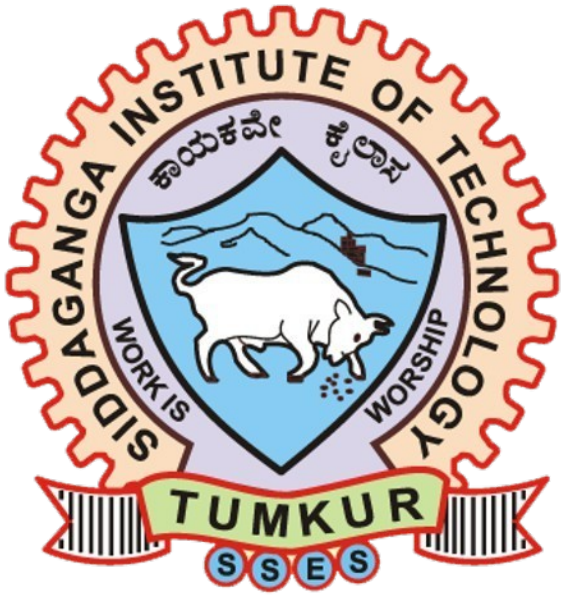
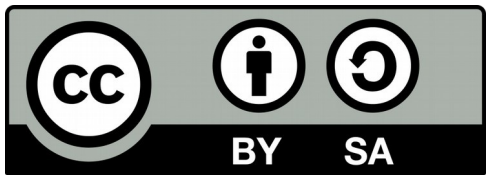
UNICAST ROUTING PROTOCOLS

SIDDAGANGA INSTITUTE OF TECHNOLOGY

Department of CSE

Prabodh C P





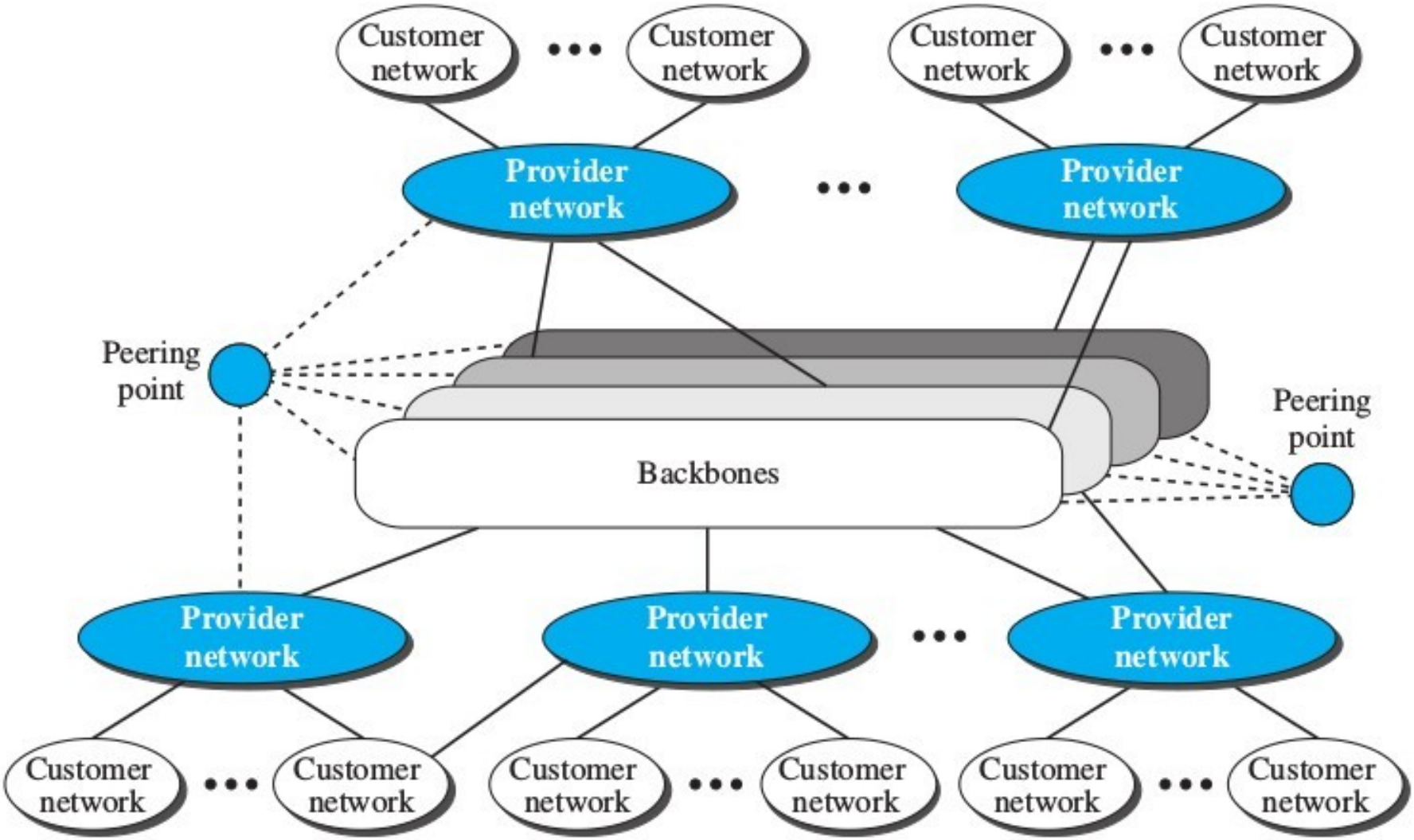
Contents

- UNICAST ROUTING PROTOCOLS
  - Internet Structure
  - Routing Information Protocol (RIP)
  - Open Shortest Path First(OSPF)
  - Border Gateway Protocol Version 4 (BGP4)

UNICAST ROUTING PROTOCOLS

- we discuss unicast routing protocols used in the Internet
- A protocol is more than an algorithm.
- A protocol needs to define its
  - domain of operation,
  - the messages exchanged,
  - communication between routers,
  - interaction with protocols in other domains.
- The three common protocols used in the Internet:
  - Routing Information Protocol (RIP), based on the distance-vector algorithm
  - Open Shortest Path First (OSPF), based on the link-state algorithm
  - Border Gateway Protocol (BGP), based on the path-vector algorithm.

Internet Structure



Internet Structure

- There are several **backbones** run by private communication companies that provide global connectivity.
- These backbones are connected by some **peering points** that allow connectivity between backbones.
- At a lower level, there are some **provider networks** that use the backbones for global connectivity but provide services to Internet customers.
- There are some **customer networks** that use the services provided by the provider networks.
- All the above can be called an Internet Service Provider or ISP. They provide services, but at different levels.

Hierarchical Routing

- It is obvious that routing in the Internet cannot be done using a single protocol for two reasons: a scalability problem and an administrative issue.
- Scalability problem
  - means that the size of the forwarding tables becomes huge,
  - searching for a destination in a forwarding table becomes time-consuming, and updating creates a huge amount of traffic.
- The administrative issue is each ISP is run by an administrative authority.
- The administrator needs to have control in its system.
- The organization must be able to use as many subnets and routers as it needs, may desire that the routers be from a particular manufacturer, may wish to run a specific routing algorithm to meet the needs of the organization, and may want to impose some policy on the traffic passing through its ISP.

Hierarchical Routing

- Hierarchical routing means considering each ISP as an autonomous system (AS).
- Each AS can run a routing protocol that meets its needs, but the global Internet runs a global protocol to glue all ASs together.
- The routing protocol run in each AS is referred to as intra-AS routing protocol, **intradomain** routing protocol, or interior gateway protocol (IGP)
  - **RIP and OSPF**
- The global routing protocol is referred to as inter-AS routing protocol, **interdomain** routing protocol, or exterior gateway protocol (EGP).
  - **BGP**

Autonomous Systems

- Each ISP is an autonomous system
- each AS is given an autonomous number (ASN) by the ICANN.
- Each ASN is a 16-bit unsigned integer that uniquely defines an AS.
- The autonomous systems, however, are not categorized according to their size; they are categorized according to the way they are connected to other ASs.
- The autonomous systems are classified as
  - Stub AS.
  - Multihomed AS.
  - Transient AS.

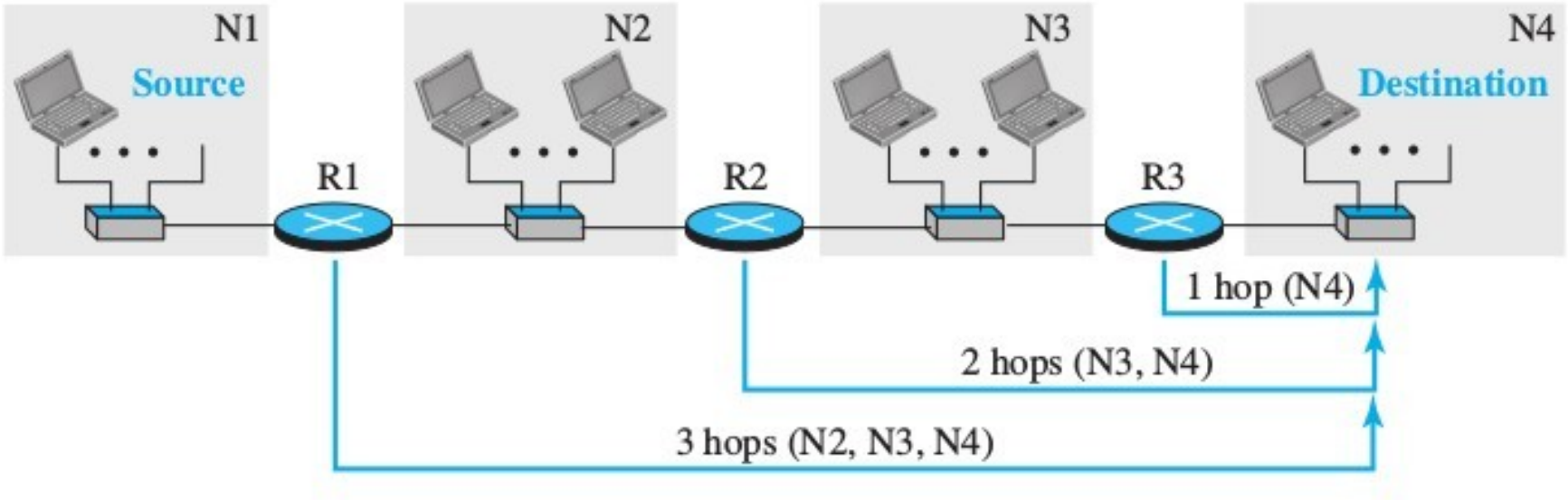
Autonomous Systems

- **Stub AS.**
  - A stub AS has only one connection to another AS. The data traffic can be either initiated or terminated in a stub AS; the data cannot pass through it. A good example of a stub AS is the customer network, which is either the source or the sink of data.
- **Multihomed AS.**
  - A multihomed AS can have more than one connection to other ASs, but it does not allow data/transit traffic to pass through it. Eg : Large corporation
- **Transient AS.**
  - A transient AS is connected to more than one other AS and also allows the traffic to pass through. The provider networks and the backbone are good examples of transient ASs.

Routing Information Protocol (RIP)

- The Routing Information Protocol (RIP) is one of the most widely used intradomain routing protocols based on the distance-vector routing algorithm.
- RIP was started as part of the Xerox Network System (XNS), but it was the Berkeley Software Distribution (BSD) version of UNIX that helped make the use of RIP widespread.

Hop Count



- a router in an AS needs to know how to forward a packet to different networks (subnets) in an AS, RIP routers advertise the cost of reaching different networks instead of reaching other nodes in a theoretical graph.
- In other words, the cost is defined between a router and the network in which the destination host is located.
- Second, to make the implementation of the cost simpler the cost is defined as the number of hops

Forwarding Tables

Forwarding table for R1

Destination network	Next router	Cost in hops
N1	——	1
N2	——	1
N3	R2	2
N4	R2	3

Forwarding table for R2

Destination network	Next router	Cost in hops
N1	R1	2
N2	——	1
N3	——	1
N4	R3	2

Forwarding table for R3

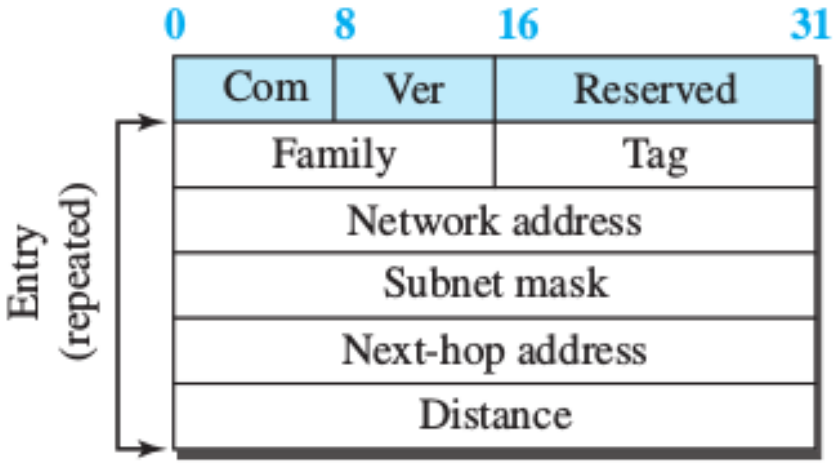
Destination network	Next router	Cost in hops
N1	R2	3
N2	R2	2
N3	——	1
N4	——	1

- In DVR we exchange Distance vectors but in RIP we exchange forwarding tables
- A forwarding table in RIP is a three-column table in which
  - the first column is the address of the destination network,
  - the second column is the address of the next router to which the packet should be forwarded,
  - and the third column is the cost (the number of hops) to reach the destination network.
- Least cost tree info from R1 to N4
- Third column required for updating forwarding table

RIP Implementation

- RIP is implemented as a process that uses the service of UDP on the well-known port number 520.
- In BSD, RIP is a daemon process (a process running in the background), named routed
- RIP is a routing protocol to help IP route its datagrams through the AS, the RIP messages are encapsulated inside UDP user datagrams, which in turn are encapsulated inside IP datagrams.
- In other words, RIP runs at the application layer, but creates forwarding tables for IP at the network later.
- RIP has gone through two versions: RIP-1 and RIP-2. The second version is backward compatible with the first section;

RIP Messages



Fields

Com: Command, request (1), response (2)
Ver: Version, current version is 2
Family: Family of protocol, for TCP/IP value is 2
Tag: Information about autonomous system
Network address: Destination address
Subnet mask: Prefix length
Next-hop address: Address length
Distance: Number of hops to the destination

RIP Messages

- RIP has two types of messages: request and response.
- 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 message can ask about specific entries or all entries.
- A response (or update)message can be either solicited or unsolicited.
- A solicited response message is sent only in answer to a request message.
- It contains information about the destination specified in the corresponding request message.
- An unsolicited response message, on the other hand, is sent periodically, every 30 seconds or when there is a change in the forwarding table.

RIP Algorithm

RIP implements the same algorithm as the distance-vector routing algorithm with modifications to enable a router to update its forwarding table:

- Instead of sending only distance vectors, a router needs to send the whole contents of its forwarding table in a response message.
- The receiver adds one hop to each cost and changes the next router field to the address of the sending router.
  - We call each route in the modified forwarding table the received route and each route in the old forwarding table the old route.

RIP Algorithm

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

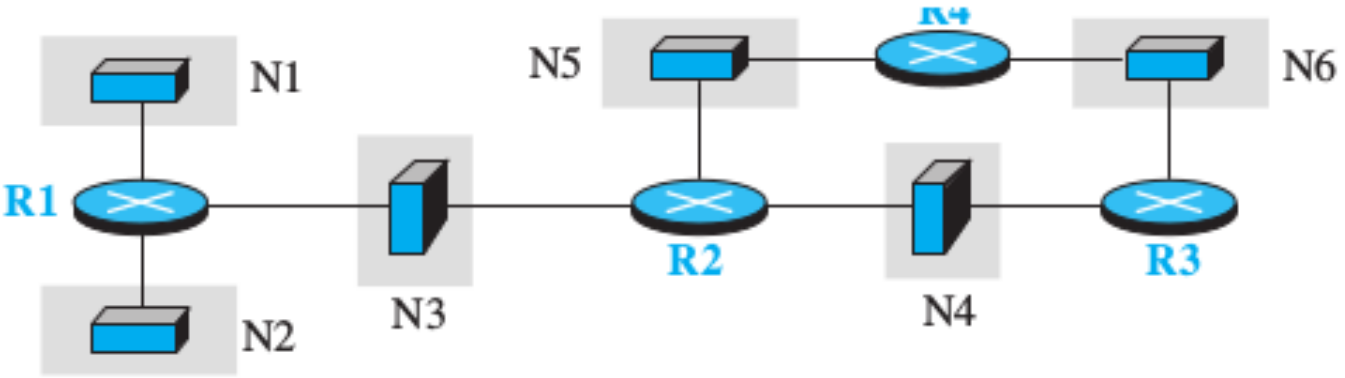
  1) If the received route does not exist in the old forwarding table, it should be added to the route.

  2) If the cost of the received route is lower than the cost of the old one, the received route should be selected as the new one.

  3) If the cost of the received route is higher than the cost of the old one, but the value of the next router is the same in both routes, the received route should be selected as the new one. This is the case where the route was actually advertised by the same router in the past, but now the situation has been changed.

- The new forwarding table needs to be sorted according to the destination route (mostly using the longest prefix first).

Example of an autonomous system using RIP



R1

Des.	N. R.	Cost
N1	——	1
N2	——	1
N3	——	1

R2

Des.	N. R.	Cost
N3	——	1
N4	——	1
N5	——	1

R3

Des.	N. R.	Cost
N4	——	1
N6	——	1

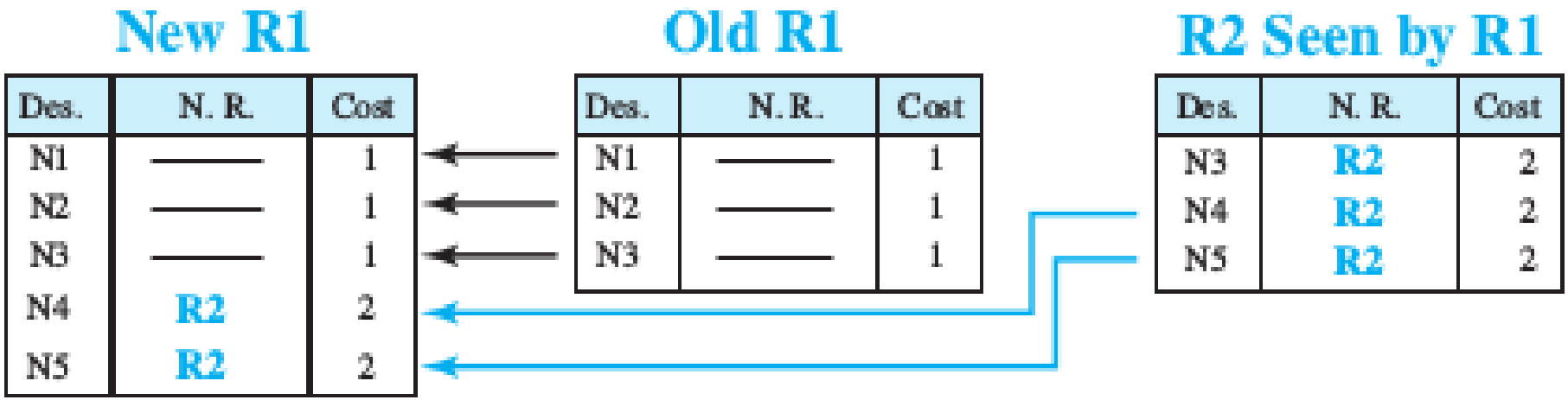
R4

Des.	N. R.	Cost
N5	——	1
N6	——	1

Forwarding tables after all routers booted

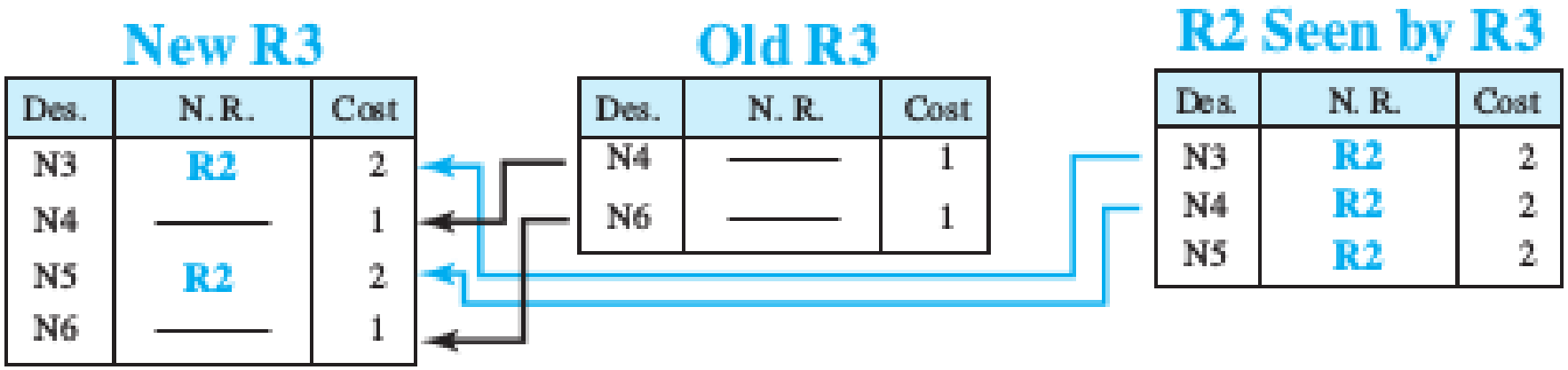
Updating Forwarding Tables

R1 after receiving update from R2



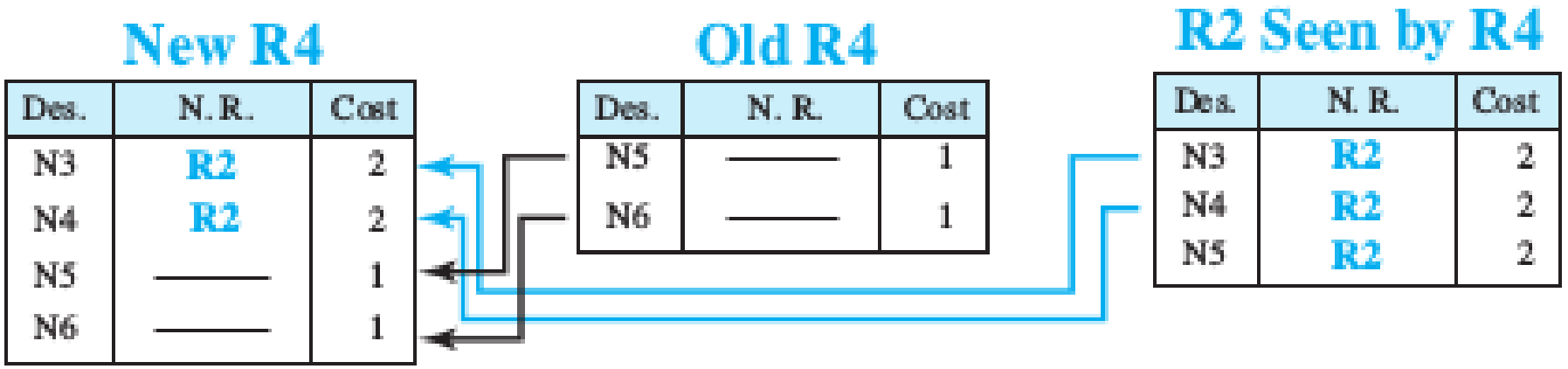
Updating Forwarding Tables

R3 after receiving update from R2



Updating Forwarding Tables

R4 after receiving update from R2



Stabilized Forwarding tables

Final R1

Des.	N. R.	Cost
N1	———	1
N2	———	1
N3	———	1
N4	R2	2
N5	R2	2
N6	R2	3

Final R2

Des.	N. R.	Cost
N1	R1	2
N2	R1	2
N3	———	1
N4	———	1
N5	———	1
N6	R3	2

Final R3

Des.	N. R.	Cost
N1	R2	3
N2	R2	3
N3	R2	2
N4	———	1
N5	R2	2
N6	———	1

Final R4

Des.	N. R.	Cost
N1	R2	3
N2	R2	3
N3	R2	2
N4	R2	2
N5	———	1
N6	———	1

Timers in RIP

RIP uses three timers to support its operation.

The **periodic timer** controls the advertising of regular update messages.

Each router has one periodic timer that is randomly set to a number between 25 and 35 seconds . The timer counts down; when zero is reached, the update message is sent, and the timer is randomly set once again.

The **expiration timer** governs the validity of a route. When a router receives update information for a route, the expiration timer is set to 180 seconds for that particular route. Every time a new update for the route is received, the timer is reset.

If there is a problem on an internet and no update is received within the allotted 180 seconds, the route is considered expired and the hop count of the route is set to 16, which means the destination is unreachable.

Timers in RIP

The **garbage collection timer** is used to purge a route from the forwarding table. When the information about a route becomes invalid, the router does not immediately purge that route from its table.

Instead, it continues to advertise the route with a metric value of 16.

At the same time, a garbage collection timer is set to 120 seconds for that route.

When the count reaches zero, the route is purged from the table.

This timer allows neighbors to become aware of the invalidity of a route prior to purging.

Performance

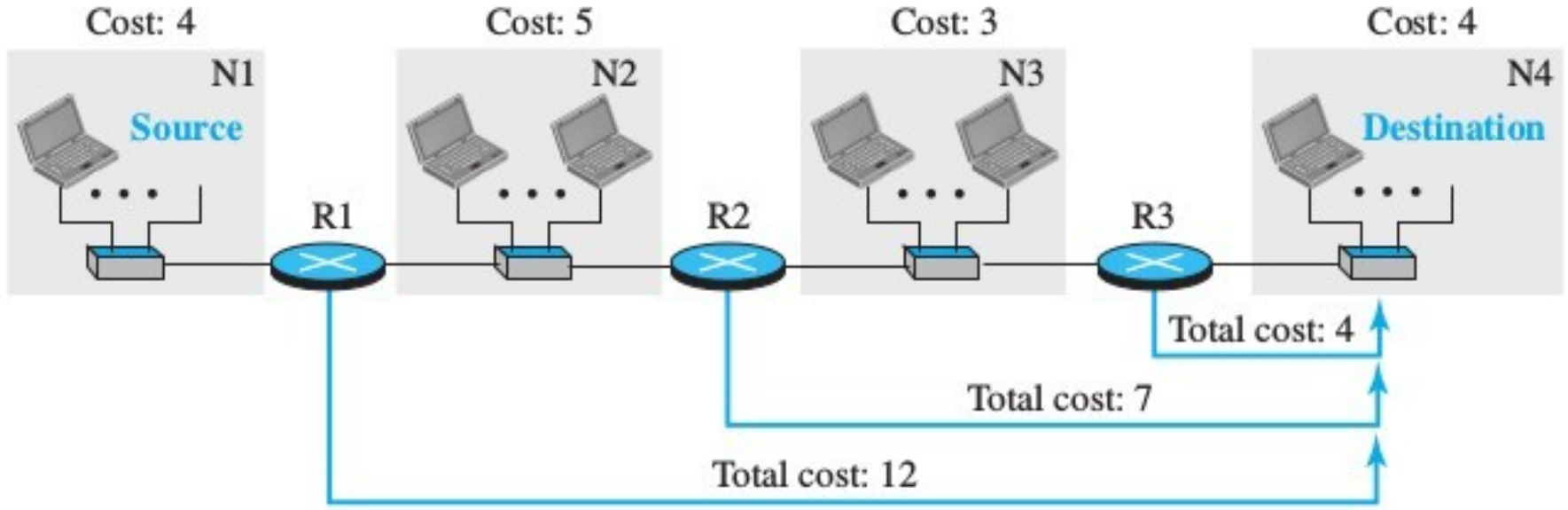
- **Update Messages:** The update messages in RIP have a very simple format and are sent only to neighbors; they are local.
- They do not normally create traffic because the routers try to avoid sending them at the same time.
- **Convergence of Forwarding Tables:** RIP uses the distance-vector algorithm, which can converge slowly if the domain is large, but, since RIP allows only 15 hops in a domain (16 is considered as infinity), there is normally no problem in convergence.
- **Robustness:** Can tolerate failures in the network

Open Shortest Path First (OSPF)

OSPF is also an intradomain routing protocol like RIP, but it is based on the link-state routing protocol

Metric

- cost of reaching a destination from the host is calculated from the source router to the destination network.
- each link (network) can be assigned a weight based on the throughput, round-trip time, reliability, and so on.
- Or even the hop count



Forwarding Tables in OSPF

- Each OSPF router can create a forwarding table after finding the shortest-path tree between itself and the destination using Dijkstra's algorithm
- Forwarding tables are same as RIP except for the cost values
- Because both RIP and OSPF uses shortest path trees in routing

Forwarding table for R1

Destination network	Next router	Cost
N1	——	4
N2	——	5
N3	R2	8
N4	R2	12

Forwarding table for R2

Destination network	Next router	Cost
N1	R1	9
N2	——	5
N3	——	3
N4	R3	7

Forwarding table for R3

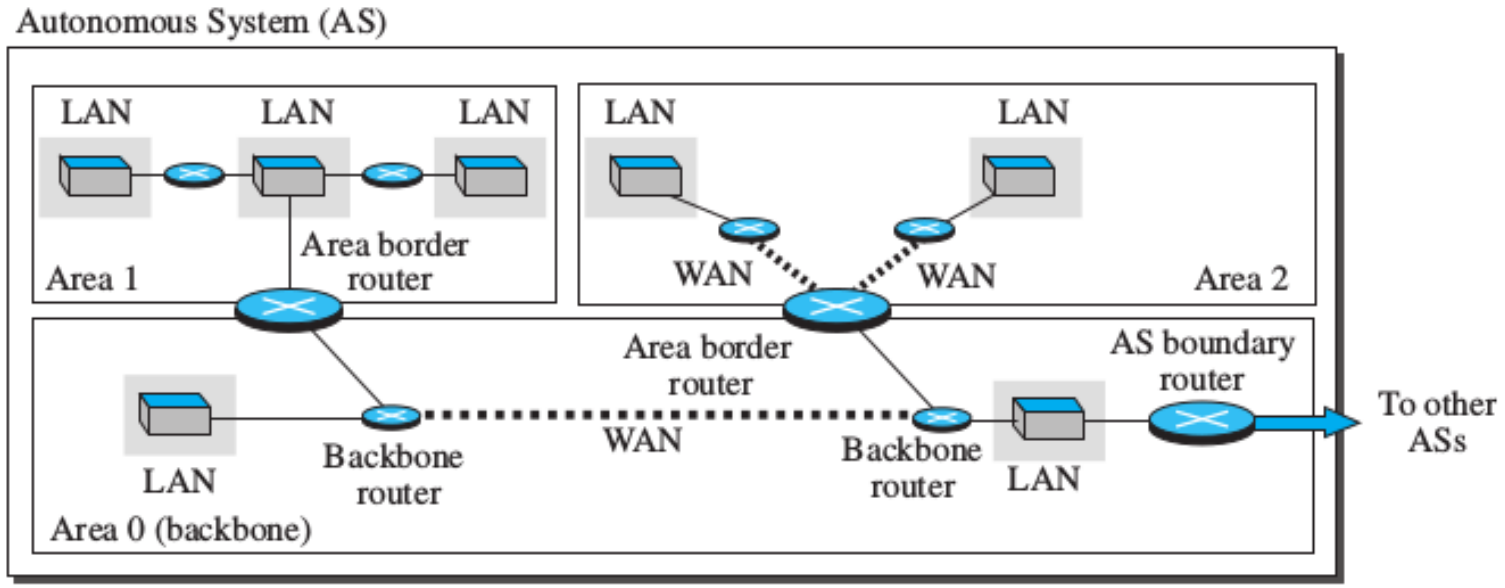
Destination network	Next router	Cost
N1	R2	12
N2	R2	8
N3	——	3
N4	——	4

Areas

- Unlike RIP, OSPF was designed to be able to handle routing in a small or large autonomous system.
- For create the global LSDB, OSPF requires flooding of LSPs
- This may create a huge traffic in a large AS.
- So an AS is divided into **areas** - small independent domain for flooding LSPs.
- Two level Hierarchy
  - AS
  - Areas
- each router in an area needs to know the information about the link states not only in its area but also in other areas.

Areas

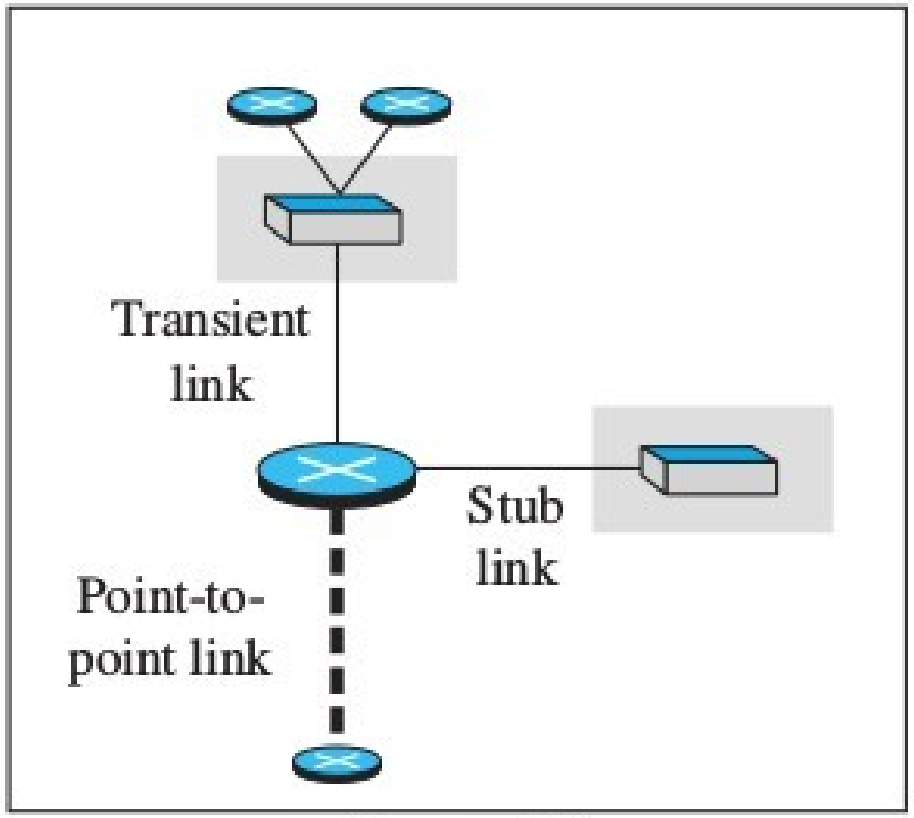
- one of the areas in the AS is designated as the **backbone area**
- The routers in the backbone area are responsible for passing the information collected by each area to all other areas.
- Each area has an area identification.
- The area identification of the backbone is zero



Link-State Advertisement

- In OSPF a router advertise the state of each link to all neighbors for the formation of the LSDB.
- we need different types of advertisements, each capable of advertising different situations.
- We can have five types of link-state advertisements:
  - Router link
  - Network link
  - Summary link to network
  - Summary link to AS border router
  - External link.

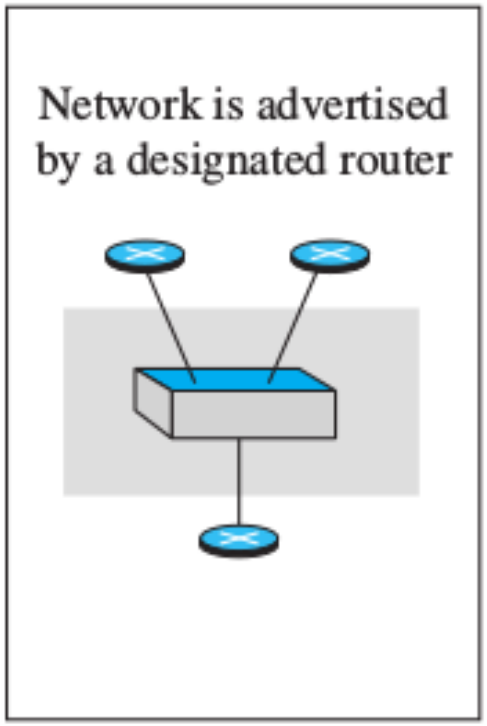
Router link



a. Router link

- Router link advertises the existence of a router as a node.
- In addition to giving the address of the announcing router, this type of advertisement can define one or more types of links that connect the advertising router to other entities.
- A transient link announces a link to a transient network
- A stub link advertises a link to a stub network
- A point-to-point link should define the address of the router at the end of the point-to-point line

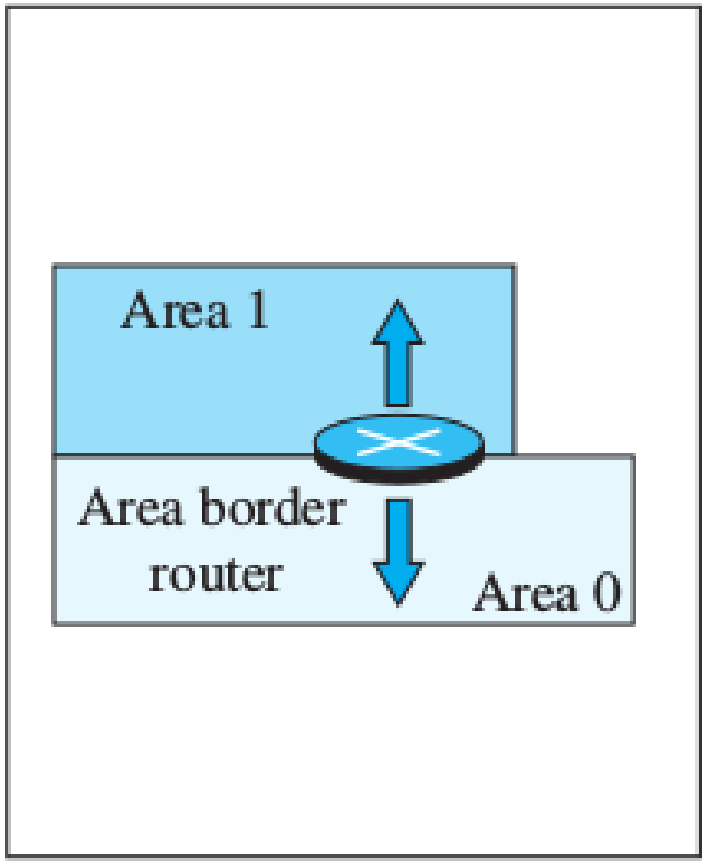
Network link



b. Network link

- A network link advertises the network as a node.
- However, since a network cannot do announcements itself (it is a passive entity), one of the routers is assigned as the designated router and does the advertising.
- It also announces the IP address of all routers but not the cost.

Summary link to network

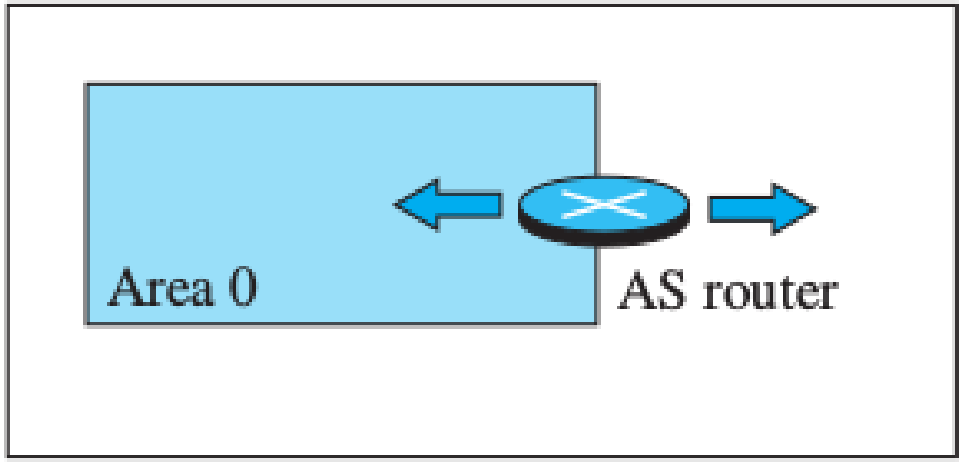


c. Summary link to network

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

Summary link to AS

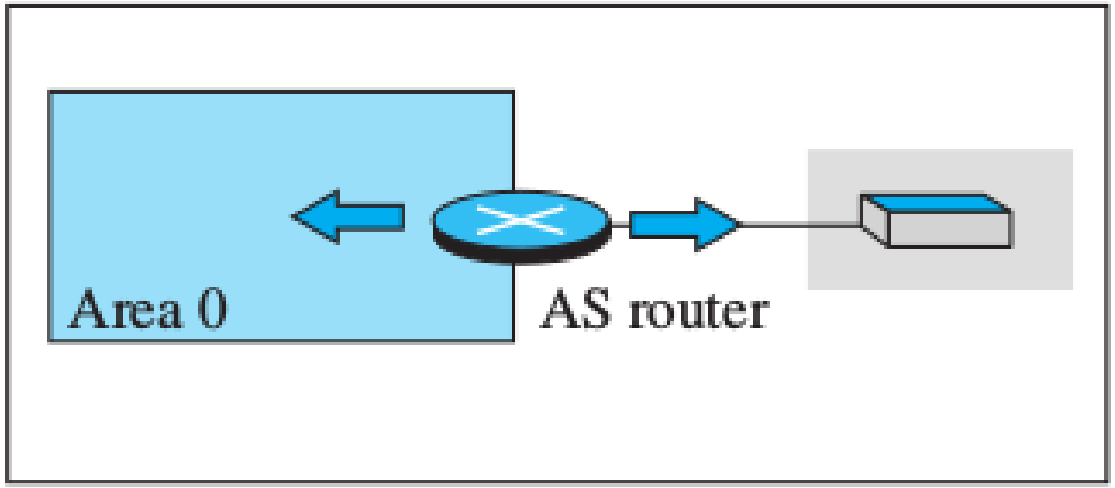
This is done by an AS router that advertises the summary links from other ASs to the backbone area of the current AS, information which later can be disseminated to the areas so that they will know about the networks in other ASs.



d. Summary link to AS

External link

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



e. External link

OSPF Implementation

OSPF is implemented as a program in the network layer, using the service of the IP for propagation.

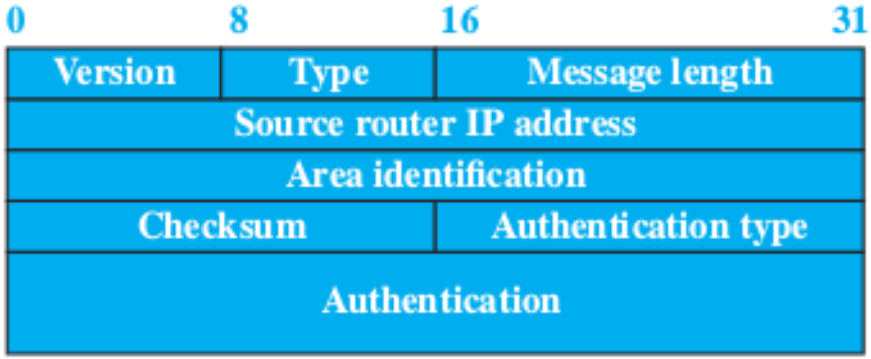
An IP datagram that carries a message from OSPF sets the value of the protocol field to 89.

This means that, although OSPF is a routing protocol to help IP to route its datagrams inside an AS, the OSPF messages are encapsulated inside datagrams.

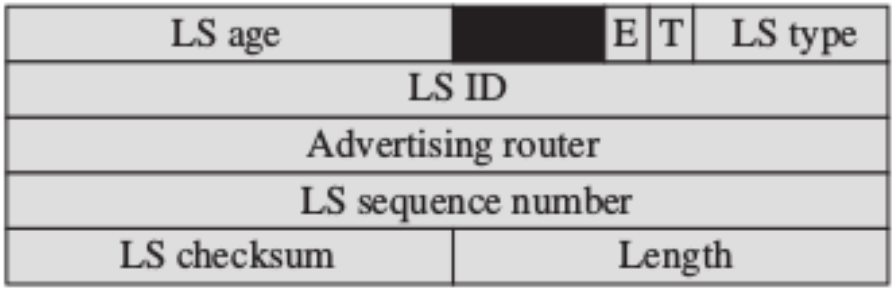
OSPF has gone through two versions: version 1 and version 2.

Most implementations use version 2.

OSPF Headers



OSPF common header



Link-state general header

- The OSPF common header is used in all messages
- Link-state general header (which is used in some messages).

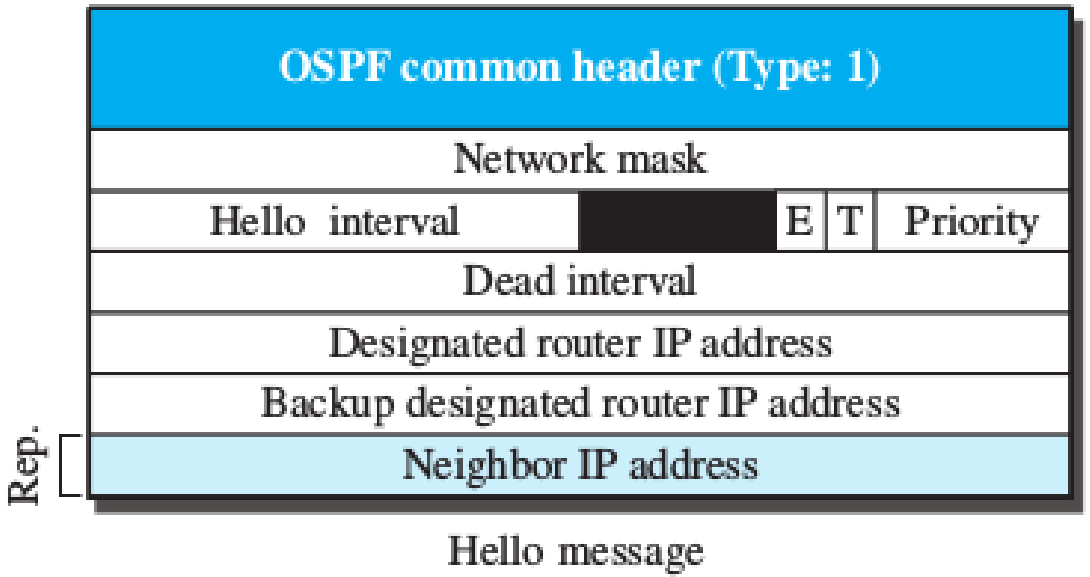
OSPF Messages

OSPF uses five different types of messages.

- The Hello message (type 1)
- Database description message (type 2)
- Link-state request message (type 3)
- Link-state update message (type 4)
- Link-state acknowledgment message (type 5)

The Hello message (type 1)

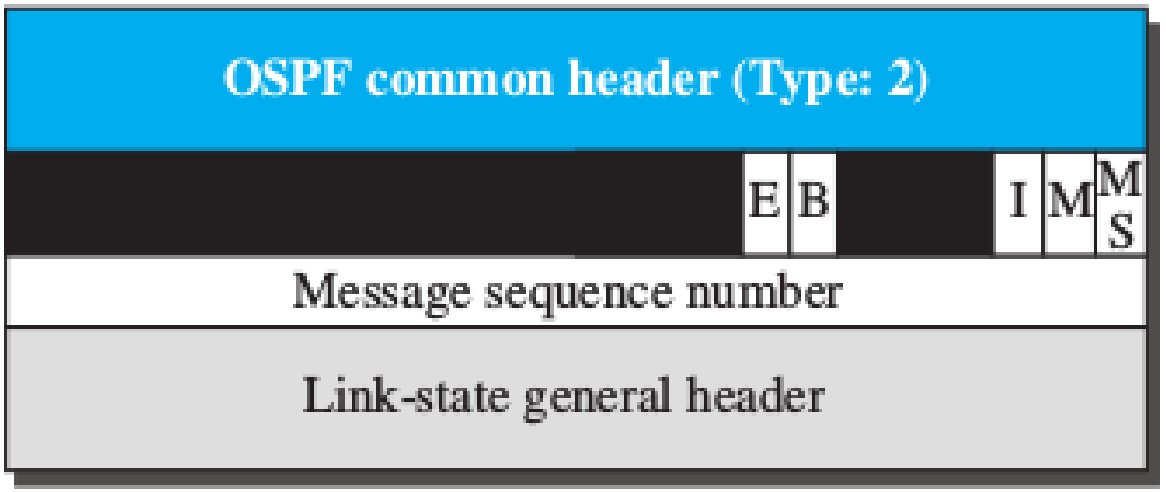
It is used by a router to introduce itself to the neighbors and announce all neighbors that it already knows.



Hello message

Database description message (type 2)

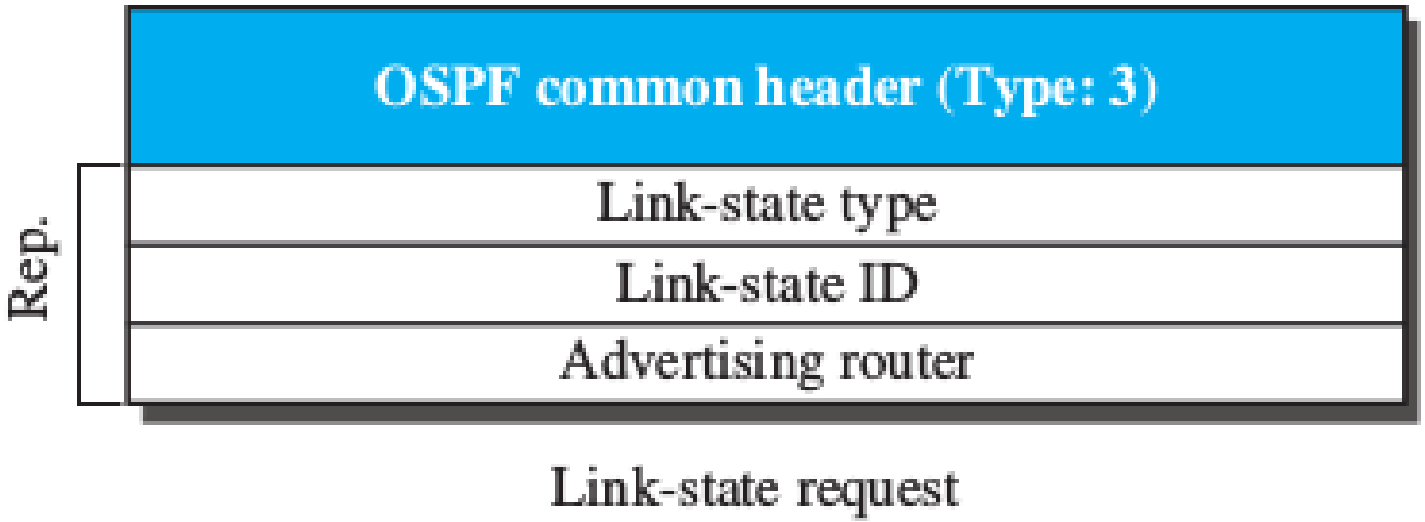
The database description message (type 2) is normally sent in response to the hello message to allow a newly joined router to acquire the full LSDB.



Database description

Link-state request message (type 3)

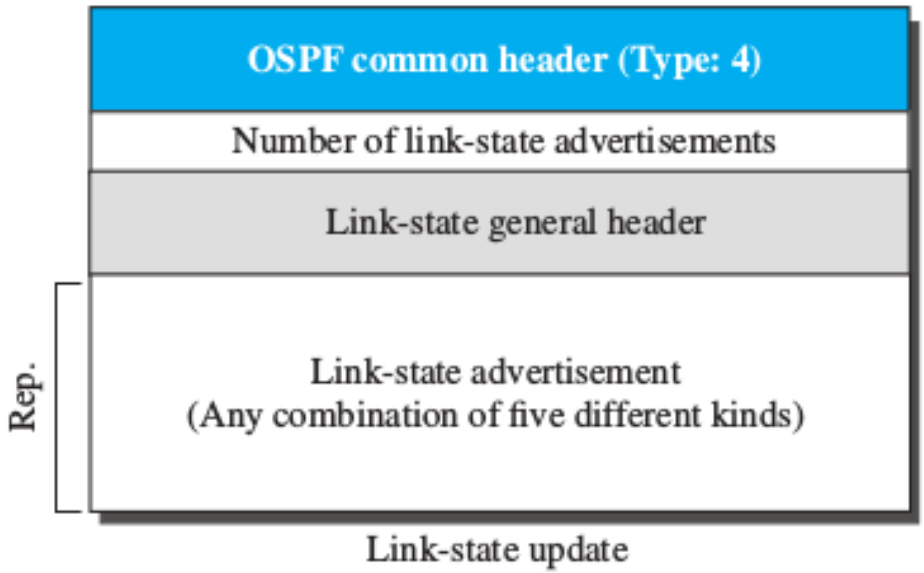
The link-state request message (type 3) is sent by a router that needs information about a specific LS.



Link-state update message (type 4)

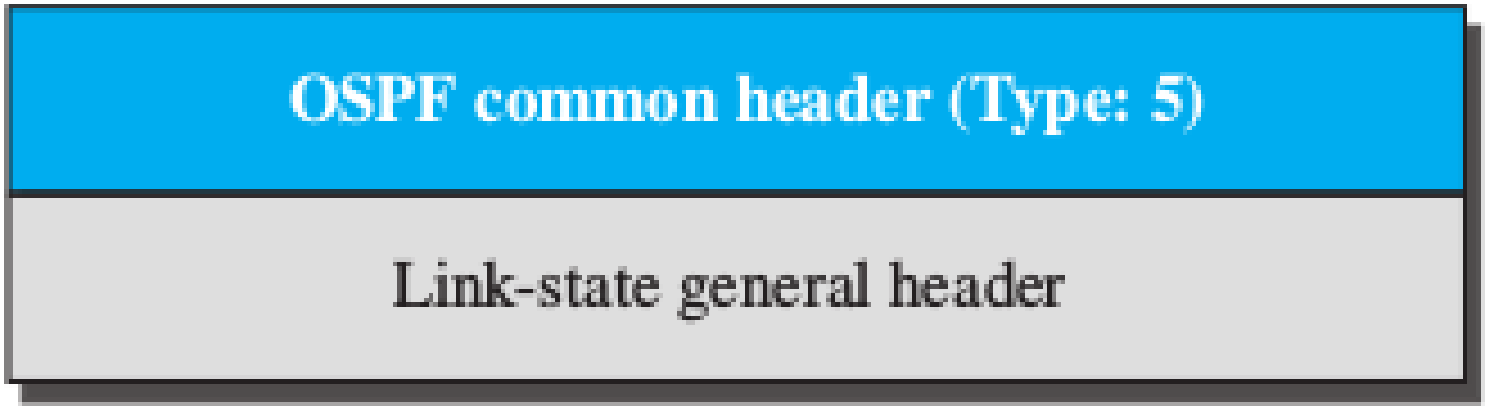
The link-state update message (type 4) is the main OSPF message used for building the LSDB.

This message, in fact, has five different versions (router link, network link, summary link to network, summary link to AS border router, and external link),



Link-state acknowledgment message (type 5)

The link-state acknowledgment message (type 5) is used to create reliability in OSPF; each router that receives a link-state update message needs to acknowledge it.



Authentication

Provision for authentication

prevents a malicious entity from sending OSPF messages to a router and causing the router to become part of the routing system to which it actually does not belong.

OSPF Algorithm

OSPF implements the link-state routing algorithm with following modifications:

- After each router has created the shortest-path tree, the algorithm needs to use it to create the corresponding routing algorithm.
- The algorithm needs to be augmented to handle sending and receiving all five types of messages.

Performance

- **Update Messages.**
  - The link-state messages in OSPF have a somewhat complex format. They also are flooded to the whole area. If the area is large, these messages may create heavy traffic and use a lot of bandwidth.
- **Convergence of Forwarding Tables.**
  - When the flooding of LSPs is completed, each router can create its own shortest-path tree and forwarding table; convergence is fairly quick. However, each router needs to run Dijkstra's algorithm, which may take some time.

Performance

- **Robustness.**

  The OSPF protocol is more robust than RIP because, after receiving the completed LSDB, each router is independent and does not depend on other routers in the area. Corruption or failure in one router does not affect other routers as seriously as in RIP.

Border Gateway Protocol Version 4 (BGP4)

The Border Gateway Protocol version 4 (BGP4) is the only interdomain routing protocol used in the Internet today.

BGP4 is based on the path-vector algorithm

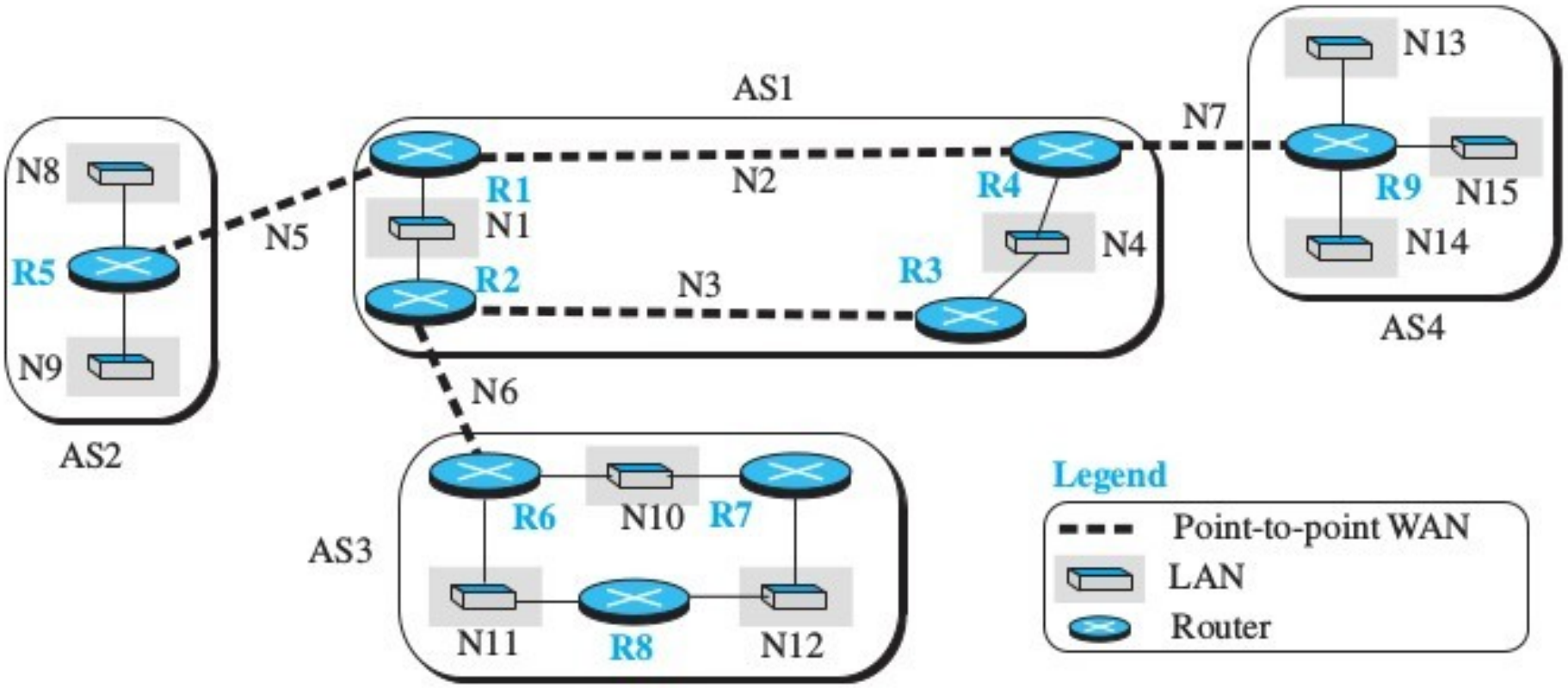
we introduce the basics of BGP and its relationship with intradomain routing protocols (RIP or OSPF).

Consider an example of an internet with four autonomous systems.

AS2, AS3, and AS4 are stub autonomous systems; AS1 is a transient one.

In our example, data exchange between AS2, AS3, and AS4 should pass through AS1.

A sample internet with four AS



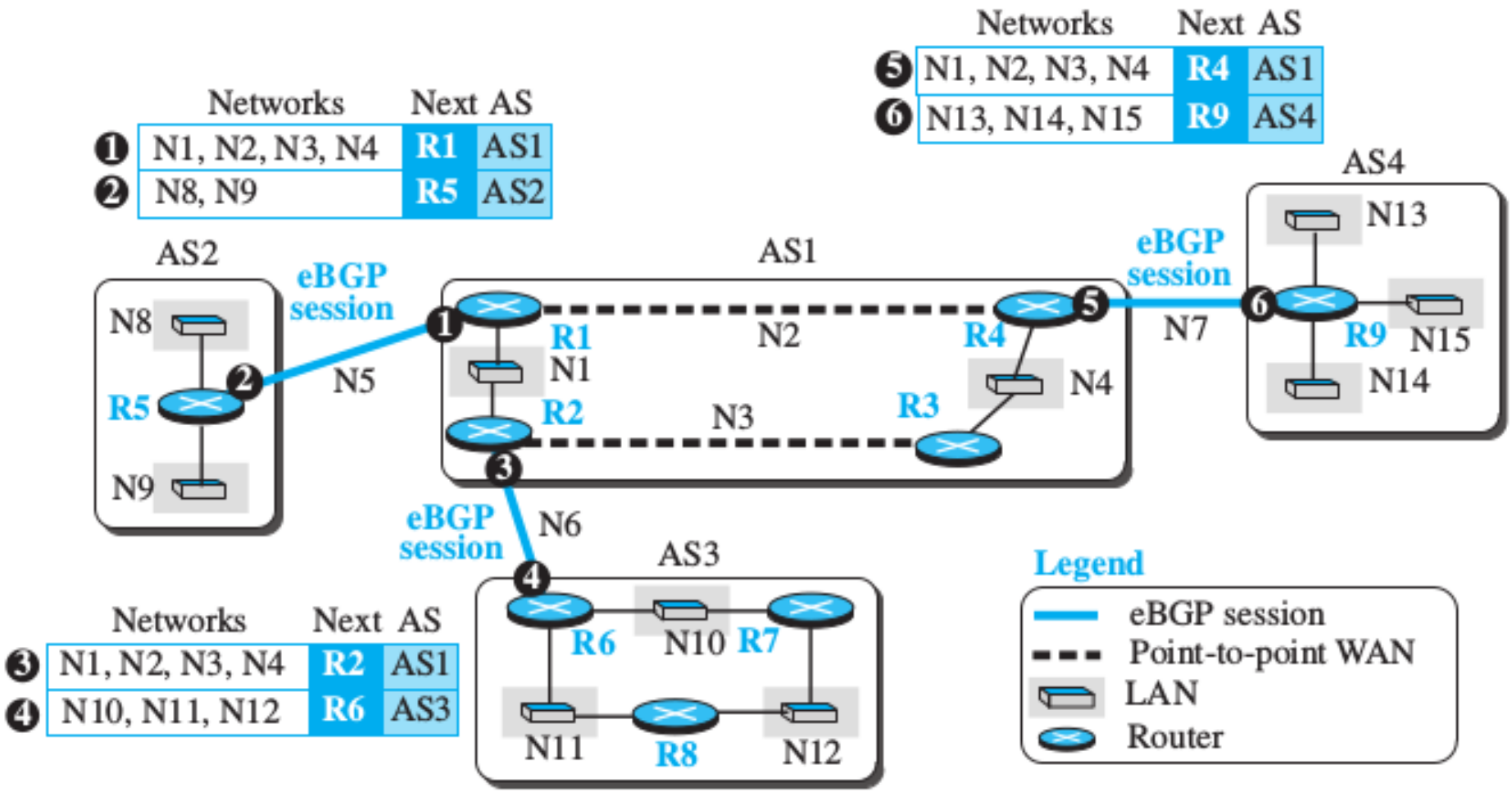
BGP

- Each autonomous system uses one of the two common intradomain protocols, RIP or OSPF.
- Each router in each AS knows how to reach a network that is in its own AS, but it does not know how to reach a network in another AS.
- To enable this we first install a variation of BGP4, called external BGP (eBGP), on each border router (the one at the edge of each AS which is connected to a router at another AS).
- We then install the second variation of BGP, called internal BGP (iBGP), on all routers.
- This means that the border routers will be running three routing protocols (intradomain, eBGP, and iBGP), but other routers are running two protocols (intradomain and iBGP).

Operation of External BGP (eBGP)

- We can say that BGP is a kind of point-to-point protocol. When the software is installed on two routers, they try to create a TCP connection using the well-known port 179.
- In other words, a pair of client and server processes continuously communicate with each other to exchange messages.
- The eBGP variation of BGP allows two physically connected border routers in two different ASs to form pairs of eBGP speakers and exchange messages.
- The three pairs are :
  - R1-R5,
  - R2-R6,
  - R4-R9.

Figure 20.25 *eBGP operation*



eBGP

- The connection between these pairs is established over three physical WANs (N5,N6, and N7).
- There is a need for a logical TCP connection to be created over the physical connection to make the exchange of information possible.
- Each logical connection in BGP parlance is referred to as a **session.** This means that we need 3 sessions in our example.
- The figure also shows the simplified update messages sent by routers involved in the eBGP sessions.
- Message number 1 is sent by router R1 and tells router R5 that N1, N2, N3 and N4 can be reached through router R1.
- Router R5 can now add these pieces of information at the end of its forwarding table. When R5 receives any packet destined for these four networks, it can use its forwarding table and find that the next router is R1.

eBGP

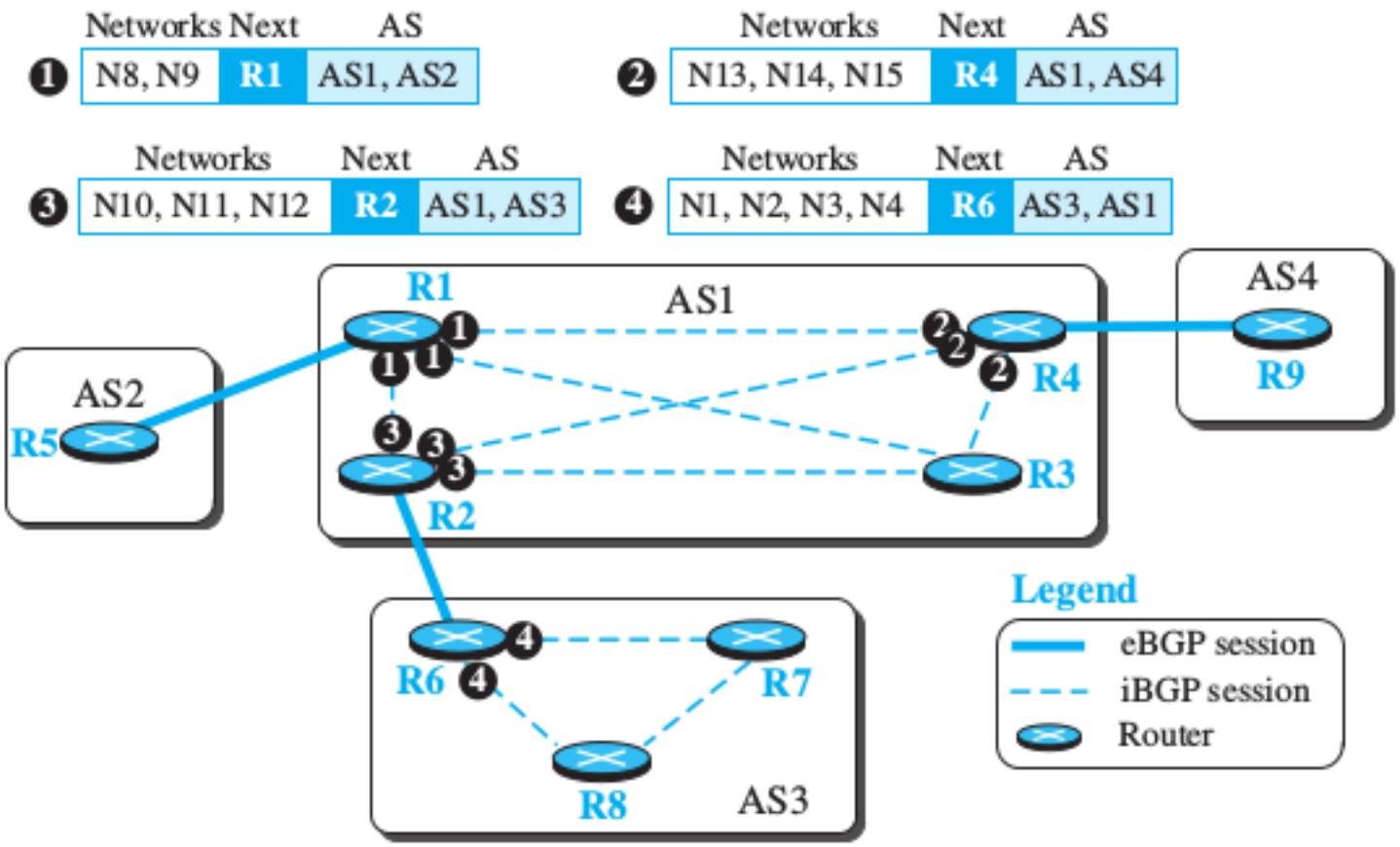
- After the three BGP sessions, the reachability information is not complete.
  - Some border routers do not know how to route a packet destined for non-neighbor As. For eg: R5 does not know how to route packets destined for networks in AS3 and AS4
  - None of the nonborder routers know how to route a packet destined for any networks in other ASs.
- To address the above two problems, we need to allow all pairs of routers (border or nonborder) to run the second variation of the BGP protocol, iBGP.

Operation of Internal BGP (iBGP)

- The iBGP protocol is similar to the eBGP protocol in that it uses the service of TCP on the well-known port 179, but it creates a session between any possible pair of routers inside an autonomous system.
- However if an AS has only one router, there cannot be an iBGP session.
- if there are n routers in an autonomous system, there should be ${}^{n}C_{2}$ iBGP sessions in that autonomous system (a fully connected mesh) to prevent loops in the system.

Operation of Internal BGP (iBGP)

Figure 20.26 *Combination of eBGP and iBGP sessions in our internet*



Operation of Internal BGP (iBGP)

- The first message (numbered 1) is sent by R1 announcing that networks N8 and N9 are reachable through the path AS1-AS2, but the next router is R1.
- This message is sent, through separate sessions, to R2, R3, and R4. Routers R2, R4, and R6 do the same thing but send different messages to different destinations.
- The interesting point is that, at this stage, R3, R7, and R8 create sessions with their peers, but they actually have no message to send.
- The updating process does not stop here. For example, after R1 receives the update message from R2, it combines the reachability information about AS3 with the reachability information it already knows about AS1 and sends a new update message to R5.
- Now R5 knows how to reach networks in AS1 and AS3.

Operation of Internal BGP (iBGP)

- The process continues with routers exchanging messages For eg : when R1 receives the update message from R4 and so on.
- Finally we obtain the BGP path tables as shown

Networks	Next	Path
N8, N9	R5	AS1, AS2
N10, N11, N12	R2	AS1, AS3
N13, N14, N15	R4	AS1, AS4

Path table for R1

Networks	Next	Path
N8, N9	R1	AS1, AS2
N10, N11, N12	R6	AS1, AS3
N13, N14, N15	R1	AS1, AS4

Path table for R2

Networks	Next	Path
N8, N9	R2	AS1, AS2
N10, N11, N12	R2	AS1, AS3
N13, N14, N15	R4	AS1, AS4

Path table for R3

Networks	Next	Path
N8, N9	R1	AS1, AS2
N10, N11, N12	R1	AS1, AS3
N13, N14, N15	R9	AS1, AS4

Path table for R4

Networks	Next	Path
N1, N2, N3, N4	R1	AS2, AS1
N10, N11, N12	R1	AS2, AS1, AS3
N13, N14, N15	R1	AS2, AS1, AS4

Path table for R5

Networks	Next	Path
N1, N2, N3, N4	R2	AS3, AS1
N8, N9	R2	AS3, AS1, AS2
N13, N14, N15	R2	AS3, AS1, AS4

Path table for R6

Networks	Next	Path
N1, N2, N3, N4	R6	AS3, AS1
N8, N9	R6	AS3, AS1, AS2
N13, N14, N15	R6	AS3, AS1, AS4

Path table for R7

Networks	Next	Path
N1, N2, N3, N4	R6	AS3, AS1
N8, N9	R6	AS3, AS1, AS2
N13, N14, N15	R6	AS3, AS1, AS4

Path table for R8

Networks	Next	Path
N1, N2, N3, N4	R4	AS4, AS1
N8, N9	R4	AS4, AS1, AS2
N10, N11, N12	R4	AS4, AS1, AS3

Path table for R9

Injection of Information into Intradomain Routing

- The role of an interdomain routing protocol such as BGP is to help the routers inside the AS to augment their routing information.
- In other words, the path tables collected and organized by BGP are not used, per se, for routing packets; they are injected into intradomain forwarding tables (RIP or OSPF) for routing packets.
- This can be done in several ways depending on the type of AS.
- In the case of a stub AS, the only area border router adds a default entry at the end of its forwarding table and defines the next router to be the speaker router at the end of the eBGP connection.

Injection of Information into Intradomain Routing

- R5 in AS2 defines R1 as the default router for all networks other than N8 and N9. (default entry)
- R9 in AS4 with the default router to be R4.
- In AS3, R6 set its default router to be R2, but R7 and R8 set their default router to be R6
- In the case of a transient AS, The router in an AS should inject the whole contents of the path table (eg: R1 in AS1 which is shown next)
- Address Aggregation can also be performed to reduce size of forwarding tables
- For example, prefixes 14.18.20.0/26, 14.18.20.64/26, 14.18.20.128/26, and 14.18.20.192/26, can be combined into 14.18.20.0/24

Forwarding tables afer injection from BGP

Des.	Next	Cost
N1	—	1
N4	R4	2
N8	R5	1
N9	R5	1
N10	R2	2
N11	R2	2
N12	R2	2
N13	R4	2
N14	R4	2
N15	R4	2

Table for R1

Des.	Next	Cost
N1	—	1
N4	R3	2
N8	R1	2
N9	R1	2
N10	R6	1
N11	R6	1
N12	R6	1
N13	R3	3
N14	R3	3
N15	R3	3

Table for R2

Des.	Next	Cost
N1	R2	2
N4	—	1
N8	R2	3
N9	R2	3
N10	R2	2
N11	R2	2
N12	R2	2
N13	R4	2
N14	R4	2
N15	R4	2

Table for R3

Des.	Next	Cost
N1	R1	2
N4	—	1
N8	R1	2
N9	R1	2
N10	R3	3
N11	R3	3
N12	R3	3
N13	R9	1
N14	R9	1
N15	R9	1

Table for R4

Des.	Next	Cost
N8	—	1
N9	—	1
0	R1	1

Table for R5

Des.	Next	Cost
N10	—	1
N11	—	1
N12	R7	2
0	R2	1

Table for R6

Des.	Next	Cost
N10	—	1
N11	R6	2
N12	—	1
0	R6	2

Table for R7

Des.	Next	Cost
N10	R6	2
N11	—	1
N12	—	1
0	R6	2

Table for R8

Des.	Next	Cost
N13	—	1
N14	—	1
N15	—	1
0	R4	1

Table for R9

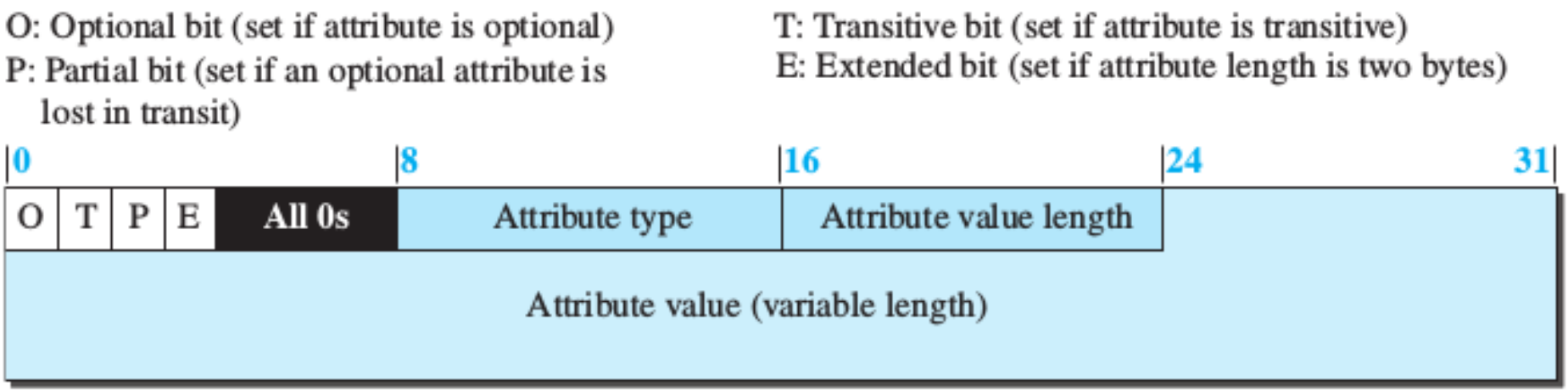
Path Attributes

- In both intradomain routing protocols (RIP or OSPF), a destination is normally associated with two pieces of information: next hop and cost.
- The first one shows the address of the next router to deliver the packet; the second defines the cost to the final destination.
- Inter-domain routing is more involved and naturally needs more information about how to reach the final destination.
- In BGP these pieces are called path attributes.
- BGP allows a destination to be associated with up to seven path attributes.

Path attributes

- Path attributes are divided into two broad categories:
  - well-known
  - optional
- A well-known attribute must be recognized by all routers; an optional attribute need not be.
- A well-known attribute can be mandatory, which means that it must be present in any BGP update message, or discretionary, which means it does not have to be.
- An optional attribute can be either transitive, which means it can pass to the next AS, or intransitive, which means it cannot.

Format of path attribute



O: Optional bit (set if attribute is optional)

P: Partial bit (set if an optional attribute is lost in transit)

T: Transitive bit (set if attribute is transitive)

E: Extended bit (set if attribute length is two bytes)

The attribute value length defines the length of the attribute value field

Attribute types

ORIGIN (type 1).

This is a well-known mandatory attribute, which defines the source of the routing information.

This attribute can be defined by one of the three values: 1, 2, 3.

Value 1 means that the information about the path has been taken from an intradomain protocol (RIP or OSPF).

Value 2 means that the information comes from BGP.

Value 3 means that it comes from an unknown source.

Attribute types

AS-PATH (type 2).

This is a well-known mandatory attribute, which defines the list of autonomous systems through which the destination can be reached.

We have used this attribute in our examples

NEXT-HOP (type 3).

This is a well-known mandatory attribute, which defines the next router to which the data packet should be forwarded.

This attribute helps to inject path information collected through the operations of eBGP and iBGP into the intradomain routing protocols such as RIP or OSPF.

Attribute types

MULT-EXIT-DISC (type 4).

The multiple-exit discriminator is an optional intransitive attribute, which discriminates among multiple exit paths to a destination.

The value of this attribute is normally defined by the metric in the corresponding intradomain protocol

LOCAL-PREF (type 5).

The local preference attribute is a well-known discretionary attribute.

It is normally set by the administrator, based on the organization policy. The routes the administrator prefers are given a higher local preference value

Attribute types

ATOMIC-AGGREGATE (type 6).

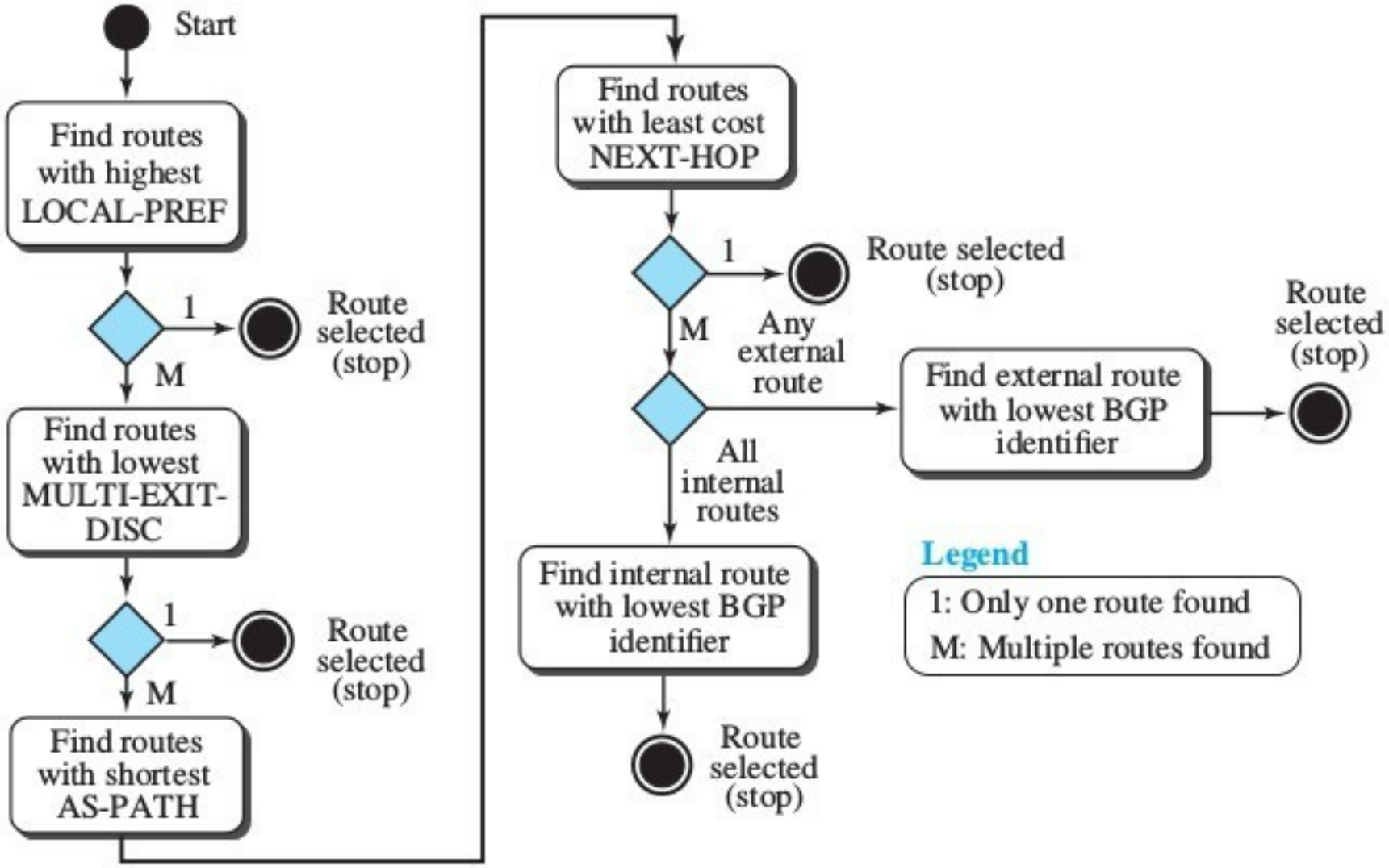
This is a well-known discretionary attribute, which defines the destination prefix as not aggregate; it only defines a single destination network.

This attribute has no value field, which means the value of the length field is zero.

AGGREGATOR (type 7).

This is an optional transitive attribute, which emphasizes that the destination prefix is an aggregate.

Flow diagram for route selection



BGP Messages

Open Message.

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

Update Message.

The update message is the heart of the BGP protocol. It is used by a router to withdraw destinations that have been advertised previously, to announce a route to a new destination, or both.

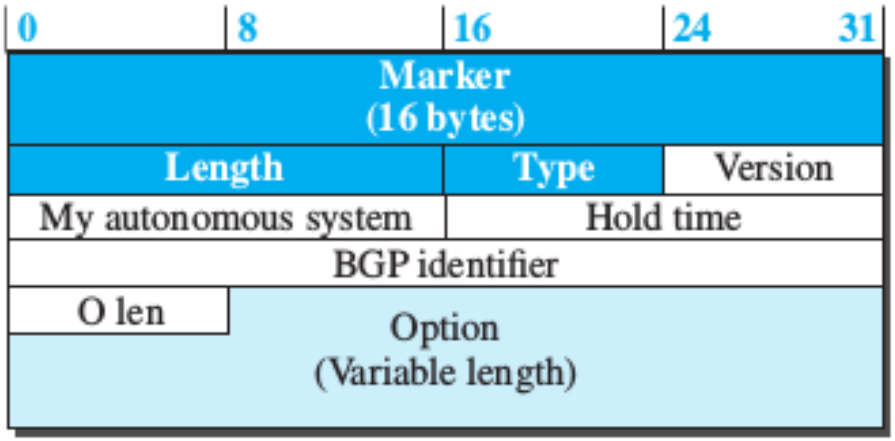
Keepalive Message.

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

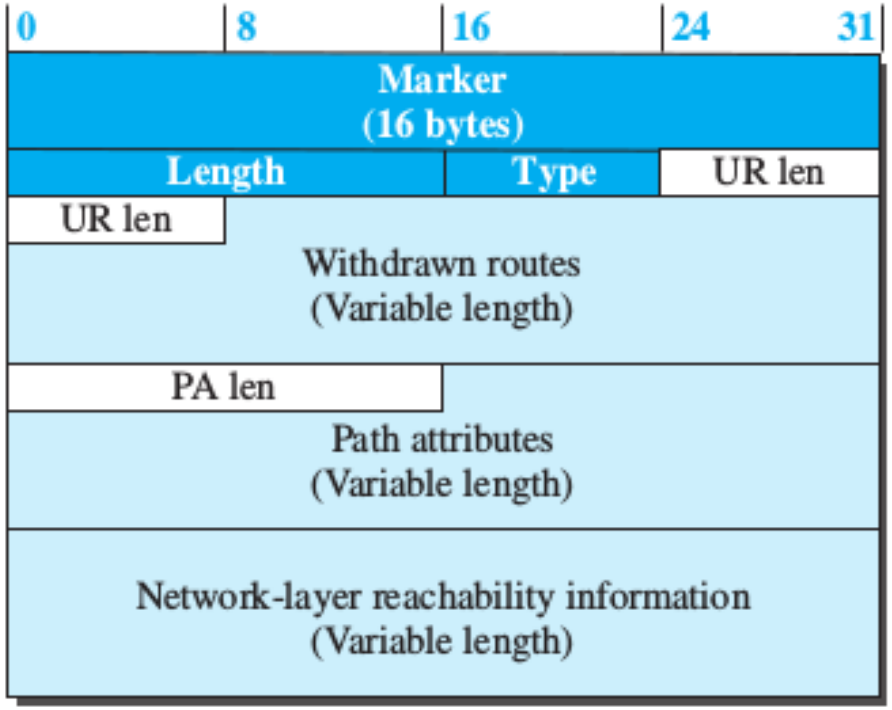
Notification.

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

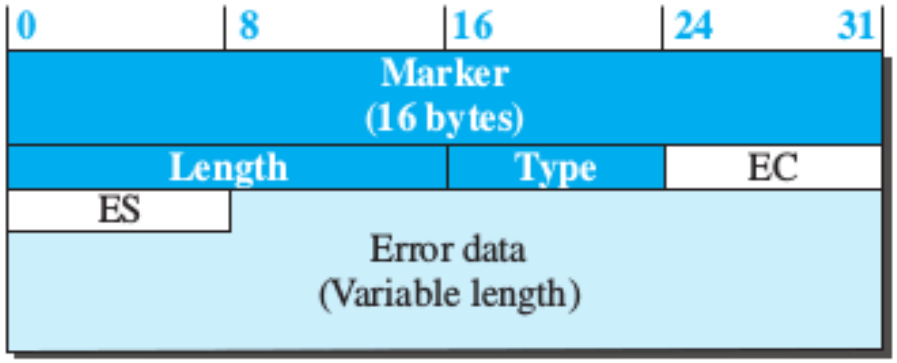
BGP Messages



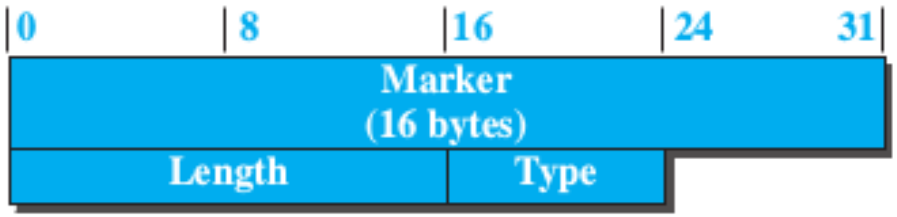
Open message (type 1)



Update message (type 2)



Notification message (type 3)



Keepalive message (type 4)

Fields in common header

Marker: Reserved for authentication
Length: Length of total message in bytes
Type: Type of message (1 to 4)

Abbreviations

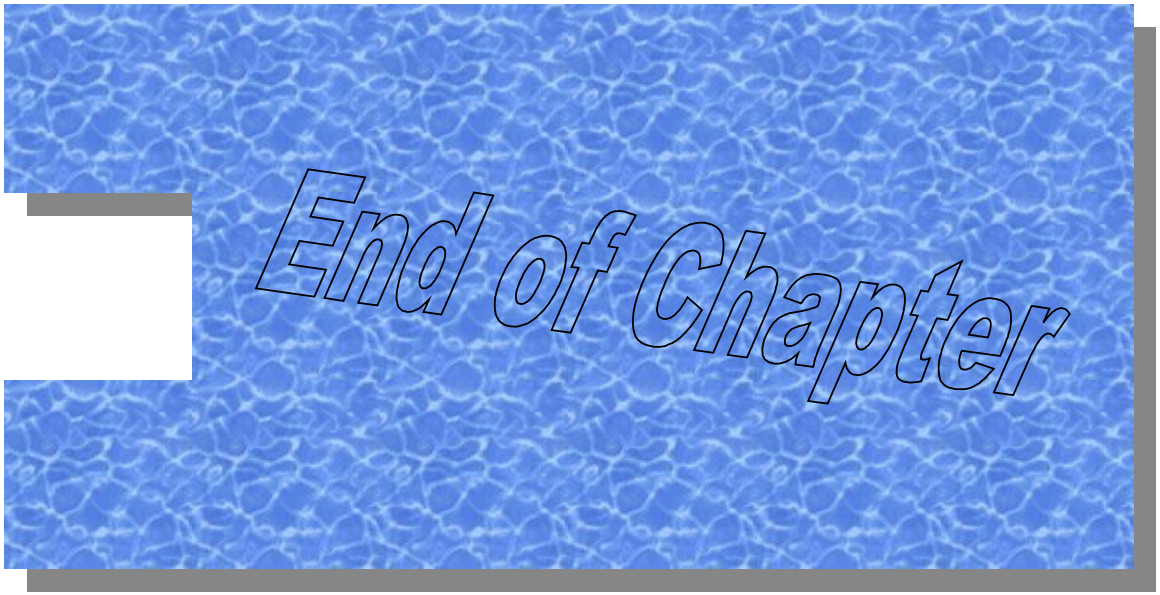
O len: Option length
EC: Error code
ES: Error subcode
UR len: Unfeasible route length
PA len: Path attribute length

BGP Performance

BGP performance can be compared with RIP.

BGP speakers exchange a lot of messages to create forwarding tables, but BGP is free from loops and count-to-infinity.

The same weakness we mention for RIP about propagation of failure and corruption also exists in BGP.


End of Chapter