Chapter 13

Routing Protocols (RIP, OSPF, BGP)

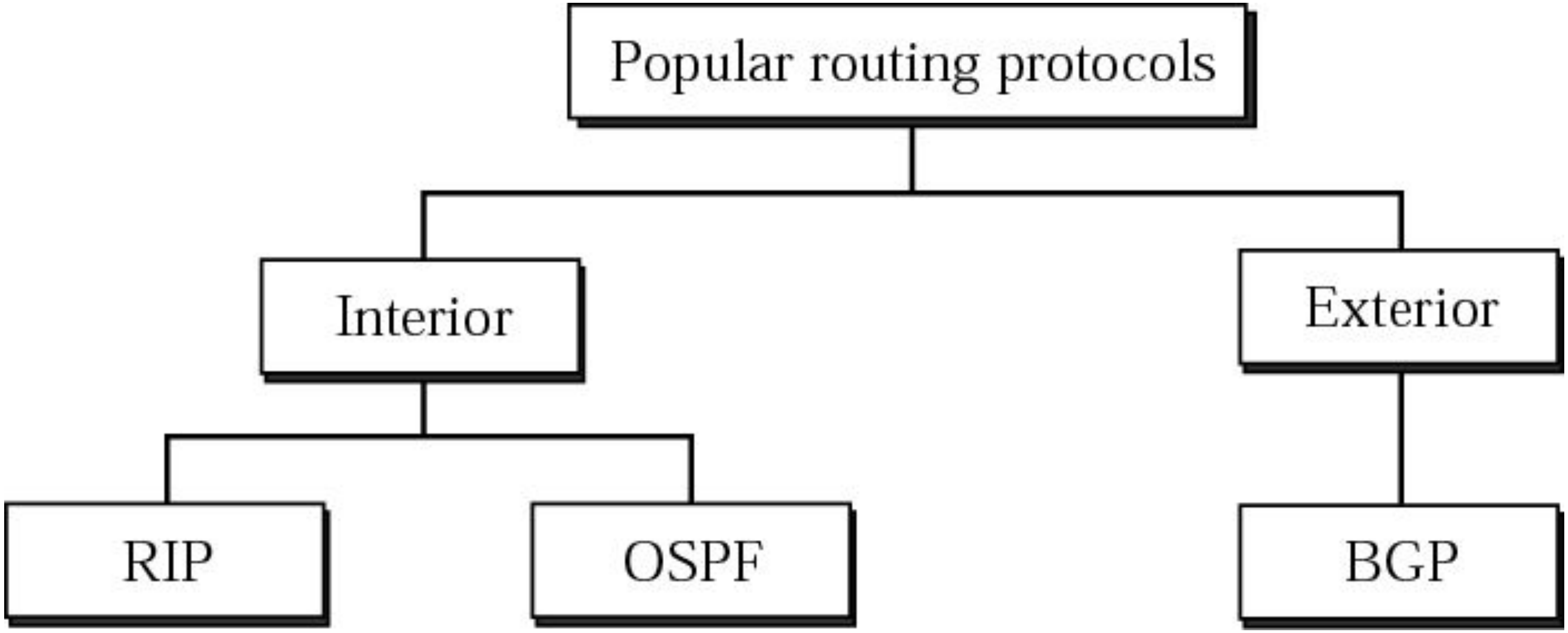
CONTENTS

- **INTERIOR AND EXTERIOR ROUTING**
- **RIP**
- **OSPF**
- **BGP**

- Today, an internet can be so large that one routing protocol cannot handle the task of updating the routing tables of all routers. For this reason, an internet is divided into autonomous systems.
- An **autonomous system (AS) is a group of networks and routers** under the authority of a single administration.
- **Routing inside an autonomous system is referred to as *intra-domain routing.***
- ***Routing between autonomous systems is* referred to as *inter-domain routing.***
- *Each autonomous system can choose one or more* intradomain routing protocols to handle routing inside the autonomous system.
- However, only one interdomain routing protocol handles routing between autonomous systems.

Figure 13-1

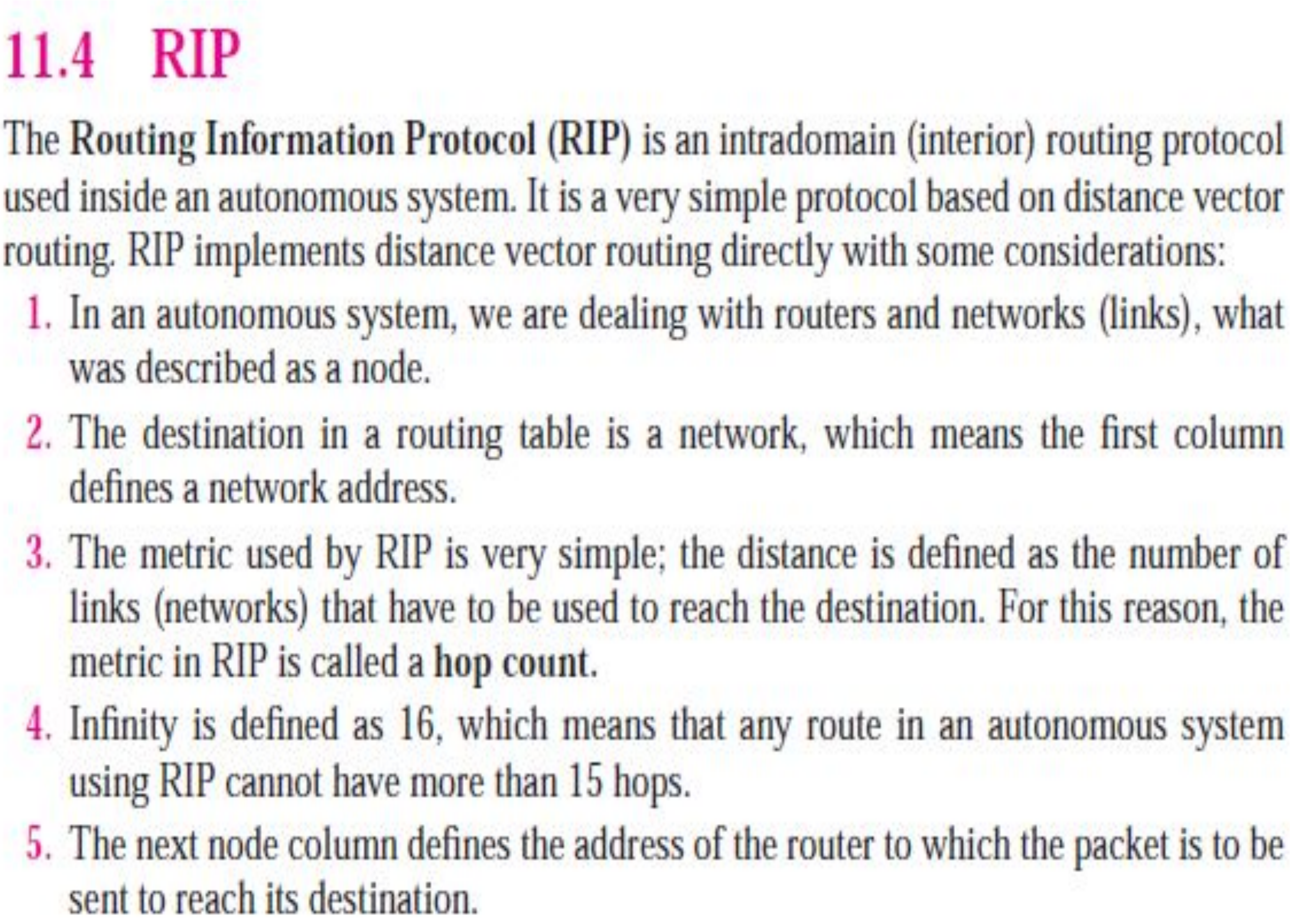
Popular routing protocols



- Routing Information Protocol (RIP) is the implementation of the distance vector protocol.
- Open Shortest Path First (OSPF) is the implementation of the link state protocol.
- Border Gateway Protocol (BGP) is the implementation of the path vector protocol.

13.2

RIP: Routing Information Protocol



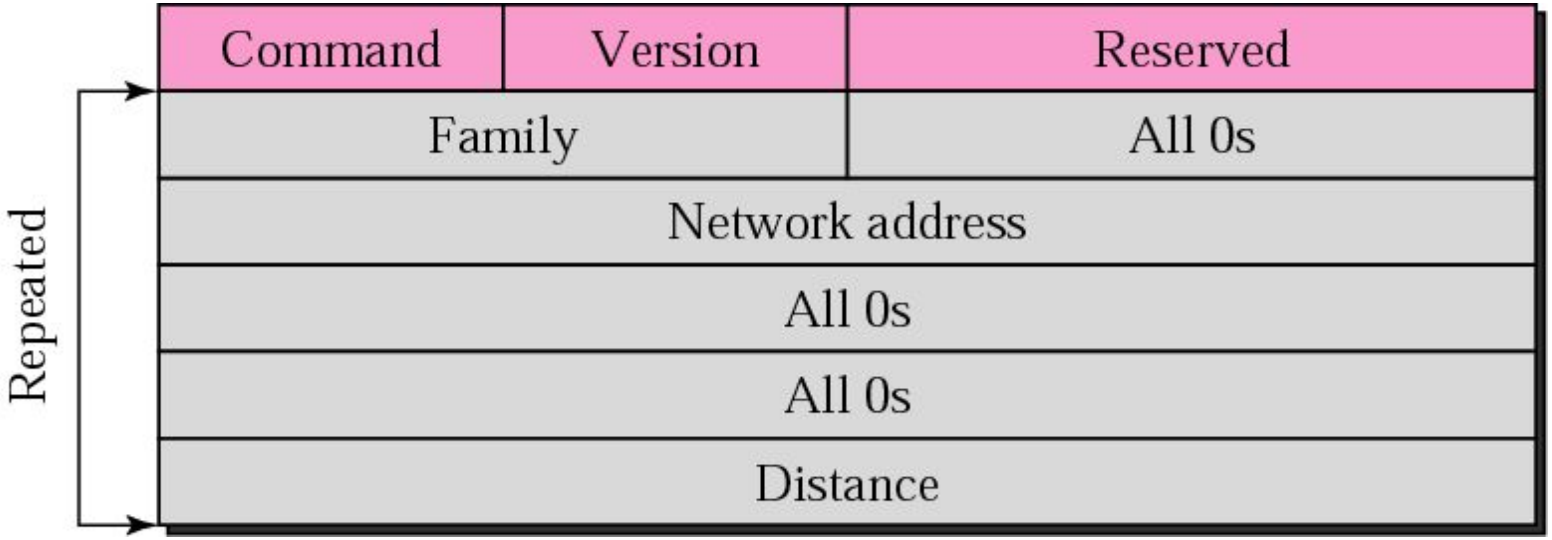
11.4 RIP

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

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

Figure 13-6

RIP message format


Command

Version

Reserved

Family

All 0s

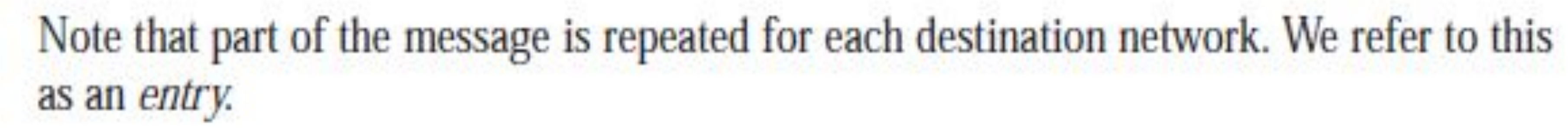
Network address

All 0s

All 0s

Distance

- Command. This 8-bit field specifies the type of message: request (1) or response (2).
- Version. This 8-bit field defines the version.
- Family. This 16-bit field defines the family of the protocol used.
- Network address. **The address field defines the address of the destination network**. RIP has allocated 14 bytes for this field to be applicable to any protocol. However, IP currently uses only 4 bytes. The rest of the address is filled with 0s
- Distance. This 32-bit field defines the hop count (cost) from the advertising router to the destination network.



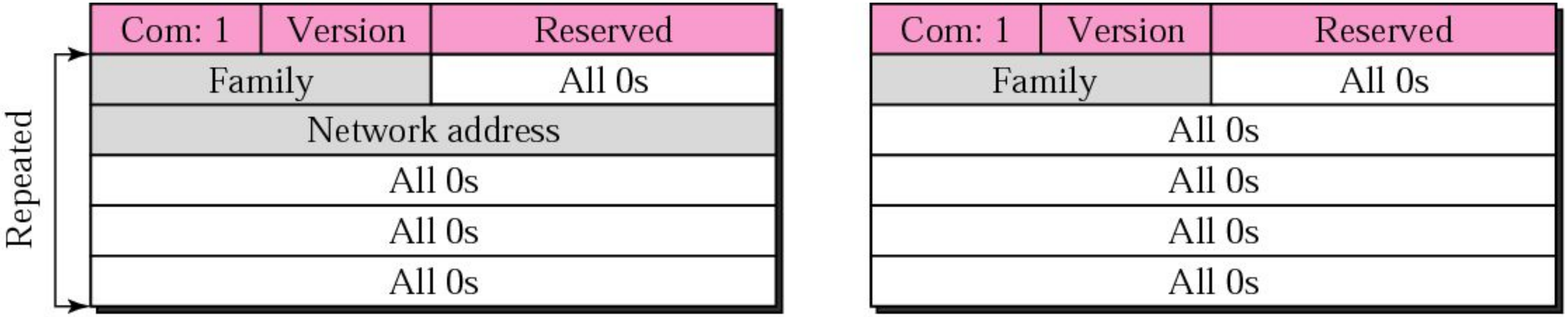
Note that part of the message is repeated for each destination network. We refer to this as an *entry.*

Figure 13-7

Request messages

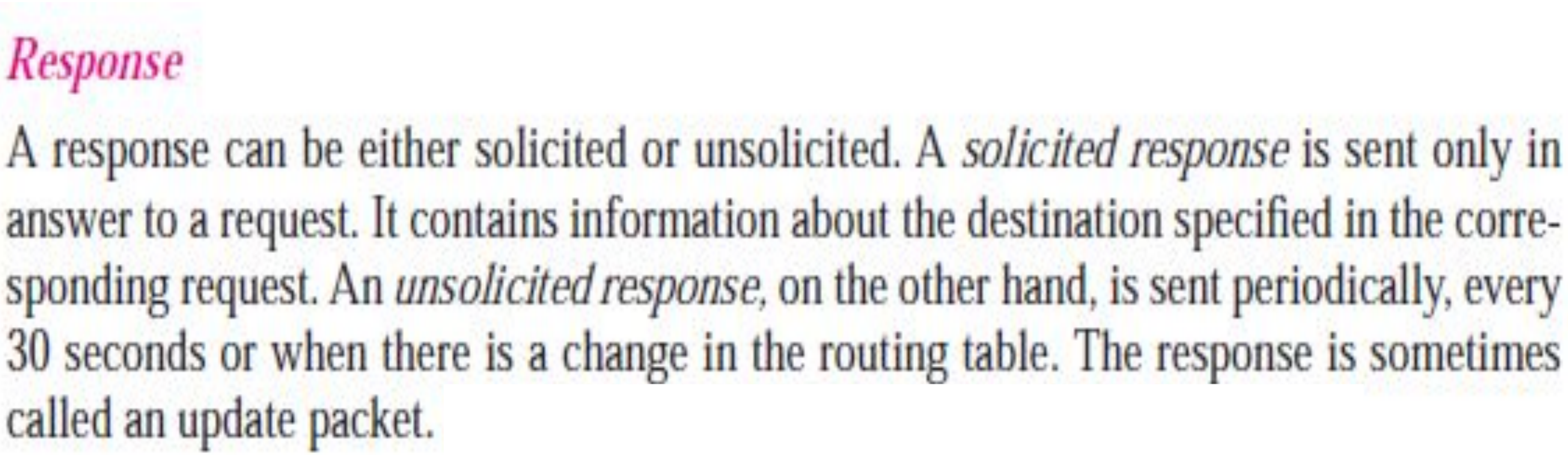
Request

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 (see Figure 11.12).



a. Request for some

b. Request for all

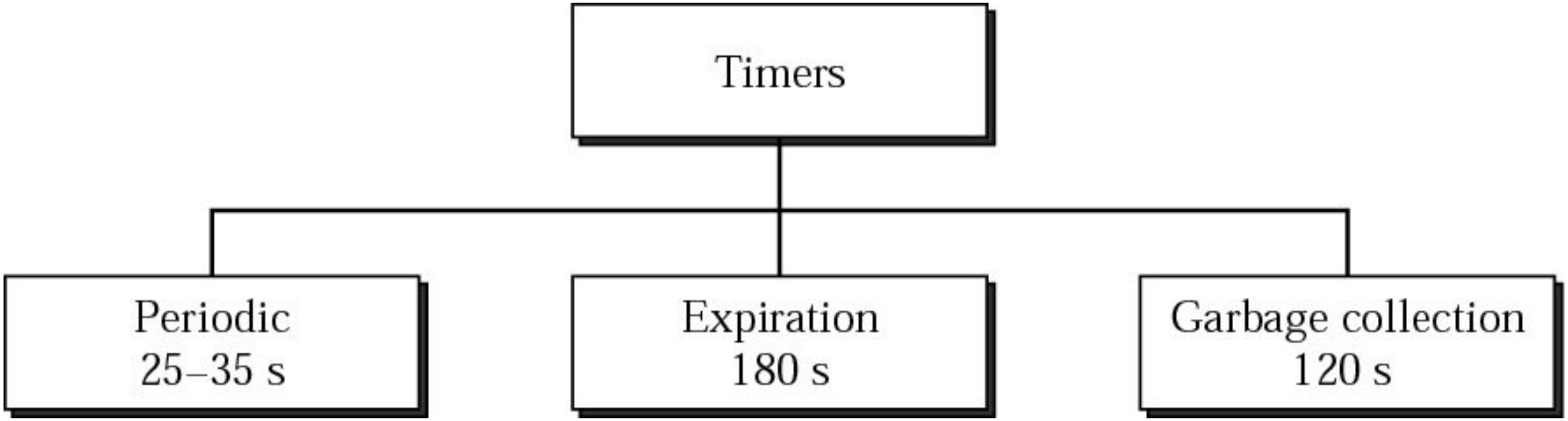


Response

A response can be either solicited or unsolicited. A *solicited response* is sent only in answer to a request. It contains information about the destination specified in the corresponding request. An *unsolicited response,* on the other hand, is sent periodically, every 30 seconds or when there is a change in the routing table. The response is sometimes called an update packet.

Figure 13-10

RIP timers

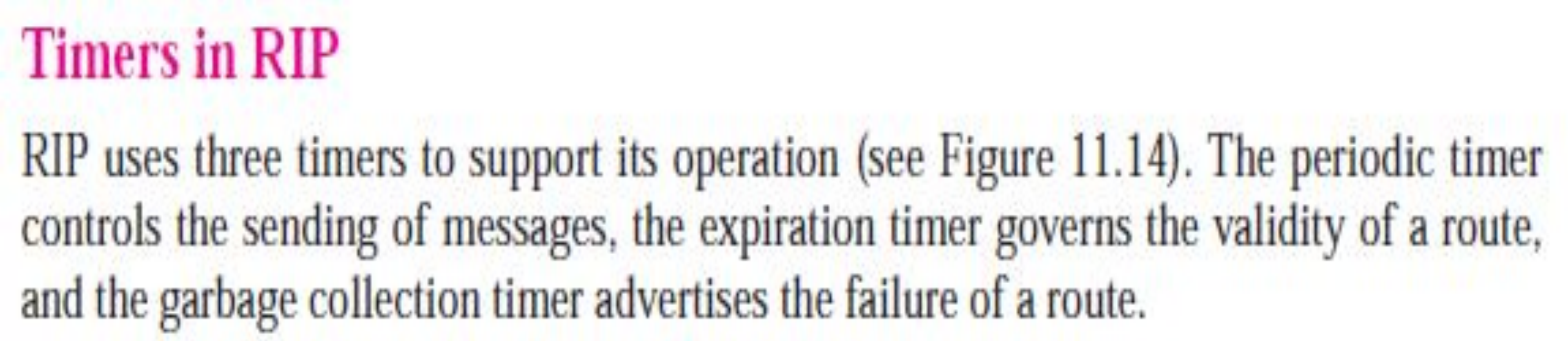


Timers

Periodic
25–35 s

Expiration
180 s

Garbage collection
120 s



Timers in RIP

RIP uses three timers to support its operation (see Figure 11.14). The periodic timer controls the sending of messages, the expiration timer governs the validity of a route, and the garbage collection timer advertises the failure of a route.

- ***Periodic Timer***
- 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.
- It counts down; when zero is reached, the update message is sent, and the timer is randomly set once again.

- *Expiration Timer*
- 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 s for that particular route.
- Every time a new update for the route is received, the timer is reset.
- However, if there is a problem on an internet and no update is received within the allotted 180 s, the route is considered expired and the hop count of the route is set to 16, which means the destination is unreachable.
- Every route has its own expiration timer.

- ***Garbage Collection Timer***
- 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 timer called the **garbage collection timer is set to 120 s** 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.

- **Example 11.5**
- A routing table has 20 entries. It does not receive information about five routes for 200 s. How many timers are running at this time?
- **Solution**
- The 21 timers are listed below:
- Periodic timer: 1
- Expiration timer: 20 − 5 = 15
- Garbage collection timer: 5

RIP Version 2

RIP version 2 was designed to overcome some of the shortcomings of version 1. The

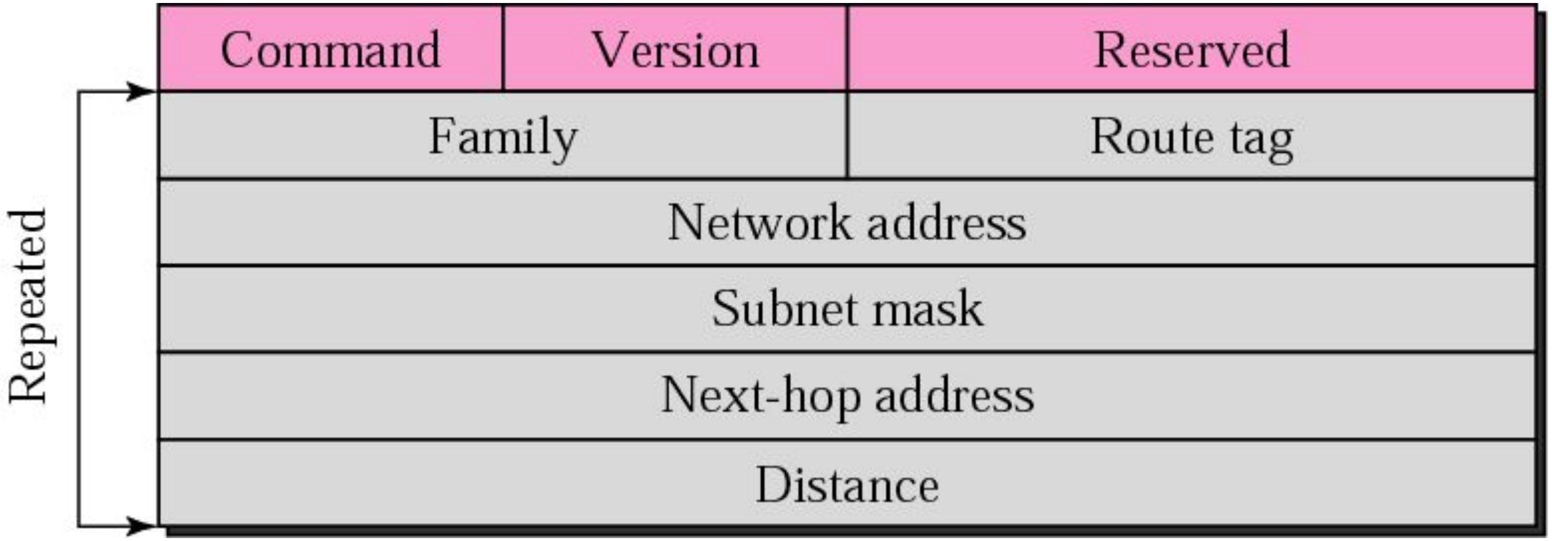
designers of version 2 have not augmented the length of the message for each entry.

They have only replaced those fields in version 1 that were filled with 0s for the TCP/IP

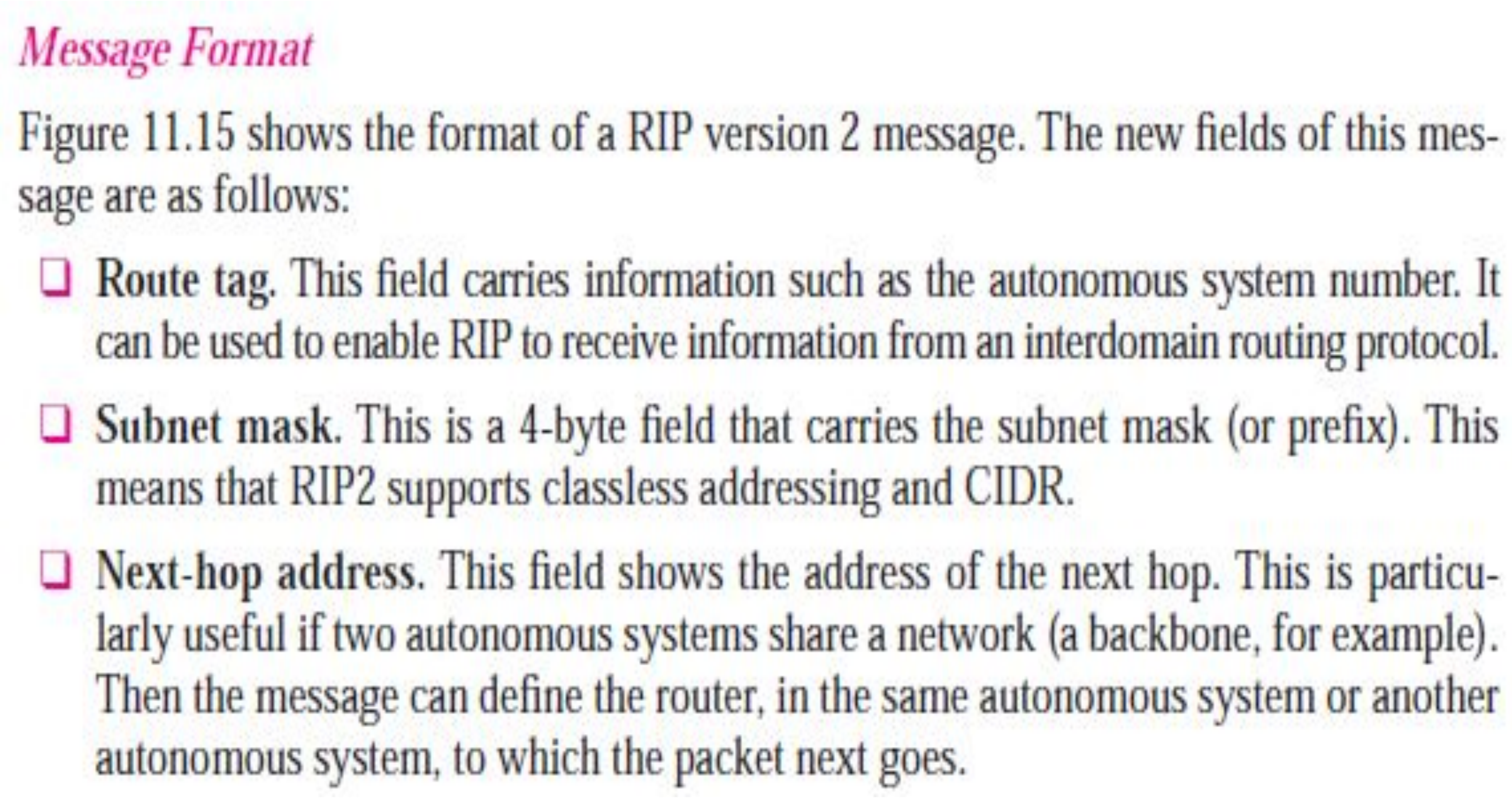
protocol with some new fields.

Figure 13-16

RIP-v2 Format



Command	Version	Reserved
Family		Route tag
Network address		
Subnet mask		
Next-hop address		
Distance		



Message Format

Figure 11.15 shows the format of a RIP version 2 message. The new fields of this message are as follows:

- **Route tag.** This field carries information such as the autonomous system number. It can be used to enable RIP to receive information from an interdomain routing protocol.
- **Subnet mask.** This is a 4-byte field that carries the subnet mask (or prefix). This means that RIP2 supports classless addressing and CIDR.
- **Next-hop address.** This field shows the address of the next hop. This is particularly useful if two autonomous systems share a network (a backbone, for example). Then the message can define the router, in the same autonomous system or another autonomous system, to which the packet next goes.

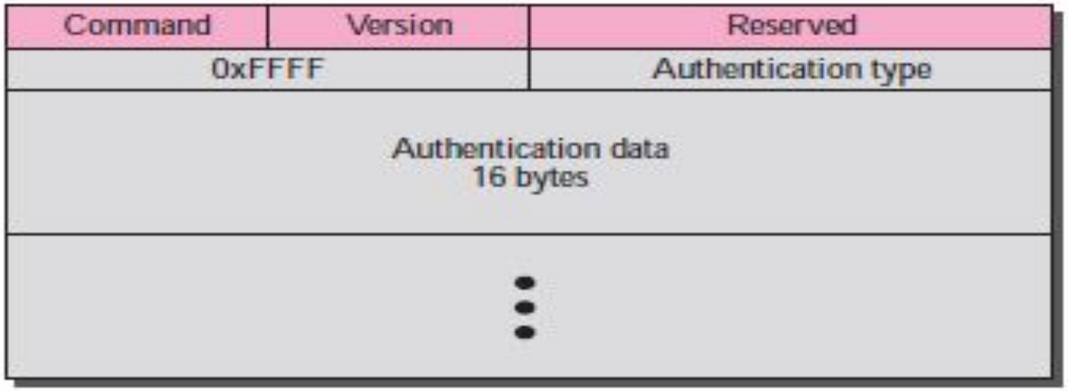
Classless Addressing

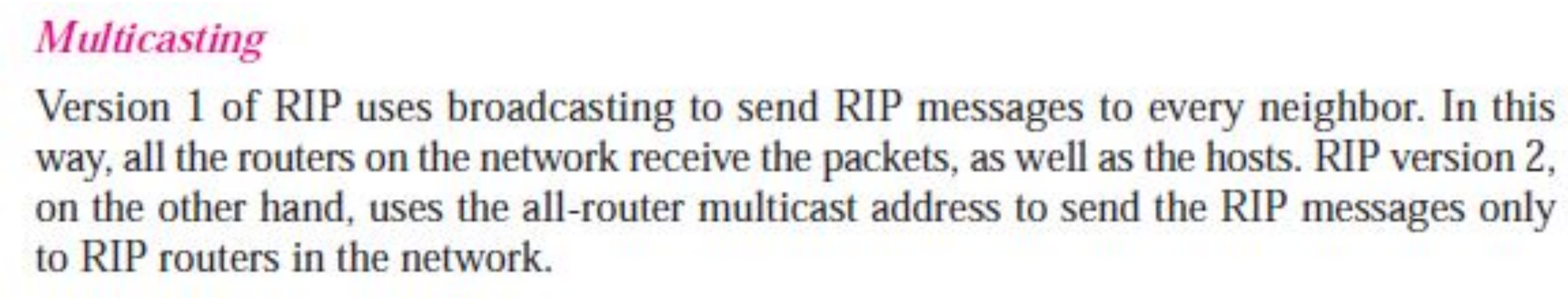
Probably the most important difference between the two versions of RIP is classful versus classless addressing. RIPv1 uses classful addressing. The only entry in the message format is the network address (with a default mask). RIPv2 adds one field for the subnet mask, which can be used to define a network prefix length. This means that in this version, we can use classless addressing. A group of networks can be combined into one prefix and advertised collectively,

Authentication

Authentication is added to protect the message against unauthorized advertisement. No new fields are added to the packet; instead, the first entry of the message is set aside for authentication information. To indicate that the entry is authentication information and not routing information, the value of $FFFF_{16}$ is entered in the family field (see Figure 11.16). The second field, the authentication type, defines the protocol used for authentication, and the third field contains the actual authentication data.

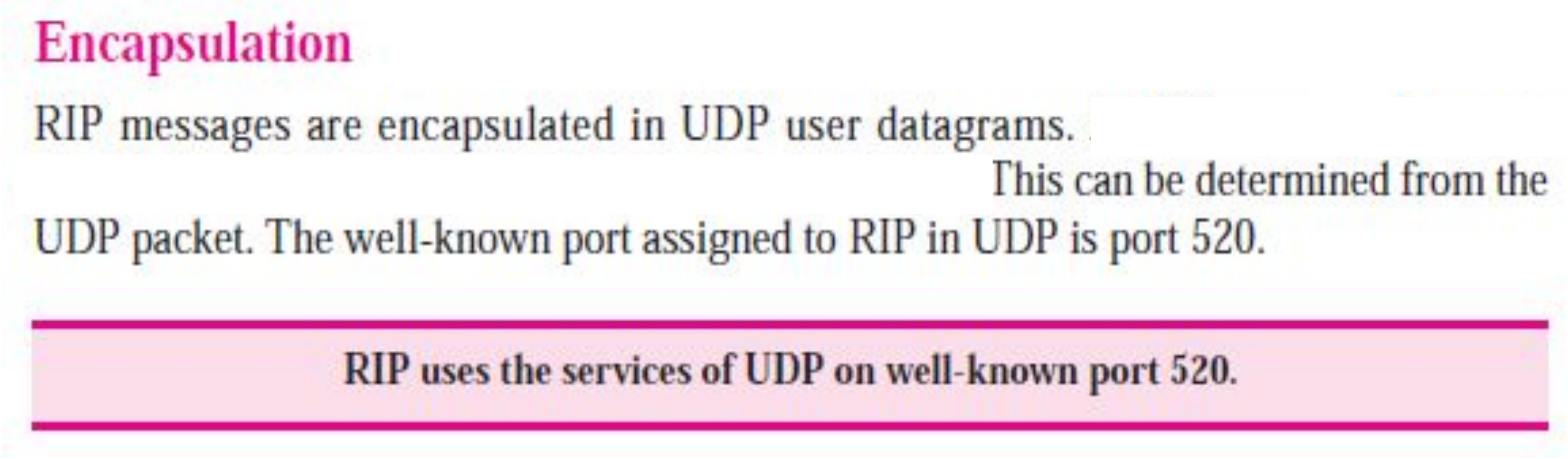
Figure 11.16 *Authentication*





Multicasting

Version 1 of RIP uses broadcasting to send RIP messages to every neighbor. In this way, all the routers on the network receive the packets, as well as the hosts. RIP version 2, on the other hand, uses the all-router multicast address to send the RIP messages only to RIP routers in the network.



Encapsulation

RIP messages are encapsulated in UDP user datagrams.

This can be determined from the UDP packet. The well-known port assigned to RIP in UDP is port 520.

RIP uses the services of UDP on well-known port 520.

Note

RIP version 2 supports CIDR.

Note

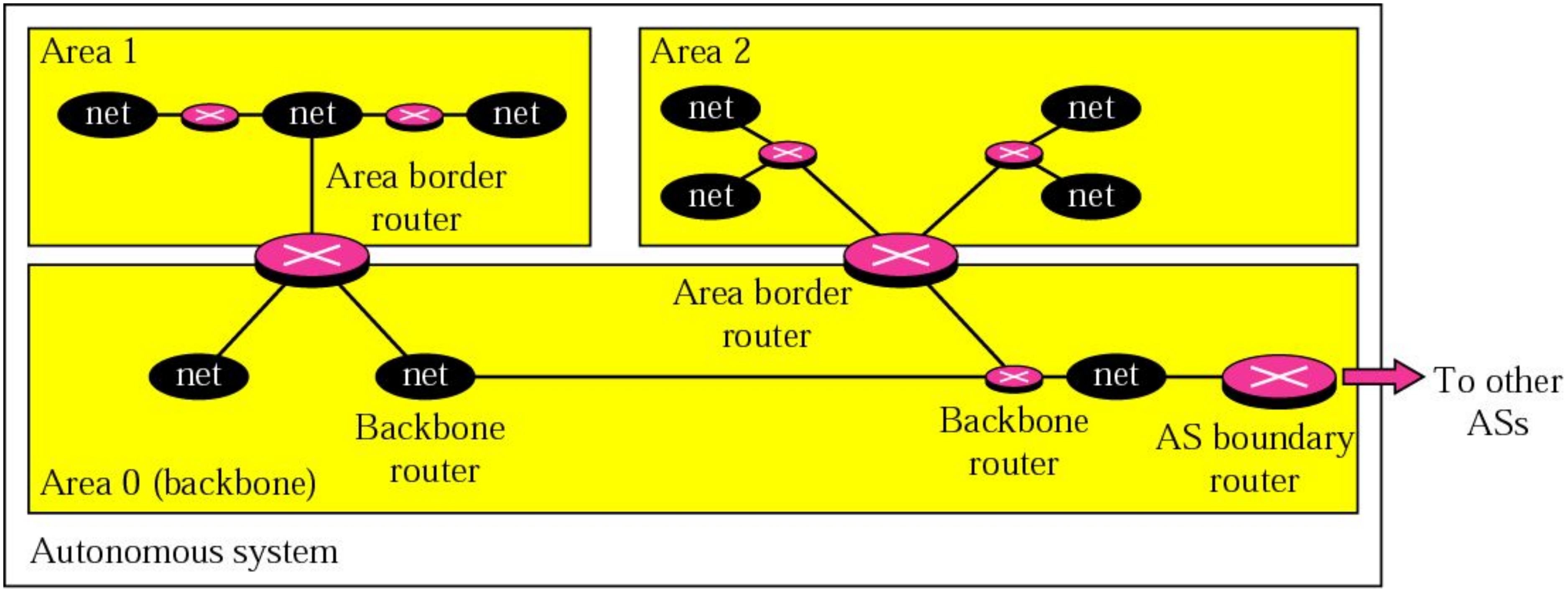
RIP uses the services of UDP on well-known port 520.

13.3

OSPF:
Open Shortest
Path First

Figure 13-18

Areas in an autonomous system



McGraw-Hill

Figure 13-19

Types of links

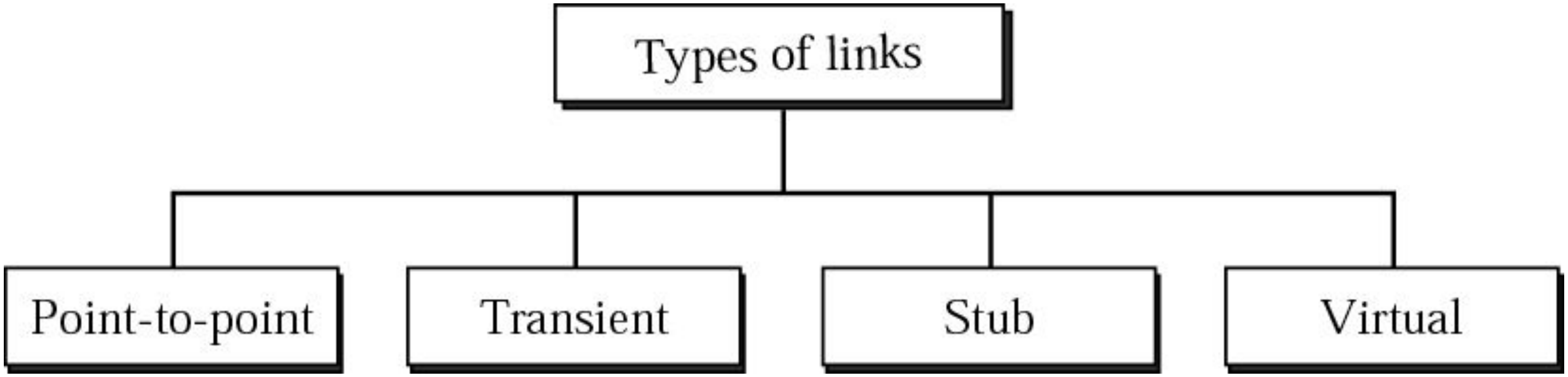


Figure 13-20

Point-to-point link

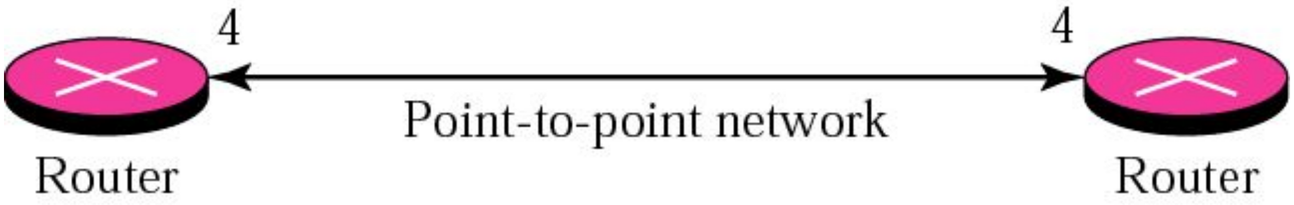
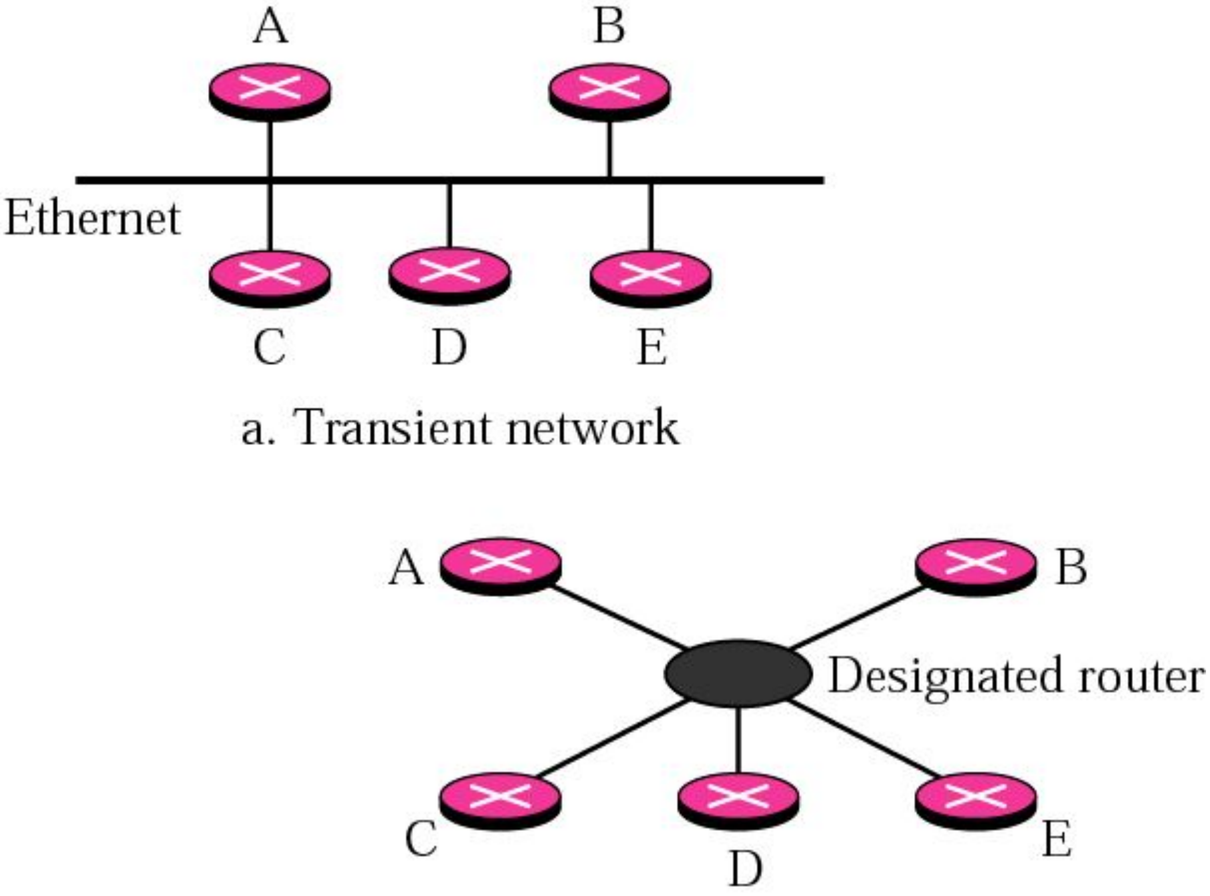


Figure 13-21

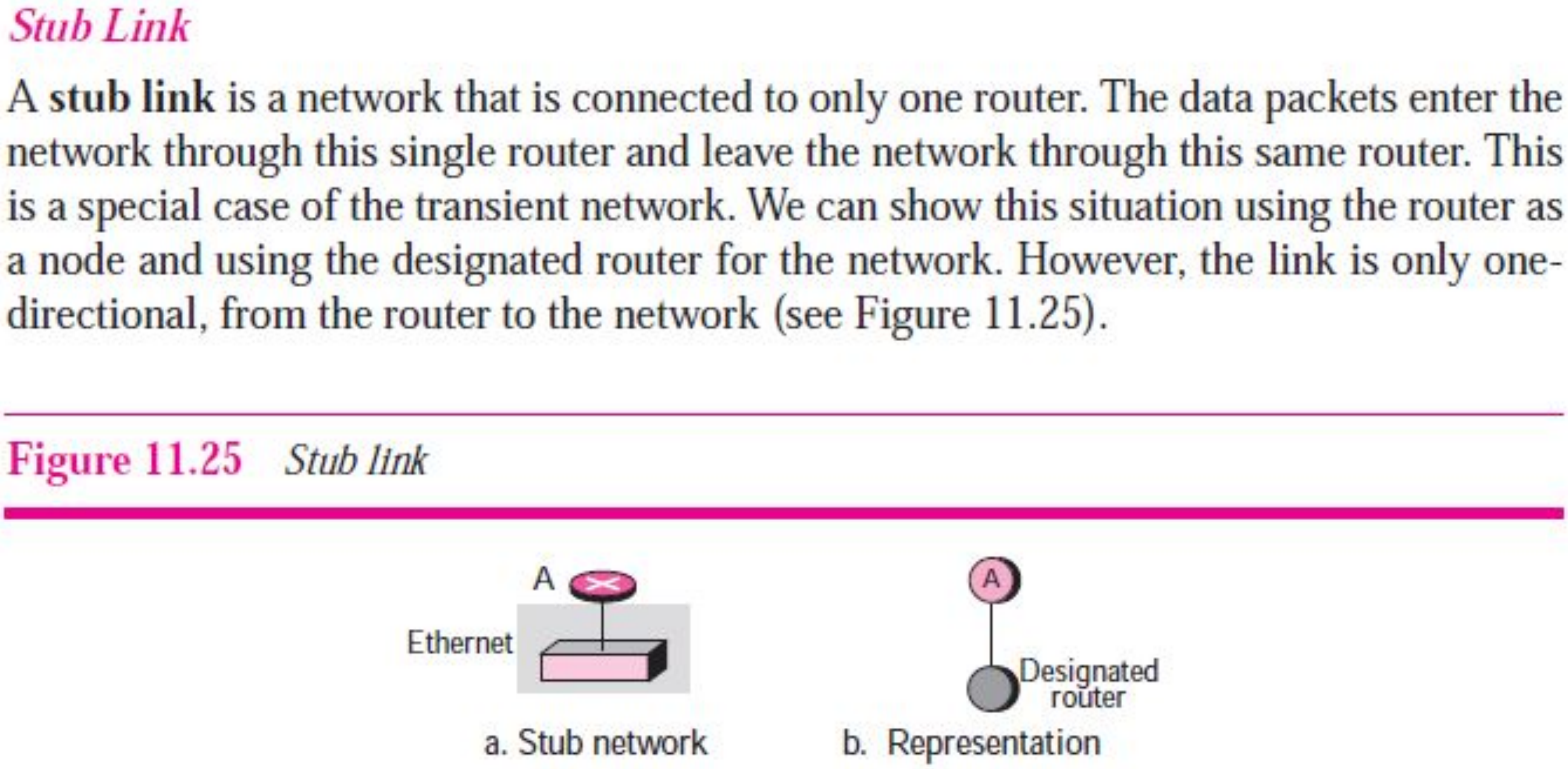
Transient link



a. Transient network

Figure 13-22

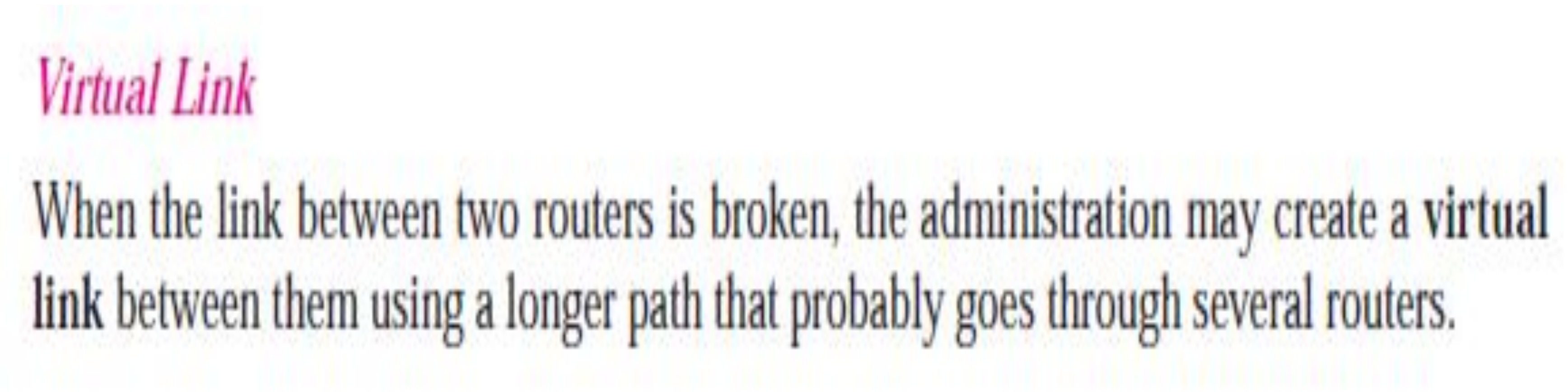
Stub link



Stub Link

A **stub link** is a network that is connected to only one router. The data packets enter the network through this single router and leave the network through this same router. This is a special case of the transient network. We can show this situation using the router as a node and using the designated router for the network. However, the link is only one-directional, from the router to the network (see Figure 11.25).

Figure 11.25 *Stub link*

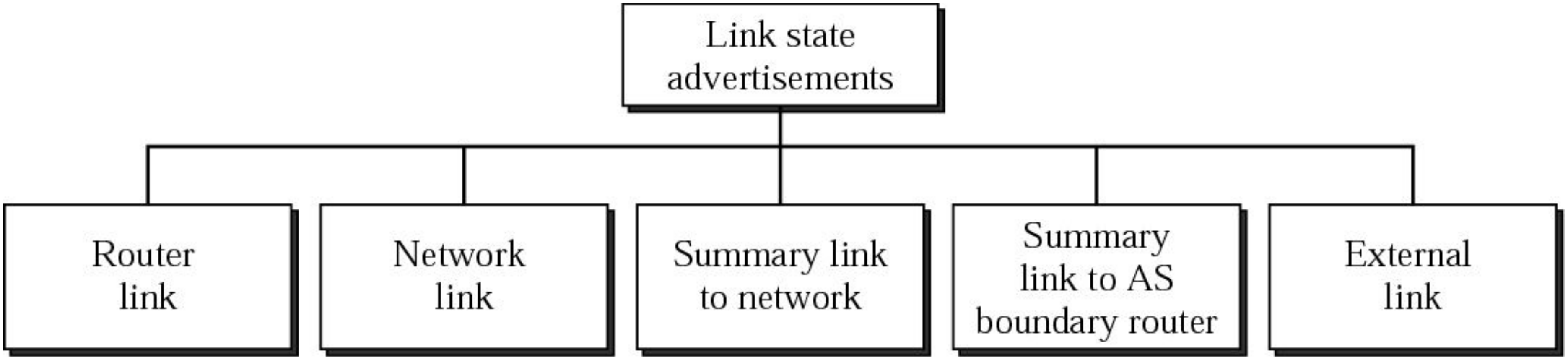


Virtual Link

When the link between two routers is broken, the administration may create a **virtual link** between them using a longer path that probably goes through several routers.

Figure 13-25

Types of LSAs (link state advertisement)


Link state advertisements

Router link

Network link

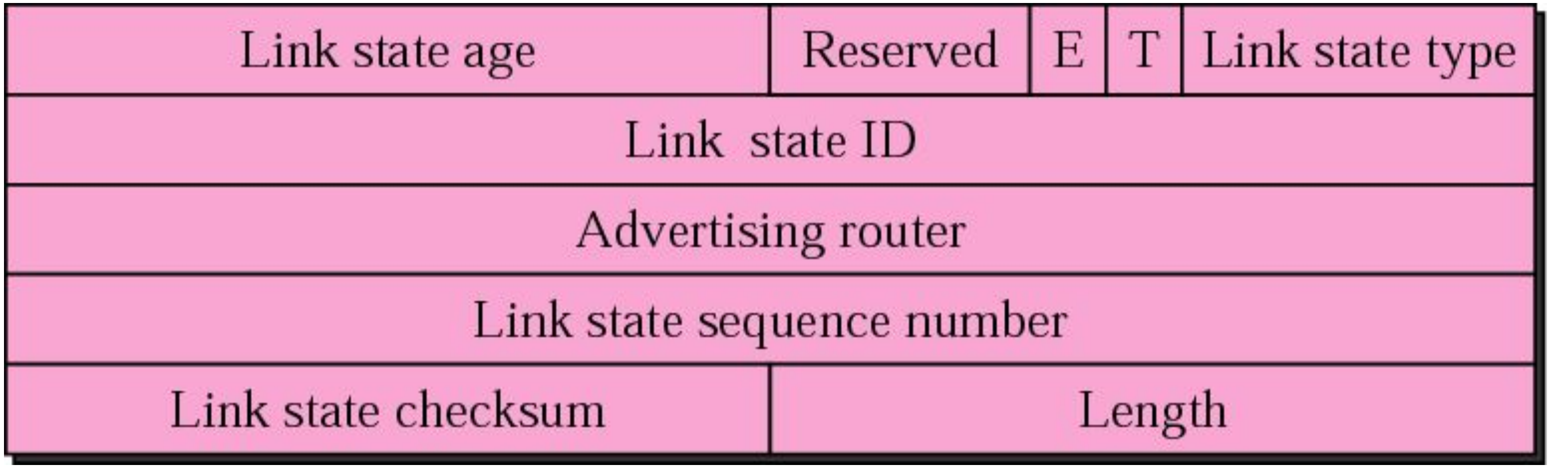
Summary link to network

Summary link to AS boundary router

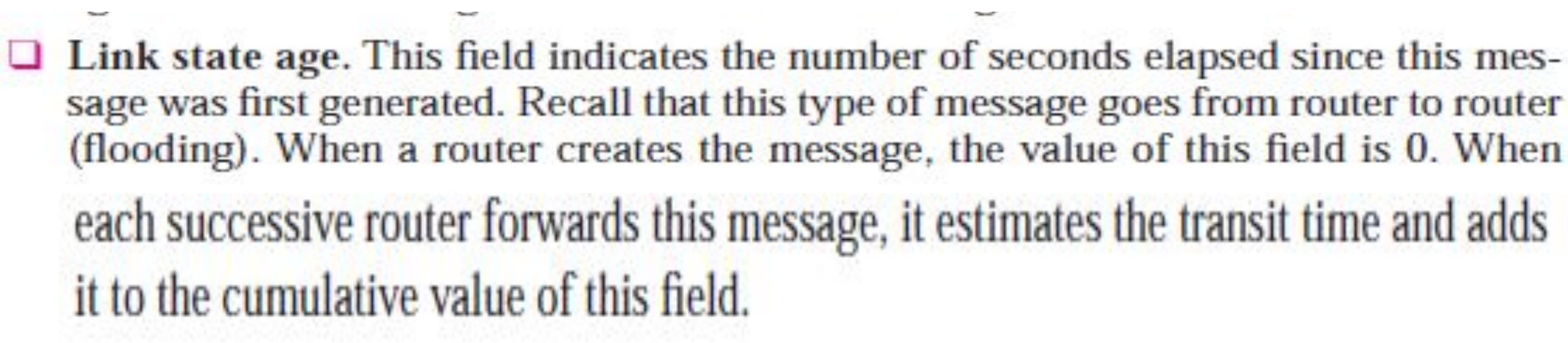
External link

Figure 13-39

LSA header



<table>
<tr><td>Link state age</td><td>Reserved</td><td>E</td><td>T</td><td>Link state type</td></tr>
<tr><td colspan="5">Link state ID</td></tr>
<tr><td colspan="5">Advertising router</td></tr>
<tr><td colspan="5">Link state sequence number</td></tr>
<tr><td>Link state checksum</td><td colspan="4">Length</td></tr>
</table>



□ **Link state age.** This field indicates the number of seconds elapsed since this message was first generated. Recall that this type of message goes from router to router (flooding). 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.

- **E flag.** If this 1-bit flag is set to 1, it means that the area is a stub area. A stub area is an area that is connected to the backbone area by only one path.
- **T flag.** If this 1-bit flag is set to 1, it means that the router can handle multiple types of service.
- **Link state type.** This field defines the LSA type. As we discussed before, there are five different advertisement types: router link (1), network link (2), summary link to network (3), summary link to AS boundary router (4), and external link (5).
- **Link state ID.** The value of this field depends on the type of link. For type 1 (router link), it is the IP address of the router. For type 2 (network link), it is the IP address of the designated router. For type 3 (summary link to network), it is the address of the network. For type 4 (summary link to AS boundary router), it is the IP address of the AS boundary router. For type 5 (external link), it is the address of the external network.
- **Advertising router.** This is the IP address of the router advertising this message.
- **Link state sequence number.** This is a sequence number assigned to each link state update message.

- **Length.** This defines the length of the whole packet in bytes.

Figure 13-25

Types of LSAs (link state advertisement)

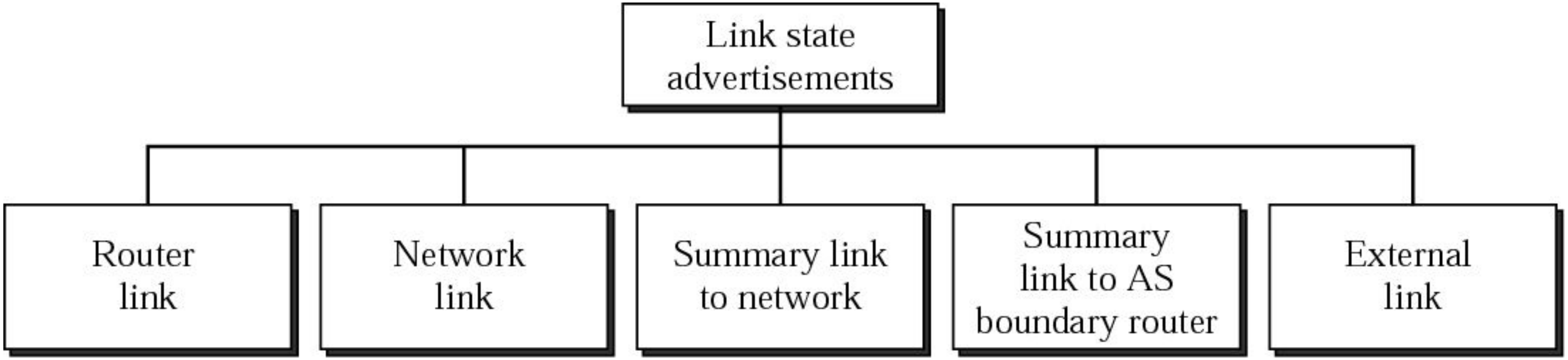


Figure 13-26

Router link

The router link LSA advertises all of the links of a router (true router).

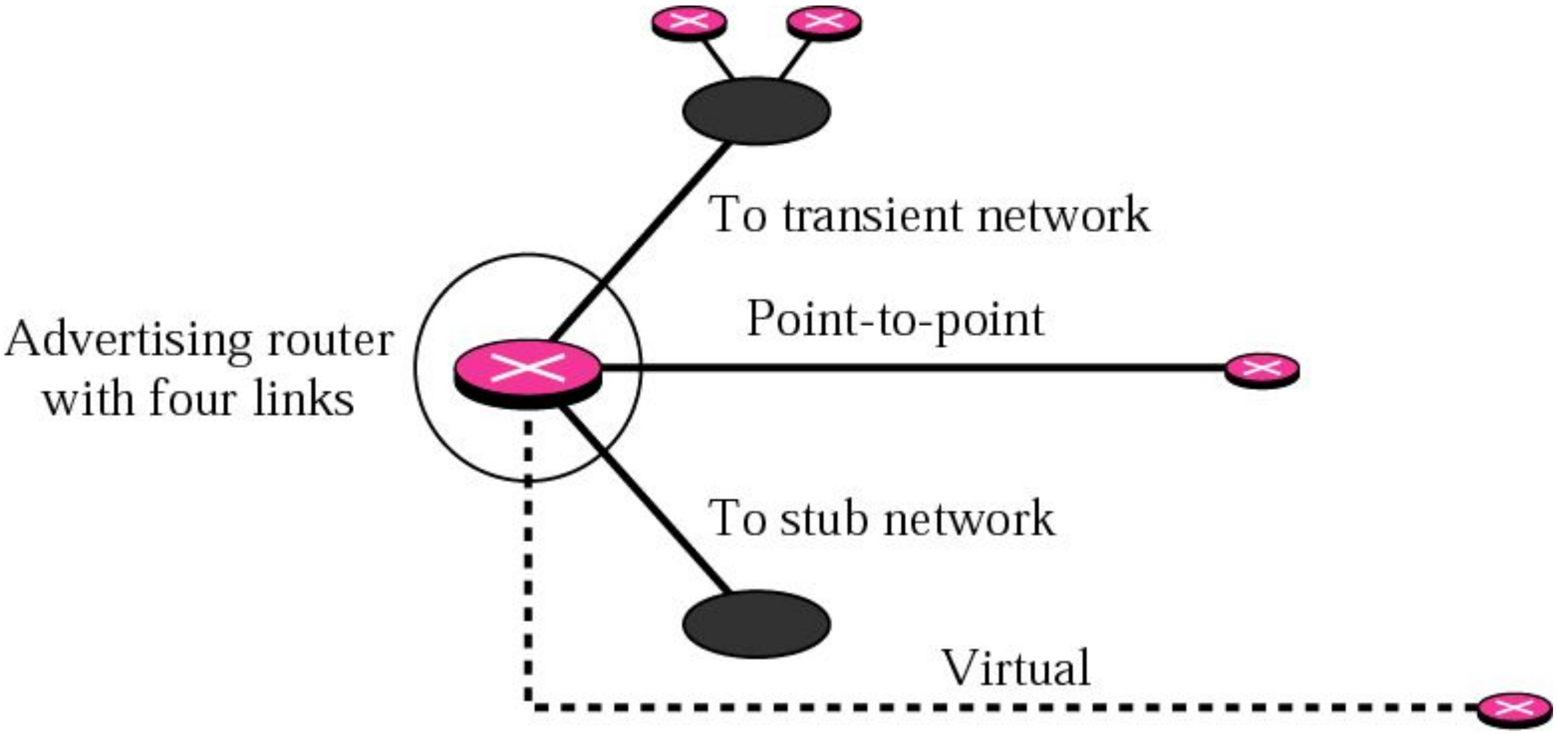


Figure 11.32 *Router link LSA*

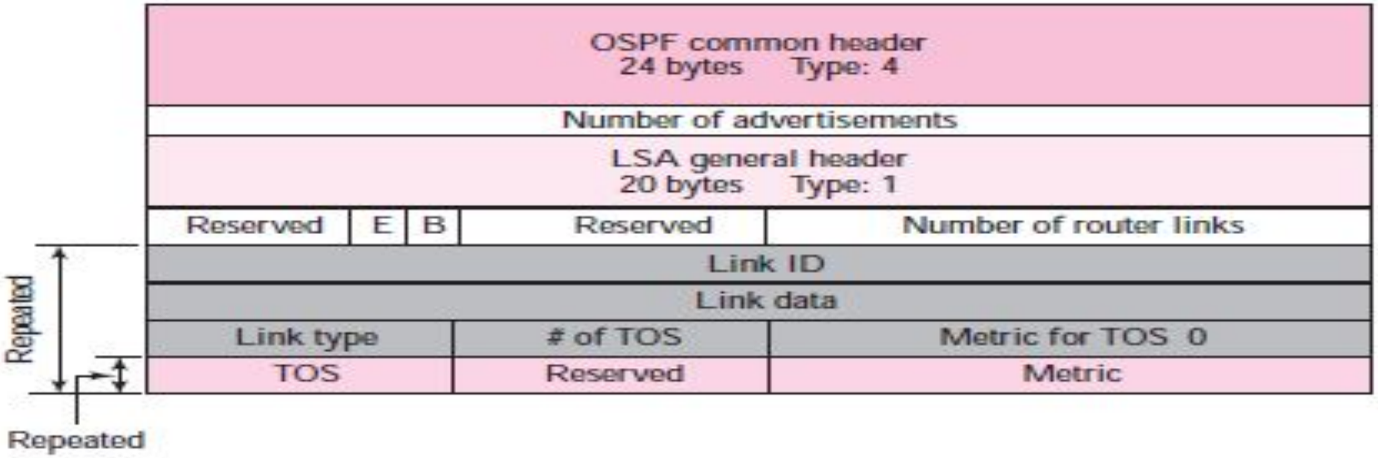


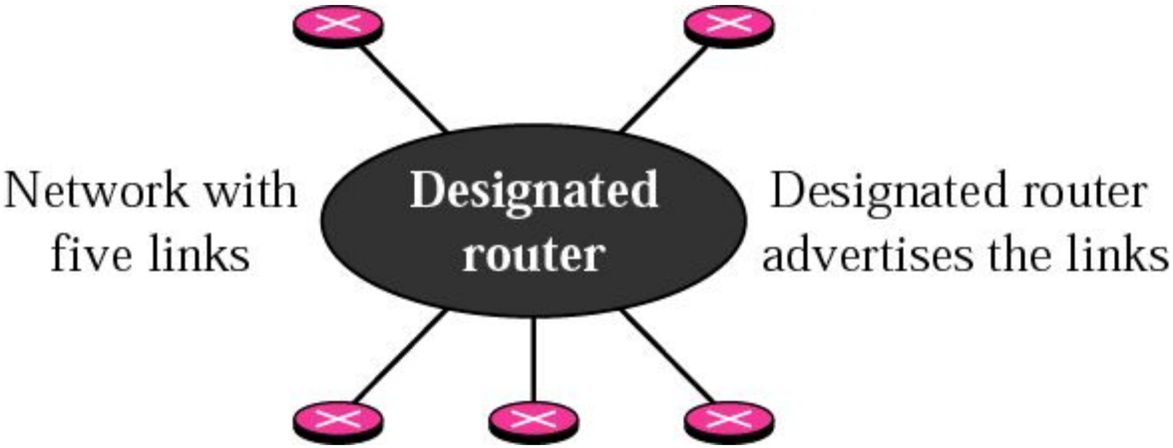
Table 11.5 *Link Types, Link Identification, and Link Data*

Link Type	*Link Identification*	*Link Data*
Type 1: Point-to-point	Address of neighbor router	Interface number
Type 2: Transient	Address of designated router	Router address
Type 3: Stub	Network address	Network mask
Type 4: Virtual	Address of neighbor router	Router address

- **Link type.** Four different types of links are defined based on the type of network to which the router is connected (see Table 11.5).
- **Number of types of service (TOS).** This field defines the number of types of services announced for each link.
- **Metric for TOS 0.** This field defines the metric for the default type of service (TOS 0).
- **TOS.** This field defines the type of service.
- **Metric.** This field defines the metric for the corresponding TOS.

Figure 13-27

Network link



Network link LSA

A network link defines the links of a network. A designated router, on behalf of the transient network, distributes this type of LSP packet.

The packet announces the existence of all of the routers connected to the network.

Figure 13-28

Summary link to network

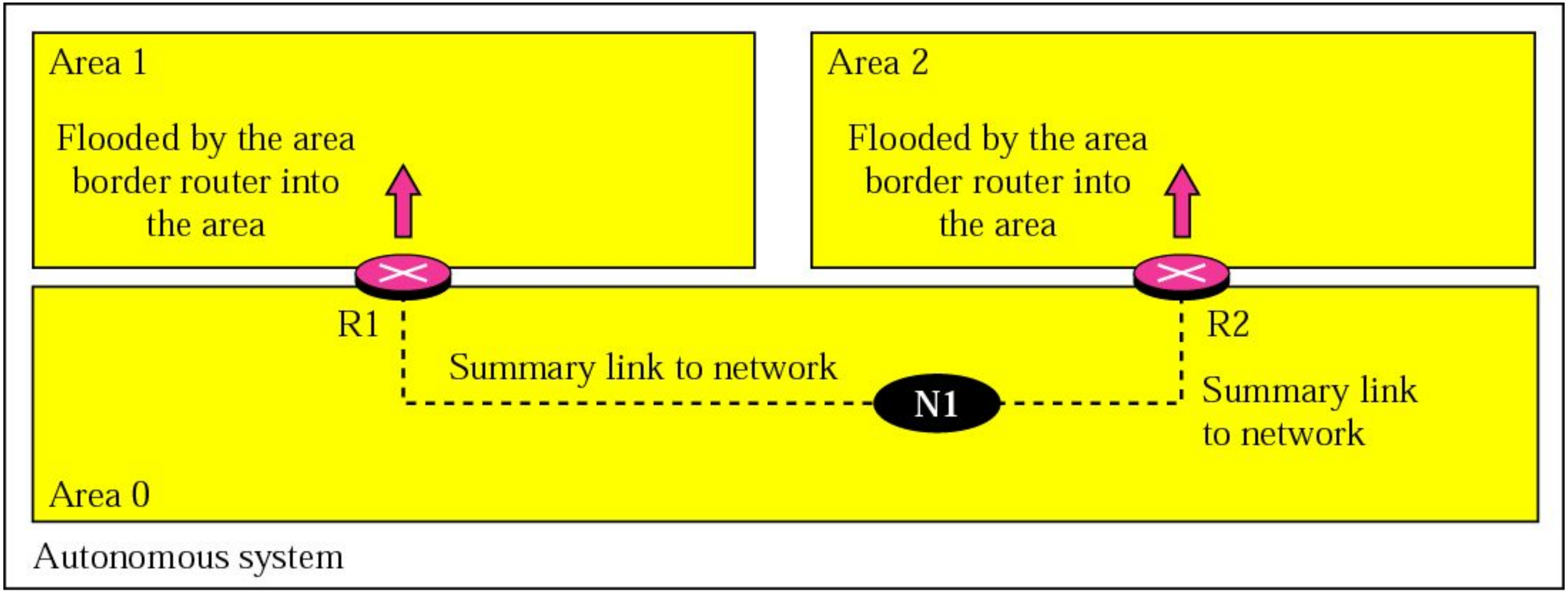

Area 1
Flooded by the area border router into the area
Area 2
Flooded by the area border router into the area
R1
R2
Summary link to network
N1
Summary link to network
Area 0
Autonomous system

Figure 13-29

Summary link to AS boundary router

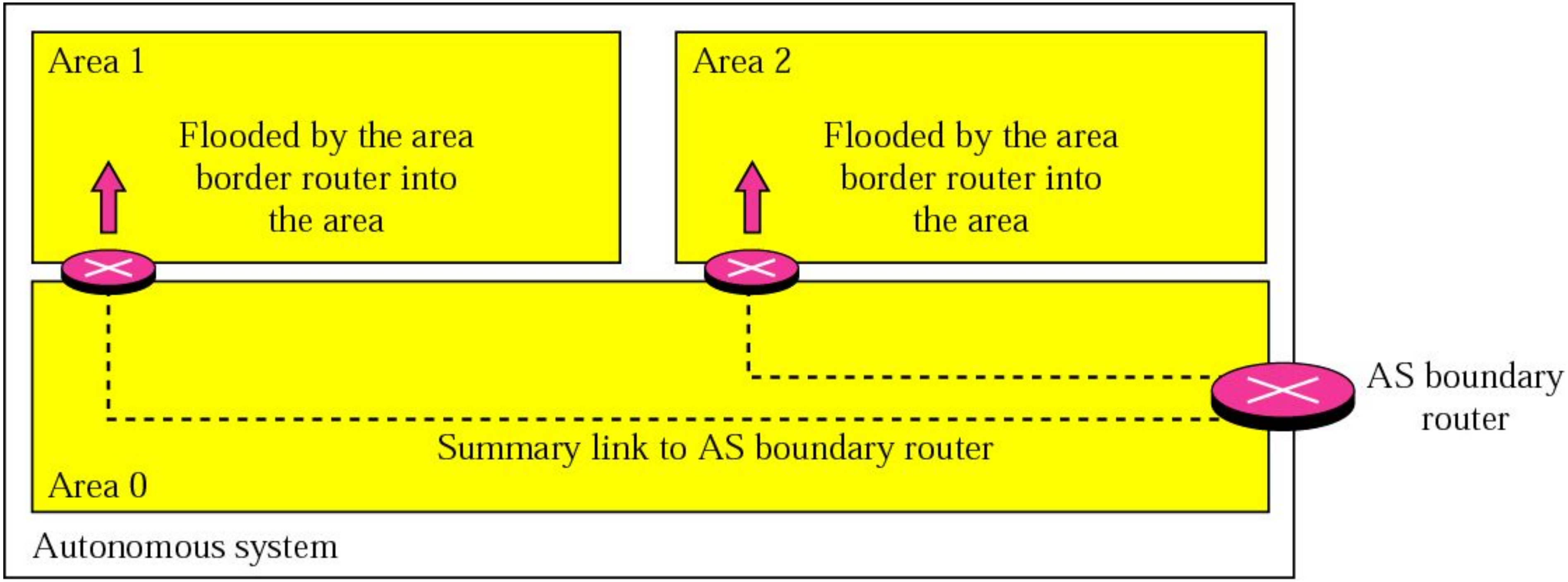
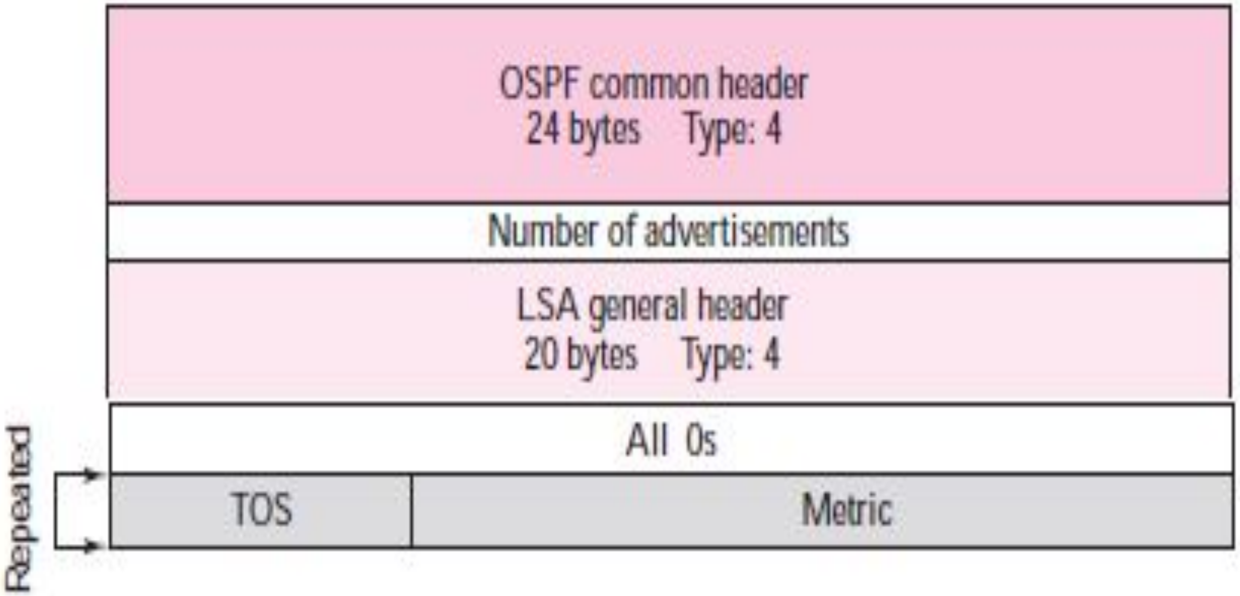

Area 1
Flooded by the area border router into the area
Area 2
Flooded by the area border router into the area
AS boundary router
Summary link to AS boundary router
Area 0
Autonomous system

Figure 11.43 *Summary link to AS boundary router LSA*



OSPF common header
24 bytes Type: 4

Number of advertisements

LSA general header
20 bytes Type: 4

All 0s

TOS

Metric

Repeated

Figure 13-30

External link

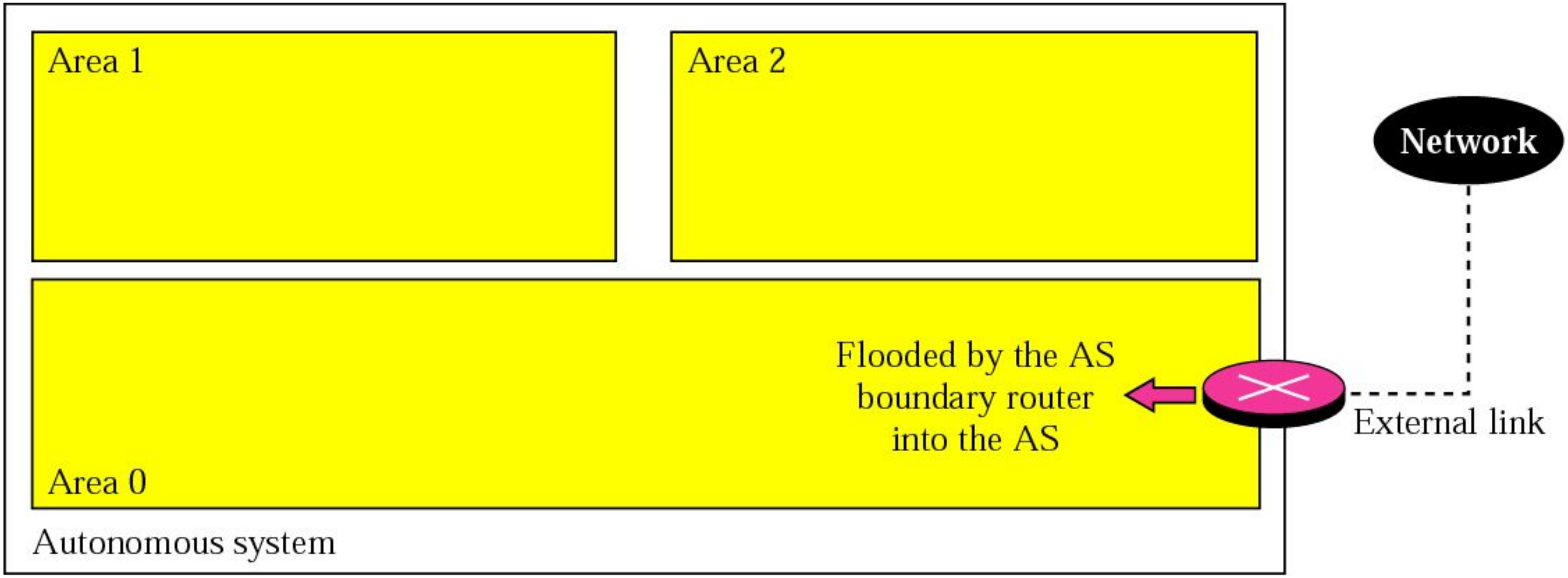
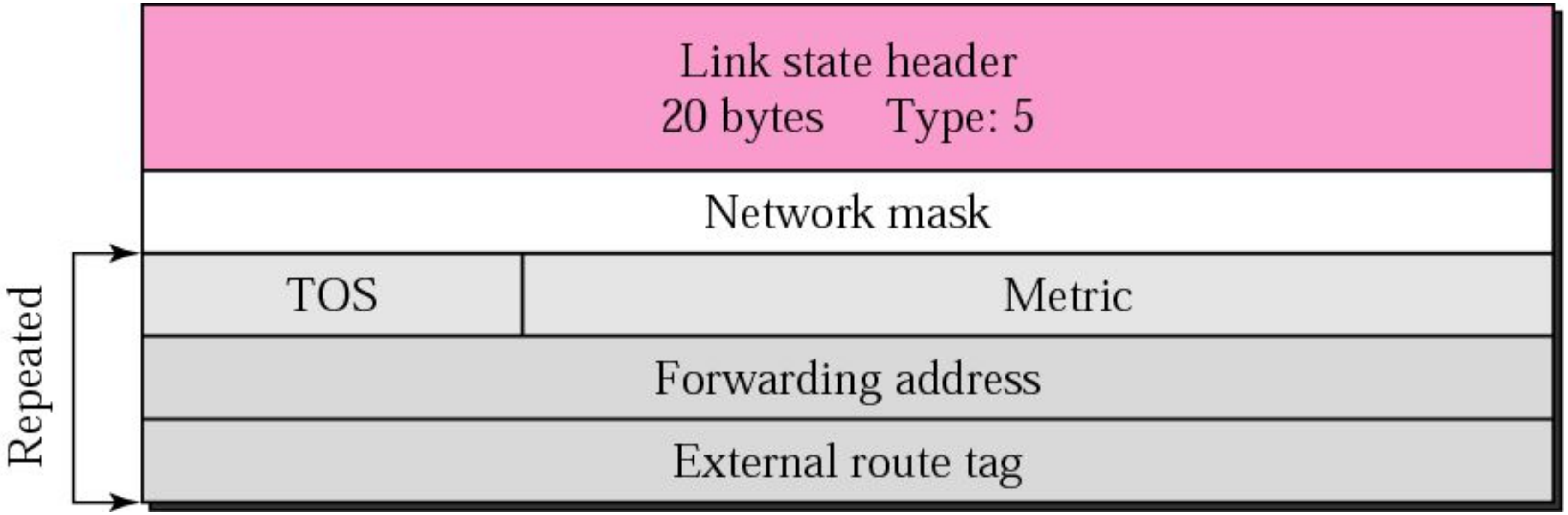


Figure 13-48

External link LSA



Link state header
20 bytes Type: 5

Network mask

TOS | Metric

Forwarding address

External route tag

Repeated

Types of OSPF packets

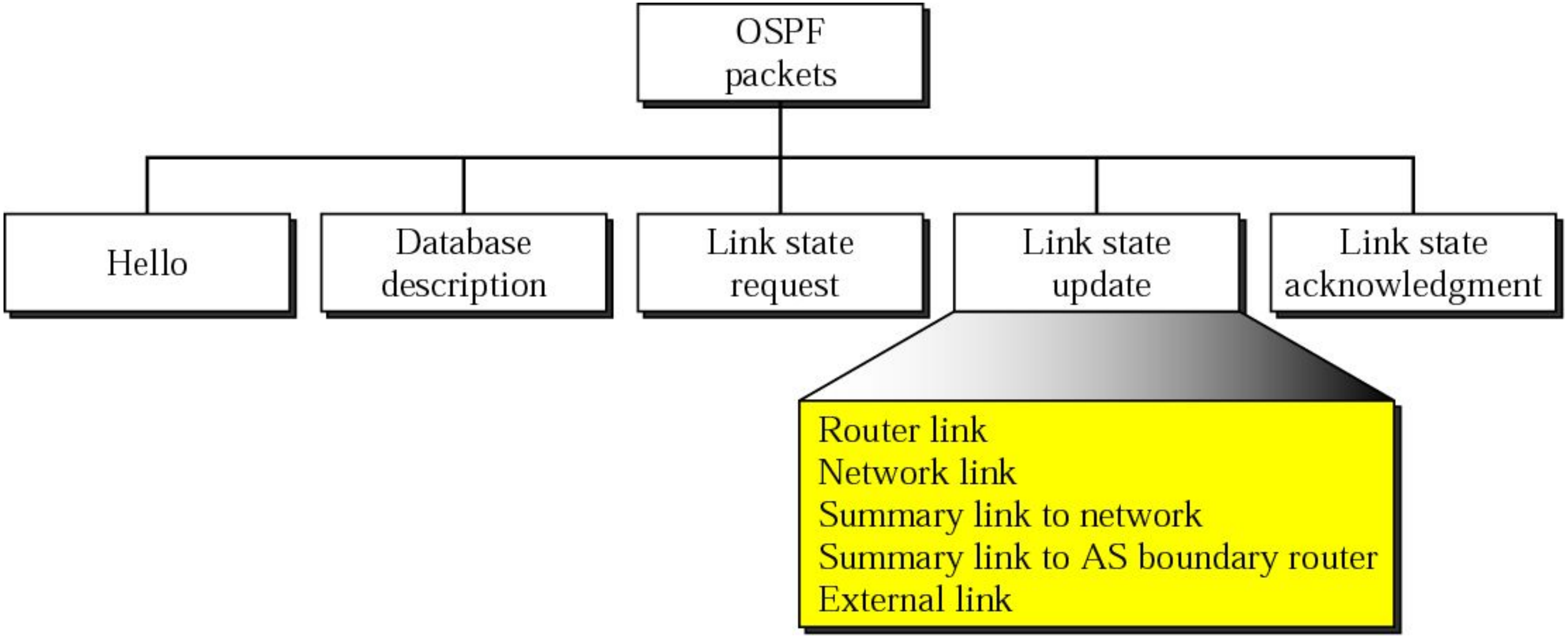
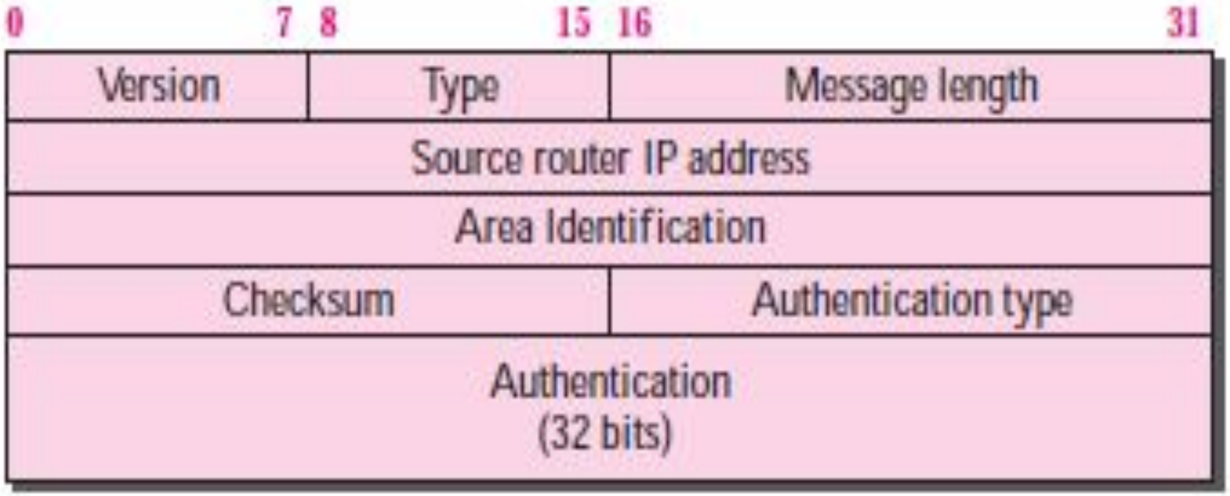
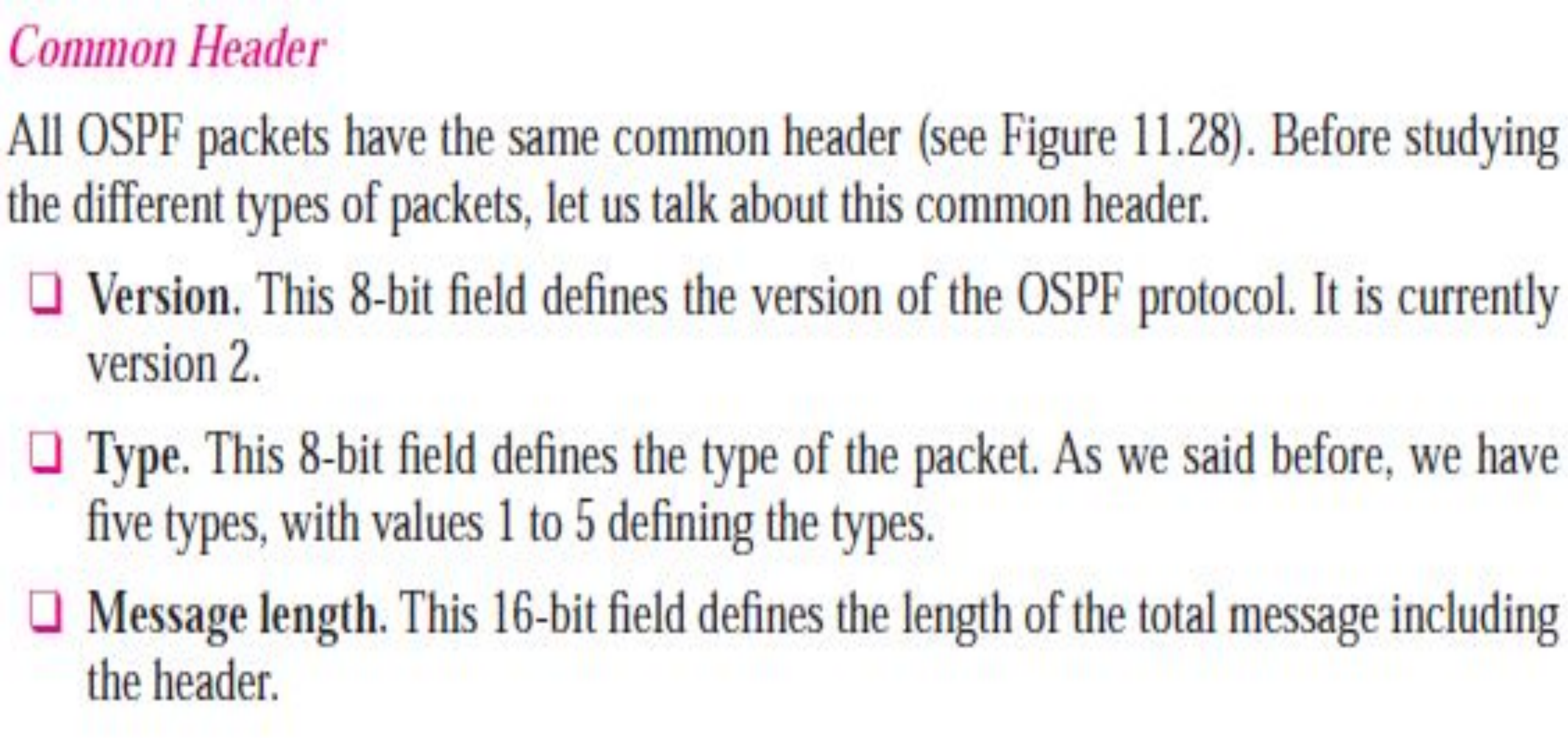


Figure 11.28 *OSPF common header*



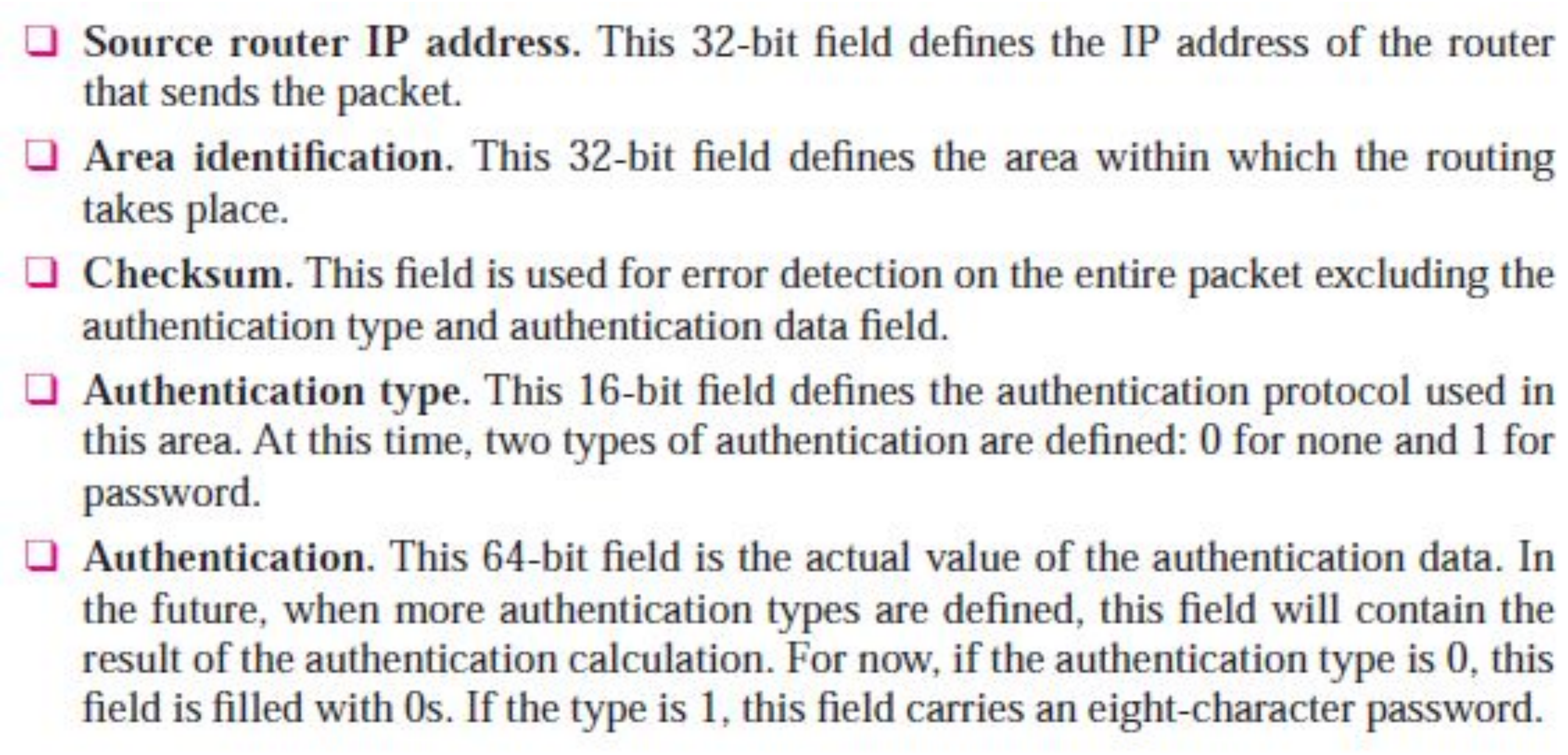
<table>
<tr><th>0 7</th><th>8 15</th><th>16 31</th></tr>
<tr><td>Version</td><td>Type</td><td>Message length</td></tr>
<tr><td colspan="3">Source router IP address</td></tr>
<tr><td colspan="3">Area Identification</td></tr>
<tr><td colspan="2">Checksum</td><td>Authentication type</td></tr>
<tr><td colspan="3">Authentication
(32 bits)</td></tr>
</table>



Common Header

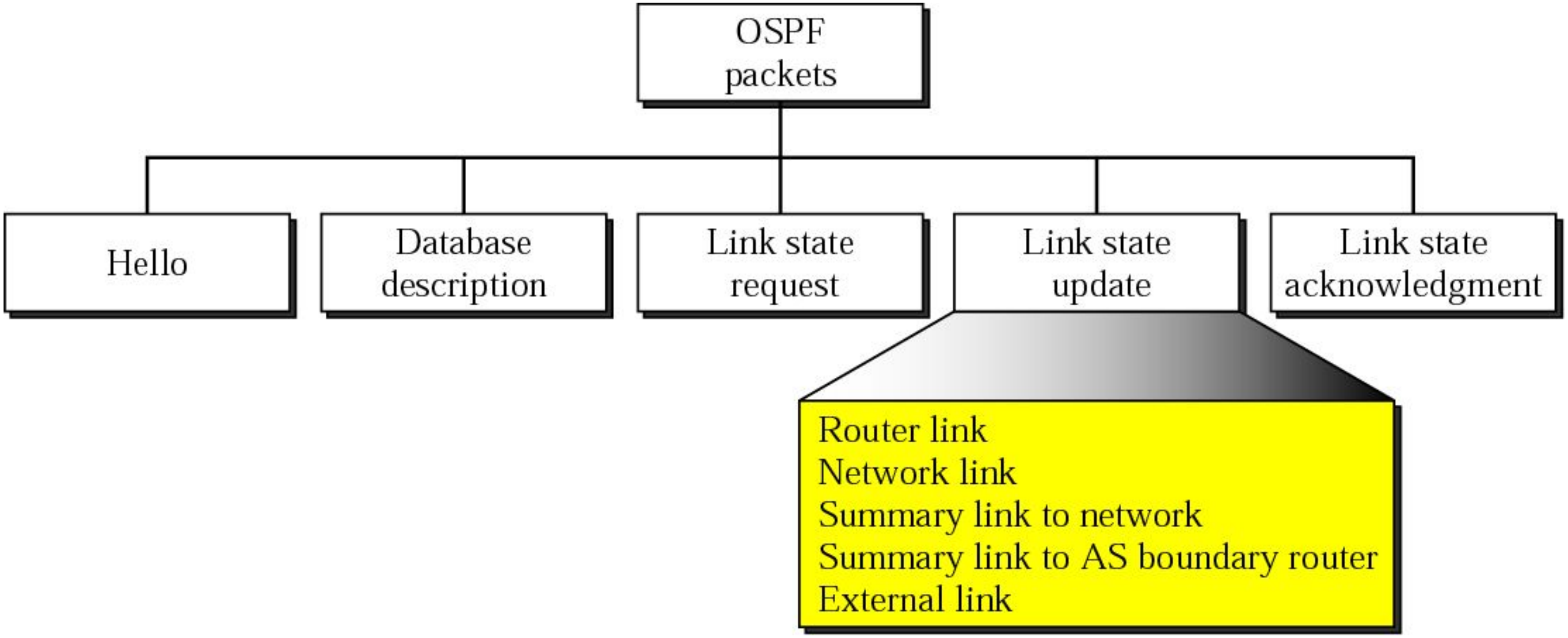
All OSPF packets have the same common header (see Figure 11.28). Before studying the different types of packets, let us talk about this common header.

- ❑ **Version.** This 8-bit field defines the version of the OSPF protocol. It is currently version 2.
- ❑ **Type.** This 8-bit field defines the type of the packet. As we said before, we have five types, with values 1 to 5 defining the types.
- ❑ **Message length.** This 16-bit field defines the length of the total message including the header.



- **Source router IP address.** This 32-bit field defines the IP address of the router that sends the packet.
- **Area identification.** This 32-bit field defines the area within which the routing takes place.
- **Checksum.** This field is used for error detection on the entire packet excluding the authentication type and authentication data field.
- **Authentication type.** This 16-bit field defines the authentication protocol used in this area. At this time, two types of authentication are defined: 0 for none and 1 for password.
- **Authentication.** This 64-bit field is the actual value of the authentication data. In the future, when more authentication types are defined, this field will contain the result of the authentication calculation. For now, if the authentication type is 0, this field is filled with 0s. If the type is 1, this field carries an eight-character password.

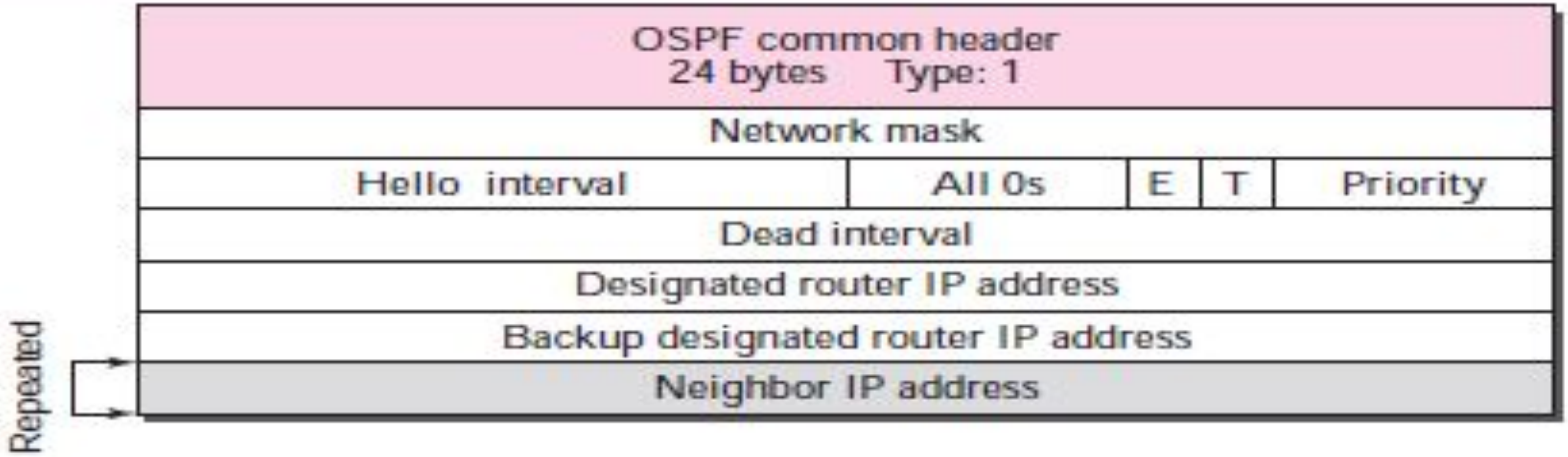
Types of OSPF packets



Hello Message

OSPF uses the **hello message** to create neighborhood relationships and to test the reachability of neighbors. This is the first step in link state routing. Before a router can flood all of the other routers with information about its neighbors, it must first greet its neighbors. It must know if they are alive, and it must know if they are reachable (see Figure 11.46).

11.46 *Hello packet*



OSPF common header
24 bytes Type: 1

Network mask

Hello interval | All 0s | E | T | Priority

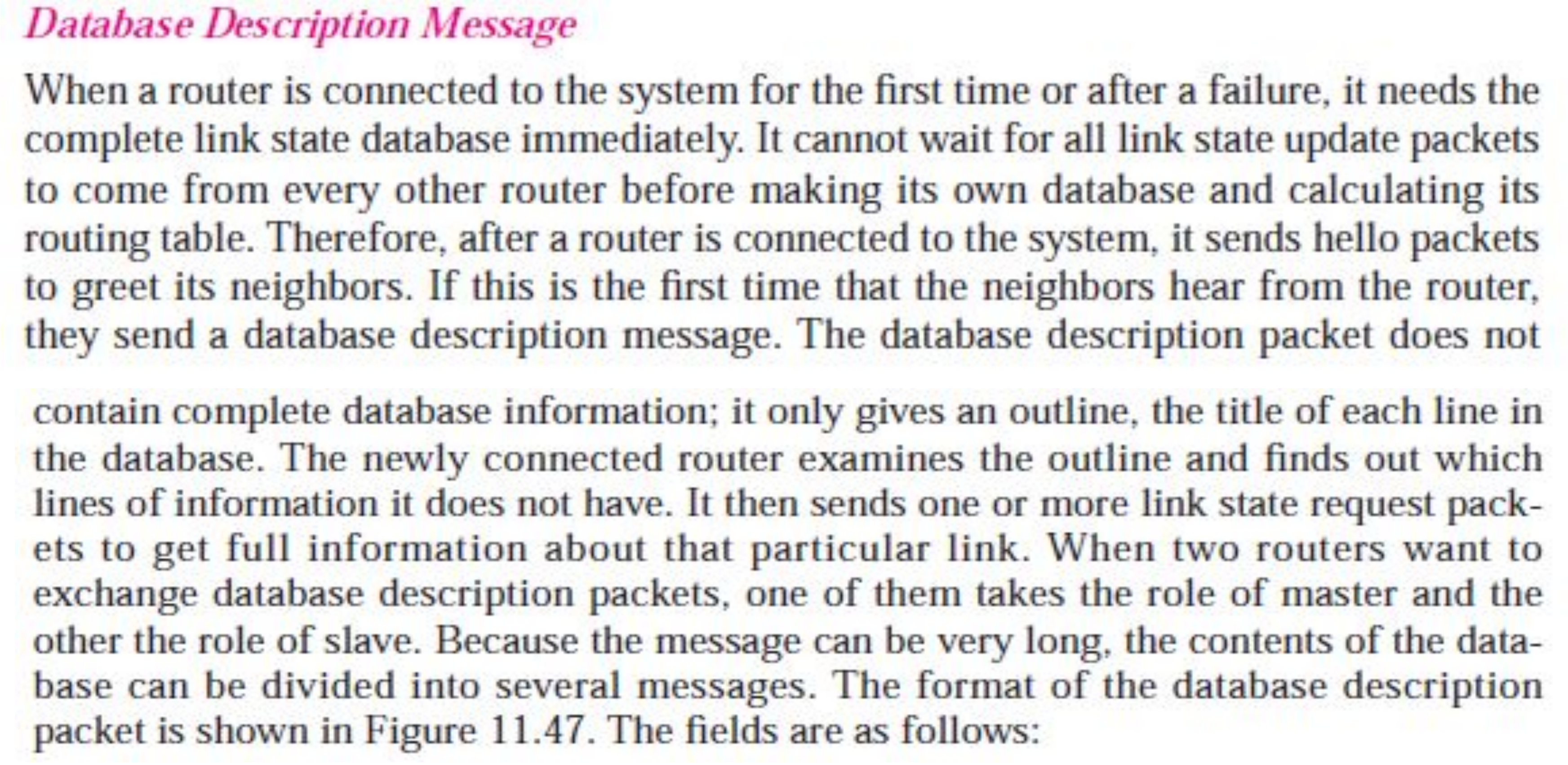
Dead interval

Designated router IP address

Backup designated router IP address

Neighbor IP address

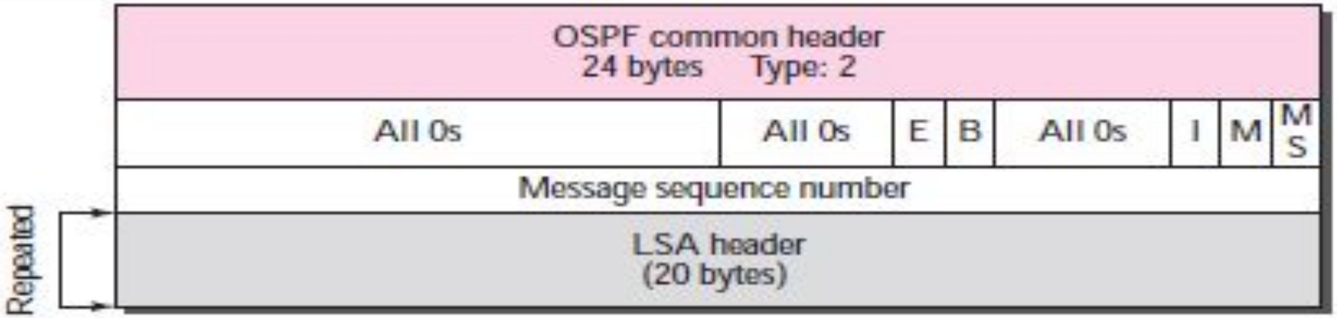
- □ **Network mask.** This 32-bit field defines the network mask of the network over which the hello message is sent.
- □ **Hello interval.** This 16-bit field defines the number of seconds between hello messages.
- □ **E flag.** This is a 1-bit flag. When it is set, it means that the area is a stub area.
- □ **T flag.** This is a 1-bit flag. When it is set, it means that the router supports multiple metrics.
- □ **Priority.** This field defines the priority of the router. The priority determines the selection of the designated router. After all neighbors declare their priorities, the router with the highest priority is chosen as the designated router. The one with the 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.
- □ **Dead interval.** This 32-bit field defines the number of seconds that must pass before a router assumes that a neighbor is dead.
- □ **Designated router IP address.** This 32-bit field is the IP address of the designated router for the network over which the message is sent.
- □ **Backup designated router IP address.** This 32-bit field is the IP address of the backup designated router for the network over which the message is sent.
- □ **Neighbor IP address.** This is a repeated 32-bit field that defines the routers that have agreed to be the neighbors of the sending router. In other words, it is a current list of all the neighbors from which the sending router has received the hello message.



Database Description Message

When a router is connected to the system for the first time or after a failure, it needs the complete link state database immediately. It cannot wait for all link state update packets to come from every other router before making its own database and calculating its routing table. Therefore, after a router is connected to the system, it sends hello packets to greet its neighbors. If this is the first time that the neighbors hear from the router, they send a database description message. The database description packet does not contain complete database information; it only gives an outline, the title of each line in the database. The newly connected router examines the outline and finds out which lines of information it does not have. It then sends one or more link state request packets to get full information about that particular link. When two routers want to exchange database description packets, one of them takes the role of master and the other the role of slave. Because the message can be very long, the contents of the database can be divided into several messages. The format of the database description packet is shown in Figure 11.47. The fields are as follows:

Figure 11.47 *Database description packet*

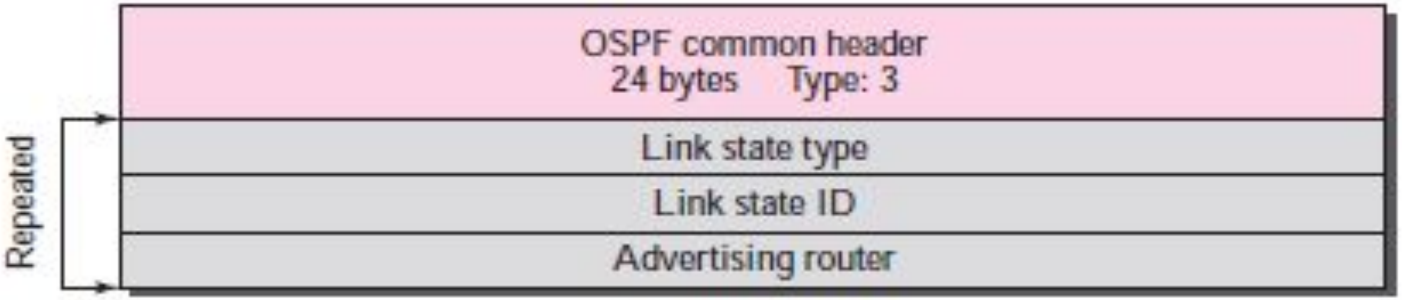


- **E flag.** This 1-bit flag is set to 1 if the advertising router is an autonomous boundary router (*E* stands for external).
- **B flag.** This 1-bit flag is set to 1 if the advertising router is an area border router.
- **I flag.** This 1-bit field, the *initialization* flag, is set to 1 if the message is the first message.
- **M flag.** This 1-bit field, the *more* flag, is set to 1 if this is not the last message.
- **M/S flag.** This 1-bit field, the *master/slave* bit, indicates the origin of the packet: master (M/S = 1) or slave (M/S = 0).
- **Message sequence number.** This 32-bit field contains the sequence number of the message. It is used to match a request with the response.
- **LSA header.** This 20-byte field is used in each LSA. The format of this header is discussed in the link state update message section. This header gives the outline of each link, without details. It is repeated for each link in the link state database.

Link State Request Packet

The format of the **link state request packet** is shown in Figure 11.48. This is a packet that is sent by a router that needs information about a specific route or routes. It is answered with a link state update packet. It can be used by a newly connected router to request more information about some routes after receiving the database description packet. The three fields here are part of the LSA header, which has already been discussed. Each set of the three fields is a request for one single LSA. The set is repeated if more than one advertisement is desired.

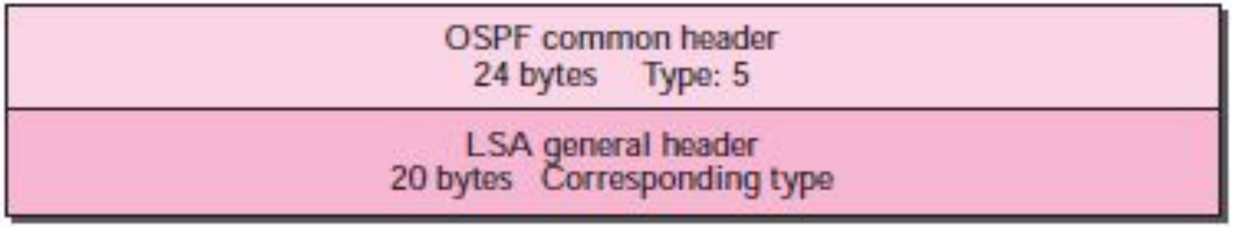
Figure 11.48 *Link state request packet*



Link State Acknowledgment Packet

OSPF makes routing more reliable by forcing every router to acknowledge the receipt of every link state update packet. The format of the **link state acknowledgment packet** is shown in Figure 11.49. It has the common OSPF header and the general LSA header. These two sections are sufficient to acknowledge a packet.

Figure 11.49 *Link state acknowledgment packet*



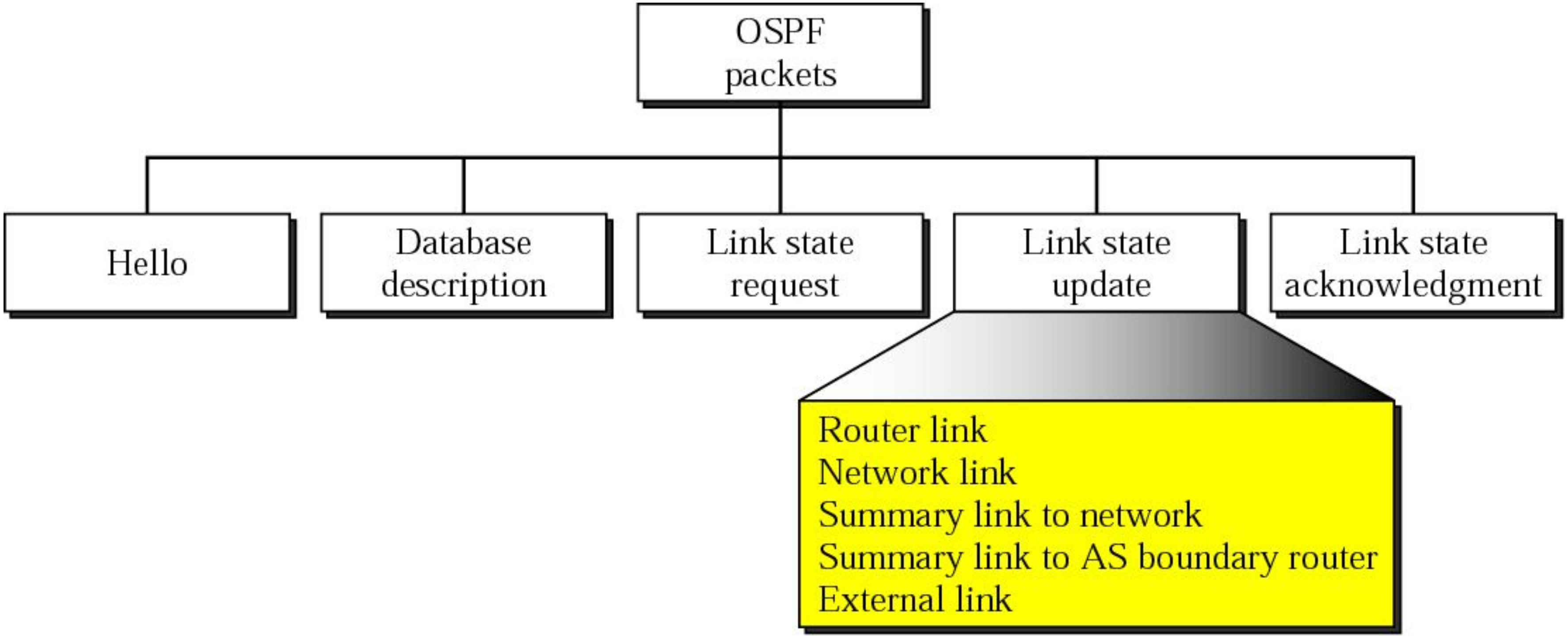
Encapsulation

OSPF packets are encapsulated in IP datagrams. They contain the acknowledgment mechanism for flow and error control. They do not need a transport layer protocol to provide these services.

OSPF packets are encapsulated in IP datagrams.

Figure 13-33

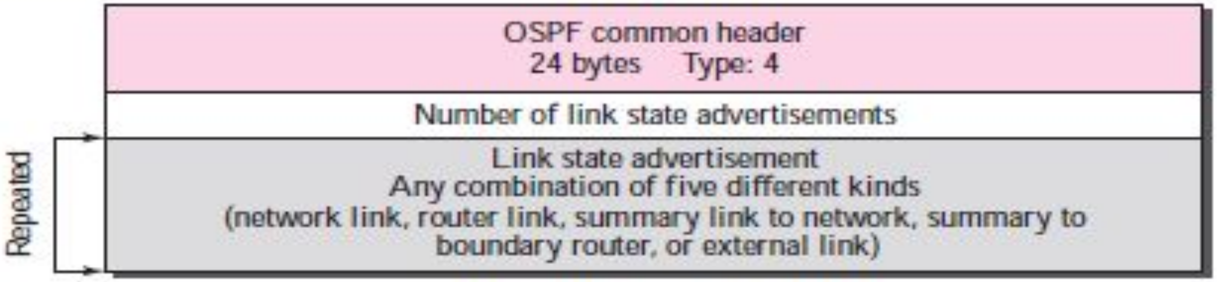
Types of OSPF packets



Link State Update Packet

We first discuss the **link state update packet,** the heart of the OSPF operation. It is used by a router to advertise the states of its links. The general format of the link state update packet is shown in Figure 11.29.

Figure 11.29 *Link state update packet*


OSPF common header
24 bytes Type: 4
Number of link state advertisements
Link state advertisement
Any combination of five different kinds
(network link, router link, summary link to network, summary to boundary router, or external link)
Repeated

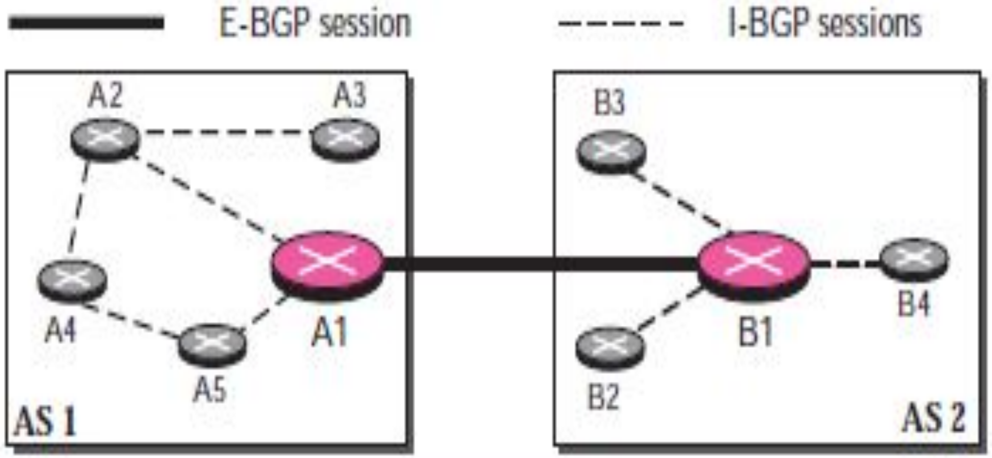
13.4

BGP: Border Gateway Protocol

BGP SESSIONS

- BGP can have two types of sessions: external BGP (E-BGP) and internal BGP (I-BGP) sessions.
- The E-BGP session is used to exchange information between two speaker nodes belonging to two different autonomous systems.
- The I-BGP session, on the other hand, is used to exchange routing information between two routers inside an autonomous system.

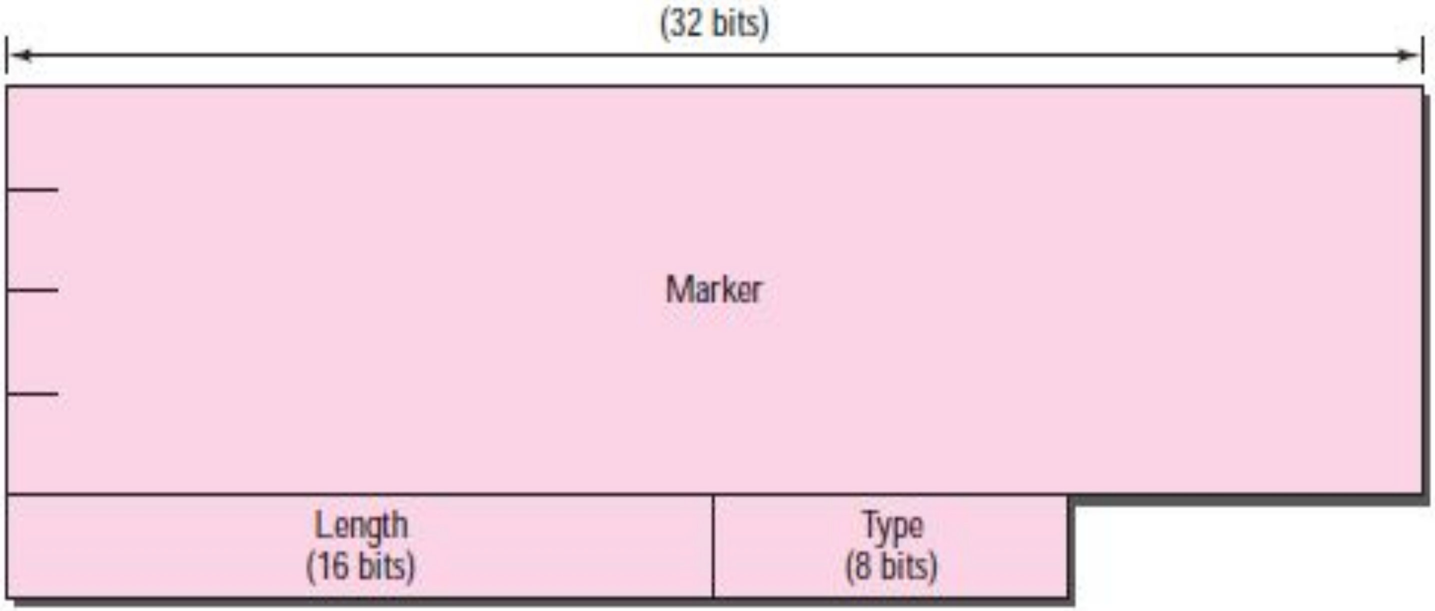
Figure 11.53 *Internal and external BGP sessions*

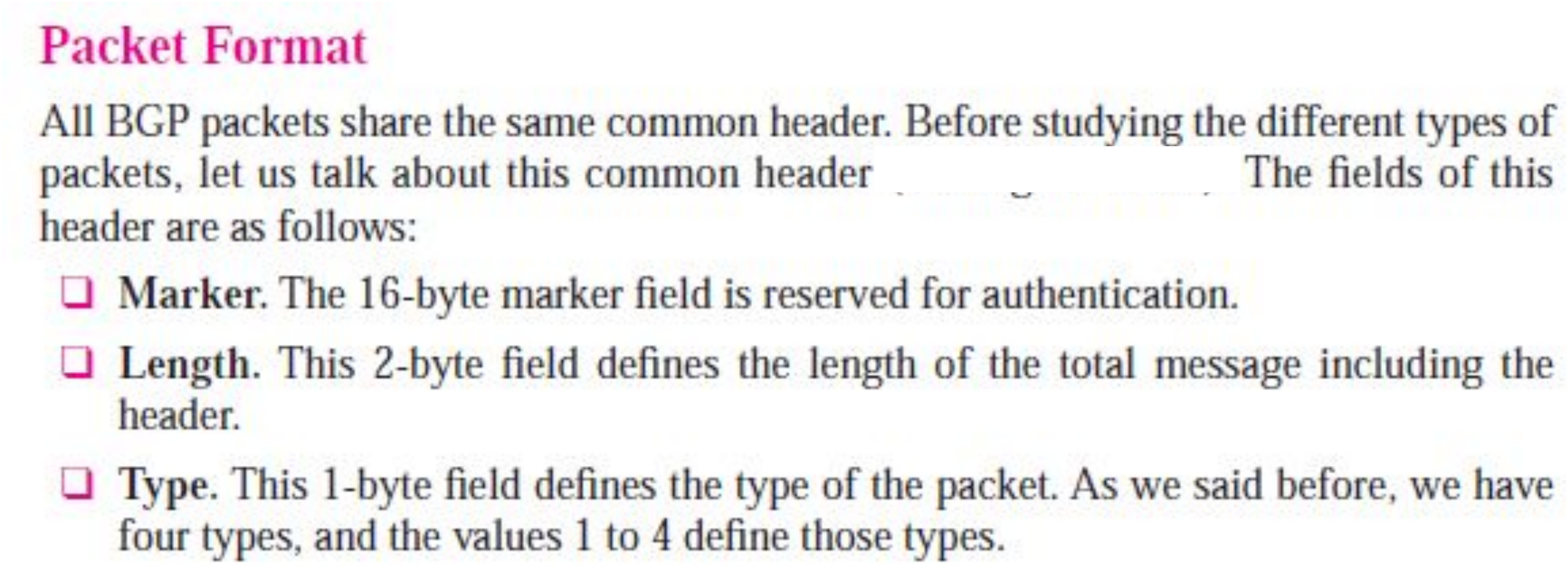


The session established between AS1 and AS2 is an E-BGP session. The two speaker routers exchange information they know about networks in the Internet. However, these two routers need to collect information from other routers in the autonomous systems. This is done using I-BGP sessions.

BGP PACKET HEADER

Figure 11.55 *BGP packet header*



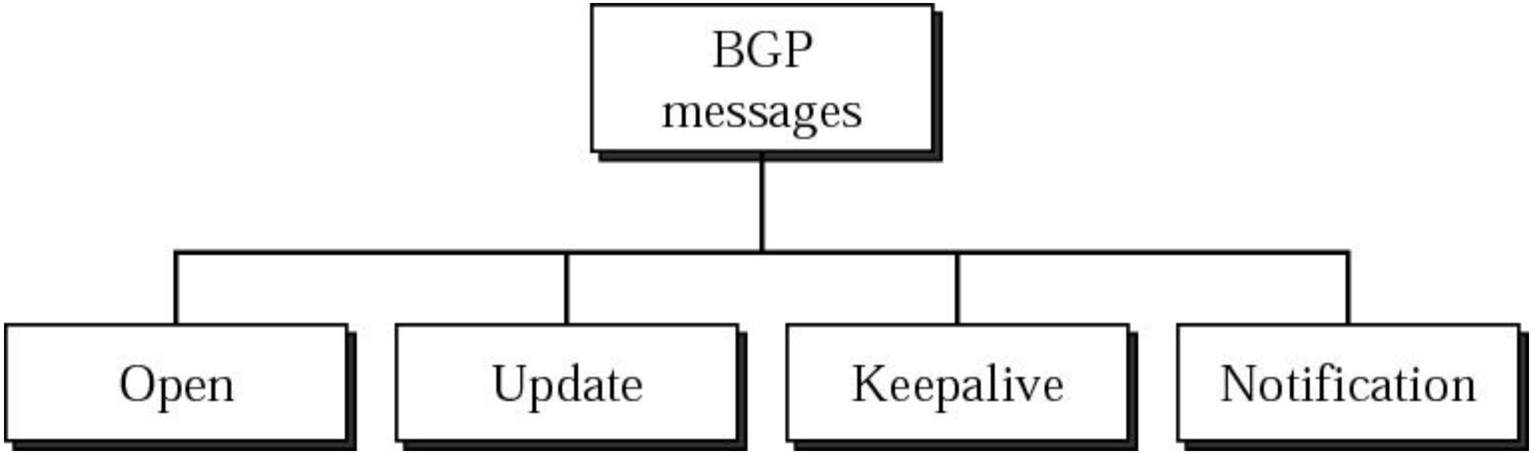


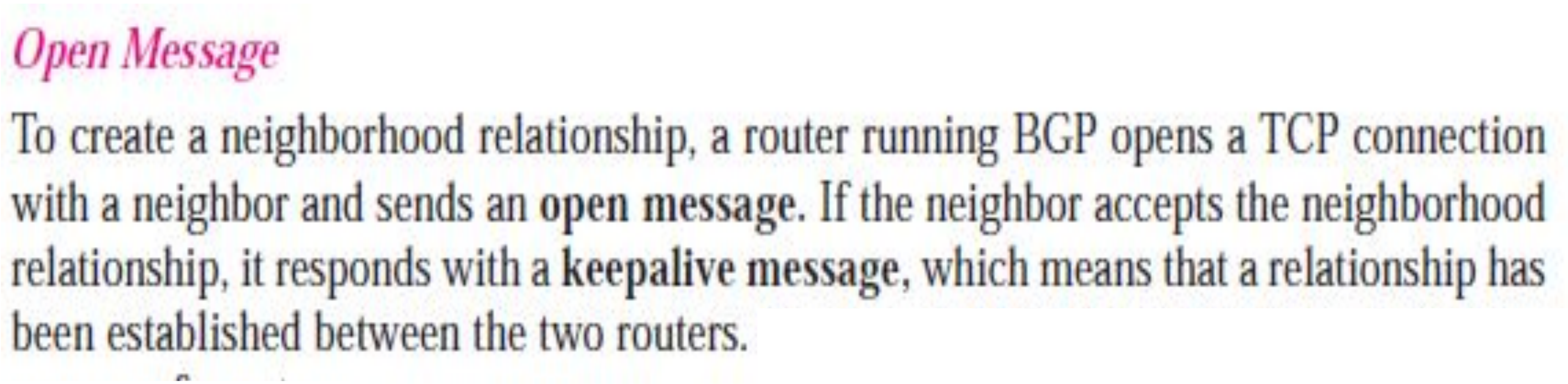
Packet Format

All BGP packets share the same common header. Before studying the different types of packets, let us talk about this common header The fields of this header are as follows:

- □ **Marker.** The 16-byte marker field is reserved for authentication.
- □ **Length.** This 2-byte field defines the length of the total message including the header.
- □ **Type.** This 1-byte field defines the type of the packet. As we said before, we have four types, and the values 1 to 4 define those types.

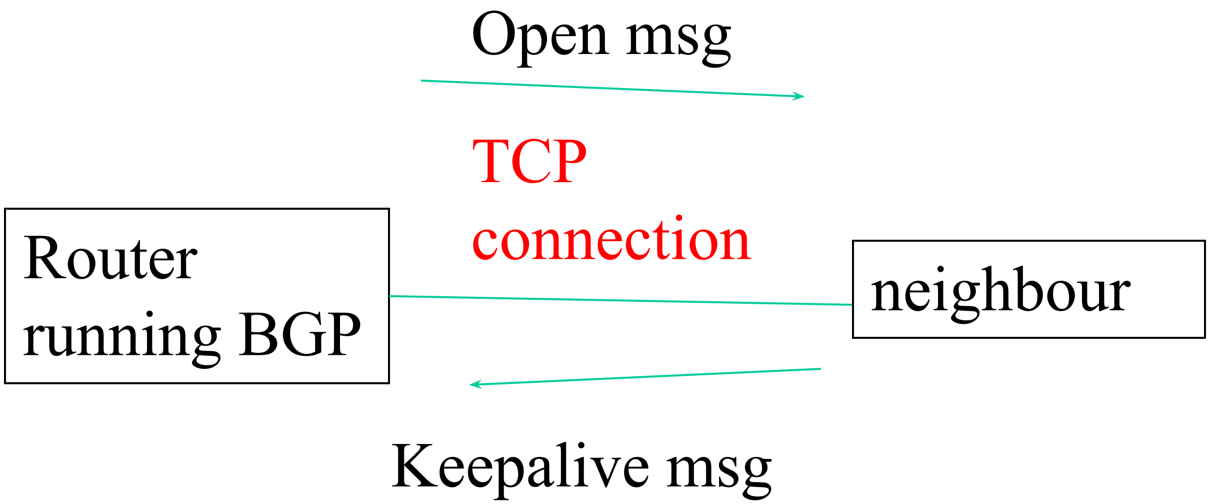
Types of BGP messages





Open Message

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



Open msg

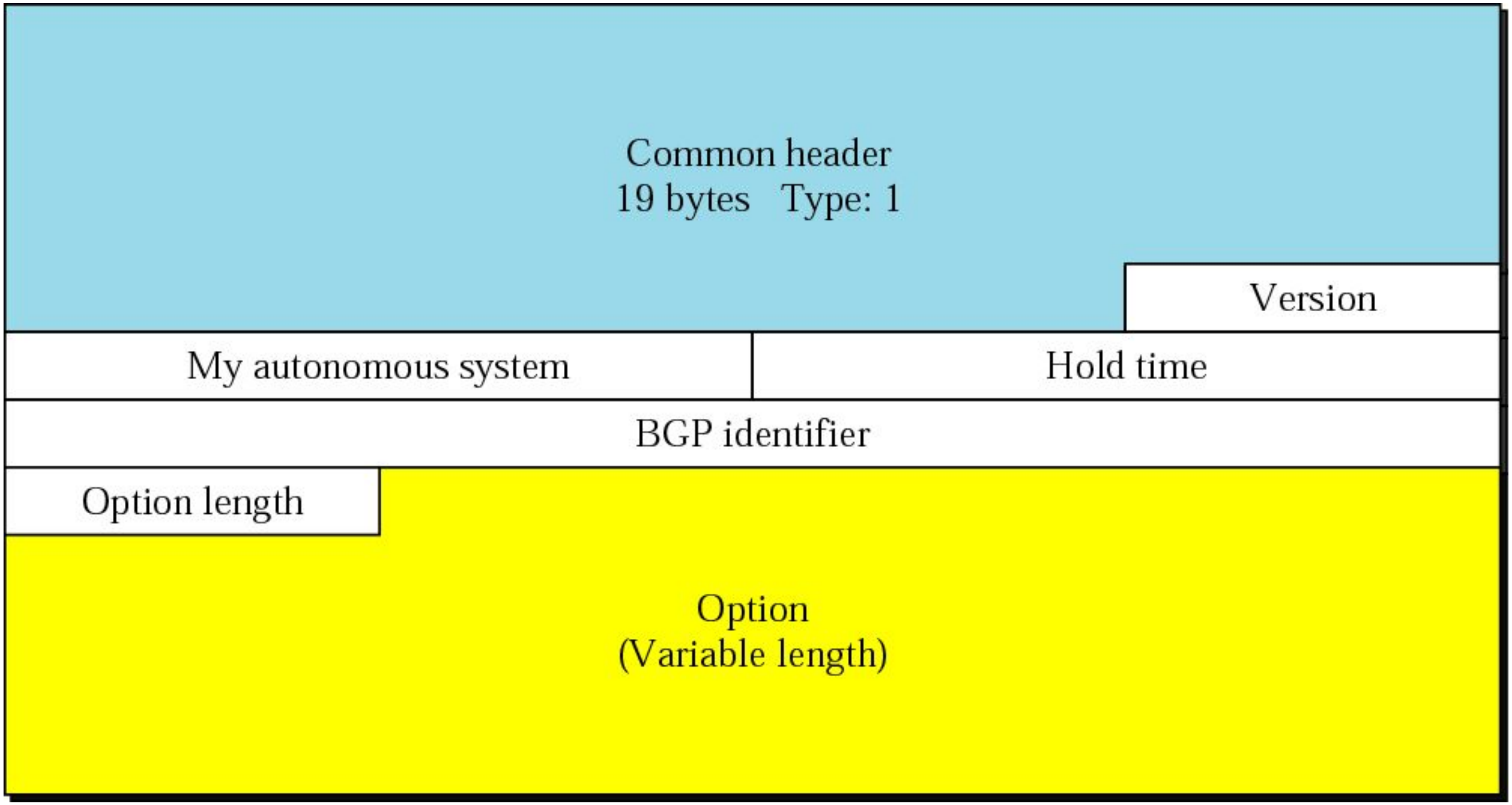
TCP connection

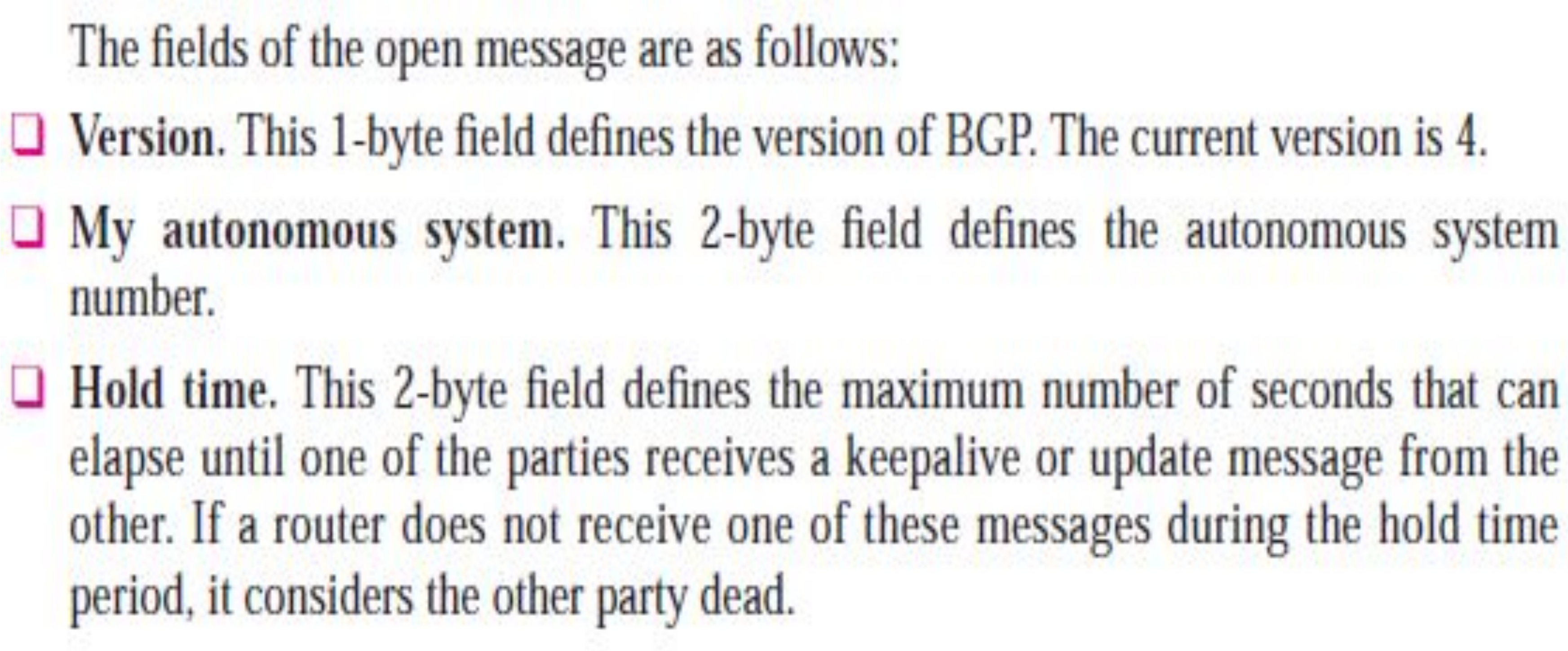
Router running BGP

neighbour

Keepalive msg

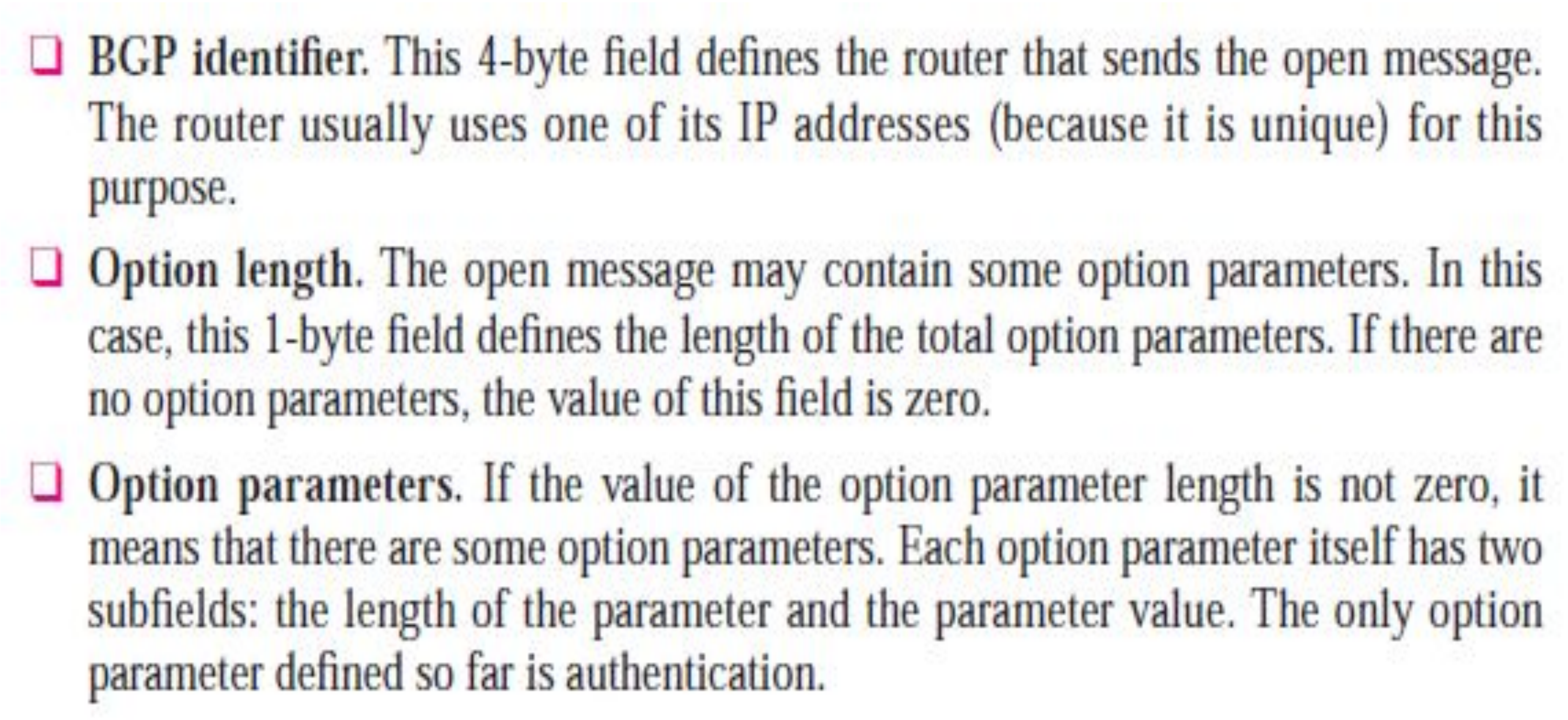
Open message



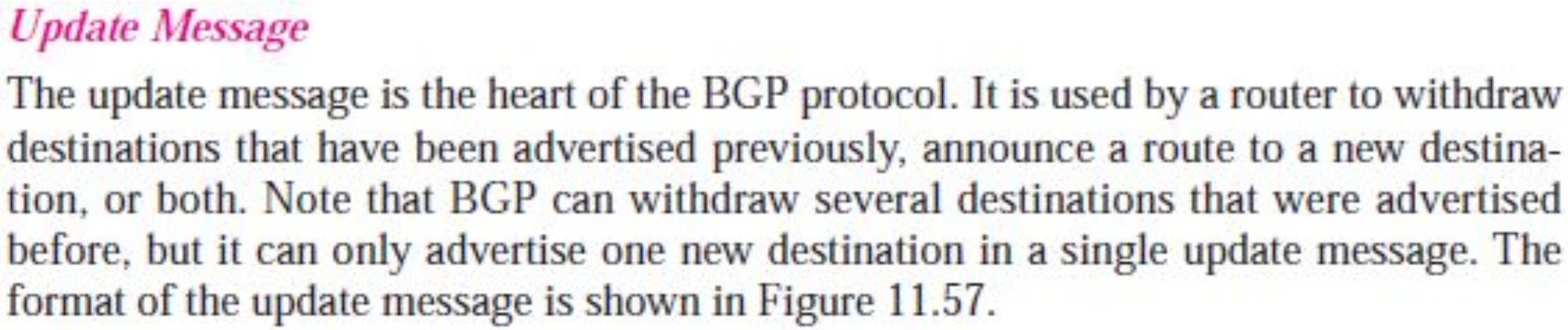


The fields of the open message are as follows:

- □ **Version.** This 1-byte field defines the version of BGP. The current version is 4.
- □ **My autonomous system.** This 2-byte field defines the autonomous system number.
- □ **Hold time.** This 2-byte field defines the maximum number of seconds that can elapse until one of the parties receives a keepalive or update message from the other. If a router does not receive one of these messages during the hold time period, it considers the other party dead.



- **BGP identifier.** This 4-byte field defines the router that sends the open message. The router usually uses one of its IP addresses (because it is unique) for this purpose.
- **Option length.** The open message may contain some option parameters. In this case, this 1-byte field defines the length of the total option parameters. If there are no option parameters, the value of this field is zero.
- **Option parameters.** If the value of the option parameter length is not zero, it means that there are some option parameters. Each option parameter itself has two subfields: the length of the parameter and the parameter value. The only option parameter defined so far is authentication.

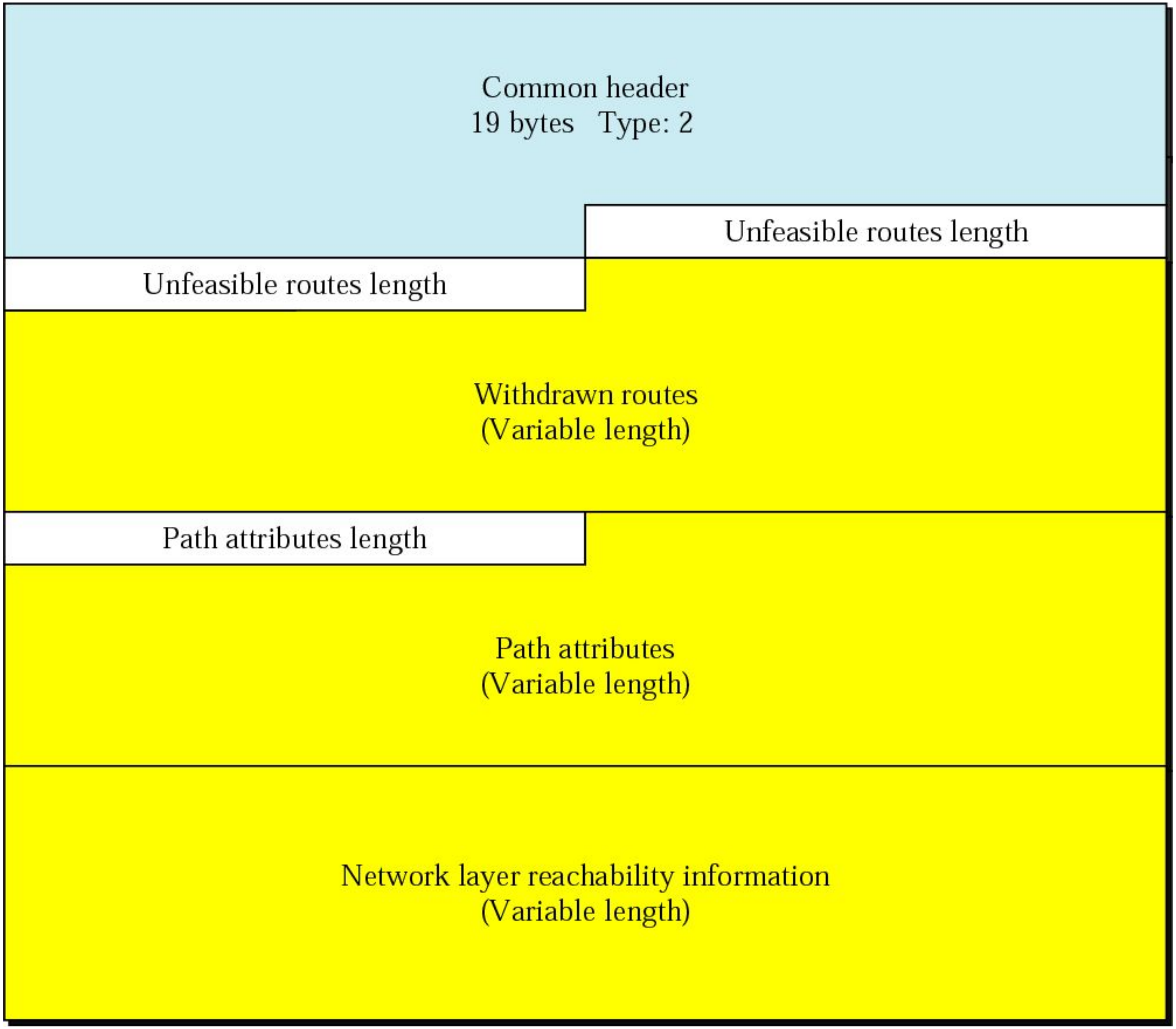


Update Message

The update message is the heart of the BGP protocol. It is used by a router to withdraw destinations that have been advertised previously, announce a route to a new destination, or both. Note that BGP can withdraw several destinations that were advertised before, but it can only advertise one new destination in a single update message. The format of the update message is shown in Figure 11.57.

Figure 13-54

Update message



Common header
19 bytes Type: 2

Unfeasible routes length

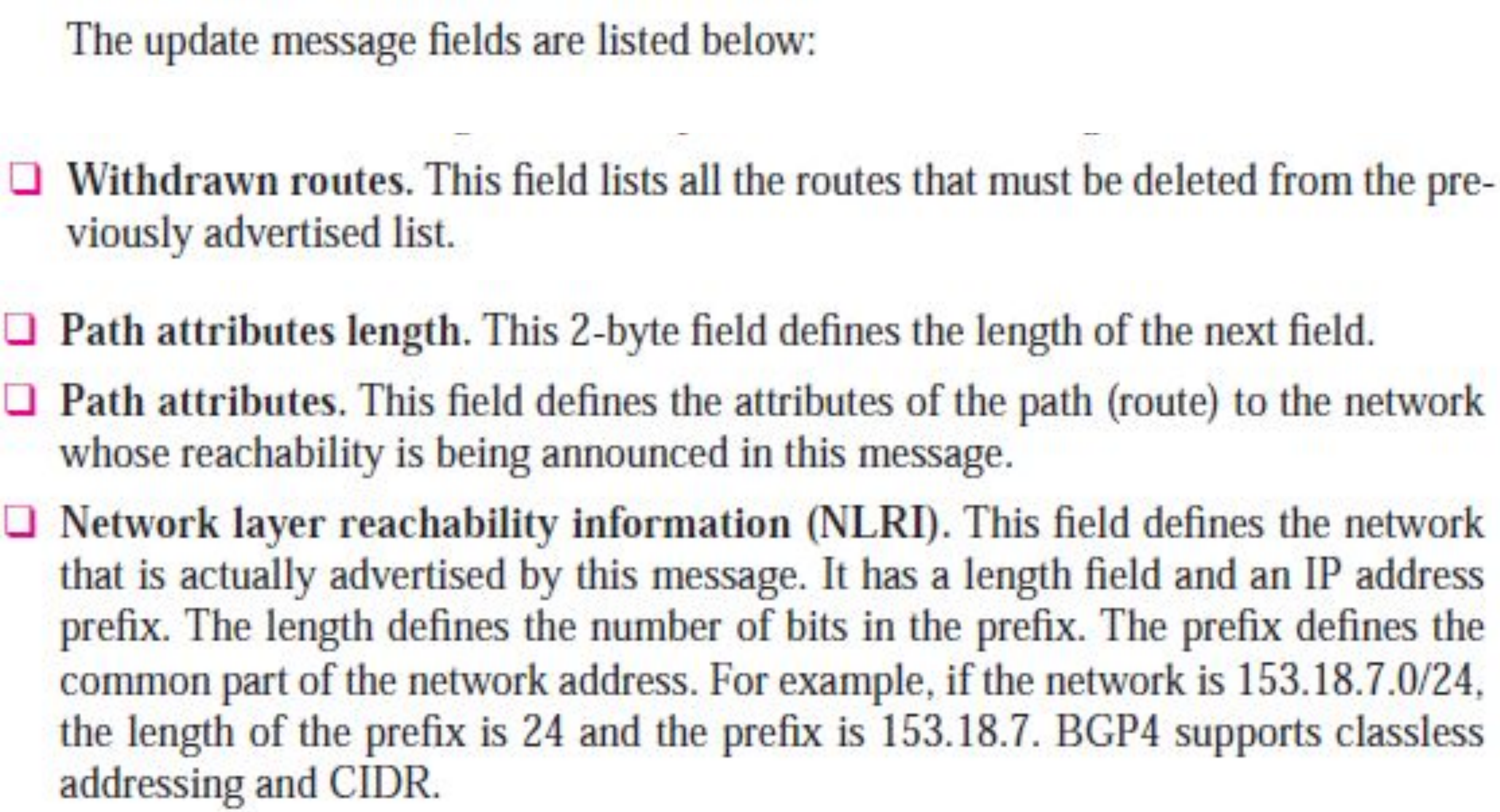
Unfeasible routes length

Withdrawn routes
(Variable length)

Path attributes length

Path attributes
(Variable length)

Network layer reachability information
(Variable length)



The update message fields are listed below:

- ❑ **Withdrawn routes.** This field lists all the routes that must be deleted from the previously advertised list.
- ❑ **Path attributes length.** This 2-byte field defines the length of the next field.
- ❑ **Path attributes.** This field defines the attributes of the path (route) to the network whose reachability is being announced in this message.
- ❑ **Network layer reachability information (NLRI).** This field defines the network that is actually advertised by this message. It has a length field and an IP address prefix. The length defines the number of bits in the prefix. The prefix defines the common part of the network address. For example, if the network is 153.18.7.0/24, the length of the prefix is 24 and the prefix is 153.18.7. BGP4 supports classless addressing and CIDR.

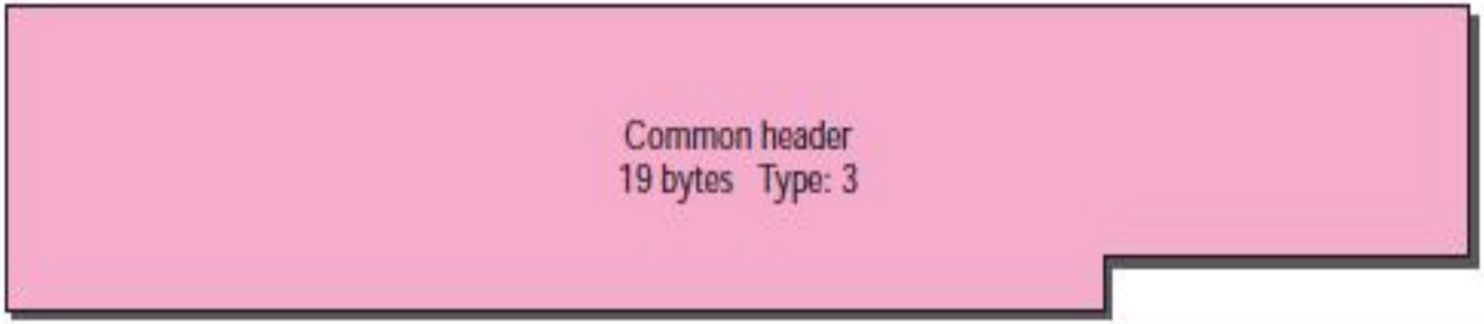
Note

BGP supports classless addressing and CIDR.

Keepalive Message

The routers running the BGP protocols exchange keepalive messages regularly (before their hold time expires) to tell each other that they are alive. The keepalive message consists of only the common header shown in Figure 11.58.

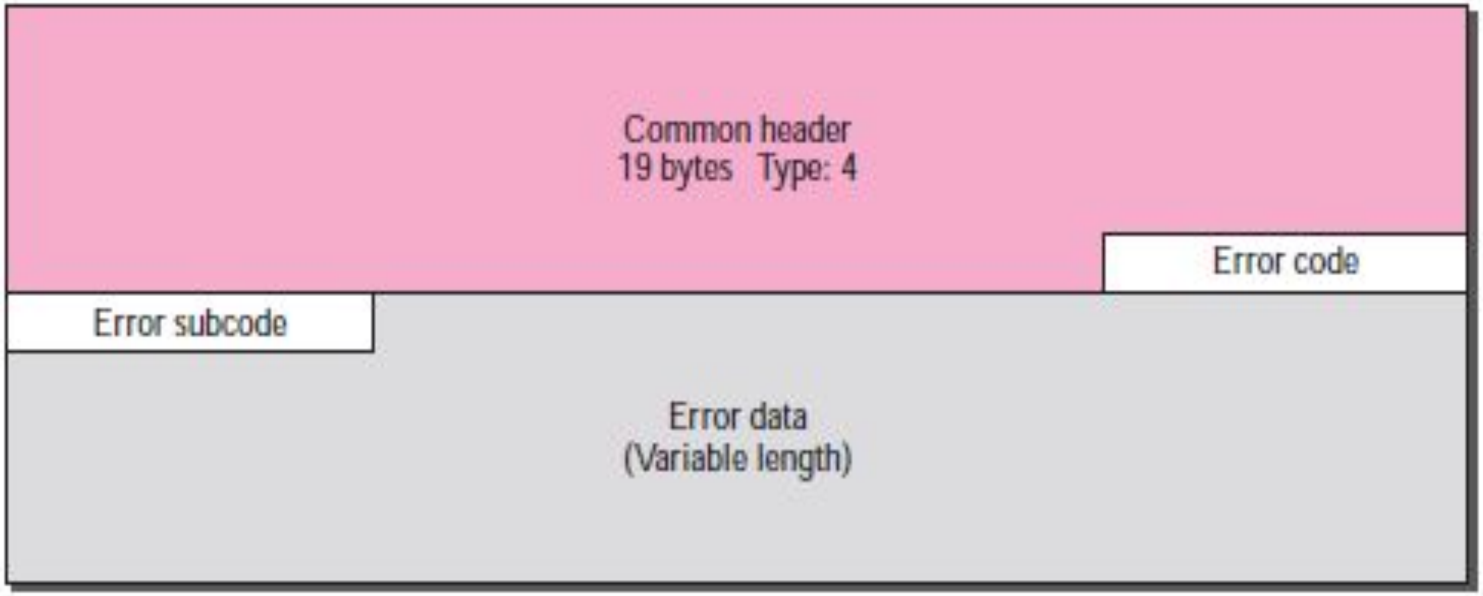
Figure 11.58 *Keepalive message*



Notification Message

A notification message is sent by a router whenever an error condition is detected or a router wants to close the connection. The format of the message is shown in Figure 11.59.The fields making up the notification message follow:

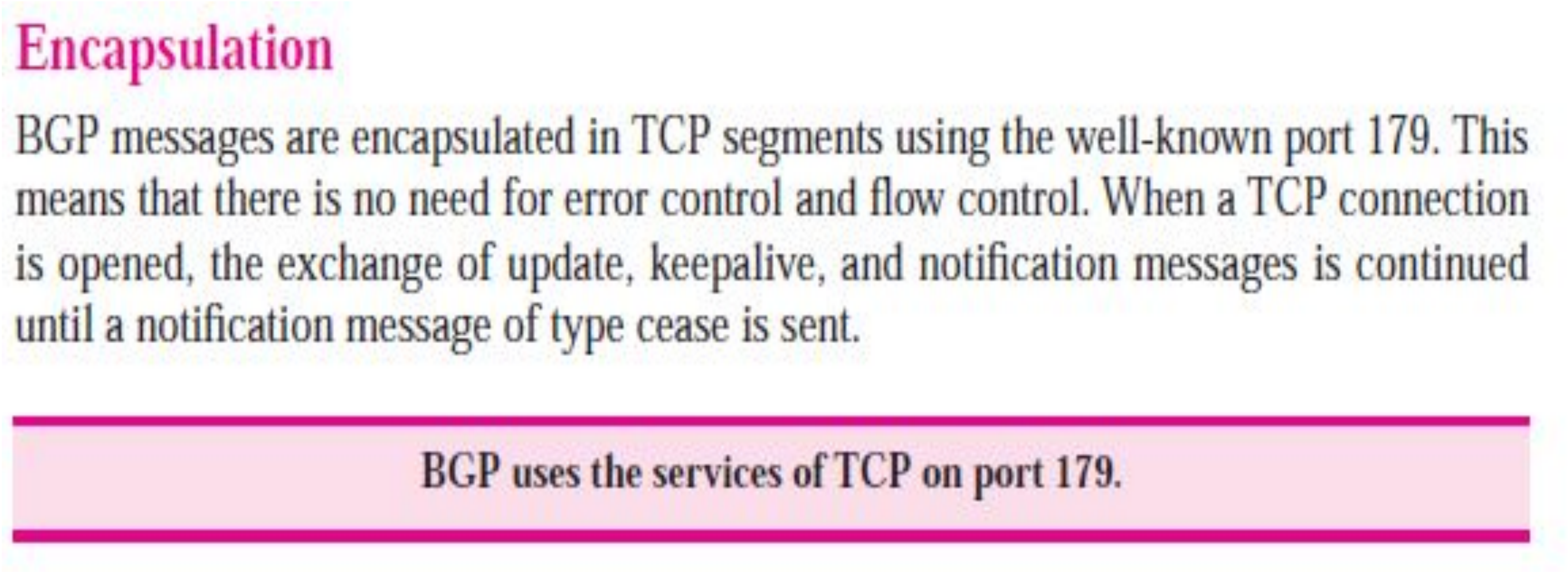
Figure 11.59 *Notification message*



- **Error code.** This 1-byte field defines the category of the error. See Table 11.6.
- **Error subcode.** This 1-byte field further defines the type of error in each category.
- **Error data.** This field can be used to give more diagnostic information about the error.

Table 11.6 *Error Codes*

Error Code	*Error Code Description*	*Error Subcode Description*
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.



Encapsulation

BGP messages are encapsulated in TCP segments using the well-known port 179. This means that there is no need for error control and flow control. When a TCP connection is opened, the exchange of update, keepalive, and notification messages is continued until a notification message of type cease is sent.

BGP uses the services of TCP on port 179.