第19章 互联网的操作

(1) 因特网路由协议

南京大学计算机系　黄皓教授

2007年10月9日 星期二

|

2007年10月12日 星期五

Reference

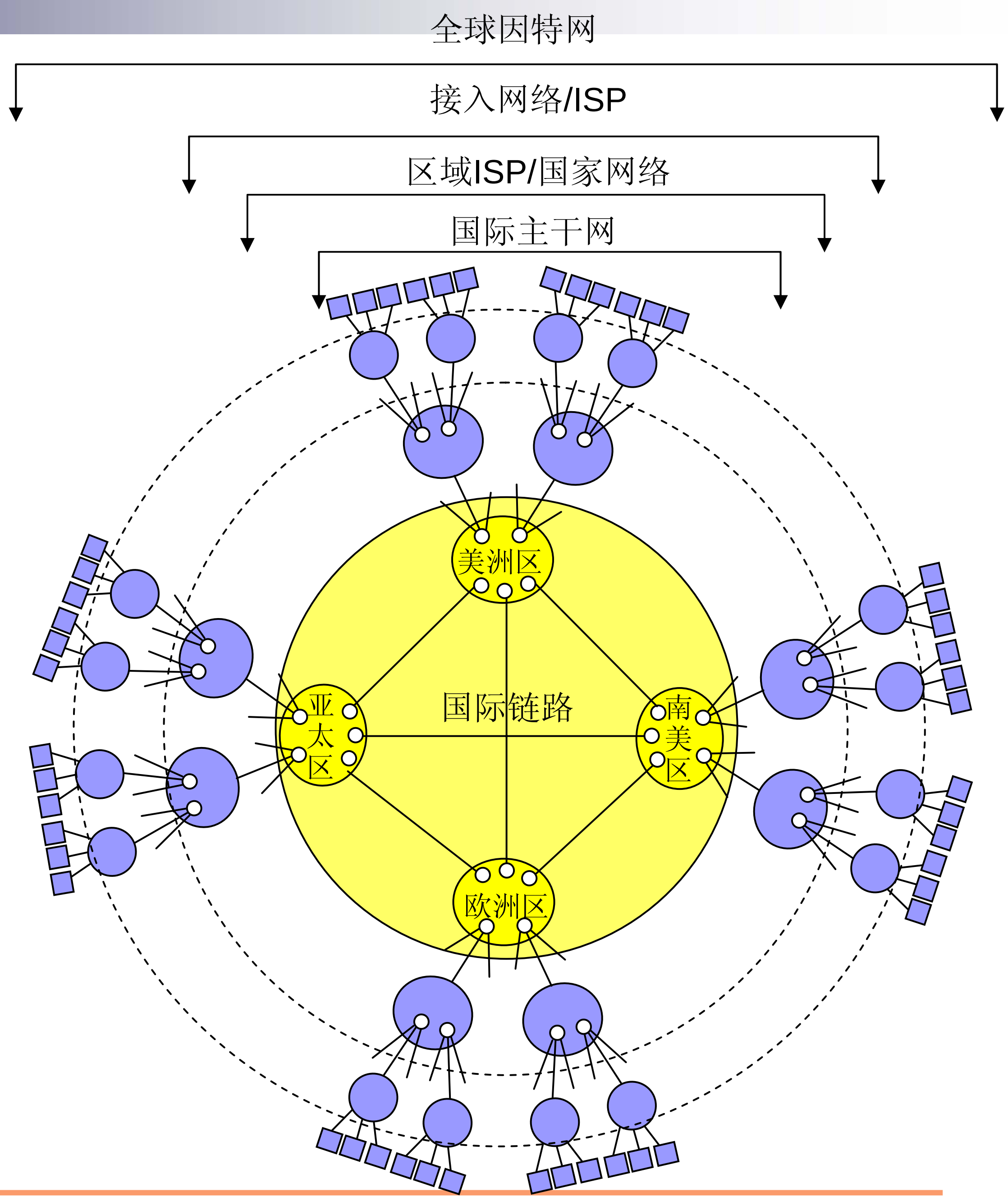
- **TCP/IP Tutorial and Technical Overview, ibm.com**/redbooks
- **Christian Huitma, Routing in the Internet.**

Routing Protocols

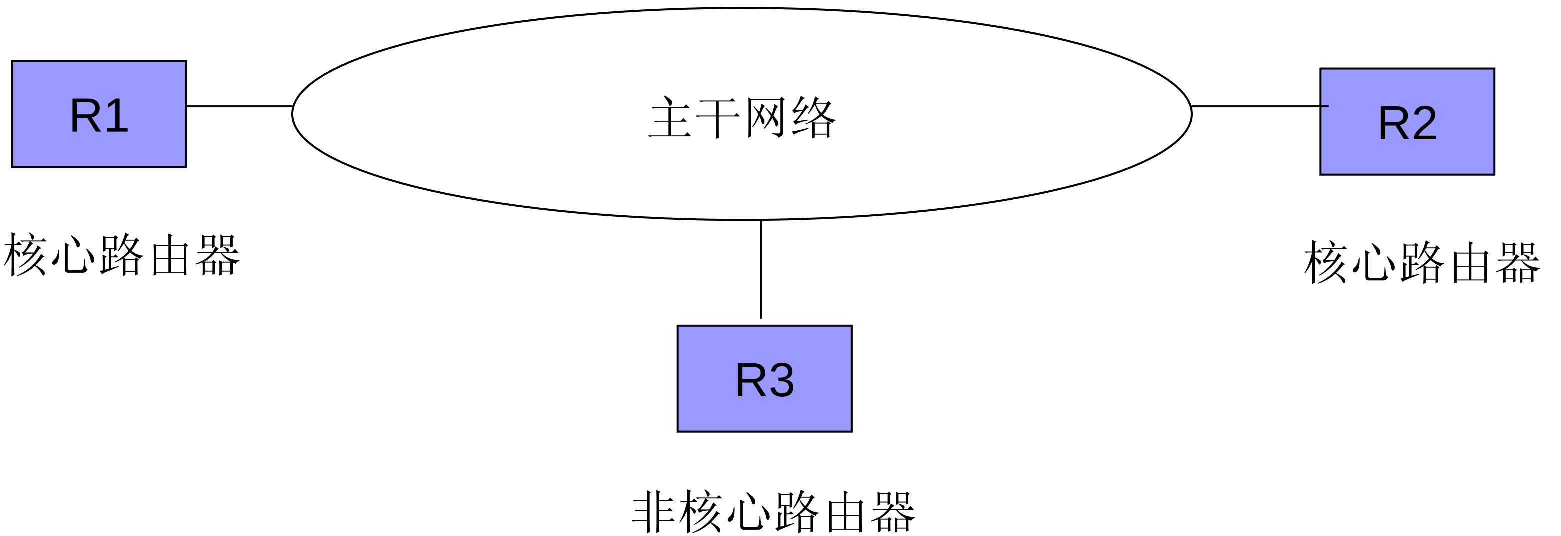
- Routing Information
  - About topology and delays in the internet
- Routing Algorithm
  - Used to make routing decisions based on information

因特网结构

- 第一层
  - 国家主干网
- 第二层
  - 区域ISP
- 第三层
  - 接入网，校园网，无线LAN

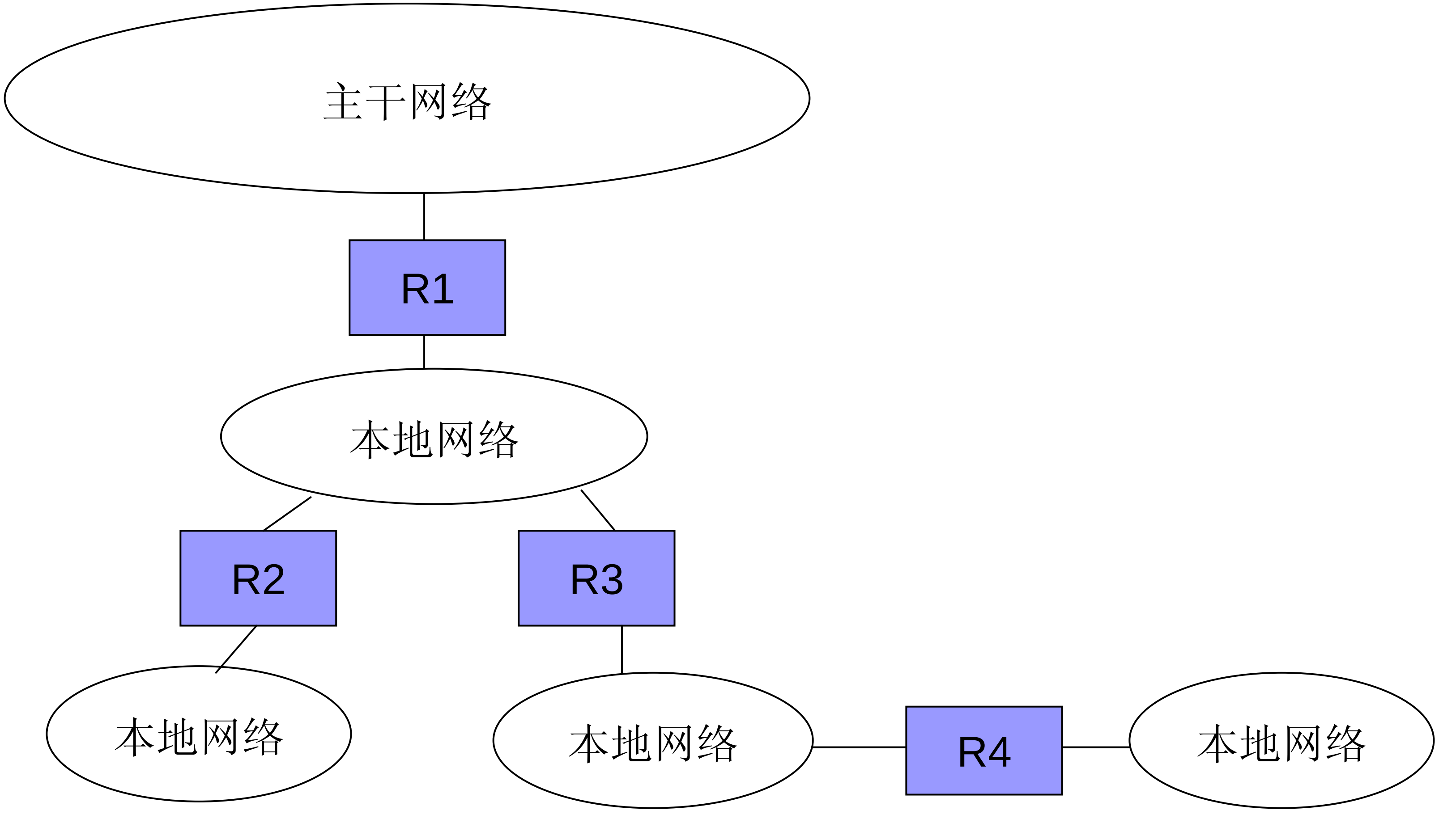


额外跳



- 限制路由器的数量
- 非核心路由其选择一个核心路由器作为默认路由器
- 额外跳

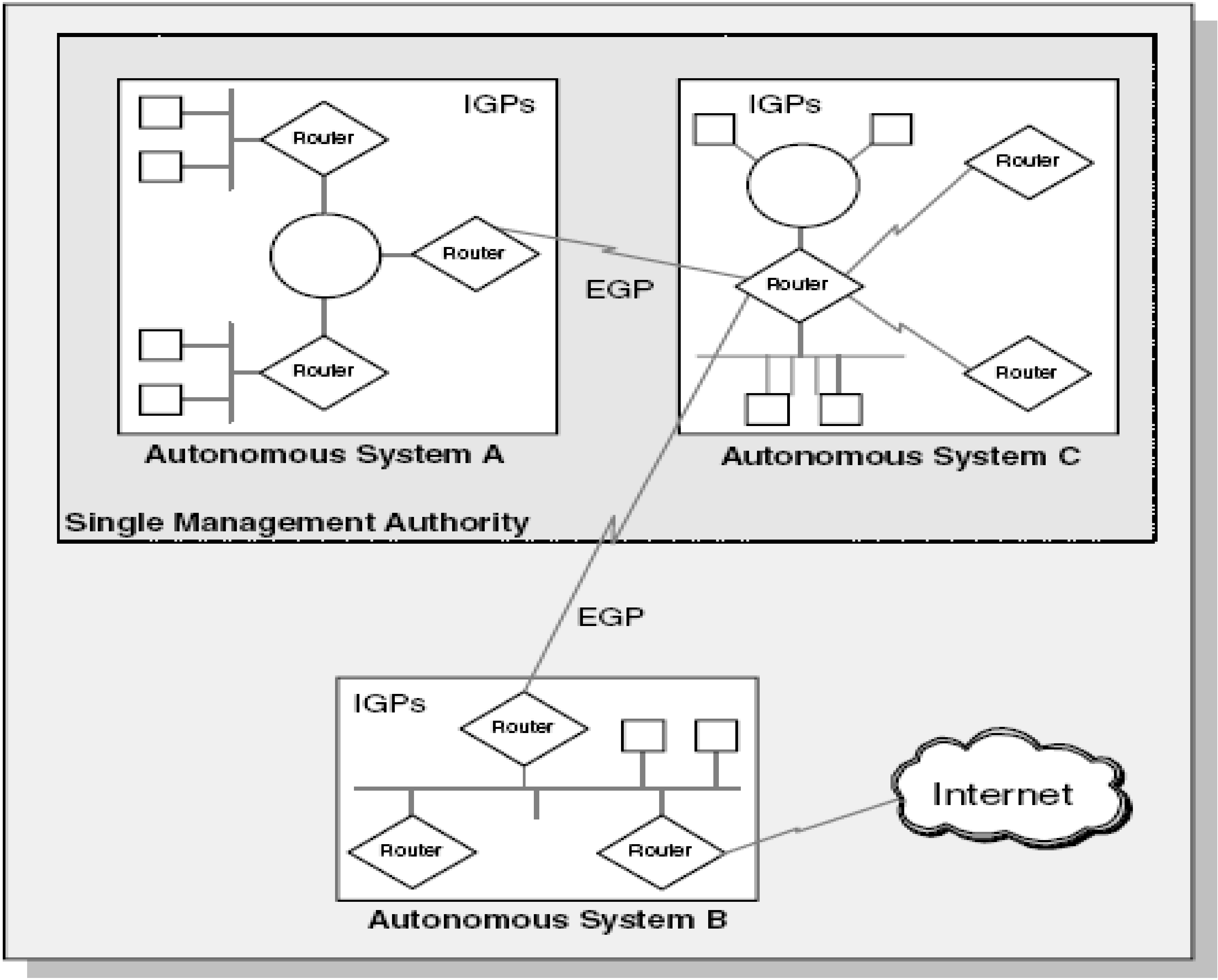
隐藏网络



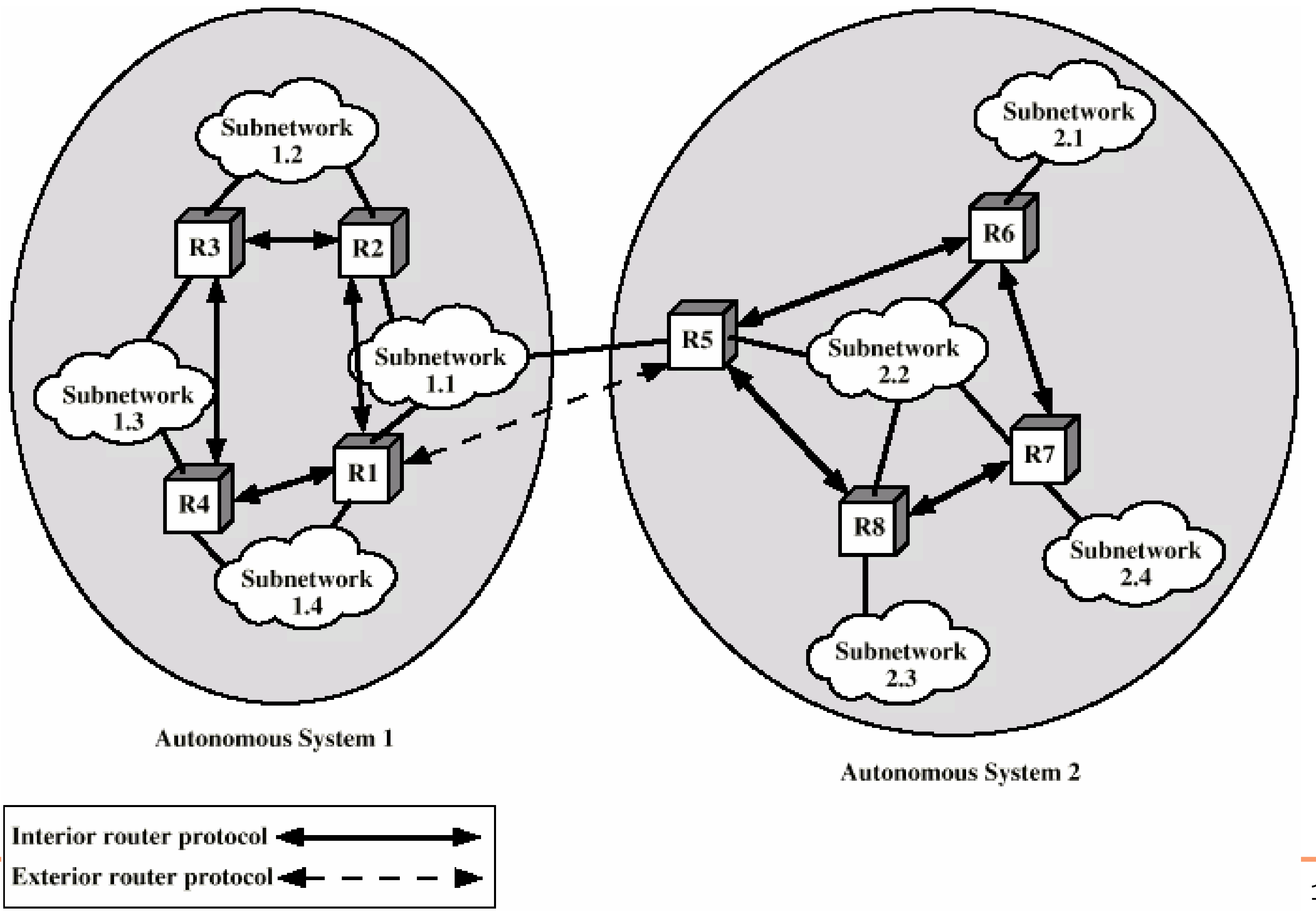
Autonomous systems

- The definition of an autonomous system (AS) is integral to understanding the function and scope of a routing protocol.
- An AS is defined as a logical portion of a larger IP network.
- AS is normally comprised of an internetwork within an organization. It is administered by a single management authority.
- Exchange information
- Common routing protocol
- A connected network
  - There is at least one route between any pair of nodes

- Interior Gateway Protocols (IGPs)
  - Interior gateway protocols allow routers to exchange information within an AS.
  - Examples of these protocols are Open Short Path First (OSPF) and Routing Information Protocol (RIP).
- Exterior Gateway Protocols (EGPs)
  - Exterior gateway protocols allow the exchange of summary information between autonomous systems.
  - An example of this type of routing protocol is Border Gateway Protocol (BGP).


IGPs
Router
Router
Router
Autonomous System A
IGPs
Router
Router
Router
Autonomous System C
EGP
Single Management Authority
EGP
IGPs
Router
Router
Router
Autonomous System B
Internet

Application of IRP and ERP



Types of IP routing

- **Static routing**
  - Static routing is manually performed by the network administrator.
  - The administrator is responsible for discovering and propagating routes through the network.
  - These definitions are manually programmed in every routing device in the environment.
  - There is no communication between routers regarding the current topology of the network.

static routes can be used:

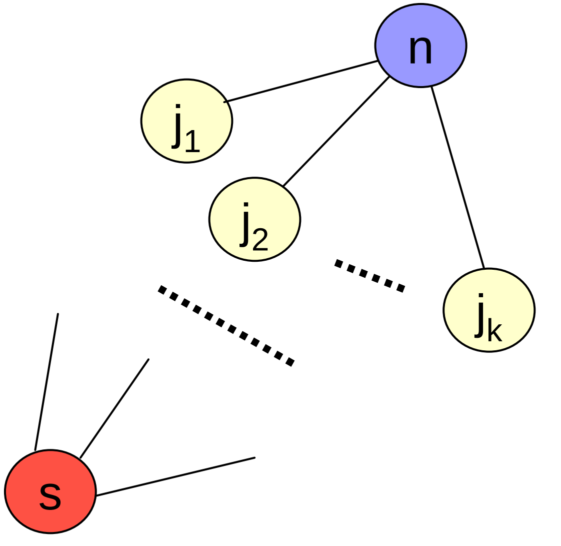
- To manually define a default route.
- To define a route that is not automatically advertised within a network.
- When complex routing policies are required.
- To provide a more secure network environment.
- To provide more efficient resource utilization.

Routing Distance-vector

- Each node (router or host) exchange information with neighboring nodes
  - Neighbors are both directly connected to same network
- **First generation routing algorithm for ARPANET**
- Node maintains vector of link costs for each directly attached network and distance and next-hop vectors for each destination
- Used by Routing Information Protocol (RIP)
- Requires transmission of lots of information by each router
  - Distance vector to all neighbors
  - Contains estimated path cost to all networks in configuration
  - Changes take long time to propagate

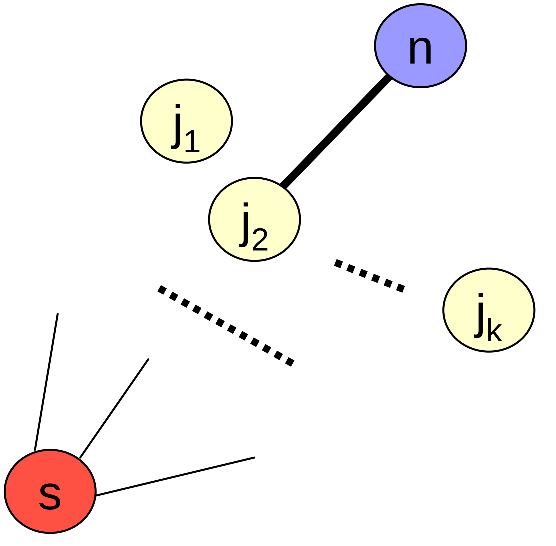
Bellman-Ford Algorithm Method

- Step 1 [Initialization]
  - $L_0(n) = \infty$, for all $n \neq s$
  - $L_h(s) = 0$, for all h
- Step 2 [Update]
- For each successive $h \geq 0$
  - For each $n \neq s$, compute
  - $L_{h+1}(n)=\min_j[L_h(j)+w(j,n)]$
- Connect n with predecessor node j that achieves minimum
- Eliminate any connection of n with different predecessor node formed during an earlier iteration
- Path from s to n terminates with link from j to n



Bellman-Ford Algorithm Method

- Step 1 [Initialization]
  - $L_0(n) = \infty$, for all $n \neq s$
  - $L_h(s) = 0$, for all h
- Step 2 [Update]
- For each successive $h \geq 0$
  - For each $n \neq s$, compute
  - $L_{h+1}(n)=\min_j[L_h(j)+w(j,n)]$
- Connect n with predecessor node j that achieves minimum
- Eliminate any connection of n with different predecessor node formed during an earlier iteration
- Path from s to n terminates with link from j to n

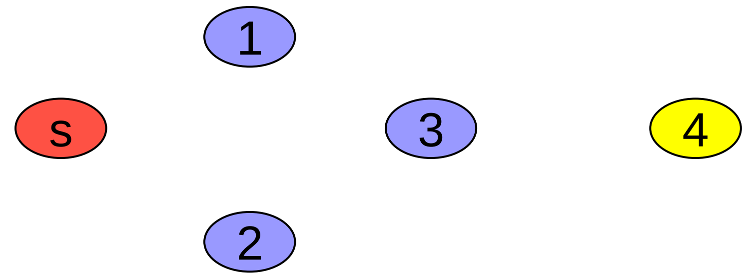


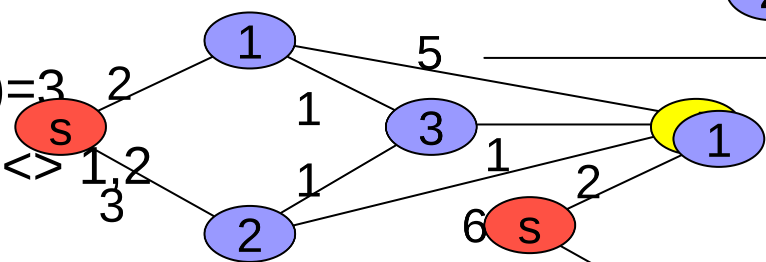
disadvantages with DV

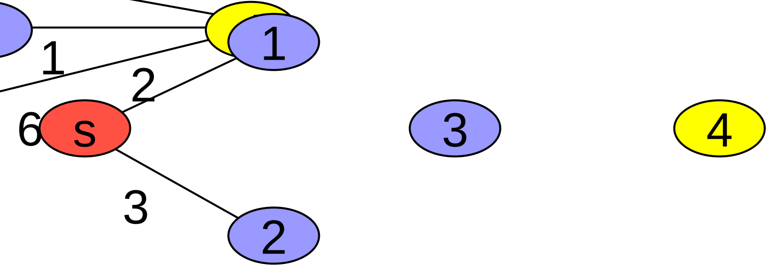
- During an adverse condition, the length of time for every device in the network to produce an accurate routing table is called the ***convergence time***.
- In large, complex internetworks using distance vector algorithms, this time can be excessive.
- To reduce convergence time, a limit is often placed on the maximum number of hops contained in a single route.
- Distance vector routing tables are periodically transmitted to neighboring devices. They are sent even if no changes have been made to the contents of the table.

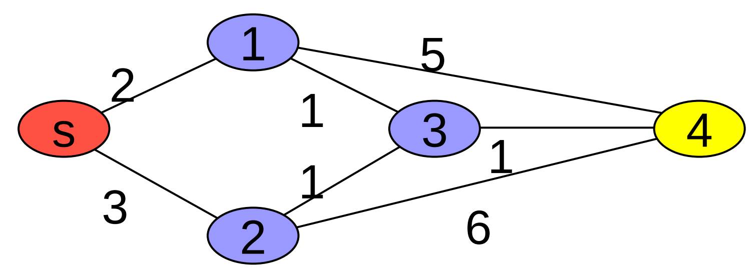
Bellman-Ford Algorithm Method

- Step 1 [Initialization]
  - $L_0(n) = \infty$, for all $n \neq s$
  - $L_h(s) = 0$, for all h
- h=1
  - L(1) =2, L(2)=3
  - L(j) = ∞, j <> 1,2
- h=2
  - L(1) = 2, L(2) = 3,
  - L(3)=3, L(4)=7
- h=3
  - L(1) = 2, L(2) = 3,
  - L(3)=3, L(4)=4







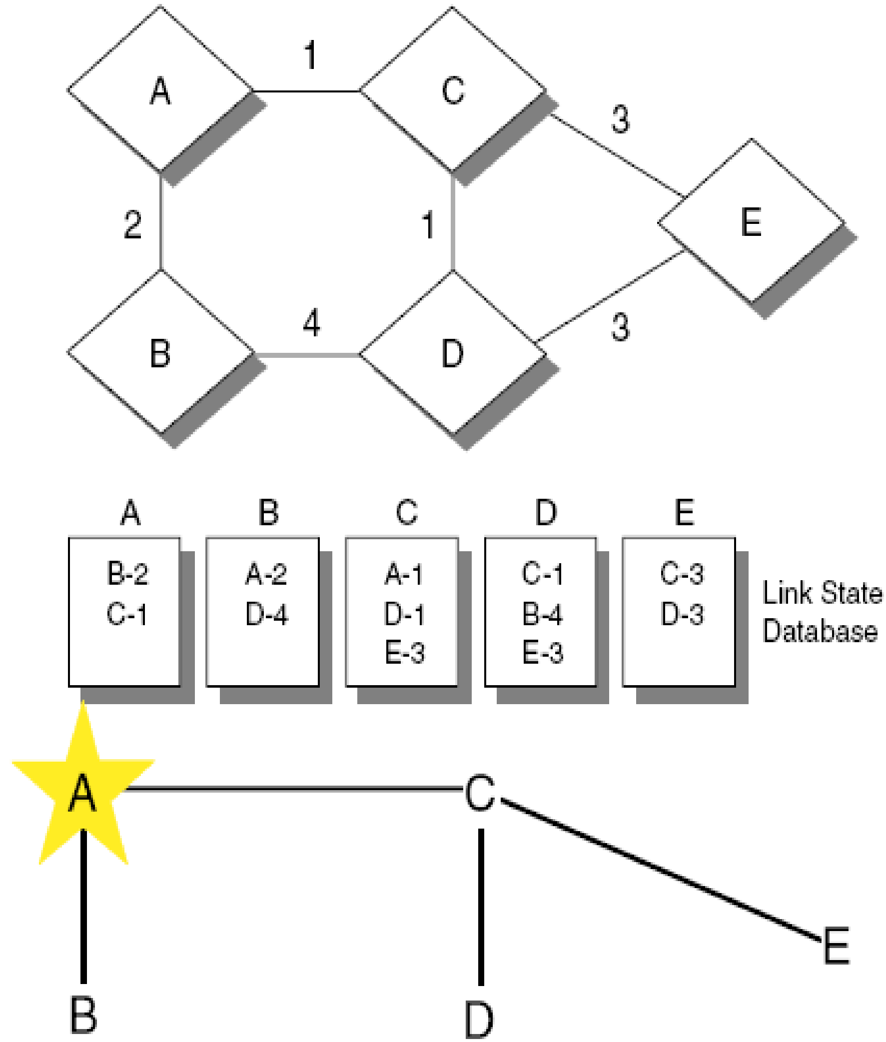


Routing – Link-state

- Designed to overcome drawbacks of distance-vector
- When router initialized, it determines link cost on each interface
- Advertises set of link costs **(Link State Advertisement, LSA)** to all other routers in topology
  - Not just neighboring routers
- From then on, monitor link costs
  - If significant change, router advertises new set of link costs
- Each router can construct topology of entire configuration
  - Can calculate shortest path to each destination network
- Router constructs routing table, listing first hop to each destination
- **Router does not use distributed routing algorithm**
  - Use any routing algorithm to determine shortest paths
  - In practice, Dijkstra's algorithm
- Open shortest path first (OSPF) protocol uses link-state routing.
- **Also second generation routing algorithm for ARPANET**

Shortest-Path First (SPF) algorithm

- The SPF algorithm is used to process the information in the topology database.
- It provides a tree-representation of the network. The device running the SPF algorithm is the root of the tree.
- The output of the algorithm is the list of shortest-paths to each destination network.



Exterior Router Protocols – Not Distance-vector

- Link-state and distance-vector not effective for exterior router protocol
  - Distance-vector assumes routers share common distance metric
  - ASs may have different priorities
    - May have restrictions that prohibit use of certain other AS
    - Distance-vector gives no information about ASs visited on route

Exterior Router Protocols – Not Link-state

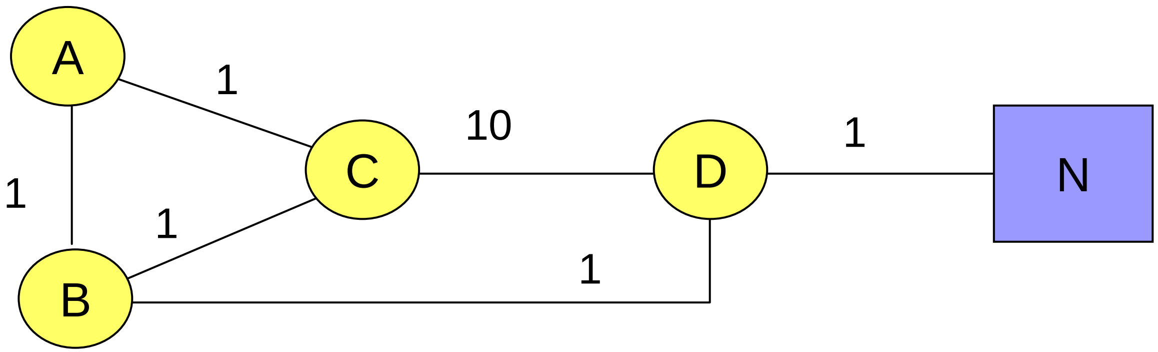
- Different ASs may use different metrics and have different restrictions
  - Impossible to perform a consistent routing algorithm.
- Flooding of link state information to all routers unmanageable

Exterior Router Protocols – Path-vector

- Dispense with routing metrics
- Provide information about which networks can be reached by a given router and ASs crossed to get there
  - Does not include distance or cost estimate
- Each block of information lists all ASs visited on this route
  - Enables router to perform policy routing
  - E.g. avoid path to avoid transiting particular AS
  - E.g. link speed, capacity, tendency to become congested, and overall quality of operation, security
  - E.g. minimizing number of transit ASs

Routing Information Protocol (RIP)

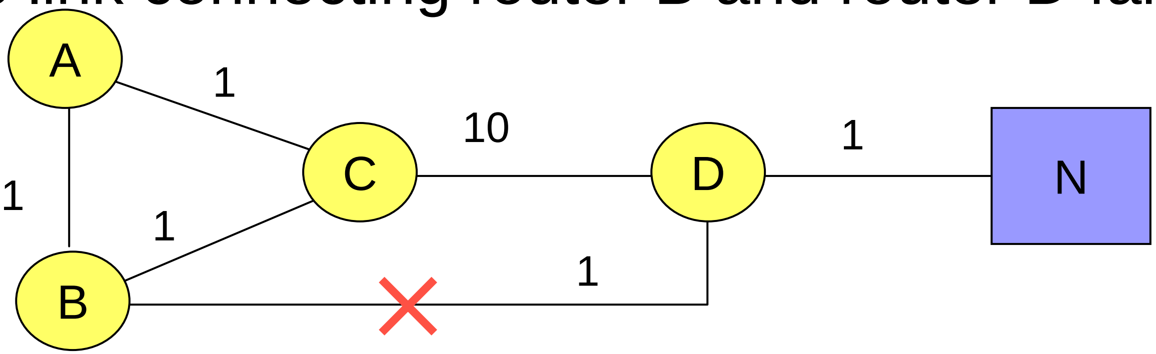
Convergence



- Router D to the target network: Directly connected network. Metric 1.
- Router B to the target network: Next hop is router D. Metric is 2.
- Router C to the target network: Next hop is router B. Metric is 3.
- Router A to the target network: Next hop is router B. Metric is 3.

counting to infinity

- the link connecting router B and router D fails.



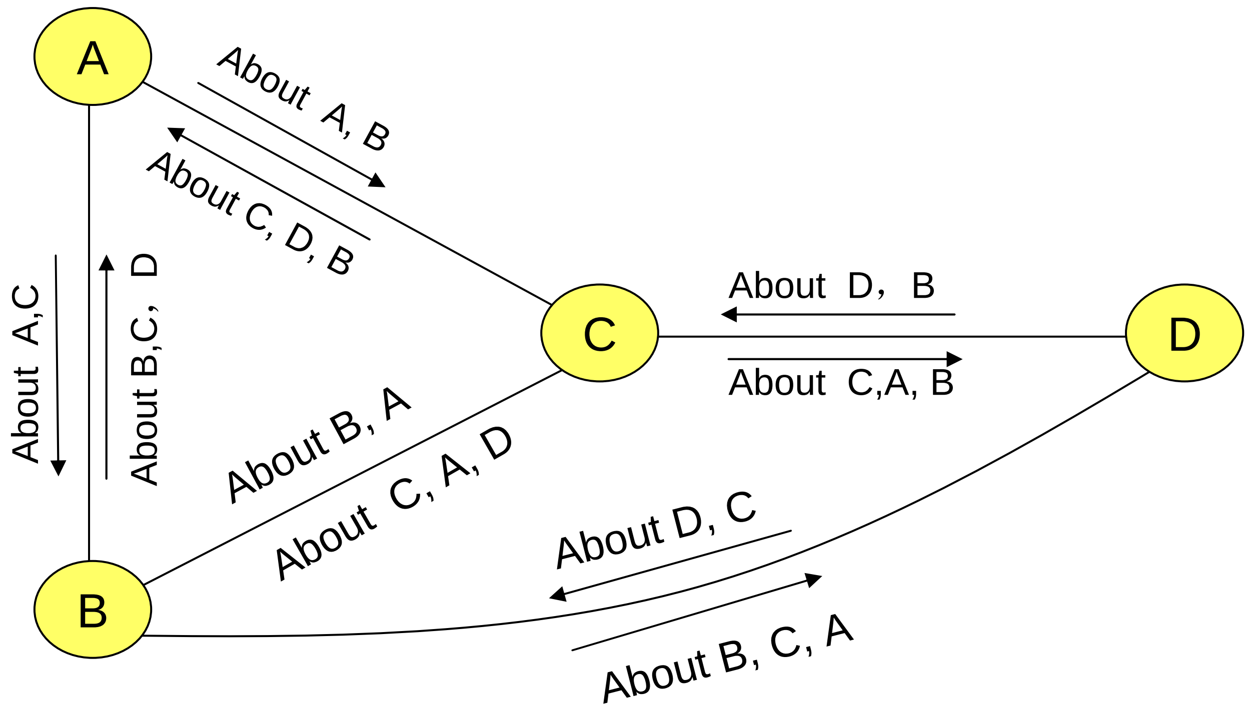
Time

D:	Direct	1	Direct	1	Direct	1	Direct	1	....	Direct	1	Direct	1
B:	Unreachable		C	4	C	5	C	6		C	11	C	12
C:	B	3	A	4	A	5	A	6		A	11	D	11
A:	B	3	C	4	C	5	C	6	....	C	11	C	12

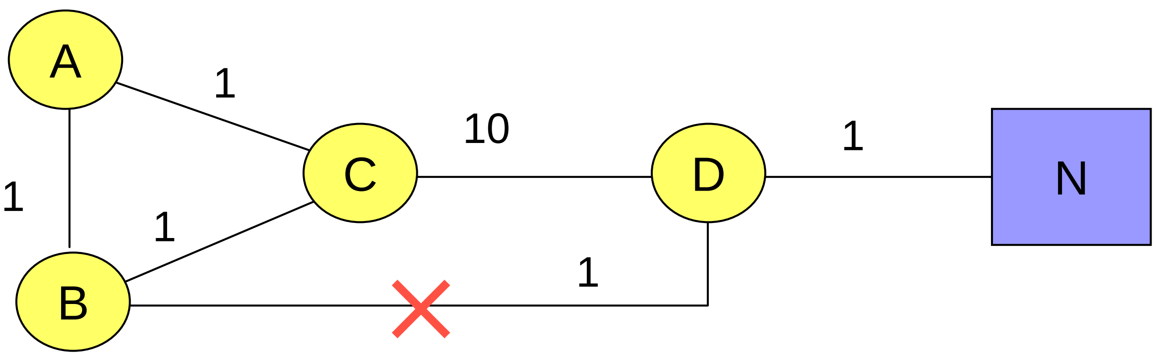
- The length of a route must be less than 15. 15 = infinity.

split horizon

- The "simple " scheme omits routes learned from one neighbor in updates sent to that neighbor.



split horizon



Time

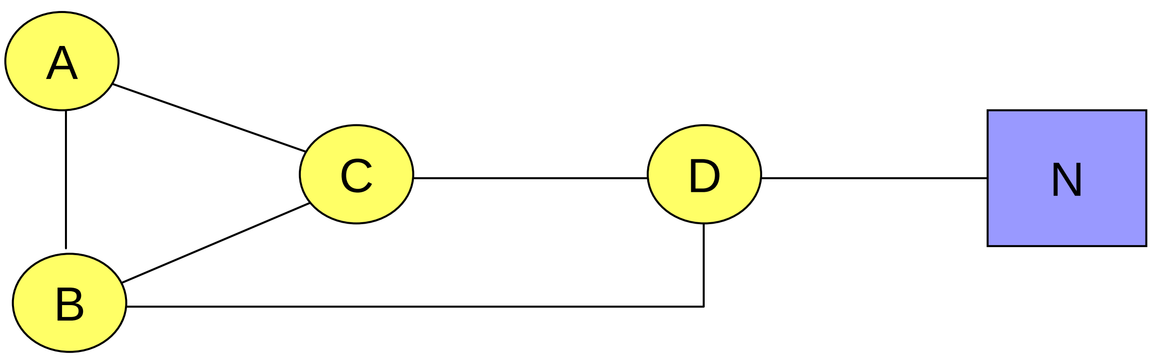
D:	Direct	1	Direct	1	Direct	1	Direct	1
B:	Unreachable		Unreachable		Unreachable		C	12
C:	B	3	A	4	D	11	D	11
A:	B	3	C	4	Unreachable		C	12

Note: Faster Routing Table Convergence

Wait for timeout

Split horizon with poisoned reverse

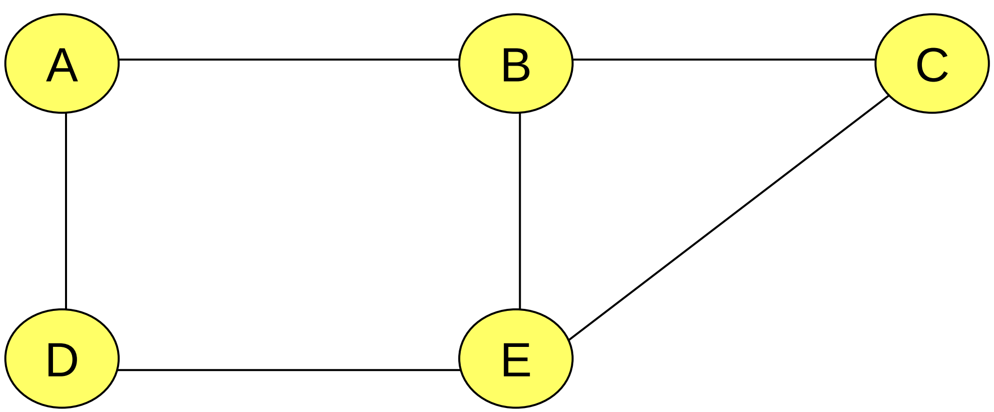
- "Split horizon with poisoned reverse" includes such routes in updates, but sets their metrics to infinity.
- If A thinks it can get to D via C, its messages to C should indicate that D is unreachable.
- If the route through C is real, then C either has a direct connection to D, or a connection through some other gateway.



With poison reverse, when a routing update indicates that a network is unreachable, routes are immediately removed from the routing table.

counting to infinity under the Split horizon with poisoned reverse

	距离	下一跳
B→D	2	E
C→D	2	E
E→D	1	



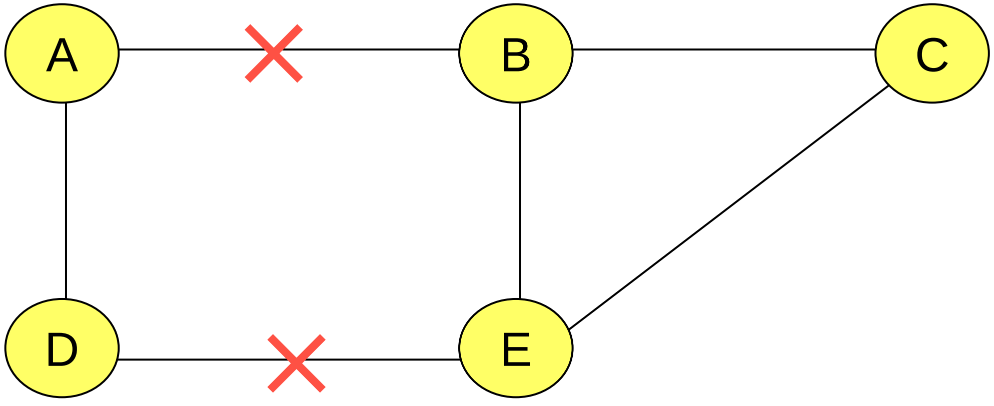
counting to infinity under the Split horizon with poisoned reverse

	距离	下一跳
B→D	2	E
C→D	2	E
E→D	无穷	

	距离	下一跳
B→D	无穷	
C→D	2	E
E→D	无穷	

	距离	下一跳
B→D	3	C
C→D	2	E
E→D	4	B

Unreachable message reached B but not reached C.



Triggered updates

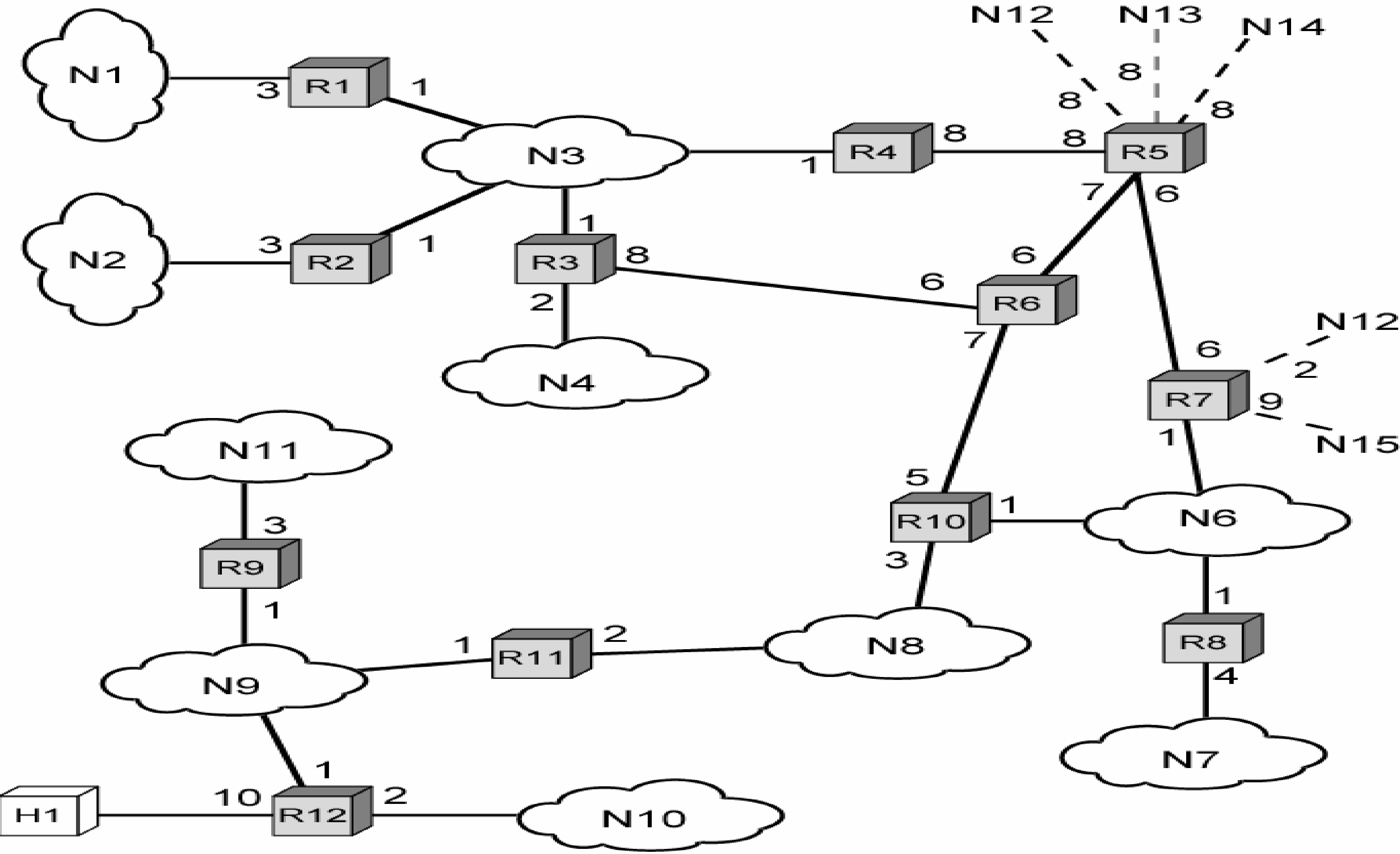
- To get triggered updates, we simply add a rule that

  whenever a gateway changes the metric for a route, it is required to send update messages almost immediately, even if it is not yet time for one of the regular update message.

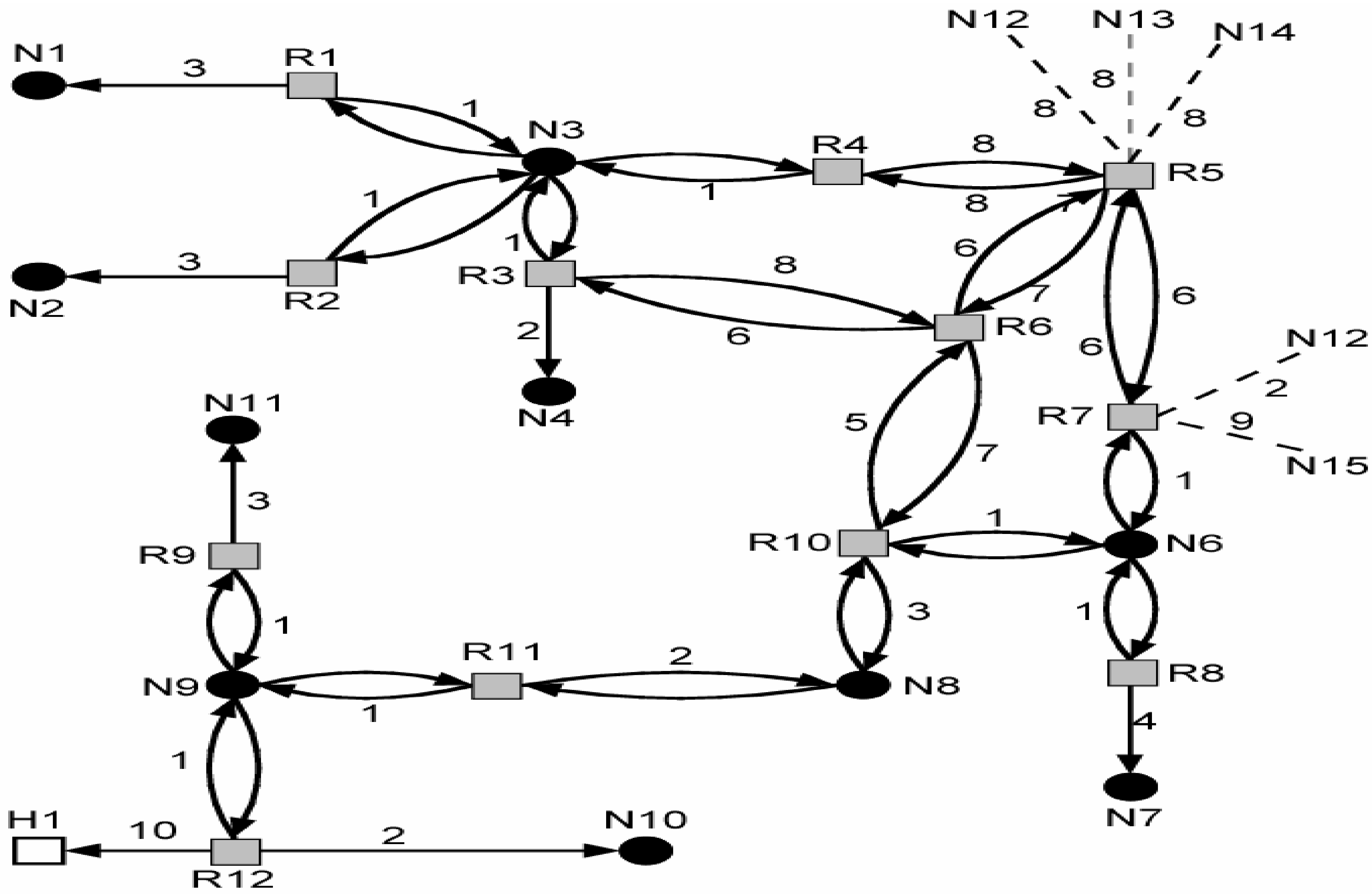
- RIP is a UDP-based protocol.
- Each host that uses RIP has a routing process that sends and receives datagrams on UDP port number 520.

OSPF

Sample AS — a OSPF network



Directed Graph of AS



SPF Tree

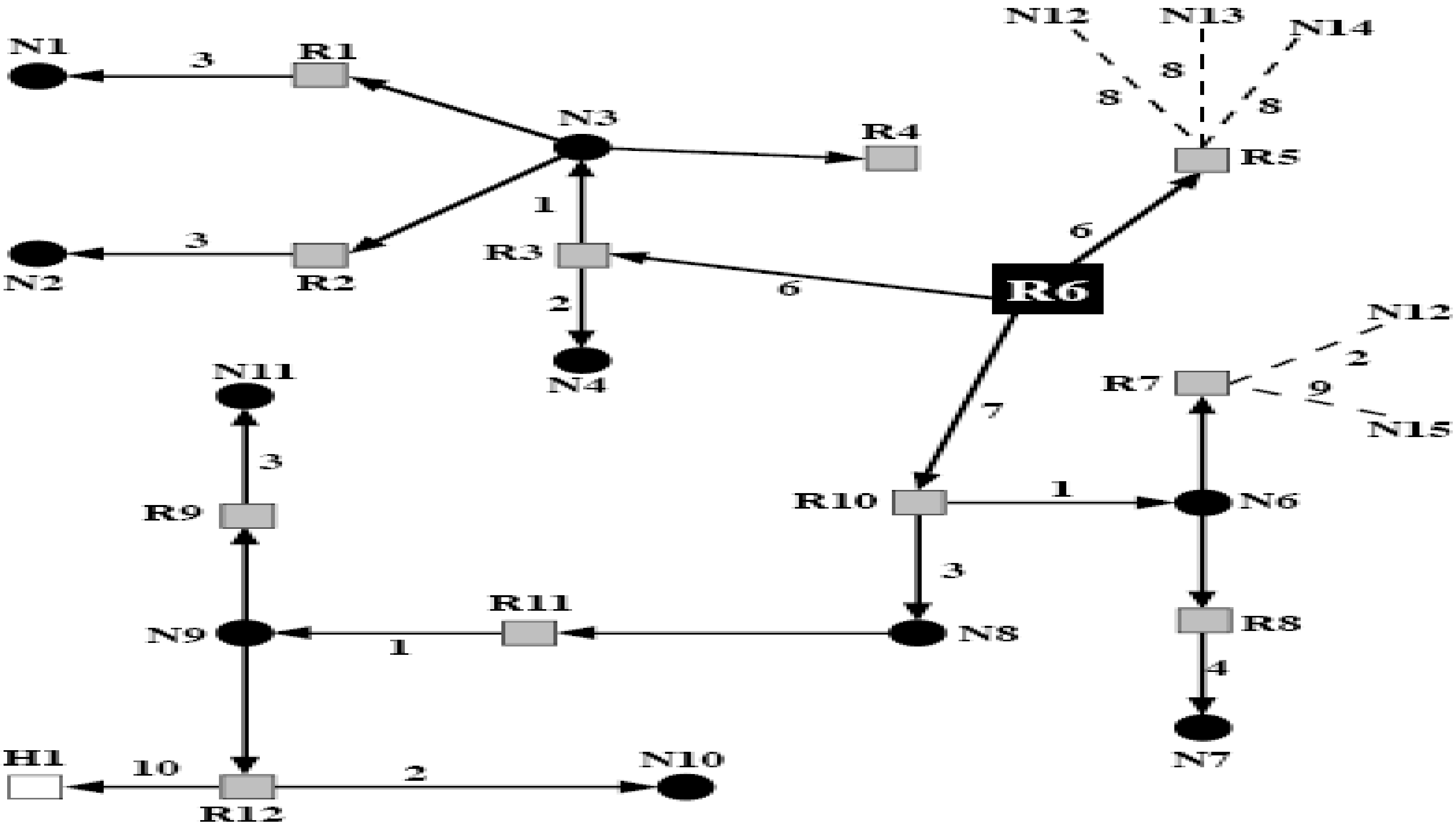


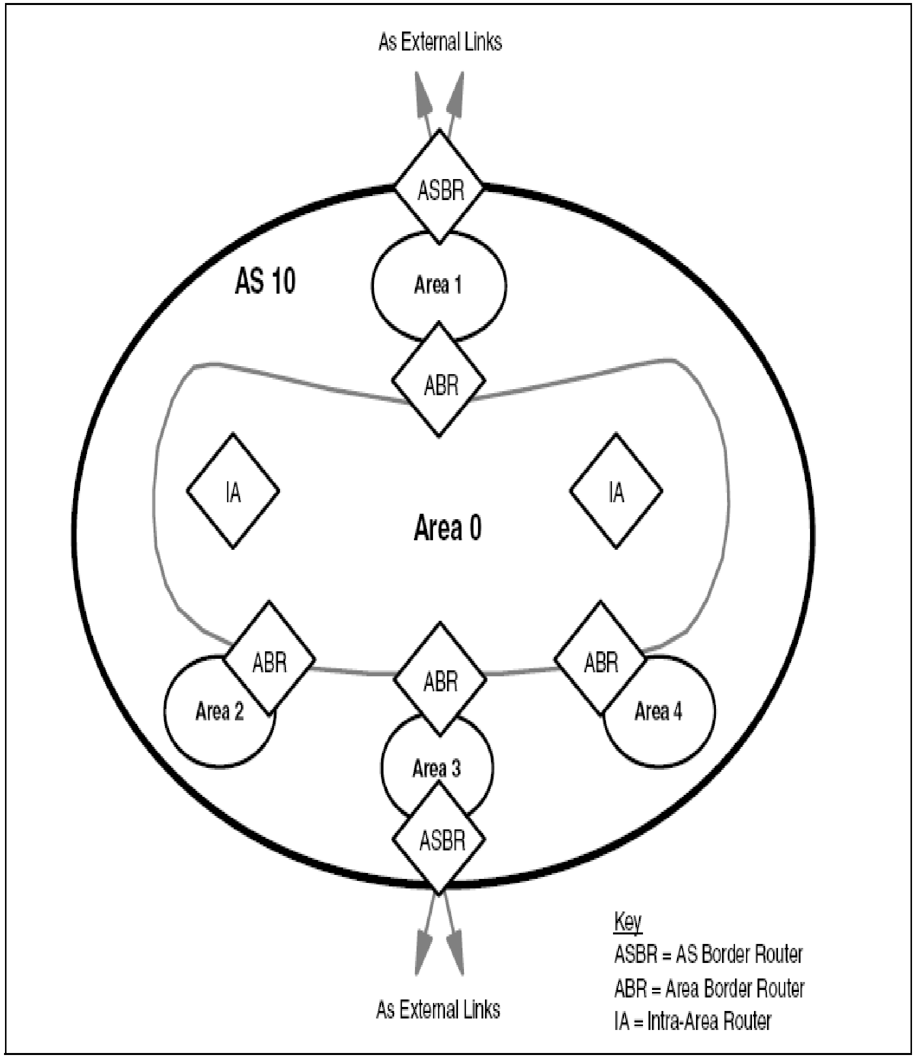
Figure 19.9 The SPF Tree for Router R6

OSPF terminology

1. OSPF areas
2. Intra-area, area border and AS boundary routers
3. Physical network types
4. Neighbor routers and adjacencies
5. Designated and backup designated router
6. Link state database
7. Link state advertisements and flooding

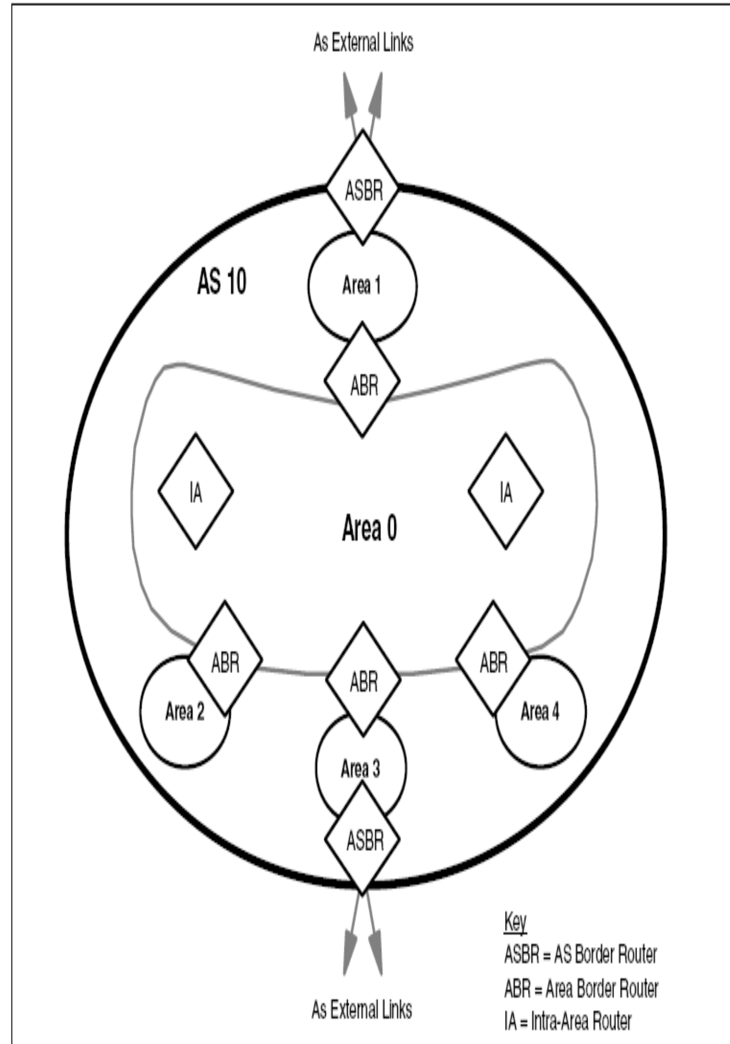
(1) OSPF areas

- OSPF networks are divided into a collection of *areas*.
- An area consists of a logical grouping of networks and routers.
- The area may coincide with geographic or administrative boundaries.
- Each area is assigned a 32-bit *area ID*.



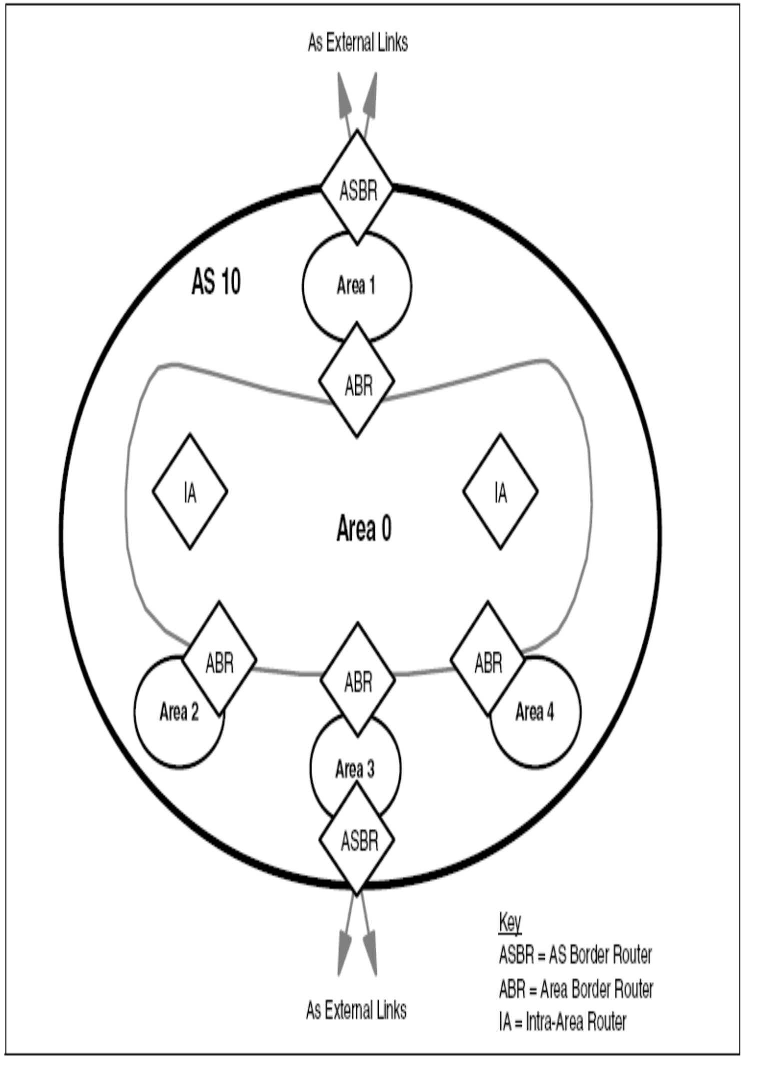
(1) OSPF areas

- benefits:
  - **Within an area, every router maintains an identical topology database, This reduces the size of the topology database maintained by each router.**
  - **Areas limit the potentially explosive growth in the number of link state updates.**
  - **Areas reduce the CPU processing required to maintain the topology database.**



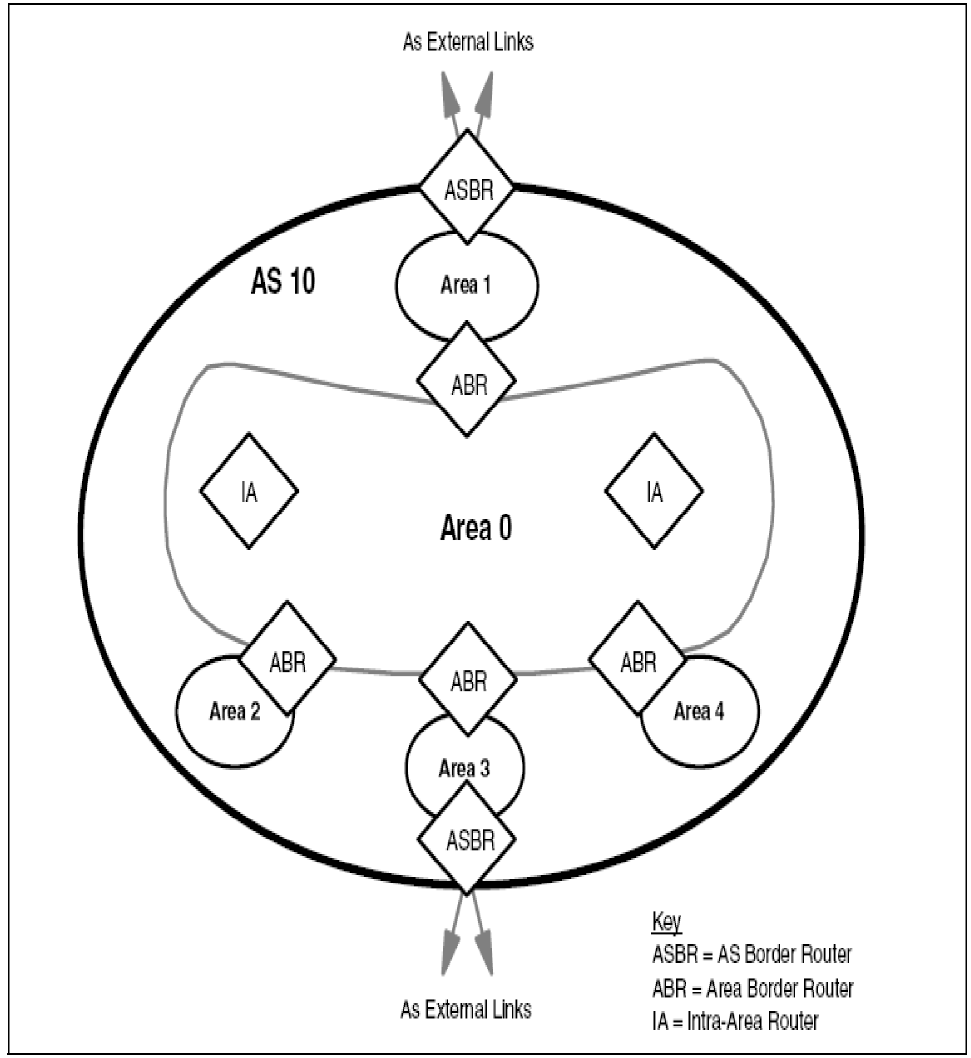
(1) OSPF areas

- ***Backbone area and area 0***
  - All OSPF networks contain at least one backbone area.
  - Additional areas may be created based on network topology or other design requirements.
  - the backbone physically connects to all other areas.
  - OSPF expects all areas to announce routing information directly into the backbone.



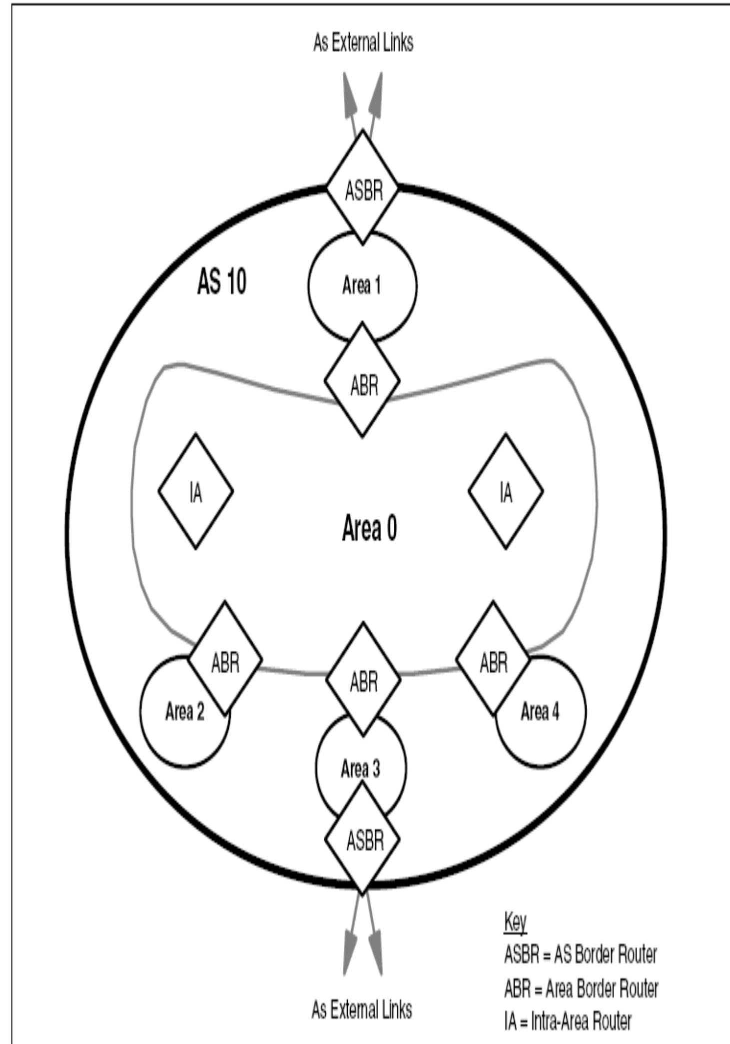
(2) Intra-area, area border and AS boundary routers

- Intra-Area Routers
  - This class of router is logically located entirely within an OSPF area. Intra-area routers maintain a topology database for their local area.



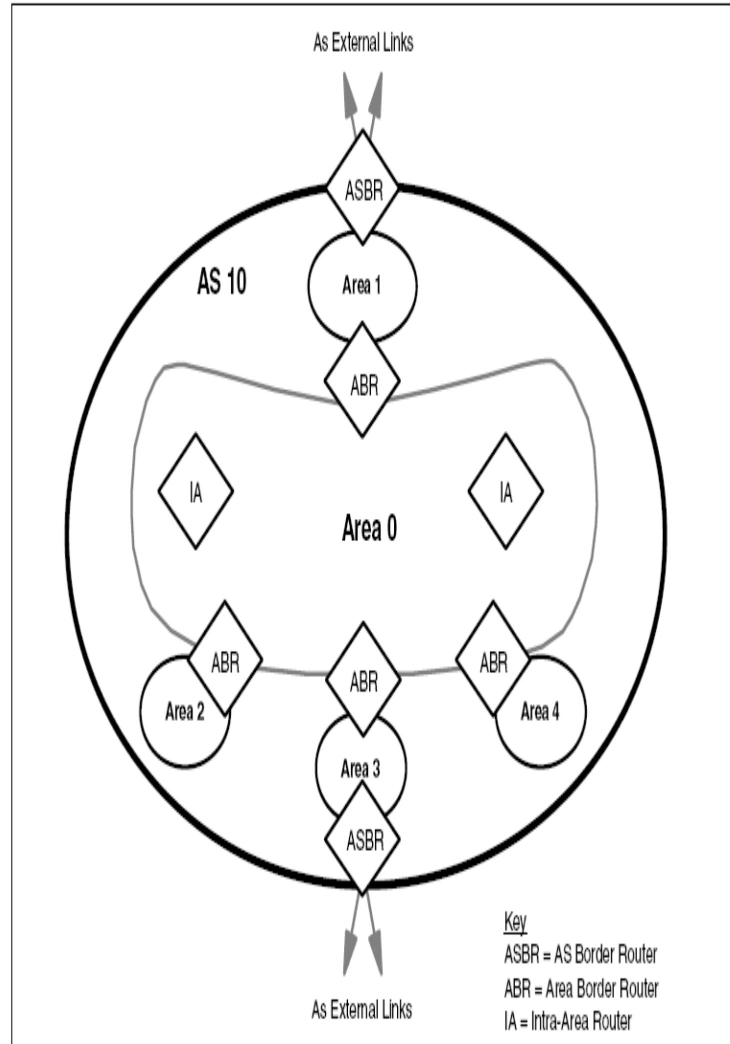
(2) Intra-area, area border and AS boundary routers

- Area Border Routers (ABR)
  - This class of router is logically connected to two or more areas. One area must be the backbone area.
  - An ABR is used to interconnect areas.
  - They maintain a separate topology database for each attached area.
  - ABRs also execute separate instances of the SPF algorithm for each area.



(2) Intra-area, area border and AS boundary routers

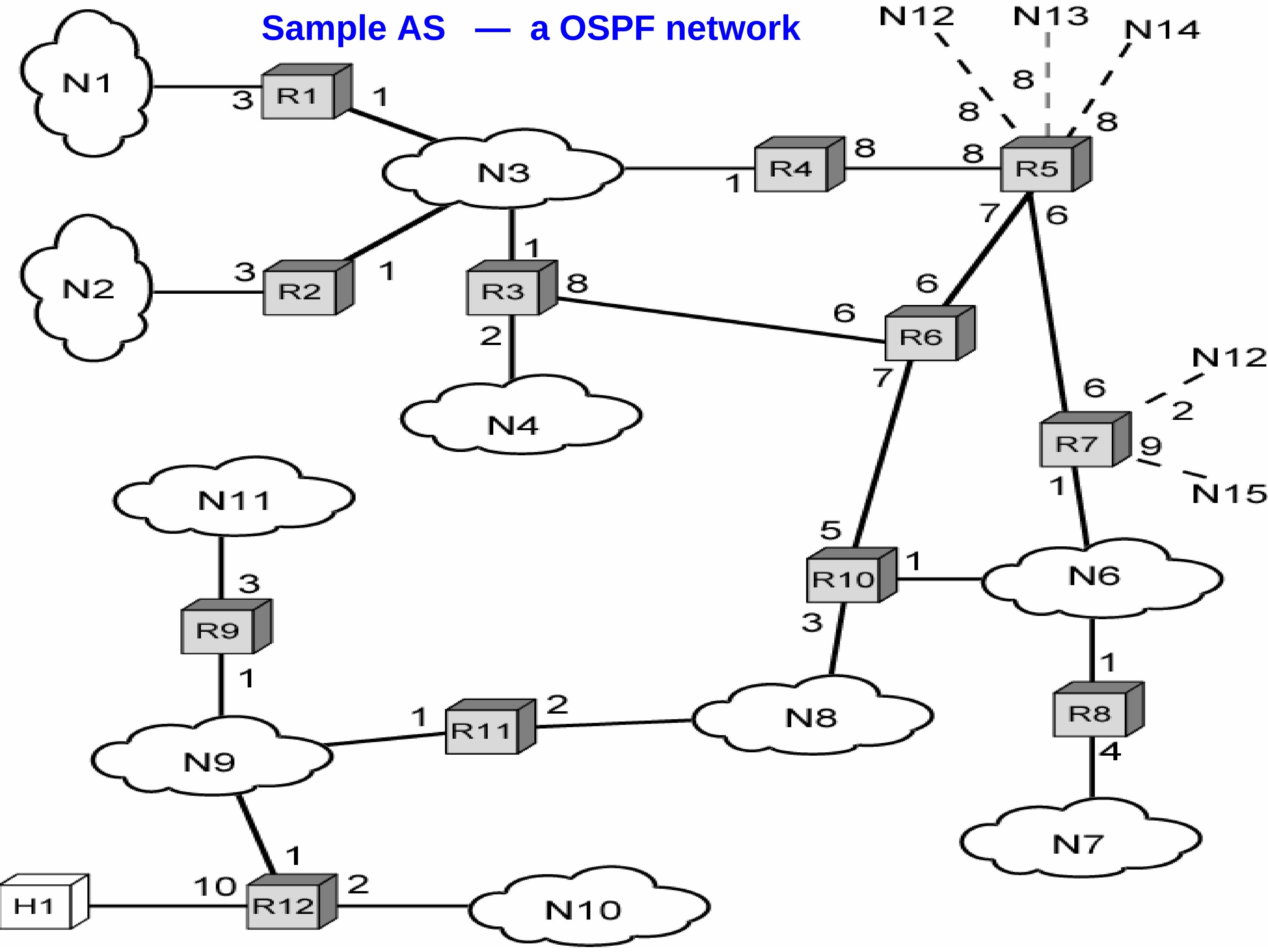
- AS Boundary Routers (ASBR)
  - This class of router is located at the periphery of an OSPF internetwork.
  - It functions as a gateway exchanging reachability between the OSPF network and other routing environments.
  - ASBRs are responsible for announcing AS external link advertisements through the AS.



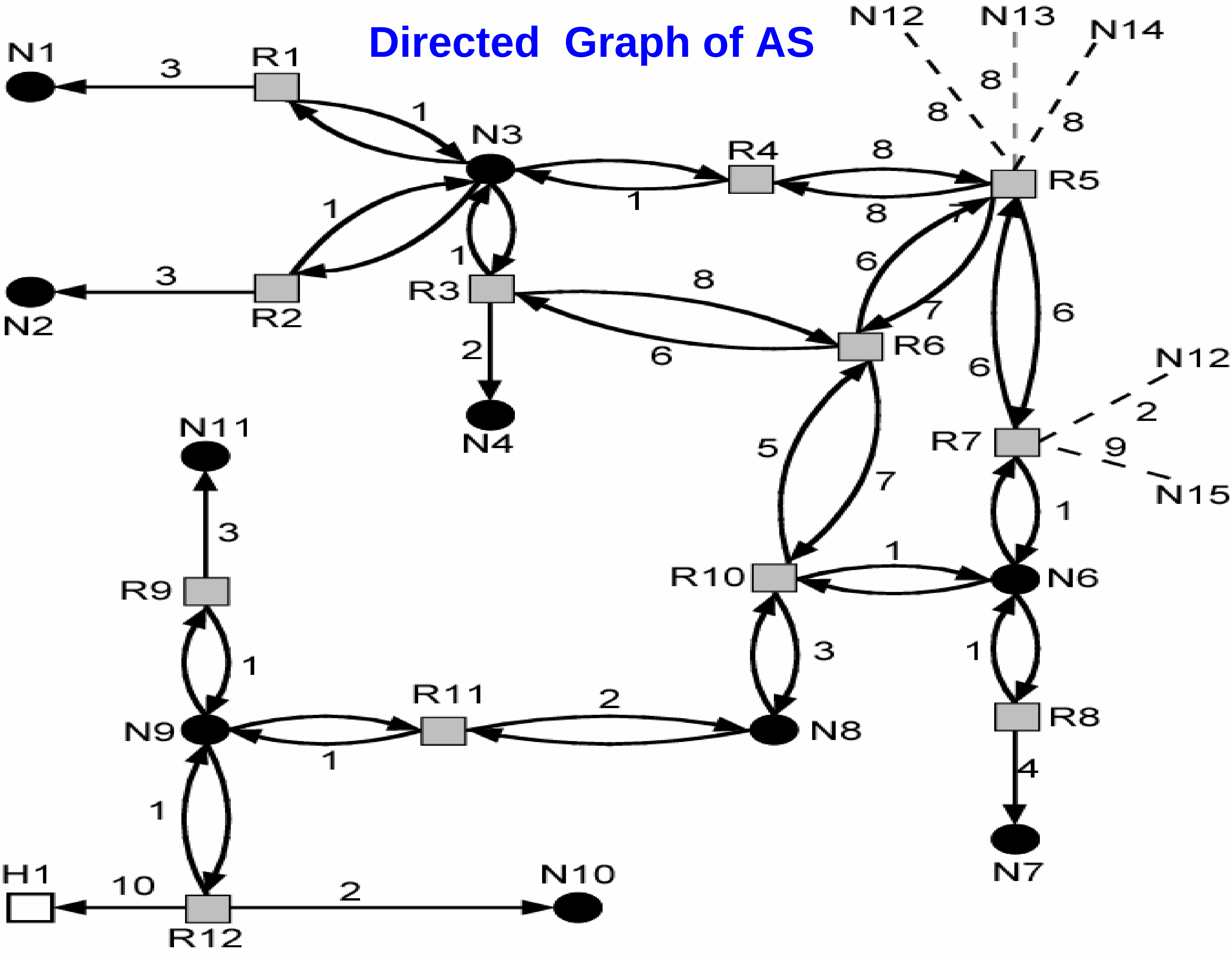
(3) Physical network types

- Point-to-point
  - Point-to-point networks directly link two routers.
- Multi-access
  - Multi-access networks support the attachment of more than two routers.
  - **Broadcast networks** have the capability of simultaneously directing a packet to all attached routers. Ethernet and token-ring LANs
  - **Non-broadcast networks**. Each packet must be specifically addressed to every router in the network. X.25 and frame relay networks.

Sample AS — a OSPF network



Directed Graph of AS



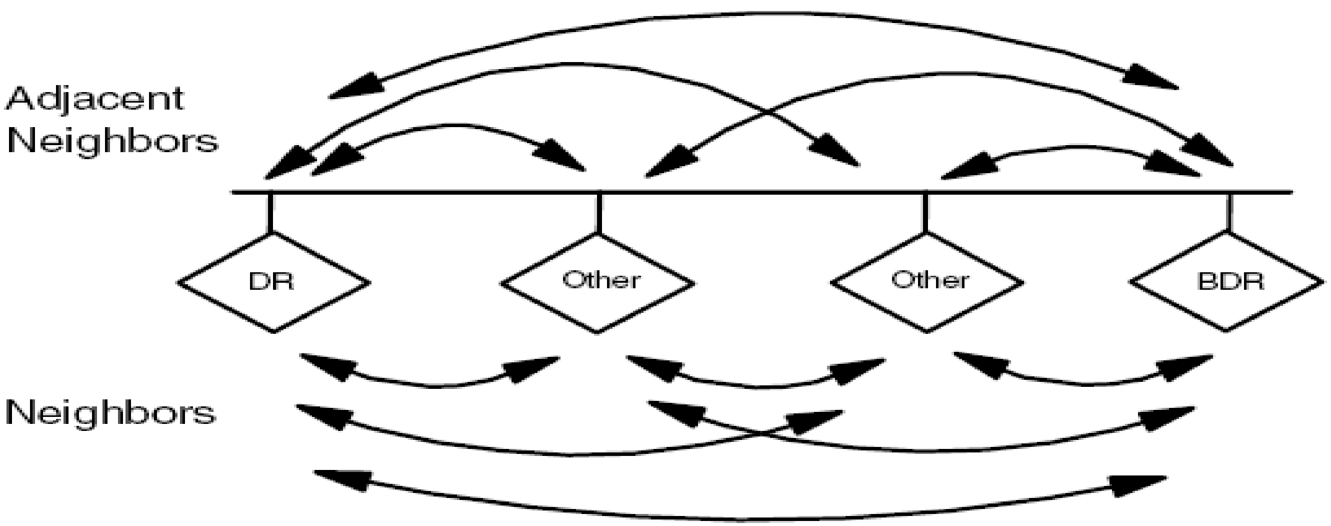
（4）Neighbor routers and adjacencies

- Routers that share a common network segment establish a neighbor relationship on the segment.
  - Area-id:The routers must belong to the same OSPF area.
  - Authentication
  - Hello and dead intervals: The routers must specify the same timer intervals used in the Hello protocol.

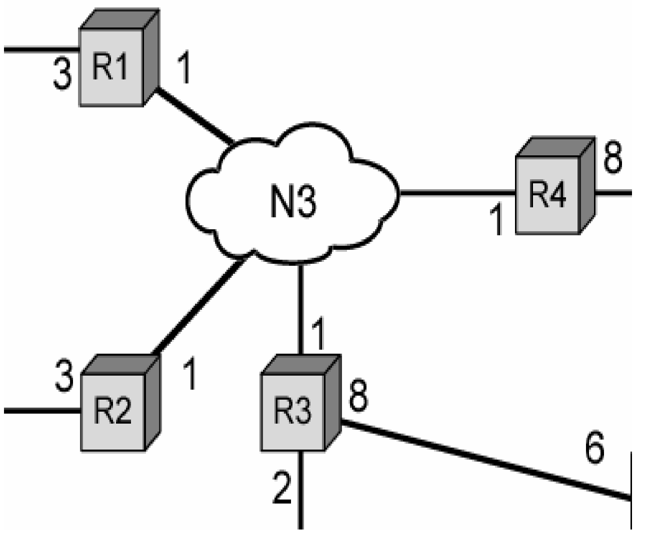
- Neighboring routers are considered adjacent when
  - they have synchronized their topology databases.
  - This occurs through the exchange of link state information.

(5) Designated and backup designated router

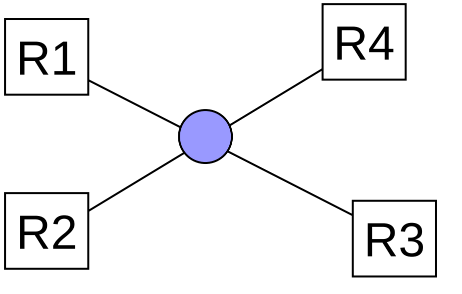
- Each multi-access network elects a designated router (DR) and backup designated router (BDR).
  - It forms adjacencies with all routers on the multi-access network.
  - It generates network link advertisements listing each router connected to the multi-access network.



- R1-R4之间有4×3/2＝6条链路要公告
- R1-R4与N3之间有4条链路要公告
- 共有10(≈$n^2/2$)条链路公告



- 指派路由公告N3到R1-R4的链路，路由器自己公布路由器到N3的链路。共8(2n)条链路。



Link state database

- The link state database is also called the *topology database*（**link state database**）.
- It contains the set of link state advertisements describing the OSPF network and any external connections.
- Each router within the area maintains an identical copy of the link state database.

Link state advertisements and flooding

- LSAs are exchanged between adjacent OSPF routers.
- *reliable flooding*.
  - Each router stores the LSA for a period of time before propagating the information to its neighbors. If, during that time, a new copy of the LSA arrives, the router replaces the stored version. However, if the new copy is outdated, it is discarded.
  - To ensure reliability, each link state advertisement must be acknowledged. Multiple acknowledgements can be grouped together into a single acknowledgement packet. If an acknowledgement is not received, the original link state update packet is retransmitted.

OSPF packet types

- OSPF packets are transmitted in IP datagrams. They are not encapsulated within TCP or UDP packets.
- OSPF uses multicast facilities to communicate with neighboring devices.
- Packets are sent to the reserved multicast address 224.0.0.5 (AllSPFRouters address).

Common header of OSPF packets

Number of Octets	Field	
1	Version	Version = 2
1	Packet Type	1 = Hello 2 = Database Description 3 = Link State Request 4 = Link State Update 5 = Link State Acknowledgment
2	Packet Length	
4	Router ID	
4	Area ID	
2	Checksum	
2	Authentication Type	0 = No Authentication 1 = Simple Password
8	Authentication Data	Password if Type 1 Selected

five possible types of OSPF

- Hello
  - This packet type is used to discover and maintain neighbor relationships.
- Database description
  - This packet type describes the set of LSAs contained in the router's link state database.
- Link state request
  - This packet type is used to request a more current instance of an LSA from a neighbor.
- Link state update
  - This packet type is used to provide a more current instance of an LSA to a neighbor.
- Link state acknowledgement
  - This packet type is used to acknowledge receipt of a newly received LSA.

activities to accomplish this information exchange

- **Neighbor communication**
- **Electing a designated router**
- **Establishing adjacencies and synchronizing databases**

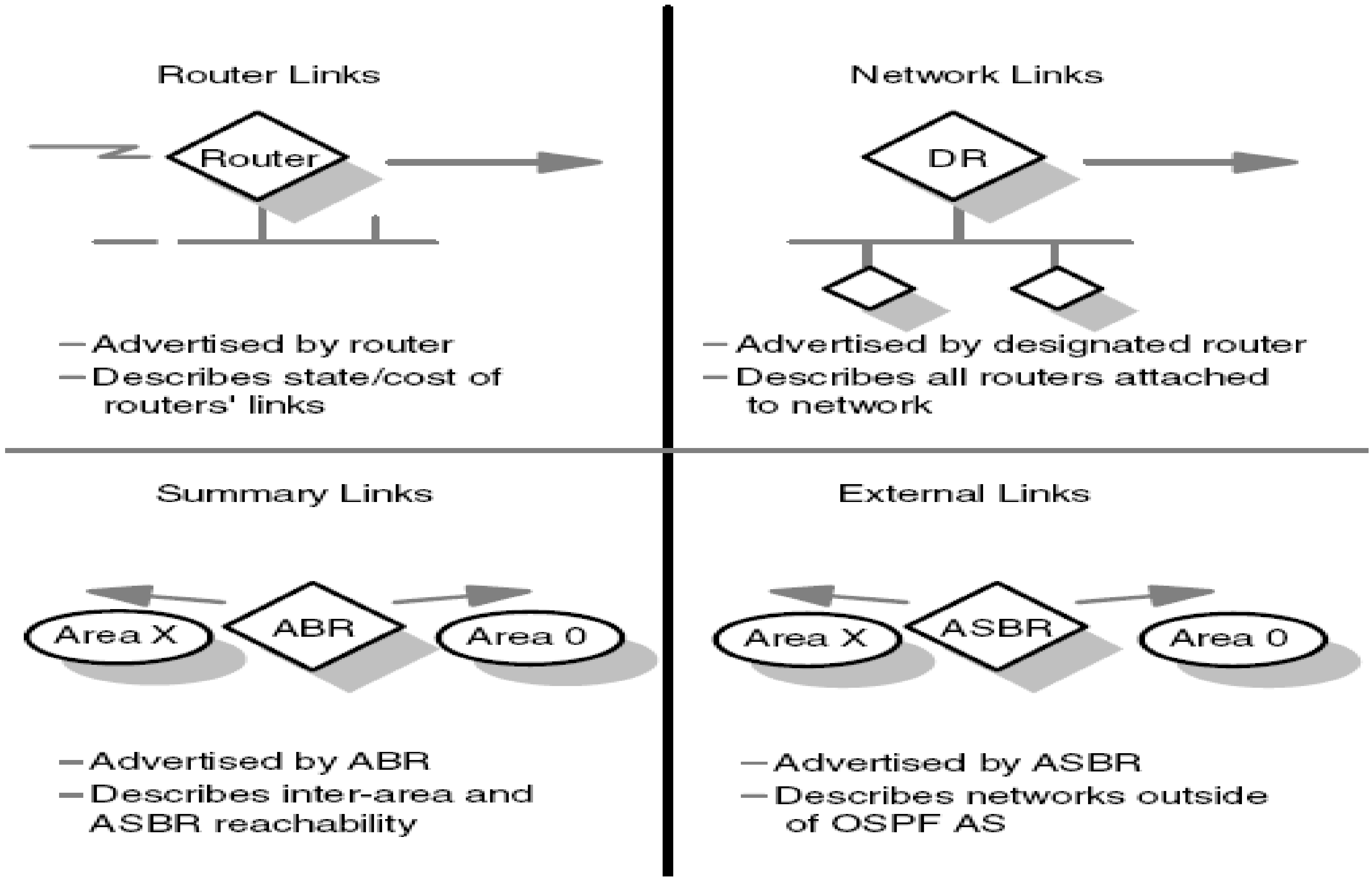
Neighbor communication

- The Hello protocol discovers and maintains relationships with neighboring routers.
- Hello packets are periodically sent out to each router interface.
- The packet contains the RID of other routers whose hello packets have already been received over the interface.
- When a device sees its own RID in the hello packet generated by another router, these devices establish a neighbor relationship.

Link state advertisements contain five types of information

- Router LSAs
  - describes the state of the router's interfaces (links) within the area.
- Network LSAs
  - lists the routers connected to a multi-access network.
  - generated by the DR
- summary LSAs describe routes to destinations in other areas within the OSPF network.
- summary LSAs describe routes to ASBRs.
- AS external LSAs
  - describes routes to destinations external to the OSPF network.

five types of LSA information



The LSA header

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |    Options    |    LS type    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The LSA header

- LS type
  - router-LSAs, network-LSAs, summary-LSAs, AS-external-LSAs
- LS age
  - The time in seconds since the LSA was originated.
- Link State ID
  - **This field identifies the portion of the internet environment that is being described by the LSA. The contents of this field depend on the LSA's LS type.**
  - For example, in network-LSAs the Link State ID is set to the IP interface address of the network's Designated Router
- Advertising Router
  - The Router ID of the router that originated the LSA.
- LS sequence number
  - Detects old or duplicate LSAs.
- length
  - This includes the 20 byte LSA header.

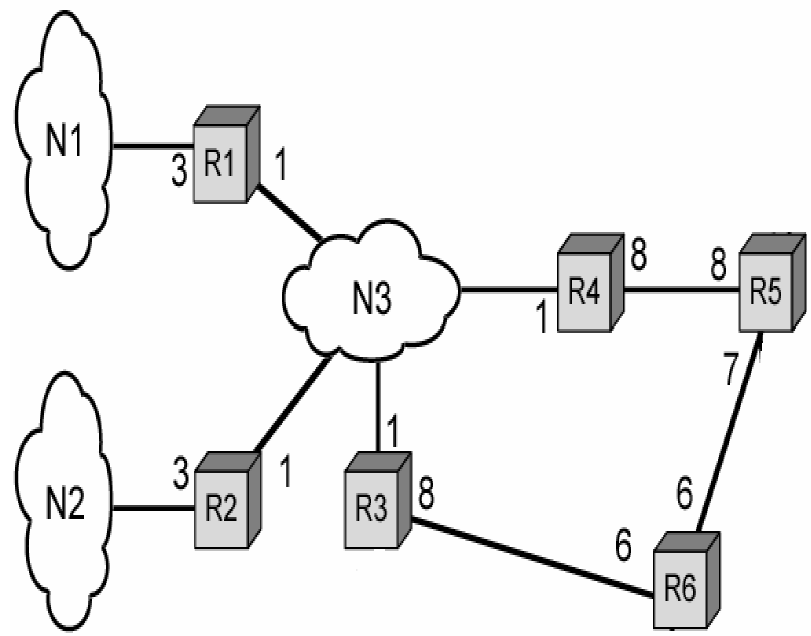
(1) Router-LSAs

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |     Options   |       1       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    0    |V|E|B|        0      |            # links            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Link ID                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Link Data                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     # TOS     |            metric             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      TOS      |        0      |          TOS  metric          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Link ID                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Link Data                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
```

(1) Router-LSAs

- bit E： When set, the router is an AS boundary router (E is for external).
- bit B： When set, the router is an area border router (B is for border).

type	Link ID
Point-to-point	Neighboring router's Router ID
to a transit network	IP address of Designated Router
to a stub network	IP network

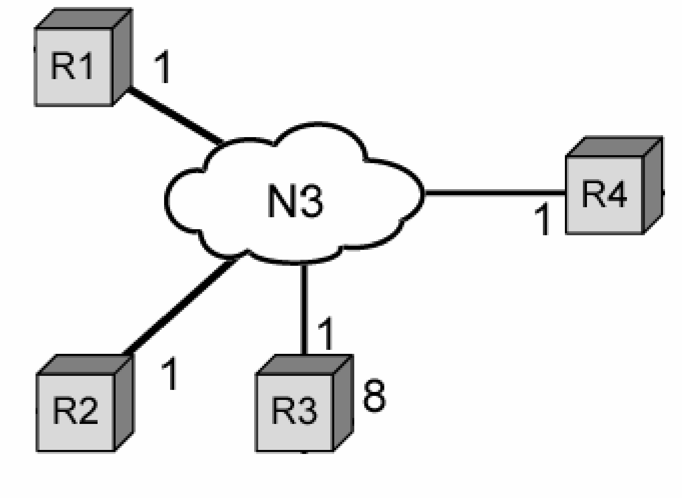


(2) Network-LSAs

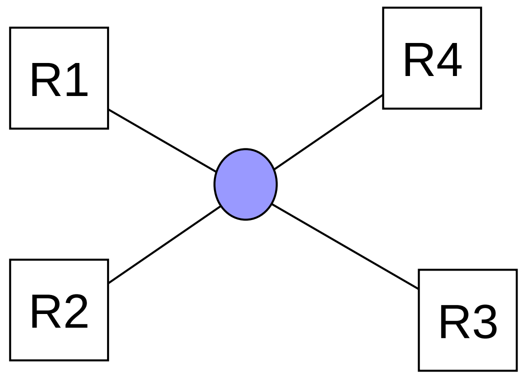
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |      Options  |      2        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Network Mask                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Attached Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
```

(2) Network-LSAs

- The network-LSA is originated by the network's Designated Router.
- The LSA describes all routers attached to the network, including the Designated Router itself.
- The LSA's Link State ID field lists the IP interface address of the Designated Router.



- Network Mask
  - The IP address mask for the network.
- Attached Router
  - The Router IDs of each of the routers attached to the network.



(3) Summary-LSAs

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |     Options   |    3 or 4     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Network Mask                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      0        |                  metric                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     TOS       |                TOS  metric                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
```

(3) Summary-LSAs

- These LSAs are originated by area border routers.
- Summary-LSAs describe inter-area destinations.
- Type 3 summary-LSAs are used when the destination is an IP network.
- When the destination is an AS boundary router, a Type 4 summary-LSA is used, and the Link State ID field is the AS boundary router’s OSPF Router ID.

(4) AS-external-LSAs

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |     Options   |      5        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Network Mask                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|E|     0       |                  metric                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Forwarding address                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      External Route Tag                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|E|    TOS      |                TOS  metric                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Forwarding address                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      External Route Tag                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
```

(4) AS-external-LSAs

- These LSAs are originated by AS boundary routers, and describe destinations external to the AS.
- For these LSAs the Link State ID field specifies an IP network number.

- Network Mask
  - The IP address mask for the advertised destination.
- Metric: The cost of this route.
- Forwarding address: Data traffic for the advertised destination will be forwarded to this address.

The Hello Protocol

- The Hello Protocol is responsible for establishing and maintaining neighbor relationships.
- Hello packets are sent periodically out all router interfaces.
- Bidirectional communication is indicated when the router sees itself listed in the neighbor's Hello Packet.
- On broadcast and NBMA networks, the Hello Protocol elects a Designated Router for the network.

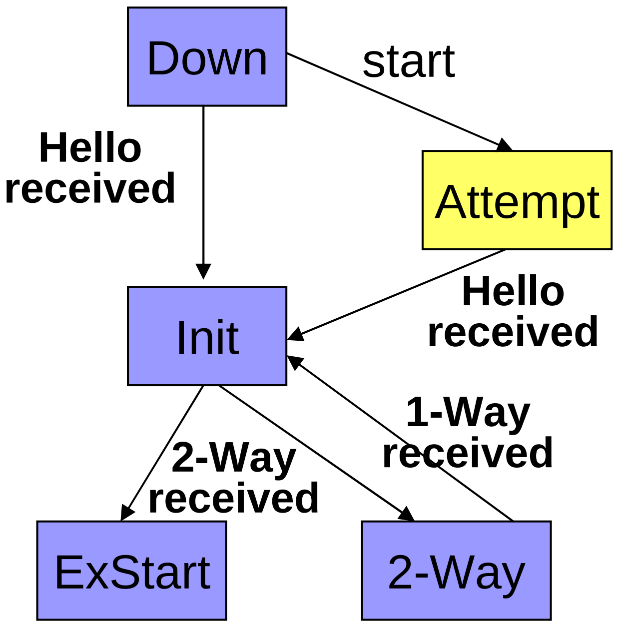
```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Version #   |       1       |         Packet length         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                          Router ID                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                           Area ID                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           Checksum            |             AuType            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Authentication                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Authentication                          |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        Network Mask                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         HelloInterval         |    Options    |    Rtr Pri    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     RouterDeadInterval                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                      Designated Router                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Backup Designated Router                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                          Neighbor                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                              ...                              |
```

- On broadcast networks, each router advertises itself by periodically multicasting Hello Packets.
- This allows neighbors to be discovered dynamically.
- These Hello Packets contain the router's view of the Designated Router's identity, and the list of routers whose Hello Packets have been seen recently.
- All routers connected to a common network must agree on certain parameters (Network mask, HelloInterval and RouterDeadInterval).

- On NBMA networks some configuration information may be necessary for the operation of the Hello Protocol.
- Each router that may potentially become Designated Router has a list of all other routers attached to the network.
- A router, having Designated Router potential, sends Hello Packets to all other potential Designated Routers when its interface to the NBMA network first becomes operational.
- This is an attempt to find the Designated Router for the network.
- If the router itself is elected Designated Router, it begins sending Hello Packets to all other routers attached to the network.

Neighbor states（1）

- Down
  - the initial state
  - there has been no recent information received from the neighbor.
- Attempt
  - This state is only valid for neighbors attached to NBMA networks. It indicates that no recent information has been received from the neighbor, but that a more concerted effort should be made to contact the neighbor. This is done by sending the neighbor Hello packets at intervals of HelloInterval.



- Init: **In this state, an Hello packet has recently been seen from the neighbor.**
- ExStart: **In this state, an Hello packet has recently been seen from the neighbor.**
- 2-Way: communication between the two routers is bidirectional.

Events causing neighbor state（1） changes

- HelloReceived
- Start
  - This is an indication that Hello Packets should now be sent to the neighbor at intervals of HelloInterval seconds. This event is generated only for neighbors associated with NBMA networks.
- 2-Way Received
  - This is indicated by the router seeing itself in the neighbor's Hello packet.
- 1-Way Received
  - An Hello packet has been received from the neighbor, in which the router is not mentioned. This indicates that communication with the neighbor is not bidirectional.

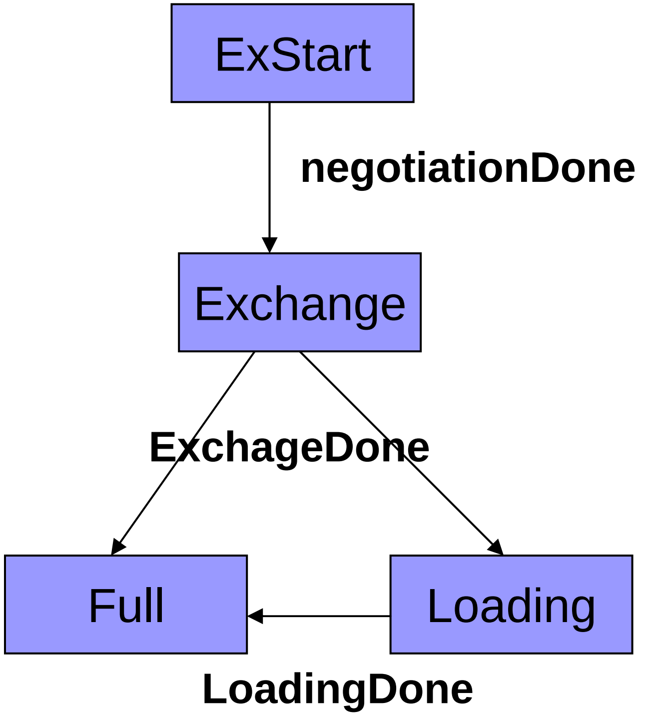
The Synchronization of Databases

- In a link-state routing algorithm, it is very important for all routers' link-state databases to stay synchronized.
- OSPF simplifies this by requiring only adjacent routers to remain synchronized.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Version #   |       2       |         Packet length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Router ID                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Area ID                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Checksum            |             AuType            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Authentication                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Authentication                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Interface MTU         |    Options    |0|0|0|0|0|I|M|MS
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     DD sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            LS age             |    Options    |    LS type    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Link State ID                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Advertising Router                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     LS sequence number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         LS checksum           |             length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              ...                              |
```

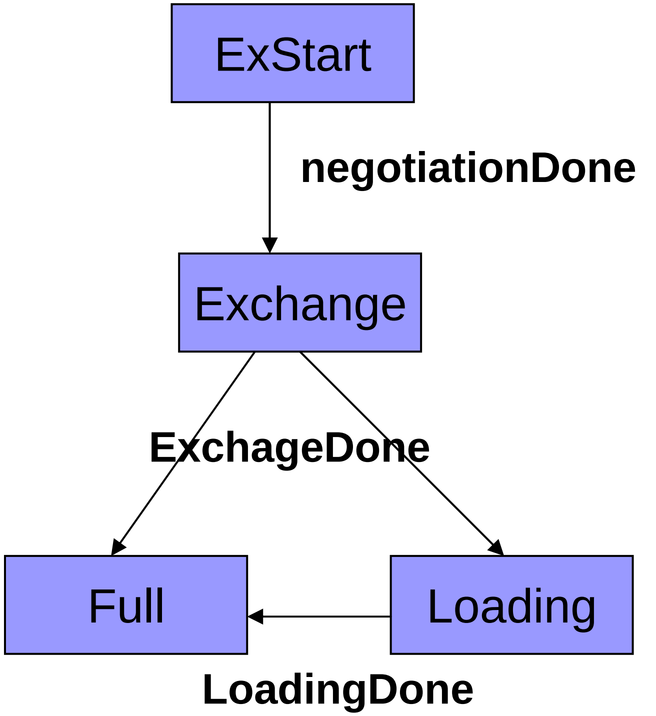
Neighbor states（2）

- Exchange
  - **In this state the router is describing its entire link state database by sending Database Description packets to the neighbor.**
- Loading
  - **In this state, Link State Request packets are sent to the neighbor asking for the more recent LSAs that have been discovered (but not yet received) in the Exchange state.**
- Full
  - In this state, the neighboring routers are fully adjacent. These adjacencies will now appear in router-LSAs and network-LSAs.



Events causing neighbor state changes(2)

- NegotiationDone
  - The Master/Slave relationship has been negotiated, and DD sequence numbers have been exchanged. This signals the start of the sending/receiving of Database Description packets.
- ExchangeDone
- Loading Done



```
+---+                                                   +---+
|RT1|                                                   |RT2|
+---+                                                   +---+

Down                                                    Down
                 Hello(DR=0,seen=0)
           ------------------------------>
              Hello (DR=RT2,seen=RT1,...)                Init
           <------------------------------
ExStart         D-D (Seq=x,I,M,Master)
           ------------------------------>
                D-D (Seq=y,I,M,Master)                   ExStart
           <------------------------------
Exchange        D-D (Seq=y,M,Slave)
           ------------------------------>
                D-D (Seq=y+1,M,Master)                   Exchange
           <------------------------------
                D-D (Seq=y+1,M,Slave)
           ------------------------------>
                          ...
                          ...
                          ...
                D-D (Seq=y+n, Master)
           <------------------------------
                D-D (Seq=y+n, Slave)
 Loading   ------------------------------>
                      LS Request                          Full
           ------------------------------>
                      LS Update
           <------------------------------
                      LS Request
           ------------------------------>
                      LS Update
           <------------------------------
 Full
```

Border Gateway Protocol (BGP)

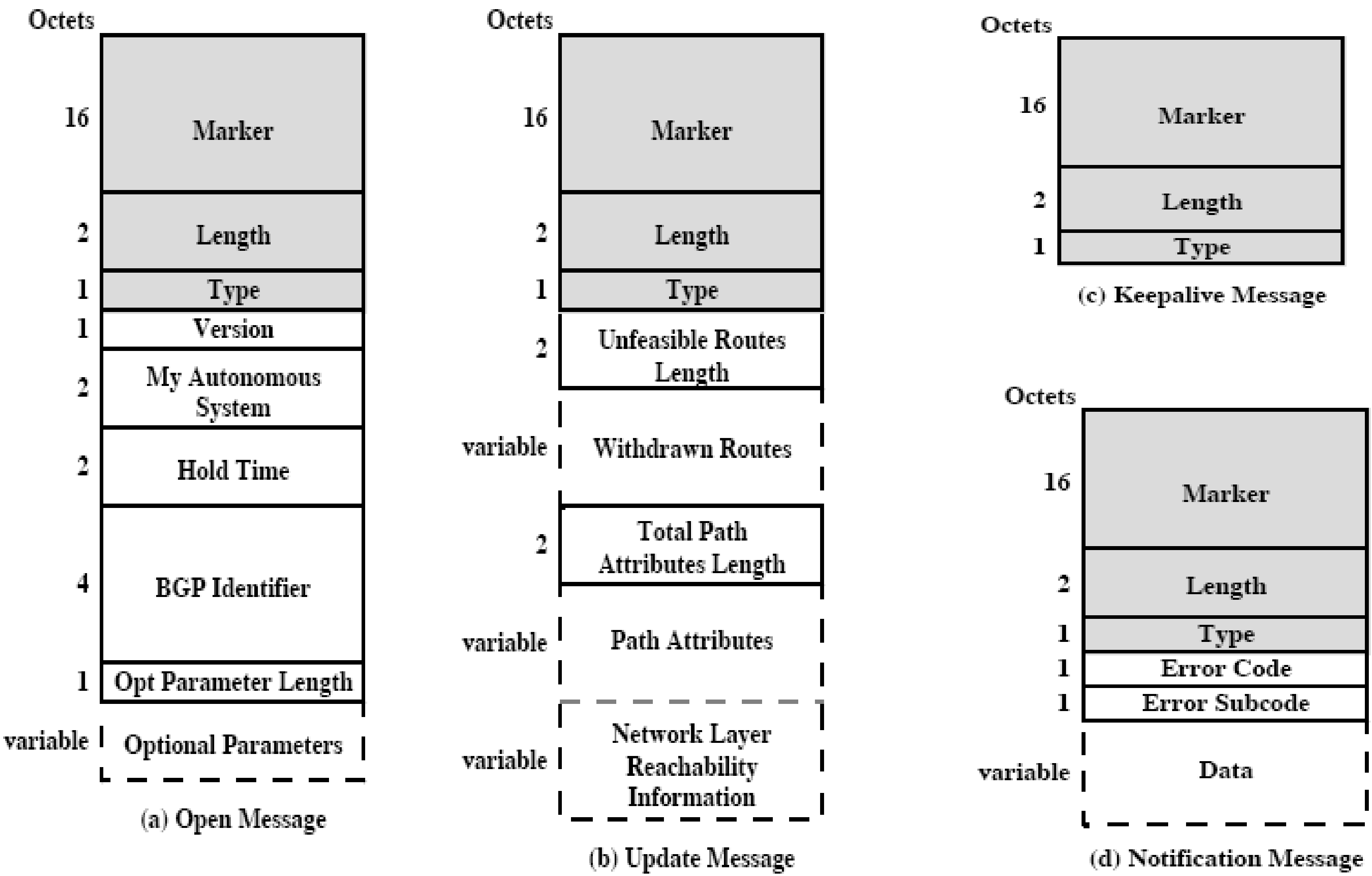
Border Gateway Protocol (BGP)

- For use with TCP/IP internets
- BGP messages are sent over TCP connections
- BGP messages
  - Open: opens TCP connection to peer and authenticates sender
  - Keep-alive: (1) ACKs OPEN request; (2) keeps connection alive in absence of UPDATES
  - Update: (1) advertises new path; (2) withdraws old
  - Notification: (1) closes connection; (2) reports errors in previous msg

Procedures of BGP

- **Neighbor acquisition**
  - One router sends an *Open* message to another
  - If the target router accepts the request, it returns a *Keep-alive* **message**
- **Neighbor reachability**
  - The two routers periodically issue *Keep-alive* or *Update* messages to each other
- **Network reachability**
  - Each router maintains a database of networks
  - That it can reach and the list of *ASs* passed
  - The router issues an *Update* message whenever a change is made to this database

BGP Messages

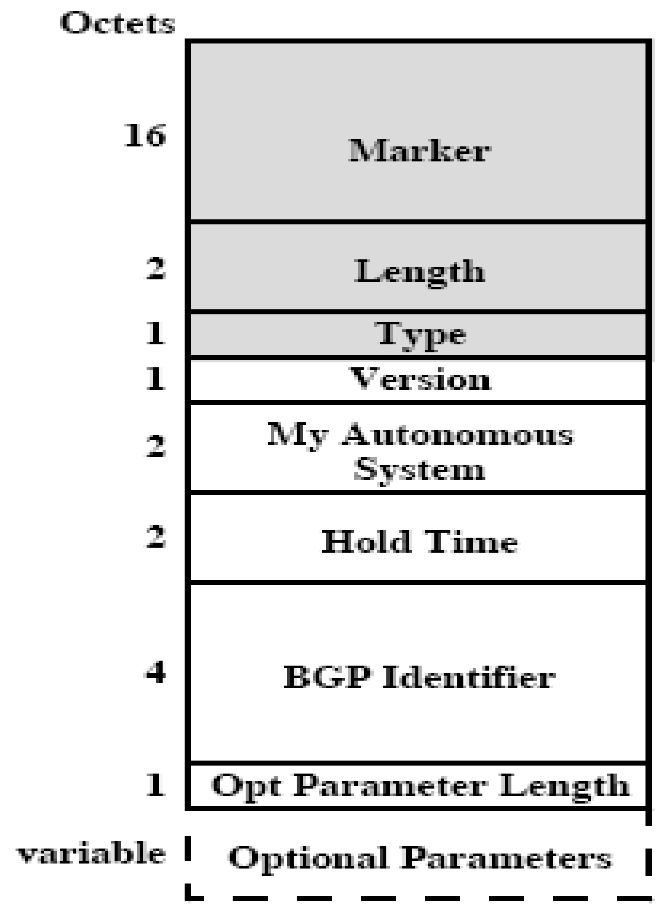


BGP Messages

- 3 common fixed-size fields in each header
- Marker (16 octets)
  - Detect loss of synchronization between a pair of BGP speaker
  - Authenticate incoming BGP messages
- Length (2 octets)
  - Length of message in octets, including the header
- Type (1 octets)
  - 1.Open, 2.Update, 3.Notification, 4.Keep-alive

Open Message

- Version (1 octet)
  - Current BGP version (v4)
- My Autonomous System (2 octets)
  - Identification of AS the sender belongs to
- Hold time (2 octets)
  - Max time between Keep-alive and/or update messages
- BGP Identifier (4 octets)
  - Identifier of the sender
- Opt parameter length (1 octet)
  - Total length of the Optional parameter field in octet

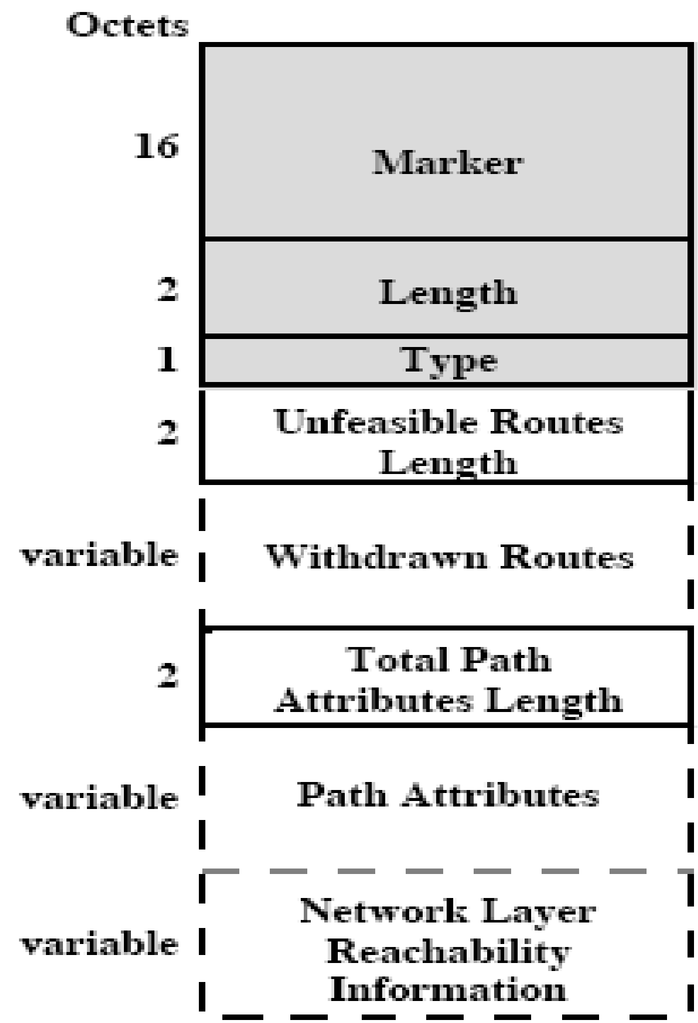


(a) Open Message

8	16	
Parm. Type	Parm. Length	Parameter Value (Variable)

Update Message (1)

- Unfeasible Routes Length (2 octets)
  - Total length of withdraw routes in octets
- Withdrawn route (variable length)
  - A list of IP address prefixes, 2-tuple of the form <length, prefix>
  - Each prefix identifies a network
  - e.g. <10, D8CA> means 16 bits length, 216.202.0.0 network
- Total Path Attribute Length (2 octets)
  - Total length of path attribute field in octets



(b) Update Message

Update Message (2)

- Path Attribute (variable length)
  - A list of path attributes, each path attribute is a triple <attribute type, attribute length, attribute value>
  - Attributes that apply to the particular router or route
- Network Layer Reachability Information (variable length)
  - A list of IP address prefixes, each one is 2-tuple of the form <length, prefix>
  - A single route through the internet

Defined Path Attributes (1)

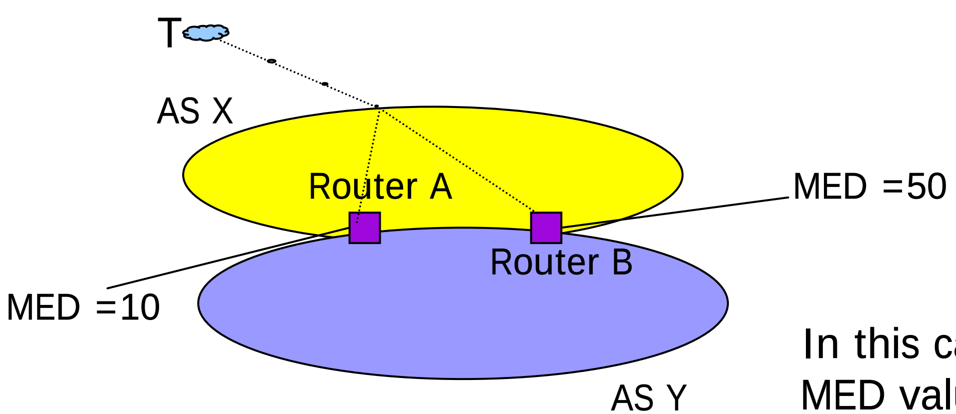
- **Well-known mandatory**
  - The attribute must be recognized by all BGP implementations. It must be sent in every UPDATE message.
- **Well-known discretionary**
  - The attribute must be recognized by all BGP implementations. However, it is not required to be sent in every UPDATE message.
- **Optional transitive**
  - It is not required that every BGP implementation recognize this type of attribute. A path with an unrecognized optional transitive attribute is accepted and simply forwarded to other BGP peers.
- **Optional non-transitive**
  - It is not required that every BGP implementation recognize this type of attribute. These attributes can be ignored and not passed along to other BGP peers.

Defined Path Attributes (2)

- Origin （ **Well-known mandatory** ）
  - Learned from IGP or EGP
- AS_Path （ **Well-known mandatory** ）
  - A list of AS traversed, in ordered or unordered way
  - Enables routing policy, such as security, performance, QOS, number of ASs, etc.
- Next_hop （ **Well-known mandatory** ）
  - IP address of the border router that are used as the next hop
  - Not all routers implement BGP
  - Responsible for informing outside routers of the route to other networks

Defined Path Attributes (3)

- Multi_Exit_Disc (MED)
  - There may be multiple border points in one *AS* available to another *AS*
  - MED is a metric value computed by certain routing policy within the *AS*
  - It may be used by another BGP router to discriminate among multiple exit points



In this case, it selects route used router A. Because MED value of router A is lower than router B's MED

Defined Path Attributes (4)

- Local_pref
  - Should be included when the 2 BGP speakers located within the same AS
  - It is used by a BGP speaker to inform other BGP speakers in its own autonomous system of the originating speaker's degree of preference for an advertised route.
- Atomic_Aggregate
  - Informs others that the local system selected a more general route without specifying some interim specific routes
- Aggregator
  - Contains the last AS number and IP address of the BGP router that formed the aggregate route

Keep Alive Message

- To tell other routers that this router is still here
- BGP speaker send *Keep-Alive* message periodically to keep connection

Notification Message (1)

- Message header error
  - Authentication and syntax, subtypes:
  - Connection Not Synchronized
  - Bad Message Length
  - Bad Message Type
- Open message error
  - Syntax and option not recognized, Unacceptable hold time, subtypes:
  - Unsupported Version Number
  - Bad peer AS
  - Bad BGP identifier
  - Unsupported Optional Parameter, …

Notification Message (2)

- Update message error
  - Syntax and validity errors
- Hold time expired
  - Connection is closed
- Finite state machine error
  - Any procedural errors: wrong message at wrong states
  - e.g. got *Open* message at *Connect* state
- Cease
  - Used to close a connection when there is no error

BGP Routing Information Exchange

- Within AS, router builds topology picture using IGP
- Router issues Update message to other routers outside AS using BGP
- These routers exchange info with other routers in other AS
- Routers must then decide best routes