Module
7

Routing and Congestion Control

Lesson 4

Border Gateway Protocol (BGP)

Specific Instructional Objectives

On completion of this lesson, the students will be able to:

- Explain the operation of the BGP protocol
- Explain the routing algorithm used in BGP
- Explain various attributes used in BGP
- State various message types used in BGP

7.4.1 Introduction

The **Border Gateway Protocol (BGP)** is an inter-autonomous system routing protocol. As discussed in section 7.1.5, an autonomous system (AS) is a network or group of networks under a common administration and with common routing policies. BGP is used to exchange routing information for the Internet and is the protocol used between Internet service providers (ISP), which are different ASes.

One of the most important characteristics of BGP is its *flexibility*. The protocol can connect together any internetwork of autonomous systems using an arbitrary topology. The only requirement is that each AS have at least one router that is able to run BGP and that this router connect to at least one other AS's BGP router. Beyond that, “the sky's the limit,” as they say. BGP can handle a set of ASs connected in a full mesh topology (each AS to each other AS), a partial mesh, a chain of ASes linked one to the next, or any other configuration. It also handles changes to topology that may occur over time.

The primary function of a BGP speaking system is to exchange network reachability information with other BGP systems. This network reachability information includes information on the list of Autonomous Systems (ASs) that reachability information traverses. BGP constructs a graph of autonomous systems based on the information exchanged between BGP routers. As far as BGP is concerned, whole Internet is a graph of ASs, with each AS identified by a Unique AS number. Connections between two ASs together form a path and the collection of path information forms a route to reach a specific destination. BGP uses the path information to ensure the loop-free inter-domain routing.

Another important assumption that BGP makes is that it doesn't know anything about what happens within the AS. This is of course an important prerequisite to the notion of an AS being *autonomous* - it has its own internal topology and uses its own choice of routing protocols to determine routes. BGP only takes the information conveyed to it from the AS and shares it with other ASs.

When a pair of autonomous systems agrees to exchange routing information, each must designate a router that will speak BGP on its behalf; the two routers are said to become *BGP peers* of one another. As a router speaking BGP must communicate with a peer in another autonomous system, usually a machine, which is near to the edge (Border) of the autonomous system is selected for this. Hence, BGP terminology calls the machine a *Border Gateway Router*.

In this lesson we shall discuss the important features of BGP. First we will look at the basics of BGP. Then in next section we shall have a look at BGP characteristics that make it stand apart from other routing protocols. Then in the fourth section we will discuss various attributes of BGP. In section 7.4.5 we shall briefly have a look at the BGP's path selection procedure. And in the last section we shall present various message formats used in BGP.

7.4.2 BGP Characteristics

BGP is different from other routing protocols in several ways. Most important being that BGP is neither a pure distance vector protocol nor a pure link state protocol. Let's have a look at some of the characteristics that stands BGP apart from other protocols.

- **Inter-Autonomous System Configuration**: BGP's primary role is to provide communication between two autonomous systems.
- **Next-Hop paradigm**: Like RIP, BGP supplies next hop information for each destination.
- **Coordination among multiple BGP speakers within the autonomous system**: If an Autonomous system has multiple routers each communicating with a peer in other autonomous system, BGP can be used to coordinate among these routers, in order to ensure that they all propagate consistent information.
- **Path information**: BGP advertisements also include path information, along with the reachable destination and next destination pair, which allows a receiver to learn a series of autonomous system along the path to the destination.
- **Policy support**: Unlike most of the distance-vector based routing, BGP can implement policies that can be configured by the administrator. For Example, a router running BGP can be configured to distinguish between the routes that are known from within the Autonomous system and that which are known from outside the autonomous system.
- **Runs over TCP**: BGP uses TCP for all communication. So the reliability issues are taken care by TCP.
- **Conserve network bandwidth**: BGP doesn't pass full information in each update message. Instead full information is just passed on once and thereafter successive messages only carries the incremental changes called **deltas**. By doing so a lot of network Bandwidth is saved. BGP also conserves bandwidth by allowing sender to aggregate route information and send single entry to represent multiple, related destinations.
- **Support for CIDR**: BGP supports classless addressing (CIDR). That it supports a way to send the network mask along with the addresses.
- **Security**: BGP allows a receiver to authenticate messages, so that the identity of the sender can be verified.

7.4.3 BGP Functionality and Route Information Management

The job of the Border Gateway Protocol is to facilitate the exchange of route information between BGP devices, so that each router can determine efficient routes to each of the networks on an IP internetwork. This means that descriptions of routes are the key data that BGP devices work with. But in a broader aspect, BGP peers perform three basic functions. The First function consists of initial peer acquisition and authentication. Both the peers establish a TCP connection and perform message exchange that guarantees both sides have agreed to communicate. The second function primarily focus on sending of negative or positive reachability information, this step is of major concern. The Third function provides ongoing verification that the peers and the network connection between them are functioning correctly. Every BGP speaker is responsible for managing route descriptions according to specific guidelines established in the BGP standards.

BGP Route Information Management Functions

Conceptually, the overall activity of route information management can be considered to encompass four main tasks:

- **Route Storage:** Each BGP stores information about how to reach networks in a set of special databases. It also uses databases to hold routing information received from other devices.

- **Route Update:** When a BGP device receives an *Update* from one of its peers, it must decide how to use this information. Special techniques are applied to determine when and how to use the information received from peers to properly update the device's knowledge of routes.

- **Route Selection:** Each BGP uses the information in its route databases to select good routes to each network on the internetwork.

- **Route Advertisement:** Each BGP speaker regularly tells its peers what it knows about various networks and methods to reach them. This is called *route advertisement* and is accomplished using BGP *Update* messages.

7.4.4 BGP Attributes

BGP Attributes are the properties associated with the routes that are learned from BGP and used to determine the best route to a destination, when multiple routes are available. An understanding of how BGP attributes influence route selection is required for the design of robust networks. This section describes the attributes that BGP uses in the route selection process:

- AS_path
- Next hop
- Weight

- Local preference
- Multi-exit discriminator
- Origin
- Community

AS_path Attribute: When a route advertisement passes through an autonomous system, the AS number is added to an ordered list of AS numbers that the route advertisement has traversed. Figure 7.4.1 shows the situation in which a route is passing through three autonomous systems.

AS1 originates the route to 172.16.1.0 and advertises this route to AS 2 and AS 3, with the AS_path attribute equal to {1}. AS 3 will advertise back to AS 1 with AS-path attribute {3, 1}, and AS 2 will advertise back to AS 1 with AS-path attribute {2, 1}. AS 1 will reject these routes when its own AS number is detected in the route advertisement. This is the mechanism that BGP uses to detect routing loops. AS 2 and AS 3 propagate the route to each other with their AS numbers added to the AS_path attribute. These routes will not be installed in the IP routing table because AS 2 and AS 3 are learning a route to 172.16.1.0 from AS 1 with a shorter AS_path list.

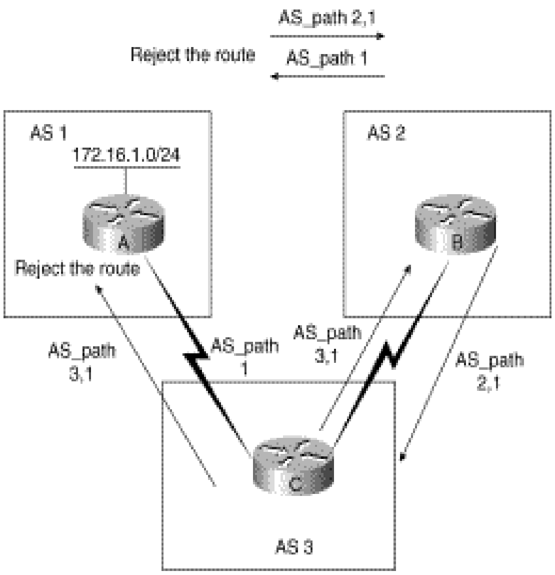


Figure 7.4.1 BGP AS-path Attribute

Next-Hop Attribute: The EBGP *next-hop* attribute is the IP address that is used to reach the advertising router. For EBGP peers, the next-hop address is the IP address of the connection between the peers. For IBGP, the EBGP next-hop address is carried into the local AS, as illustrated in Fig. 7.4.2.

Router C advertises network 172.16.1.0 with a next hop of 10.1.1.1. When Router A propagates this route within its own AS, the EBGP next-hop information is preserved. If Router B does not have routing information regarding the next hop, the route will be discarded. Therefore, it is important to have an IGP running in the AS to propagate next-hop routing information.

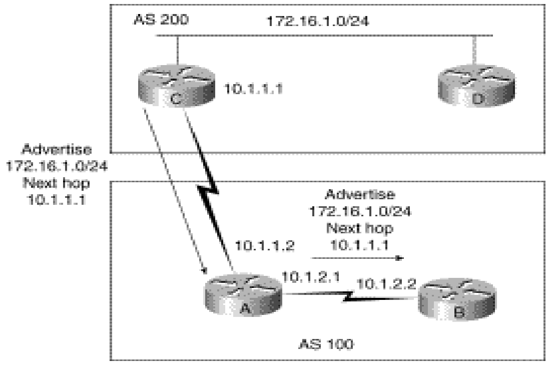


Figure 7.4.2 BGP Next-Hop Attribute

Weight Attribute: *Weight* is a Cisco-defined attribute that is local to a router. The weight attribute is not advertised to neighboring routers. If the router learns about more than one route to the same destination, the route with the highest weight will be preferred.

In Fig. 7.4.3, Router A is receiving an advertisement for network 201.12.23.0 from routers B and C. When Router A receives the advertisement from Router B, the associated weight is set to 50. When Router A receives the advertisement from Router C, the associated weight is set to 100. Both paths for network 201.12.23.0 will be in the BGP routing table, with their respective weights. The route with the highest weight will be installed in the IP routing table

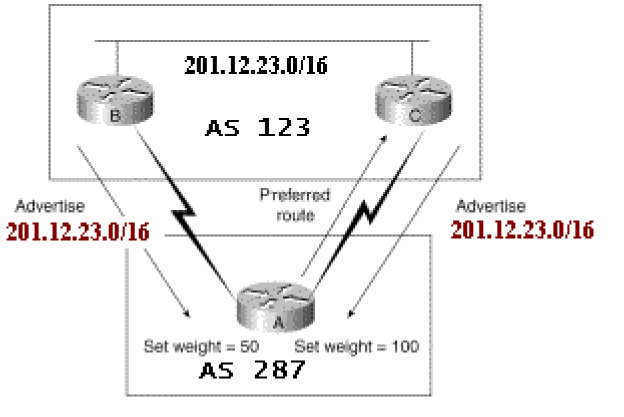


Figure 7.4.3 BGP Weight Attribute

Local Preference Attribute: The *local preference* attribute is used to prefer an exit point from the local autonomous system (AS). Unlike the weight attribute, the local preference attribute is propagated throughout the local AS. If there are multiple exit points from the AS, the local preference attribute is used to select the exit point for a specific route.

In Fig. 7.4.4, AS 287 is receiving two advertisements for network 201.12.23.0 from AS 123. When Router A receives the advertisement for network 201.12.23.0, the corresponding local preference is set to 150. When Router B receives the advertisement for same network, the corresponding local preference is set to 251. These local preference values will be exchanged between routers A and B. Because Router B has a higher local preference than Router A, Router B will be used as the exit point from AS 287 to reach network 201.12.23.0 in AS 123.

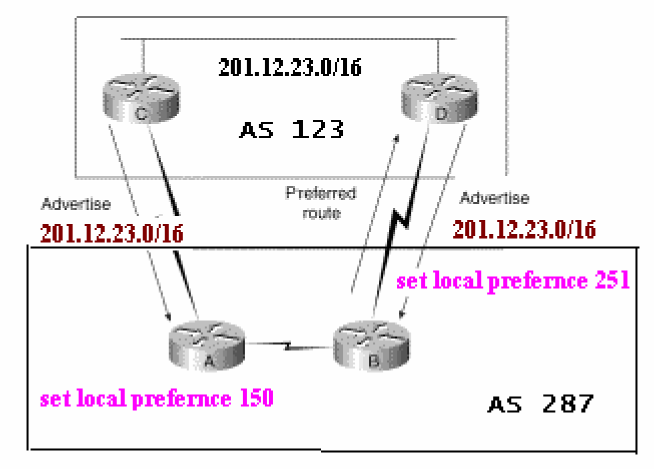


Figure 7.4.4 Local Preference Attribute

Multi-Exit Discriminator Attribute: The *multi-exit discriminator (MED)* or *metric attribute* is used as a suggestion to an external AS regarding the preferred route into the AS that is advertising the metric. The term *suggestion* is used because the external AS that is receiving the MEDs may be using other BGP attributes for route selection.

In Fig. 7.4.5, Router C is advertising the route 201.12.23.0 with a metric of 23, while Route D is advertising 201.12.23.0 with a metric of 5. The lower value of the metric is preferred, so AS 287 will select the route to router D for network 201.12.23.0 in AS 123. MEDs are advertised throughout the local AS.

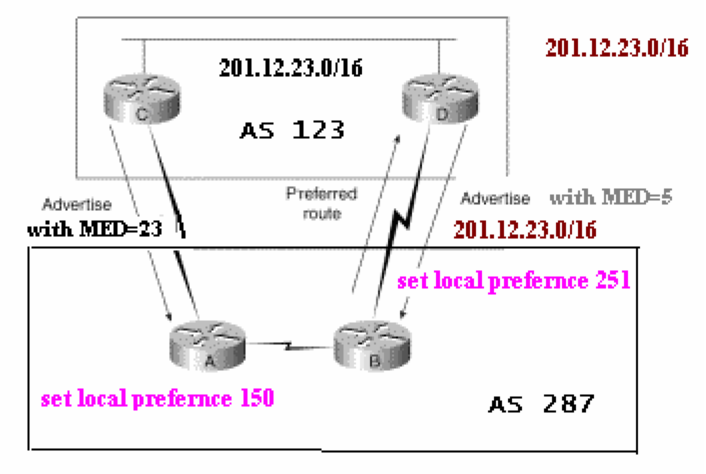


Figure 7.4.5   Multi-Exit Discriminator Attribute

Origin Attribute: The *origin attribute* indicates how BGP learned about a particular route. The origin attribute can have one of three possible values:

- **IGP—**The route is interior to the originating AS. This value is set when the network router configuration command is used to inject the route into BGP.
- **EGP—**The route is learned via the Exterior Border Gateway Protocol (EBGP).
- **Incomplete—**The origin of the route is unknown or learned in some other way. An origin of incomplete occurs when a route is redistributed into BGP.

The origin attribute is used for route selection.

Community Attribute: The community attribute provides a way of grouping destinations, called communities, to which routing decisions (such as acceptance, preference, and redistribution) can be applied. Route maps are used to set the community attribute. Predefined community attributes are listed here:

- **No-export**—Do not advertise this route to EBGP peers.
- **No-advertise—**Do not advertise this route to any peer.
- **Internet**—Advertise this route to the Internet community; all routers in the network belong to it.

Figure 7.4.8 demonstrates the third community attribute namely, Internet community attribute. There are no limitations to the scope of the route advertisement from AS 1.

Figure 7.4.6 illustrates the no-export community. AS 1 advertises 172.16.1.0 to AS 2 with the community attribute no-export. AS 2 will propagate the route throughout AS 2 but will not send this route to AS 3 or any other external AS.

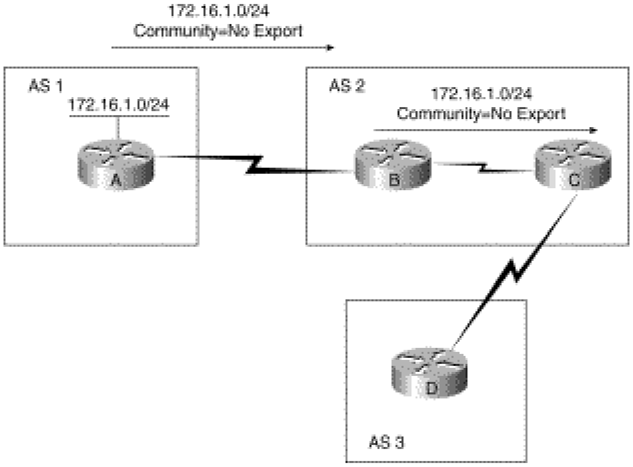


Figure 7.4.6 BGP no-export Community Attribute

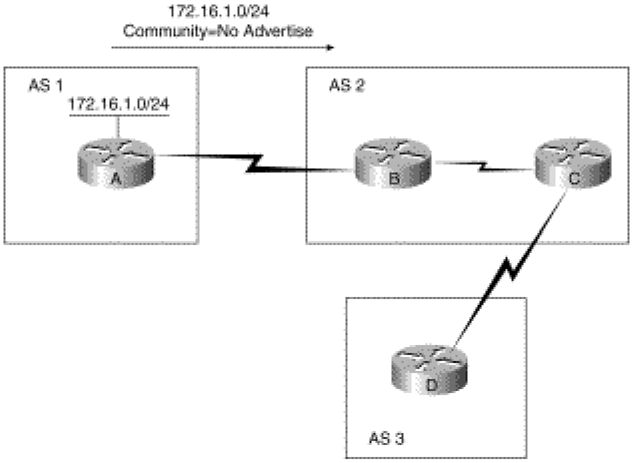


Figure 7.4.7 BGP no-advertise Community Attribute

In Fig. 7.4.7, AS 1 advertises 172.16.1.0 to AS 2 with the community attribute no-advertise. Router B in AS 2 will not advertise this route to any other router, i.e. the advertisement for this route is not even made within the Autonomous system, it would be restricted just to the Router B.

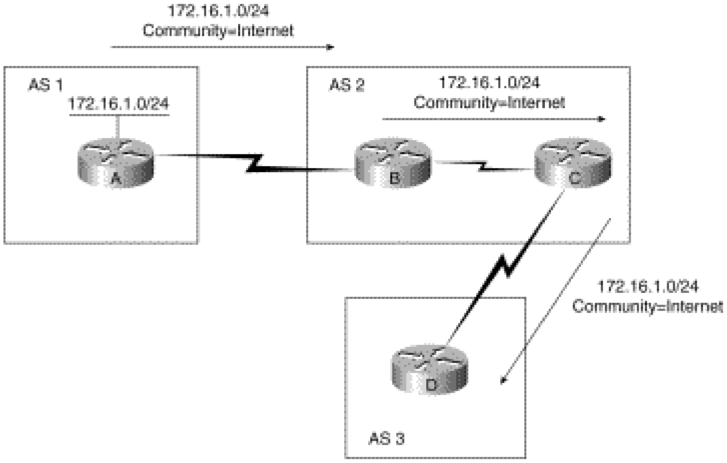


Figure 7.4.8 BGP Internet Community Attribute

7.4.5 BGP Path Selection

BGP could possibly receive multiple advertisements for the same route from multiple sources. BGP selects only one path as the best path. When the path is selected, BGP puts the selected path in the IP routing table and propagates the path to its neighbors. BGP uses the following criteria, in the order presented, to select a path for a destination:

- If the path specifies a next hop that is inaccessible, drop the update.
- Prefer the path with the largest weight.
- If the weights are the same, prefer the path with the largest local preference.
- If the local preferences are the same, prefer the path that was originated by BGP running on this router.
- If no route was originated, prefer the route that has the shortest AS_path.
- If all paths have the same AS_path length, prefer the path with the lowest origin type (where IGP is lower than EGP, and EGP is lower than incomplete).
- If the origin codes are the same, prefer the path with the lowest MED attribute.
- If the paths have the same MED, prefer the external path to the internal path.
- If the paths are still the same, prefer the path through the closest IGP neighbor.
- Prefer the path with the lowest IP address, as specified by the BGP router ID.

7.4.6 BGP Message type

For all the functions described above, BGP defines four basic message types namely, *OPEN*, *UPDATE*, *NOTIFICATION*, *KEEPALIVE*. In this section we shall discuss these message formats.

7.4.6.1 BGP Fixed Header Format

Each BGP message begins with a fixed header that identifies the message type. Figure 7.4.9 illustrates this fixed header format.

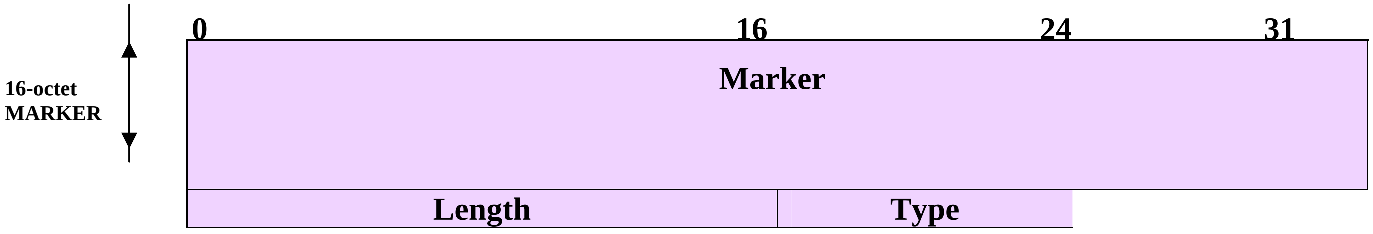


Figure 7.4.9 BGP Fixed header Format

Fields in the fixed header

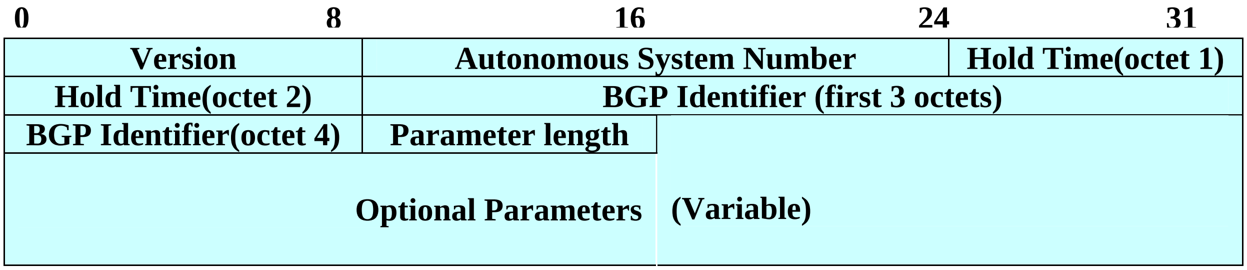
- **MARKER**: The 16-octect MARKER field contains a value that both sides agree to use to mark the beginning of the message. This is basically used for synchronization. In the initial message it contains all 1's and if the peers agree to use authentication mechanism, the marker can contain the authentication information.

- **LENGTH**: The 2-octect LENGTH field Specifies the Total message length measured in octets. The minimum message size is 19 octets (i.e. only fixed header), and the maximum allowable length is 4096 octets.
- **TYPE**: 1-octet field contains one of the 4 values of the message type listed below:

Type Code	Message Type	Description
1	**OPEN**	**Initialize communication**
2	**UPDATE**	**Advertise or withdraw routes**
3	**NOTIFICATION**	**Response to an Incorrect message**
4	**KEEPALIVE**	**Actively test peer connectivity**

5.4.6.2 BGP OPEN Message

This is the first message send after the BGP peers establishes a TCP connection. Both of the peers send OPEN message to declare their autonomous system number and other operating parameters. Figure 7.4.10 illustrates the OPEN message format.



0	8	16	24 31
Version	**Autonomous System Number**		**Hold Time(octet 1)**
Hold Time(octet 2)	**BGP Identifier (first 3 octets)**		
BGP Identifier(octet 4)	**Parameter length**		
Optional Parameters (Variable)			

Figure 7.4.10 BGP OPEN Message Format

Fields in the message header has been explained below:

- **Version**: It identifies the protocol version used.
- **Autonomous System Number**: Gives the autonomous system of the sender's system.
- **Hold Time**: it specifies maximum time receiver should wait for a message from sender. The receiver implements a timer using this value. The value is reset each time a message arrives; if timer expires it assumes that sender is not available.
- **BGP identifier**: It is a 32-bit value that uniquely identifies the sender. It is the IP address and the router must choose one of its IP addresses to use with all the BGP peers.
- **Parameter Length**: If Optional parameters are specified then this fields contains the length of optional parameters, in octets.
- **Optional Parameters**: It contains a list of parameters. Authentication is also a kind of parameter in BGP. It's done in this way so that the BGP peers can choose the authentication method without making it a part of BGP fixed header.

When it accepts an incoming OPEN message, a machine speaking BGP responds by sending a KEEPALIVE message. Each side must send an OPEN message and receive a

KEEPALIVE message before they can actually exchange routing information. Thus, KEEPALIVE messages are a kind of acknowledgement for OPEN message.

7.4.6.3 BGP UPDATE Message

After the BGP peers have established a TCP connection, send the OPEN message, and acknowledge them, peers use UPDATE message for advertisements. Peers use UPDATE message to advertise new destination that are reachable or to withdraw previously advertised destination, which have become unreachable. Figure 7.4.11 illustrates the UPDATE message format.

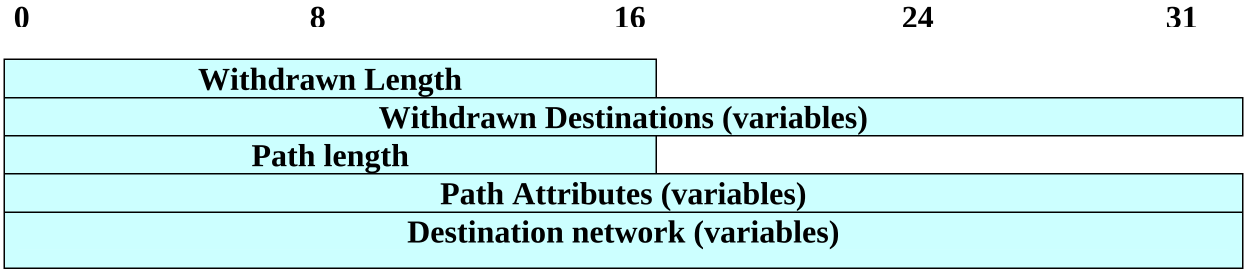

0
8
16
24
31
Withdrawn Length
Withdrawn Destinations (variables)
Path length
Path Attributes (variables)
Destination network (variables)

Figure 5.31 BGP UPDATE Message Format

Update message is divided into two parts:

- First part lists the previously advertised routes that are being withdrawn
- Second specifies new destination being advertised.

Description about the fields in the header is given below:

- **Withdrawn Length**: it is a 2-octet field that specifies the size of the withdrawn destination field that follows.
- **Path Length**: specifies the size of the Path Attributes. These path attributes are associated with the new advertisements.
- Both **Withdrawn Destination** and **Destination network**, in the message format, contains a list of IP addresses. BGP supports classless addressing in a different way, instead of sending subnet masks separately with each IP address; it uses a compress representation to reduce the message size. BGP doesn't send a bit mask; instead it encodes information about the mask into a single octet that precedes each address. The Mask octet contains a binary integer that specifies number of bits in the mask (mask bits are assumed to be contiguous). Address that follows the mask is also compressed, only octets covered by the mask are included. For example, only two address octets follow a mask value of 9 to 16 and so on. It is illustrated in Fig. 7.4.12.

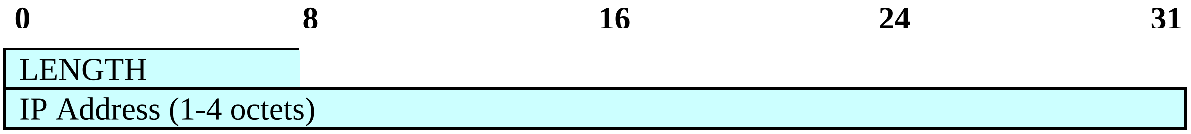


Figure 7.4.12 Compressed form that BGP uses to store destination IP and Mask

7.4.6.4 BGP NOTIFICATION Message

BGP supports NOTIFICATION message type for control purposes and when error occurs. Once BGP detects a problem it sends a notification message and then closes TCP connection. Figure 7.4.13 illustrates the NOTIFICATION message format. Following tables list the possible values of Error code (Fig. 7.4.14) and Error subcodes (Fig. 7.4.15):

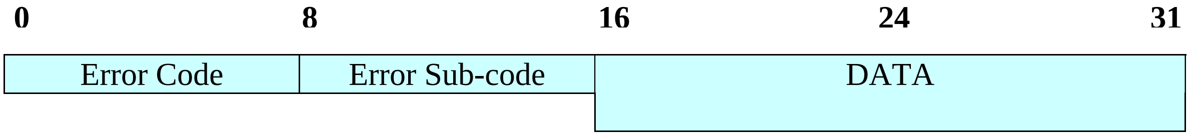


Figure 7.4.13 BGP Notification Message Format

Error Code	Meaning
1	Error in Message header
2	Error in OPEN Message
3	Error in UPDATE Message
4	Hold Timer Expired
5	Finite State machine Error
6	Cease (terminate connection)

Figure 7.4.14 Possible values of Error Code

SubCode For Message Header Errors (Error Code =1)	
1	Connection not synchronized
2	Incorrect message length
3	Incorrect message Type
SubCode For OPEN Message Errors (Error Code = 2)	
1	Version number Unsupported
2	Peer AS Invalid
3	BGP Identifier Invalid
4	Unsupported optional parameter
5	Authentication failure
6	Hold Time Unacceptable

SubCode For UPDATE Message Errors (Error Code =3)	
1	Attribute List Malformed
2	Unrecognized Attribute
3	Missing Attribute
4	Attribute Flags Error
5	Attribute Length Error
6	Invalid Origin Attribute
7	AS Routing loop
8	Next Hop Invalid
9	Error in Optional Attribute
10	Invalid network Field
11	Malformed AS Path

Figure 7.4.15 Possible values of Error Sub-code

7.4.6.5 BGP KEEPALIVE Message

Two BGP peers periodically exchange KEEPALIVE messages to test the network connectivity between them and to verify that both are functioning well. A KEEPALIVE message consists of standard message header with no extra data (19 octets), as discussed in section 7.4.6.1.

Fill in the Blanks

1. BGP is abbreviated as __________.
2. BGP is a ___________-autonomous system routing protocol.
3. The protocol can connect together any internetwork of autonomous systems using an ___________ topology
4. The overall activity of route information management can be considered to encompass four main tasks: ____________, Route Update, ________, Route Selection.
5. The _________ indicates how BGP learned about a particular route.
6. Origin attribute can have one of three possible values namely, ________,________ and ________
7. _________ Field contains a value that both sides agree to use to mark the beginning of the message.
8. The minimum message size is _____ octets, and the maximum allowable length is ______ octets
9. Type Code equals 1 means _________ Message and _________ communication.
10. BGP identifier is a _____-bit value.
11. ______________ is a 2-octet field that specifies the size of the withdrawn destination field that follows.
12. Error Code Value equal to 3 signifies error in _________ message.

Answers

1. Border Gateway protocol
2. Inter
3. arbitrary
4. Route storage, route advertisement
5. origin attribute
6. IGP, EGP, Incomplete
7. MARKER
8. 19, 4096
9. OPEN, initialize
10. 32
11. Withdrawn Length
12. UPDATE

Short Answer Questions

1. **Explain few characteristics of BGP.**

Ans: Most important characteristic of BGP is that it is nor a pure distance-vector protocol nor a pure link-state protocol. Some other characteristics that stand BGP apart from other protocols are:

- **Inter-Autonomous System Configuration**: BGP's primary role is to provide communication between two autonomous systems.
- **Next-Hop paradigm**: Like RIP, BGP supplies next hop information for each destination.
- **Coordination among multiple BGP speakers within the autonomous system**: If an Autonomous system has multiple routers each communicating with a peer in other autonomous system, BGP can be used to coordinate among these routers, in order to ensure that they all propagate consistent information.
- **Path information**: BGP advertisements also include path information, along with the reachable destination and next destination pair, which allows a receiver to learn a series of autonomous system along the path to the destination.
- **Runs over TCP**: BGP uses TCP for all communication. So the reliability issues are taken care by TCP.
- **Conserve network bandwidth**: BGP doesn't pass full information in each update message. Instead full information is just passed on once and thereafter successive messages only carries the incremental changes called **deltas**. By doing so a lot of network Bandwidth is saved. BGP also conserves bandwidth
- **Support for CIDR**: BGP supports classless addressing (CIDR). That it supports a way to send the network mask along with the addresses.
- **Security**: BGP allows a receiver to authenticate messages, so that the identity of the sender can be verified.

2. **Explain the three basic functions performed by BGP peers.**

Ans: BGP peers perform three basic functions.

- The First function consists of initial peer acquisition and authentication. Both the peers establish a TCP connection and perform message exchange that guarantees both sides have agreed to communicate.
- The second function primarily focus on sending of negative or positive reachability information, this step is of major concern.
- The Third function provides ongoing verification that the peers and the network connection between them are functioning correctly.

3. **What are the basic activities of route information management?**

Ans: The overall activity of route information management can be considered to encompass four main tasks:

- **Route Storage:** Each BGP stores information about how to reach networks in a set of special databases. It also uses databases to hold routing information received from other devices.

- **Route Update:** When a BGP device receives an *Update* from one of its peers, it must decide how to use this information. Special techniques are applied to determine when and how to use the information received from peers to properly update the device's knowledge of routes.
- **Route Selection:** Each BGP uses the information in its route databases to select good routes to each network on the internetwork.
- **Route Advertisement:** Each BGP speaker regularly tells its peers what it knows about various networks and methods to reach them. This is called *route advertisement* and is accomplished using BGP *Update* messages.

4. **Explain the Original Attribute.**

Ans: The *origin attribute* indicates how BGP learned about a particular route. The origin attribute can have one of three possible values:

- **IGP—**The route is interior to the originating AS. This value is set when the network router configuration command is used to inject the route into BGP.
- **EGP—**The route is learned via the Exterior Border Gateway Protocol (EBGP).
- **Incomplete—**The origin of the route is unknown or learned in some other way. An origin of incomplete occurs when a route is redistributed into BGP.

5. Explain BGP path Selection procedure.

Ans: BGP uses the following criteria, in the order presented, to select a path for a destination:

- If the path specifies a next hop that is inaccessible, drop the update.
- Prefer the path with the largest weight.
- If the weights are the same, prefer the path with the largest local preference.
- If the local preferences are the same, prefer the path that was originated by BGP running on this router.
- If no route was originated, prefer the route that has the shortest AS_path.
- If all paths have the same AS_path length, prefer the path with the lowest origin type (where IGP is lower than EGP, and EGP is lower than incomplete).
- If the origin codes are the same, prefer the path with the lowest MED attribute.
- If the paths have the same MED, prefer the external path to the internal path.
- If the paths are still the same, prefer the path through the closest IGP neighbor.
- Prefer the path with the lowest IP address, as specified by the BGP router ID.

6. Explain Type field in Fixed Header of BGP.

Ans: TYPE: is a 1-octet field contains one of the 4 values of the message type listed below:

Type Code	Message Type	Description
1	OPEN	Initialize communication
2	UPDATE	Advertise or withdraw routes
3	NOTIFICATION	Response to an Incorrect message
4	KEEPALIVE	Actively test peer connectivity