CHAPTER 8

Border Gateway Protocol—The Glue That Holds the Internet Together

In 1991, the first website was developed for the Stanford Linear Accelerator Center. By 1993 there were approximately 600 websites. The number of websites grew to 10,000 by 1994, 100,000 by 1995, and more than 1 million by 1997. The number of websites continues to grow at a phenomenal rate, and today there are millions of websites accessible through the Internet. Millions of websites means there are millions of IP addresses— and a scalable IP routing protocol is needed to tie them all together to form the World Wide Web.

In Chapter 1, "Routing and Switching in Everyday Life," one of the systems you examined was the highway system. If you consider a house to be a website, and a city to be a company containing one or more websites, the highway system is what ties them all together. You used an interstate highway map to get to a particular city, and then used a city map to get to a specific location. The city map is analogous to using an interior routing protocol (RIP, IGRP, EIGRP, OSPF, or IS-IS) to determine a route within a specific organization. The interstate highway map is analogous to an exterior gateway protocol (EGP) used for Internet routing. The Border Gateway Protocol (BGP) is the protocol used today for Internet routing, and this chapter discusses the concepts and operation of BGP and how BGP is used spin the web that is the Internet.

Understanding the Need for BGP

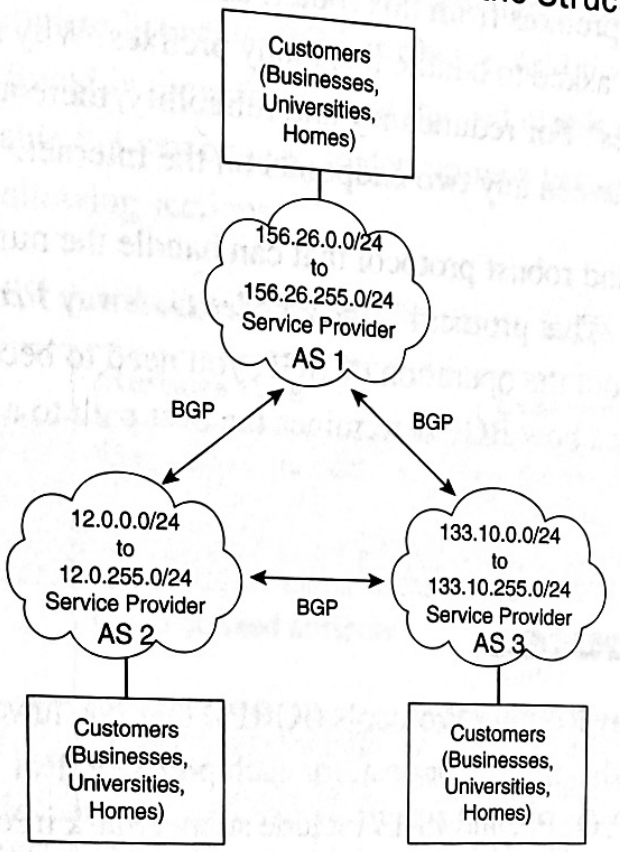
When you get home from work, school, or play, many of you connect to the Internet. You may check your e-mail, see how your investments did that day, or just surf the Internet. Have you ever wondered how your e-mail gets to your computer or how your favorite website appears on your screen? If you have read the first seven chapters of this book, your answer should be "IP routing takes care of getting information to my computer." The interior gateway protocols (RIP, IGRP, EIGRP, OSPF, and IS-IS) cannot scale to the number of IP prefixes existing on the Internet. The interior protocols were designed for use within an organization, and their best path algorithms cannot handle the number of routes required for Internet routing. BGP uses a simple algorithm for determining the best path to a destination; using new properties of routes in addition to the network prefix and subnet mask. These new properties, or route attributes, along with BGP route summarization, allow BGP to scale to the number of routes required for Internet routing.

Figure 8-1 shows a simplified representation of the structure of the Internet. Service providers, such as your cable company, own a block or blocks of IP addresses. These IP addresses are assigned to enterprise customers such as businesses, universities, or individual users. There are many service providers around the world and each is assigned a unique autonomous system (AS) number, the use of which you will soon learn about in the later section entitled, "Autonomous System—AS-PATH." For example, the service provider assigned AS 1 owns the 156.26.0.0 Class B address space (156.26.0.0/24 to 156.26.255.0/24). The service provider is free to subnet this IP address block and assign subnets to customers. For the customers of AS 1 to reach any IP destination on the Internet, the entire world needs to know how to route an IP packet to the customers of this service provider; and the customers of AS 1 need to know how to route an IP packet from their networks to anywhere in the world. For Internet routing, you can eliminate interior routing protocols RIP version 1 (RIPv1) and IGRP because these protocols have too many limitations to use in an enterprise network, and there is no way you would use them for Internet routing. RIPv2 improves on RIPv1, but it has the limitations of a maximum hop count of 15 and a slow convergence time. A maximum hop count of 15 eliminates RIP from consideration for Internet routing. EIGRP, OSPF, and IS-IS are the recommended routing protocols to use in an

enterprise network. They all have faster convergence time, and are able to determine a loop-free routing topology. The amount of time and computer resources a router needs to exchange routes and determine a loop-free topology depends on the number of network

prefixes that have to be dealt with. A protocol for Internet routing must be able to handle a large number of routes as shown in Example 8-1.

Figure 8-1 Simplified Representation of the Structure of the Internet



The output shown in Example 8-1 is from one of AT&T's Internet routers, and shows the number of Internet routes on the day I wrote this chapter (the number of routes increases everyday).

Example 8-1 Routing Table on an Internet Router

```
*****************************************************

     AT&T Enhanced Network Services
                (CERFnet)
               route-server
```

continues

Example 8-1 Routing Table on an Internet Router (continued)

```
*****************************************************

route-server>show ip bgp summary | include network entries
127990 network entries and 383407 paths using 29796898 bytes of memory
```

As highlighted in the output, there are 127,990 network prefixes and 383,407 paths to reach these prefixes from this router. EIGRP, OSPF, and IS-IS would self-destruct if they were asked to handle this many prefixes. Why are there more paths than network prefixes? For redundancy and reliability, there are multiple paths that can be taken between any two endpoints on the Internet.

You need a simple and robust protocol that can handle the number of routes that exist on the Internet. That protocol is the ***Border Gateway Protocol (BGP)***. Before you learn about the operation of BGP, you need to become familiar with BGP terminology and how BGP determines the best path to a network prefix.

BGP Attributes

The Interior Gateway Routing Protocols (IGRPs) that you have learned advertised a network prefix with a metric, or cost, for each prefix (RIPv1 and IGRP). In addition, RIPv2, EIGRP, OSPF, and IS-IS include subnet mask information along with the advertised network prefix. If there are multiple paths from a router to the same destination, the best path is the one with the total lowest cost or metric. Each of the interior protocols uses a different metric, and each protocol considers the route with the lowest total cost to be the best path.

BGP also advertises prefix and mask information to a neighboring BGP router; but these advertisements include other parameters, or attributes, BGP can use to determine a best path. There are four types of BGP attributes:

- **Well-known mandatory**—An attribute that must be advertised to and understood by all BGP neighbors.

- **Well-known discretionary**–An optional attribute that is not required to be advertised to a BGP neighbor.
- **Optional transitive**–If a BGP neighbor receives an attribute it does not understand, the attribute is passed on to other BGP neighbors.
- **Optional nontransitive**–If a BGP neighbor receives an attribute it does not understand, the attribute is not passed on to other BGP neighbors.

Table 8-1 lists the basic BGP attributes discussed in this chapter and their type. Table 8-1 is not a complete list of the BGP attributes. Additional BGP attributes and their use can be found in the references at the end of this chapter. The attributes listed in Table 8-1 can be used to determine a best path to a destination as discussed in the following sections.

Table 8-1 BGP Attributes and Their Attribute Type

Attribute Name	Attribute Type	Description
AS_PATH	Well-known mandatory	An ordered list of each AS that a BGP route advertisement has passed through
WEIGHT	Cisco defined attribute	Used to select a path from a router
LOCAL_PREF	Well-known discretionary	Used to select an exit path from the local AS
MULTI_EXIT_DISC	Optional nontransitive	Used to select a path into the local AS
ORIGIN	Well-known mandatory	Indicates how BGP learned the route IGP—Network statement EGP—Learned from an EBGP neighbor Incomplete—Learned through redistribution

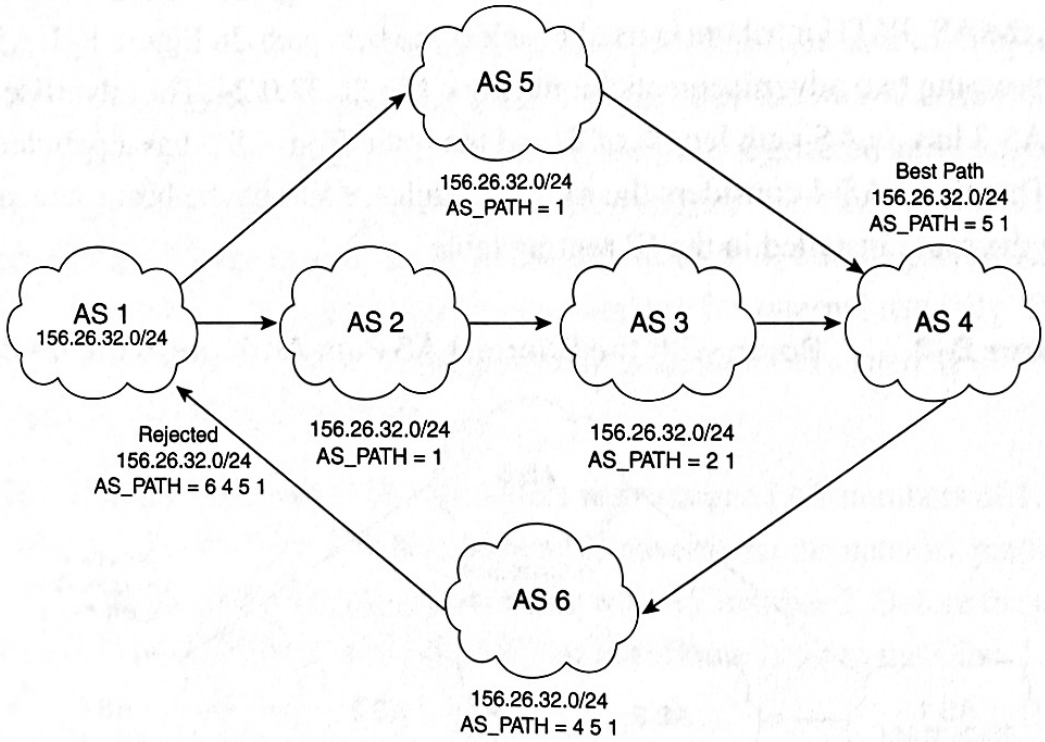
continues

An AS path is an ordered list of the number of autonomous systems the advertisement has passed through. This is similar to the hop count used by RIP because an AS number in the path can be considered to be a hop toward the destination.

In addition to using the AS path to determine the best path to a particular IP prefix if there is more than one path, BGP also uses the AS_PATH attribute for loop detection.

In Figure 8-4, AS 4 is advertising the best path for 156.26.32.0/24 to AS 6. AS 6 advertises this prefix to AS 1 with an AS_PATH attribute of 6 4 5 1. When AS 1 receives the advertisement from AS 6, this route will be rejected because AS 1 is contained in the AS_PATH attribute. BGP routers reject any advertisement containing their own AS number.

Figure 8-4 BGP Loop Detection



As mentioned earlier in this section, all things being equal, the AS_PATH attribute is used to select the best path. "All things" refer to other BGP attributes that can also influence the selection of the best path.

When the advertisement is received by AS 2, AS 2 knows the prefix originated in AS 1 because the last and only AS number in the AS_PATH attribute is equal to 1. AS 2 also knows that the prefix originated in AS 1 because there is only one AS number listed in the AS_PATH attribute. When AS 2 advertises the 156.26.32.0/24 prefix to AS 3, AS 2 will prepend, or put at the beginning, its own AS number in the AS_PATH attribute. When AS 3 receives the advertisement, the AS_PATH attribute now has a value of 2 1 with a space between the AS numbers. AS 3 knows the prefix originated in AS 1 (the last number in the attribute is 1), and the prefix has passed through AS 2. Finally, AS 3 prepends its AS number to the AS_PATH attribute, and advertises the prefix to AS 4. AS 4 receives an AS_PATH attribute of 3 2 1 for the 156.26.32.0/24 prefix.

If all the other BGP attributes are equal, the router with the shortest AS path is considered to be the best path. In Figure 8-2, there is only one advertisement for network 156.26.32.0/24. By default, it has to be the best path. If there are multiple paths, the AS_PATH attribute is used to select the best path. In Figure 8-3, AS 4 is now receiving two advertisements for network 156.26.32.0/24. The advertisement from AS 3 has an AS path length of 3, and the route from AS 5 has a path length of 2. Therefore, AS 4 considers the route through AS 5 to be the best route, and this is the route installed in the IP routing table.

Figure 8-3 Route with the Shortest AS Path Attribute Is the Best Path

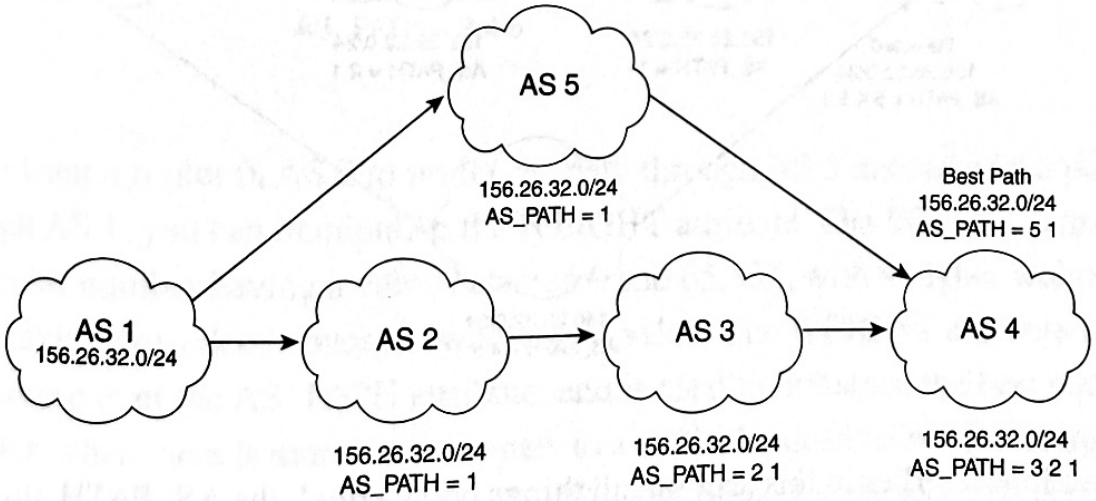


Table 8-1 BGP Attributes and Their Attribute Type (continued)

Attribute Name	Attribute Type	Description
NEXT_HOP	Well-known mandatory	EBGP—Set to the address of the interface used to communicate with an external neighbor IBGP—EBGP next hop is sent unchanged to IBGP neighbors
Community	Optional transitive	A number used to group routes so a policy can be applied to the group

Autonomous System Path Attribute—AS_PATH

The AS_PATH attribute is used by BGP for best path selection and routing loop detection. Every service provider is assigned a unique ***autonomous system (AS)*** number in the range 1 to 64,511. These AS numbers are registered and assigned in the same way that unique IP address blocks are assigned. Service providers can also use private AS numbers in the range 64,512 to 65,534, but the private AS numbers cannot be advertised on the Internet and are for internal use only. The private AS numbers are similar to the private IP address blocks such as the 10.0.0.0/8 private IP address space.

In Figure 8-2, there are four service providers with assigned AS numbers of 1, 2, 3, and 4. The service provider with AS number 1 is advertising the network prefix and mask of 156.26.32.0/24 to the service provider with AS number 2. Before the prefix and mask from AS 1 is advertised, the AS_PATH attribute is set to the value 1.

Figure 8-2 AS_PATH Attribute

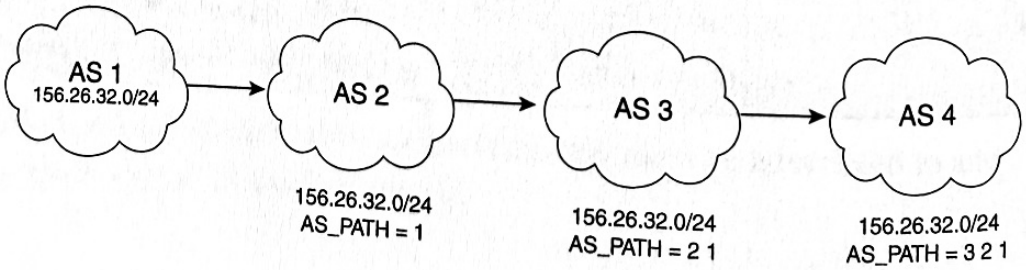
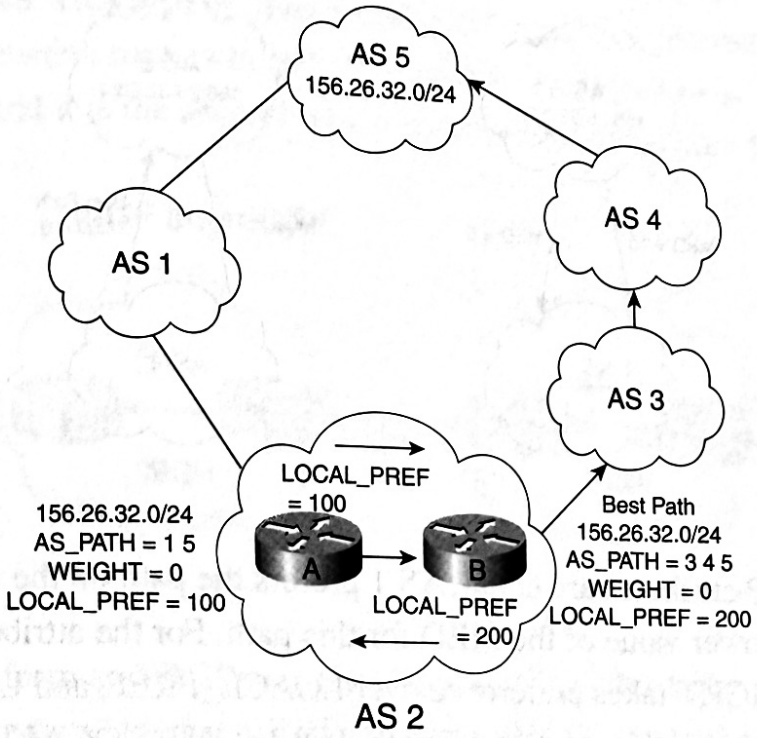


Figure 8-7 LOCAL_PREF Attribute

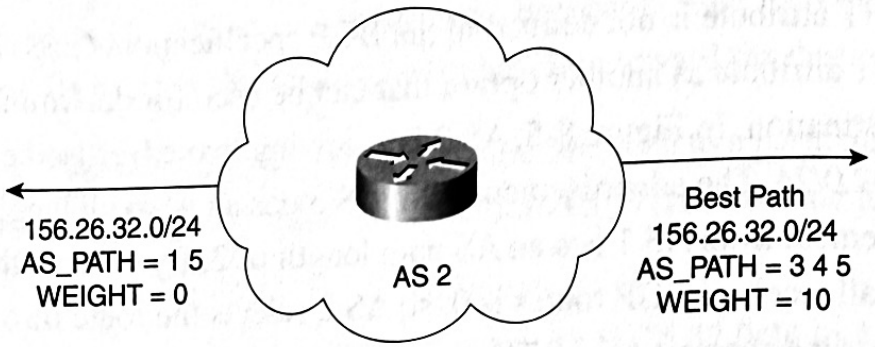


Metric or MULTI_EXIT_DISC (MED) Attribute

The WEIGHT attribute is used to prefer a path from a particular router. The LOCAL_PREF attribute is used to prefer a path leaving a particular AS. The metric, or MULTI_EXIT_DISC (multi-exit discriminator or MED), is used to prefer a path into an AS. MED is a 32-bit number having a value between 0 and 4,294,967,295 with a lower value being preferred over a higher value.

In Figure 8-8, AS 5 owns the network prefix 156.26.32.0/24 and there are two paths to reach this prefix from AS 1. AS 5 wants AS 1 to use the path on the right for traffic from AS 1 to AS 5. AS 5 sets the MED on the left to any value higher than the default value of 0 and advertises this value to AS 1.

Figure 8-6 WEIGHT Attribute



The WEIGHT attribute has only local significance. This means a router does not advertise the WEIGHT attribute to neighboring BGP routers. When a router receives a BGP prefix, the WEIGHT attribute is set to 0. Any BGP route originated by a router has a default weight attribute of 32,768.

Local Preference Attribute—LOCAL_PREF

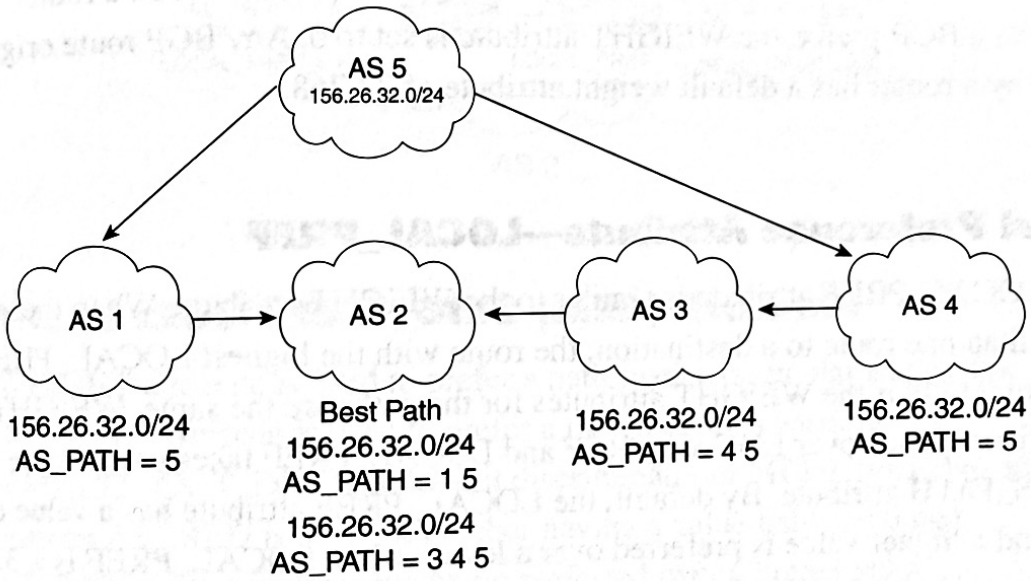
The LOCAL_PREF attribute is similar to the WEIGHT attribute. When there is more than one route to a destination, the route with the highest LOCAL_PREF is preferred only if the WEIGHT attributes for the paths are the same. WEIGHT takes precedence over LOCAL_PREF and LOCAL_PREF takes precedence over the AS_PATH attribute. By default, the LOCAL_PREF attribute has a value of 100 and a higher value is preferred over a lower value. LOCAL_PREF is a 32-bit number with a value in the range of 0 to 4,294,967,295.

In Figure 8-7, Routers A and B have two paths to network 156.26.32.0/24. Router A is using the default LOCAL_PREF value of 100 for both paths. Router B has set the LOCAL_PREF attribute for the path shown to 200. Therefore, Router B prefers the path through AS 3. Unlike the WEIGHT attribute, the LOCAL_PREF attribute is advertised throughout the local AS. Router A receives an advertisement for prefix 156.26.32.0/24 from Router B with a LOCAL_PREF of 200. Therefore, Router A also prefers the path through AS 3 to reach the network prefix 156.26.32.0/24.

WEIGHT Attribute

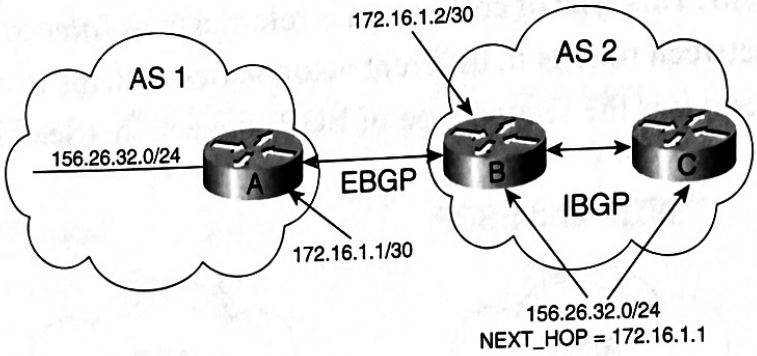
The WEIGHT attribute is not defined in the BGP specification. Cisco has defined the WEIGHT attribute as another option that can be used for determining a best path to a destination. In Figure 8-5, AS 2 is receiving two advertisements for prefix 156.26.32.0/24. The advertisement from AS 3 has an AS path length of 3, and the advertisement from AS 1 has an AS path length of 2. By default, the weight attribute of all received BGP routes is 0, so AS 2 selects the route through AS 1 because it has the shortest AS_PATH attribute.

Figure 8-5 AS 2 Selects the Path to 156.26.32.0/24 Through AS 1 Because It Has the Shortest AS_PATH Attribute



If you want a router in AS 2 to prefer the path through AS 3 instead of the path through AS 1, you can manipulate the WEIGHT attribute. The WEIGHT attribute is a 16-bit number having a value between 0 and 65,535, with a higher weight value taking precedence over a lower weight value. The WEIGHT attribute takes precedence over the AS_PATH attribute, and is used to influence the best path selection when there is more than one path to a particular destination. In Figure 8-6, a router in AS 2 has manipulated the WEIGHT attribute for the route received from AS 3. This path, although it has a longer AS_PATH attribute, will be considered the best path because of the higher WEIGHT attribute of the path.

Figure 8-10 NEXT_HOP Attribute



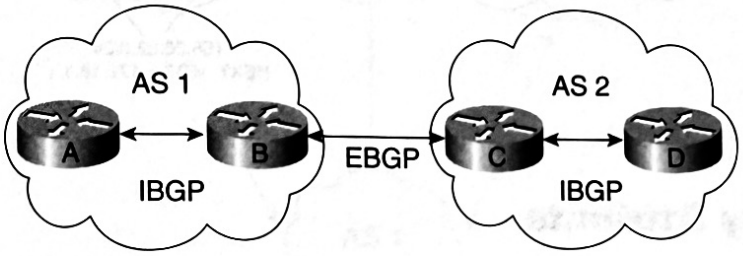
Community Attribute

A router has the ability to manipulate routes sent to or received from a neighbor. All routing protocols can perform route filtering on sent or received routes. With output route filtering, a router can prevent specific routes from being advertised to a neighbor. With input route filtering, a router can reject, or filter out, routes received from a neighbor. For protocols other than BGP, route filtering is based on a prefix and mask. The router is configured with one or more conditions consisting of a prefix and a mask, and any routes that match one of the conditions will be filtered. For example, the condition using the prefix mask combination 156.26.32.0/24 can filter any route that matches 156.26.32.0/24, or is a subnet of this IP address block.

Input and output route filtering can also be used with BGP, but BGP has additional route attributes that can be used to identify the routes to be filtered or manipulated. Instead of simply filtering a route, you may want to manipulate one or more of the attributes associated with a BGP route. In Figure 8-11, AS 1 is receiving 11 routes from three different autonomous systems. Assume AS 1 wants to either filter out (reject) the boxed routes in Figure 8-11, or modify one or more of the same attributes of the indicated routes. You can accomplish this by using four conditions that identify the prefix/mask pairs that you are interested in filtering or manipulating.

There are two types of BGP connections. One is a connection between BGP routers in the same AS. This type of connection is referred to as ***Internal BGP (IBGP)***. A connection between routers in different autonomous systems is called ***External BGP (EBGP)***, and it is the second type of BGP connection. (See Figure 8-9.)

Figure 8-9 IBGP and EBGP



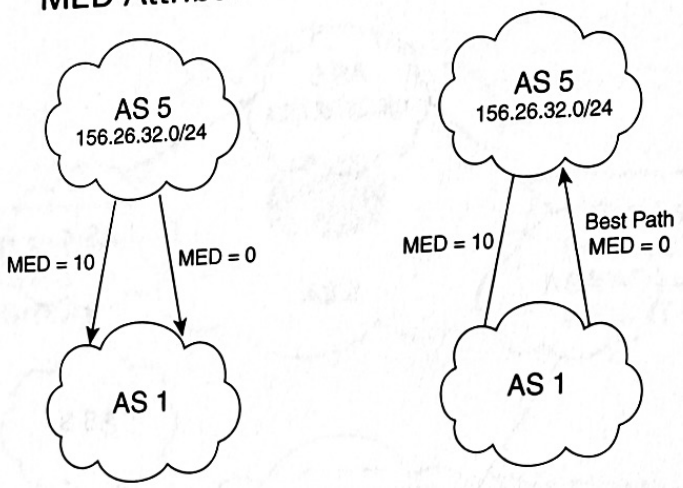
Routes learned from an EBGP neighbor have an ORIGIN attribute of EGP. Any routes that are redistributed into BGP have an ORIGIN attribute of incomplete. In order of preference, IGP routes are preferred over EGP routes and EGP routes are preferred over incomplete routes. EBGP routes are preferred over IBGP routes.

NEXT_HOP Attribute

When a router advertises a prefix to an EBGP neighbor, the NEXT_HOP attribute is set to the IP address of the interface used to connect to the neighbor. When a BGP router advertises a route learned through EBGP to an IBGP neighbor, the NEXT_HOP attribute is not modified. (See Figure 8-10.)

When a router receives a route from a BGP neighbor, the first attribute that is checked is the NEXT_HOP attribute. If the NEXT_HOP is not reachable (not in the local IP routing table), the route is rejected. EBGP neighbors are usually directly connected by a physical interface, so the NEXT_HOP will be in the IP routing table as a connected interface. IBGP neighbors do not have to be directly connected. In Figure 8-10, Router B accepts the route from Router A because the NEXT_HOP network is a directly connected network. Router C will accept the route from Router B if the NEXT_HOP network is in the IP routing table on Router C.

Figure 8-8 MED Attribute



If the other BGP attributes are equal, AS 1 prefers the path on the right to AS 5 because of the lower value of the MED for this path. For the attributes you have seen so far, WEIGHT takes preference over LOACL_PREF, and LOCAL_PREF takes precedence over the shortest AS path. If WEIGHT, LOCAL_PREF, and the AS_PATH attributes are the same for the paths, MED is used to make a best path determination.

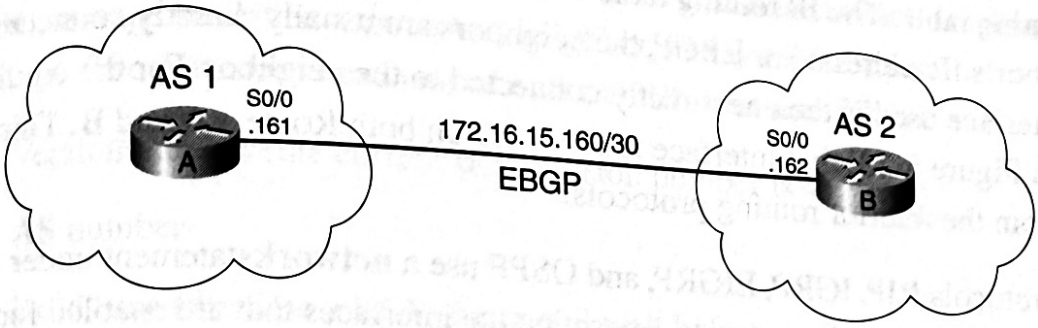
ORIGIN Attribute

The ORIGIN attribute indicates one of the three sources of a route in the BGP routing table. The three ways BGP can learn a route are

- Using the **network**statement
- From an EBGP neighbor
- Through redistribution

The **network**statement is used to cause BGP to advertise a route in the local IP routing table to a BGP neighbor. Routes injected into BGP through the **network** statement have their ORIGIN attribute set to IGP, indicating the route is interior to the originating AS.

Figure 8-13 Configuring EBGP



The first step in configuring EBGP is to configure a loopback interface on Routers A and B. (See Example 8-2.) As with OSPF, the IP address assigned to the loopback interface is used as the BGP router ID.

Example 8-2 Configuring Loopback Interfaces on Routers A and B

```
Router A
interface Loopback0
 ip address 172.16.15.254 255.255.255.255

Router B
interface Loopback0
 ip address 172.16.16.254 255.255.255.255
```

The next step is to enable the BGP process on both routers. (See Example 8-3.)

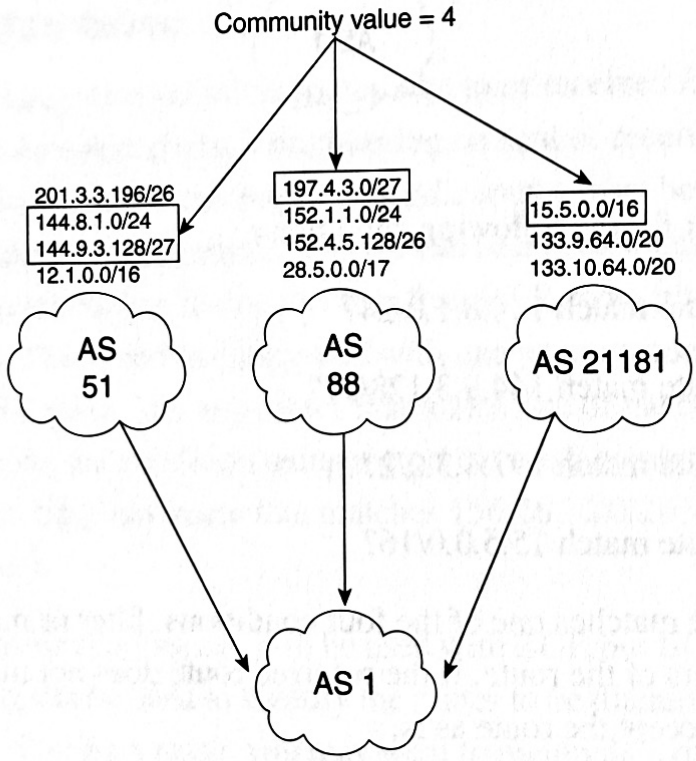
Example 8-3 Enabling the BGP Process on the Routers

```
Router A
router bgp 1
 neighbor 172.16.15.162 remote-as 2

Router B
router bgp 2
 neighbor 172.16.15.161 remote-as 1
```

The general form of the command to enable the BGP process is **router bgp***as-number*. *as-number* is the number that has been assigned to the AS. Only one AS number can be assigned to a service provider, so there can be only one BGP process configured on a router. The command **neighbor** *neighbor-ip-address* enables BGP on an interface. The interface that is enabled for BGP is determined from the

them to do you a favor. That favor is to attach a number to the routes you are interested in manipulating. The number you pick is 4, and you ask your friends to send only the number 4 with the routes you have identified. In BGP, this extra number is called a ***community attribute***, and it is used to group routes so a common policy can be applied to all routes belonging to the same community. (See Figure 8-12.) Therefore, instead of having to configure 4 conditions, or 400 conditions, you need only to configure one condition. That condition is to match any route that has a particular community value. For this example, that community value is 4.

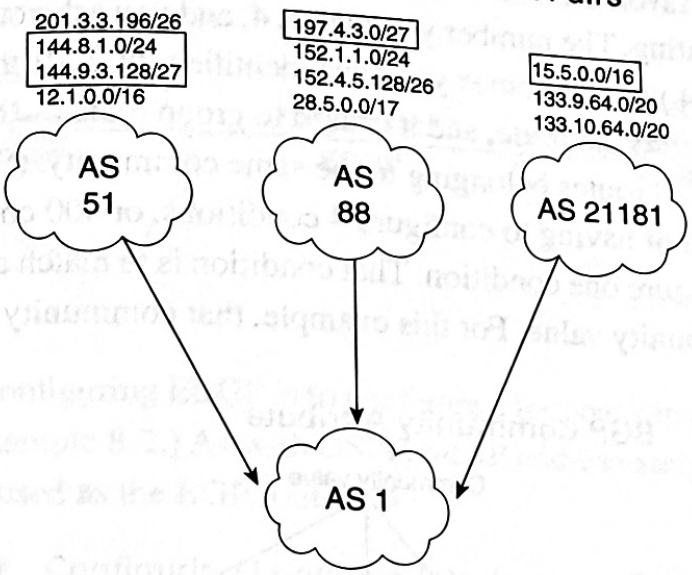
Figure 8-12 BGP Community Attribute



Basic BGP Configuration—EBGP

External BGP (EBGP) neighbors are usually directly connected. In Figure 8-13, you configure EBGP between Routers A and B and investigate the operation and configuration of EBGP neighbors.

Figure 8-11 Filtering Using Prefix and Mask Pairs



You need to check for the following conditions:

- Does the route match 144.8.1.0/24?
- Does the route match 144.9.3.128/27?
- Does the route match 197.4.3.0/27?
- Does the route match 15.5.0.0/16?

If a received route matches one of the four conditions, filter or manipulate the selected parameters of the route. If the received route does not match one of the four conditions, accept the route as is.

For this example, you did not have to worry about too many conditions; there are only four. But there are more than 100,000 BGP routes in the Internet routing table, and if you want to apply the same filter to a large number of routes, your list of conditions could grow to a size that would be difficult to manage.

Although you can filter or manipulate BGP routes based on prefix/mask combinations, you can use another technique to reduce the number of conditions the router needs to check to determine if a particular route is a candidate for filtering or manipulation. You could contact your friends in AS 51, 88, and 21181, and ask

configuration. Even for the most basic BGP configuration there can be up to six errors. The BGP AS numbers may be wrong (two errors), the neighbor's IP address may be wrong (two errors), or the neighbor's remote-AS number may be wrong (two errors). If the configuration parameters are correct, the EBGP neighbors should reach the Established State as shown in the partial listing in Example 8-4.

Example 8-4 Verifying the BGP Connection

```
A#show ip bgp neighbors
BGP neighbor is 172.16.15.162,  remote AS 2, external link
  BGP version 4, remote router ID 172.16.16.254
  BGP state = Established, up for 02:02:11
!Output omitted for brevity
```

The EBGP neighbor for Router A is 172.16.15.162. This is the IP address used in the **neighbor** command on Router A. The neighbor's AS number is 2 and this is an external link because Router A is in AS 1. The BGP version number is 4, which is the current version in use today. The BGP state is Established, and any other state indicates either the routers are still transitioning to the Established State or there are configuration errors. The remote router ID is the BGP router ID of the neighbor. In this case, the router ID is taken from the address of the loopback interface, similar to OSPF. Finally, the keepalive interval is 60 seconds. If a KEEPALIVE message is not received from a neighbor within the holdtime (180 seconds), the neighbor is declared down.

Advertising IP Prefixes

The EBGP neighbors are in the Established State, but neither router is advertising routing information. This is in contrast to the interior routing protocols where neighbor discovery and route advertisement are both tied to the use of the **network** statement. With BGP, there are three ways IP prefixes can be injected into the BGP routing table, as follows:

- **network** command
- Route redistribution
- From a BGP neighbor

In the Connect State, a BGP router sends an OPEN message to each neighbor. The OPEN message identifies a router to a neighbor, and is used for parameter exchange. The parameters exchanged in the OPEN message follow:

- Version number (the current BGP version number is 4)
- AS number
- Hold time (the time a BGP router will wait before declaring a neighbor down)
- BGP router ID

After a BGP router sends an OPEN message to a neighbor, the router transitions to the OpenSent State, and waits to receive an OPEN message from the neighbor. When the OPEN message is received from the neighbor, the routers transition to the OpenConfirm State, and wait for a KEEPALIVE message. By default, KEEPALIVE messages are sent every 60 seconds, and are used to verify the neighbor is still there. After the neighbors have received a KEEPALIVE message, the routers transition to the Established State, and exchange KEEPALIVE, UPDATE, and NOTIFICATION messages.

An UPDATE message advertises or withdraws network prefixes. An UPDATE message advertises the following:

- IP prefixes and subnet mask lengths; also referred to as Network Layer Reachability Information (NLRI)
- BGP attributes
- Withdrawn routes

A NOTIFICATION message is used when an error occurs that causes the BGP session to be shut down.

If the BGP neighbors do not reach the Established State, there is a problem. The problem could be as simple as one of the BGP neighbors has not yet been configured. For this case, the configured neighbor will be in the Active State waiting to create a TCP connection with the neighbor. There could be errors in the BGP

IP routing table. The IP routing table determines the interface used to reach the neighbor's IP address. For EBGP, the neighbors are usually directly connected, so the interface used is the one directly connected to the neighbor. For the configuration in Figure 8-13, that interface is serial 0/0 on both Routers A and B. This differs from the interior routing protocols.

The protocols RIP, IGRP, EIGRP, and OSPF use a **network**statement under the routing process configuration to determine the interfaces that are enabled for a protocol, and which prefixes and masks are advertised. The enabled interfaces for the interior routing protocols are also used for neighbor discovery (EIGRP, OSPF, and IS-IS). Interfaces are enabled for IS-IS under the interface configuration using the **ip router isis**command. With EBGP you don't need to discover EBGP neighbors because you already know the interface used to reach the neighbor. Therefore, you can form an EBGP neighbor relationship without advertising any routes. When EBGP is enabled on an interface, the routers attempt to form a BGP relationship by progressing through the following states:

- **Idle State**–Initialize a TCP connection with the BGP neighbor.
- **Connect State**–Wait for a TCP connection to be established with the BGP neighbor. If a TCP connection is established, send an OPEN message and transition to the OpenSent State. If a connection timeout occurs, transition to the Active State.
- **Active State**–Wait for a TCP connection to be established with the BGP neighbor. If a TCP connection is established, transition to the Connect State.
- **OpenSent State**–Wait for an OPEN message from the BGP neighbor. If errors occur, transition to the Idle State. If an OPEN message is received from the BGP neighbor, transition to the OpenConfirm State.
- **OpenConfirm State**–Wait for a KEEPALIVE message from the BGP neighbor. When a KEEPALIVE is received, transition to the Established State.
- **Established State**–Exchange routing information.

In Example 8-6, a next hop of 0.0.0.0 and a blank AS path attribute indicate that 172.16.0.0 originates on Router A. You can also see that the Metric, or MED, is 0 and the WEIGHT is 32768 (default values). After a valid route is injected into the BGP routing table, the route will be advertised to Router B as shown in Example 8-7.

Example 8-7 BGP Routing Table on Router B

```
B#show ip bgp
BGP table version is 4, local router ID is 172.16.16.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.0.0       172.16.15.161            0             0 1 i
```

In Example 8-7, the 172.16.0.0/16 prefix has been successfully advertised to Router B. Notice that the NEXT_HOP attribute is equal to the IP address used to reach Router B, the AS path attribute is 1, and the Metric and WEIGHT attributes have a value of 0. This BGP route is a valid route and is transferred to the IP routing table on Router B, as shown in Example 8-8.

Example 8-8 IP Routing Table on Router B

```
B#show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is not set

     172.16.0.0/16 is variably subnetted, 4 subnets, 3 masks
C       172.16.16.164/30 is directly connected, Serial0/1
C       172.16.15.160/30 is directly connected, Serial0/0
C       172.16.16.254/32 is directly connected, Loopback0
B       172.16.0.0/16 [20/0] via 172.16.15.161, 00:09:41
```

You have successfully advertised your first BGP route. From the routing
you see the administrative

Example 8-5 Router A Contains Only Connected Routes (continued)

```
Gateway of last resort is not set

     172.16.0.0/16 is variably subnetted, 4 subnets, 3 masks
C       172.16.15.160/30 is directly connected, Serial0/0
C       172.16.15.128/27 is directly connected, Ethernet0/0
C       172.16.15.254/32 is directly connected, Loopback0
C       172.16.15.64/27 is directly connected, Ethernet0/1
```

The operation of the **network**command depends on the state of autosummarization configured under the BGP routing process. By default, autosummarization is enabled, so we examine this case first. The format of the network command is

network*prefix* **mask** *subnet-mask*

The *subnet-mask* parameter is optional. If the *subnet-mask* parameter is not used, BGP assumes a classful prefix. With autosummarization enabled, the **network** command can be used to inject a classful prefix into the BGP routing table. For this operation to be successful, there must exist at least one subnet of the classful prefix in the IP routing table. The routing table on Router A contains subnets of 172.16.0.0/16. Therefore, you can inject the classful prefix 172.16.0.0/16 using the following form of the **network**command:

```
router bgp 1
 network 172.16.0.0
 neighbor 172.16.15.162 remote-as 2
```

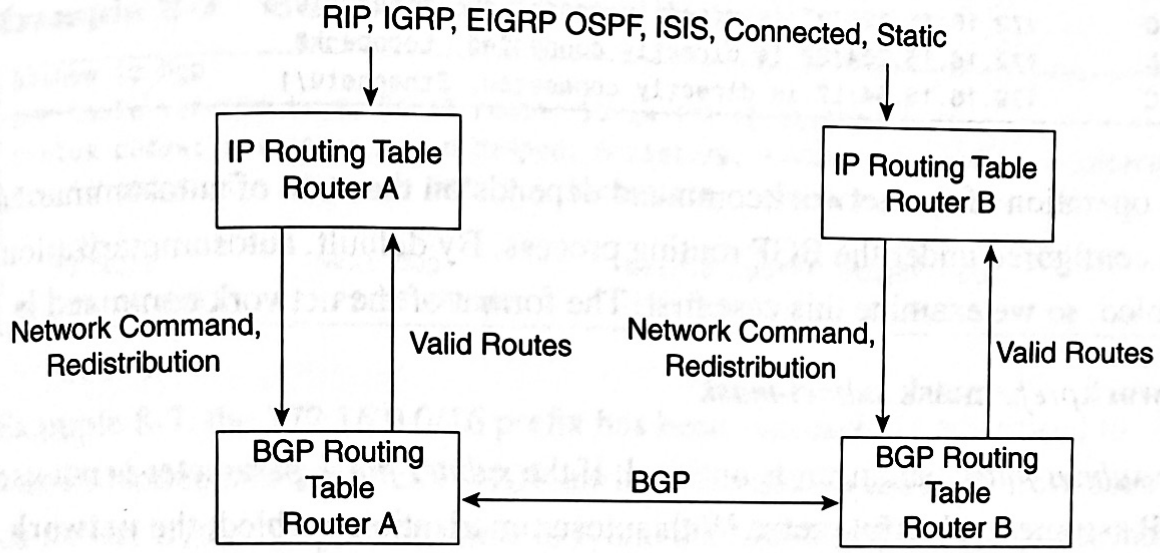
You should see this classful prefix in the BGP routing table on Router A. (See Example 8-6.)

Example 8-6 BGP Routing Table on Router A

```
A#show ip bgp
BGP table version is 8, local router ID is 172.16.15.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.0.0       0.0.0.0                  0         32768 i
```

Figure 8-14 illustrates the three methods.

Figure 8-14 Prefixes Enter the Local BGP Routing Table Using the **network** Command, Route Redistribution, or from a BGP Neighbor



Using the network Command to Inject Routes

The first way to enter prefixes into the BGP routing table is using the **network** command under the BGP configuration. The **network** command is used to transfer prefixes from the IP routing table to the BGP routing table as BGP routes. Prefixes in the IP routing table are learned from a dynamic interior routing protocol (RIP, IGRP, EIGRP, OSPF, or IS-IS), by configuring static routes, or from the IP address and mask assigned to a directly connected interface. The IP routing table on Router A contains only connected routes because no other IP routing protocols have been configured, as shown in Example 8-5.

Example 8-5 Router A Contains Only Connected Routes

```
A#show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route
```

continues

Example 8-13 Four Prefixes in Router A and B Routing Tables (continued)

```
*> 172.16.15.128/27 0.0.0.0                    0          32768 i
*> 172.16.15.160/30 0.0.0.0                    0          32768 i
*> 172.16.15.254/32 0.0.0.0                    0          32768 I

B#show ip bgp
BGP table version is 9, local router ID is 172.16.16.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.15.64/27  172.16.15.161        0             0 1 i
*> 172.16.15.128/27 172.16.15.161        0             0 1 i
*> 172.16.15.160/30 172.16.15.161        0             0 1 i
*> 172.16.15.254/32 172.16.15.161        0             0 1 i
```

Example 8-14 verifies the BGP routes have been transferred to the IP routing table on Router B.

Example 8-14 IP Routing Table on Router B

```
B#show ip route
     172.16.0.0/16 is variably subnetted, 6 subnets, 3 masks
C       172.16.16.164/30 is directly connected, Serial0/1
C       172.16.15.160/30 is directly connected, Serial0/0
B       172.16.15.128/27 [20/0] via 172.16.15.161, 00:02:36
B       172.16.15.254/32 [20/0] via 172.16.15.161, 00:02:36
C       172.16.16.254/32 is directly connected, Loopback0
B       172.16.15.64/27 [20/0] via 172.16.15.161, 00:02:36
```

There are four prefixes in the BGP routing table on Router B, but only three BGP learned prefixes in the IP routing table. Router A is advertising network 172.16.15.160/30 to Router B via BGP. This is a directly connected network on Router B, so it is in the IP routing table as a connected network and not a BGP learned prefix. (See Example 8-15.)

command must exactly match a prefix and mask in the IP routing table. If there is not an exact match, the route will not be transferred to the BGP routing table. The command used was **network 172.16.0.0**and the router now looks for an exact match in the IP routing table. There are subnets of 172.16.0.0 in the IP routing table on Router A, but not the specific 172.16.0.0/16 route. To inject the connected networks on Router A into the BGP routing table, you must use the exact prefix and mask with the **network**command. Router A has four directly connected networks that you want to turn into BGP routes. (See Example 8-11.)

Example 8-11 Router A's Directly Connected Routes

```
A#show ip route | include directly connected
C       172.16.15.160/30 is directly connected, Serial0/0
C       172.16.15.128/27 is directly connected, Ethernet0/0
C       172.16.15.254/32 is directly connected, Loopback0
C       172.16.15.64/27 is directly connected, Ethernet0/1
```

The **network**commands needed to inject these prefixes into BGP are shown in Example 8-12.

Example 8-12 Required **network** Commands When Autosummary Is Disabled

```
Router A
router bgp 1
 network 172.16.15.64 mask 255.255.255.224
 network 172.16.15.128 mask 255.255.255.224
 network 172.16.15.160 mask 255.255.255.252
 network 172.16.15.254 mask 255.255.255.255
 neighbor 172.16.15.162 remote-as 2
 no auto-summary
```

You should now see four prefixes in the BGP routing tables on Routers A and B. (See Example 8-13.)

Example 8-13 Four Prefixes in Router A and B Routing Tables

```
A#show ip bgp
BGP table version is 13, local router ID is 172.16.15.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.15.64/27  0.0.0.0                  0         32768 i
```

continues

BGP route is 0. BGP does not use a cost to determine the best path but instead uses the BGP attributes of WEIGHT, LOCAL_PREF, AS_PATH, MED, and ORIGIN.

Cisco recommends that you disable autosummarization when using BGP. This is accomplished using the **no auto-summary** command under the BGP configuration, as shown in Example 8-9.

Example 8-9 Disabling Autosummarization

```
Router A
router bgp 1
 network 172.16.0.0
 neighbor 172.16.15.162 remote-as 2
 no auto-summary

Router B
router bgp 2

 neighbor 172.16.15.161 remote-as 1
 no auto-summary
```

What happens when you disable autosummarization on the BGP routers? Will network 172.16.0.0/16 be advertised? If you examine the BGP table on Routers A and B and the IP routing table on Router B, you see that this prefix is no longer being advertised. (See Example 8-10.)

Example 8-10 Routing Tables on Routers A and B

```
A#show ip bgp
!no output

B#show ip bgp
!no output

B#show ip route
     172.16.0.0/16 is variably subnetted, 3 subnets, 2 masks
C       172.16.16.164/30 is directly connected, Serial0/1
C       172.16.15.160/30 is directly connected, Serial0/0
C       172.16.16.254/32 is directly connected, Loopback0
```

When autosummarization is disabled, this changes the operation of the **network** command. The new rule is that the prefix and mask used with the **network**

Example 8-20 Verifying the Redistributed Routes (continued)

```
B#show ip bgp
BGP table version is 21, local router ID is 172.16.16.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.15.64/27  172.16.15.161            0             0 1 ?
*> 172.16.15.128/27 172.16.15.161            0             0 1 ?
*> 172.16.15.160/30 172.16.15.161            0             0 1 ?
*> 172.16.15.254/32 172.16.15.161            0             0 1 ?

B#show ip route
     172.16.0.0/16 is variably subnetted, 6 subnets, 3 masks
C       172.16.16.164/30 is directly connected, Serial0/1
C       172.16.15.160/30 is directly connected, Serial0/0
B       172.16.15.128/27 [20/0] via 172.16.15.161, 00:01:44
B       172.16.15.254/32 [20/0] via 172.16.15.161, 00:01:44
C       172.16.16.254/32 is directly connected, Loopback0
B       172.16.15.64/27 [20/0] via 172.16.15.161, 00:01:44
```

You can see from the BGP routing tables on Routers A and B that the redistributed routes have an origin of ? or incomplete.

Using BGP to Inject Routes

The final method of injecting routes into the BGP routing table is through BGP itself. You have seen this on Router B because the routes in Router B's BGP routing table were learned through EBGP from Router A.

The network in Figure 8-15 demonstrates additional EBGP properties, and illustrates the BGP AS_PATH attribute. An additional AS has been added as a BGP neighbor to Router B. The BGP configuration on Router B contains the new neighbor in AS 3. (See Example 8-21.)

The connected routes should now be in the BGP routing table on Router A, as shown in Example 8-18.

Example 8-18 Connected Routes in Router A's BGP Routing Table

```
A#show ip bgp
BGP table version is 2, local router ID is 172.16.15.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.0.0       0.0.0.0                  0         32768 ?
```

Notice that the BGP routing table has just one route. When autosummarization is enabled, BGP automatically summarizes redistributed routes to a classful boundary. All of the connected interfaces on Router A are subnets of the Class B address block 172.16.0.0/16 so Router A automatically summarizes the connected networks. Disable autosummarization on Router A so the specific prefixes assigned to the connected interfaces are injected into BGP. (See Example 8-19.)

Example 8-19 Disabling Autosummarization on Router A

```
Router A
router bgp 1
 redistribute connected
 neighbor 172.16.15.162 remote-as 2
 no auto-summary
```

The specific prefixes and masks for the connected networks should now be in the BGP routing tables on Routers A and B, and in the IP routing table on Router B. (See Example 8-20.)

Example 8-20 Verifying the Redistributed Routes

```
A#show ip bgp
BGP table version is 7, local router ID is 172.16.15.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.15.64/27  0.0.0.0                  0         32768 ?
*> 172.16.15.128/27 0.0.0.0                  0         32768 ?
*> 172.16.15.160/30 0.0.0.0                  0         32768 ?
*> 172.16.15.254/32 0.0.0.0                  0         32768 ?
```

Example 8-15 Network Between Routers A and B Is Directly Connected

```
B#show ip route
     172.16.0.0/16 is variably subnetted, 6 subnets, 3 masks
C       172.16.16.164/30 is directly connected, Serial0/1
C       172.16.15.160/30 is directly connected, Serial0/0
B       172.16.15.128/27 [20/0] via 172.16.15.161, 00:02:36
B       172.16.15.254/32 [20/0] via 172.16.15.161, 00:02:36
C       172.16.16.254/32 is directly connected, Loopback0
B       172.16.15.64/27 [20/0] via 172.16.15.161, 00:02:36
```

Using Redistribution to Inject Routes

The second method for injecting routes into the BGP routing table is to use redistribution. On Router A, re-enable autosummarization using the **auto-summary** command, and delete the network statements using the **no** form of the **network** command. (See Example 8-16.)

Example 8-16 Removing Commands from the Configuration

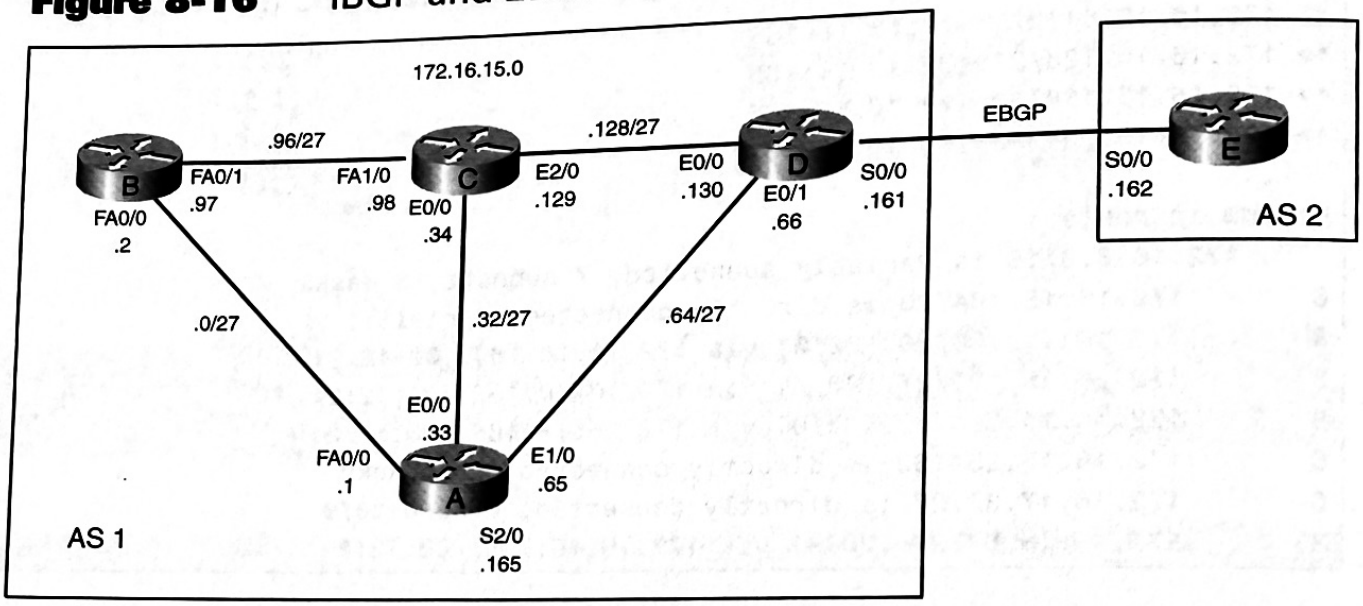
```
Router A
router bgp 1
 no network 172.16.15.64 mask 255.255.255.224
 no network 172.16.15.128 mask 255.255.255.224
 no network 172.16.15.160 mask 255.255.255.252
 no network 172.16.15.254 mask 255.255.255.255
 neighbor 172.16.15.162 remote-as 2
 auto-summary
```

Configure Router A to inject the connected networks into BGP using redistribution, as shown in Example 8-17.

Example 8-17 Configuring Redistribution

```
Router A
router bgp 1
 redistribute connected
 neighbor 172.16.15.162 remote-as 2
```

Figure 8-16 IBGP and EBGP Connections



EBGP and IBGP use the same set of configuration commands for enabling the BGP process and for advertising network prefixes. The differences between EBGP and IBGP are in their operation, and not in their configuration. To begin, EBGP is configured between Routers D and E with Router E advertising its directly connected interfaces using route redistribution. (See Example 8-23.)

Example 8-23 Configuring EBGP Between Routers D and E

```
Router D
router bgp 1

 neighbor 172.16.15.162 remote-as 2
 no auto-summary

Router E
router bgp 2
 redistribute connected
 neighbor 172.16.15.161 remote-as 1
 no auto-summary
```

Router D should have the prefixes advertised by Router E in both the BGP and IP routing tables. (See Example 8-24.)

Example 8-22 BGP and IP Routing Tables on Router C (continued)

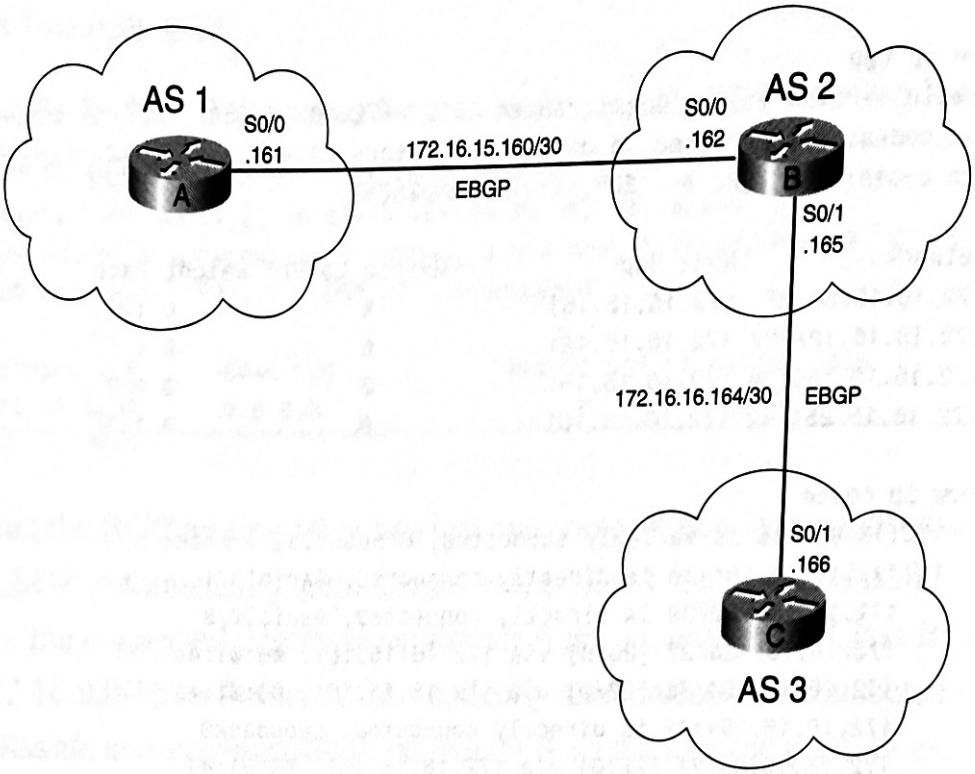
```
*> 172.16.15.64/27   172.16.16.165
*> 172.16.15.128/27  172.16.16.165                    0 2 1 ?
*> 172.16.15.160/30  172.16.16.165                    0 2 1 ?
*> 172.16.15.254/32  172.16.16.165                    0 2 1 ?
                                                      0 2 1 ?

C#show ip route
     172.16.0.0/16 is variably subnetted, 7 subnets, 3 masks
C       172.16.16.164/30 is directly connected, Serial0/1
B       172.16.15.160/30 [20/0] via 172.16.16.165, 00:12:21
B       172.16.15.128/27 [20/0] via 172.16.16.165, 00:12:21
B       172.16.15.254/32 [20/0] via 172.16.16.165, 00:12:21
C       172.16.17.254/32 is directly connected, Loopback0
C       172.16.17.32/27 is directly connected, Ethernet0/0
B       172.16.15.64/27 [20/0] via 172.16.16.165, 00:12:21
```

The AS_PATH attribute on Router C has a value of 2 1, as expected. One of the main points of this exercise was to demonstrate that an EBGP router advertises routes learned from one EBGP neighbor to another EBGP neighbor. The other main point is Router B sets the NEXT_HOP attribute to the IP address of the interface it is using to communicate with Router C. In the next section, you learn this is not the case when using internal or IBGP.

Basic BGP Configuration—IBGP

The network in Figure 8-16 will be used to investigate the configuration and operation of Internal BGP (IBGP). Router D has an EBGP connection to AS 2 and IBGP connections to the routers in AS 1.

Figure 8-15 EBGP Route Advertisement



Example 8-21 New Neighbor for Router B

```
Router B
router bgp 2
 neighbor 172.16.15.161 remote-as 1
 neighbor 172.16.16.166 remote-as 3
 no auto-summary

Router C
router bgp 3

 neighbor 172.16.16.165 remote-as 2
```

Router B in AS 2 advertises the prefixes learned through EBGP from Router A to Router C. (See Example 8-22.)

Example 8-22 BGP and IP Routing Tables on Router C

```
C#show ip bgp
BGP table version is 5, local router ID is 172.16.17.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
```

Example 8-29 NEXT_HOP Attribute Is Unchanged by IBGP

```
Router D
*> 172.16.15.160/30 172.16.15.162                    0                0 2 ?
Router C
* i172.16.15.160/30 172.16.15.162                    0    100         0 2 ?
```

Both routers have the same next hop for network 172.16.15.160/30, but Router D considers the route valid (*>) and Router C considers the route valid (*), but there is not a best path to the prefix. The specifics of the route can be checked using the **show ip bgp** *ip-prefix* command on Router C. (See Example 8-30.)

Example 8-30 The NEXT_HOP Is Unknown to Router C

```
C#show ip bgp 172.16.15.160
BGP routing table entry for 172.16.15.160/30, version 0
Paths: (1 available, no best path)
  Not advertised to any peer
  2
    172.16.15.162 (inaccessible) from 172.16.15.130 (172.16.15.254)
      Origin incomplete, metric 0, localpref 100, valid, internal, not
        synchronized
```

The prefix 172.16.15.160 and the other prefixes received from Router D are inaccessible. This means Router C does not know how to reach the NEXT_HOP. Recall BGP does not change the NEXT_HOP information received from an EBGP peer when the prefix is advertised to an IBGP peer. Router D does not change the NEXT_HOP of 172.16.15.162 when the prefixes are advertised to Router C. These routes are considered inaccessible because there is not a route in the IP routing table for prefix 172.16.15.160/30. One way to overcome this problem is to run an IGP in AS 1 and advertise not only the next-hop network but also all networks belonging to AS 1. For this exercise, use single-area OSPF on Routers A, B, C, and D. The OSPF configurations are identical for all routers in AS 1. (See Example 8-31.)

Example 8-31 OSPF Configuration for Routers in AS 1

```
router ospf 1
 network 172.16.0.0 0.0.255.255 area 51
```

Example 8-26 Verifying a Specific BGP Neighbor

```
D#show ip bgp neighbors 172.16.15.129
BGP neighbor is 172.16.15.129, remote AS 1, internal link
  BGP version 4, remote router ID 172.16.15.253
  BGP state = Established, up for 00:28:24
  Last read 00:00:24, hold time is 180, keepalive interval is 60 seconds
!Output omitted for brevity
```

Routers C and D have established an IBGP neighbor relationship indicated by "internal link" displayed in the output in Example 8-26. Router D should be advertising the prefixes received from Router E to Router C as shown in Example 8-27.

Example 8-27 BGP Routing Table on Router C

```
C#show ip bgp
BGP table version is 1, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
* i172.16.15.160/30 172.16.15.162          0    100      0 2 ?
* i172.16.16.164/30 172.16.15.162          0    100      0 2 ?
* i172.16.16.254/32 172.16.15.162          0    100      0 2 ?
```

Verify that the BGP routes have been transferred to the IP routing table on Router C. (See Example 8-28.)

Example 8-28 IP Routing Table on Router C

```
C#show ip route
     172.16.0.0/16 is variably subnetted, 5 subnets, 3 masks
C       172.16.15.128/27 is directly connected, Ethernet2/0
C       172.16.15.253/32 is directly connected, Loopback0
C       172.16.15.32/27 is directly connected, Ethernet0/0
C       172.16.9.0/30 is directly connected, Serial0/1
C       172.16.15.96/27 is directly connected, FastEthernet1/0
```

The routes have not been transferred from the BGP routing table to the IP routing table on Router C. Any idea what the problem is? If you look at one of the entries in the BGP routing table on Routers C and D, you will notice a slight difference, as highlighted in Example 8-29.

Example 8-24 Router D BGP and IP Routing Tables

```
D#show ip bgp
BGP table version is 4, local router ID is 172.16.15.254
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 172.16.15.160/30 172.16.15.162            0             0 2 ?
*> 172.16.16.164/30 172.16.15.162            0             0 2 ?
*> 172.16.16.254/32 172.16.15.162            0             0 2 ?

D#show ip route
     172.16.0.0/16 is variably subnetted, 6 subnets, 3 masks
B       172.16.16.164/30 [20/0] via 172.16.15.162, 00:27:23
C       172.16.15.160/30 is directly connected, Serial0/0
C       172.16.15.128/27 is directly connected, Ethernet0/0
C       172.16.15.254/32 is directly connected, Loopback0
B       172.16.16.254/32 [20/0] via 172.16.15.162, 00:27:23
C       172.16.15.64/27 is directly connected, Ethernet0/1
```

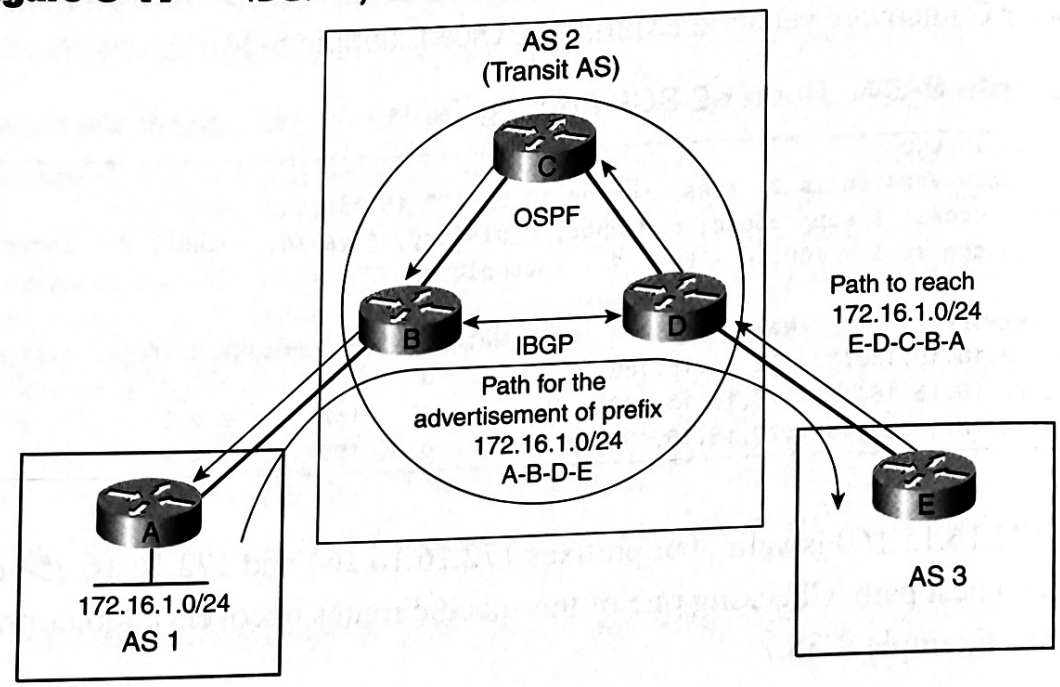
Configuring IBGP between Routers C and D is no different than configuring EBGP except the AS numbers on Routers C and D are the same. (See Example 8-25.)

Example 8-25 Configuring IBGP Between Routers C and Router D

```
Router C
router bgp 1
 neighbor 172.16.15.130 remote-as 1
 no auto-summary

Router D
router bgp 1
 neighbor 172.16.15.129 remote-as 1
 neighbor 172.16.15.162 remote-as 2
 no auto-summary
```

After the IBGP processes and neighbors are configured, check to see if the neighbors have established a BGP neighbor relationship. (See Example 8-26.). If you use the **show ip bgp neighbors,**the router displays all BGP neighbors. To use the command to display one neighbor, use the neighbor's IP address.

Figure 8-17 IBGP Synchronization



Routers A and B have established an EBGP connection and Router A is advertising prefix 172.16.1.0/24 to Router B. In AS 2, the physical path between Routers B and D is through Router C. A physical connection between IBGP routers is not required, and Routers B and D have established an IBGP connection. BGP packets exchanged between Routers B and D physically pass through Router C, but Router C has no knowledge of their contents.

Routers B and D need to know how to reach each other through IP so AS 2 is running OSPF and advertising all networks internal to AS 2. All that is required to establish an IBGP connection is an IP path between the two routers. Assume synchronization is not a property of IBGP and trace the chain of events in Figure 8-17:

1. Router A advertises prefix 172.16.1.0/24 to Router B through EBGP.
2. Router B accepts the route because the next hop is known (it's a directly connected network), and the prefix is placed in the routing table on Router B.
3. Router B advertises the prefix to Router D through IBGP without changing the next hop information.
4. Router D accepts the route because the next hop is known through OSPF.

No BGP routes appear in the IP routing table. Looking at the BGP routing table on Router C uncovers yet another problem. (See Example 8-34.)

Example 8-34 Router C BGP Routing Table

```
C#show ip bgp
BGP table version is 2, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*>i172.16.15.160/30 172.16.15.162            0    100      0 2 ?
* i172.16.16.164/30 172.16.15.162            0    100      0 2 ?
* i172.16.16.254/32 172.16.15.162            0    100      0 2 ?
```

Prefix 172.16.15.160 is valid, but prefixes 172.16.16.164 and 172.16.16.254 do not have a best path. Checking one of the specific routes uncovers another problem. (See Example 8-35.)

Example 8-35 Prefix in AS 1 Is Not Synchronized

```
C#show ip bgp 172.16.16.164
BGP routing table entry for 172.16.16.164/30, version 0
Paths: (1 available, no best path)
  Not advertised to any peer
  2
    172.16.15.162 (metric 74) from 172.16.15.130 (172.16.15.254)
      Origin incomplete, metric 0, localpref 100, valid, internal,
        not synchronized
```

The two invalid BGP routes are not synchronized. Synchronization is a property of IBGP. A BGP prefix received from an IBGP neighbor is not considered valid unless the prefix is already in the IP routing table.

The BGP prefix 172.16.15.160 is synchronized because this prefix is in the IP routing table as an OSPF route. The BGP prefixes 172.16.16.164 and 172.16.16.254 are not in the IP routing table, therefore, they are not synchronized. The reason for synchronization can be explained using the network in Figure 8-17.

The network connecting Routers D and E is now being advertised by OSPF making the BGP next hop advertised by Router D accessible. (See Example 8-32.)

Example 8-32 Verifying the Next Hop Is Being Advertised in AS 1

```
C#show ip route | include 172.16.15.160
O        172.16.15.160/30 [110/74] via 172.16.15.130, 00:05:45, Ethernet2/0

C#show ip bgp 172.16.15.160
BGP routing table entry for 172.16.15.160/30, version 2
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Not advertised to any peer
  2
    172.16.15.162 (metric 74) from 172.16.15.130 (172.16.15.254)
     Origin incomplete, metric 0, localpref 100, valid, internal, synchronized,
       best
```

The BGP routes should now be in the IP routing table on Router C, as shown in Example 8-33.

Example 8-33 Router C IP Routing Table

```
C#show ip route
     172.16.0.0/16 is variably subnetted, 12 subnets, 3 masks
O        172.16.15.160/30 [110/74] via 172.16.15.130, 00:08:45, Ethernet2/0
O        172.16.15.164/30 [110/50] via 172.16.15.97, 00:08:45, FastEthernet1/0
C        172.16.15.128/27 is directly connected, Ethernet2/0
O        172.16.15.251/32 [110/3] via 172.16.15.97, 00:08:45, FastEthernet1/0
O        172.16.15.254/32 [110/11] via 172.16.15.130, 00:08:45, Ethernet2/0
C        172.16.15.253/32 is directly connected, Loopback0
O        172.16.15.252/32 [110/2] via 172.16.15.97, 00:08:46, FastEthernet1/0
C        172.16.15.32/27 is directly connected, Ethernet0/0
O        172.16.15.0/27 [110/2] via 172.16.15.97, 00:08:46, FastEthernet1/0
C        172.16.9.0/30 is directly connected, Serial0/1
C        172.16.15.96/27 is directly connected, FastEthernet1/0
O        172.16.15.64/27 [110/12] via 172.16.15.97, 00:08:46, FastEthernet1/0
```

Prefix 172.16.15.160 is in the IP routing table as an OSPF route. Why is this prefix not an IBGP route? Router C is learning about prefix 172.16.15.160 through OSPF and IBGP. OSPF has an administrative distance of 110, and IBGP has an administrative distance of 200. Therefore, the OSPF route is preferred over the IBGP route.

Before inspecting the BGP routing table on Router B, check to see if a BGP neighbor relationship has been established. (See Example 8-40.)

Example 8-40 Verifying the BGP Neighbor Relationship

```
B#show ip bgp neighbors 172.16.15.98 | include Established
  BGP state = Established, up for 00:04:22
```

To answer the question of which routes Router C is advertising, display the contents of the BGP routing table on Router B.

Example 8-41 BGP Routing Table on Router B

```
B#show ip bgp
!no output
```

Router C has not advertised any BGP routes to Router B. This illustrates another property of IBGP that differs from the operation of EBGP. All IBGP routes have the same AS number so the AS path attribute cannot be used for loop detection. Therefore, an IBGP router will not advertise routes learned from one IBGP neighbor to another IBGP neighbor. This restriction means that IBGP routers have to be fully meshed. In other words, all IBGP routers in an AS must have an IBGP connection to all other IBGP routers. In Figure 8-16, each router must have an IBGP connection to all other routers in the AS. This requires three IBGP connections on Routers A, B, C, and D, for a total of six IBGP connections. (See Example 8-42.) Because this is IBGP, the routers do not have to be physically connected.

Example 8-42 Configuring a Full IBGP Mesh

```
Router A
router bgp 1
 no synchronization
 neighbor 172.16.15.2 remote-as 1
 neighbor 172.16.15.34 remote-as 1
 neighbor 172.16.15.66 remote-as 1

Router B
router bgp 1
 no synchronization
 neighbor 172.16.15.1 remote-as 1
 neighbor 172.16.15.98 remote-as 1
 neighbor 172.16.15.130 remote-as 1
```

Now our problem is solved. Or is it? Router C has routes to the EBGP routes, but Router E does not have a route to Router C. A static route can be used on Router E to reach all the networks in AS 1. (See Example 8-38.)

Example 8-38 Static Route on Router E

```
Router E
ip route 172.16.15.0 255.255.255.0 172.16.15.161
```

The main point of this exercise is that simply turning off synchronization does not solve our problem. We have to ensure that any routes injected into the IP routing table can be reached by the router.

With synchronization enabled, routes learned through IBGP are not considered valid unless the routes are already in the IP routing table. If the routes are already in the IP routing table, why do we need BGP? In Figure 8-17, AS 2 is a BGP transit system between AS 1 and AS 3. You could redistribute the BGP routes into OSPF on Router B, advertise the redistributed routes to Routers C and D through OSPF, and redistribute the routes back into BGP on Router D. This works and routing will be complete, but you lose the BGP attributes when the routes are redistributed into OSPF. The problem in Figure 8-17 is that all the routers in the physical path between AS 1 and AS 3 are not running IBGP. If you configure IBGP on Router C, synchronization can be disabled in AS 2 as long as the next hop advertised by the IBGP routers is reachable.

In Figure 8-16, IBGP is configured between Routers B and C. What BGP routes does Router C advertise to Router B? (See Example 8-39.)

Example 8-39 Routes Advertised from Router C to Router B

```
Router C
router bgp 1
 no synchronization
 neighbor 172.16.15.97 remote-as 1
 neighbor 172.16.15.130 remote-as 1
 no auto-summary

Router B
router bgp 1
 neighbor 172.16.15.98 remote-as 1
```

5. Router D advertises 172.16.1.0/24 to Router E in AS 3.
6. Router E accepts the prefix because the next hop is known, and the prefix is placed in the IP routing table on Router E.
7. Assume a host in AS 3 sends a packet to a host on network 172.16.1.0/24. The host sends the packet to Router E, and Router E forwards the packet to Router D after consulting the IP routing table.
8. Router D receives the packet and looks at the destination IP address to make a forwarding decision. The next hop is Router C, and the packet is forwarded to Router C.
9. Router C has no knowledge of network 172.16.1.0/24 because this prefix was not learned through OSPF or BGP, so Router C drops the packet.

Synchronization can be disabled using the **no synchronization** command. With synchronization disabled on Router C in Figure 8-16, the BGP routes advertised by Router D are transferred to the IP routing table on Router C. (See Example 8-36.)

Example 8-36 Disabling Synchronization

```
Router C
router bgp 1
 no synchronization
 neighbor 172.16.15.130 remote-as 1
 no auto-summary
```

With synchronization disabled, the routes in the BGP table on Router C are now valid and the BGP routes are transferred to the IP routing table. (See Example 8-37.)

Example 8-37 Transferring BGP Routes to the IP Routing Table

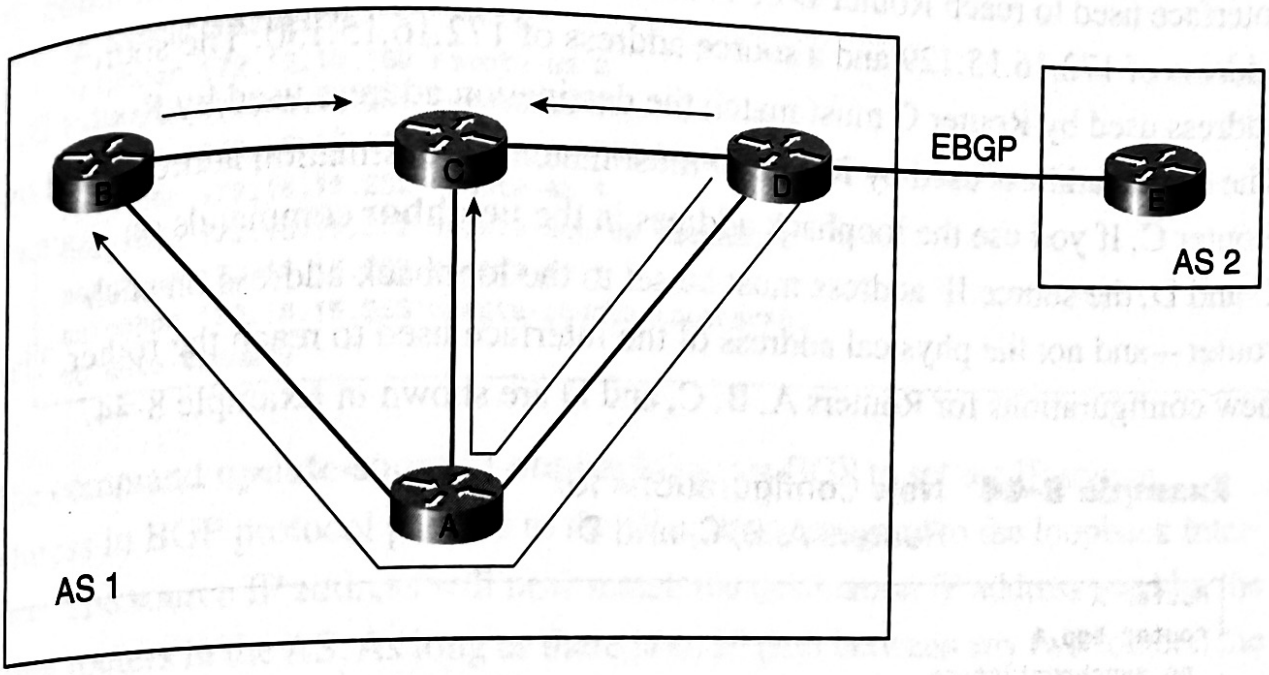
```
C#show ip bgp
BGP table version is 4, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*>i172.16.15.160/30 172.16.15.162            0    100      0 2 ?
*>i172.16.16.164/30 172.16.15.162            0    100      0 2 ?
*>i172.16.16.254/32 172.16.15.162            0    100      0 2 ?

C#show ip route | i B
B       172.16.16.164/30 [200/0] via 172.16.15.162, 00:02:28
B       172.16.16.254/32 [200/0] via 172.16.15.162, 00:02:28
```

172.16.15.130. The source IP address used by Router C is the IP address of the interface used to reach Router D or 172.16.15.129. Router D uses a destination IP address of 172.16.15.129 and a source address of 172.16.15.130. The source address used by Router C must match the destination address used by Router D. The source address used by Router D must match the destination address used by Router C. If you use the loopback address in the **neighbor** commands on Routers C and D, the source IP address must be set to the loopback address on each router—and not the physical address of the interface used to reach the router. The new configurations for Routers A, B, C, and D are shown in Example 8-44.

Example 8-44 New Configurations for Routers A, B, C, and D

```
Router A
router bgp 1
 no synchronization
 neighbor 172.16.15.252 remote-as 1
 neighbor 172.16.15.252 update-source Loopback0
 neighbor 172.16.15.253 remote-as 1
 neighbor 172.16.15.253 update-source Loopback0
 neighbor 172.16.15.254 remote-as 1
 neighbor 172.16.15.254 update-source Loopback0
 no auto-summary

Router B
router bgp 1
 no synchronization
 neighbor 172.16.15.251 remote-as 1
 neighbor 172.16.15.251 update-source Loopback0
 neighbor 172.16.15.253 remote-as 1
 neighbor 172.16.15.253 update-source Loopback0
 neighbor 172.16.15.254 remote-as 1
 neighbor 172.16.15.254 update-source Loopback0
 no auto-summary

Router C
router bgp 1
 no synchronization
 neighbor 172.16.15.251 remote-as 1
 neighbor 172.16.15.251 update-source Loopback0
 neighbor 172.16.15.252 remote-as 1
 neighbor 172.16.15.252 update-source Loopback0
 neighbor 172.16.15.254 remote-as 1
 neighbor 172.16.15.254 update-source Loopback0
 no auto-summary

Router D
router bgp 1
```

Figure 8-18 IBGP Does Not Require a Physical Connection Between Neighbors



Instead of using a physical address in the **neighbor** command, you can use a router's loopback address. In Figure 8-18, the best path from Router D to Router C's loopback address is through the direct connection between Routers C and D. If this connection fails, the IGP, in this case OSPF, converges and Router D uses the path through Router A to Router C. If the link between A and C fails, OSPF converges again and Router D uses the path through A and B to reach C. Before you start using loopback addresses in the neighbor commands, you need to be aware of how these addresses are used by IBGP.

The **neighbor** commands used to establish the IBGP connection between Routers C and D use physical addresses. (See Example 8-43.)

Example 8-43 Establishing the IBGP Connection

```
Router C
router bgp 1
 neighbor 172.16.15.130 remote-as 1
Router D
router bgp 1
 neighbor 172.16.15.129 remote-as 1
```

When Router C sends a BGP protocol packet to Router D, the destination IP address is set to the address used in the **neighbor** command on Router C or

Example 8-42 Configuring a Full IBGP Mesh (continued)

```
Router C
router bgp 1
 no synchronization
 neighbor 172.16.15.33 remote-as 1
 neighbor 172.16.15.97 remote-as 1
 neighbor 172.16.15.130 remote-as 1

Router D
router bgp 1
 no synchronization
 neighbor 172.16.15.65 remote-as 1
 neighbor 172.16.15.97 remote-as 1
 neighbor 172.16.15.129 remote-as 1
 neighbor 172.16.15.162 remote-as 2
```

IBGP neighbors do not need a physical connection between them to form a neighbor relationship. All that is required is an IP path between the IBGP neighbors. The configurations in this section used the IP address of a physical interface in the neighbor configuration statements. If the physical interface of the neighbor fails, the BGP connection fails—even though there may be another IP path to another physical interface available. The next section examines a technique to form an IBGP neighbor relationship without specifying a physical interface.

IBGP and Loopback Interfaces

The IBGP configurations on the routers in AS 1 use a physical address in the neighbor commands. What happens if the physical address in the **neighbor** command is no longer available due to a network failure? In Figure 8-18, Router D has three paths to Router C that can be used to form an IBGP connection. If the physical address that is used in the **neighbor** command is unavailable, the other paths will not be used because the route to the physical address is gone.

Example 8-45 Routers A, B, D as Route Reflector Clients

```
Router A
router bgp 1
 no synchronization
 neighbor 172.16.15.253 remote-as 1
 neighbor 172.16.15.253 update-source Loopback0
 no auto-summary

Router B
router bgp 1
 no synchronization
 neighbor 172.16.15.253 remote-as 1
 neighbor 172.16.15.253 update-source Loopback0
 no auto-summary

Router D
router bgp 1
 no synchronization
 neighbor 172.16.15.162 remote-as 2
 neighbor 172.16.15.253 remote-as 1
 neighbor 172.16.15.253 update-source Loopback0
 no auto-summary

Router C
router bgp 1
 no synchronization
 neighbor 172.16.15.251 remote-as 1
 neighbor 172.16.15.251 update-source Loopback0
 neighbor 172.16.15.251 route-reflector-client
 neighbor 172.16.15.252 remote-as 1
 neighbor 172.16.15.252 update-source Loopback0
 neighbor 172.16.15.252 route-reflector-client
 neighbor 172.16.15.254 remote-as 1
 neighbor 172.16.15.254 update-source Loopback0
 neighbor 172.16.15.254 route-reflector-client
 no auto-summary
```

A route reflector advertises, or reflects, prefixes learned from a client to all other clients.

Confederations

Another technique to reduce the number of IBGP connections is called ***confederations***. The AS is divided into two or more subautonomous systems and the

The number of IBGP connections required for a full mesh based on the number of IBGP routers is given by the following formula where n = number of IBGP routers:

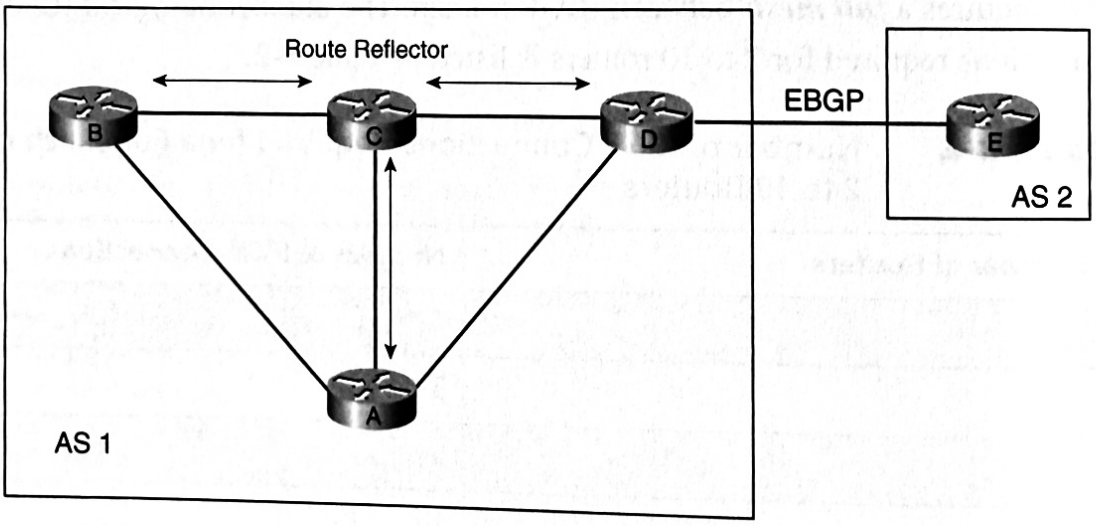
$$\text{Connections} = [(n)(n-1)]/2$$

The number of required IBGP connections grows as the square of the number of routers. For 20 IBGP routers, the number of required connections is 190. The following sections look at techniques—route reflector and confederation—for reducing the number of connections for an IBGP network.

Route Reflector

A ***route reflector*** can reduce the number of required IBGP connections from [(n)(n-1)]/2 to n – 1. In Figure 8-19, Router C is a route reflector for Routers A, B, and D. Routers A, B, and D are referred to as clients of the route reflector and establish only an IBGP connection with the route reflector.

Figure 8-19 Route Reflector Reduces the Required Number of IBGP Connections to n - 1



Routers A, B, and D ***peer*** (form a neighbor connection) only with Router C. The configuration on Router C designates Routers A, B, and D as route reflector clients. (See Example 8-45.) This is the same network used in Figure 8-16, but with Router C as the route reflector.

Example 8-44 New Configurations for Routers A, B, C, and D (continued)

```
no synchronization
neighbor 172.16.15.162 remote-as 2
neighbor 172.16.15.251 remote-as 1
neighbor 172.16.15.251 update-source Loopback0
neighbor 172.16.15.252 remote-as 1
neighbor 172.16.15.252 update-source Loopback0
neighbor 172.16.15.253 remote-as 1
neighbor 172.16.15.253 update-source Loopback0
no auto-summary
```

The command **update-source Loopback0** causes BGP to set the IP source address in BGP protocol packets to the IP address assigned to the loopback interface. The source IP address will now match the destination IP address used by the other routers in the AS. As long as there is an IP path between any two routers, the BGP connection will be active.

Scaling IBGP

IBGP requires a ***full mesh*** between IBGP routers. The number of logical IBGP connections required for 2 to 10 routers is listed in Table 8-2.

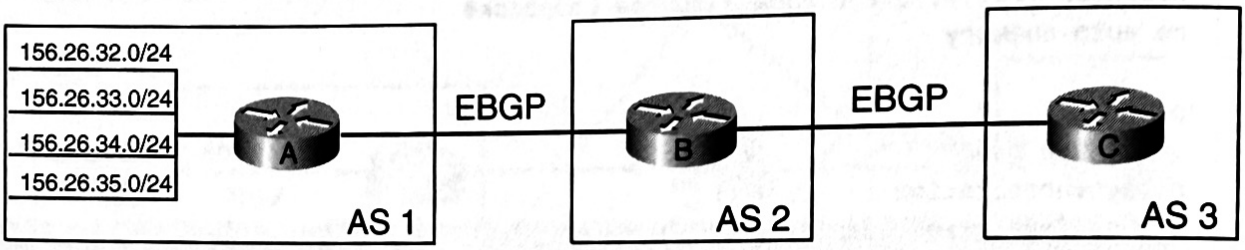
Table 8-2 Number of IBGP Connections Required for a Full Mesh of 2 to 10 Routers

Number of Routers	Number of IBGP Connections
2	1
3	3
4	6
5	10
6	15
7	21
8	28
9	36
10	45

BGP Route Summarization

In Figure 8-22, Router A is advertising four prefixes to Router B through EBGP, and Router B advertises these prefixes to Router C.

Figure 8-22 BGP Route Summarization



The AS path attribute for the 156.26.0.0 prefixes on Router B equals 1, and on Router C the AS path attribute is 2 1. (See Example 8-47.)

Example 8-47 AS_PATH Attributes on Routers B and C

```
B#show ip bgp
BGP table version is 5, local router ID is 172.16.15.252
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/24   172.16.15.1              0             0 1 i
*> 156.26.33.0/24   172.16.15.1              0             0 1 i
*> 156.26.34.0/24   172.16.15.1              0             0 1 i
*> 156.26.35.0/24   172.16.15.1              0             0 1 I

C#show ip bgp
BGP table version is 5, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/24   172.16.15.97                          0 2 1 i
*> 156.26.33.0/24   172.16.15.97                          0 2 1 i
*> 156.26.34.0/24   172.16.15.97                          0 2 1 i
*> 156.26.35.0/24   172.16.15.97                          0 2 1 i
```

The reason for pointing out the value of the AS_PATH attribute is so you can see what happens to this attribute when you create a BGP summary address. A BGP

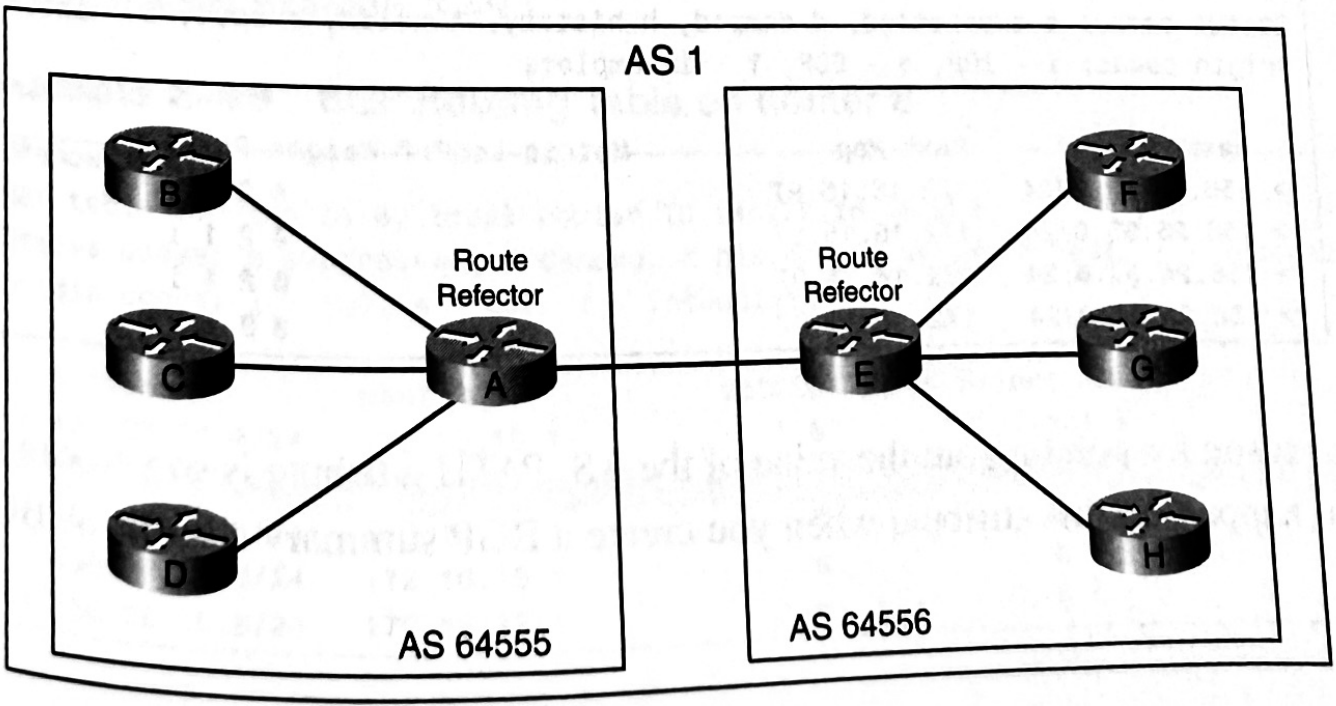
Example 8-46 Configuring a Confederation (continued)

```
Router C
router bgp 64555
 no synchronization
 bgp confederation peers 64556
 neighbor 172.16.15.33 remote-as 64556
 neighbor 172.16.15.252 remote-as 64555
 neighbor 172.16.15.252 update-source Loopback0
 no auto-summary

Router D
router bgp 64556
 no synchronization
 bgp confederation identifier 1
 neighbor 172.16.15.162 remote-as 2
 neighbor 172.16.15.251 remote-as 64556
 neighbor 172.16.15.251 update-source Loopback0
 no auto-summary
```

The private sub-AS number is used to configure the BGP process and neighbor statements. The **bgp confederation peers** command is used on a router having a neighbor in a different sub-AS. Router E should see Router D in AS 1 because this is the assigned AS number. The command **bgp confederation identifier** sets the AS number advertised to an external BGP peer.

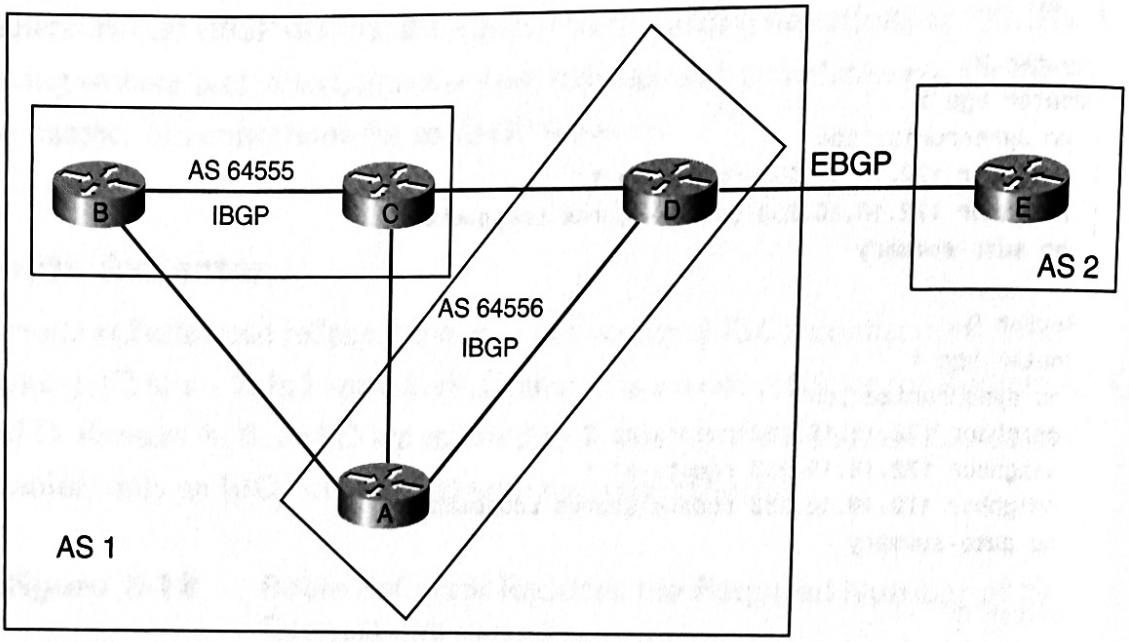
Route reflectors and confederations can be combined to further reduce the number of IBGP connections. The AS in Figure 8-21 uses two subautonomous systems, and two route reflectors to reduce the number of IBGP connections from 28 to 7.

Figure 8-21 Route Reflectors and Confederations Used Together



subautonomous systems are usually assigned a private AS number in the range of 64,512 to 65,534. In Figure 8-20, Routers B and C are in sub-AS 64555, and Routers A and D are in sub-AS 64556.

Figure 8-20 BGP Confederation



Configuring a confederation is more complicated than configuring a route reflector as seen in Example 8-46.

Example 8-46 Configuring a Confederation

```
Router A
router bgp 64556
 no synchronization
 bgp confederation peers 64555
 neighbor 172.16.15.34 remote-as 64555
 neighbor 172.16.15.254 remote-as 64556
 neighbor 172.16.15.254 update-source Loopback0
 no auto-summary

Router B
router bgp 64555
 no synchronization
 neighbor 172.16.15.253 remote-as 64555
 neighbor 172.16.15.253 update-source Loopback0
 no auto-summary
```

within the aggregate address block. To prevent the specific prefixes from being advertised, use the **summary-only** option with the **aggregate-address**command. (See Example 8-51.)

Example 8-51 Suppressing the More Specific Prefixes

```
Router B
router bgp 2
 aggregate-address 156.26.32.0 255.255.252.0 summary-only
 neighbor 172.16.15.1 remote-as 1
 neighbor 172.16.15.98 remote-as 3
```

The BGP routing table on Router B now shows the specific routes as being suppressed. (See Example 8-52.) A suppressed route is a route that is not advertised by BGP.

Example 8-52 Router B BGP Routing Table

```
B#show ip bgp
BGP table version is 10, local router ID is 172.16.15.252
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
s> 156.26.32.0/24   172.16.15.1              0             0 1 i
*> 156.26.32.0/22   0.0.0.0                                32768 i
s> 156.26.33.0/24   172.16.15.1              0             0 1 i
s> 156.26.34.0/24   172.16.15.1              0             0 1 i
s> 156.26.35.0/24   172.16.15.1              0             0 1 i
```

Router C should have only the aggregate prefix in its BGP routing table. (See Example 8-53.)

Example 8-53 Aggregate Prefix in Router C's BGP Routing Table

```
C#show ip bgp
BGP table version is 10, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/22   172.16.15.97                           0 2 i
```

The aggregate prefix has a next hop of 0.0.0.0 because the prefix originates on Router B and is referred to as a *locally originated route*. Also, the AS path attribute for the aggregate prefix has been initialized to an empty string. Router B advertises the aggregate prefix to both Routers A and C. (See Example 8-50.)

Example 8-50 Summarized Route as Seen by Router A

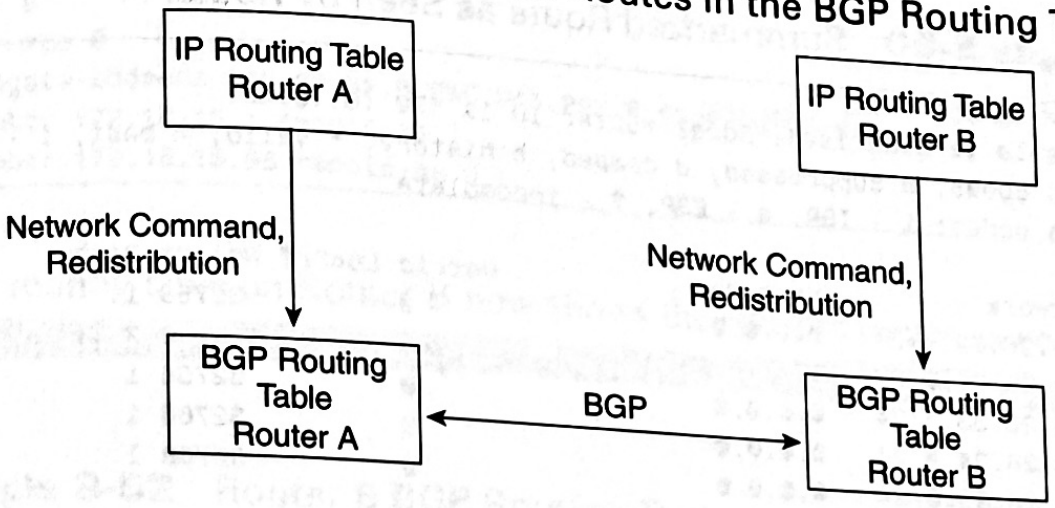
```
A#show ip bgp
BGP table version is 6, local router ID is 172.16.15.251
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/24   0.0.0.0                  0         32768 i
*> 156.26.32.0/22   172.16.15.2                            0 2 i
*> 156.26.33.0/24   0.0.0.0                  0         32768 i
*> 156.26.34.0/24   0.0.0.0                  0         32768 i
*> 156.26.35.0/24   0.0.0.0                  0         32768 I

C#show ip bgp
BGP table version is 6, local router ID is 172.16.15.253
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/24   172.16.15.97                           0 2 1 i
*> 156.26.32.0/22   172.16.15.97                           0 2 i
*> 156.26.33.0/24   172.16.15.97                           0 2 1 i
*> 156.26.34.0/24   172.16.15.97                           0 2 1 i
*> 156.26.35.0/24   172.16.15.97                           0 2 1 i
```

The summary address created by Router B does not include the path information of the specific prefixes that are part of the summary. The specific prefixes originated in AS 1, but the only AS number in the AS path attribute for the prefix is 2.

The AS path attribute for the aggregate route on Routers A and C has a value of 2, indicating this route originated in AS 2. The four specific prefixes are in the BGP routing table on Router A because these prefixes are originated by Router A. Why are the four specific prefixes in the routing table on Router C? When a summary address was created with EIGRP, OSPF, or IS-IS, the specific prefixes making up the summary were not advertised—only the summary address was advertised. By default, BGP advertises the aggregate address and all the prefixes contained

summary address is configured using the command **aggregate-address**under the BGP router configuration. BGP can only summarize the addresses in the BGP routing table. (See Figure 8-23.)

Figure 8-23 BGP Can Summarize Routes in the BGP Routing Table



A summary address is created on Router B and includes only the four subnets advertised by Router A. (See Example 8-48.)

Example 8-48 Summary Address on Router B

```
Router B
router bgp 2
 aggregate-address 156.26.32.0 255.255.252.0
 neighbor 172.16.15.1 remote-as 1
 neighbor 172.16.15.98 remote-as 3
```

The BGP routing table on Router B contains the aggregate and specific prefixes of 156.26.32.0. (See Example 8-49.)

Example 8-49 BGP Routing Table on Router B

```
B#show ip bgp
BGP table version is 6, local router ID is 172.16.15.252
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 156.26.32.0/24   172.16.15.1              0             0 1 i
*> 156.26.32.0/22   0.0.0.0                                32768 i
*> 156.26.33.0/24   172.16.15.1              0             0 1 i
*> 156.26.34.0/24   172.16.15.1              0             0 1 i
*> 156.26.35.0/24   172.16.15.1              0             0 1 i
```

9. Name all the ways for installing a prefix in the BGP routing table.
10. What is the first thing that BGP checks to determine if a prefix is accessible?
11. What is synchronization?
12. Name two methods for reducing the number of IBGP connections.
13. Why does IBGP require a full mesh?
14. What is the major difference between BGP and IGP route summarization?
15. Why is BGP a better choice for Internet routing than the IGPs?
16. What are the four general types of BGP attributes?

References

Use the following references to learn more about the concepts described in this chapter:

- Doyle, J. and J.D. Carroll. 2001. *Routing TCP/IP,* Volume II. Indianapolis: Cisco Press.
- Halabi, Sam. *Internet Routing Architectures*, Second Edition. 2000. Indianapolis: Cisco Press.
- Parkhurst, William. *Cisco BGP-4 Command and Configuration Handbook*. 2001. Indianapolis: Cisco Press.
- Rekhter, Y. and T. Li. March 1995. RFC 1771, A Border Gateway Protocol 4 (BGP-4).
- Cisco Documentation, Cisco.com.

Summary

BGP is an exterior gateway protocol (EGP) used to route between the autonomous systems that make up the Internet. The best path algorithms of the interior routing protocols (RIP, IGRP, EIGRP, OSPF, and IS-IS) do not scale to the number of IP prefixes required for Internet routing. The BGP best path algorithm is a simple and elegant mechanism that can handle well over 100,000 routes.

In addition to advertising network prefixes and subnet masks, BGP uses a set of attributes in the best path decision process. These attributes are: AS_PATH, WEIGHT, LOCAL_PREF, ORIGIN, and MULTI_EXIT_DISC. The BGP attributes give BGP the flexibility to make a best path decision based on any, or all, of the attributes.

Although the BGP decision algorithm is relatively simple, the configuration and mastery of BGP is not. This chapter covered the basic concepts, operation, and configuration of BGP, and has provided a foundation for exploring BGP further.

Chapter Review Questions

You can find the answers to these questions in Appendix A.

1. Describe four differences between the operation of IBGP and EBGP.
2. What is the purpose of the AS_PATH attribute?
3. What is the purpose of the WEIGHT attribute?
4. What is the scope of the WEIGHT attribute?
5. What is the purpose of the LOCAL_PREF attribute?
6. What is the scope of the LOCAL_PREF attribute?
7. What is the purpose of the metric or MULTI_EXIT_DISC attribute?
8. What is the order of preference for the BGP attributes AS_PATH, LOCAL_PREF, MED, and WEIGHT?

Additional options are available for the manipulation of BGP aggregate addresses. You can choose to suppress only a subset of the specific prefixes. You can also, on a neighbor-by-neighbor basis, unsuppress routes that have been suppressed using the **aggregate-address**command. In fact, there are more than 100 BGP configuration commands that can be used to fine-tune a BGP configuration. For specific information regarding these additional commands, see the references at the end of this chapter.

BGP Decision Process

For every prefix a BGP router receives, the first thing checked is the NEXT_HOP attribute. If the next hop is not accessible (not in the IP routing table), the prefix is rejected. For routes received from an IBGP peer, the next check after the next hop check is to see if the routes are synchronized. Unsynchronized routes are rejected only if synchronization is enabled. The following steps in the BGP decision process apply only if there is more than one route to a particular prefix:

1. Prefer the path with the highest weight
2. Prefer the path with the highest local preference
3. Prefer locally originated routes (Next Hop = 0.0.0.0)
4. Prefer the path with the shortest AS path
5. Prefer the path with the lowest origin type: IGP is lower than EGP, and EGP is lower than INCOMPLETE
6. Prefer the path with the lowest metric or MED
7. Prefer EBGP routes over IBGP routes
8. Prefer the lowest IGP metric to the next hop
9. Prefer the path from the router with the lowest router ID