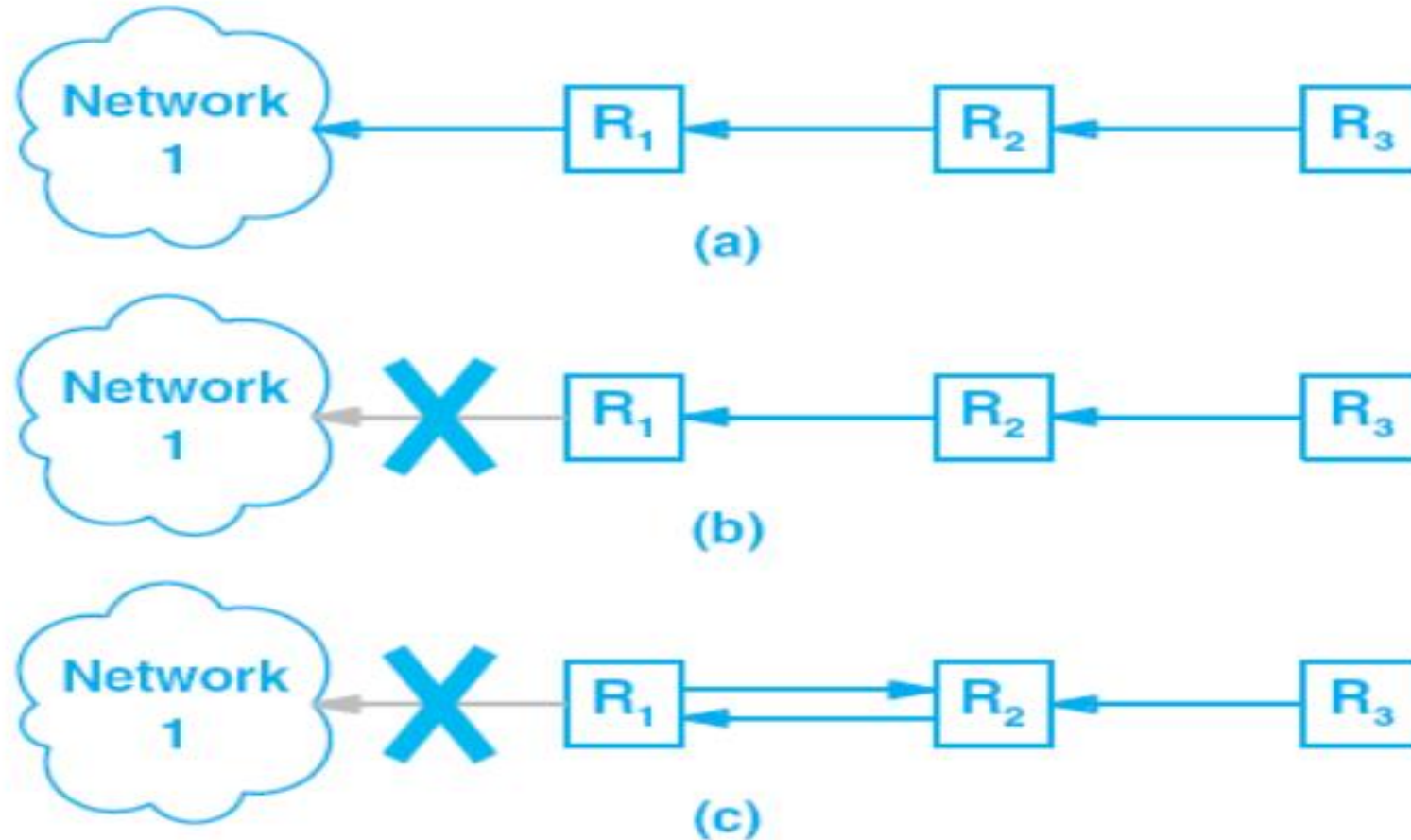
...

Finally



- 
- 
- ➡ Refer Page No : 291
    - TCP/IP Protocol Suite, Forouzan

# Illustration of the Slow Convergence Problem



**Figure 14.4** Illustration of the slow convergence problem with (a) three routers that have a route to network 1, (b) the connection to network 1 has failed and  $R_1$  has lost its route, and (c) a routing loop caused because  $R_2$  advertises a route to network 1.

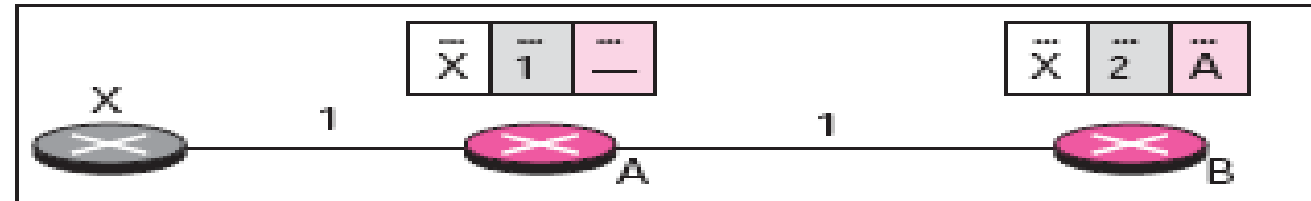
# Slow Convergence Problem

- A conventional distance-vector algorithm can **form a routing loop after a failure** occurs because routing information that a router sent can reach the router again.
- **Refer Page No : 294**
  - **Internetworking with TCP/IP , Comer**

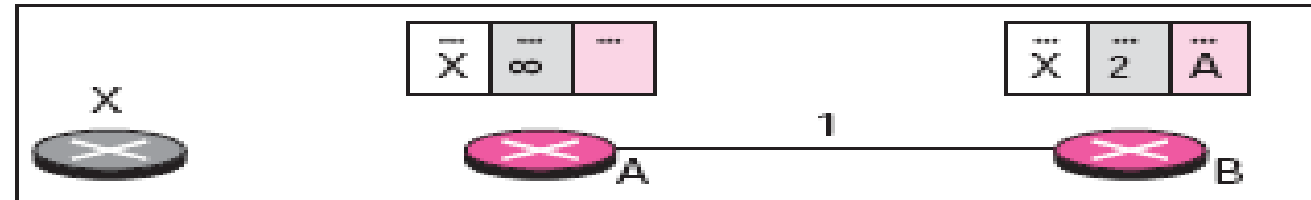
# Slow Convergence Problem/Count to infinity

## Example

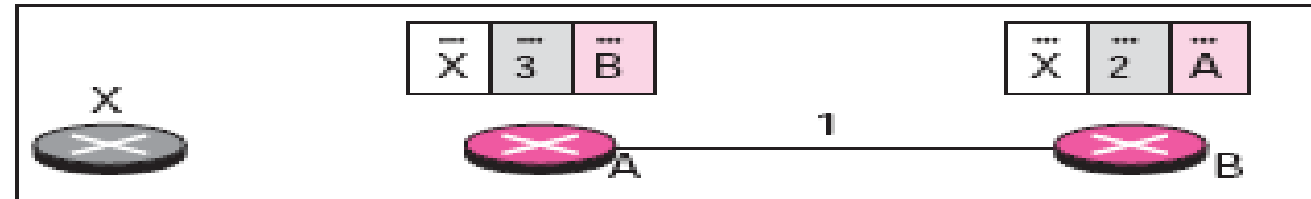
Before failure



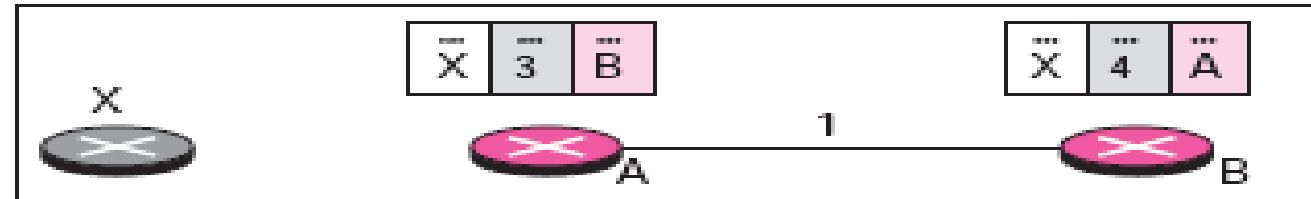
After failure



After A receives update from B

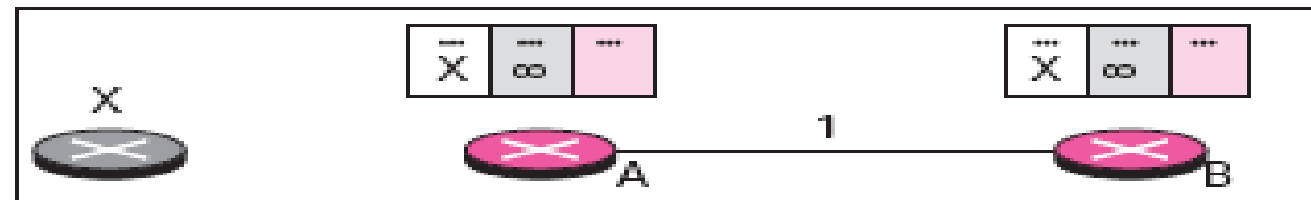


After B receives update from A



...

Finally



# Solving The Slow Convergence Problem

## ➡ Defining Infinity

- ➡ Redefine infinity to a small number, such as 16
- ➡ Most implementations of DVR define 16 as infinity
- ➡ DVR cannot be used in large system.
- ➡ **The size of the n/w, in each direction, cannot exceed 15 hops**

# Solving The Slow Convergence Problem

## ➤ Split Horizon

- **Instead of advertising the table through each interface, each node sends only part of its table through each interface**
- E.g. node B thinks that the optimum route to reach X is via A, it does not need to advertise this piece of information to A; the information has come from A
- Taking information from Node A, modifying it, and sending it back to Node A is what creates the confusion
- Node B eliminates a line of its routing table before it sends it to A.
- Later, when node A sends its routing table to B, B also corrects its routing table.
- The system becomes stable after the first update; both A and B know that X is not reachable.



# Solving The Slow Convergence Problem

## ► One drawback of Split Horizon

- Normally, the DV protocol uses a timer and if there is no news about a route, the node deletes the route from its table
- In the e.g., if node B eliminates the route to X from its advertisement to A, node A cannot guess that this is due to split horizon strategy or because B has not received any news about X recently

# Solving The Slow Convergence Problem

- Another technique used to solve the slow convergence problem employs **hold down**.
- Hold down **forces a participating router to ignore information about a network for a fixed period of time** following receipt of a message that claims the network is unreachable.
- For RIP, the hold down period is set to **60 seconds**, twice as long as a normal update period.

# Solving The Slow Convergence Problem

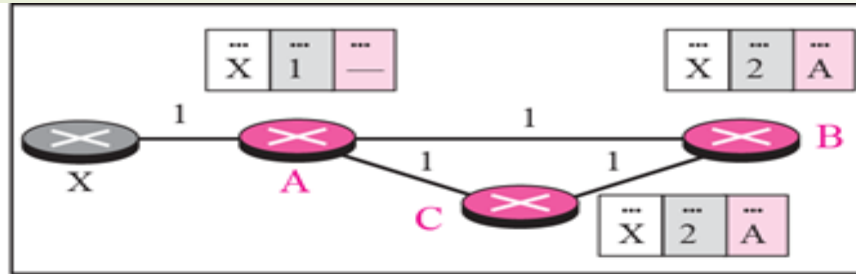
- **Split Horizon** can be combined with **Poison Reverse**
  - Node B can **still advertise the value for X**, but if the source of information is A, it can **replace the distance with infinity** as a warning:  
“Do not use this value; what I know about this route comes from you”

# Solving The Slow Convergence Problem

- **To make poison reverse most effective**, it must be combined with **triggered updates**.
- The triggered update mechanism **forces a router to broadcast routing information immediately after receiving bad news**.
- ie: the router does not wait until its next periodic broadcast.
- **By sending an update immediately, a router minimizes the time it is vulnerable.** i.e., the time during which neighbors might advertise short routes because they have not received the bad news.

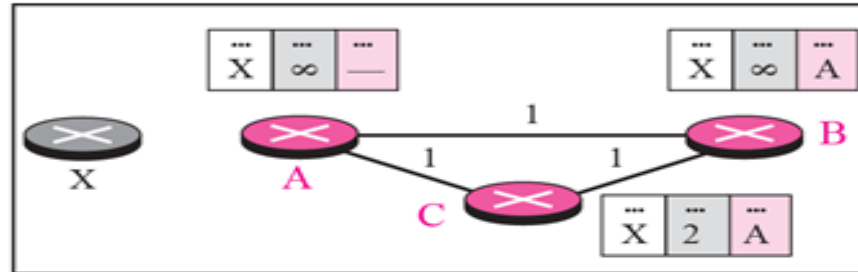
# Three Node Instability

Before failure

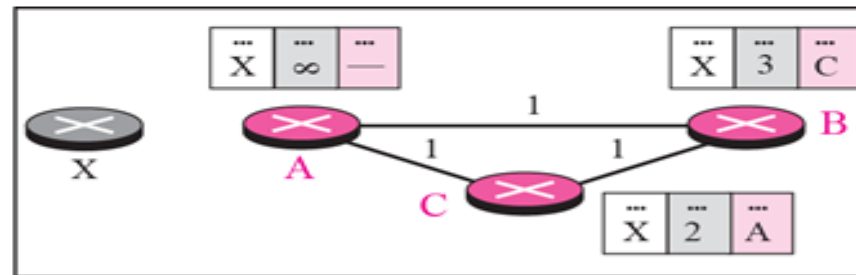


*If the instability is btw three nodes, stability cannot be guaranteed*

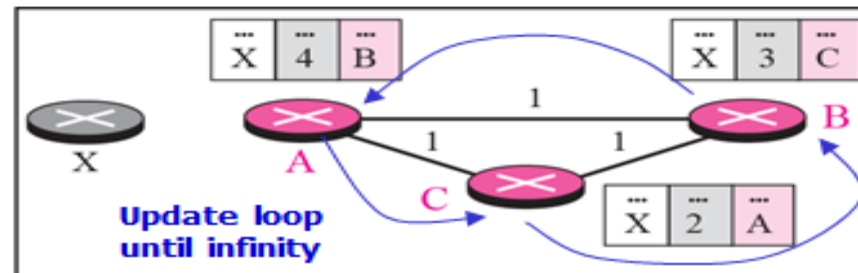
After A sends the route to B and C, but the packet to C is lost



After C sends the route to B



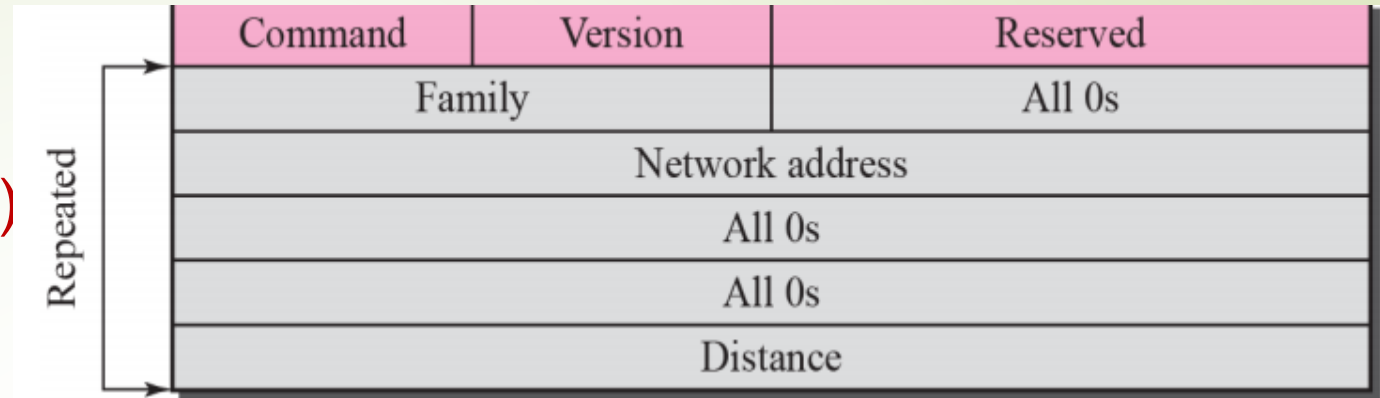
After B sends the route to A



# RIP Message Format

**I. Command** : 8 – bit field;

Type of message : **Request(1) /Response(2)**



## ➡ Request

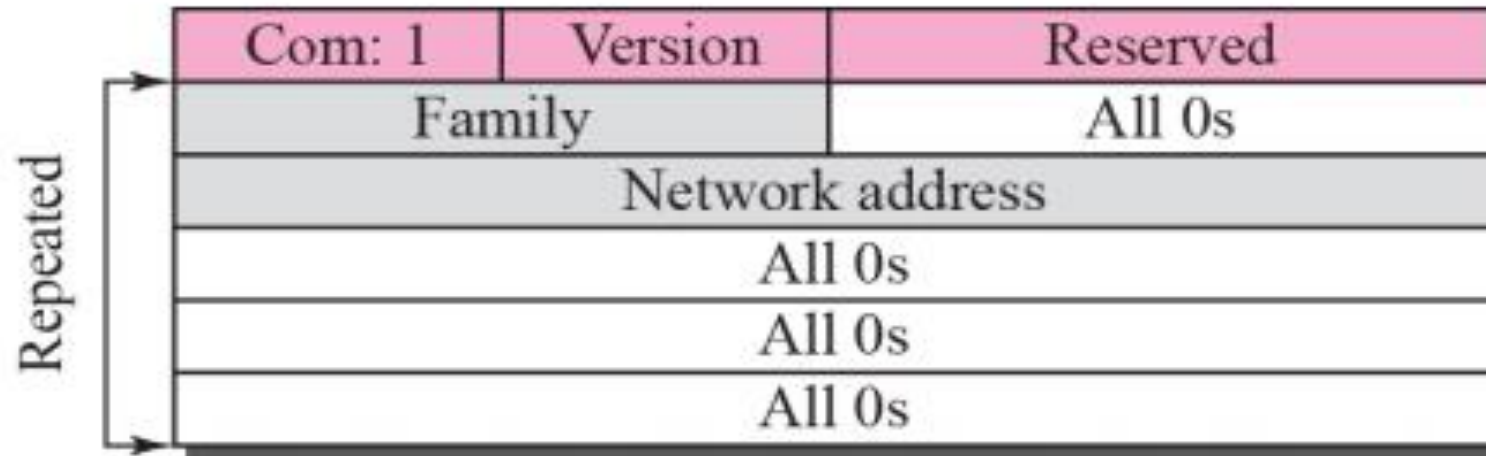
- ➡ A request message is sent by a router that has just come up or by a router that has some time-out entries
- ➡ A request can ask about specific entries or all entries and doesn't have cost

## ➡ Response

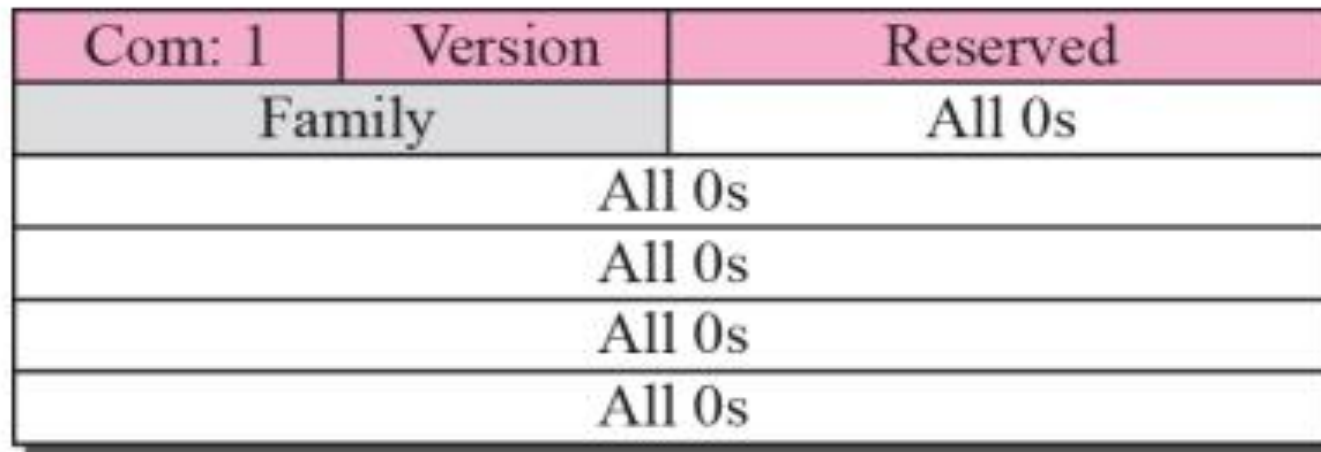
- ➡ A response can be either
  - ➡ sent only in answer to a request; contains info about destination specified in the corresponding request.
  - ➡ sent periodically, or when there is a change in the routing table
- ➡ Response is sometimes called an **update packet**.



# RIP MESSAGE FORMAT



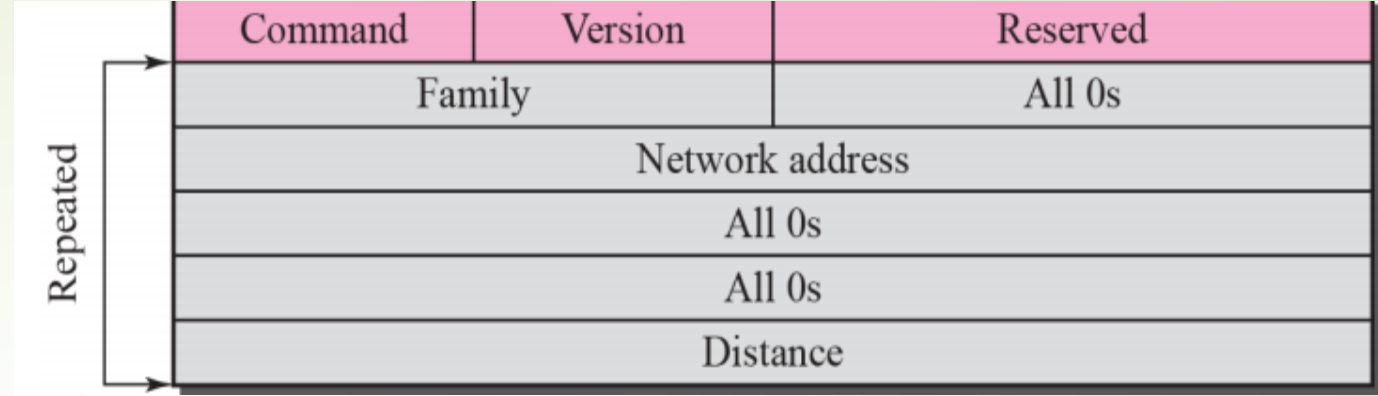
a. Request for some



b. Request for all



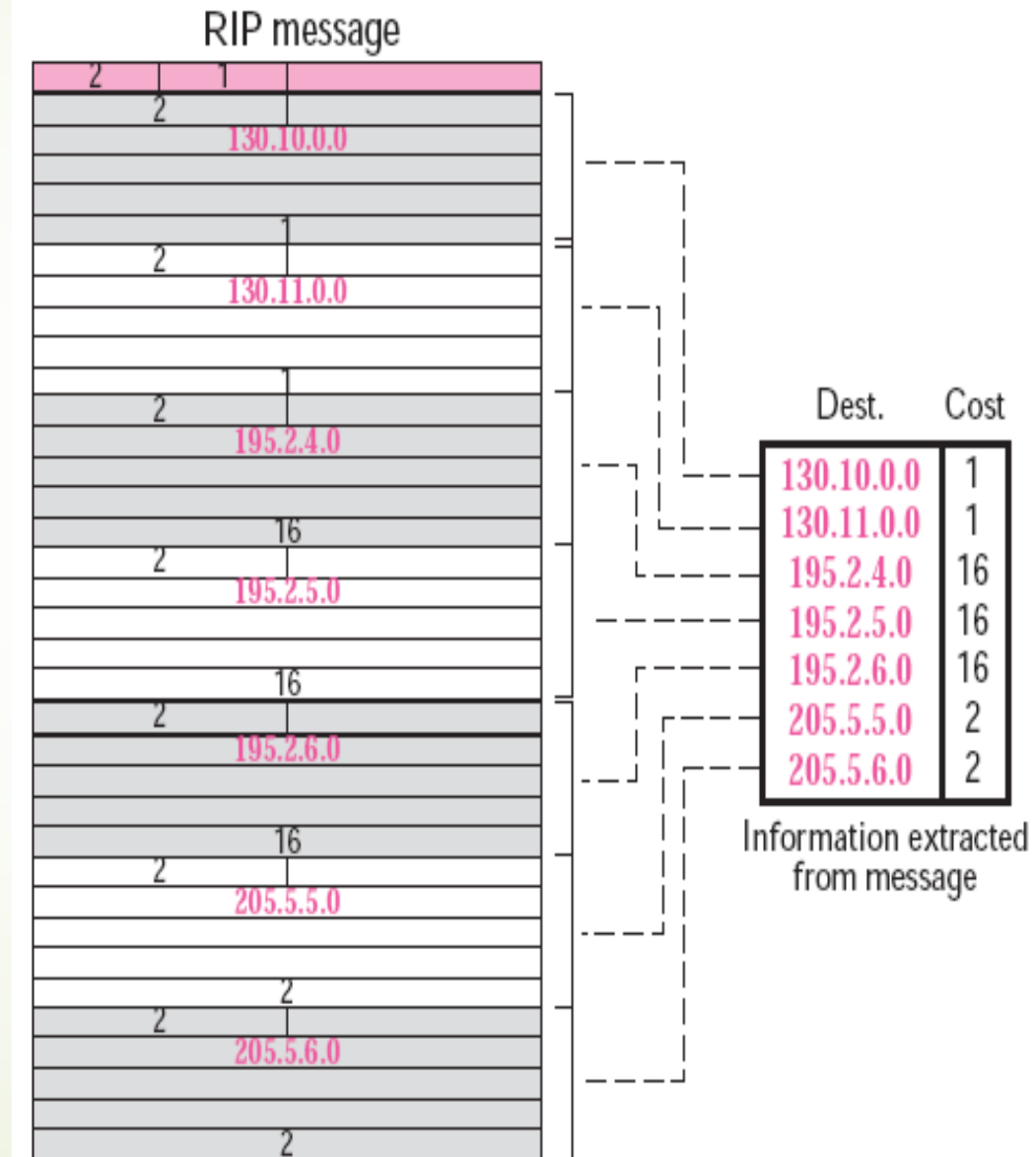
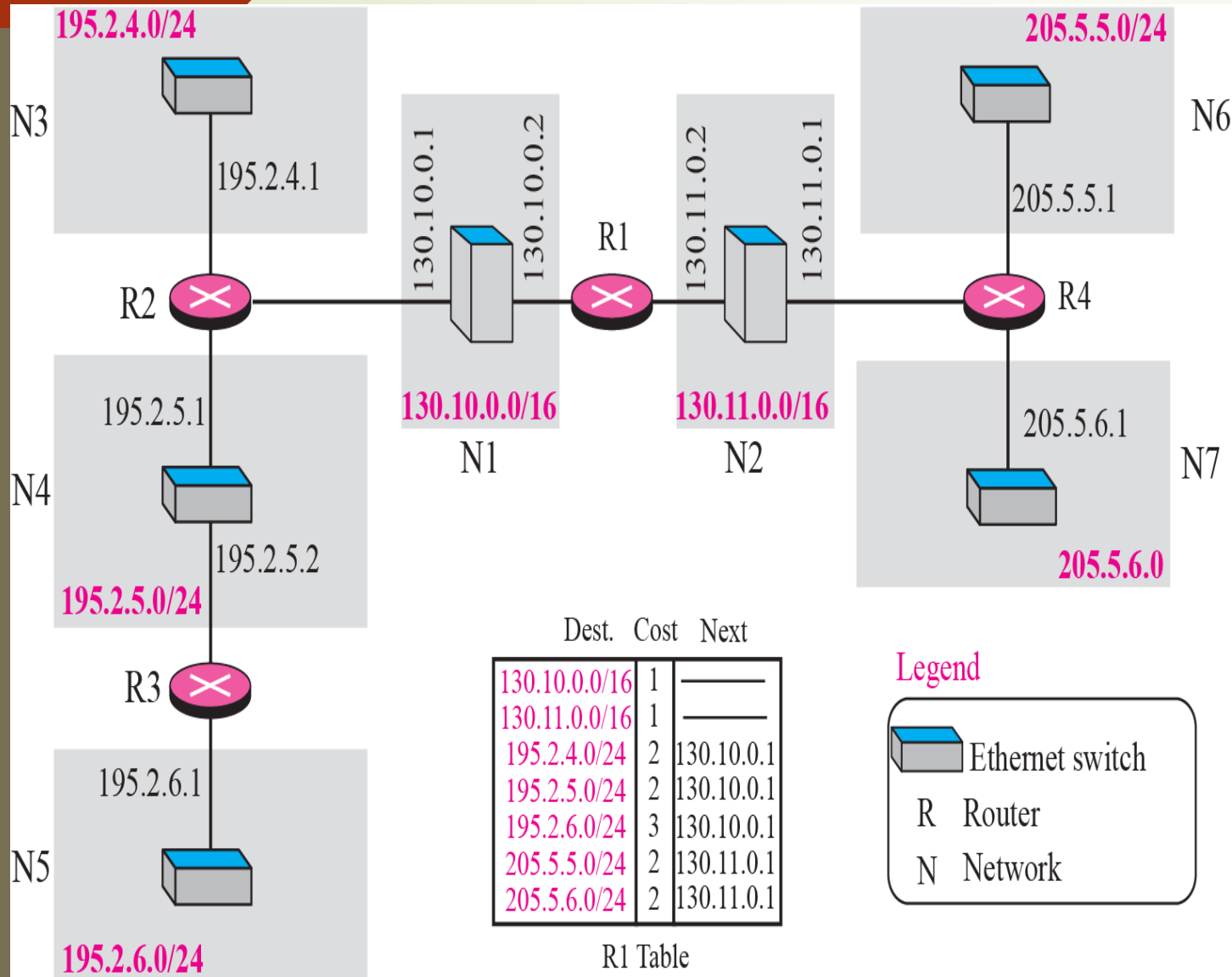
# RIP Message Format



2. **Version** : 8-bit field ; 1 or 2 (**RIPv1 or RIPv2**)
  3. **Family**: 16-bit field; Defines the family of the protocol used
    - TCP/IP, the value is 2
  4. **Network address**: defines the address of the destination network
    - RIP allocated 14 bytes, to be applicable to any protocol
    - IP uses only 4 bytes. Rest is filled with 0s.
  5. **Distance**: 32-bit field; **hop count** from the router to the destination network
- Part of the message repeated for each destination network referred as entry**

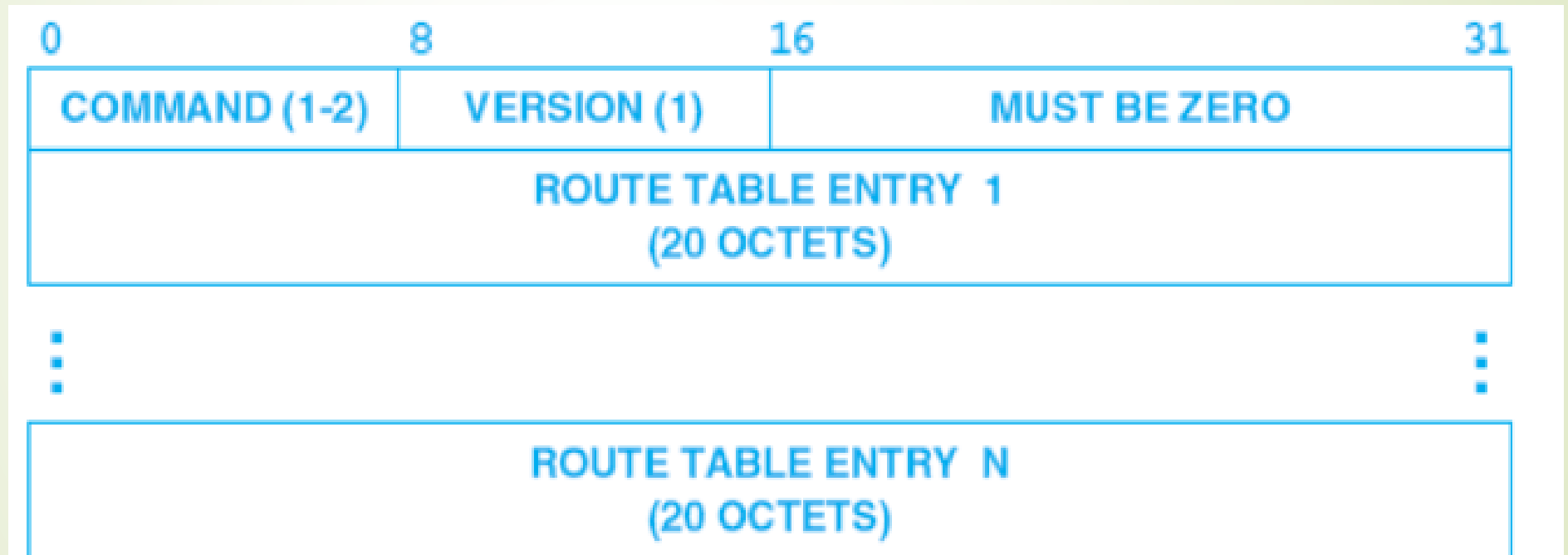
➤ Figure shows the **update message sent from router R1 to router R2.**

➤ The message is sent out of interface 130.10.0.2



# RIP For IPv6 (RIPng)

- The suffix ng stands for “next generation”;



**Figure 14.7** The overall format of a RIPng message used to carry IPv6 routing information.

# Format of Individual ROUTE TABLE ENTRY



- A route table entry with a **metric field** of **0xFF** specifies a next hop rather than a destination.
- RIPng transmits a **routing update every 30 seconds** and uses a **timeout of 180 seconds** before considering a route expired.
- RIPng also preserves the techniques of **split horizon, poison reverse, and triggered updates**.

# The Disadvantage Of Using Hop Counts

- Using RIP or RIPng as an interior gateway protocol **restricts routing to a hop-count metric.**
- Using hop counts does not always yield routes with least delay or highest capacity.
- Computing routes on the basis of minimum hop counts has the severe disadvantage
- Therefore, it may seem odd that a protocol would be designed to use a hop-count metric.

# Delay Metric (HELLO)

- **HELLO protocol is an example of an IGP** that was once deployed in the Internet and uses a routing metric other than hop count.
- Each HELLO message carried timestamp information as well as routing information, which allowed routers using HELLO to synchronize their clocks.
- HELLO used the **synchronized clocks** to find the delay on the link between each pair of routers so that each router could compute shortest delay paths to all destinations.
- The basic idea behind HELLO :
  - use a distance-vector algorithm to propagate routing information.
  - Instead of having routers report a hop count, however, HELLO reports an estimate of the delay to the destination.
  - Having synchronized clocks allows a router to estimate delay by placing a timestamp on each packet.



## Delay Metrics, Oscillation, And Route Flapping





# The Open Shortest Path First Protocol (OSPF)

- An interior gateway protocol (intra domain protocol) that uses the link-state algorithm.
- Currently, the standard version of OSPF is version 2.
- Version 2 was created for IPv4 and cannot handle IPv6.
- Write OSPFv2

# Characteristics - OSPF

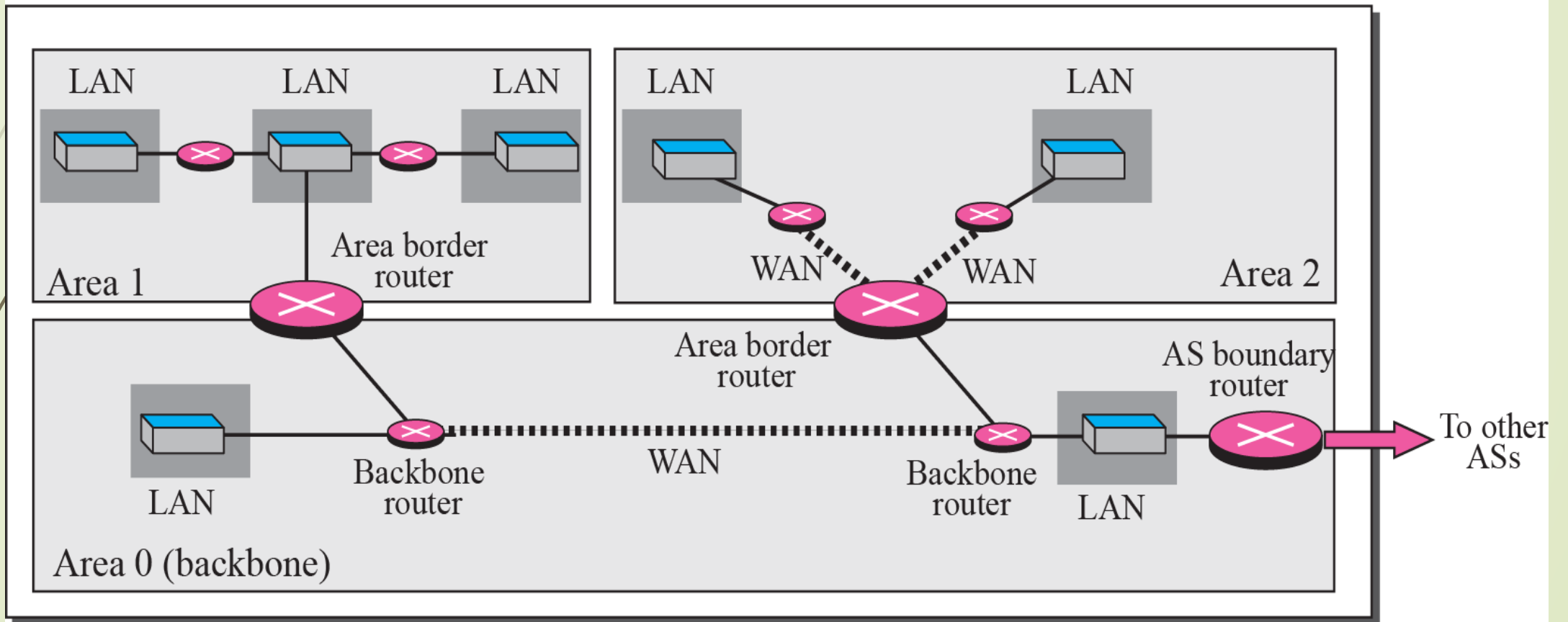
- **Open Standard:** Anyone can implement without paying license fees that encouraged many vendors to support OSPF.
- **Type Of Service Routing :** Managers can install multiple routes to a given destination, based on priority or type of service. A router running OSPF can use both the destination address and type of service when choosing a shortest route.
- **Load Balancing:** If a manager specifies multiple routes to a given destination at the same cost, OSPF distributes traffic over all routes equally.

# Characteristics - OSPF

- **Hierarchical Subdivision Into Areas:** OSPF allows a site to partition its networks and routers into subsets called areas. Each area is self-contained; knowledge of an area's topology remains hidden from other areas.
- **Support For Authentication:** OSPF supports a variety of authentication schemes that allows all exchanges between routers to be authenticated.
- **Arbitrary Granularity:** OSPF includes support for host-specific, subnet-specific, network-specific, and default routes.
- **Route Importation:** OSPF can import and disseminate routing information learned from external sites (i.e., from routers that do not use OSPF).

# AREA

## Autonomous System (AS)



# OSPFv2 Message Formats (IPv4)

- Each OSPFv2 message begins with a fixed, 24-octet header.

0	8	16	24	31
VERSION (2)	TYPE	MESSAGE LENGTH		
SOURCE ROUTER IP ADDRESS				
AREA ID				
CHECKSUM		AUTHENTICATION TYPE		
AUTHENTICATION (octets 0–3)				
AUTHENTICATION (octets 4–7)				

# OSPFv2 Message Formats (IPv4)

- **Version** : 8-bit - Version of OSPF - **2**
- **Type** : 8-bit - type of the packet (**1-5**)
- **Message length** : 16 bit - **length of the total message** including header

Type	Meaning
1	Hello (used to test reachability)
2	Database description (topology)
3	Link-status request
4	Link-status update
5	Link-status acknowledgement

- **Source router IP address**: 32-bit - **IP address of router** that sends the packet
- **Area ID** : 32-bit - Defines the **area** within which the routing takes place
- **Checksum**: used for **error detection** on the entire packet
- **Authentication type**: 16 bit - defines the authentication protocol used in this area(**0 for none, 1 for password**)
- **Authentication**: 64-bit –**actual value of authentication data**
  - If the authentication type is 0, this field is filled with 0s. If the type is 1, this field carries an eight-character password.

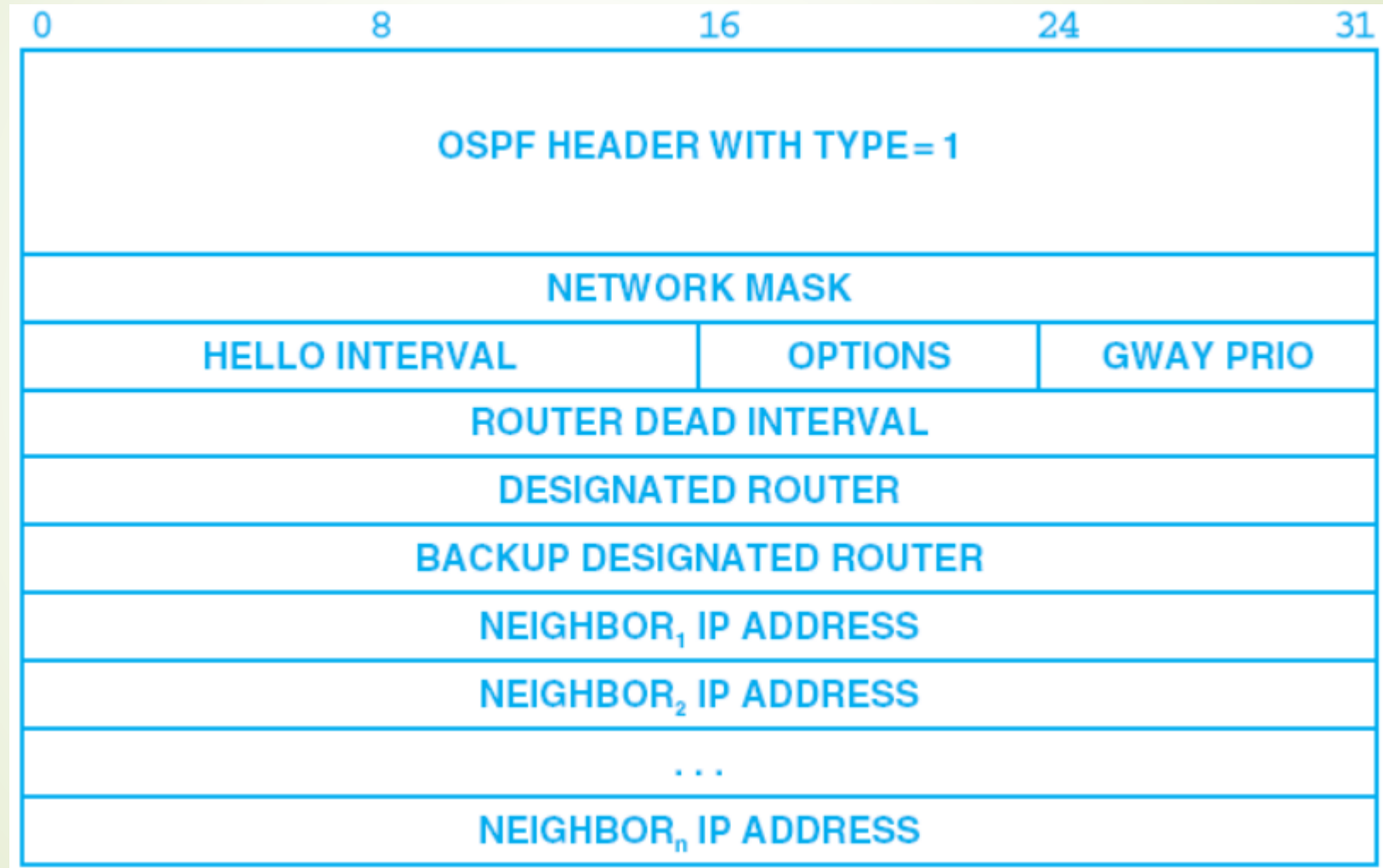


# 1. OSPFv2 Hello Message Format

- **First step** in link state routing
- OSPF uses the **hello message** to create neighborhood relationships and to test the reachability of neighbors
- Before a router can flood all of the other routers with information about its neighbors, it must **first greet its neighbors**.
- OSPFv2 sends hello messages on each link periodically
- ie: a pair of neighbor routers exchanges hello messages periodically to test reachability



# Hello message format



# Hello message format

- **Network mask:** 32-bit field defines the network mask of the network over which the hello message is sent
- **Hello interval:** 16-bit field - number of seconds between hello messages-normal period
- **Priority:** Defines the priority of the router
  - Determines the selection of the designated router.
  - The router with the highest priority is chosen as the designated router.
  - Second highest priority is chosen as the backup designated router.
  - If the value of this field is 0, it means that the router never wants to be a designated or a backup designated router.
- **Router Dead interval:** defines the number of seconds that must pass before a router assumes neighbour is dead

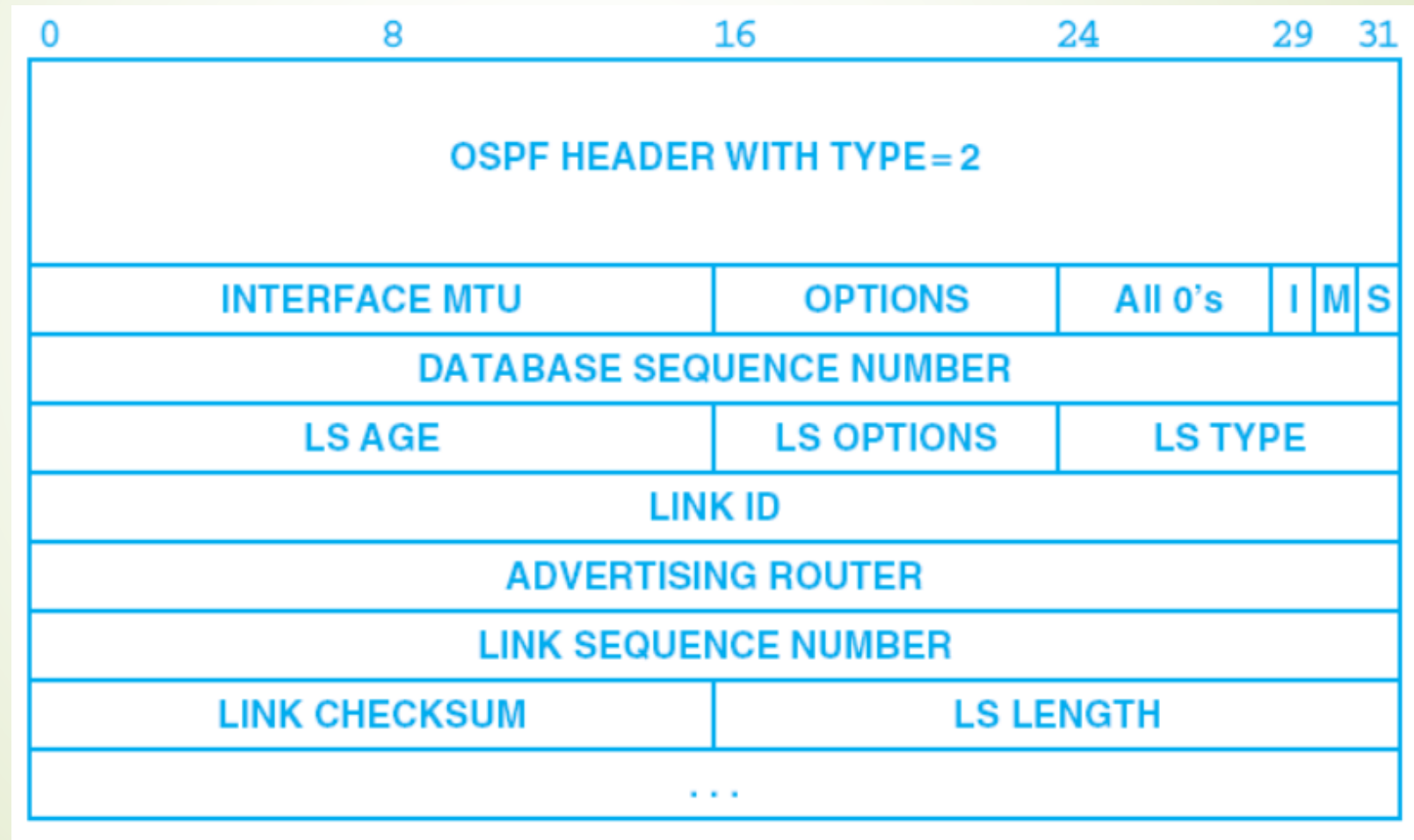
# Hello packet



- **Designated router** : 32-bit field - IP address of the designated router for the network over which the message is sent.
- **Backup designated router IP address** : 32-bit field - IP address of the backup designated router
- **Neighbor IP address.** : 32-bit field that defines the IP addresses of the routers that have agreed to be the neighbors of the sending router.
  - It is a current list of all the neighbors from which the sending router has received the hello messages.

## 2. OSPFv2 Database Description message

- When a router connected to the system for the first time or after a failure, it **needs the complete database about network immediately**
- It **send hello packets to greet its neighbors**
- If this is the first time that the neighbors hear from the router, they send a hello message followed by a **database description message**.
- Routers exchange OSPFv2 **database description message to initialize their network topology database**
- Does not contain complete database information; it only gives an outline, the title of each line in the database.
- In the exchange, one router serves as a master, while the other is a slave.

# OSPFv2 Database Description Message Format



- 
- 
- **Interface MTU** : gives the size of the largest IP datagram that can be transmitted over the interface without fragmentation.
  - **I flag** : **Initialization flag**. Set to 1 if the message is the first message
  - **M flag** : **More flag**. Set to 1 if it is not the last message, means additional messages follow
  - **S flag** : **master/slave bit**, Indicates the origin of the message; master(1) or slave (0)
  - **database sequence number** : sequence number of the message
    - Numbers the messages sequentially so that receiver can tell if one is missing
    - Initial message : random integer R
    - Subsequent messages : sequential integers starting at R



➤ The fields from *LS AGE* through *LS LENGTH* describe one link in the network topology; they are repeated for each link.

➤ **LS TYPE** describes the type of a link.

➤ **LINK ID** gives an identification for the link which can be the IP address of a router or a network, depending on the link type

➤ **LS AGE** helps order messages — it gives The time in seconds since the message was established.


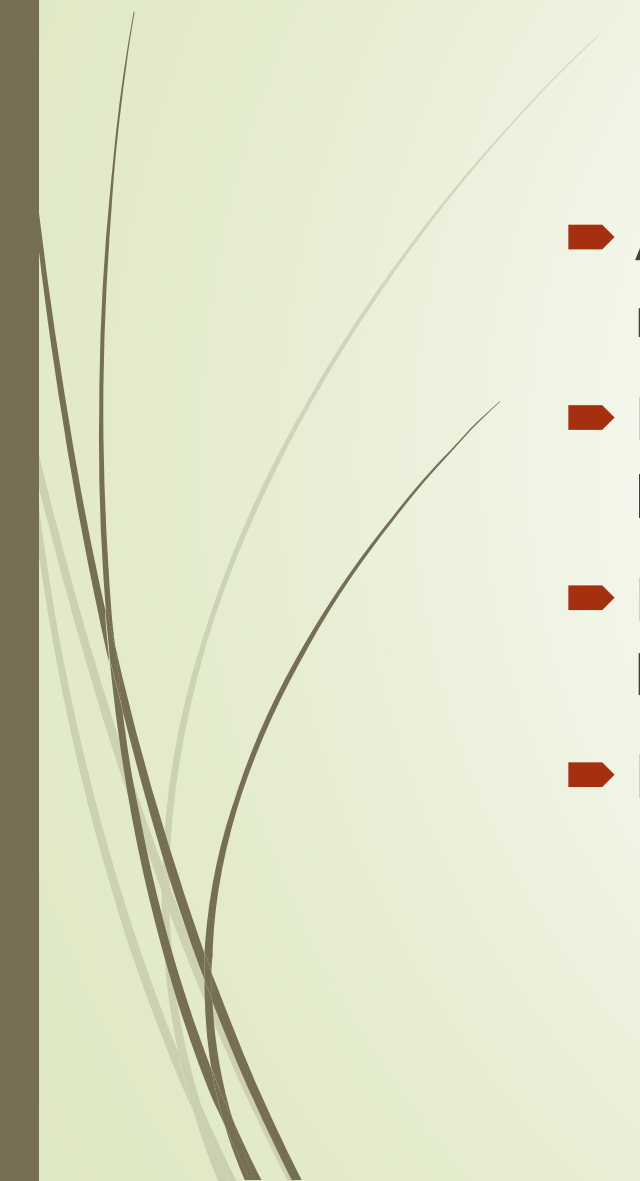
➤ When a router creates the message, the value of this field is 0.

➤ When each successive router forwards this message, it estimates the transit time and adds it to the cumulative value of this field

LS Type	Meaning
1	Router link
2	Network link
3	Summary link (IP network)
4	Summary link (link to border router)
5	External link (link to another site)

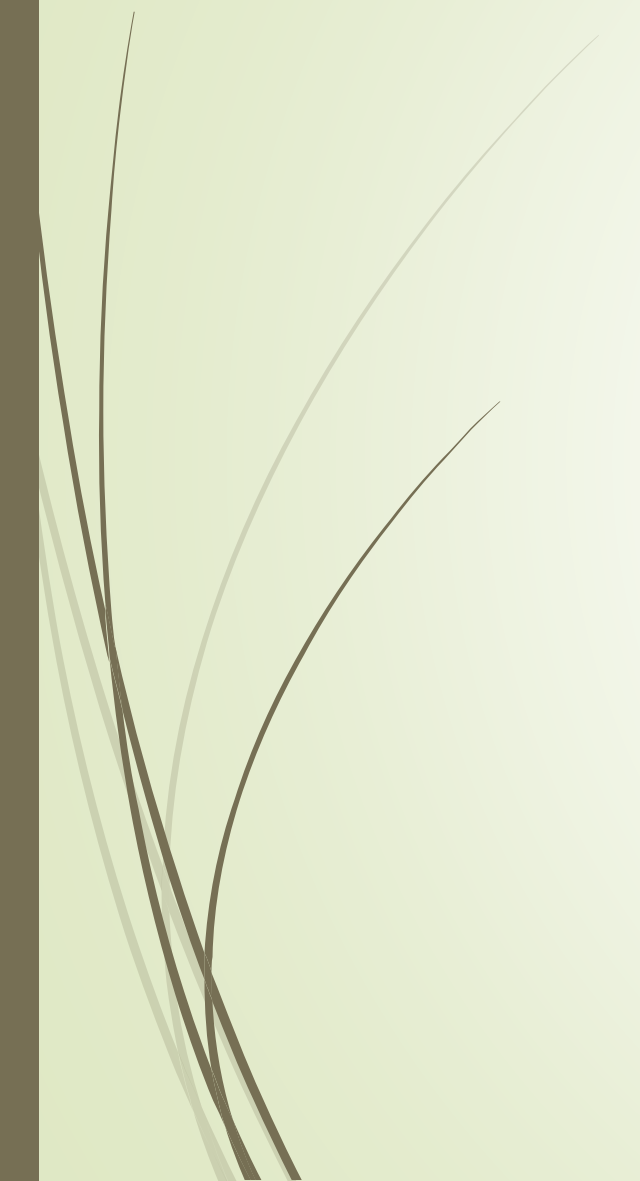
Link state type	Link state ID
Router link	IP address of the router
Network link	IP address of the designated router
Summary link to network	Address of the network
Summary link to AS boundary	IP address of the boundary router
External link	Address of the network



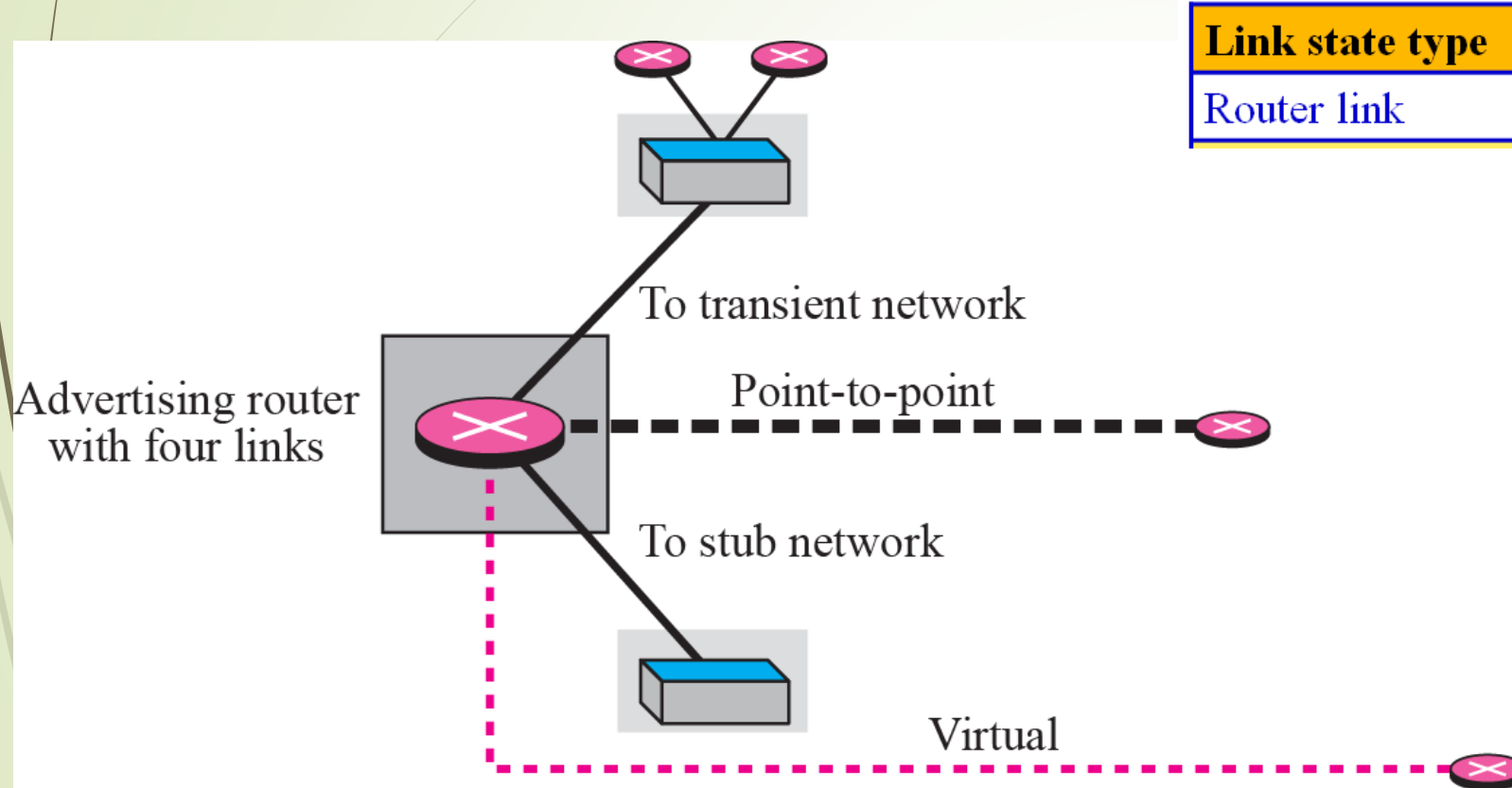
- 
- 
- **Advertising router:** The **IP address of the router** advertising this message
  - **Link sequence number:** A sequence number assigned to each link to ensure that messages are not missed or received out of order.
  - **Link checksum** provides further assurance that the link information has not been corrupted.
  - **LS Length:** defines the **length of the whole packet** in bytes



# LS TYPES

- 
1. Router link
  2. Network link
  3. Summary link to network
  4. Summary link to AS boundary router
  5. External link

# Router Link



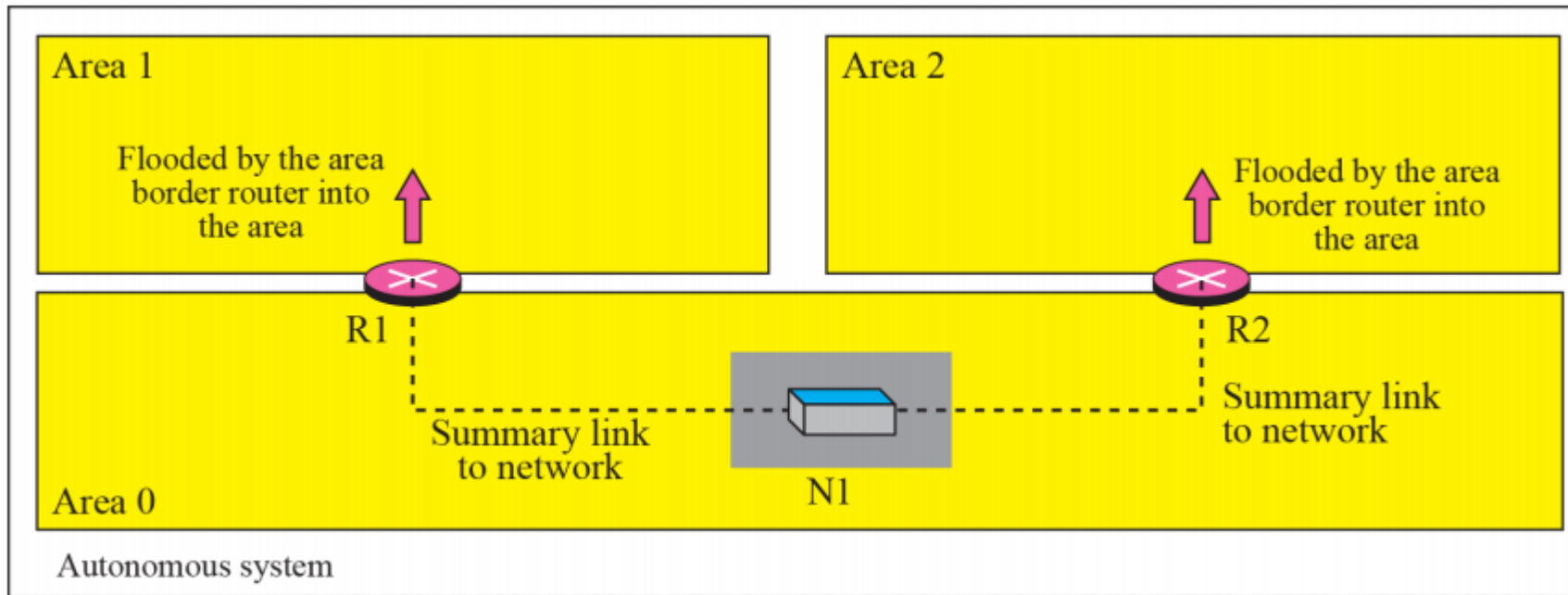
Link state type	Link state ID
Router link	IP address of the router

# Network Link



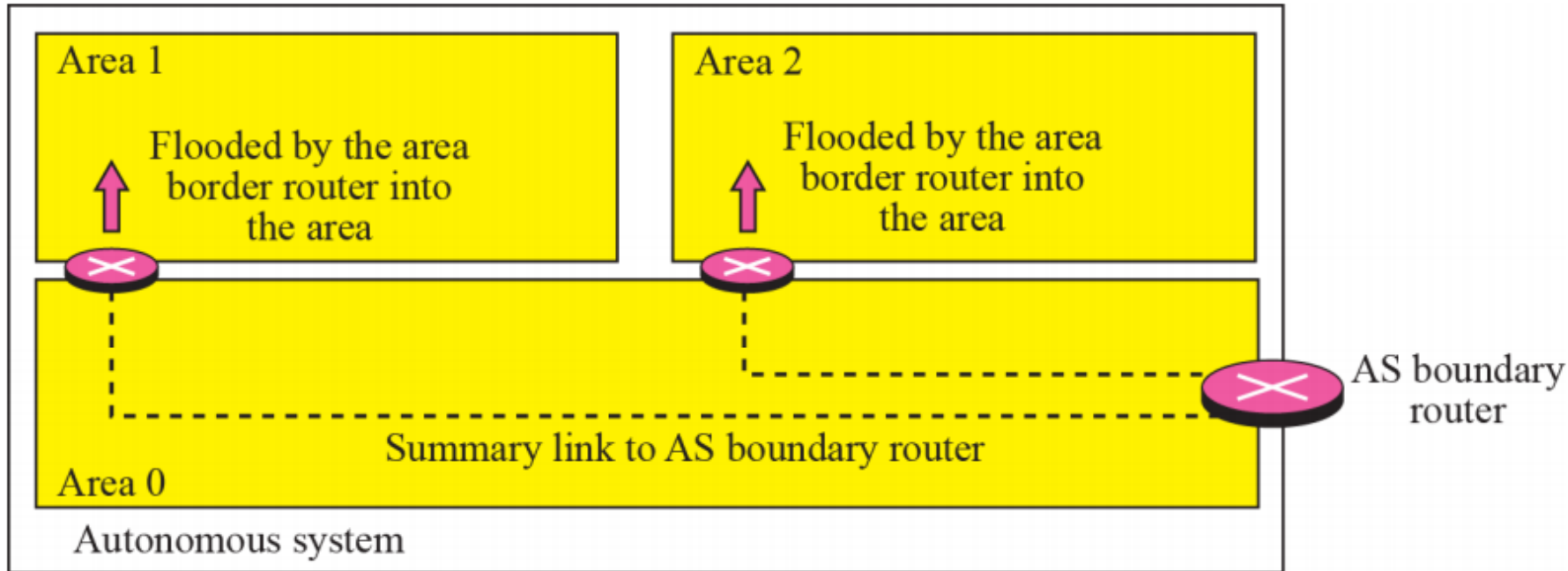
Link state type	Link state ID
Router link	IP address of the router
Network link	IP address of the designated router

# Summary link to network



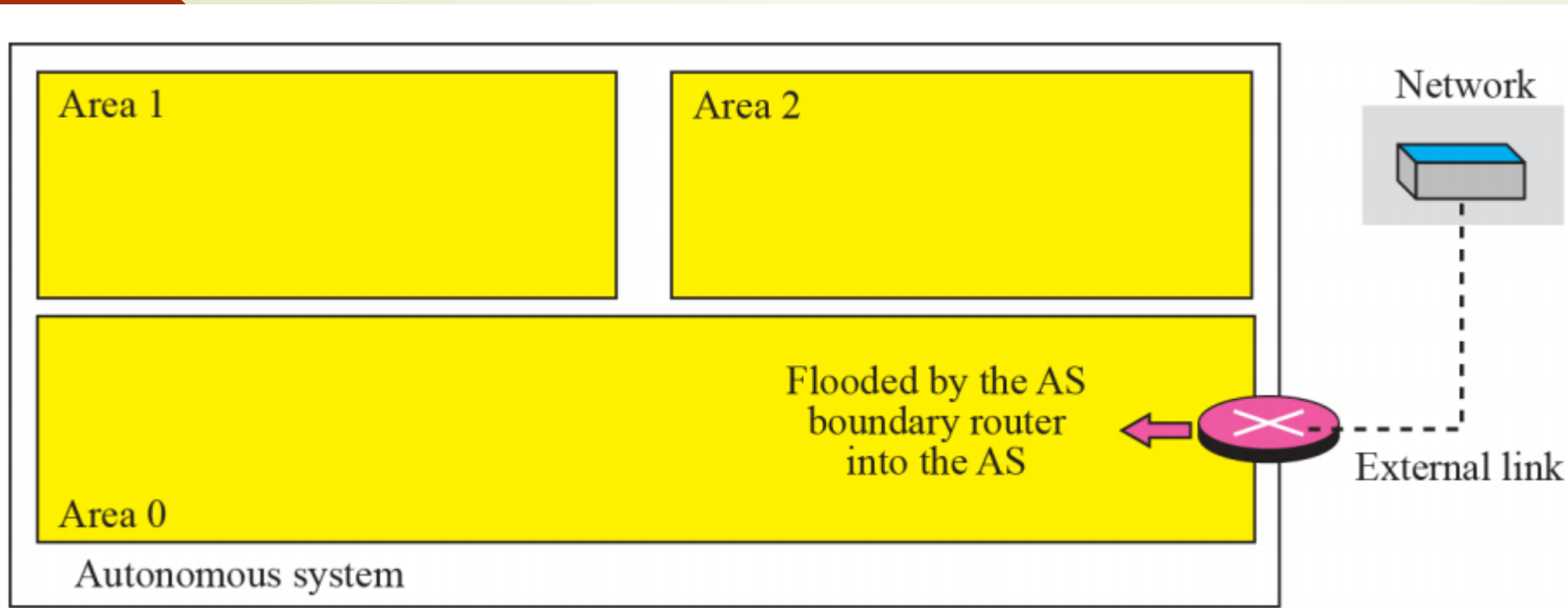
Link state type	Link state ID
Router link	IP address of the router
Network link	IP address of the designated router
Summary link to network	Address of the network

# Summary link to AS boundary router



Link state type	Link state ID
Router link	IP address of the router
Network link	IP address of the designated router
Summary link to network	Address of the network
Summary link to AS boundary	IP address of the boundary router

# External link



Link state type	Link state ID
Router link	IP address of the router
Network link	IP address of the designated router
Summary link to network	Address of the network
Summary link to AS boundary	IP address of the boundary router
External link	Address of the network

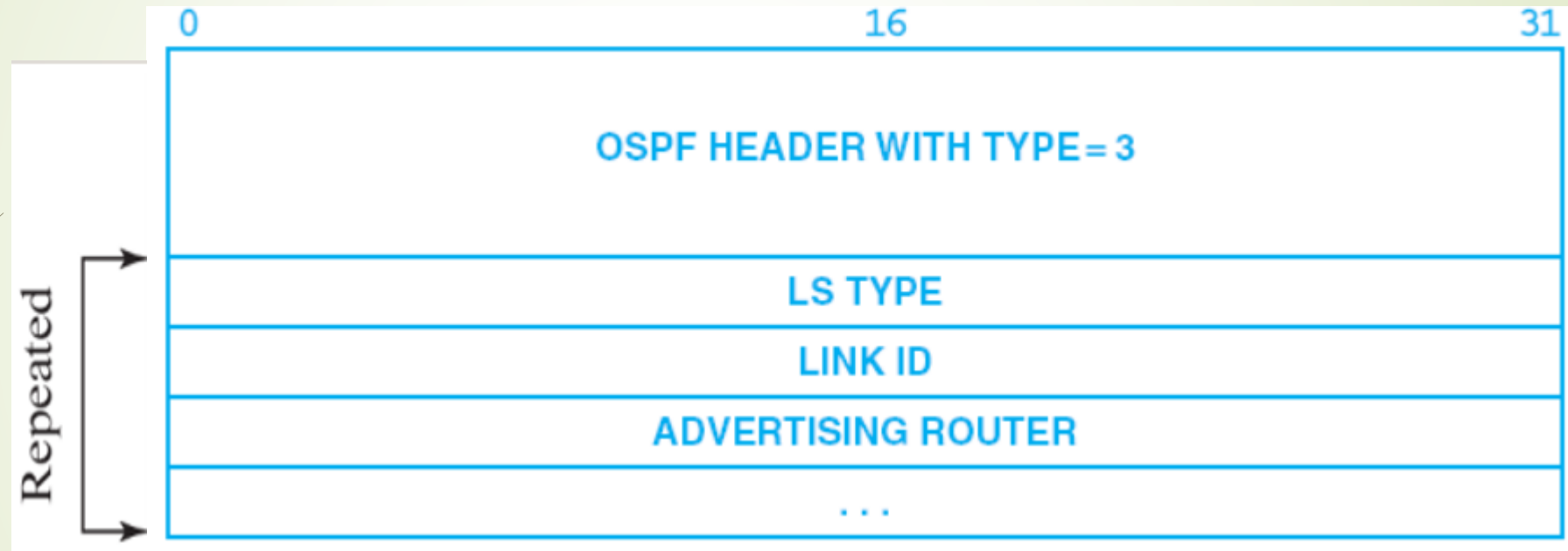




# OSPFv2 Link-Status Request Message Format

- It can be used by a newly connected router to request more information about some routes after receiving the database description packet.
- This is a packet that is sent by a router to a other routers that needs information about a specific route or routes.
- The neighbor responds with the most current information it has about the links in the request message.
- It is answered with a link state update packet.

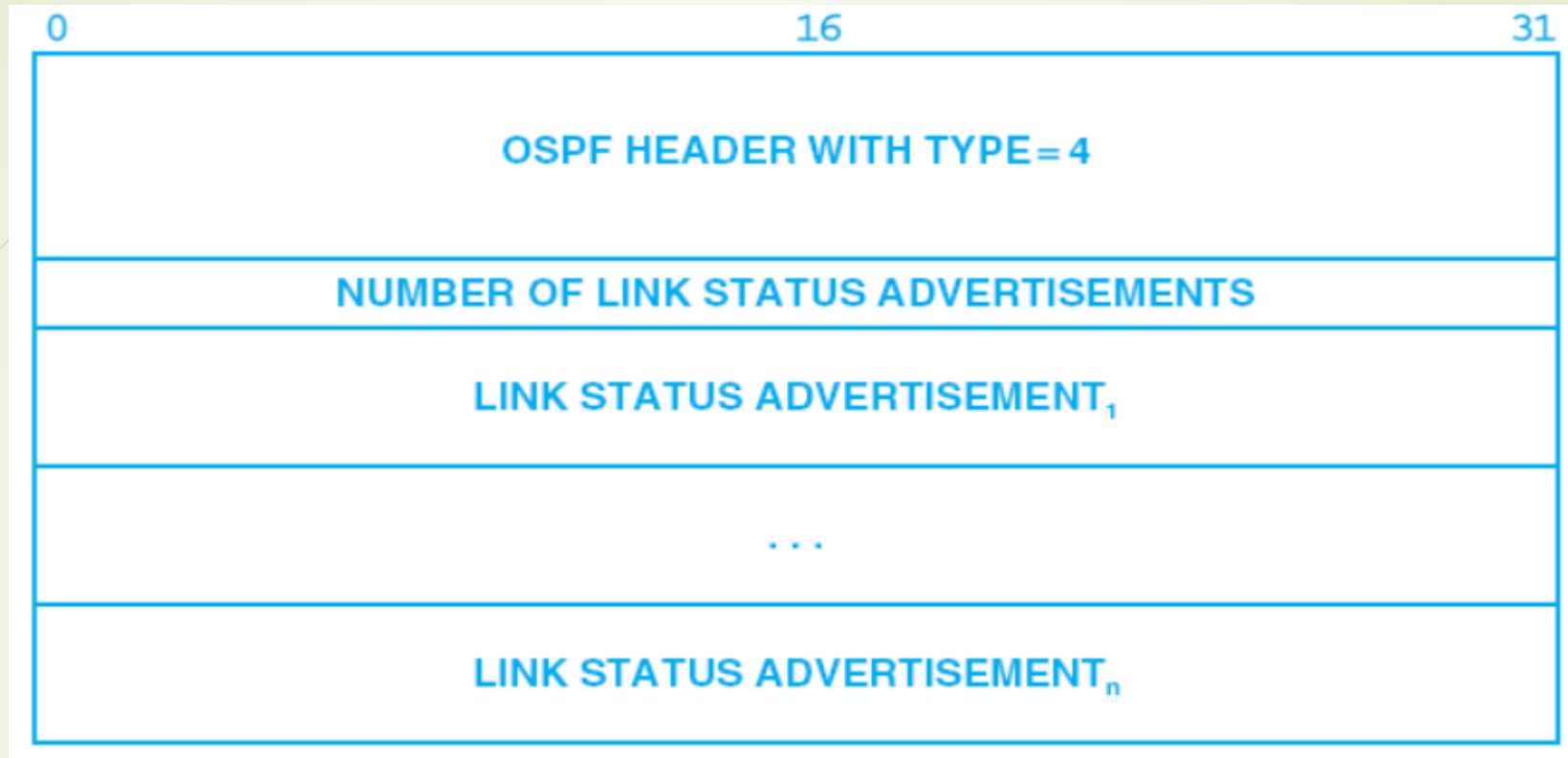
# OSPFv2 Link-Status Request Message Format



The three fields shown in the figure are repeated for each link about which status is requested.

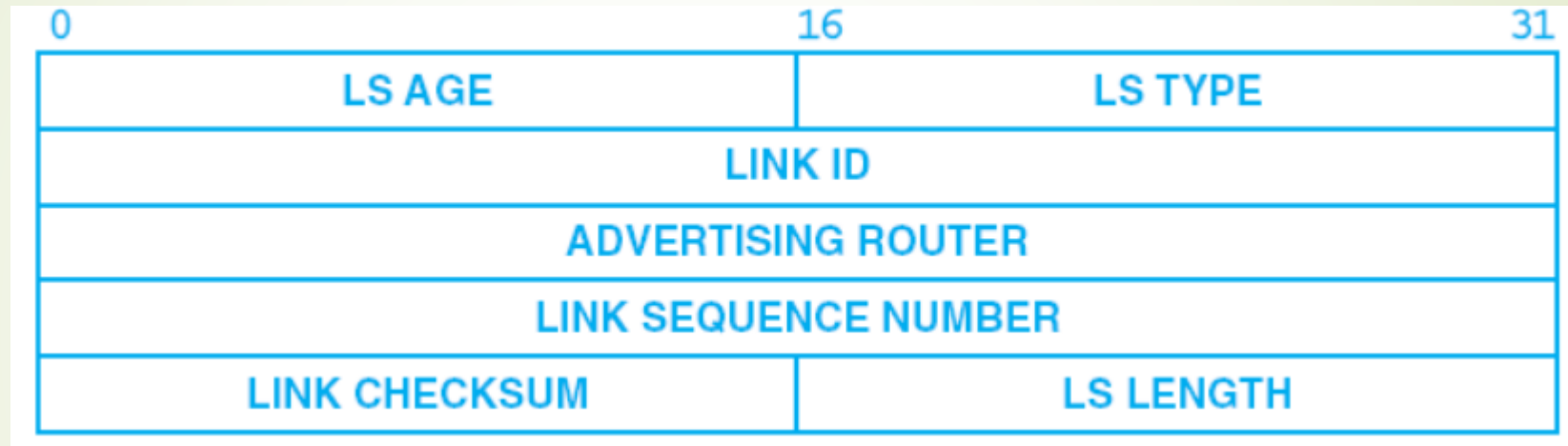
# OSPFv2 Link-Status Update Message Format

- **Heart of the OSPF operation**
- **Used by a router to advertise the states of its links**
- Because uses a link-state algorithm, routers must periodically broadcast messages that specify the status of directly-connected links.
- To do so, routers use a type 4 OSPFv2 message that is named a *link-status update*.
- **A router sends such a message to broadcast information about its directly connected links to all other routers.**
- Each update message consists of a count of advertisements followed by a list of advertisements



- Each *link-status advertisement* (LSA) has a format that specifies information about the network being advertised.

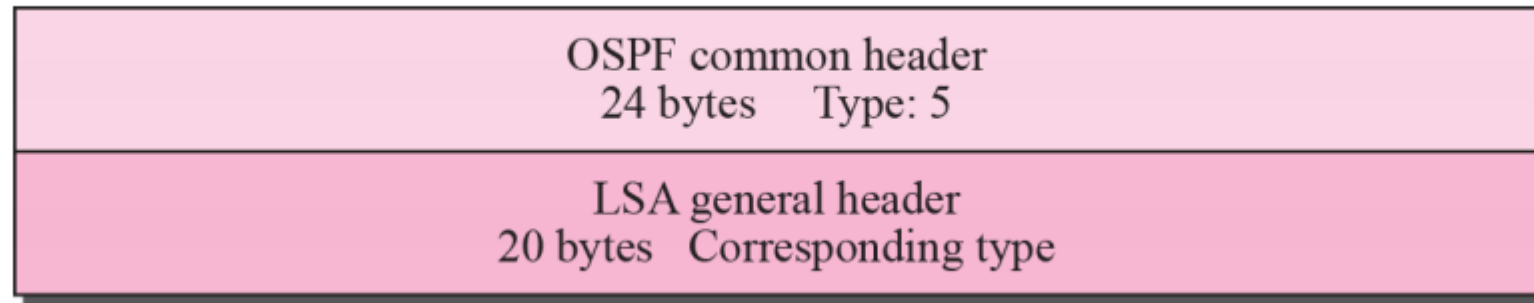
# Format of an OSPFv2 *Link-Status Advertisement (LSA)*



- **LS TYPE** field in the link-status advertisement specifies which of the links has been used.
- Thus, a router that receives a link-status update message knows exactly which of the described destinations lie inside the site and which are external.

## 5. Link status acknowledgement packet

- Every router acknowledge the receipt of the link state update packet



# OSPFv3 to support IPv6

- The protocol still uses the **link-state approach**
- OSPFv3 combines and generalizes many of the facilities and features that have been defined for OSPFv2.
- **OSPFv2 used a 32-bit IP address** to identify a router; **OSPFv3 uses a 32-bit router ID**.
- **Area identifiers** remain at **32 bits**,
- **OSPFv3 removes all authentication** from individual messages
- Messages must be changed to accommodate IPv6 addresses.
- Each occurrence of an IPv4 address with an IPv6 address would make messages too large.
- OSPFv3 minimizes the number of IPv6 addresses carried in a message and **substitutes 32-bit identifiers** for any identifier that does not need to be an IPv6 address.
- OSPFv3 supports **router LSAs, link LSAs, interarea prefix LSAs, inter-area router LSAs, AS-external LSAs, intra-area prefix LSAs, and Not So Stubby Area (NSSA) LSAs**.
- Support large autonomous systems that have a complex topology and complex rules for areas.



# OSPFv3 Message Formats

- Begins with a fixed, **16-octet header**
- **Fixed header is smaller** than the OSPFv2 header because **authentication information has been removed.**

0	8	16	24	31
VERSION (3)	TYPE	MESSAGE LENGTH		
SOURCE ROUTER ID				
AREA ID				
CHECKSUM		INSTANCE ID	0	



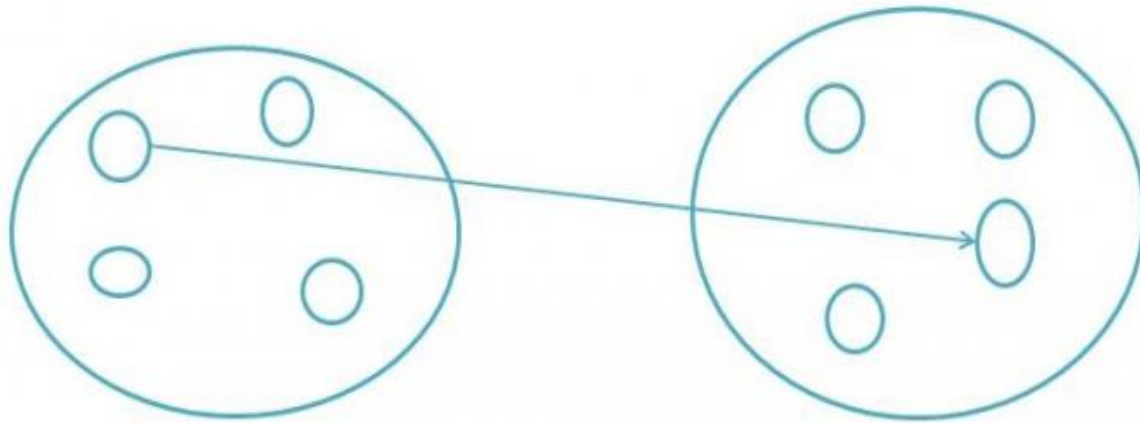
# **Internet Multicasting**

# Introduction

- **IP datagram forwarding and delivery** can be : unicast or multipoint delivery.
- Multipoint delivery : **multicast** or **broadcast**.
- **Unicast:** from one source to one destination i.e. **One-to-One**
- **Broadcast:** from one source to all possible destinations i.e. **One-to-All**
- **Multicast:** from one source to multiple destinations i.e. **One-to-Many**

**Network A**  
11.1.2.2

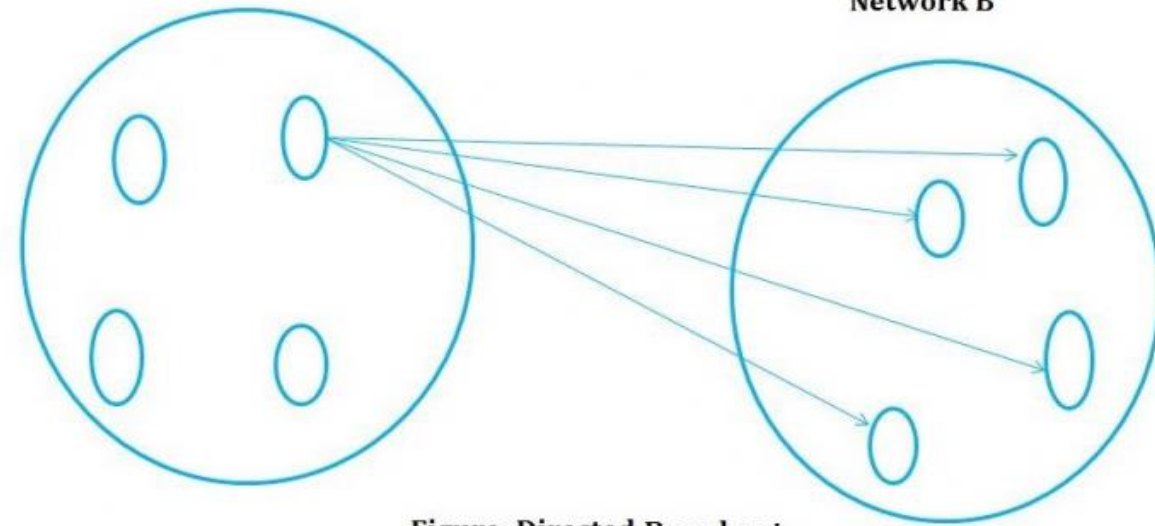
**Network B**  
20.12.4.3



**Figure: Unicast**

**Network A**

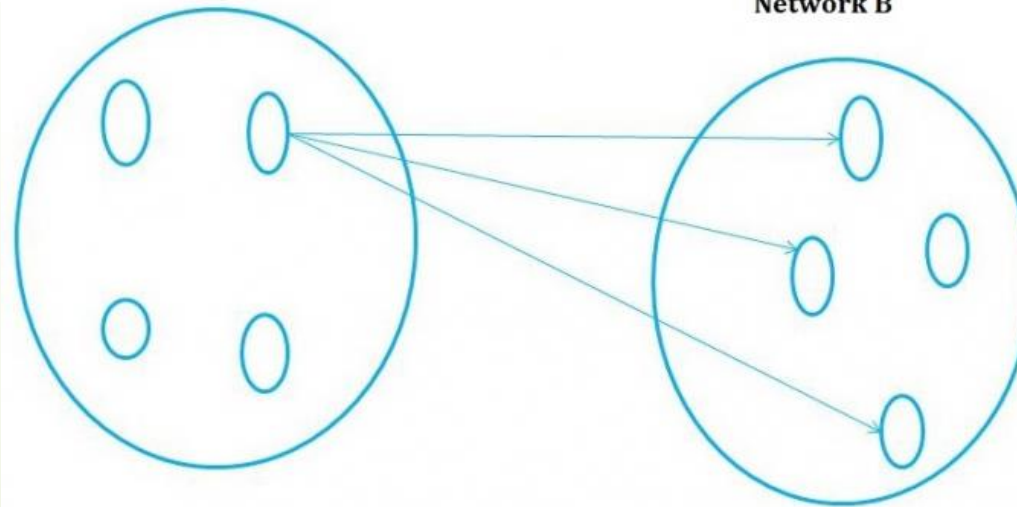
**Network B**



**Figure: Directed Broadcast**

**Network A**

**Network B**



**Figure: Multicast**

# Hardware Broadcast

- 1) The hardware **sends a single copy of a packet** - arranges for each attached computer to receive a copy.
- 2) The networking equipment implements broadcast - forwarding an independent copy of a broadcast packet to each individual computer.
- Most - **a special, reserved destination address** called a ***broadcast address***.
- To specify **broadcast delivery**, create a frame where the destination address field contains the broadcast address.
- Eg: **Ethernet uses the all 1s hardware address as a broadcast address**; each computer attached to an Ethernet network accepts frames sent to the broadcast address

# Hardware Multicast

- When a host want to use hardware multicast, they choose one particular **multicast address** to use for communication.
- The application running on a computer must ask the operating system to **configure the NIC to recognize the multicast address** that has been selected.
- After the hardware has been configured, the computer will receive a copy of any packet sent to the multicast address.
- The term **multicast group** to denote the **set of computers that are listening to a particular multicast address**.
- Eg: If applications on six computers are listening to a particular multicast address, the multicast group is said to have six members.
- A multicast address identifies an arbitrary subset of computers, and members of the group can change at any time.



# Ethernet Multicast

- An example of **hardware multicasting**.
- **Determined by low-order bit of high-order byte**
- Example in dotted hexadecimal:

**$01.00.00.00.00.00_{16}$**

- Eg : Suppose the driver configures the Ethernet multicast address:  **$01-5E-00-00-00-01_{16}$**
- After configuration, the interface hardware will accept any packet sent to the **computer's unicast MAC address, the broadcast MAC address, or the example multicast MAC address** (the hardware will continue to ignore packets sent to other multicast addresses).





# Conceptual Building Blocks Of Internet Multicast

- A multicast addressing scheme
- An effective notification and delivery mechanism
- An efficient internetwork forwarding facility

# IP Multicast Scheme

- **IP multicasting mechanism** includes all three building blocks.
- It
  - defines **multicast addressing** for both **IPv4** and **IPv6**,
  - provides a **mechanism that allows hosts to join and leave IP multicast groups**,
  - specifies **how multicast datagrams are transferred** across individual hardware networks,
  - provides a **set of protocols** routers can use to **exchange multicast routing information** and
  - **construct forwarding tables** for multicast groups.

# IP Multicast Scheme - Characteristics

## 1. One IP Multicast Address Per Group

- **IP multicast group** : A subset of computers listening to a given IP multicast address
- Each IP multicast group assigned a unique IP multicast address - **Group address**
  - permanently assigned by the Internet authority,
  - temporary, available for private use.

## 2. Number Of Groups

- Up to  $2^{28}$  simultaneous multicast groups in IPv4
- IPv6 provides many more

# IP Multicast Scheme - Characteristics

## 3. Dynamic group membership

- Host can join or leave an IP multicast group at any time.
- Host may be a member of an arbitrary number of multicast groups simultaneously.

## 4. Use Of Hardware

- IP uses hardware multicast to deliver an IP multicast datagram on the network where available

## 5. Internetwork Forwarding

- Special multicast routers/multicast capability is added to conventional routers.

# IP Multicast Scheme - Characteristics

## 6. Delivery Semantics

- Uses best-effort delivery semantics same as other IP datagram delivery, multicast datagrams can be lost, delayed, duplicated, or delivered out of order.

## 7. Membership And Transmission

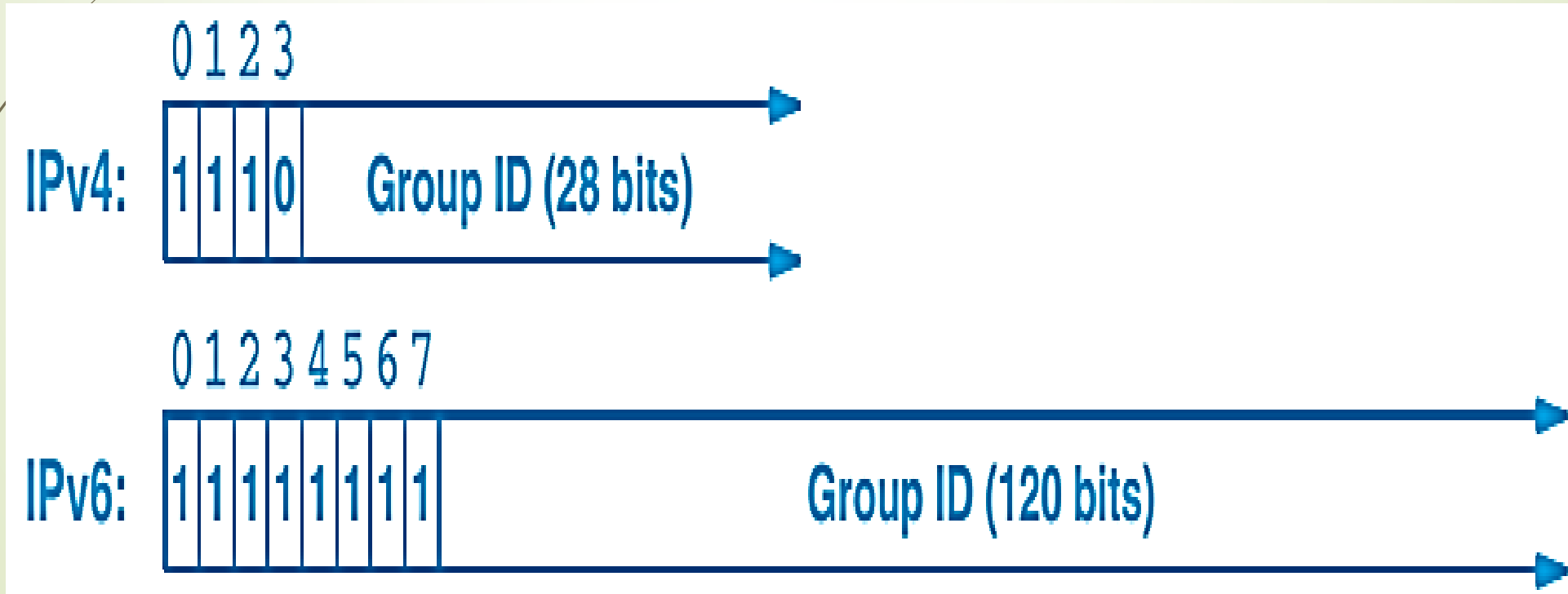
- Arbitrary sender (does not need to be a group member)
- An arbitrary host may send datagrams to any multicast group.
- Group membership is only used to determine whether the host receives datagrams sent to the group.

# IPv4 And IPv6 Multicast Addresses

- IP multicast address : Two types
  - Permanent : **Well-known** (address reserved for specific protocol)
  - Temporary : **Transient** (allocated as needed and discarded)

# IPv4 And IPv6 Multicast Addresses

- ➡ IPv4 reserves **class D** addresses for multicast :
  - ➡ the first 4 bits contain **1110** and identify the address as a multicast address.
- ➡ In IPv6, a multicast address has the first **8 bits set to 1**






# IPv4 Multicast Address Space

- IPv4 multicast addresses range from  
**224.0.0.0 through 239.255.255.255**
- Lowest address, 224.0.0.0, is reserved; it cannot be assigned to any group.
- Addresses up through 224.0.0.255 are restricted to a single network
- Addresses 239.0.0.0 through 239.255.255.255 are restricted to one organization

# The division of the IPv4 multicast address space according to scope

Address Range	Meaning
224.0.0.0	Base Address (Reserved)
224.0.0.1 – 224.0.0.255	Scope restricted to one network
224.0.1.0 – 238.255.255.255	Scope is global across the Internet
239.0.0.0 – 239.255.255.255	Scope restricted to one organization



## **Example of specific IPv4 multicast address; scope restricted to one network.**

<b>Address</b>	<b>Assigned Purpose</b>
<b>224.0.0.1</b>	<b>All Systems on this Subnet</b>
<b>224.0.0.2</b>	<b>All Routers on this Subnet</b>
<b>224.0.0.5</b>	<b>OSPFv2 All Routers</b>
<b>224.0.0.6</b>	<b>OSPFv2 Designated Routers</b>
<b>224.0.0.9</b>	<b>RIP2 Routers</b>
<b>224.0.0.12</b>	<b>DHCP Server / Relay Agent</b>
<b>224.0.0.22</b>	<b>IGMP</b>

# IPv6 Multicast Address Space

- The **first octet** of an IPv6 multicast address **contains all 1s**.
- IPv6 uses the **second octet** of the address to **specify the scope**.

Second Octet	Meaning
0x?0	Reserved
0x?1	Scope is restricted to a computer (loopback)
0x?2	Scope is restricted to the local network
0x?3	Scope is equivalent to IPv4 local scope
0x?4	Scope is administratively configured
0x?5	Scope is restricted to a single site
0x?8	Scope is restricted to a single organization
0x?E	Scope is global across the Internet

# IPv6 Multicast Address Space

- In the figure, constants starting with 0x are hexadecimal.
- The question mark denotes an arbitrary nibble.
- Thus, 0x?1 refers to 0x01, 0x11, 0x21... 0xF1.
- **Network Time Protocol (NTP)** has been assigned the multicast group ID **0x101**.
- FF02::101 (all NTP servers on a single network)
- FF08::101 (all NTP servers in an organization)

# Multicast Address Semantics

- A multicast address **can only be used as a destination address**.
- If a router finds a multicast address in the source address field of a datagram the router drops the datagram.
- **No ICMP error messages** can be generated about multicast datagrams.
- A **ping** sent to a multicast address **will go unanswered**.



# Mapping An IP Multicast Address To An Ethernet Multicast Address

- IANA owns the Ethernet address prefix **0x01005E**
- To map an IPv4 multicast address to the corresponding Ethernet multicast address, place the **low-order 23 bits** of the IPv4 multicast address into the **low-order 23 bits of the special Ethernet multicast address 01-00-5E-00-00-00<sub>16</sub>**.
- For example, the IPv4 multicast address **224.0.0.2** becomes Ethernet multicast address **01-00-5E-00-00-02<sub>16</sub>**.



# Mapping An IP Multicast Address To An Ethernet Multicast Address

- IPv6 uses the Ethernet prefix **0x3333** and selects 32 bits of the IP multicast group ID
- To map an IPv6 multicast address to the corresponding Ethernet multicast address, place the **low-order 32 bits of the IPv6 multicast address** into the **low-order 32 bits of the special Ethernet multicast address**  
**33-33-00-00-00-00<sub>16</sub>**.
- For example, IPv6 multicast address **FF02:09:09:1949::DC::1** would map to the Ethernet MAC address **33-33-00-DC-00-01**.

# Hosts And Multicast Delivery

- IP multicasting can be used on a **single physical network** or **throughout an internet**.

1. A host can send directly to a destination host by placing the datagram in a frame and using a hardware multicast address.

2. **Multicast routers** are needed to forward copies of multicast datagrams across multiple networks to all hosts participating in a multicast group.

- A host does not need to install a route to a multicast router, nor does IP software use a default route to reach a multicast router
- **Multicast routers listen for all IP multicast transmissions**; if a multicast router is present on the network, it will receive the datagram and forward it on to another network if necessary

# Multicast Scope

- The term multicast scope is used for two concepts. We use the term to clarify the set of hosts that are listening to a given multicast group or to specify a property of a multicast address.
- IP uses two techniques to control multicast scope. The first technique relies on the datagram's hop limit field to control its range. By setting the hop limit to a small value, a host can limit the distance the datagram will be forwarded.
- For example, the standard specifies that control messages, which are used for communication between a host and a router on the same network, must have a hop limit of 1.
- The second technique, which is known as administrative scoping, consists of choosing multicast addresses that have limited scope. According to the standard, routers in the Internet are forbidden from forwarding any datagram that has an address chosen from the restricted space.
- Thus, to prevent multicast communication among group members from accidentally reaching outsiders, an organization can assign the group an address that has local scope.