Configuring BGP on Cisco Routers

Volume 2

Version 3.2

Student Guide

CLS Production Servies: 12.29.05

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica
Croatia • Cyprus • Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece
Hong Kong SAR • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia
Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Puerto Rico • Romania
Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden • Switzerland
Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED "AS IS." CISCO MAKES AND YOU RECEIVE NO WARRANTIES IN CONNECTION WITH THE CONTENT PROVIDED HEREUNDER, EXPRESS, IMPLIED, STATUTORY OR IN ANY OTHER PROVISION OF THIS CONTENT OR COMMUNICATION BETWEEN CISCO AND YOU. CISCO SPECIFICALLY DISCLAIMS ALL IMPLIED WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE, OR ARISING FROM A COURSE OF DEALING, USAGE OR TRADE PRACTICE. This learning product may contain early release content, and while Cisco believes it to be accurate, it falls subject to the disclaimer above.

Students, this letter describes important course evaluation access information!

Welcome to Cisco Systems Learning. Through the Cisco Learning Partner Program, Cisco Systems is committed to bringing you the highest-quality training in the industry. Cisco learning products are designed to advance your professional goals and give you the expertise you need to build and maintain strategic networks.

Cisco relies on customer feedback to guide business decisions; therefore, your valuable input will help shape future Cisco course curricula, products, and training offerings. We would appreciate a few minutes of your time to complete a brief Cisco online course evaluation of your instructor and the course materials in this student kit. On the final day of class, your instructor will provide you with a URL directing you to a short post-course evaluation. If there is no Internet access in the classroom, please complete the evaluation within the next 48 hours or as soon as you can access the web.

On behalf of Cisco, thank you for choosing Cisco Learning Partners for your Internet technology training.

Sincerely,

Cisco Systems Learning

Table of Contents

Volume 2

Module 4

Route Selection Using Attributes

Overview

Routes learned via Border Gateway Protocol (BGP) have properties that are associated with them that aid a router in determining the best route to a destination when multiple paths to that particular destination exist. These properties are referred to as BGP attributes. This module introduces the role of BGP attributes, and explains how their presence influences route selection in BGP. Understanding how BGP attributes influence route selection is required for the design of robust networks.

This module provides advanced information on how to connect Internet customers to multiple service providers. It includes an in-depth description of BGP attributes that are used in route selection, including weight, local preference, autonomous system (AS)-path prepending, multi-exit discriminator (MED), and BGP communities.

Module Objectives

Upon completing this module, you will be able to complete the correct BGP configuration to successfully connect the customer network to the Internet in a network scenario in which multiple connections must be implemented This ability includes being able to meet these objectives:

- Successfully configure BGP to influence route selection by using the weight attribute in a customer scenario in which you must support multiple ISP connections
- Use the local preference attribute to influence route selection in a customer scenario in which you must support multiple ISP connections
- Use AS-path prepending to influence the return path that is selected by the neighboring autonomous systems in a customer scenario in which you must support multiple ISP connections
- Use the MED attribute to influence route selection in a customer scenario in which you must support multiple ISP connections
- Use BGP community attributes to influence route selection in a customer scenario in which you must support multiple ISP connections

Lesson 1

Influencing BGP Route Selection with Weights

Overview

When connections to multiple providers are required, it is important that Border Gateway Protocol (BGP) select the optimum route for traffic to use. The optimum, or best, route may not be what the network designer intended based on design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides many tools for administrators to influence route selection. One of these tools is the weight attribute.

This lesson discusses how to influence BGP route selection by setting the weight attribute of incoming BGP routes. Two methods that are used to set the weight attribute, default weight and route-maps, are discussed in this lesson. This lesson also explains how to monitor the BGP table to verify correct weight configuration and properly influence path selection.

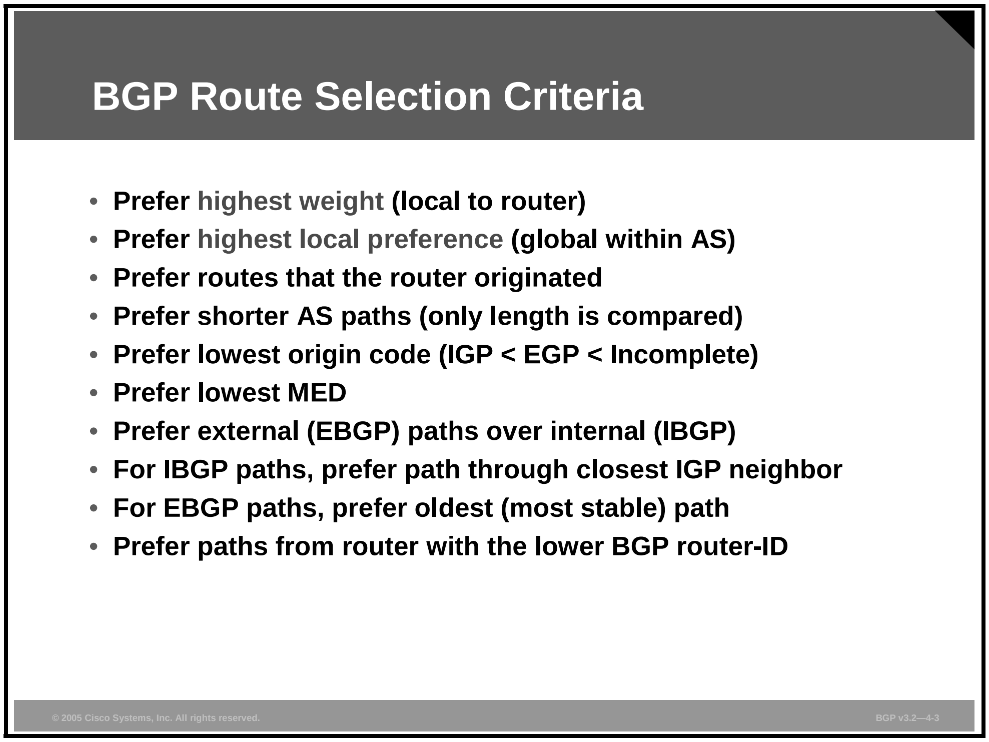
Objectives

Upon completing this lesson, you will be able to successfully configure BGP to influence route selection by using the weight attribute in a customer scenario in which you must support multiple ISP connections. This ability includes being able to meet these objectives:

- List BGP route selection criteria for best-path route selection
- Describe the use of BGP weights to influence the BGP route selection process
- Influence the BGP route selection process by configuring per-neighbor weights
- Influence the BGP route selection process by configuring BGP weights with route-maps
- Identify the Cisco IOS commands that are required to monitor BGP route selection and weights
- Summarize BGP route selection and filtering tools

BGP Route Selection Criteria

This topic lists the criteria that are used by BGP for best-path route selection.



BGP Route Selection Criteria

- Prefer highest weight (local to router)
- Prefer highest local preference (global within AS)
- Prefer routes that the router originated
- Prefer shorter AS paths (only length is compared)
- Prefer lowest origin code (IGP < EGP < Incomplete)
- Prefer lowest MED
- Prefer external (EBGP) paths over internal (IBGP)
- For IBGP paths, prefer path through closest IGP neighbor
- For EBGP paths, prefer oldest (most stable) path
- Prefer paths from router with the lower BGP router-ID

© 2005 Cisco Systems, Inc. All rights reserved.

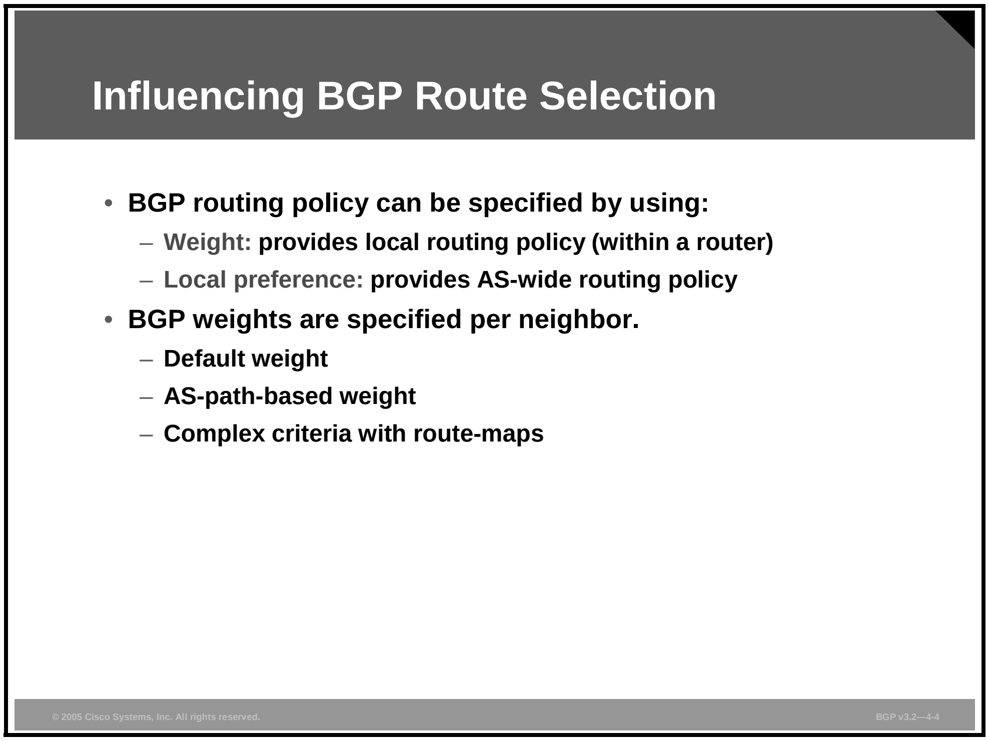
BGP v3.2—4-3

BGP route selection criteria take the weight parameter into consideration first. If a router has two alternative paths to the same destination, and their weight values are different, BGP selects the route with the highest weight value as the best. Only when the two alternatives have equal weight is the next criterion, local preference, checked.

A high local preference value is preferred over a low value. Only when the two alternatives have an equal local preference is the next criterion checked.

Influencing BGP Route Selection

This topic describes how network administrators can use BGP weights to influence the BGP route selection process.



Influencing BGP Route Selection

- **BGP routing policy can be specified by using:**
  - **Weight: provides local routing policy (within a router)**
  - **Local preference: provides AS-wide routing policy**
- **BGP weights are specified per neighbor.**
  - **Default weight**
  - **AS-path-based weight**
  - **Complex criteria with route-maps**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-4

The weight attribute is local to a single router only. The weight value is never propagated by the BGP protocol, and this value constitutes a routing policy local to the router.

Local preference is assigned to a route as an attribute. This attribute is carried with the route on all internal BGP sessions. In this situation all other BGP-speaking routers within the autonomous system (AS) receive the same information. Normally, a router assigns a local preference to a route that is received on an external BGP session before it is accepted and entered in the BGP table of the border router. Routers propagate the local preference attribute on internal BGP sessions only. This policy constitutes a routing policy for the entire AS.

The router can assign the weight attribute to a route in two ways:

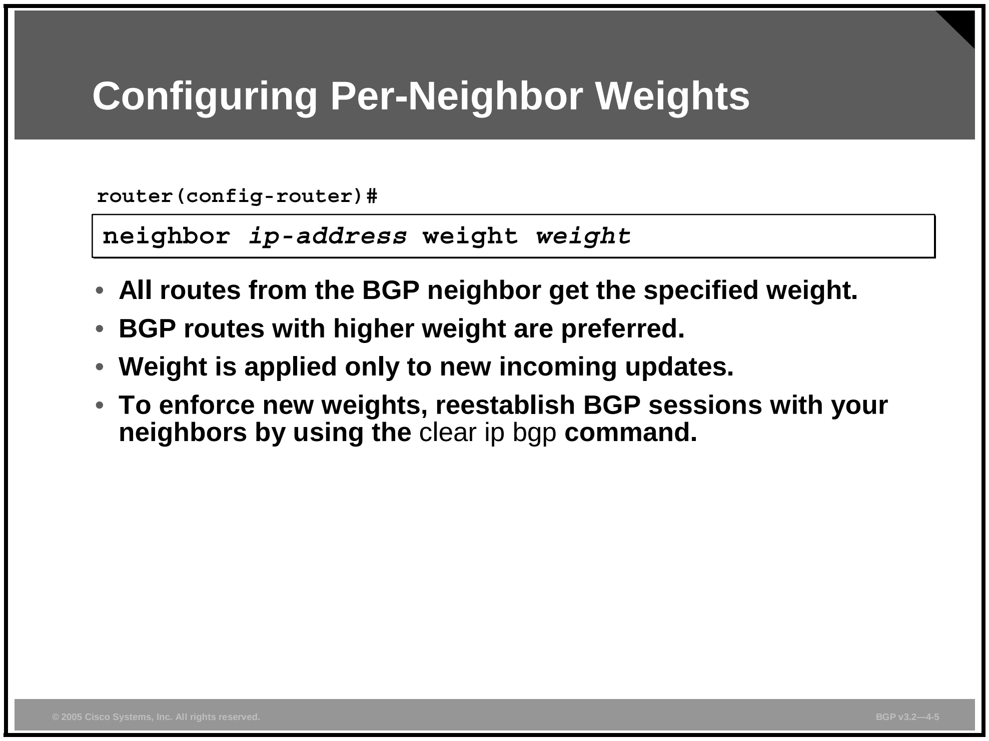
- All routes that are received from a specific neighbor can be assigned a default weight value. This weight value indicates that the neighbor is preferred over the other neighbors.
- A route-map that is applied on incoming routes from a neighbor can be used to select some routes and assign them weight values. Remember that a route-map also acts as a filter and will silently drop the routes that are not permitted by any statement in the route-map.

If configured, the default weight assignment on routes that are received from a neighbor is applied first. All routes that are received from the neighbor are assigned a weight value as defined by the default weight.

When a route-map is applied, it is configured on the router. The route-map can be arbitrarily complex and select routes based on various selection criteria, such as a network number or AS path. The selected routes can have some attributes altered. The route-map can set the weight values of permitted routes. Selection can be done in several route-map statements, giving the opportunity to assign a certain weight value to some routes and another weight value to others. A route-map can also completely filter out routes.

Configuring Per-Neighbor Weights

This topic describes how to influence the BGP route selection process by configuring per-neighbor weights.



Configuring Per-Neighbor Weights

```
router(config-router)#
neighbor ip-address weight weight
```

- **All routes from the BGP neighbor get the specified weight.**
- **BGP routes with higher weight are preferred.**
- **Weight is applied only to new incoming updates.**
- **To enforce new weights, reestablish BGP sessions with your neighbors by using the** clear ip bgp **command.**

BGP v3.2—4-5

neighbor weight

To assign a weight to a neighbor connection, use the **neighbor weight** router configuration command.

- **neighbor** {*ip-address* | *peer-group-name*} **weight** *weight*

To remove a weight assignment, use the **no** form of this command.

- **no neighbor** {*ip-address* | *peer-group-name*} **weight** *weight*

Syntax Description

Parameter	Description
ip-address	IP address of neighbor.
peer-group-name	Name of a BGP peer group.
weight	Weight to assign. Acceptable values are 0 to 65535.

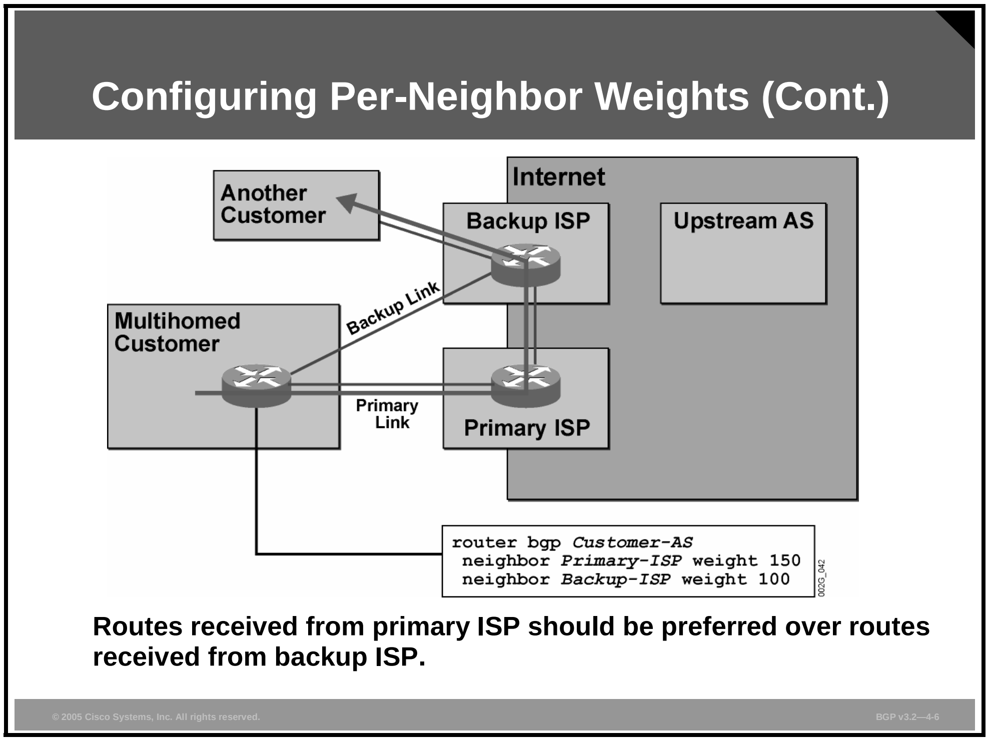
All routes that are received from the neighbor after the configuration line is in place are assigned the weight value. To make sure that all routes from the neighbor receive the new weight value, you can restart the BGP session, forcing the neighbor to resend all routes.

If no weight value is specified, the default value of 0 is applied.

Restarting BGP sessions might be necessary after making a configuration change in the routing policy. The configuration change itself will not alter the already-received routes. The **clear ip bgp** EXEC command tears down the BGP session, and the session automatically restarts.

Example: Configuring Per-Neighbor Weights

In this example, the multihomed customer would like to use the primary link to the primary Internet service provider (ISP) for all destinations.



Configuring Per-Neighbor Weights (Cont.)

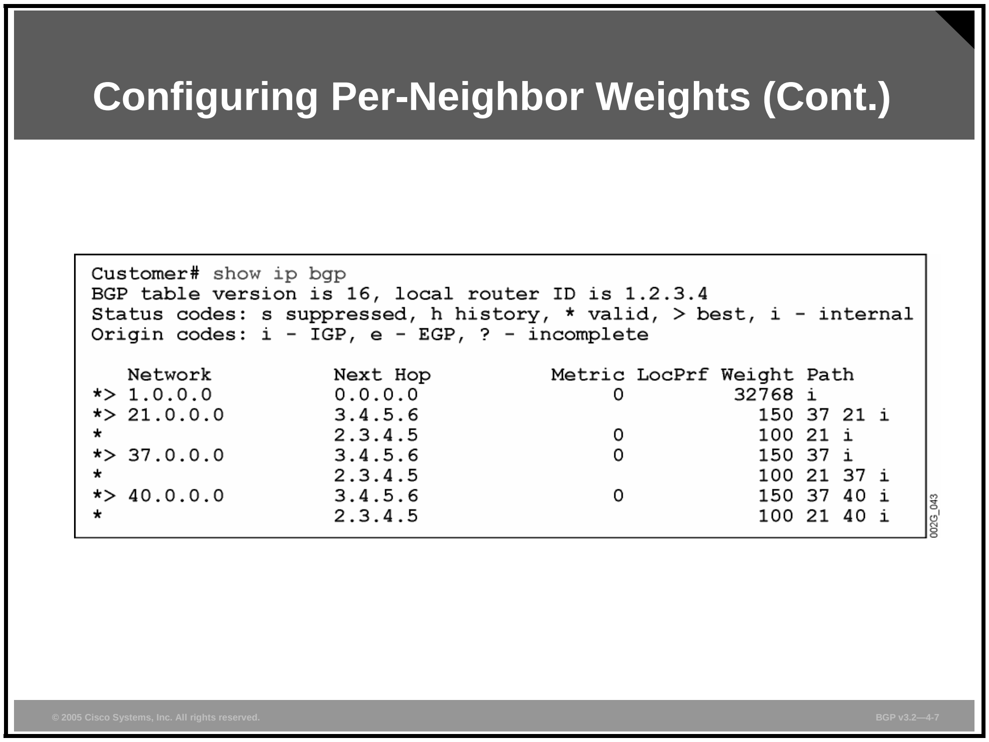
```
router bgp Customer-AS
 neighbor Primary-ISP weight 150
 neighbor Backup-ISP weight 100
```

Routes received from primary ISP should be preferred over routes received from backup ISP.

The weight is configured by the customer on both BGP sessions, giving a higher weight to the routes that are received from the primary ISP compared to those that are received from the backup ISP.

Any time that the multihomed customer receives routing information about the same IP network number from both the ISPs, the customer compares the weights assigned to the routes. Those received from the primary ISP will always win this comparison. The multihomed customer sends the outgoing IP packets to the destination network via the primary ISP regardless of the other BGP attributes that are assigned to both alternatives.

Consequently, the other customer that is directly connected to the backup ISP will also be reached via the primary ISP.



Configuring Per-Neighbor Weights (Cont.)

```
Customer# show ip bgp
BGP table version is 16, local router ID is 1.2.3.4
Status codes: s suppressed, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric LocPrf Weight Path
*> 1.0.0.0          0.0.0.0                0         32768 i
*> 21.0.0.0         3.4.5.6                            150 37 21 i
*                   2.3.4.5                0           100 21 i
*> 37.0.0.0         3.4.5.6                0           150 37 i
*                   2.3.4.5                            100 21 37 i
*> 40.0.0.0         3.4.5.6                0           150 37 40 i
*                   2.3.4.5                            100 21 40 i
```

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-7

In this example, the multihomed customer has received routes to three different class A networks outside of its own AS (network 21.0.0.0/8, network 37.0.0.0/8, and network 40.0.0.0/8). The customer has received all three routes from both the primary ISP and the backup ISP.

When the routes were received from the primary ISP, the weight value 150 was assigned to each of the routes. When the routes were received from the backup ISP, the weight value 100 was assigned to each of the routes.

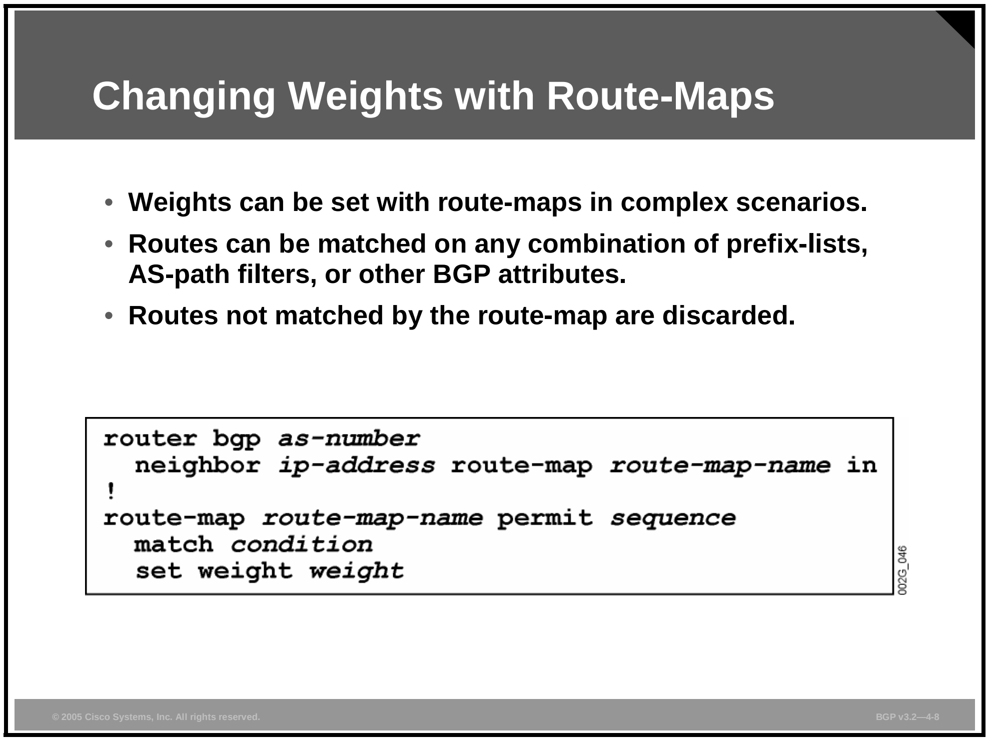
The customer router now makes the route selection. It has two alternative paths for each destination network. For each of them, the router selects the path via the primary ISP as the best. It makes this selection regardless of other BGP attributes, such as AS-path length.

The network 21.0.0.0/8 is reached via the primary ISP although it is actually a network in the AS of the backup ISP (AS 21).

The class A network 1.0.0.0/8 in this example is injected into the BGP table by this router. By default, locally sourced routes are assigned a weight of 32768.

Changing Weights with Route-Maps

This topic describes how to influence the BGP route selection process by configuring BGP weights with route-maps.



Changing Weights with Route-Maps

- **Weights can be set with route-maps in complex scenarios.**
- **Routes can be matched on any combination of prefix-lists, AS-path filters, or other BGP attributes.**
- **Routes not matched by the route-map are discarded.**

```
router bgp as-number
  neighbor ip-address route-map route-map-name in
!
route-map route-map-name permit sequence
  match condition
  set weight weight
```

The route-map is a powerful tool to select and alter routing information. When a route-map is applied to incoming information from a BGP neighbor, each received update is examined as it passes through the route-map. Statements in the route-map are executed in the order that is specified by their sequence numbers.

The first statement in the route-map for which all the match clauses indicate a match is the one that is used. If the route-map says “permit,” the set clauses are applied to the route, the route is accepted, and the weight is changed.

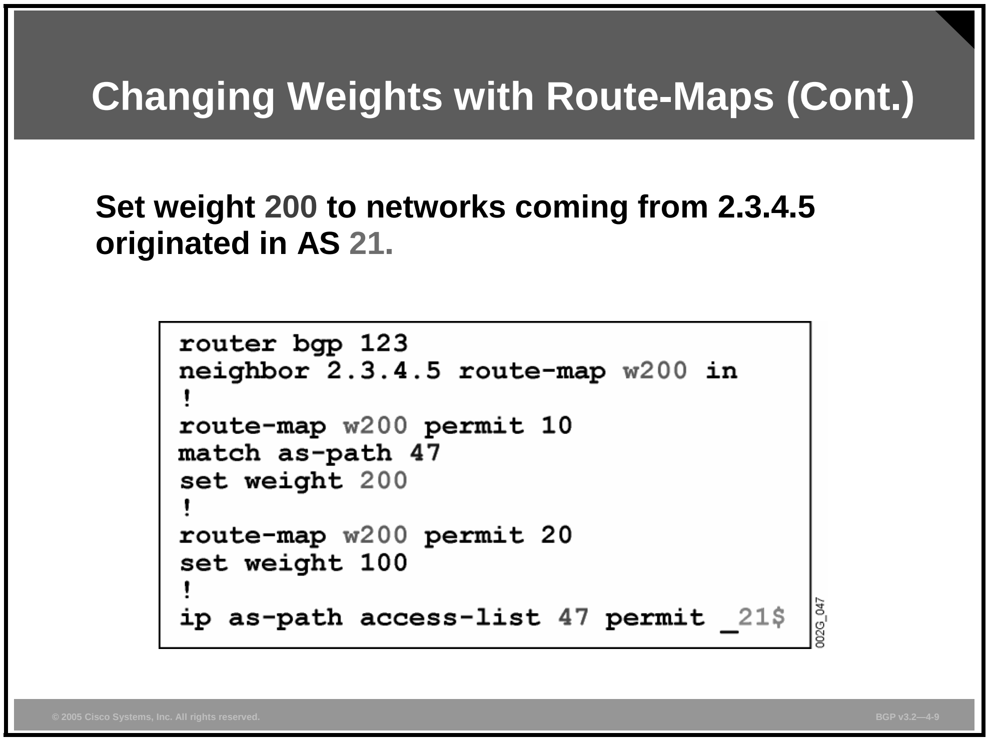
Match clauses can be arbitrarily complex. One of them can refer to an AS-path access-list that does matching on AS paths. Another can refer to a prefix-list that does matching on the announced network number. Only when all configured match clauses are evaluated is the route-map statement used and its result, permit or deny, applied.

If a received route is not matched by any of the route-map statements, and the end of the route-map is reached, the route-map logic has an “implicit deny” rule. This rule means that if no statement selects a route, the route is discarded.

If the “implicit deny” rule is not desired, an “explicit permit all” at the end of the route-map can overrule it. To ensure that such a route-map statement is the last statement, you should assign it a very high sequence number. It should not have any match clause at all. The lack of a match clause means “match all.” By not configuring any set clause, you can ensure that no attributes are altered by the statement.

Example: Changing Weights with Route-Maps

This example shows a route-map that sets the weight value to each route that it receives from a neighbor.



Changing Weights with Route-Maps (Cont.)

Set weight 200 to networks coming from 2.3.4.5 originated in AS 21.

```
router bgp 123
neighbor 2.3.4.5 route-map w200 in
!
route-map w200 permit 10
match as-path 47
set weight 200
!
route-map w200 permit 20
set weight 100
!
ip as-path access-list 47 permit _21$
```

002G_047

BGP v3.2—4-9

All received routes have their AS paths checked against the AS-path access-list 47. Those routes with an AS path that indicates that they originated in AS 21 are permitted by the AS-path access-list 47 as referenced by route-map statement number 10. Routes that are permitted and selected by route-map statement number 10 in the w200 route-map will have their weight set to 200 as indicated by the set clause in the route-map.

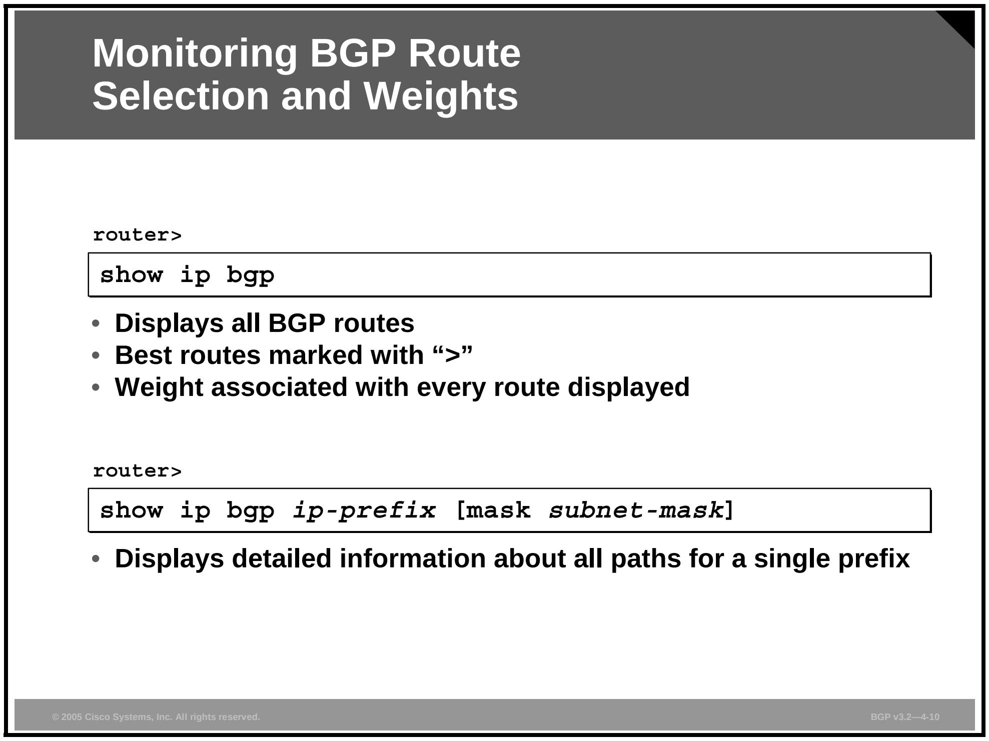
The routes that are not originated in AS 21 (routes that are not permitted by AS-path access-list 47) are then tested by route-map statement number 20. This statement does not include a match clause, indicating that all routes are matched. Therefore, all routes that are not matched by route-map statement 10 are matched by route-map statement 20. The route-map has been configured with an “explicit permit all” statement at the end of the route-map.

Routes that are matched by route-map statement 20 have their weight set to 100. The result is that the routes that originated in AS 21 are accepted by the router and assigned the weight 200. All others are accepted and assigned the weight 100. No route is discarded by this route-map.

Note Specifying weights with filter-lists is no longer supported in Cisco IOS Software Release 12.1, and the command has already been removed from Cisco IOS Software Release 12.1T. These releases use an incoming route-map, where you match an AS path with the **match as-path** command and set the weight with the **set weight** command. When you are using a route-map as a replacement for the filter-list with the **weight** option, make sure that specifying a “permit” entry in the route-map without an associated match condition does not filter all other routes. Using route-maps as a weight-setting mechanism is explained later in this lesson.

Monitoring BGP Route Selection and Weights

This topic lists the Cisco IOS commands that are required to monitor BGP route selection and weights.



Monitoring BGP Route Selection and Weights

```
router>
show ip bgp
```

- **Displays all BGP routes**
- **Best routes marked with “>”**
- **Weight associated with every route displayed**

```
router>
show ip bgp ip-prefix [mask subnet-mask]
```

- **Displays detailed information about all paths for a single prefix**

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-10

show ip bgp

To display entries in the BGP routing table, use the **show ip bgp** EXEC command.

- **show ip bgp** [*network*] [*network-mask*] [**longer-prefixes**]

Syntax Description

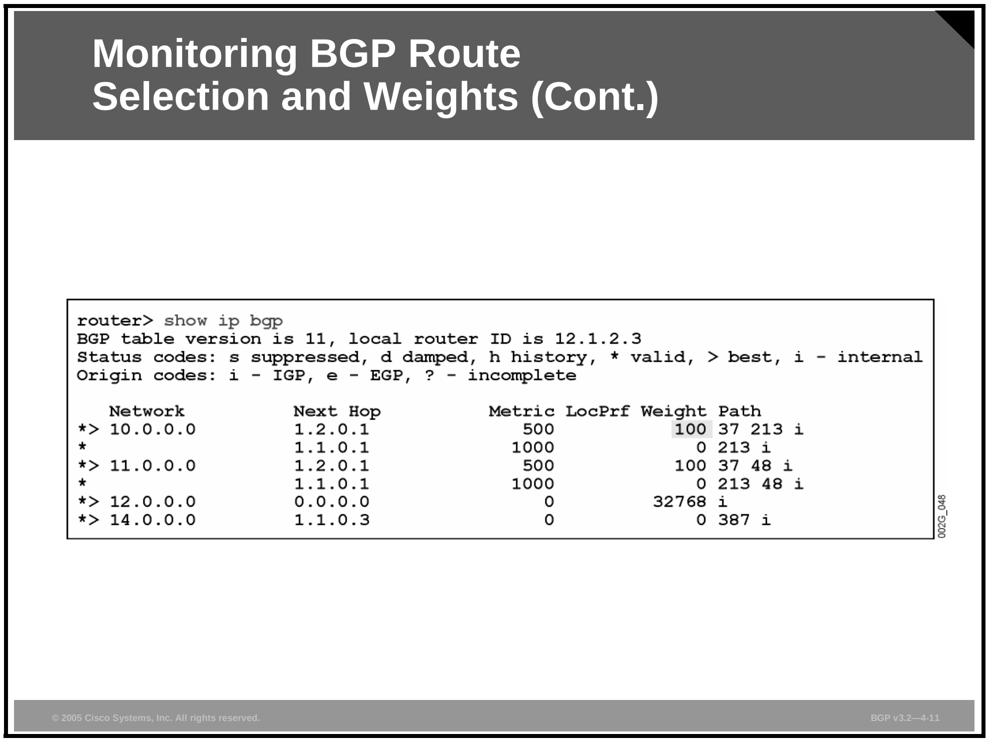
Parameter	Description
network	(Optional) Network number that is entered to display a particular network in the BGP routing table
network-mask	(Optional) Displays all BGP routes that match the address-mask pair
longer-prefixes	(Optional) Displays a route and its more specific routes

Without any argument, the **show ip bgp** command displays the entire BGP table. The routes that are selected as the best are indicated by the greater-than (“>”) character.

To get more detailed information about routes to a specific destination network, you can use the network number, and optionally the subnet mask, as an argument on the command line. These additions display more detailed information about that specific network.

Example: Monitoring BGP Route Selection and Weights

The figure shows all routes in the BGP table.



Monitoring BGP Route Selection and Weights (Cont.)

```
router> show ip bgp
BGP table version is 11, local router ID is 12.1.2.3
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
*> 10.0.0.0         1.2.0.1              500           100 37 213 i
*                   1.1.0.1             1000             0 213 i
*> 11.0.0.0         1.2.0.1              500           100 37 48 i
*                   1.1.0.1             1000             0 213 48 i
*> 12.0.0.0         0.0.0.0                0         32768 i
*> 14.0.0.0         1.1.0.3                0             0 387 i
```

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-11

The **show ip bgp** command provides a printout of all routes in the BGP table. Each route is displayed on one line. This one-line limitation means that more detailed information about the route cannot be displayed because of lack of space.

The network number is displayed, and if the subnet mask differs from the natural mask, the prefix length is indicated. The BGP next-hop attribute, multi-exit discriminator (MED), local preference, weight, AS-path, and origin code are displayed on the line. Local preference is displayed only if it is not the default value.

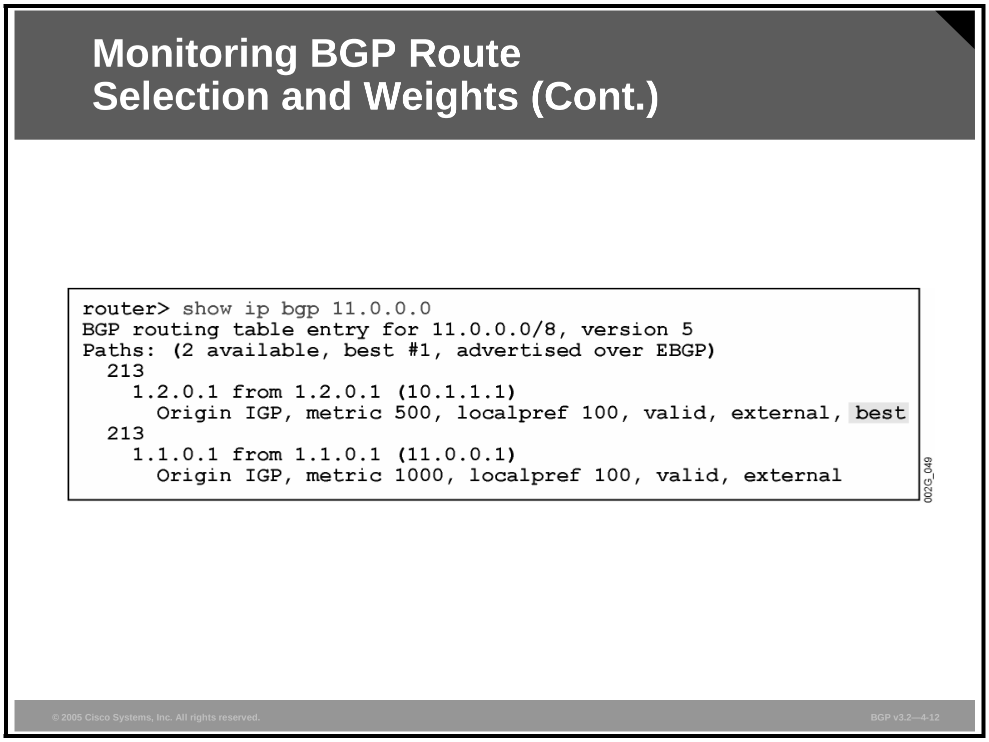
The printout is sorted in network number order. If there is more than one route to the same network, the network number is printed on the first line only. The other routes to the same network have their network field left blank on the output.

Routes that are selected as the best to reach a certain destination network are indicated by the greater than (“>”) character.

In this example, weight has been used to prefer routes received from the neighbor in AS 37. Therefore, although the AS path is shorter via AS 213, the class A network 10.0.0.0/8 is reached via AS 37 (because the weight is higher).

Network 12.0.0.0 is local with the self-originated path selected as best resulting in the next-hop address of 0.0.0.0 and a weight of 32768.

Information about network 14.0.0.0/8 is received only from the neighbor in AS 387. Because there is no alternative, the route is selected as best.



Monitoring BGP Route Selection and Weights (Cont.)

```
router> show ip bgp 11.0.0.0
BGP routing table entry for 11.0.0.0/8, version 5
Paths: (2 available, best #1, advertised over EBGP)
  213
    1.2.0.1 from 1.2.0.1 (10.1.1.1)
      Origin IGP, metric 500, localpref 100, valid, external, best
  213
    1.1.0.1 from 1.1.0.1 (11.0.0.1)
      Origin IGP, metric 1000, localpref 100, valid, external
```

The **show ip bgp** command with a network number as an argument displays more detailed information about that network only. First, a short summary that indicates the network number and prefix length is displayed, along with the table version number for this route. The next line indicates how many alternative routes have been received and which one of them has been selected by the router as the best.

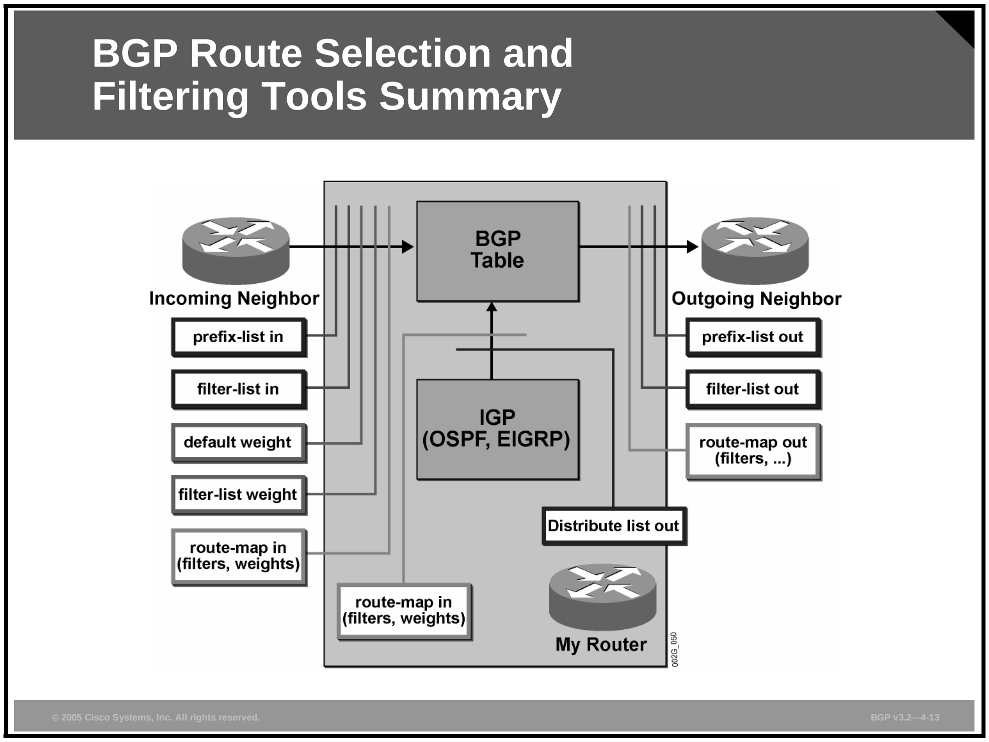
Next, there are a couple of lines for each of the received routes to reach the network. For each of the routes, all attributes are displayed. The one selected as the best also has the word "best" displayed.

In this example, there are two alternatives to reach network 11.0.0.0/8. Each of them is received from different neighbors in AS 213. The network 11.0.0.0 is created in AS 213.

The route selection mechanism has selected the first route that is listed as the best. It was chosen because the MED (metric) value is lower.

BGP Route Selection and Filtering Tools Summary

This topic presents a summary of all BGP filtering tools in the order in which they are applied.



BGP Route Selection and Filtering Tools Summary

BGP
Table

Incoming Neighbor

prefix-list in

filter-list in

default weight

filter-list weight

route-map in
(filters, weights)

IGP
(OSPF, EIGRP)

route-map in
(filters, weights)

Distribute list out

My Router

Outgoing Neighbor

prefix-list out

filter-list out

route-map out
(filters, ...)

002G_050

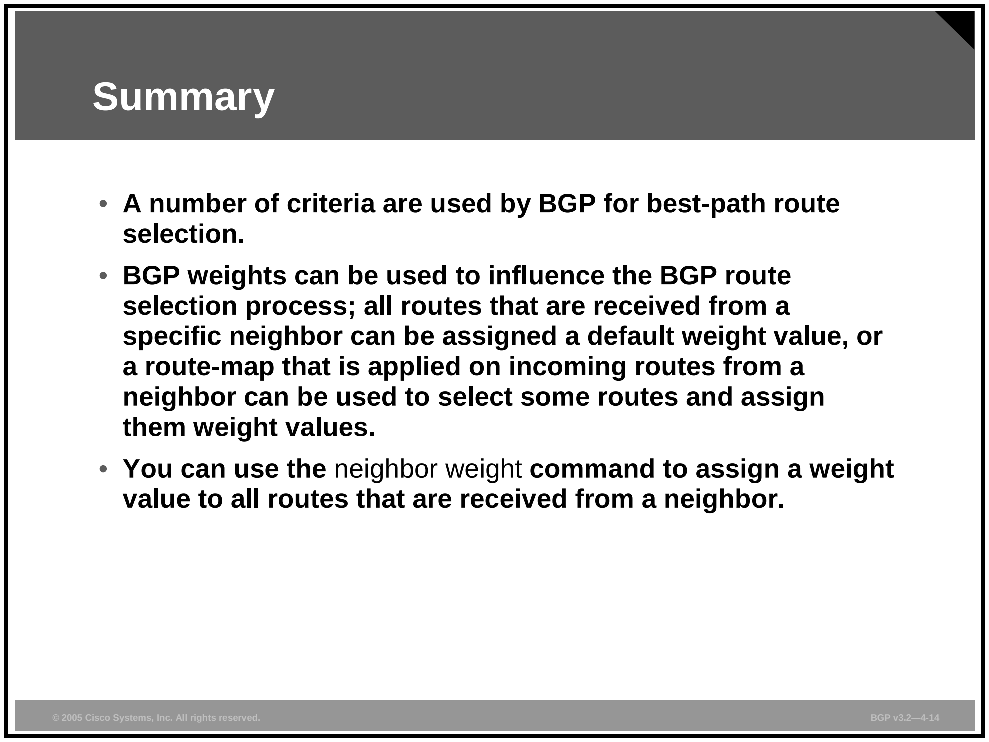
BGP v3.2—4-13

The figure shows all the possible applications of prefix-lists, filter-lists, weights, and route-maps. They are applied in the order indicated.

Prefix-lists and filter-lists, both in and out, filter out routes and discard those that are not permitted. Weight setting is applicable only on incoming routes because a router never propagates the weight attribute to its neighbors. Route-maps can be filters that discard routes but can also be used to modify and set various attributes on both incoming and outgoing routes.

Summary

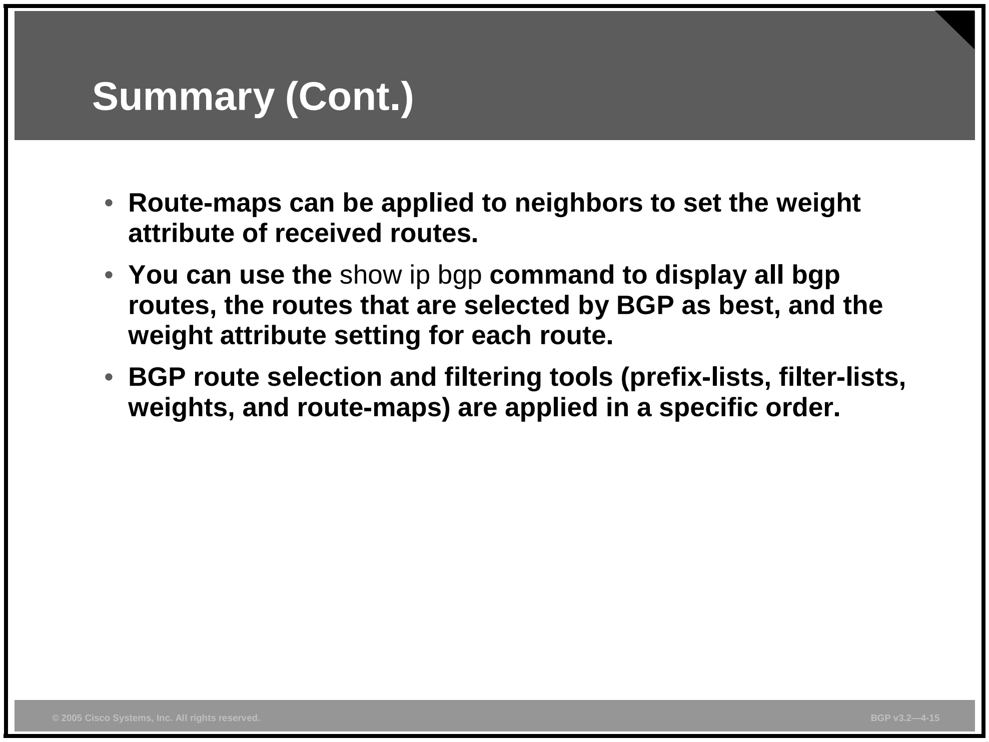
This topic summarizes the key points discussed in this lesson.



Summary

- A number of criteria are used by BGP for best-path route selection.
- BGP weights can be used to influence the BGP route selection process; all routes that are received from a specific neighbor can be assigned a default weight value, or a route-map that is applied on incoming routes from a neighbor can be used to select some routes and assign them weight values.
- You can use the neighbor weight command to assign a weight value to all routes that are received from a neighbor.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-14



Summary (Cont.)

- Route-maps can be applied to neighbors to set the weight attribute of received routes.
- You can use the show ip bgp command to display all bgp routes, the routes that are selected by BGP as best, and the weight attribute setting for each route.
- BGP route selection and filtering tools (prefix-lists, filter-lists, weights, and route-maps) are applied in a specific order.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-15

Lesson 2

Setting BGP Local Preference

Overview

When connections to multiple providers are required, it is important that Border Gateway Protocol (BGP) select the optimum route for traffic to use. The optimum, or best, route may not be what the network designer intended based on design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides many tools for administrators to influence route selection. One of these tools is the local preference attribute.

This lesson discusses how to influence BGP route selection by setting the BGP local preference attribute of incoming BGP routes. Local preference is similar to the weight attribute but differs from the BGP weight attribute in that weight is local to a specific router on which it is configured. Two methods that are used to set the local preference attribute, default local preference and route-maps, are discussed in this lesson. This lesson also explains how to monitor the BGP table to verify correct local preference configuration and to properly influence path selection.

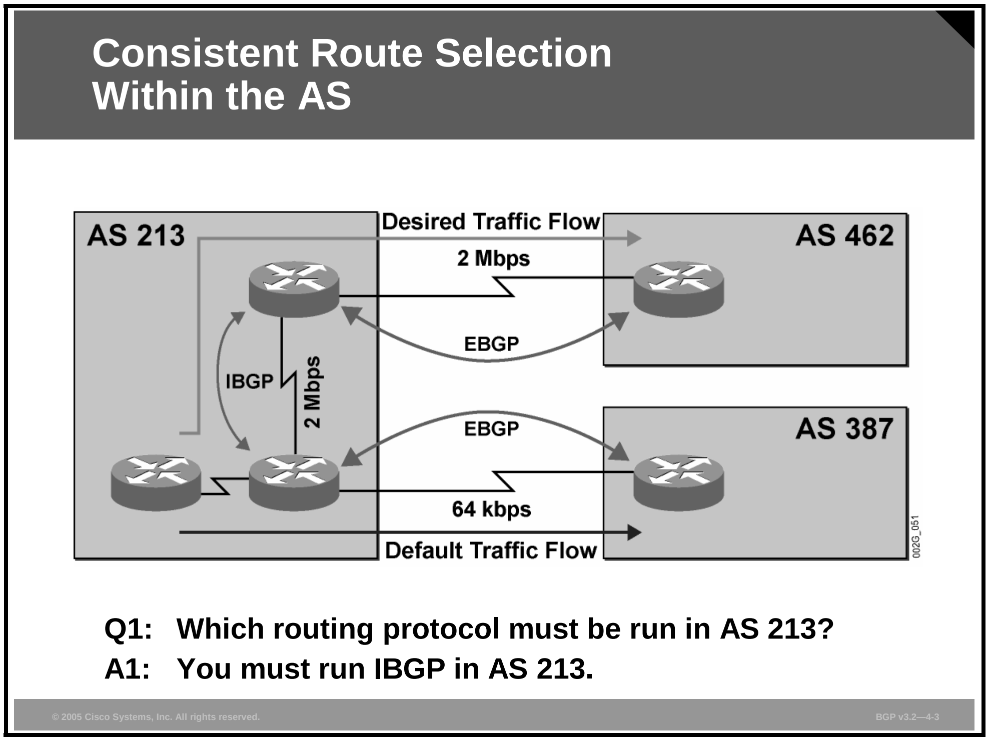
Objectives

Upon completing this lesson, you will be able to use the local preference attribute to influence route selection in a customer scenario in which you must support multiple ISP connections. This ability includes being able to meet these objectives:

- Explain why using BGP weights might not provide consistent BGP route selection in an AS
- Describe how the BGP local preference attribute influences BGP route selection
- Identify the Cisco IOS command that is required to configure default BGP local preference on a router
- Identify the Cisco IOS commands that are required to configure BGP local preference using route-maps
- Identify the Cisco IOS commands that are required to monitor BGP local preference

Consistent Route Selection Within the AS

This topic explains why using BGP weights might not provide consistent BGP route selection inside an autonomous system (AS).



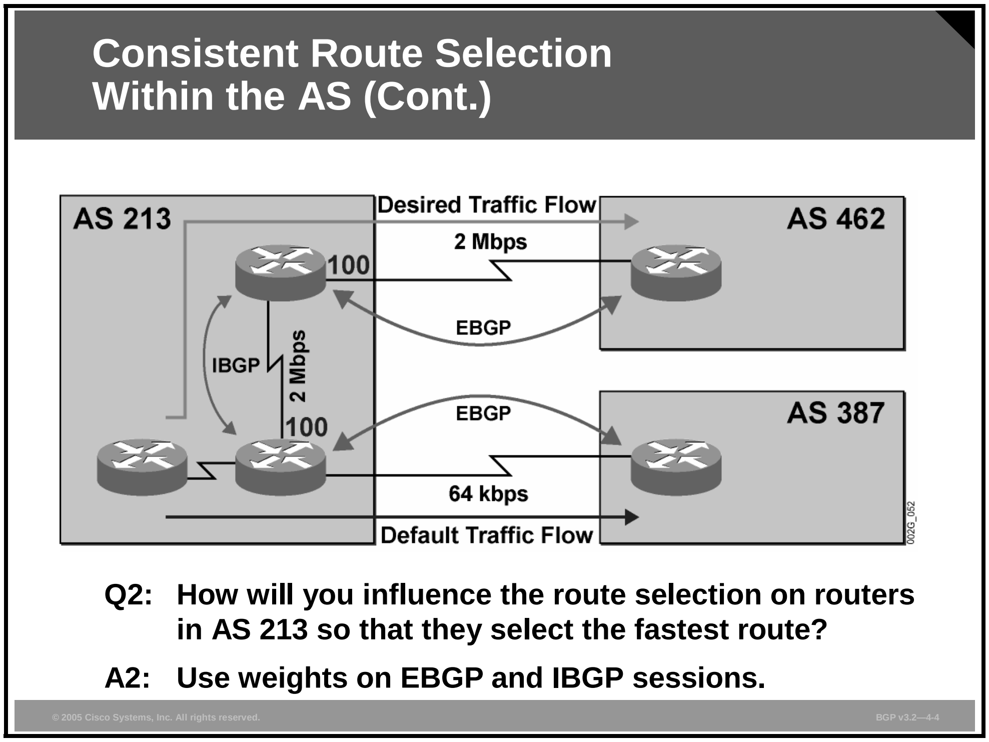
Consistent Route Selection Within the AS

Q1: Which routing protocol must be run in AS 213?

A1: You must run IBGP in AS 213.

Using BGP in autonomous systems with a single neighbor relationship usually does not require any advanced features. In situations like the one shown in the figure, however, it is important to ensure that customer routers choose the correct link. Obviously, the router should choose the 2-Mbps link and use the 64-kbps link only for backup purposes.

To make sure that the router selects the upper link (2-Mbps link) as its primary link and has the ability to switch over to the backup if a failure occurs, you must configure an Internal Border Gateway Protocol (IBGP) session between the two border routers in AS 213.



Consistent Route Selection Within the AS (Cont.)

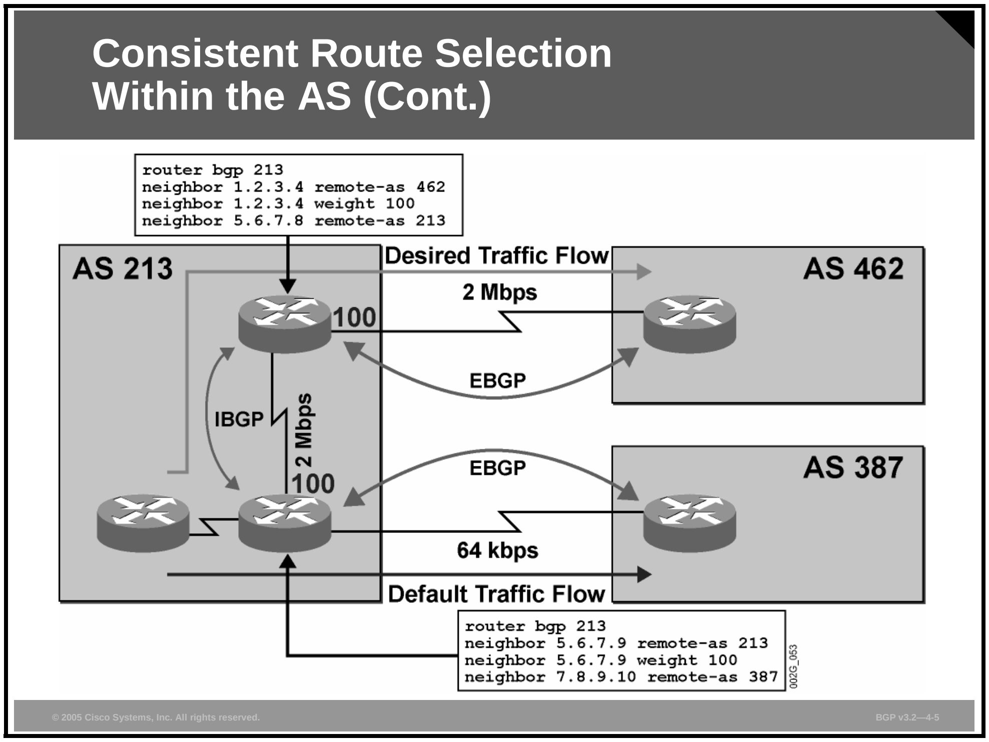
Q2: How will you influence the route selection on routers in AS 213 so that they select the fastest route?

A2: Use weights on EBGP and IBGP sessions.

BGP v3.2—4-4

One way of changing the default route selection is to use weights. Weight is an attribute that is locally significant to a router. Weight is a property, or parameter, and is, therefore, not seen on any neighboring routers. When designing BGP networks using weights, network administrators should set the weights on every router. If there is more than one path for the same network, a router will choose the one with the highest weight. The default value for weight is 0.

In this example, the upper router in AS 213 sets a weight of 100 for routes that are received over the 2-Mbps link from AS 462 (primary link) and prefers them to possible internal updates from the bottom router, where the default weight is 0. The bottom router sets a weight of 100 for internal routes that are received from the upper router and prefers them to routes that are received from AS 387. As a result, all packets will leave the AS through the primary 2-Mbps link.



Consistent Route Selection Within the AS (Cont.)

```
router bgp 213
neighbor 1.2.3.4 remote-as 462
neighbor 1.2.3.4 weight 100
neighbor 5.6.7.8 remote-as 213
```

```
router bgp 213
neighbor 5.6.7.9 remote-as 213
neighbor 5.6.7.9 weight 100
neighbor 7.8.9.10 remote-as 387
```

The configurations that are shown in the figure demonstrate how to change the default weight on a per-neighbor basis. If you use the **neighbor weight** command, all newly arrived updates will have a weight of 100. Updates coming from the other neighbor will still have the default weight of 0.

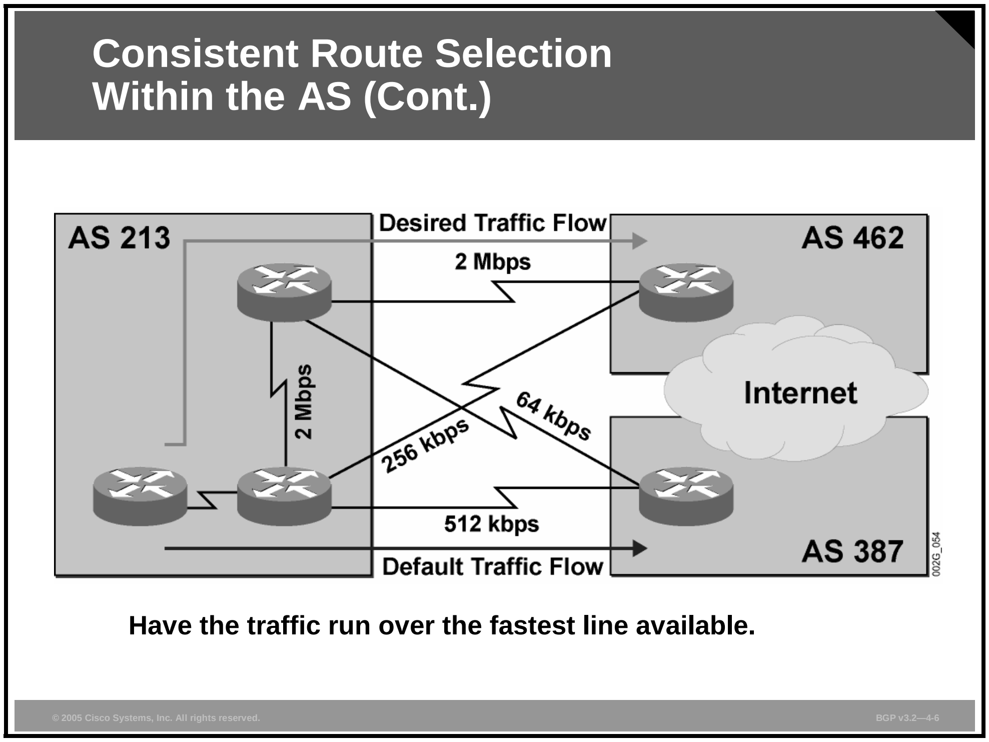
After you have applied the **neighbor weight** command, a refresh is needed from the neighbor. There are three ways of doing this, depending on the Cisco IOS version:

- Use the **clear ip bgp neighbor address** command to clear the neighbor relationship and re-establish it to refresh the BGP entries and apply the weight.
- Configure soft reconfiguration for the neighbor and use the **clear** command. You can perform all subsequent clearing by using the **clear ip bgp neighbor address soft in** command, which does not reset the neighbor relationship. The soft reconfiguration feature is supported by Cisco starting with Cisco IOS Software Release 11.2.
- Use the **clear ip bgp neighbor address in** command if both neighboring routers support the route refresh. The route refresh feature is available from Cisco starting with Cisco IOS Software Release 12.1.

See the module “Route Selection Using Policy Controls” for a detailed description of the commands here.

Example: Consistent Route Selection Within the AS

The figure shows two of the routers in AS 213 with two route-maps.



Consistent Route Selection
Within the AS (Cont.)

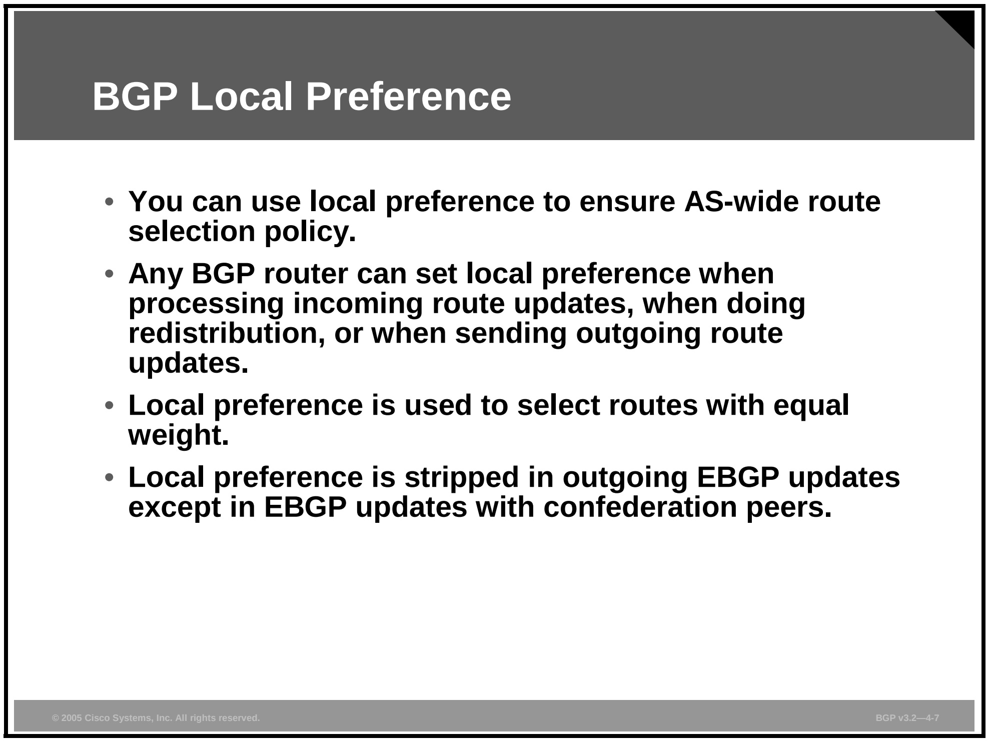
Have the traffic run over the fastest line available.

BGP v3.2—4-6

This example is more complex. When you are trying to implement this example with weights, it requires two route-maps on each router within AS 213. Luckily, BGP has a similar mechanism that you can use for consistent AS-wide route selection: local preference.

BGP Local Preference

This topic describes how the BGP local preference attribute influences BGP route selection.



BGP Local Preference

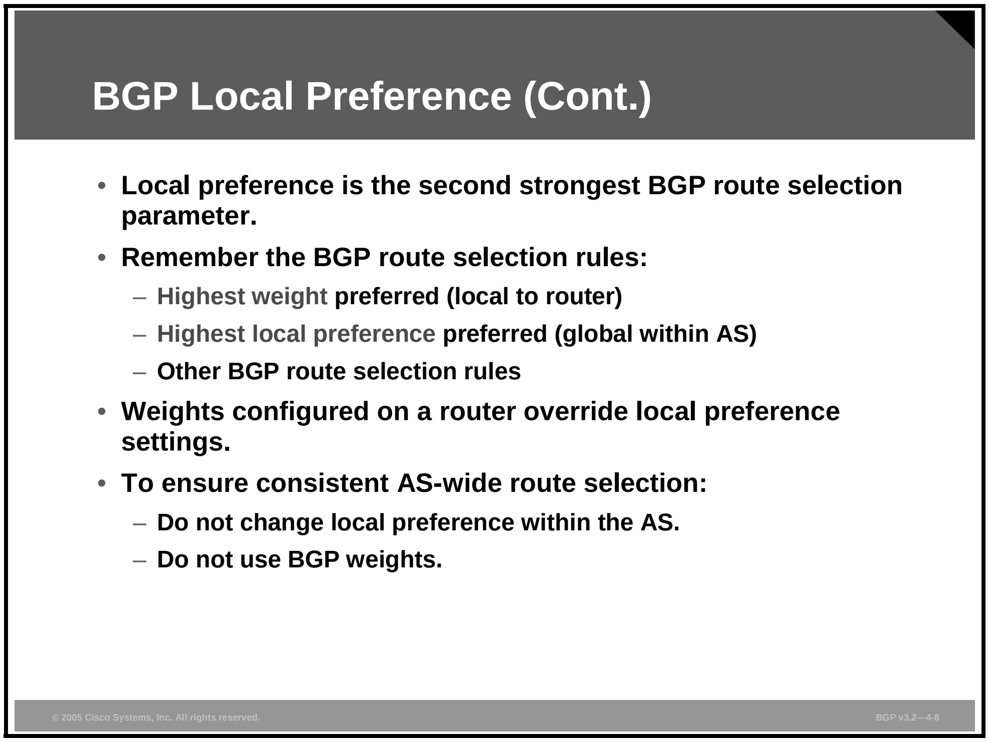
- You can use local preference to ensure AS-wide route selection policy.
- Any BGP router can set local preference when processing incoming route updates, when doing redistribution, or when sending outgoing route updates.
- Local preference is used to select routes with equal weight.
- Local preference is stripped in outgoing EBGP updates except in EBGP updates with confederation peers.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-7

Local preference is similar to weight; because it is an attribute, you can set it once and then view it on neighboring routers without having to reset it. This attribute has a default value of 100, which the router will apply to locally originated routes and updates that come in from external neighbors. Updates that come from internal neighbors already have the local preference attribute.

Local preference is the second-strongest criterion in the route selection process. If there are two or more paths available for the same network, a router will first compare weight, and if the weights are equal for all paths, the router will then compare the local preference attributes. The path with the highest local preference value will be preferred.

The local preference attribute is automatically stripped out of outgoing updates to External Border Gateway Protocol (EBGP) sessions. This practice means that you can use this attribute only within a single AS to influence the route selection process.



BGP Local Preference (Cont.)

- Local preference is the second strongest BGP route selection parameter.
- Remember the BGP route selection rules:
  - Highest weight preferred (local to router)
  - Highest local preference preferred (global within AS)
  - Other BGP route selection rules
- Weights configured on a router override local preference settings.
- To ensure consistent AS-wide route selection:
  - Do not change local preference within the AS.
  - Do not use BGP weights.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-8

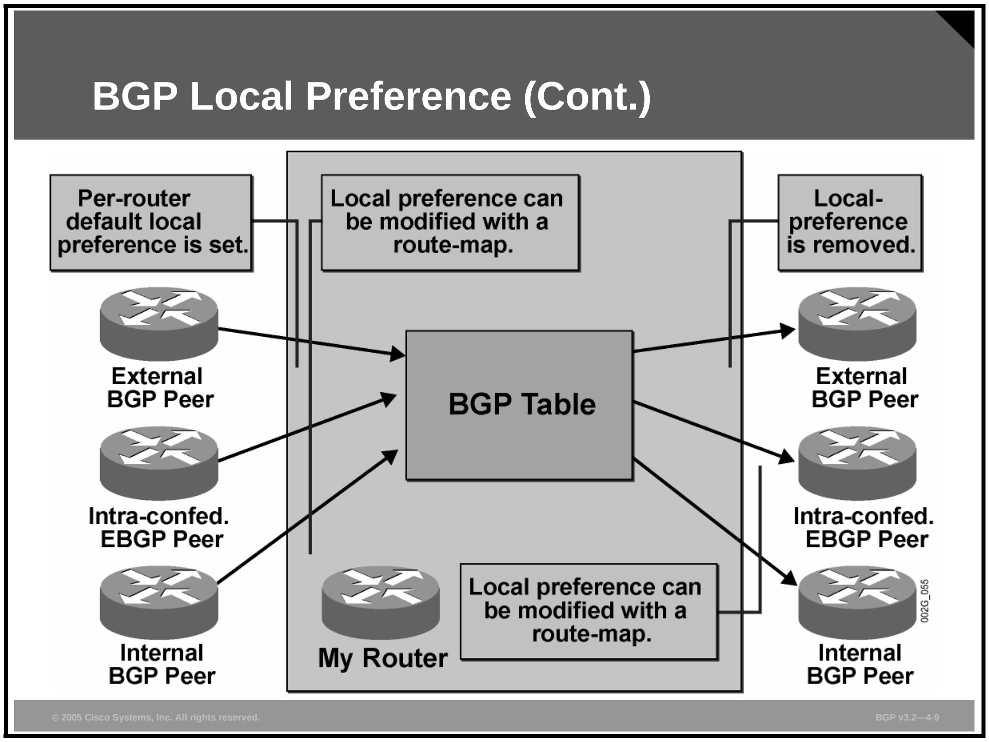
Local preference is the second-strongest BGP route selection parameter. Remember the route selection rules:

1. Prefer the highest weight (local to router).
2. Prefer the highest local preference (global within AS).
3. Process all remaining BGP route selection rules.

Because network administrators can use both weight and local preference to manipulate the route selection process, they must decide which one to use. If local preference is used, the weight should be the same for all paths.

Network administrators can use weight on an individual router to override local preference settings that are used in the rest of the AS.

In most cases, it is enough to change the default local preference on updates coming from external neighbors. Network administrators should avoid changing the local preference attribute on internal sessions to prevent unnecessary complexity and unpredictable behavior.



BGP Local Preference (Cont.)

Per-router default local preference is set.

Local preference can be modified with a route-map.

Local-preference is removed.

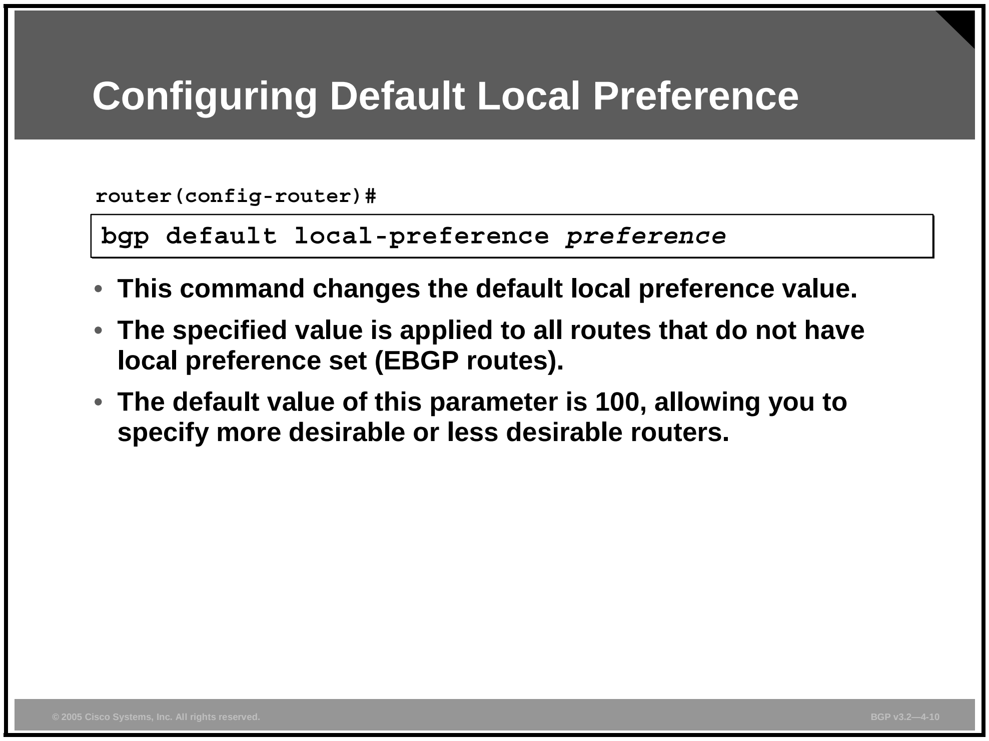
Local preference can be modified with a route-map.

Network administrators can apply local preference in the following ways:

- Use a route-map with the **set local-preference** command. You can use the route-map on incoming updates from all neighbors or on outgoing updates to internal neighbors (not recommended).
- Use the **bgp default local-preference** command to change the default local preference value that is applied to all updates that come from external neighbors or that originate locally.

Configuring Default Local Preference

This topic describes the Cisco IOS command that is required to configure default BGP local preference on a Cisco router.



Configuring Default Local Preference

```
router(config-router)#
bgp default local-preference preference
```

- **This command changes the default local preference value.**
- **The specified value is applied to all routes that do not have local preference set (EBGP routes).**
- **The default value of this parameter is 100, allowing you to specify more desirable or less desirable routers.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-10

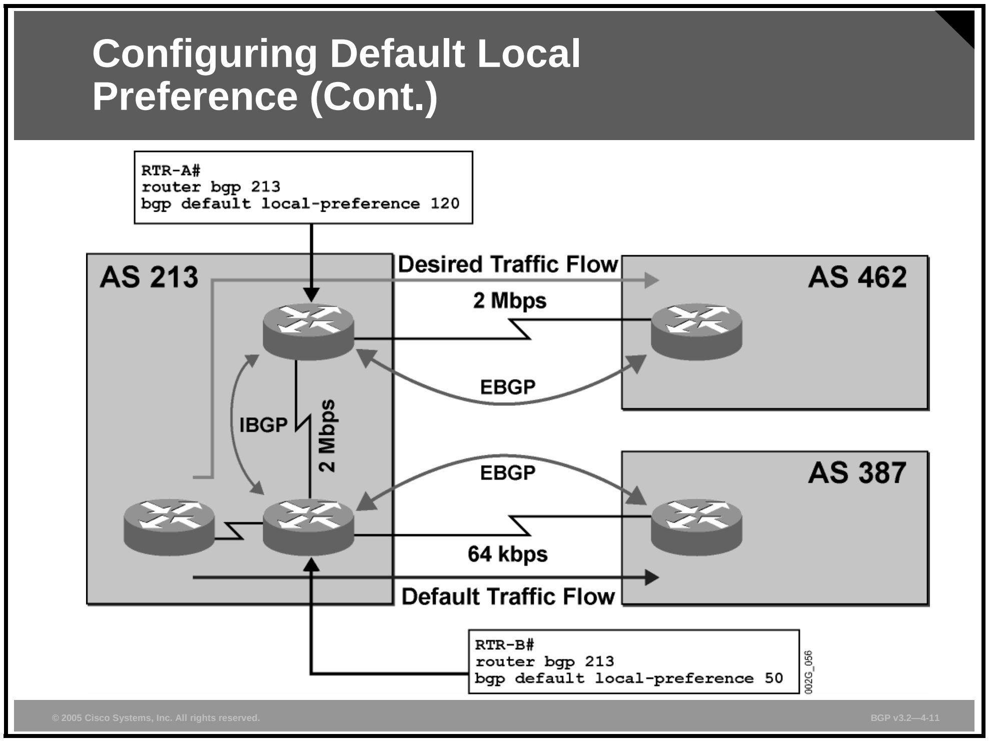
You can use the **bgp default local-preference** command in BGP configuration mode to change the default value of local preference. The new default value applies only to locally originated routes and those that are received from external neighbors.

Setting a value lower than the default of 100 will result in the router preferring internal paths to external (normally a router would prefer external routes).

Setting a value higher than 100 will result in external paths being preferred to all internal paths (also those with a shorter AS path).

Example: Configuring Default Local Preference

In the figure, the local preference attribute is used instead of weights.



Configuring Default Local Preference (Cont.)

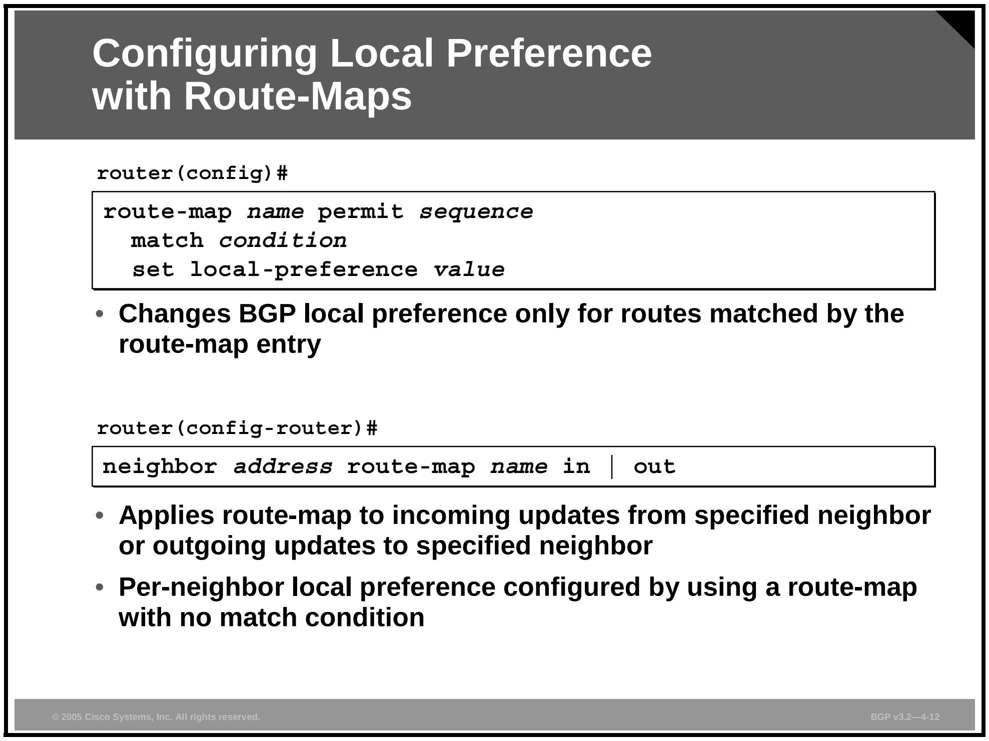
```
RTR-A#
router bgp 213
bgp default local-preference 120
```

```
RTR-B#
router bgp 213
bgp default local-preference 50
```

The two indicated routers in AS 213 have different default local preference values that are applied to external updates. The bottom router receives updates from the external neighbor and applies local preference to them. The same router then receives updates from the upper router, which set a local preference of 120 to all external updates. The bottom router then compares all paths and, where two paths exist, chooses the one with the higher local preference (120).

Configuring Local Preference with Route-Maps

This topic lists the Cisco IOS commands that are required to configure BGP local preference with route-map statements.



Configuring Local Preference with Route-Maps

```
router(config)#
```

```
route-map name permit sequence
  match condition
  set local-preference value
```

- **Changes BGP local preference only for routes matched by the route-map entry**

```
router(config-router)#
```

```
neighbor address route-map name in | out
```

- **Applies route-map to incoming updates from specified neighbor or outgoing updates to specified neighbor**
- **Per-neighbor local preference configured by using a route-map with no match condition**

BGP v3.2—4-12

To have more control over setting local preference, you may be forced to use a route-map. A route-map can have more sequenced statements, each with a different **set local-preference** command and a different match condition. If there is no **match** command, the route-map statement will apply local preference to all routes. The route-map can then be applied to BGP route updates in either the incoming or outgoing direction.

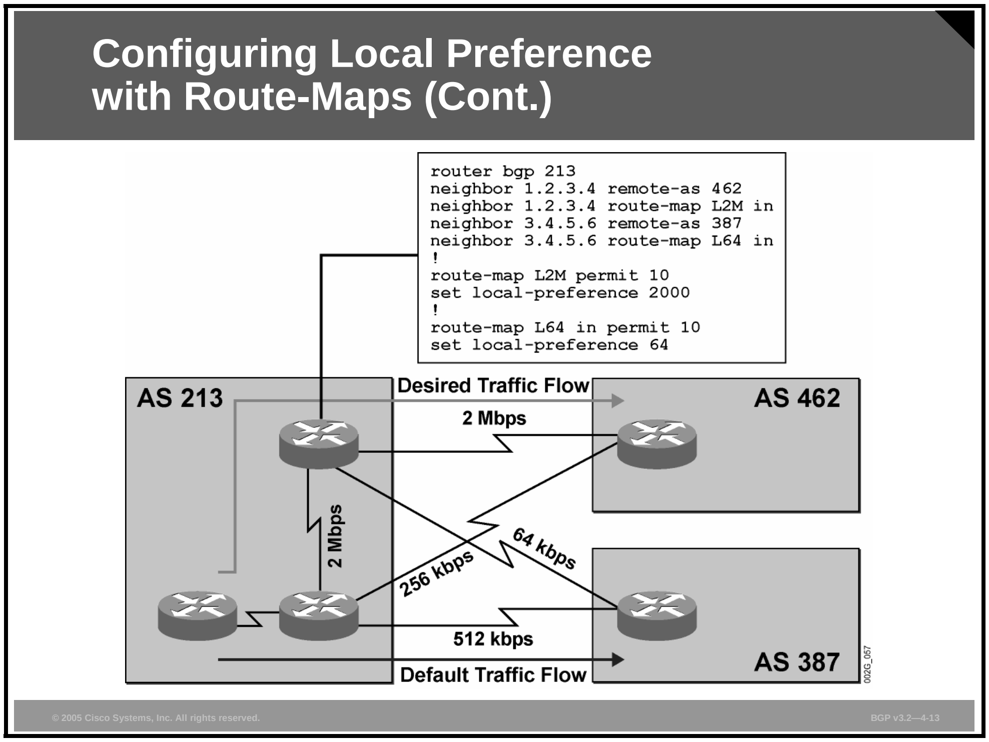
Note	Applying a route-map to outgoing updates on external sessions will have no effect on local preference in the neighboring AS.

When routers use a route-map to set local preference, the route-map is typically applied to incoming BGP routes that are advertised by an EBGP neighbor. The local router uses the local preference attribute in BGP route selection. In addition, the router also propagates the attribute to all IBGP sessions in the local AS. Normally, no modifications of local preference are made on IBGP sessions. This restriction ensures that all routers in the local AS use the same local preference value and make the same decision in the route selection process.

Note	If a network is not matched in any of the route-map statements, the network will be filtered. To permit unmatched networks without setting the local preference attribute, you should add another route-map statement without **match** or **set** commands to the end of the route-map. This statement should simply permit the remaining networks.

Example: Configuring Local Preference with Route-Maps

In this example, both routers in AS 213 have two external sessions.



Configuring Local Preference with Route-Maps (Cont.)

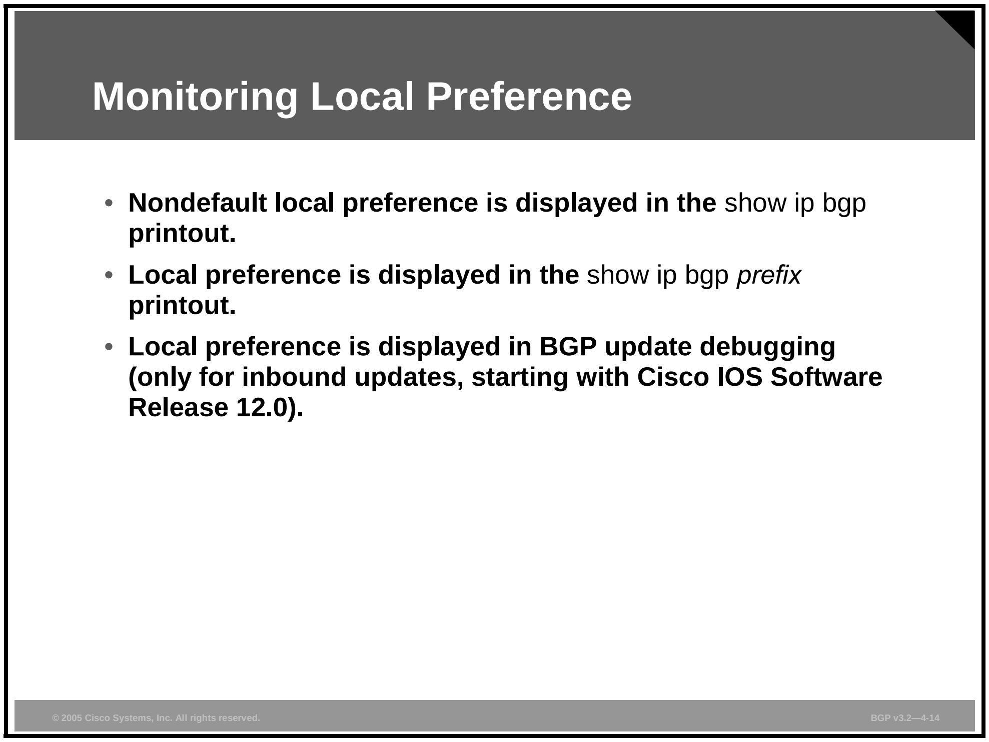
```
router bgp 213
neighbor 1.2.3.4 remote-as 462
neighbor 1.2.3.4 route-map L2M in
neighbor 3.4.5.6 remote-as 387
neighbor 3.4.5.6 route-map L64 in
!
route-map L2M permit 10
set local-preference 2000
!
route-map L64 in permit 10
set local-preference 64
```

Using the **bgp default local-preference** command is no longer possible because the second-fastest link is on another router.

The configuration here sets local preference according to the bandwidth of the link. A similar configuration exists on the bottom router. If the primary (2-Mbps) link fails, the paths that are learned through the bottom router in AS 213 (routes with a local preference of 512) will be used.

Monitoring Local Preference

This topic lists the Cisco IOS commands that are necessary to monitor BGP local preference.



Monitoring Local Preference

- **Nondefault local preference is displayed in the** show ip bgp **printout.**
- **Local preference is displayed in the** show ip bgp *prefix* **printout.**
- **Local preference is displayed in BGP update debugging (only for inbound updates, starting with Cisco IOS Software Release 12.0).**

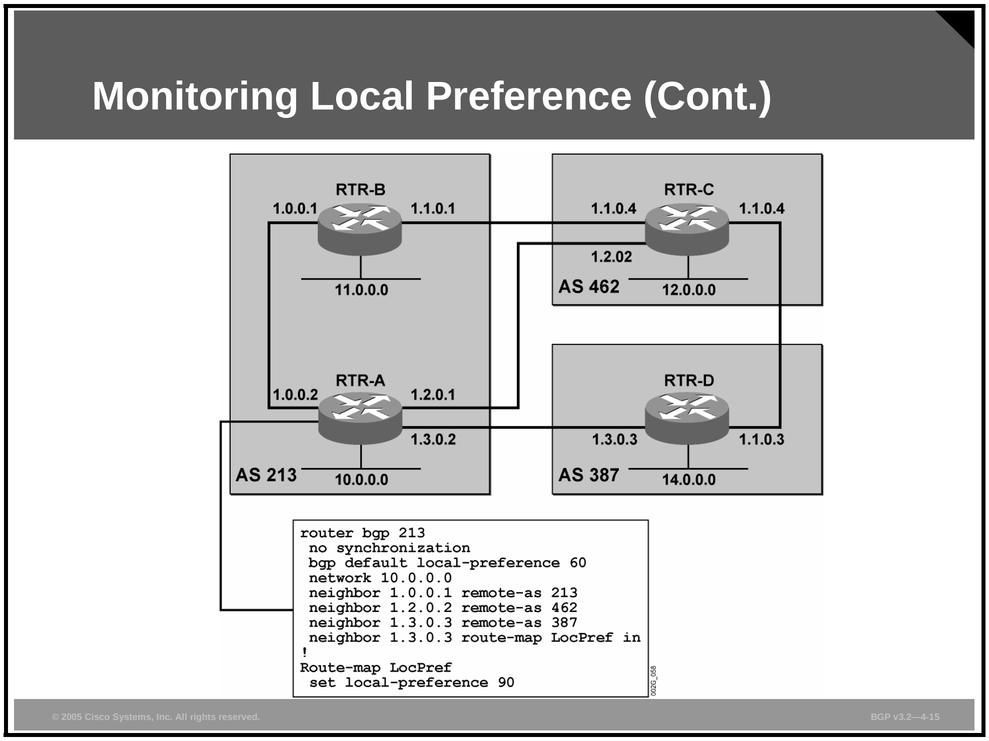
BGP v3.2—4-14

Although local preference is not a mandatory attribute, it is applied to every route. When you are using the **show ip bgp** command, a locally applied default value is not shown. All other values are displayed. You should use the **show ip bgp** *prefix* command to also display the locally applied value.

The output that is displayed from **show** and **debug** commands will vary depending on the Cisco IOS version. Newer versions typically display more information. In Cisco IOS Software Release 12.0 and in later versions, enabling debugging of incoming routing updates will also display the local preference attribute.

Example: Monitoring Local Preference

In the figure, the network shown was used to collect output from the **show** and **debug** commands in the next few examples.



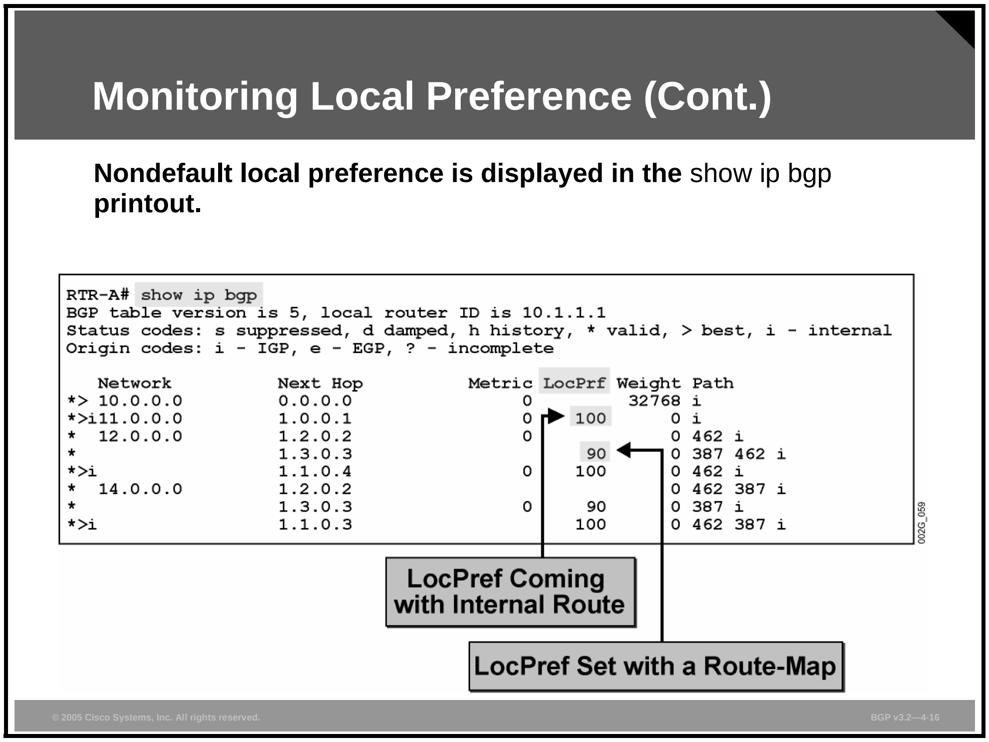
Monitoring Local Preference (Cont.)

```
router bgp 213
 no synchronization
 bgp default local-preference 60
 network 10.0.0.0
 neighbor 1.0.0.1 remote-as 213
 neighbor 1.2.0.2 remote-as 462
 neighbor 1.3.0.3 remote-as 387
 neighbor 1.3.0.3 route-map LocPref in
!
Route-map LocPref
 set local-preference 90
```

BGP v3.2—4-15

Every physical connection also includes a BGP session. All monitoring and troubleshooting commands were used on router RTR-A.

RTR-A has one internal and two external neighbors. RTR-B is setting local preference 100 for all updates, and RTR-A is setting a default local preference (value 60) for all external updates except for those coming from router RTR-D, where a route-map is used to set a local preference of 90. The following pages show the output of **show** and **debug** commands on router RTR-A.



Monitoring Local Preference (Cont.)

Nondefault local preference is displayed in the show ip bgp **printout.**

```
RTR-A# show ip bgp
BGP table version is 5, local router ID is 10.1.1.1
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
*> 10.0.0.0         0.0.0.0                0         32768 i
*>i11.0.0.0         1.0.0.1                0    100      0 i
*  12.0.0.0         1.2.0.2                0             0 462 i
*                   1.3.0.3                      90      0 387 462 i
*>i                 1.1.0.4                0    100      0 462 i
*  14.0.0.0         1.2.0.2                              0 462 387 i
*                   1.3.0.3                0     90      0 387 i
*>i                 1.1.0.3                     100      0 462 387 i
```

LocPrf Coming with Internal Route

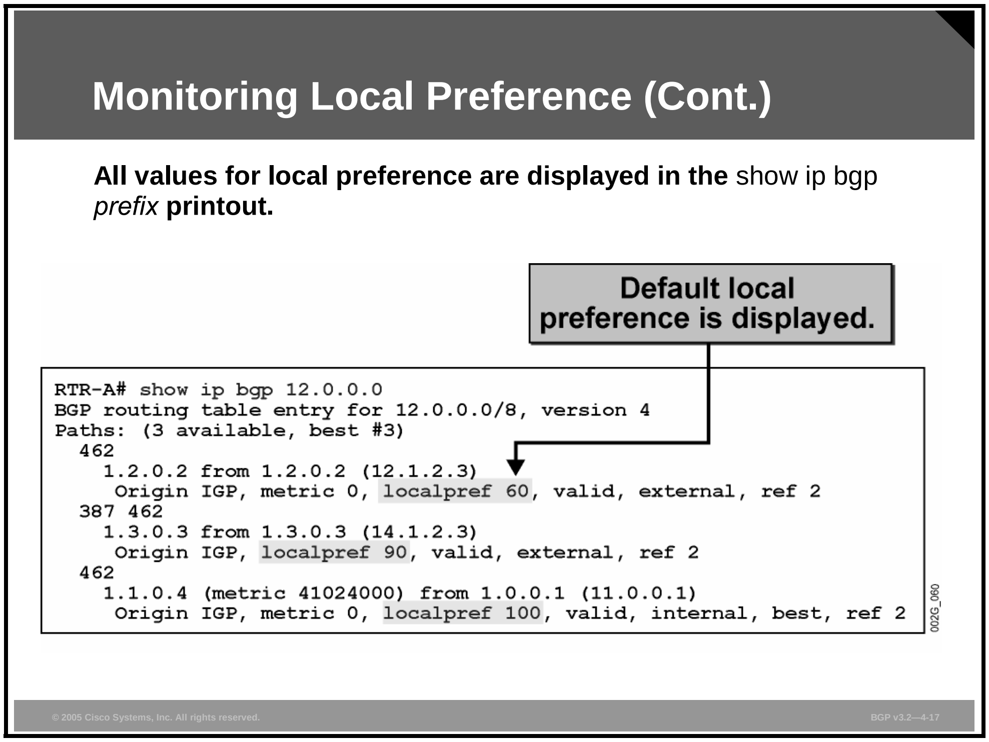
LocPrf Set with a Route-Map

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-16

The output in the figure contains routes with three different local preference values:

- Network 10.0.0.0/8 originates locally on RTR-A, and the applied default local preference 60 is not displayed.
- The second path for network 12.0.0.0/8 was received from RTR-D and received a local preference value of 90 by the route-map.
- All routes that are received from router RTR-B are marked as internal and have a local preference value of 100 set on RTR-B.

Note The output of the **show ip bgp** command will not display the local preference value if the value is the same as the **bgp default local-preference** value in the local router. In the example, RTR-B uses its default local preference value (100). When these routes are propagated to RTR-A, RTR-A displays the local preference value of 100 because it is different from the default local preference value that is configured on RTR-A.



Monitoring Local Preference (Cont.)

All values for local preference are displayed in the show ip bgp *prefix* **printout.**

Default local preference is displayed.

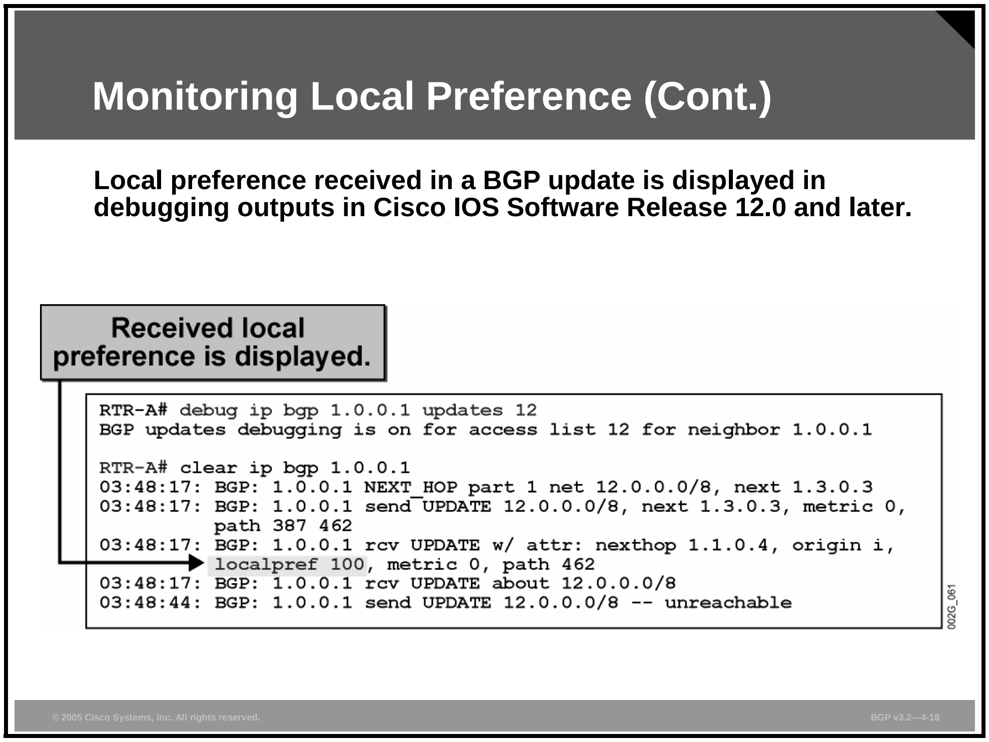
```
RTR-A# show ip bgp 12.0.0.0
BGP routing table entry for 12.0.0.0/8, version 4
Paths: (3 available, best #3)
  462
    1.2.0.2 from 1.2.0.2 (12.1.2.3)
     Origin IGP, metric 0, localpref 60, valid, external, ref 2
  387 462
    1.3.0.3 from 1.3.0.3 (14.1.2.3)
     Origin IGP, localpref 90, valid, external, ref 2
  462
    1.1.0.4 (metric 41024000) from 1.0.0.1 (11.0.0.1)
     Origin IGP, metric 0, localpref 100, valid, internal, best, ref 2
```

Use the **show ip bgp** *prefix* command to see more detailed information about a specific network, including the locally applied default local preference.

In this example, there are three paths to reach the same network:

- The first path is external and was received from router RTR-C. The new default local preference value 60 was applied to the update.
- The second path is external and was received from router RTR-D. The route-map was used to set a local preference of 90.
- The third path is internal and was received from RTR-B. The update already contained a local preference attribute with a value of 100.

Router RTR-A chose the last path as best because it has the highest local preference.



Monitoring Local Preference (Cont.)

Local preference received in a BGP update is displayed in debugging outputs in Cisco IOS Software Release 12.0 and later.

Received local preference is displayed.

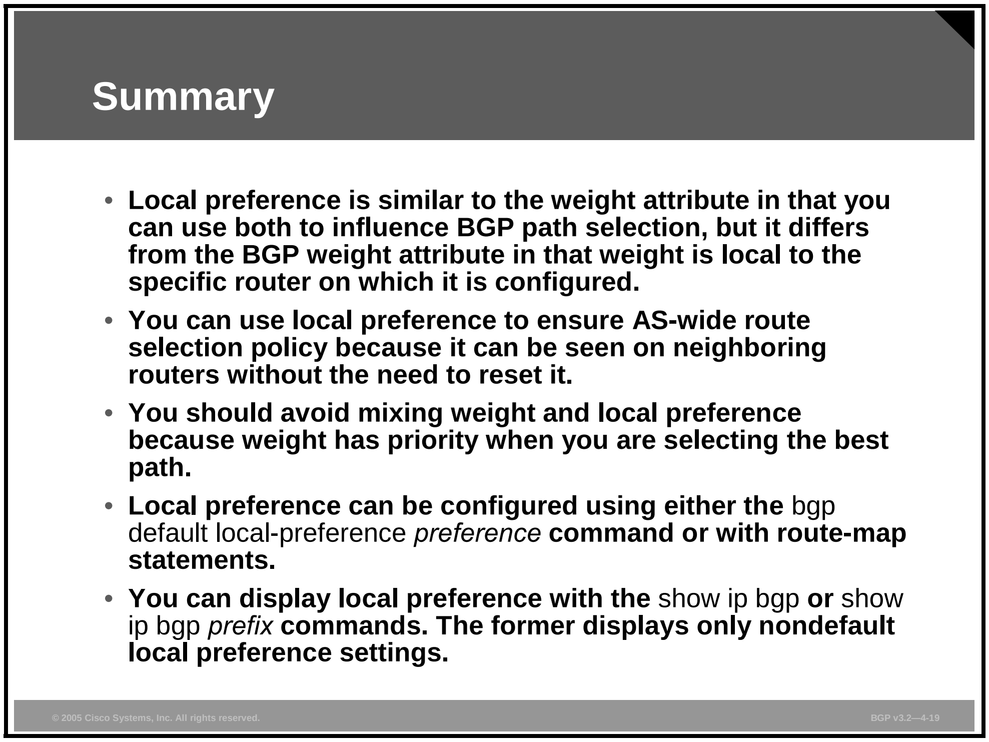
```
RTR-A# debug ip bgp 1.0.0.1 updates 12
BGP updates debugging is on for access list 12 for neighbor 1.0.0.1

RTR-A# clear ip bgp 1.0.0.1
03:48:17: BGP: 1.0.0.1 NEXT_HOP part 1 net 12.0.0.0/8, next 1.3.0.3
03:48:17: BGP: 1.0.0.1 send UPDATE 12.0.0.0/8, next 1.3.0.3, metric 0,
          path 387 462
03:48:17: BGP: 1.0.0.1 rcv UPDATE w/ attr: nexthop 1.1.0.4, origin i,
          localpref 100, metric 0, path 462
03:48:17: BGP: 1.0.0.1 rcv UPDATE about 12.0.0.0/8
03:48:44: BGP: 1.0.0.1 send UPDATE 12.0.0.0/8 -- unreachable
```

BGP v3.2—4-18

This figure shows the debugging output of incoming BGP updates. Because a router propagates the local preference attribute to other routers in the same AS only, local preference will be associated with routes sent from internal neighbors.

Summary

This topic summarizes the key points discussed in this lesson.



Summary

- **Local preference is similar to the weight attribute in that you can use both to influence BGP path selection, but it differs from the BGP weight attribute in that weight is local to the specific router on which it is configured.**
- **You can use local preference to ensure AS-wide route selection policy because it can be seen on neighboring routers without the need to reset it.**
- **You should avoid mixing weight and local preference because weight has priority when you are selecting the best path.**
- **Local preference can be configured using either the** bgp default local-preference *preference* **command or with route-map statements.**
- **You can display local preference with the** show ip bgp **or** show ip bgp *prefix* **commands. The former displays only nondefault local preference settings.**

Lesson 3

Using AS-Path Prepending

Overview

When connections to multiple providers are required it is important that Border Gateway Protocol (BGP) select the optimum route for traffic to use. The optimum, or best, route may not be what the network designer intended based on design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides many tools for administrators to influence route selection. One of these tools is autonomous system (AS)-path prepending.

Problems can arise when administrative policies mandate a specific return path be used for traffic returning to the AS, but AS-path prepending potentially allows the customer to influence the route selection of its service providers. This lesson describes AS-path prepending and the Cisco IOS commands required to properly configure and monitor AS-path configurations and filtering requirements for influencing route selection using AS-path prepending.

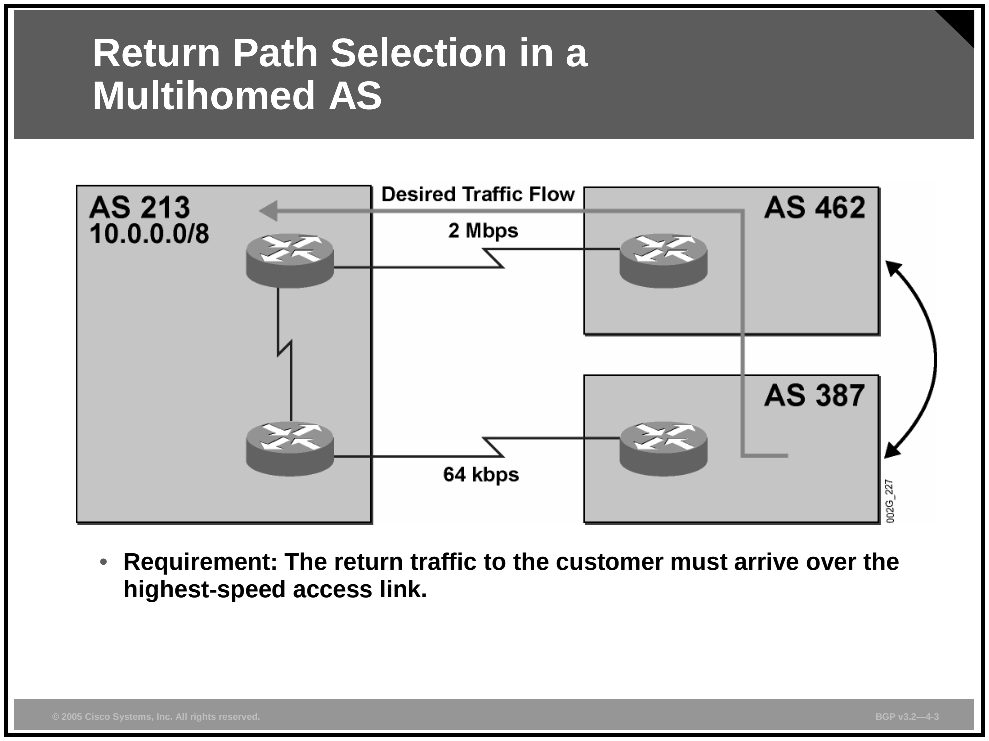
Objectives

Upon completing this lesson, you will be able to use AS-path prepending to influence the return path that is selected by the neighboring autonomous systems in a customer scenario where you must support multiple ISP connections. This ability includes being able to meet these objectives:

- Explain the need to influence BGP return path selection in a service provider environment
- Describe the function of AS-path prepending and how you can use it to facilitate proper return path selection
- Identify design considerations when you are implementing AS-path prepending to influence return path selection
- Identify the Cisco IOS commands that are required to configure AS-path prepending in a multihomed network
- Identify the Cisco IOS commands that are required to monitor the operation of AS-path prepending
- Describe the concerns with using AS-path filters when neighboring autonomous systems require AS-path prepending
- Describe the function of the BGP Hide Local-Autonomous System feature in combining separate BGP networks under a single AS

Return Path Selection in a Multihomed AS

This topic describes the need to influence BGP return path selection in a service provider environment.



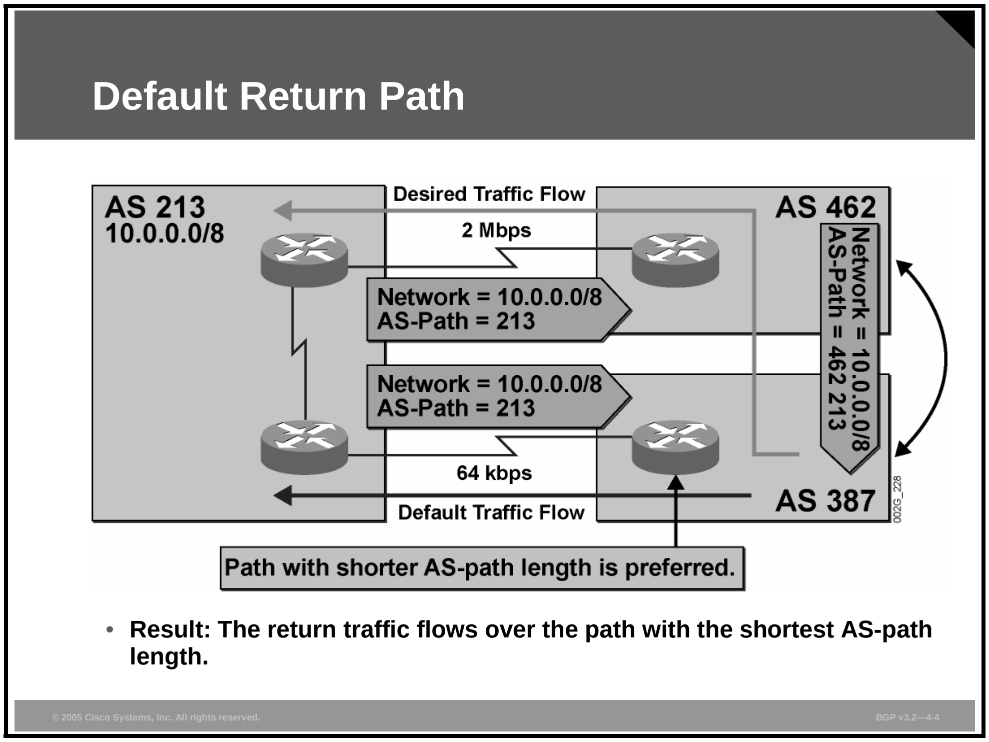
Return Path Selection in a Multihomed AS

- **Requirement: The return traffic to the customer must arrive over the highest-speed access link.**

It is fairly easy for an AS to select the appropriate path for outgoing traffic. It is much more complicated to influence other autonomous systems to select the appropriate path for traffic that is returning to a specific AS.

To configure the preferred path only for outgoing traffic and not for incoming (return) traffic is likely to result in asymmetrical traffic flow as well as suboptimal performance of the return traffic. In the figure, outgoing traffic is directed to the high-speed line (2 Mbps) as a result of configuring local preference or weight. However, the return traffic from AS 387 would take the default path over the low-speed line (64 kbps). The low-speed line would be a limiting factor in the overall performance that the network could achieve.

In this example, AS 213 requests AS 387 to send packets toward network 10.0.0.0/8 via AS 462. The reason for this request is to improve network performance and minimize delay (assuming, of course, that the connectivity between AS 387 and AS 462 is better than the direct 64-kbps link between AS 387 and AS 213).



Default Return Path

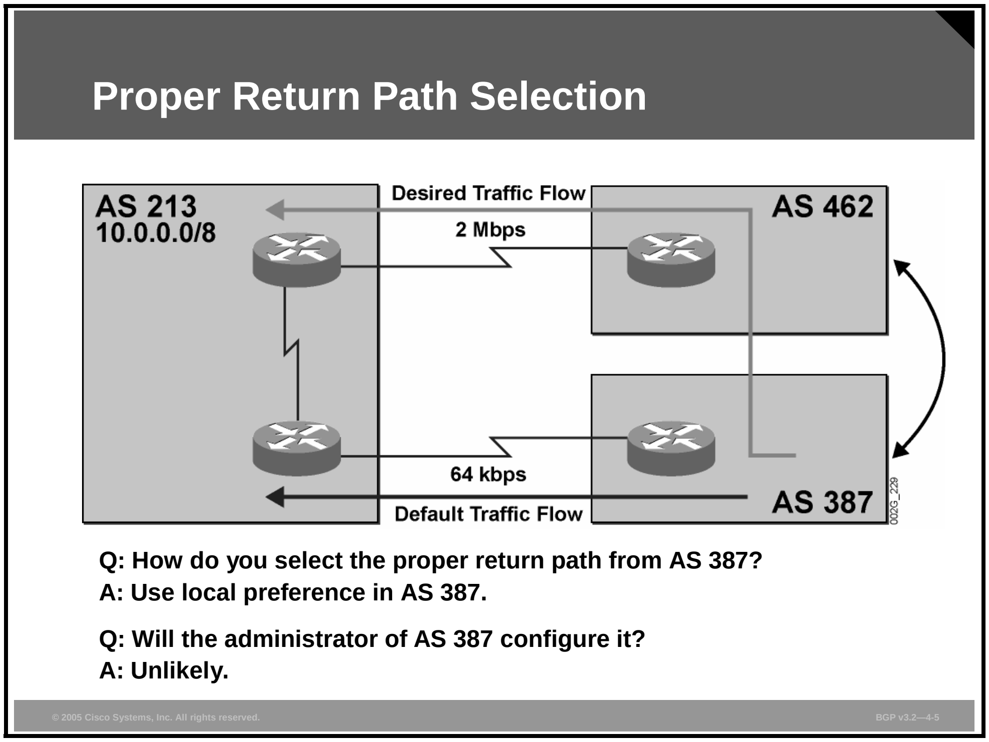
- **Result: The return traffic flows over the path with the shortest AS-path length.**

If no BGP path selection tools are configured on the route to influence the traffic flow, AS 387 will use the shortest AS path. This action will result in unwanted behavior because the return traffic to AS 213 will be sent over the low-speed WAN link.

AS 213 announces network 10.0.0.0/8 over EBGP sessions to both AS 462 and AS 387. When AS 213 sends EBGP updates, it changes the AS-path attribute according to BGP specifications. Both AS 462 and AS 387 receive a BGP update for network 10.0.0.0/8 with the AS path set to 213.

Because AS 462 selects the route for network 10.0.0.0/8 that it received from AS 213 as its best route, AS 462 uses that route and forwards it to AS 387. According to BGP specifications, AS 462 also changes the AS-path attribute. AS 387 receives the route to network 10.0.0.0/8 from AS 462 with an AS path set to 462 213.

AS 387 has now received two alternative routes to network 10.0.0.0/8 (the direct route from AS 213 and the route through AS 462). Because nothing is configured in AS 387 to influence the flow of traffic, the router will use the BGP route selection rule of shortest AS path to select the best return path to network 10.0.0.0/8.



Proper Return Path Selection

Q: How do you select the proper return path from AS 387?
A: Use local preference in AS 387.

Q: Will the administrator of AS 387 configure it?
A: Unlikely.

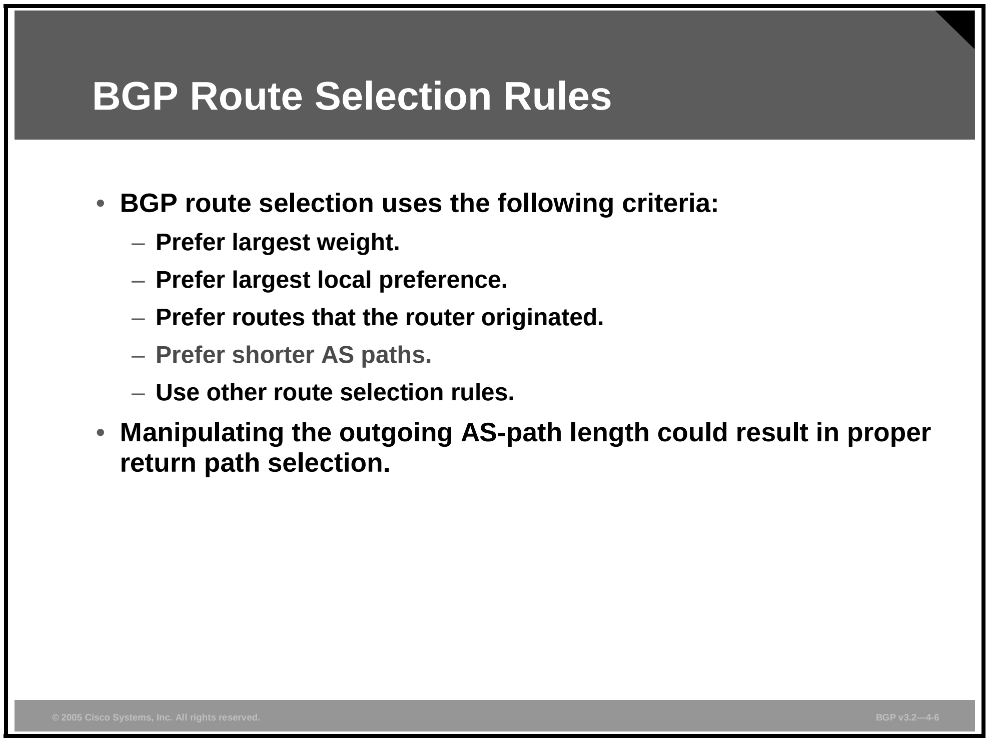
Remember that the incoming traffic flow (from the perspective of AS 213) will be a result of the route selection for outgoing traffic in AS 387. The traffic that is going out from AS 387 will end up as incoming traffic in AS 213.

If AS 387 configures some changes that cause the route selection process for outgoing traffic to prefer to reach network 10.0.0.0/8 via AS 462, the changes would result in behavior matching the desired administrative policy for AS 213, which specifies that incoming traffic to the AS should be received over the high-speed link.

One way to accomplish the desired administrative policy in AS 213 is to configure the router in AS 387, which is receiving EBGP updates directly from AS 213, to assign a local preference value less than the default value (100) to all routes that are received from AS 213. The router in AS 387 is also configured specifically not to set local preference on External Border Gateway Protocol (EBGP) routes that are received from AS 462. This configuration results in assignment of the default value of 100 to all routes received from AS 462. When the route selection process in AS 387 selects the best route to reach network 10.0.0.0/8, the difference in local preference values causes AS 387 routers to select the path via AS 462 as the best.

However, all the configuration work to complete this process must be performed in AS 387. The network administrators of AS 387 would be required to modify the router configurations in AS 387 to satisfy the administrative policy requirements of AS 213. All changes must be documented and maintained according to the rules and procedures that have been adopted by AS 387.

If AS 387 is a major Internet service provider (ISP), the network administrators most likely are too busy doing other things to tailor router configurations that are based on the demand of a single leaf (nontransit) AS that lacks bandwidth on a redundant connection.



BGP Route Selection Rules

- **BGP route selection uses the following criteria:**
  - **Prefer largest weight.**
  - **Prefer largest local preference.**
  - **Prefer routes that the router originated.**
  - **Prefer shorter AS paths.**
  - **Use other route selection rules.**
- **Manipulating the outgoing AS-path length could result in proper return path selection.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-6

Recall that BGP route selection uses the following criteria:

- Prefer the largest weight.
- Prefer the largest local preference.
- Prefer routes that the router originated.
- Prefer shorter AS paths.
- Then, prefer all other route selection criteria.

It is unlikely that the operator of an AS can request changes in router configurations in another AS. This limitation makes it virtually impossible to influence another AS to select the desired path based on the weight and local preference attributes, because both options would require configuration changes in the neighboring AS.

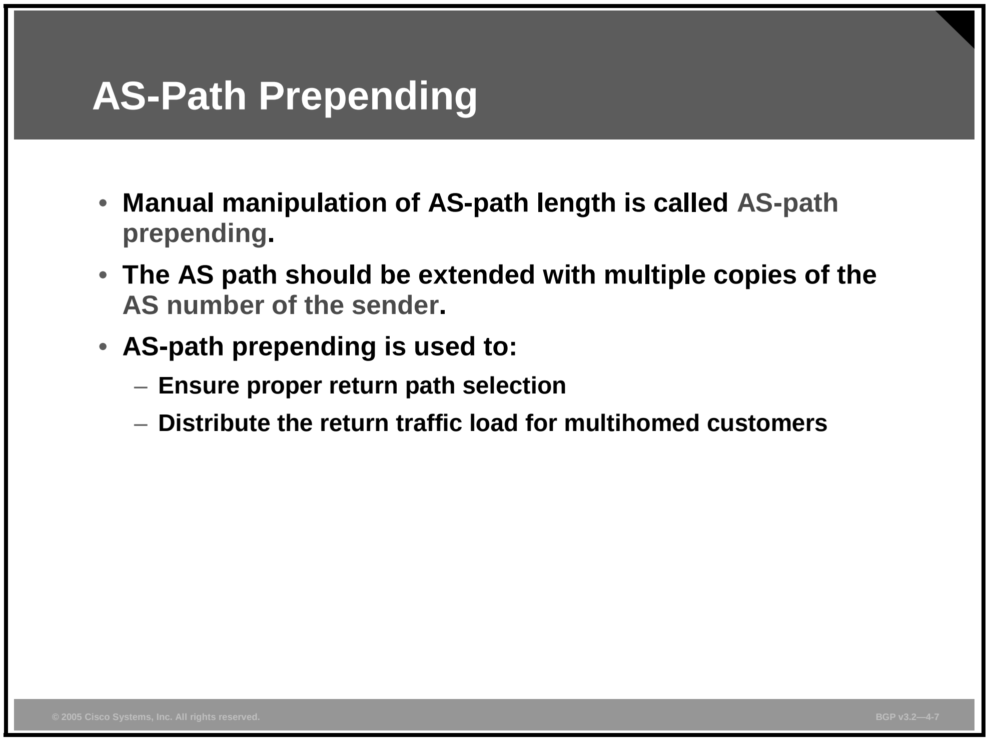
But if both the weight and the local preference parameters are left at their default settings, they will not indicate a difference. This configuration causes the route selection process to continue down the list of selection criteria. The third criterion for selection will not influence route selection in this scenario, because none of the routes originated at the router that is performing the route selection. The fourth criterion will apply, however, because the AS paths have different lengths.

If the AS path is not manually manipulated by some administrative means, the path going over the fewest number of autonomous systems is selected by the router regardless of available bandwidth. However, if the AS that is attempting to influence the incoming traffic flow is sending out EBGP updates with a manipulated AS-path attribute over that undesired path, the receiver of this update is less likely to select it as the best because the AS path now appears to be longer.

The benefit of manipulating AS paths to influence the route selection is that the configuration that is needed is done in the AS that is requesting a desired return path.

AS-Path Prepending

This topic describes the function of AS-path prepending and how you can use it to facilitate proper return path selection.



AS-Path Prepending

- **Manual manipulation of AS-path length is called AS-path prepending.**
- **The AS path should be extended with multiple copies of the AS number of the sender.**
- **AS-path prepending is used to:**
  - **Ensure proper return path selection**
  - **Distribute the return traffic load for multihomed customers**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-7

You can manipulate AS paths by prepending AS numbers to existing AS paths. Normally, you perform AS-path prepending on outgoing EBGP updates over the nondesired return path. Because the AS paths sent out over the nondesired link become longer than the AS path sent out over the preferred path, the nondesired link is now less likely to be used as the return path.

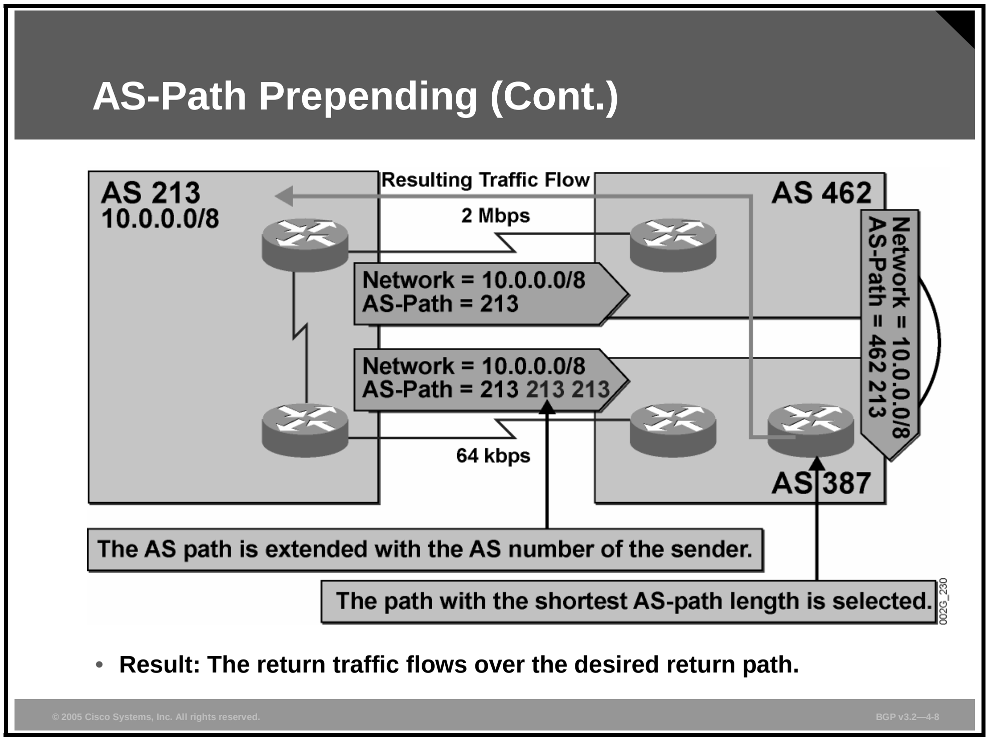
The length of the AS path is extended because additional copies of the AS number of the sender are prepended to (added to the beginning of) the AS-path attribute. To avoid clashes with BGP loop prevention mechanisms, no other AS number, except that of the sending AS, should be prepended to the AS-path attribute.

If another AS number is prepended in the AS path, the routers in the AS that has been prepended will reject the update because of BGP loop prevention mechanisms.

You can configure prepending on a router for all routing updates that you send to a neighbor or only on a subset of them.

Example: AS-Path Prepending

In this example, administrative policy in AS 213 prefers that the low-speed link be used for backup purposes only.



AS-Path Prepending (Cont.)

The AS path is extended with the AS number of the sender.

The path with the shortest AS-path length is selected.

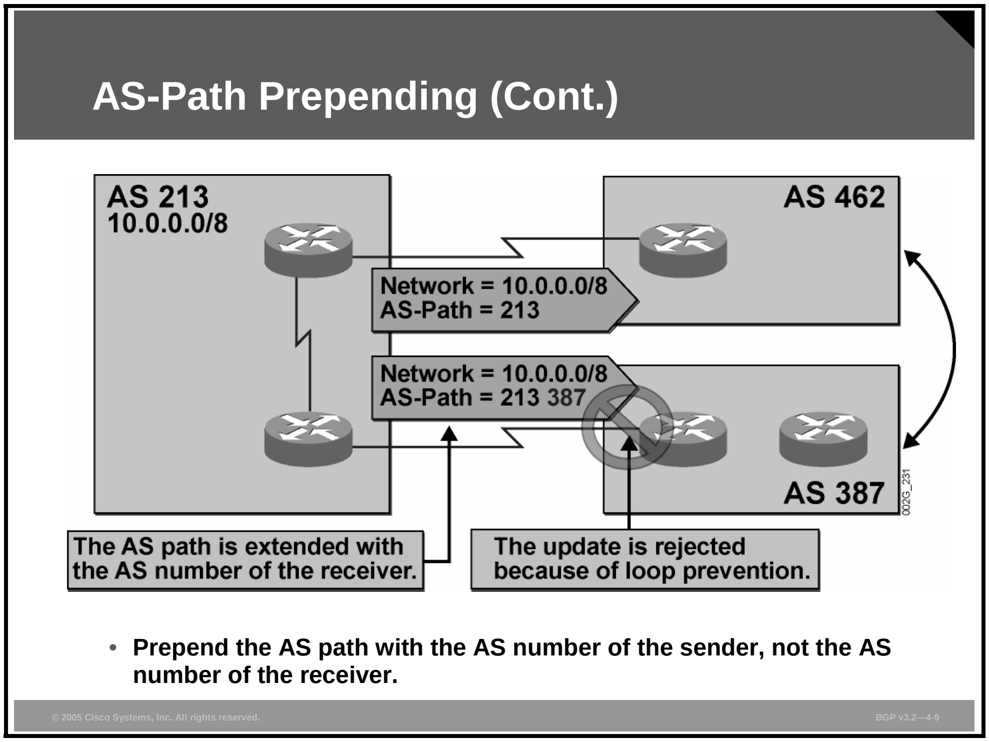
- **Result: The return traffic flows over the desired return path.**

As long as the high-speed link between AS 213 and AS 462 is available, all traffic should flow toward AS 213 using the high-speed link.

To accomplish this goal, you can configure the router in AS 213 that sends EBGP updates to AS 387 by prepending the AS path with two copies of the AS number 213. AS 387 receives two alternative routes to reach network 10.0.0.0/8: the update that it has received directly from AS 213 (that has a manipulated AS path with a length of three) and the update that it has received via AS 462 (that was not manually manipulated and therefore contains an AS-path length of two).

When AS 387 starts the route selection process to determine which route to use to reach network 10.0.0.0/8, it checks the AS-path length after the weight and local preference parameters. In this case, neither weight nor local preference has been configured, so the length of the AS path will be the deciding factor in the route selection process. Consequently, AS 387 prefers the shortest AS path and thus forwards packets toward network 10.0.0.0/8 via AS 462. The desired administrative policy has been met, and AS 213 will receive incoming traffic over the high-speed link.

If the forwarding path from AS 387 via AS 462 to AS 213 and network 10.0.0.0/8 is later broken, the BGP update to reach network 10.0.0.0/8 is revoked. In case of such a network failure, AS 387 will have only one remaining path to reach network 10.0.0.0/8. The route selection process now has only one choice, the route directly to AS 213 over the low-speed WAN link. The low-speed link will therefore serve as backup to the high-speed WAN link.



AS-Path Prepending (Cont.)

AS 213
10.0.0.0/8

AS 462

AS 387

Network = 10.0.0.0/8
AS-Path = 213

Network = 10.0.0.0/8
AS-Path = 213 387

The AS path is extended with the AS number of the receiver.

The update is rejected because of loop prevention.

- **Prepend the AS path with the AS number of the sender, not the AS number of the receiver.**

BGP v3.2—4-9

When you are manually manipulating AS paths, the only valid AS number that you can prepend is the AS number of the sender. Prepending any other AS number will cause problems.

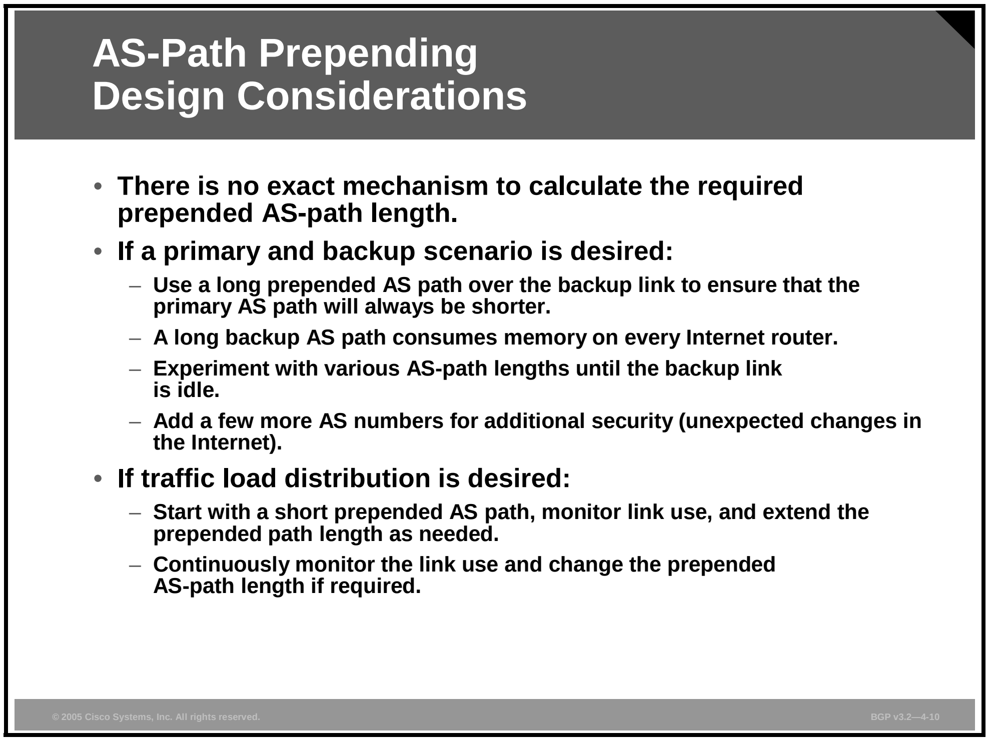
In the example, AS 213 is prepending AS number 387. The egress router performs AS-path prepending when the route is on its way to be transmitted to AS 387. After the manual manipulation, BGP automatically changes the AS path according to the BGP specifications. The local AS number should always be added first when updates are sent over an EBGP session. Therefore, when AS 387 receives the BGP update, the AS path contains the value 213 387. The AS number 387 was set by the manual manipulation, and the AS number 213 was prepended automatically by BGP because the update was sent over an EBGP session.

When the edge router in AS 387 receives the BGP update, it checks the AS path to verify that the BGP updates were not propagated accidentally by a routing loop. Because the edge router finds its own AS number in the AS path, it assumes that the BGP update has already been in AS 387. According to the BGP specification, the update will be silently ignored.

Now assume that AS 213 had, for the manual manipulation, used a different AS number, not its own and not AS number 387. Would AS 387 now have accepted the update? The answer is yes. However, in this scenario, a problem would have appeared at a later stage when the route finally reached the actual AS belonging to the manually prepended AS number. This AS would have rejected the route because it would have found its own AS number somewhere in the AS path.

AS-Path Prepending Design Considerations

This topic identifies design considerations when you are implementing AS-path prepending to influence return path selection.



AS-Path Prepending Design Considerations

- There is no exact mechanism to calculate the required prepended AS-path length.
- If a primary and backup scenario is desired:
  - Use a long prepended AS path over the backup link to ensure that the primary AS path will always be shorter.
  - A long backup AS path consumes memory on every Internet router.
  - Experiment with various AS-path lengths until the backup link is idle.
  - Add a few more AS numbers for additional security (unexpected changes in the Internet).
- If traffic load distribution is desired:
  - Start with a short prepended AS path, monitor link use, and extend the prepended path length as needed.
  - Continuously monitor the link use and change the prepended AS-path length if required.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-10

How many copies of the AS number of the sender should you prepend to the AS path? The answer depends on the goals of the administrative policy. In the general case, it is not easy to determine the exact number of required AS numbers to prepend. The sending AS does not know what alternative paths are available to other autonomous systems.

The following are two typical cases in which you can use AS-path prepending for return path selection:

- **Establishing a primary/backup link:** As an announced backup (prepended) route propagates through the Internet, all the routers along the way that receive the route need to store it together with its AS-path attribute. If this information is long, it will consume extra memory in these routers. However, because routers forward only routes that are selected as best, an AS that receives multiple alternatives to a destination will select the route with the shortest AS path and forward only that route.

  In the case where both the primary and the secondary link are up, the neighboring AS will receive two routes to the same destination that differ only in the AS-path length. The route with the shorter AS path will be subsequently advertised through the Internet.

  In the case where the primary link fails, the route with the longer AS path is the only remaining route. As a result, the primary route is withdrawn, and the prepended route is advertised through the Internet. In this case, extra memory will be consumed in each Internet router because of the storage of the prepended (longer) AS path.

The longer the AS path that is announced to the EBGP neighbor on the other side of the backup link, the less likely it is that incoming traffic will be received from that neighbor. The network administrator can make a clever guess about how many copies of the AS number to prepend. After the prepending is implemented, the network administrator has to examine the result. If the expected result is not achieved, the configuration can be changed and a few more copies of the AS number can be prepended.

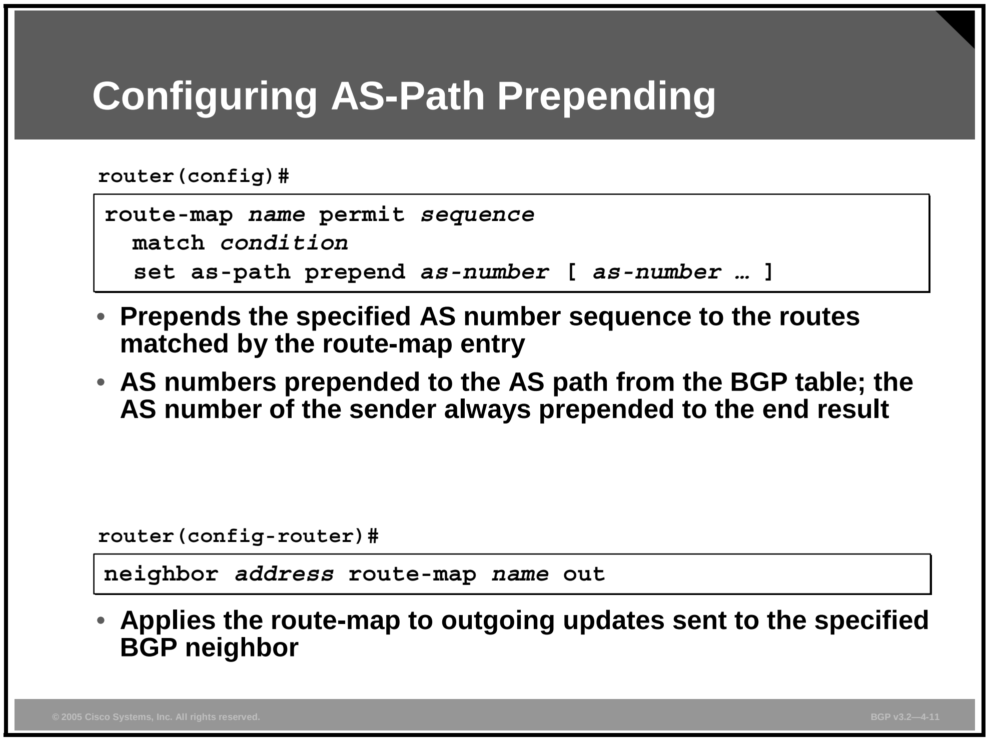
After AS-path prepending has generated the desired results, the network administrator may take the precaution of prepending a few more copies of the AS number to the AS path. This action protects the customer from packets being routed over the backup link at a possible later stage when the topology between remote autonomous systems has unexpectedly changed, yielding a longer AS path to reach the primary link.

- **Distributing the load of return traffic:** In a multihomed scenario, there is no way to exactly predetermine the volume of traffic that will be received over a particular link. The traffic load on different links will change depending on where the senders are located (which autonomous systems they belong to). The network topology and the way that different remote autonomous systems are interconnected may also change with time, changing the load distribution. Only constant monitoring and fine-tuning will ensure that the desired results are achieved.

  In a first attempt at load distribution, the network administrator can configure a router that is connected to an overused link to prepend only a few extra copies of the local AS number. After the network has been given time to converge, the network administrator must check the change in load distribution. Monitoring of the load must be done for a period long enough to be statistically significant (several days or more). If enough volume of traffic has not moved from the overused link to the underused link, the administrator must prepend more copies of the local AS number, and the process of resending local routes and monitoring the results starts all over again.

Configuring AS-Path Prepending

This topic identifies the Cisco IOS commands that are required to configure AS-path prepending in a multihomed network.



Configuring AS-Path Prepending

router(config)#

```
route-map name permit sequence
  match condition
  set as-path prepend as-number [ as-number … ]
```

- **Prepends the specified AS number sequence to the routes matched by the route-map entry**
- **AS numbers prepended to the AS path from the BGP table; the AS number of the sender always prepended to the end result**

router(config-router)#

```
neighbor address route-map name out
```

- **Applies the route-map to outgoing updates sent to the specified BGP neighbor**

BGP v3.2—4-11

You can configure manual manipulation of the AS-path attribute (prepending) using a route-map with the **set as-path prepend** command. The route-map is used to prepend the specified AS numbers to outgoing EBGP route updates that are matched with the match condition, as specified in the route-map. AS-path prepending is completed first, and then the route is subject to the normal AS-path modification procedures when it is sent over an EBGP session.

You can also use the route-map to select only a subset of routes that should have their AS path manually manipulated. Use the **set as-path prepend** command with the appropriate **route-map permit** statement.

Note Changing an outgoing route-map affects only the BGP updates that are sent after the change. To propagate the new and longer AS path with all announced routes, the routes must all be resent by the router. To do this, use the privileged EXEC command **clear ip bgp** with the **soft out** qualifier.

set as-path

To modify an AS path for BGP routes, use the **set as-path** route-map configuration command.

- **set as-path** {**tag** | **prepend** *as-path-string*}

To not modify the AS path, use the **no** form of this command.

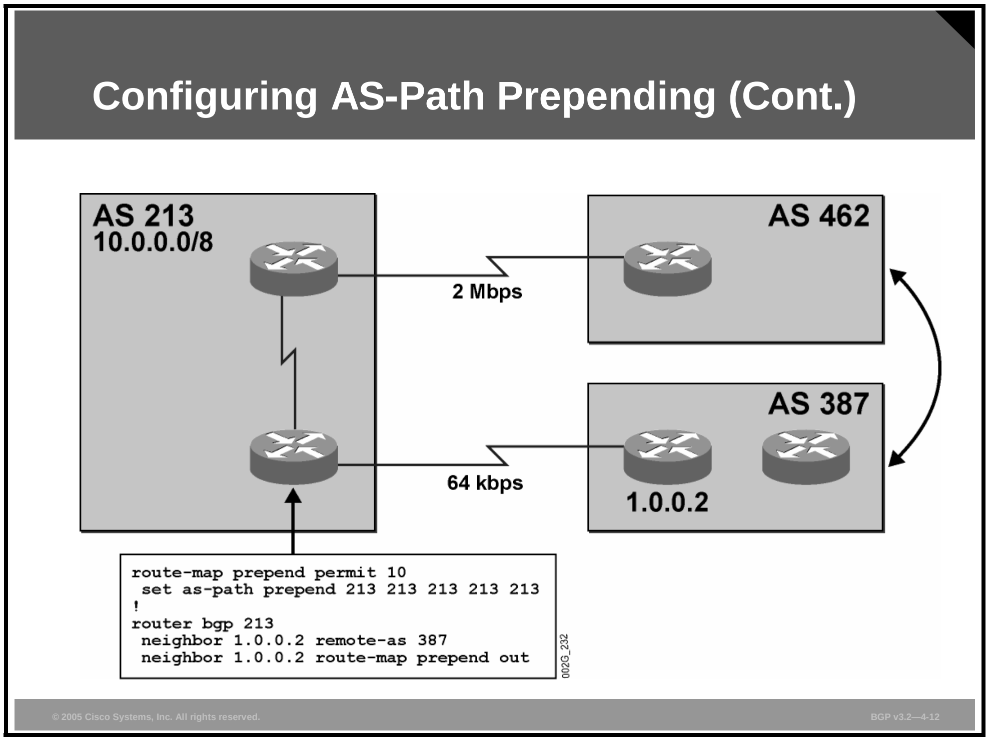
- **no set as-path** {**tag** | **prepend** *as-path-string*}

Syntax Description

Parameter	Description
tag	Converts the tag of a route into an AS path Applies only when redistributing routes into BGP
prepend *as-path-string*	Appends the string that follows the keyword **prepend** to the AS path of the route that is matched by the route-map Applies to inbound and outbound BGP route-maps

Example: Configuring AS-Path Prepending

In this example, the lower router in AS 213 is configured to prepend its own AS number five times for all updates that are sent to the EBGP neighbor 1.0.0.2 in AS 387.



Configuring AS-Path Prepending (Cont.)

```
route-map prepend permit 10
 set as-path prepend 213 213 213 213 213
!
router bgp 213
 neighbor 1.0.0.2 remote-as 387
 neighbor 1.0.0.2 route-map prepend out
```

BGP v3.2—4-12

This configuration will result in AS 387 receiving a route to network 10.0.0.0/8 with an AS path containing the AS number 213 six times (213 213 213 213 213 213). In accordance with BGP specifications, Cisco IOS software automatically adds the sixth copy of the AS number as the route leaves AS 213.

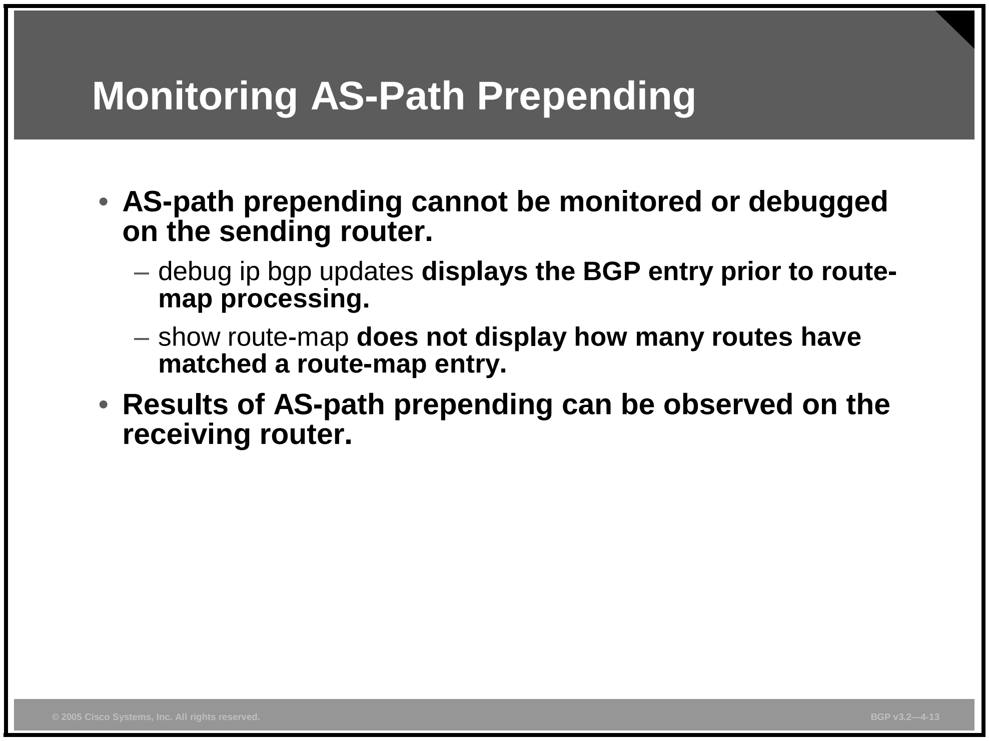
The configuration of the AS 213 router is completed in two steps:

Step 1 First, a route-map named “prepend” is created. The route-map selects all BGP routes and prepends five copies of 213 to the existing AS-path attribute that is already attached to each route. The lack of match conditions in the route-map indicates that all routes are matched.

Step 2 The route-map is applied to all outgoing updates to the EBGP neighbor 1.0.0.2.

Monitoring AS-Path Prepending

This topic identifies the Cisco IOS commands that are required to monitor the operation of AS-path prepending.



Monitoring AS-Path Prepending

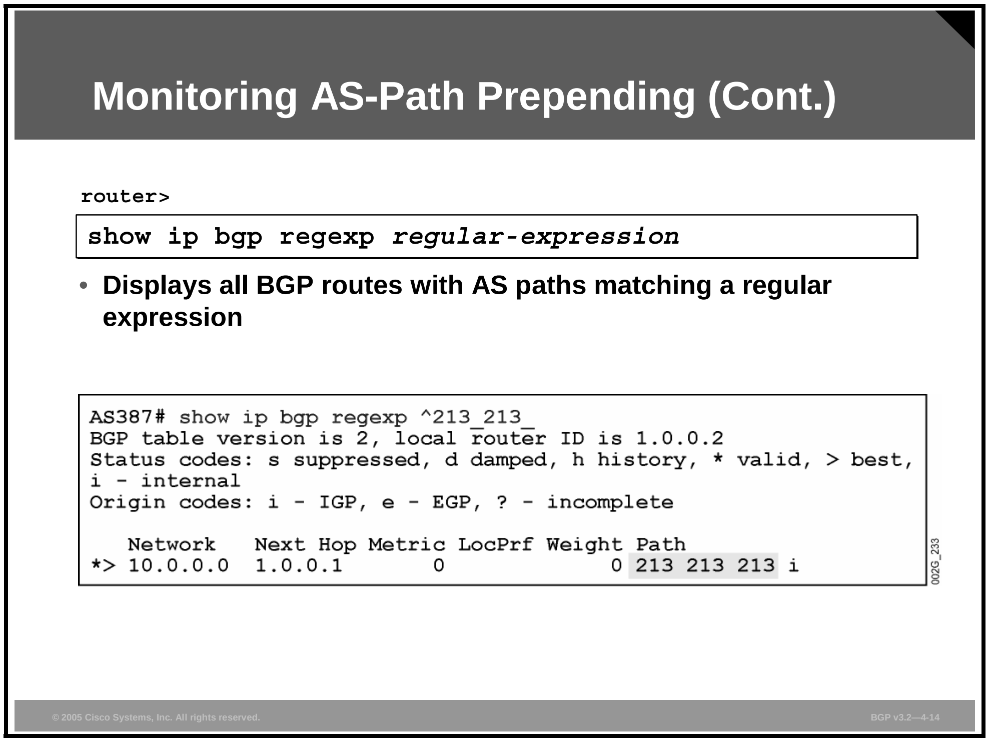
- **AS-path prepending cannot be monitored or debugged on the sending router.**
  - debug ip bgp updates **displays the BGP entry prior to route-map processing.**
  - show route-map **does not display how many routes have matched a route-map entry.**
- **Results of AS-path prepending can be observed on the receiving router.**

BGP v3.2—4-13

When you are monitoring AS-path prepending, the router doing the prepending is not the proper point to observe the results of the AS-path prepend operation. For instance, output from the **debug ip bgp updates** command does not display the prepended paths, because the route-map doing the prepending is applied afterward.

The **show route-map** command displays the configuration details of a route-map. The matching criteria and AS-path manipulation are displayed as output of the command. However, there is no indication of how many routes have been matched by a route-map statement and thus had their AS paths manipulated.

A better place for observing AS-path prepending is on the router receiving the BGP update that contains the prepended AS path. At that point, you can use the pattern of AS number sequences in the received AS-path attribute of received routes to find the routes that have a prepended AS path.



Monitoring AS-Path Prepending (Cont.)

```
router>
show ip bgp regexp regular-expression
```

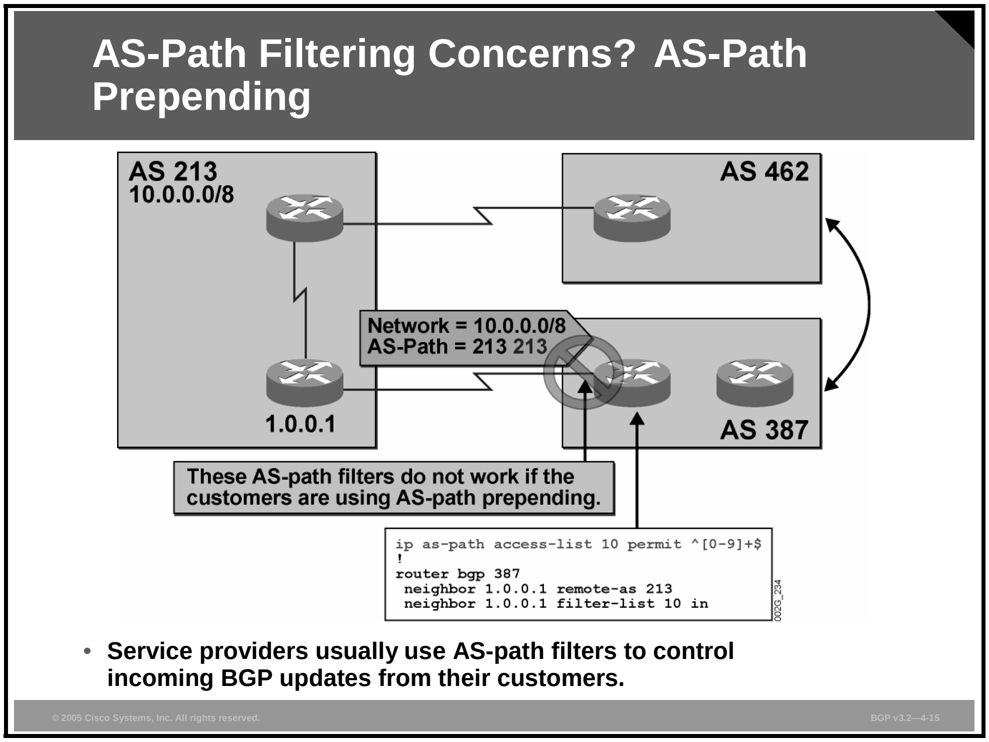
- **Displays all BGP routes with AS paths matching a regular expression**

```
AS387# show ip bgp regexp ^213_213_
BGP table version is 2, local router ID is 1.0.0.2
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
*> 10.0.0.0  1.0.0.1       0             0 213 213 213 i
```

In the figure, the **show ip bgp regexp** command is used to find all the routes in the BGP table of the receiving router in AS 387 that are directly received from AS 213 and have at least one extra copy of AS number 213 in their AS path. Network 10.0.0.0/8 is displayed as the only route meeting this selection criterion. The AS path has been prepended with two extra copies of AS 213. The egress router in AS 213 automatically added the third copy of AS 213 because the route was sent across an EBGP session.

AS-Path Filtering Concerns with AS-Path Prepending

This topic describes the concerns of using AS-path filters when neighboring autonomous systems require AS-path prepending.



AS-Path Filtering Concerns? AS-Path Prepending

These AS-path filters do not work if the customers are using AS-path prepending.

```
ip as-path access-list 10 permit ^[0-9]+$
!
router bgp 387
 neighbor 1.0.0.1 remote-as 213
 neighbor 1.0.0.1 filter-list 10 in
```

- **Service providers usually use AS-path filters to control incoming BGP updates from their customers.**

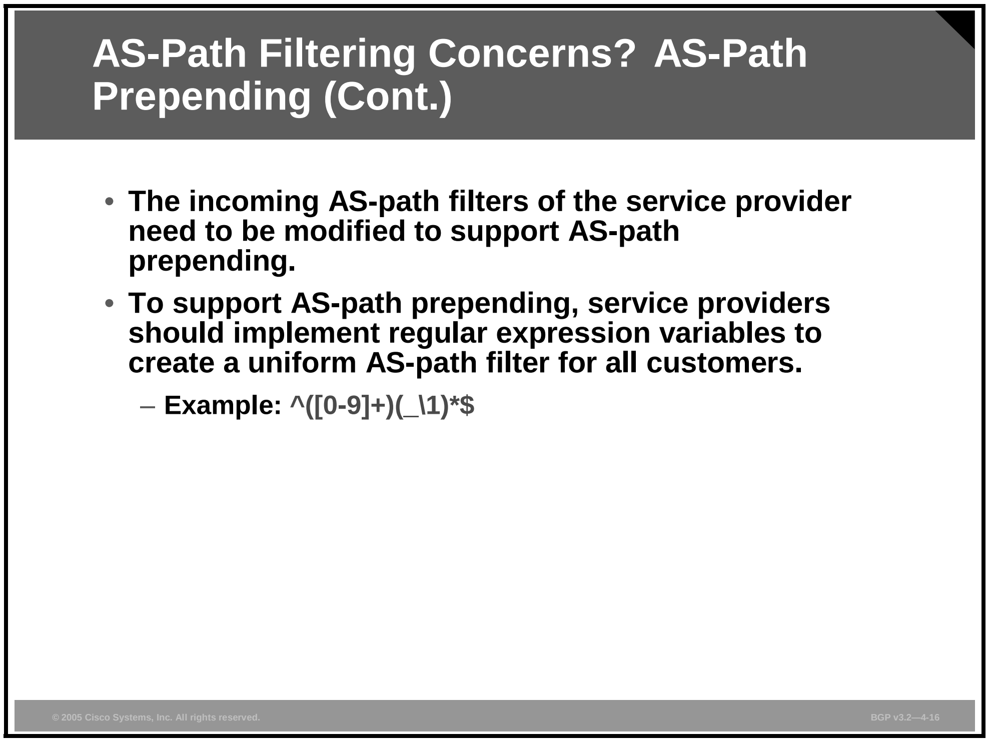
Service providers normally expect their customers to send routes that originate only in the AS of the customer. However, because customers might not do so, proactive thinking and care for the rest of the Internet cause the service provider to implement AS-path filters on incoming updates that are received from their customers.

The network administrator of the service provider in the figure could configure individual filters for each neighbor. However, a single AS-path access-list permitting only AS paths with a length of exactly one AS number would be a better solution because the service provider can uniformly apply it to all incoming routes from all customers (possibly using a peer group).

In the figure, the service provider (AS 387) has configured a filter-list, which allows only AS paths that have a length of one AS number. When the customer changes its router configuration and starts to announce network 10.0.0.0/8 with a prepended AS path, the filter-list for incoming routes to AS 387 in the service provider router will filter those routes out. This filtering results in a situation where the network 10.0.0.0/8 is not reachable over the link between AS 213 and AS 387. Therefore, the backup function is not available.

Network 10.0.0.0/8 is, however, still reachable via the path going through AS 462. This situation means that AS 387 can send packets to network 10.0.0.0/8 but not over the direct link to AS 213. This failure may be hard to detect because, during normal conditions, all autonomous systems in the figure can exchange traffic.

After AS 387 loses the route to network 10.0.0.0/8 via AS 462, possibly because the primary link between AS 213 and AS 462 is gone, the problem will be obvious. AS 387 can now no longer reach network 10.0.0.0/8 at all, although the physical link between AS 213 and AS 387 is available.



AS-Path Filtering Concerns? AS-Path Prepending (Cont.)

- **The incoming AS-path filters of the service provider need to be modified to support AS-path prepending.**
- **To support AS-path prepending, service providers should implement regular expression variables to create a uniform AS-path filter for all customers.**
  - **Example: ^([0-9]+)(_\1)*$**

BGP v3.2—4-16

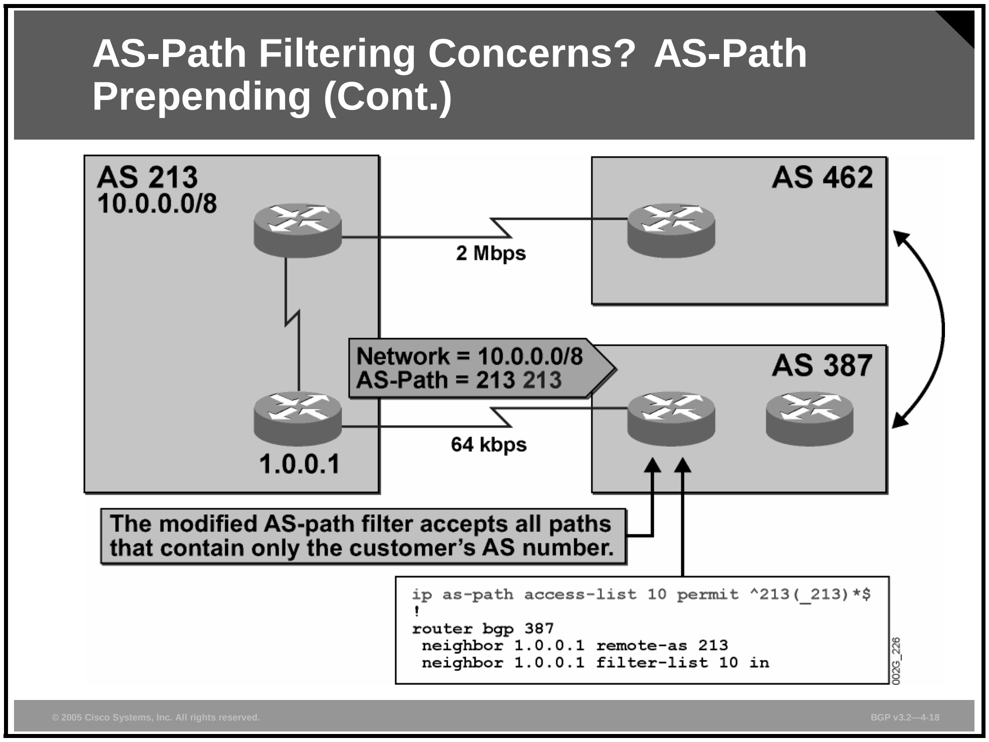
Because the AS of the service provider will receive customer routes with prepended AS-paths that have a length greater than one AS number, the provider must modify its incoming filters.

The service provider needs to create a new inbound regular expression filter, using regular expression variables and parentheses for recall.

What is needed is a filter that will allow any AS path containing one or multiple copies of the same AS number. An example of such a filter is as follows:

- ^([0-9]+)(_\1)*$

This filter matches any AS path beginning with any AS number and continues with no or multiple repetitions of that same AS number (the variable "\1" repeats the value in the brackets). The regular expression would therefore match AS paths 99 99 99, 2 2 2, or 100, but it would not match AS path 100 99.



AS-Path Filtering Concerns? AS-Path Prepending (Cont.)

The modified AS-path filter accepts all paths that contain only the customer’s AS number.

```
ip as-path access-list 10 permit ^213(_213)*$
!
router bgp 387
 neighbor 1.0.0.1 remote-as 213
 neighbor 1.0.0.1 filter-list 10 in
```

BGP v3.2—4-18

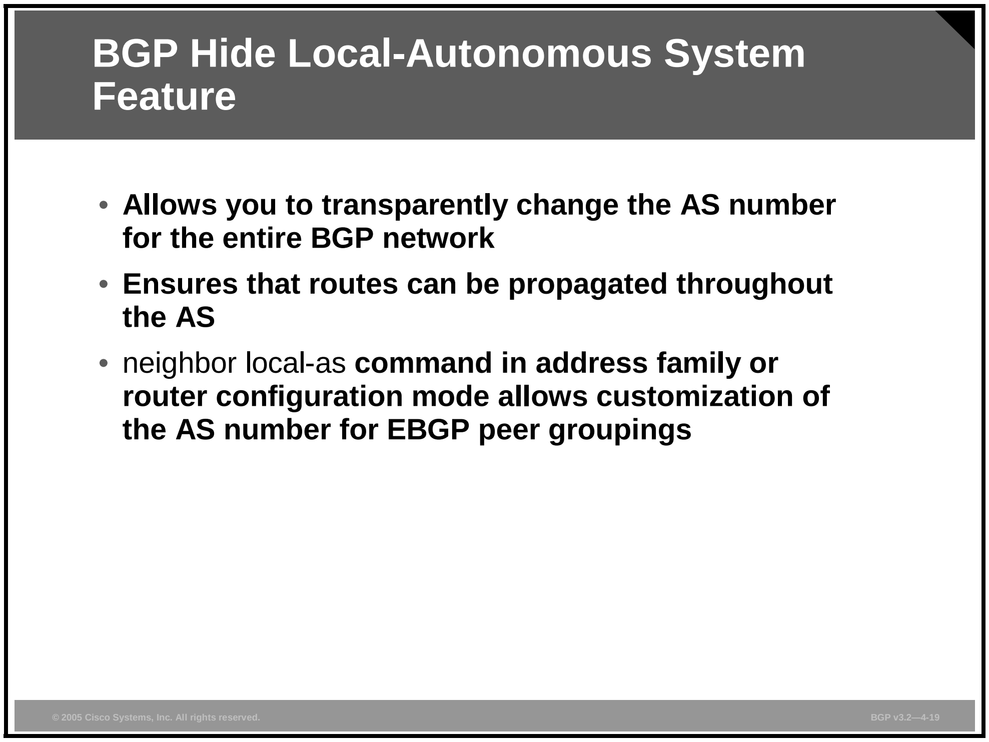
In the figure, the service provider (AS 387) has configured an individual filter for all routes that are received directly from AS 213. The AS path is required to start with 213. Then multiple copies of 213 may follow it. The asterisk allows for zero occurrences, permitting the AS path with a single copy of 213 as well.

If the same service provider router has more customers that are attached to it, they will all require an individual filter-list because the AS number of the customer is explicitly indicated in the regular expression.

An alternative would be to implement the AS-path filter by using regular expression variables.

BGP Hide Local-Autonomous System

This topic describes how the BGP Hide Local-Autonomous System feature simplifies the task of changing the AS number in a BGP network.



BGP Hide Local-Autonomous System Feature

- **Allows you to transparently change the AS number for the entire BGP network**
- **Ensures that routes can be propagated throughout the AS**
- neighbor local-as **command in address family or router configuration mode allows customization of the AS number for EBGP peer groupings**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-19

Changing the AS number may be necessary when two separate BGP networks are combined under a single AS, a situation that typically occurs when one ISP purchases another ISP. Changing the AS number in a BGP network can be difficult because, during the transition, Internal Border Gateway Protocol (IBGP) peers will reject external routes from peers with a local AS number in the AS number path to prevent routing loops. The BGP Hide Local-Autonomous System feature allows you to transparently change the AS number for the entire BGP network and ensure that routes can be propagated throughout the AS, while the AS number transition is incomplete.

neighbor local-as

To allow customization of the AS number for EBGP peer groupings, use the **neighbor local-as** command in address family or router configuration mode.

- **neighbor** {*ip-address* | *peer-group-name*} **local-as** *as-number* [**no-prepend**]

To disable this function, use the **no** form of this command.

- **no neighbor** {*ip-address* | *peer-group-name*} **local-as** *as-number*

Syntax Description

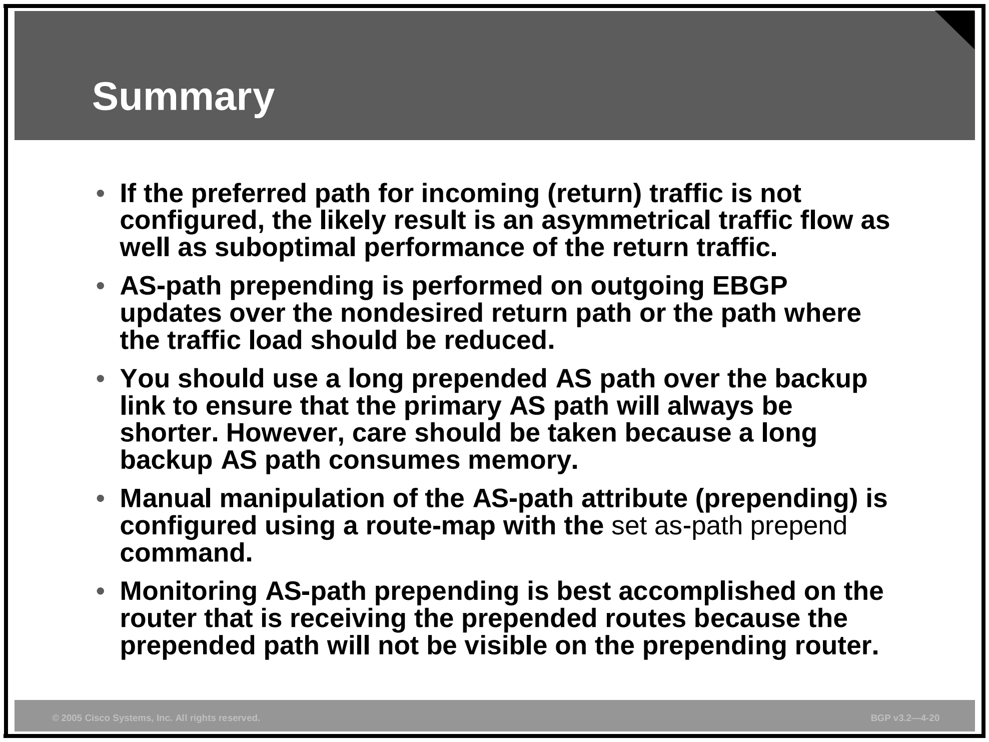
Parameter	Description
ip-address	IP address of the local BGP-speaking neighbor.
peer-group-name	Name of a BGP peer group.
as-number	Valid AS number from 1 to 65535. Do not specify the AS number to which the neighbor belongs.
no-prepend	Configures the router to not prepend the local AS number to any routes received from an external peer.

The **neighbor local-as** command is used initially to configure BGP peers to support two local AS numbers to maintain peering between two separate BGP networks. This configuration allows the ISP to immediately make the transition without any impact on existing customer configurations. When the customer configurations have been updated, the next step is to complete the transition from the old AS number to the new AS number.

When the **neighbor local-as** command is configured on a BGP peer, the local AS number is automatically prepended to all routes that are learned from EBGP peers by default. This behavior, however, makes changing the AS number for a service provider or large BGP network difficult because routes with the prepended AS number will be rejected by IBGP peers that are configured with the same AS number. For example, if you configure an IBGP peer with the **neighbor 10.0.0.2 local-as 20** statement, all routes that are learned from the 10.0.0.2 external peer will automatically have the AS number 20 prepended. Internal routers that are configured with the AS number 20 will detect these routes as routing loops and reject them. This behavior requires you to change the AS number for all IBGP peers at the same time.

Summary

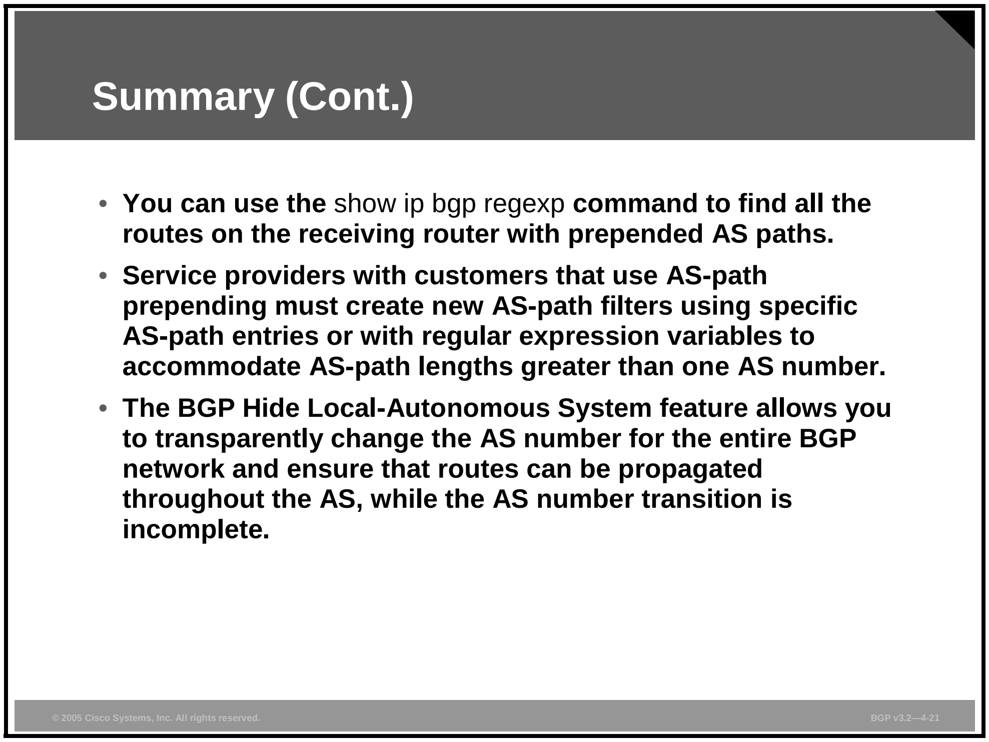
This topic summarizes the key points discussed in this lesson.



Summary

- **If the preferred path for incoming (return) traffic is not configured, the likely result is an asymmetrical traffic flow as well as suboptimal performance of the return traffic.**
- **AS-path prepending is performed on outgoing EBGP updates over the nondesired return path or the path where the traffic load should be reduced.**
- **You should use a long prepended AS path over the backup link to ensure that the primary AS path will always be shorter. However, care should be taken because a long backup AS path consumes memory.**
- **Manual manipulation of the AS-path attribute (prepending) is configured using a route-map with the** set as-path prepend **command.**
- **Monitoring AS-path prepending is best accomplished on the router that is receiving the prepended routes because the prepended path will not be visible on the prepending router.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-20



Summary (Cont.)

- **You can use the** show ip bgp regexp **command to find all the routes on the receiving router with prepended AS paths.**
- **Service providers with customers that use AS-path prepending must create new AS-path filters using specific AS-path entries or with regular expression variables to accommodate AS-path lengths greater than one AS number.**
- **The BGP Hide Local-Autonomous System feature allows you to transparently change the AS number for the entire BGP network and ensure that routes can be propagated throughout the AS, while the AS number transition is incomplete.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-21

Understanding BGP Multi-Exit Discriminators

Overview

When connections to multiple providers are required, it is important that Border Gateway Protocol (BGP) select the optimum route for traffic to use. It is equally important that the return path that is selected be the optimum return path into the autonomous system (AS). The optimum, or best, route may not be what the network designer intended based on design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides a tool for administrators to influence route selection, the multi-exit discriminator (MED) attribute.

This lesson discusses how to influence BGP route selection by setting the BGP MED attribute of outgoing BGP routes. Two methods that are used to set the MED attribute, the default MED and route-maps, are discussed in this lesson. In addition to basic MED attribute configuration, advanced commands to manipulate MED properties are discussed, as well as how to monitor and troubleshoot the BGP table to verify correct MED configuration and to properly influence path selection.

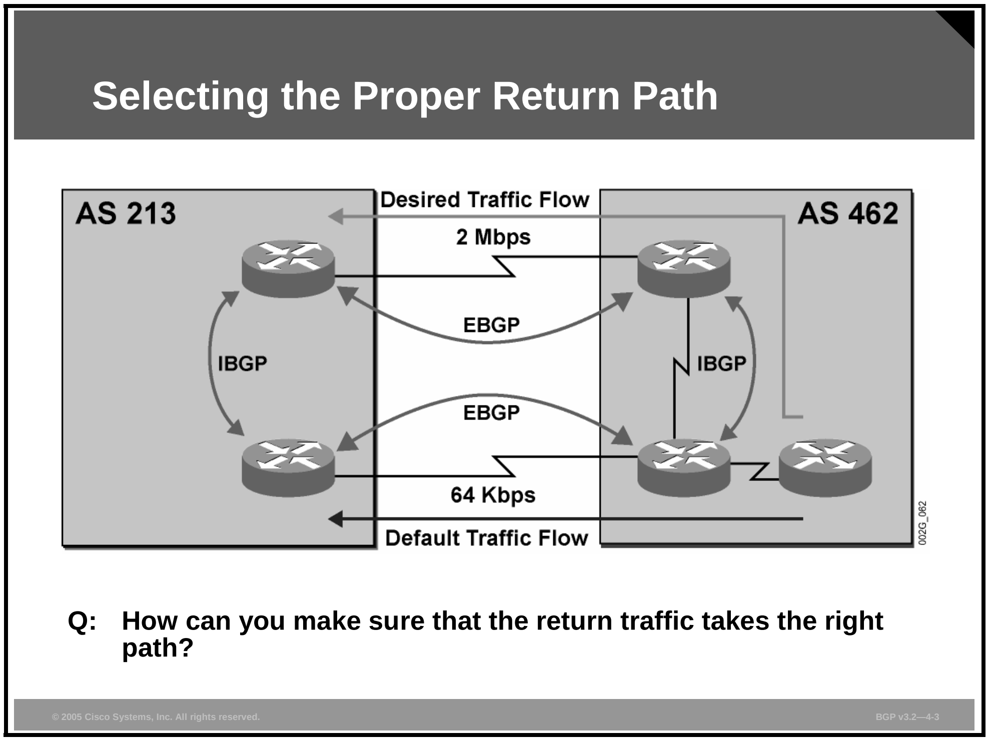
Objectives

Upon completing this lesson, you will be able to use the MED attribute to influence route selection in a customer scenario in which you must support multiple ISP connections. This ability includes being able to meet these objectives:

- Describe how the MED can be used to facilitate proper return path selection
- Explain how the value of the MED attribute changes inside a BGP AS and between different BGP autonomous systems
- Identify the Cisco IOS command that is required to configure changes to the default BGP MED on a Cisco IOS router
- Identify the Cisco IOS commands that are required to configure the BGP MED using route-maps
- Identify the Cisco IOS commands that are required to configure advanced MED features on Cisco routers
- Identify the Cisco IOS commands that are required to monitor the BGP MED
- Identify the Cisco IOS commands that are required to troubleshoot the BGP MED configurations

Selecting the Proper Return Path

This topic describes how you can use the MED attribute to facilitate proper return path selection.

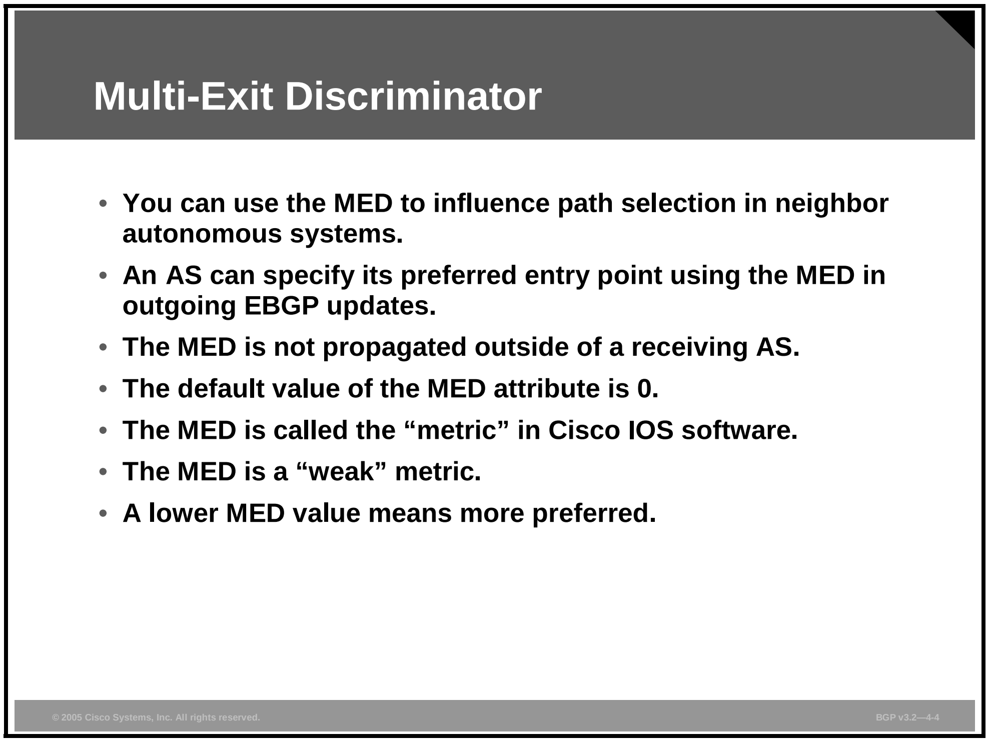


Selecting the Proper Return Path

Q: How can you make sure that the return traffic takes the right path?

BGP v3.2—4-3

When multiple connections between providers are required, BGP attributes such as weight and local preference solve only half the problem: how to choose the right path out of the AS. This topic focuses on the second, more complex half of the problem: how to influence neighboring autonomous systems to choose the correct return path back into the AS.



Multi-Exit Discriminator

- You can use the MED to influence path selection in neighbor autonomous systems.
- An AS can specify its preferred entry point using the MED in outgoing EBGP updates.
- The MED is not propagated outside of a receiving AS.
- The default value of the MED attribute is 0.
- The MED is called the “metric” in Cisco IOS software.
- The MED is a “weak” metric.
- A lower MED value means more preferred.

BGP v3.2—4-4

The MED attribute is a hint to external neighbors about the preferred path into an AS when multiple entry points exist. You can apply the MED attribute on outgoing updates to a neighboring AS to influence the route selection process in that AS. The MED attribute is useful only when there are multiple entry points into an AS.

The MED attribute, which is sent to an external neighbor, will be seen only within that AS. An AS that receives a route that contains the MED attribute will not advertise that MED beyond its local AS.

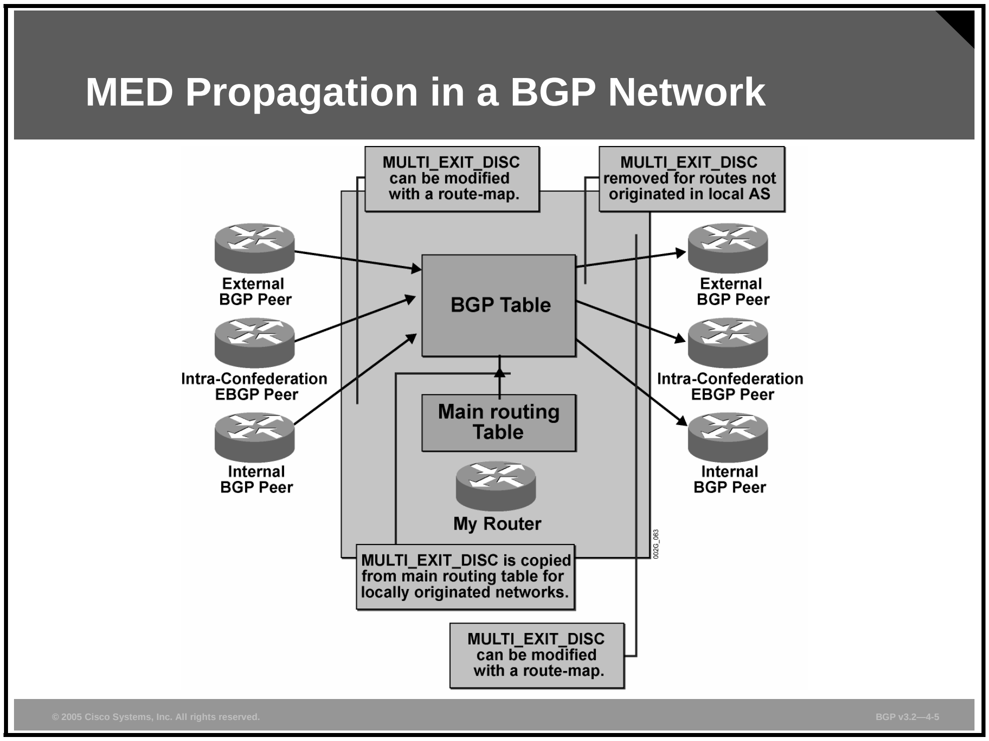
The default value of the MED attribute is 0. A lower value of the MED attribute indicates a more preferred path.

The MED attribute is considered a “weak” metric. In contrast with weight and local preference, a router will prefer a path with the smallest MED value but only if the weight, local preference, AS-path, and origin code attributes are equal. Using the MED may not yield the expected result if the neighboring AS modifies any of the stronger BGP route selection mechanisms.

Note The term that is used in Cisco IOS software for MED is “metric.” The term “metric” also applies to the **set** command that is used in route-maps as well as all **show** and **debug** commands.

MED Propagation in a BGP Network

This topic explains how the value of the MED attribute changes inside a BGP AS and between different BGP autonomous systems.



MED Propagation in a BGP Network

MULTI_EXIT_DISC can be modified with a route-map.

MULTI_EXIT_DISC removed for routes not originated in local AS

MULTI_EXIT_DISC is copied from main routing table for locally originated networks.

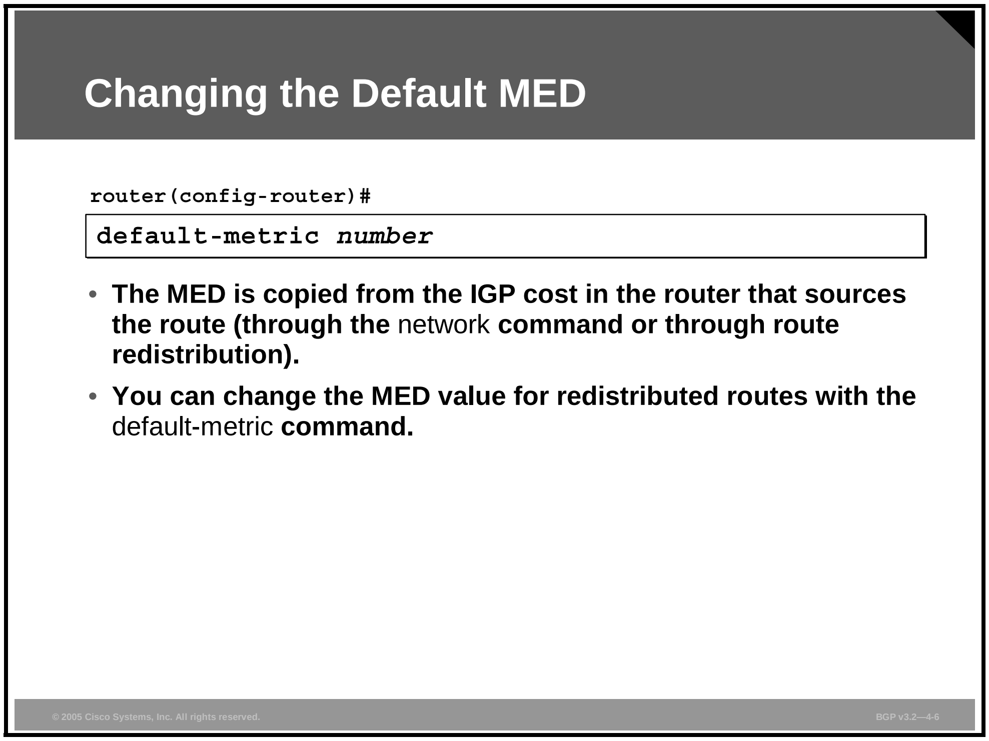
MULTI_EXIT_DISC can be modified with a route-map.

BGP v3.2—4-5

The figure shows how the value of the MED attribute is assigned, depending upon the routing information source. A route-map must be configured on a router to manually assign a value to the MED attribute. For the networks that are also present in the BGP table, the router assigns a default value from the metric in the routing table and copies it into the MED attribute. The MED attribute is automatically removed on external sessions if the attribute did not originate in the local AS.

Changing the Default MED

This topic describes the Cisco IOS command that is required to configure changes to the default BGP MED on a Cisco IOS router.



Changing the Default MED

```
router(config-router)#
default-metric number
```

- **The MED is copied from the IGP cost in the router that sources the route (through the** network **command or through route redistribution).**
- **You can change the MED value for redistributed routes with the** default-metric **command.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-6

The MED is not a mandatory attribute, and there is no MED attribute that is attached to a route by default. The only exception is if the router is originating networks that have an exact match in the routing table (through the **network** command or through redistribution). In that case, the router uses the metric in the routing table as the MED attribute value.

Using the **default-metric** command in BGP configuration mode causes all redistributed networks to have the specified MED value.

default-metric (BGP)

To set the default metric (MED) value for BGP routes, use the **default-metric** command in router configuration mode.

- **default-metric** *number*

To return to the default state, use the **no** form of this command.

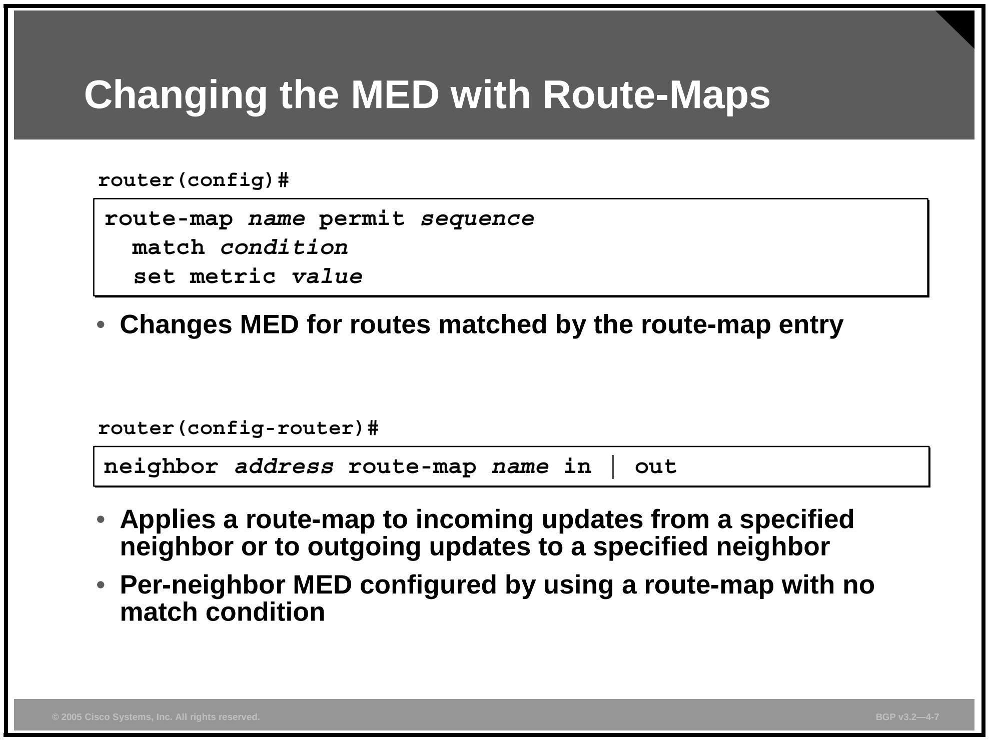
- **no default-metric** *number*

Syntax Description

Parameter	Description
number	Default metric value that is appropriate for the specified routing protocol

Changing the MED with Route-Maps

This topic lists the Cisco IOS commands that are required to configure changes to the BGP MED attribute with route-map statements.



Changing the MED with Route-Maps

```
router(config)#
```

```
route-map name permit sequence
  match condition
  set metric value
```

- Changes MED for routes matched by the route-map entry

```
router(config-router)#
```

```
neighbor address route-map name in | out
```

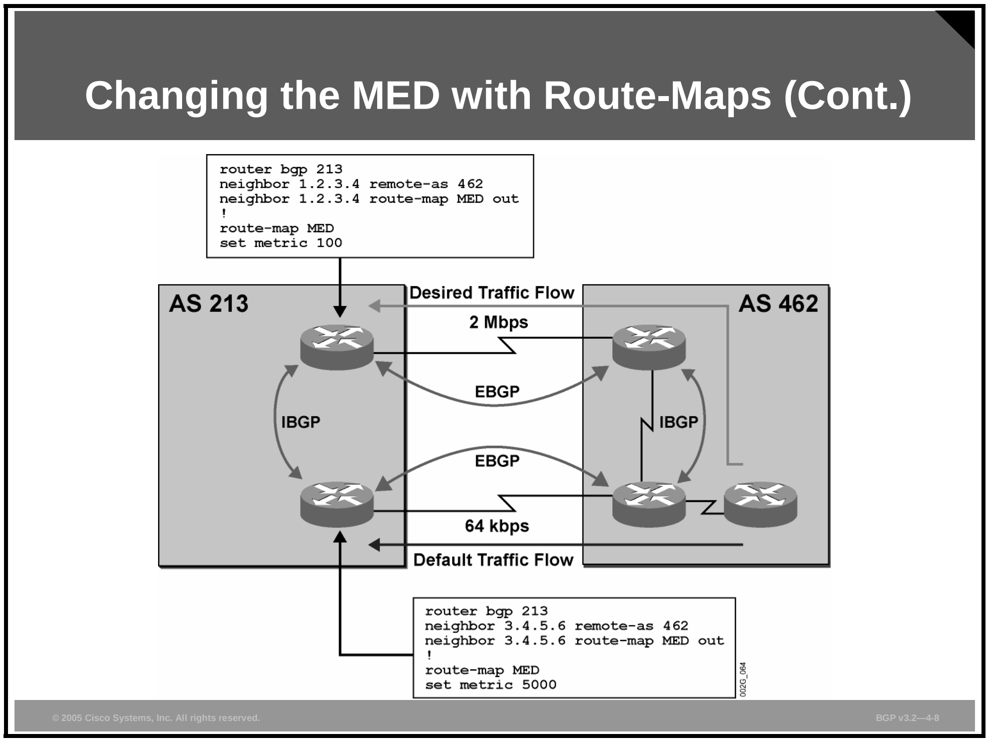
- Applies a route-map to incoming updates from a specified neighbor or to outgoing updates to a specified neighbor
- Per-neighbor MED configured by using a route-map with no match condition

BGP v3.2—4-7

You can use a route-map to set the MED on incoming or outgoing updates. Use the **set metric** command within route-map configuration mode to set the MED attribute.

Example: Changing the MED with Route-Maps

The example shows how to set a per-neighbor MED on an outgoing update.



Changing the MED with Route-Maps (Cont.)

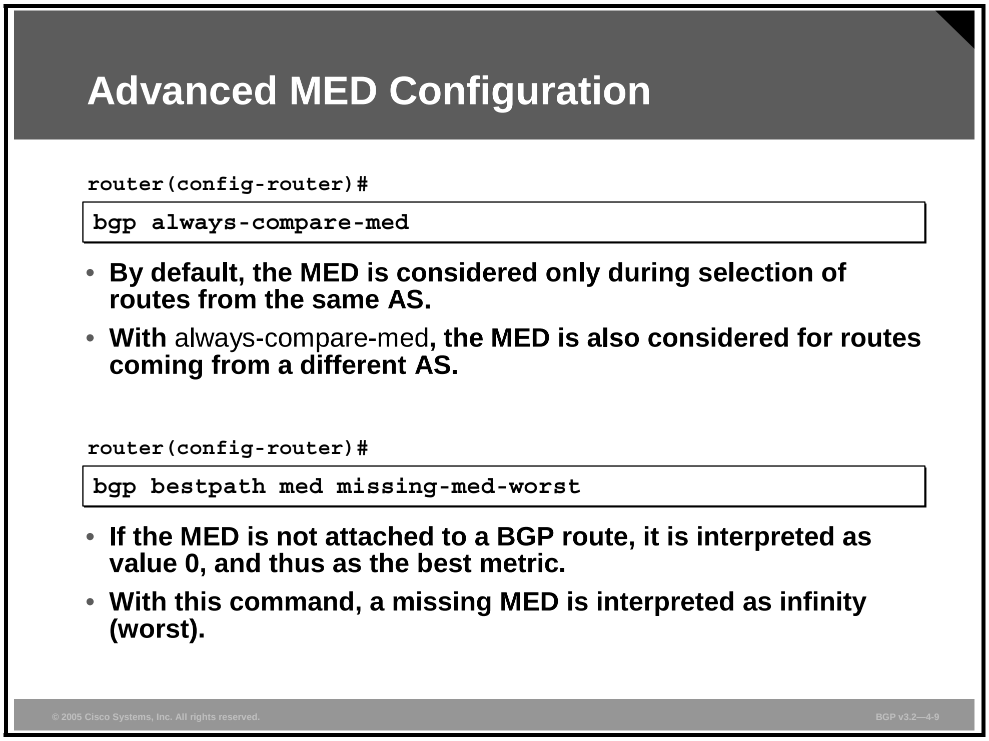
```
router bgp 213
neighbor 1.2.3.4 remote-as 462
neighbor 1.2.3.4 route-map MED out
!
route-map MED
set metric 100
```

```
router bgp 213
neighbor 3.4.5.6 remote-as 462
neighbor 3.4.5.6 route-map MED out
!
route-map MED
set metric 5000
```

The result of this action is that the neighboring AS will prefer the upper link to AS 213. The solution, of course, relies on the neighboring AS not changing the weight, local preference, AS-path, or origin code attributes in updates that it receives from AS 213.

Advanced MED Configuration

This topic lists the Cisco IOS commands that are required to configure advanced MED features on Cisco routers.



Advanced MED Configuration

```
router(config-router)#
bgp always-compare-med
```

- **By default, the MED is considered only during selection of routes from the same AS.**
- **With** always-compare-med**, the MED is also considered for routes coming from a different AS.**

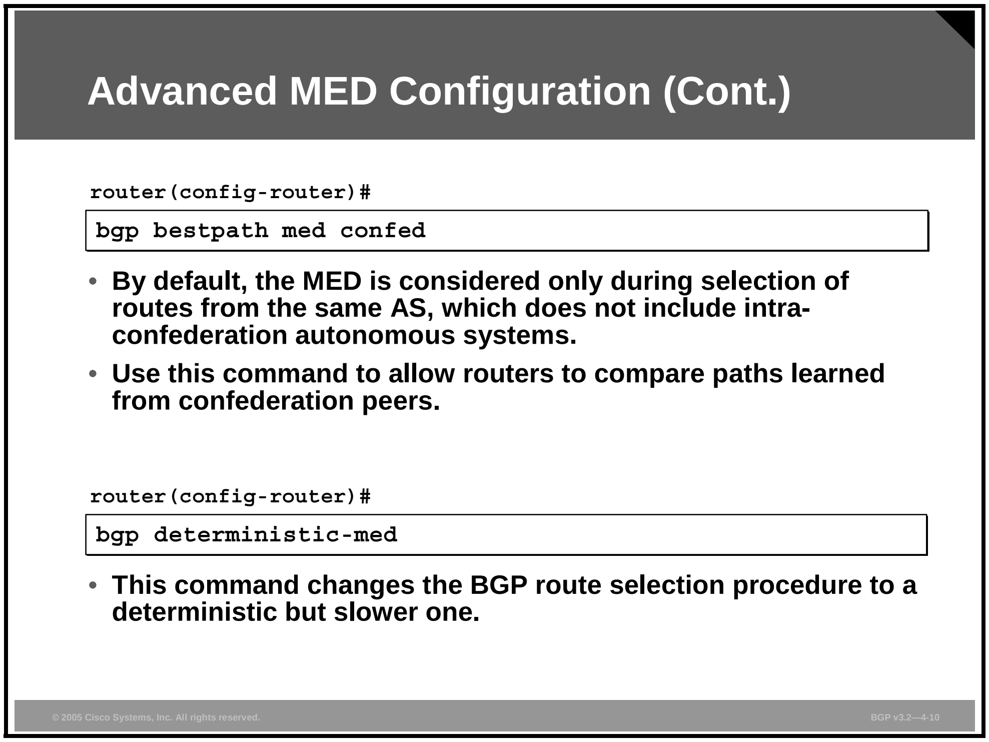
```
router(config-router)#
bgp bestpath med missing-med-worst
```

- **If the MED is not attached to a BGP route, it is interpreted as value 0, and thus as the best metric.**
- **With this command, a missing MED is interpreted as infinity (worst).**

BGP v3.2—4-9

Several rules exist on when and how you should use the MED attribute:

- You should use the MED in the route selection process only if both (all) paths come from the same AS. Use the **bgp always-compare-med** command to force the router to compare the MED even if the paths come from different autonomous systems. You need to enable this option in the entire AS; otherwise, routing loops can occur.
- According to a BGP standard describing MED, you should regard a missing MED attribute as an infinite value. Cisco IOS software, on the other hand, regards a missing MED attribute as having a value of 0. Use the **bgp bestpath med missing-med-worst** command when combining equipment from different vendors. An even better solution is to make sure that every update carries a MED attribute.



Advanced MED Configuration (Cont.)

router(config-router)#

```
bgp bestpath med confed
```

- By default, the MED is considered only during selection of routes from the same AS, which does not include intra-confederation autonomous systems.
- Use this command to allow routers to compare paths learned from confederation peers.

router(config-router)#

```
bgp deterministic-med
```

- This command changes the BGP route selection procedure to a deterministic but slower one.

BGP v3.2—4-10

You must use the **bgp bestpath med confed** command when you use the MED within a confederation to influence the route selection process. A router will compare MED values for the routes that originate in the confederation.

When you enable a deterministic MED comparison, you allow a router to compare MED values before it considers BGP route type (external or internal) and Interior Gateway Protocol (IGP) metric to the next-hop address. The router will compare MED values immediately after the AS-path length.

Note Cisco recommends enabling the **bgp deterministic-med** command in all new network rollouts. For existing networks, you must deploy the command either on all routers at the same time or incrementally, with care to avoid possible IBGP routing loops.

Example: Advanced MED Configuration

The following example demonstrates how the **bgp deterministic-med** and **bgp always-compare-med** commands can influence MED-based path selection. Consider the following BGP routes for network 172.16.0.0/16 in the order that they are received:

```
entry 1: AS(PATH) 65500, med 150, external, rid 192.168.13.1
entry 2: AS(PATH) 65100, med 200, external, rid 1.1.1.1
entry 3: AS(PATH) 65500, med 100, internal, rid 192.168.8.4
```

Note BGP compares multiple routes to a single destination in pairs, starting with the newest entry and moving toward the oldest entry (starting at the top of the list and moving down). For example, entry 1 and entry 2 are compared. The better of these two is then compared to entry 3, and so on.

In the case where both commands are disabled, BGP compares entry 1 and entry 2. Entry 2 is chosen as the better of these two because it has a lower router-ID. The MED is not checked because the paths are from a different neighbor AS. Next, entry 2 is compared to entry 3. BGP chooses entry 2 as the best path because it is external.

In the case where the **bgp deterministic-med** command is disabled and the **bgp always-compare-med** command has been enabled, BGP compares entry 1 to entry 2. These entries are from different autonomous systems, but because the **bgp always-compare-med** command is enabled, the MED is used in the comparison. Entry 1 is the better of these two entries because it has a lower MED value. Next, BGP compares entry 1 to entry 3. The MED is checked again because the entries are now from the same AS. BGP chooses entry 3 as the best path.

In the case where the **bgp deterministic-med** command has been enabled and the **bgp always-compare-med** command has been disabled, BGP groups routes from the same AS together and compares the best entries of each group. The BGP table looks like the following:

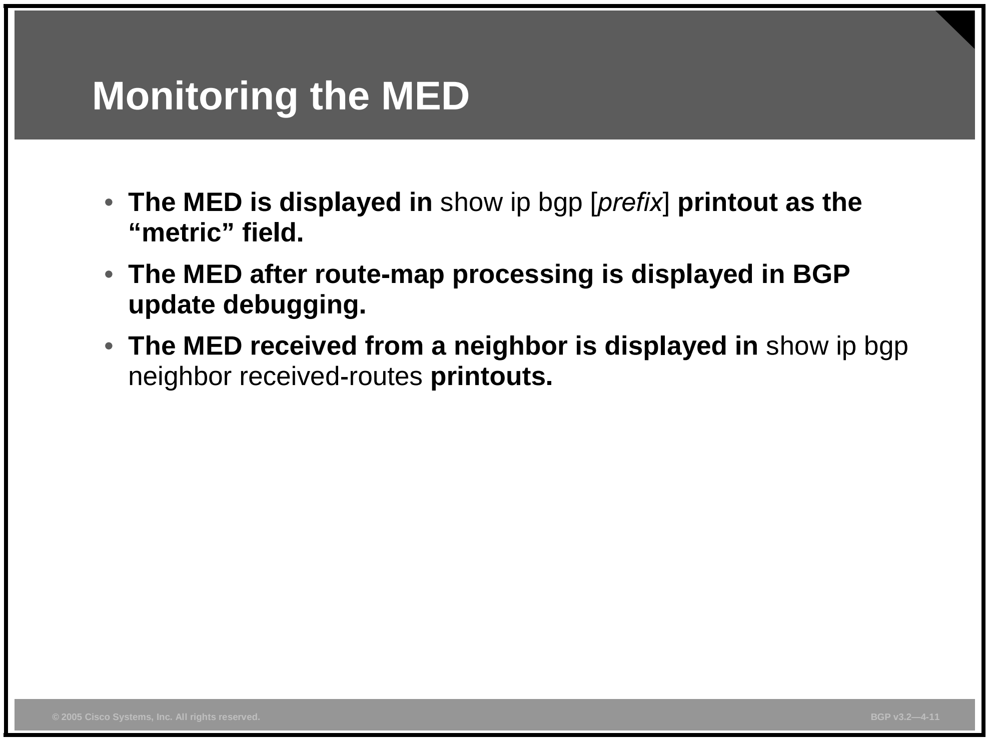
```
entry 1: AS(PATH) 65100, med 200, external, rid 1.1.1.1
entry 2: AS(PATH) 65500, med 100, internal, rid 192.168.8.4
entry 3: AS(PATH) 65500, med 150, external, rid 192.168.13.1
```

There is a group for AS 65100 and a group for AS 65500. BGP compares the best entries for each group. Entry 1 is the best of its group because it is the only route from AS 100. BGP compares entry 1 to the best of group AS 65500, entry 2 (because it has the lowest MED). Because the two entries are not from the same neighbor AS, the MED is not considered in the comparison. The EBGP route wins over the IBGP route, making entry 1 the best route.

If the **bgp always-compare-med** command were also enabled, BGP would have taken the MED into account for the last comparison and have selected entry 2 as the best path.

Monitoring the MED

This topic lists the Cisco IOS commands that are required to monitor the BGP MED attribute on a Cisco router.



Monitoring the MED

- **The MED is displayed in** show ip bgp [*prefix*] **printout as the "metric" field.**
- **The MED after route-map processing is displayed in BGP update debugging.**
- **The MED received from a neighbor is displayed in** show ip bgp neighbor received-routes **printouts.**

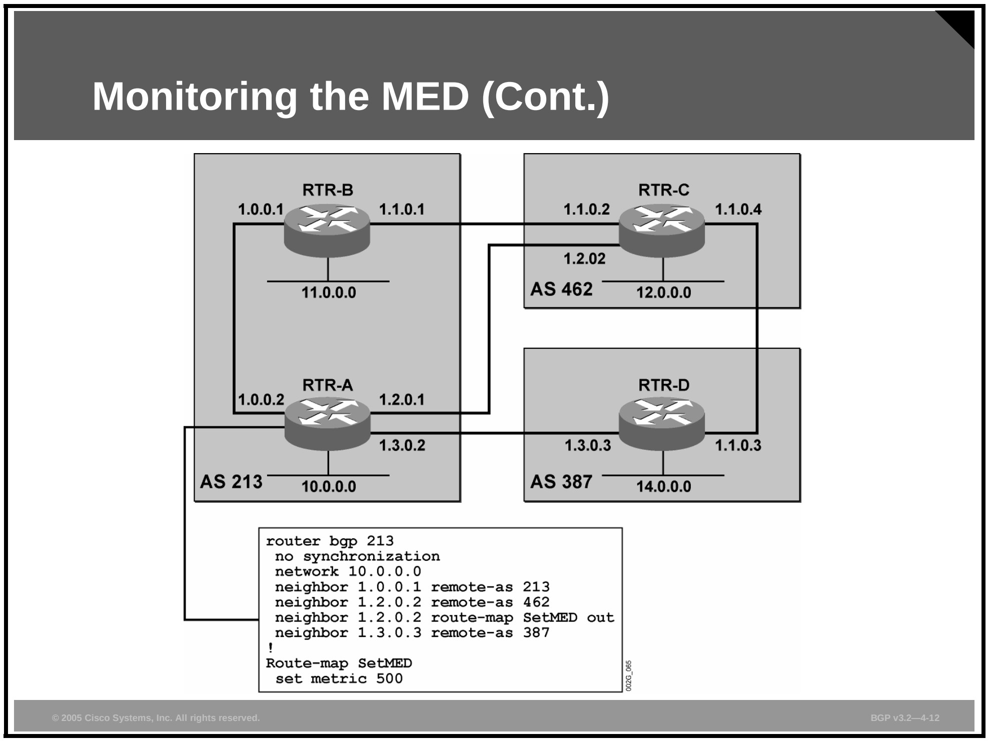
© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-11

All BGP-related **show** and **debug** commands display the value of the MED attribute. If the inbound soft reconfiguration feature is enabled on the router, the original MED attribute that is received by the router is also displayed. The following examples demonstrate command output for Cisco **show ip bgp** commands.

Example: Monitoring the MED

This example illustrates monitoring the BGP table to verify correct MED configurations.

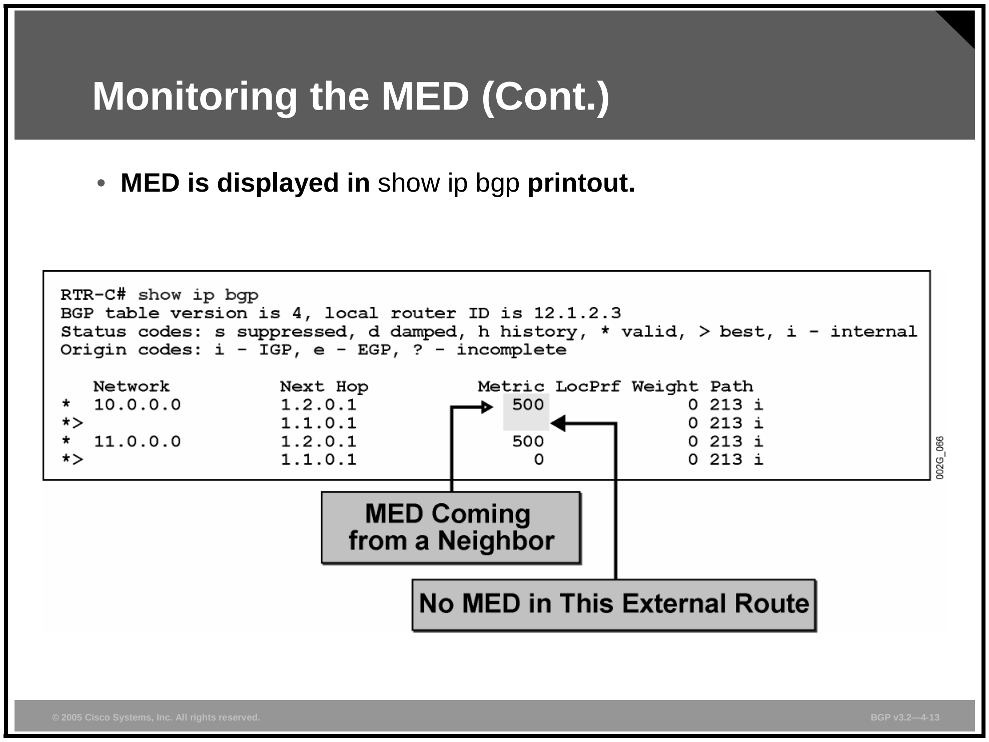


Monitoring the MED (Cont.)

```
router bgp 213
 no synchronization
 network 10.0.0.0
 neighbor 1.0.0.1 remote-as 213
 neighbor 1.2.0.2 remote-as 462
 neighbor 1.2.0.2 route-map SetMED out
 neighbor 1.3.0.3 remote-as 387
!
Route-map SetMED
 set metric 500
```

The same network used in the previous lesson on BGP local preference is used in this topic to produce the sample output that follows. All commands were executed on router RTR-C.

Some routing updates from router RTR-B are sent to router RTR-C with a MED of 500. Some updates from RTR-B to RTR-C have the MED set to 0, and some are without a MED attribute. Inbound soft reconfiguration is used on router RTR-C.



Monitoring the MED (Cont.)

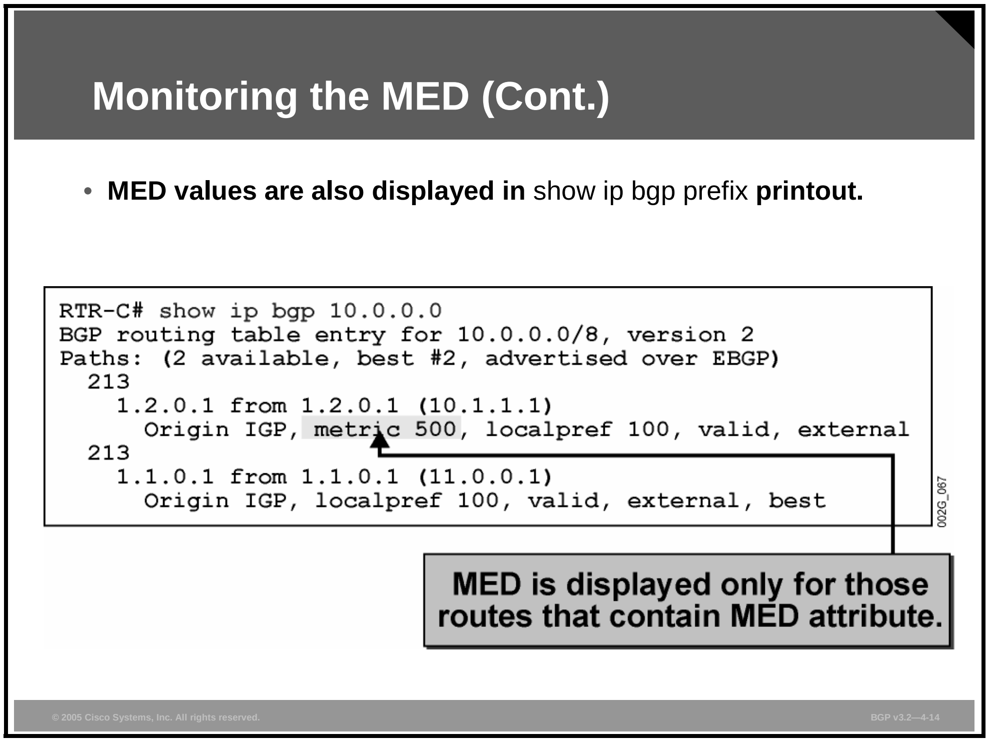
- **MED is displayed in** show ip bgp **printout.**

```
RTR-C# show ip bgp
BGP table version is 4, local router ID is 12.1.2.3
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
*  10.0.0.0         1.2.0.1             500              0 213 i
*>                  1.1.0.1                              0 213 i
*  11.0.0.0         1.2.0.1             500              0 213 i
*>                  1.1.0.1               0              0 213 i
```

MED Coming from a Neighbor

No MED in This External Route

Both networks that are received from router RTR-B have a MED of 500. Network 10.0.0.0/8, which is received from RTR-A, has no MED attribute, while network 11.0.0.0/8 has a MED value of 0.



Monitoring the MED (Cont.)

- **MED values are also displayed in** show ip bgp prefix **printout.**

```
RTR-C# show ip bgp 10.0.0.0
BGP routing table entry for 10.0.0.0/8, version 2
Paths: (2 available, best #2, advertised over EBGP)
  213
    1.2.0.1 from 1.2.0.1 (10.1.1.1)
      Origin IGP, metric 500, localpref 100, valid, external
  213
    1.1.0.1 from 1.1.0.1 (11.0.0.1)
      Origin IGP, localpref 100, valid, external, best
```

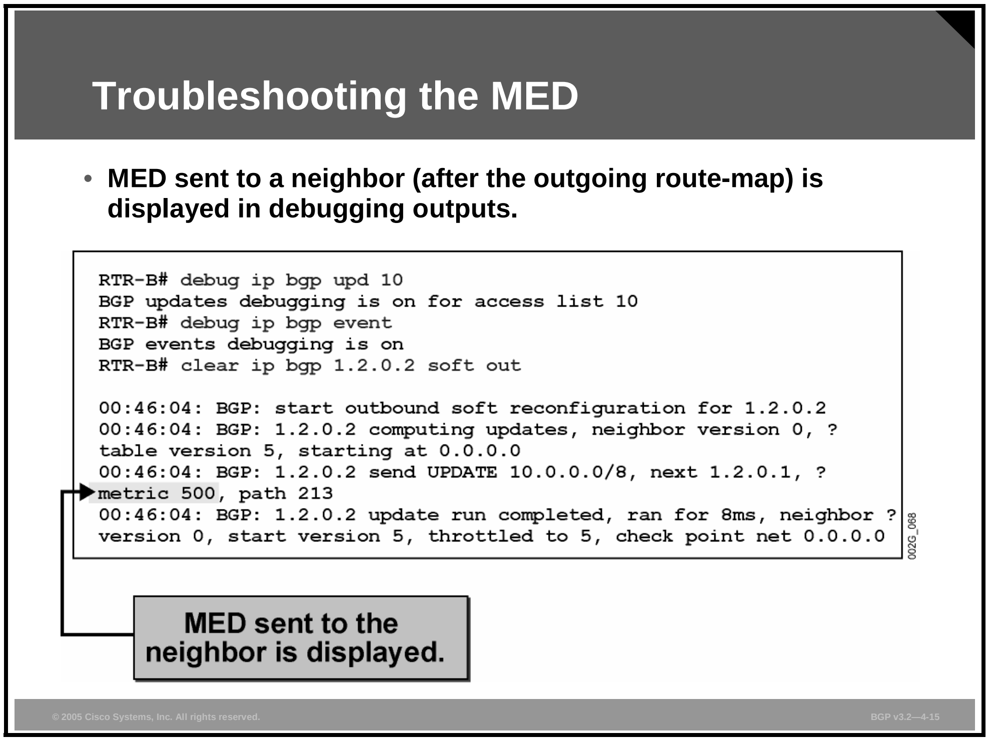
MED is displayed only for those routes that contain MED attribute.

BGP v3.2—4-14

When looking at detailed information for a specific network, you will see the MED (via the **show ip bgp** *prefix* command) only if the attribute exists.

Troubleshooting the MED

This topic lists the Cisco IOS commands that are required to troubleshoot BGP MED configurations on a Cisco router.



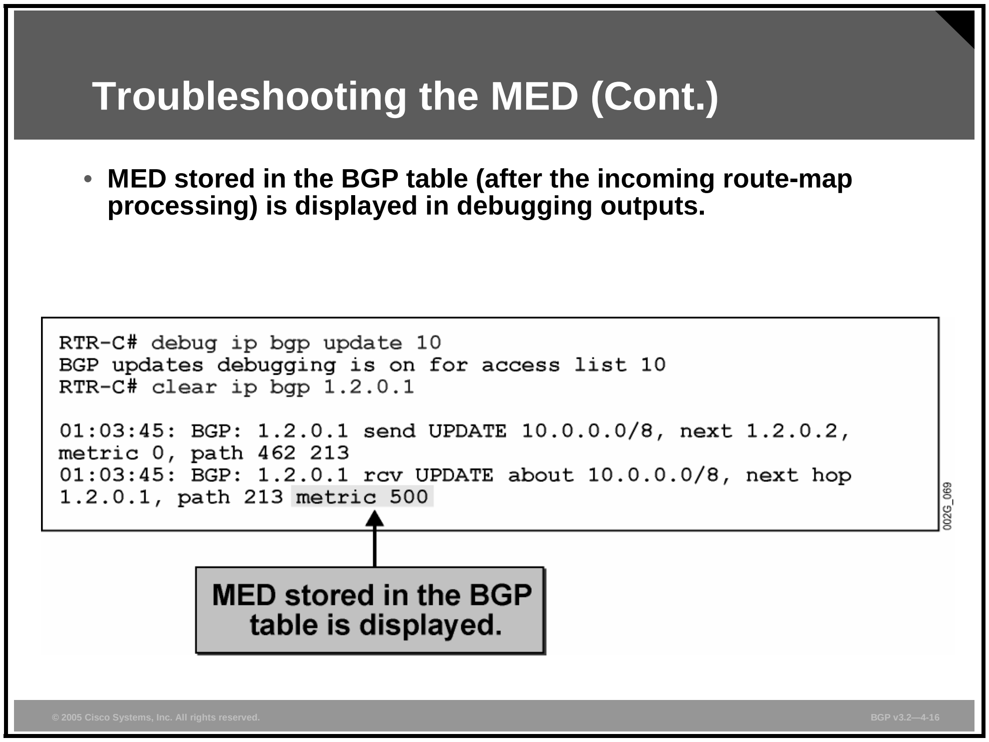
Troubleshooting the MED

- **MED sent to a neighbor (after the outgoing route-map) is displayed in debugging outputs.**

```
RTR-B# debug ip bgp upd 10
BGP updates debugging is on for access list 10
RTR-B# debug ip bgp event
BGP events debugging is on
RTR-B# clear ip bgp 1.2.0.2 soft out

00:46:04: BGP: start outbound soft reconfiguration for 1.2.0.2
00:46:04: BGP: 1.2.0.2 computing updates, neighbor version 0, ?
table version 5, starting at 0.0.0.0
00:46:04: BGP: 1.2.0.2 send UPDATE 10.0.0.0/8, next 1.2.0.1, ?
metric 500, path 213
00:46:04: BGP: 1.2.0.2 update run completed, ran for 8ms, neighbor ?
version 0, start version 5, throttled to 5, check point net 0.0.0.0
```

MED sent to the neighbor is displayed.

If debugging is necessary to troubleshoot a problem, the MED, among other attributes, is displayed. This example shows the MED attribute set with an outgoing route-map.



Troubleshooting the MED (Cont.)

- **MED stored in the BGP table (after the incoming route-map processing) is displayed in debugging outputs.**

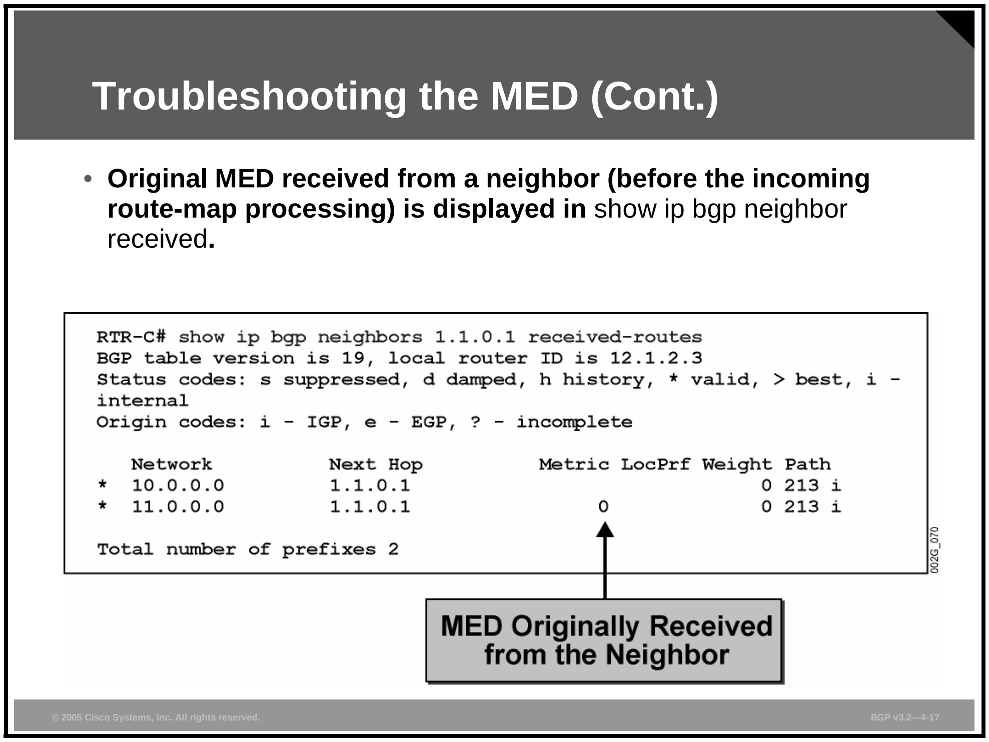
```
RTR-C# debug ip bgp update 10
BGP updates debugging is on for access list 10
RTR-C# clear ip bgp 1.2.0.1

01:03:45: BGP: 1.2.0.1 send UPDATE 10.0.0.0/8, next 1.2.0.2,
metric 0, path 462 213
01:03:45: BGP: 1.2.0.1 rcv UPDATE about 10.0.0.0/8, next hop
1.2.0.1, path 213 metric 500
```

MED stored in the BGP table is displayed.

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-16

This debugging example shows the MED attribute value after the update has been processed by an incoming route-map.



Troubleshooting the MED (Cont.)

- **Original MED received from a neighbor (before the incoming route-map processing) is displayed in** show ip bgp neighbor received**.**

```
RTR-C# show ip bgp neighbors 1.1.0.1 received-routes
BGP table version is 19, local router ID is 12.1.2.3
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
*  10.0.0.0         1.1.0.1                               0 213 i
*  11.0.0.0         1.1.0.1                0             0 213 i

Total number of prefixes 2
```

MED Originally Received from the Neighbor

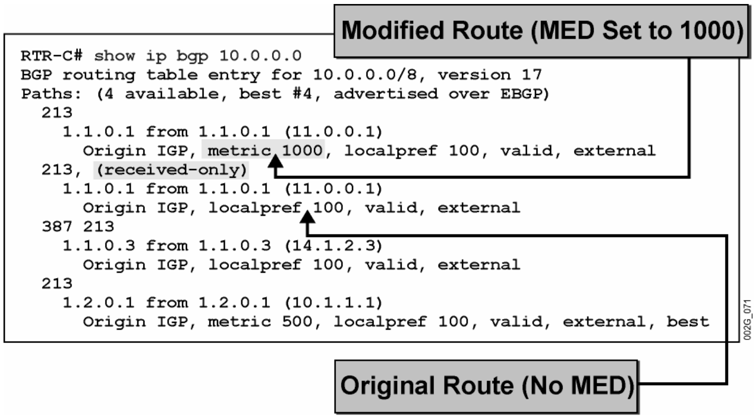
© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-17

To see the original MED, you need to enable soft reconfiguration on the router. The **show ip bgp neighbor** *address* **received-routes** command displays the original updates before any filters or route-maps have filtered or changed them.

Troubleshooting the MED (Cont.)

- **Both original route and modified route are displayed with a route-map when inbound soft reconfiguration is configured.**



Modified Route (MED Set to 1000)

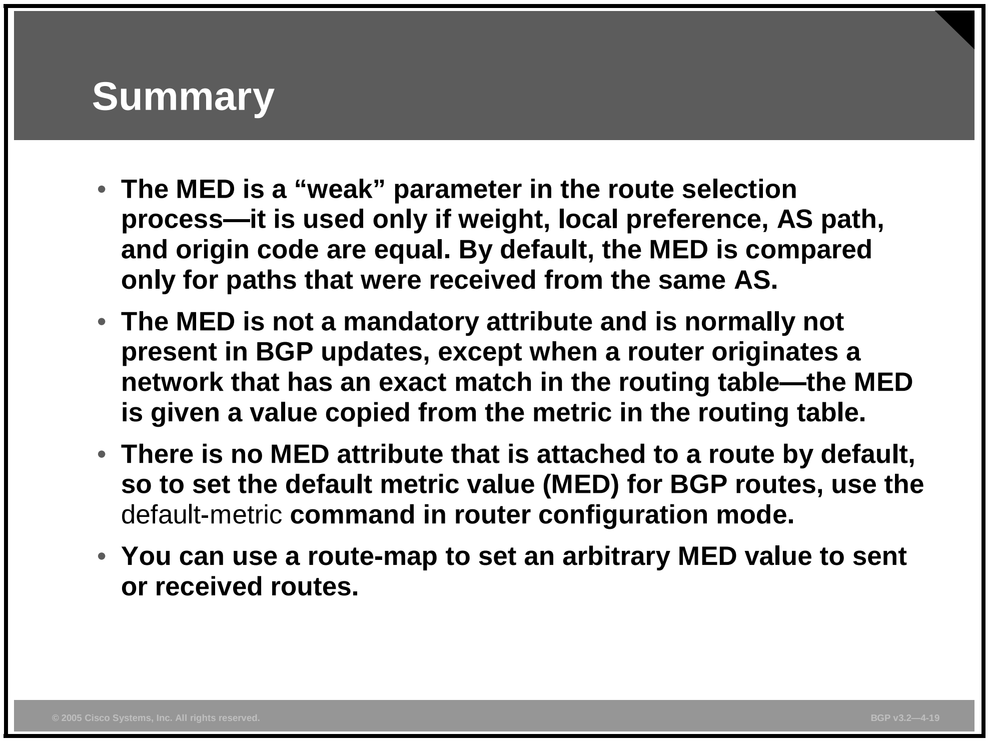
```
RTR-C# show ip bgp 10.0.0.0
BGP routing table entry for 10.0.0.0/8, version 17
Paths: (4 available, best #4, advertised over EBGP)
  213
    1.1.0.1 from 1.1.0.1 (11.0.0.1)
      Origin IGP, metric 1000, localpref 100, valid, external
  213, (received-only)
    1.1.0.1 from 1.1.0.1 (11.0.0.1)
      Origin IGP, localpref 100, valid, external
  387 213
    1.1.0.3 from 1.1.0.3 (14.1.2.3)
      Origin IGP, localpref 100, valid, external
  213
    1.2.0.1 from 1.2.0.1 (10.1.1.1)
      Origin IGP, metric 500, localpref 100, valid, external, best
```

Original Route (No MED)

If soft reconfiguration is enabled, the original updates to the MED attribute are available by using the **show ip bgp** *prefix* command. The original versions are marked with the **received-only** keyword and follow the version that is in the global BGP table. In the figure, the received update had no MED attribute but a value of 1000 was later applied through a route-map.

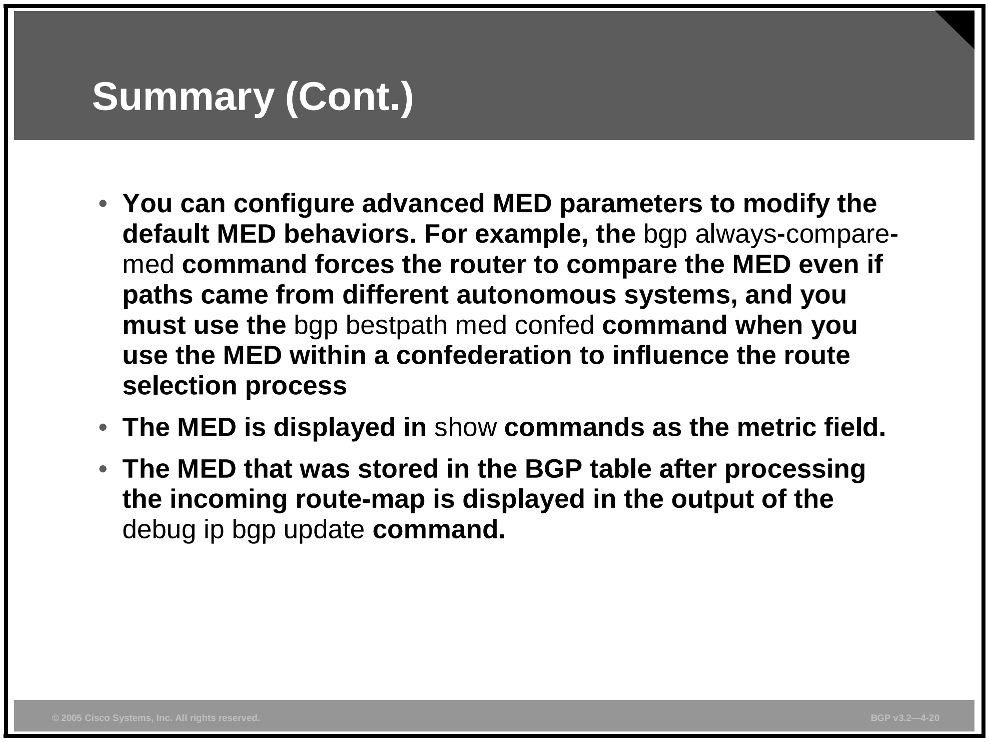
Summary

This topic summarizes the key points discussed in this lesson.



Summary

- **The MED is a “weak” parameter in the route selection process—it is used only if weight, local preference, AS path, and origin code are equal. By default, the MED is compared only for paths that were received from the same AS.**
- **The MED is not a mandatory attribute and is normally not present in BGP updates, except when a router originates a network that has an exact match in the routing table—the MED is given a value copied from the metric in the routing table.**
- **There is no MED attribute that is attached to a route by default, so to set the default metric value (MED) for BGP routes, use the** default-metric **command in router configuration mode.**
- **You can use a route-map to set an arbitrary MED value to sent or received routes.**



Summary (Cont.)

- **You can configure advanced MED parameters to modify the default MED behaviors. For example, the** bgp always-compare-med **command forces the router to compare the MED even if paths came from different autonomous systems, and you must use the** bgp bestpath med confed **command when you use the MED within a confederation to influence the route selection process**
- **The MED is displayed in** show **commands as the metric field.**
- **The MED that was stored in the BGP table after processing the incoming route-map is displayed in the output of the** debug ip bgp update **command.**

Lesson 5

Addressing BGP Communities

Overview

When connections to multiple Internet service providers (ISPs) are required, it is important that Border Gateway Protocol (BGP) select the optimum route for traffic to use. It is equally important that the return path that is selected be the optimum return path into the autonomous system (AS). The optimum, or best, route may not be what the network designer intended based on design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides tools for administrators to influence route selection. The BGP community attribute is one such tool.

This lesson discusses how to influence BGP route selection by setting the BGP community attribute on outgoing BGP routes and describes BGP communities and their use to facilitate proper return path selection. The configuration details of BGP communities and the use of community-lists and route-maps to influence route selection are also discussed. This lesson concludes by explaining how to monitor BGP community attributes.

Objectives

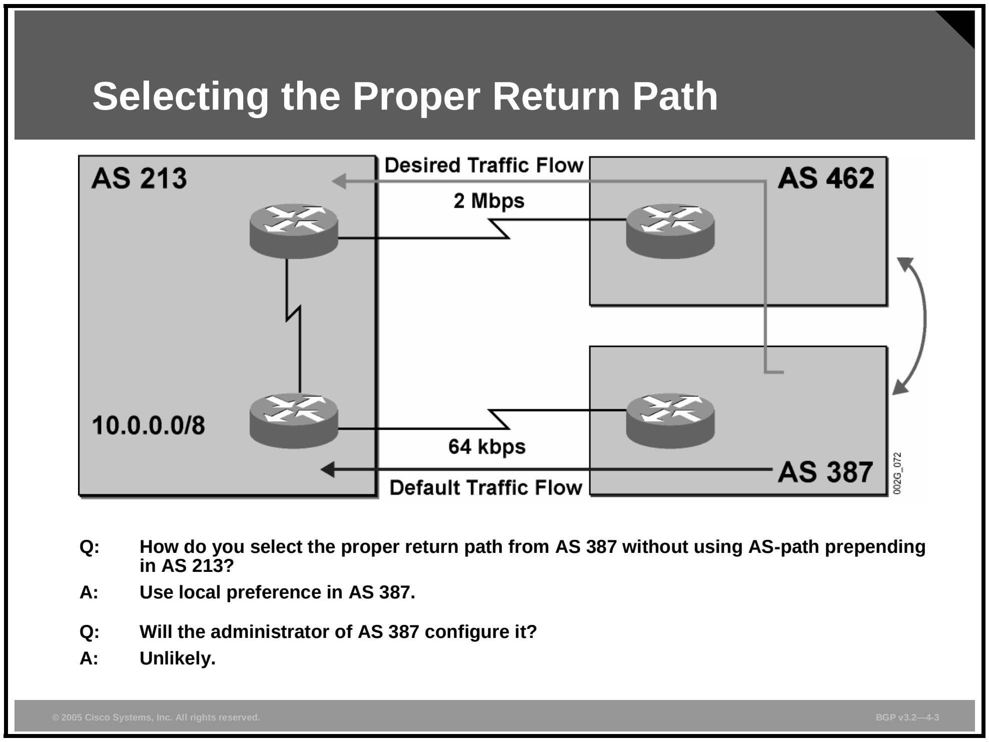
Upon completing this lesson, you will be able to use BGP community attributes to influence route selection in a customer scenario where you must support multiple ISP connections. This ability includes being able to meet these objectives:

- Describe the issues of return path selection for multihomed customers and why you cannot use the BGP attributes of weight, local preference, and MED to solve these issues
- Describe the basic qualities of BGP communities
- Describe how BGP communities facilitate proper return path selection
- List the steps that are required to successfully deploy communities in a BGP-based network
- Identify the Cisco IOS commands that are required to configure route tagging with BGP communities
- Identify the Cisco IOS command that is required to enable BGP community propagation
- Identify the Cisco IOS commands that are required to match routes based on attached BGP communities using community-lists
- Describe the function of the BGP Named Community Lists feature

- Describe the function of the BGP Cost Community feature
- Describe the function of the BGP Link Bandwidth feature
- Describe how BGP supports sequenced entries in extended community-lists
- Identify the Cisco IOS command that is required to match routes based on attached BGP communities using route-maps
- Identify the Cisco IOS commands that are required to monitor BGP communities

Selecting the Proper Return Path

This topic describes the issues of return path selection for multihomed customers and why you cannot use the BGP attributes of weight, local preference, and multi-exit discriminator (MED) to resolve these issues.



Selecting the Proper Return Path

Q: How do you select the proper return path from AS 387 without using AS-path prepending in AS 213?

A: Use local preference in AS 387.

Q: Will the administrator of AS 387 configure it?

A: Unlikely.

BGP v3.2—4-3

In this example, the customer and the backup service provider would like to avoid AS-path prepending and rely on other BGP tools to properly route the return traffic over the highest-speed WAN link.

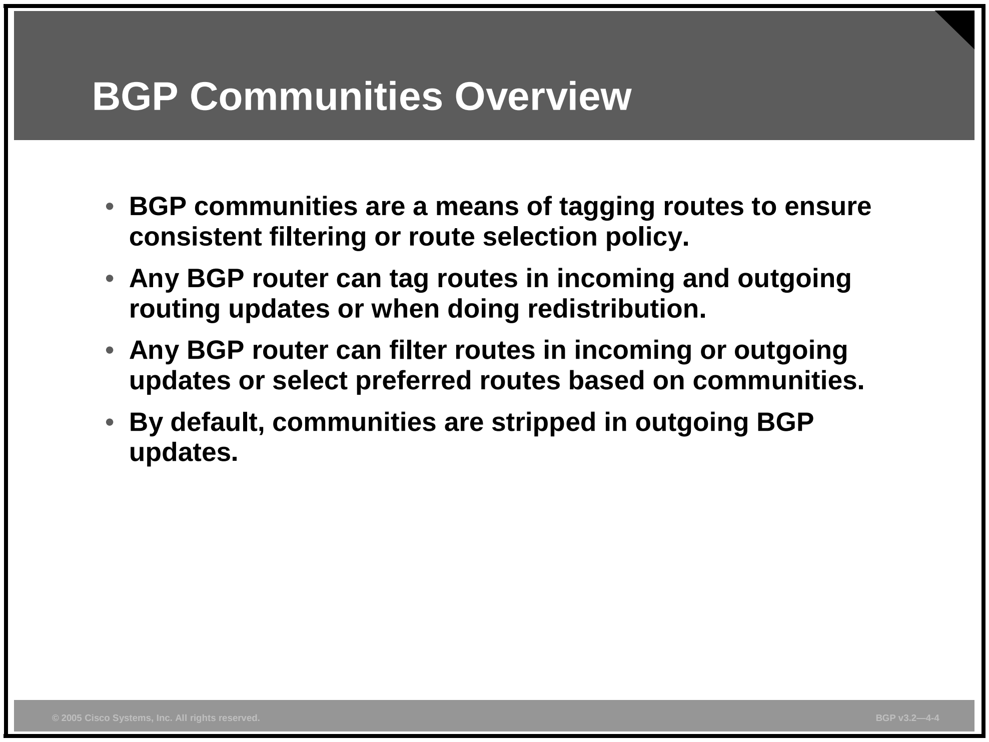
Using the MED to influence the preferred return path is not possible because the MED cannot be propagated across several autonomous systems. AS 387 would, therefore, receive networks from AS 213 directly with the MED attribute but would receive networks without a MED attribute from AS 462. In any case, BGP route selection would be based on the length of the AS path, and even if the MED was present and used the shortest path, the path would still be through the slow 64-kbps link.

The only option for resolving this issue is to use local preference in AS 387. The problem with this solution is that service providers normally do not rush to implement every wish that their customers might have.

This lesson describes a solution to this case study that uses the transitive optional attribute called "BGP community" in conjunction with local preference.

BGP Communities Overview

This topic describes the basic qualities of BGP communities.



BGP Communities Overview

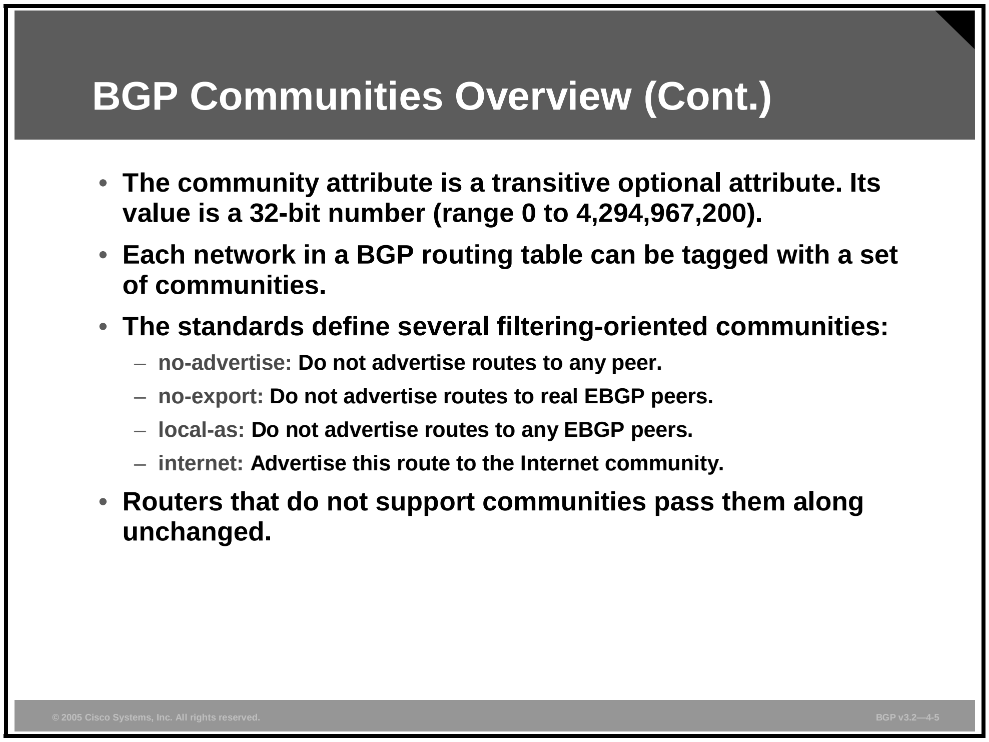
- **BGP communities are a means of tagging routes to ensure consistent filtering or route selection policy.**
- **Any BGP router can tag routes in incoming and outgoing routing updates or when doing redistribution.**
- **Any BGP router can filter routes in incoming or outgoing updates or select preferred routes based on communities.**
- **By default, communities are stripped in outgoing BGP updates.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-4

BGP communities are attributes that are used to group and filter routes. Communities are designed to give the network operator the ability to apply policies to large numbers of routes by using match and set clauses in the configuration of route maps. Community-lists are used in this process to identify and filter routes by their common attributes.

A community is an attribute that is used to tag BGP routes. A router can apply it to any BGP route by using a route-map. Other routers can then perform any action based on the tag (community) that is attached to the route.

There can be more than one BGP community that is attached to a single route, but the routers, by default, remove communities in outgoing BGP updates.



BGP Communities Overview (Cont.)

- **The community attribute is a transitive optional attribute. Its value is a 32-bit number (range 0 to 4,294,967,200).**
- **Each network in a BGP routing table can be tagged with a set of communities.**
- **The standards define several filtering-oriented communities:**
  - no-advertise: **Do not advertise routes to any peer.**
  - no-export: **Do not advertise routes to real EBGP peers.**
  - local-as: **Do not advertise routes to any EBGP peers.**
  - internet: **Advertise this route to the Internet community.**
- **Routers that do not support communities pass them along unchanged.**

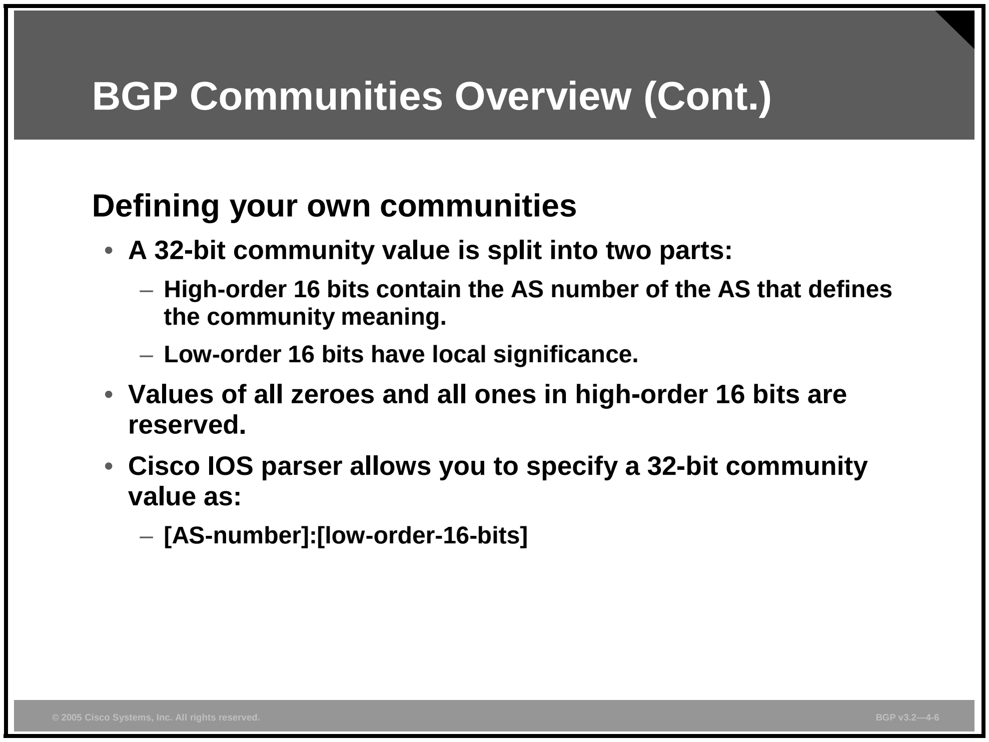
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-5

The community attribute is a 32-bit transitive optional BGP attribute that was designed to group destinations and apply routing decision (accept, prefer, redistribute, and so on) according to communities to allow the easy application of administrative policies. BGP communities provide a mechanism to reduce BGP configuration complexity on a router controlling the distribution of routing information.

A set of community values has been predefined. When a router receives a route that has been marked with a predefined community, the router will perform a specific, predefined action that is based on that community setting as follows:

- **no-advertise:** If a router receives an update carrying this community, it will not forward it to any neighbor.
- **no-export:** If a router receives an update carrying this community, it will not propagate it to any external neighbors except intra-confederation external neighbors. This is the most widely used predefined community attribute.
- **local-as:** This community has a similar meaning to **no-export**, but it keeps a route within the local AS (or member-AS within the confederation). The route is not sent to external BGP neighbors or to intra-confederation external neighbors.
- **internet:** Advertise this route to the Internet community. All routers belong to it.

Routers that do not support the community attribute will pass the attribute to other neighbors because it is a transitive attribute.



BGP Communities Overview (Cont.)

Defining your own communities

- A 32-bit community value is split into two parts:
  - High-order 16 bits contain the AS number of the AS that defines the community meaning.
  - Low-order 16 bits have local significance.
- Values of all zeroes and all ones in high-order 16 bits are reserved.
- Cisco IOS parser allows you to specify a 32-bit community value as:
  - [AS-number]:[low-order-16-bits]

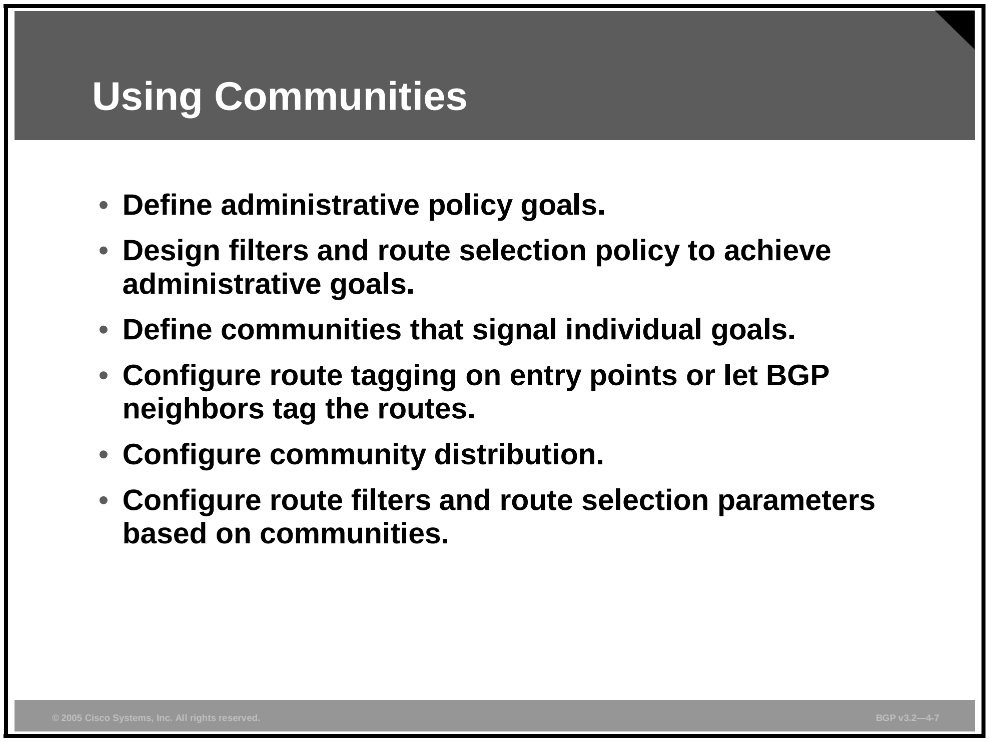
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-6

Community attributes are usually used between neighboring autonomous systems. For the BGP communities to be globally unique, a public AS number should be part of the community value. For this reason, you can enter the community value as two 16-bit numbers separated by a colon. The first number (high-order 16 bits) should be the AS number of the AS that defines the community value, and the second number should be a value that is assigned a certain meaning (that is, translation of a community value into local preference in the neighboring AS).

Communities can also be used internally within an AS (to ensure AS-wide routing policy), in which case the first 16 bits should contain the AS number of the local AS.

Using Communities

This topic describes how BGP communities can facilitate proper return path selection.



Using Communities

- **Define administrative policy goals.**
- **Design filters and route selection policy to achieve administrative goals.**
- **Define communities that signal individual goals.**
- **Configure route tagging on entry points or let BGP neighbors tag the routes.**
- **Configure community distribution.**
- **Configure route filters and route selection parameters based on communities.**

Designing a BGP solution around BGP communities usually requires the following steps:

Step 1 Define administrative policy goals that you need to implement.

Step 2 Define the filters and route selection policy that will achieve the required goals.

Step 3 Assign a community value to each goal.

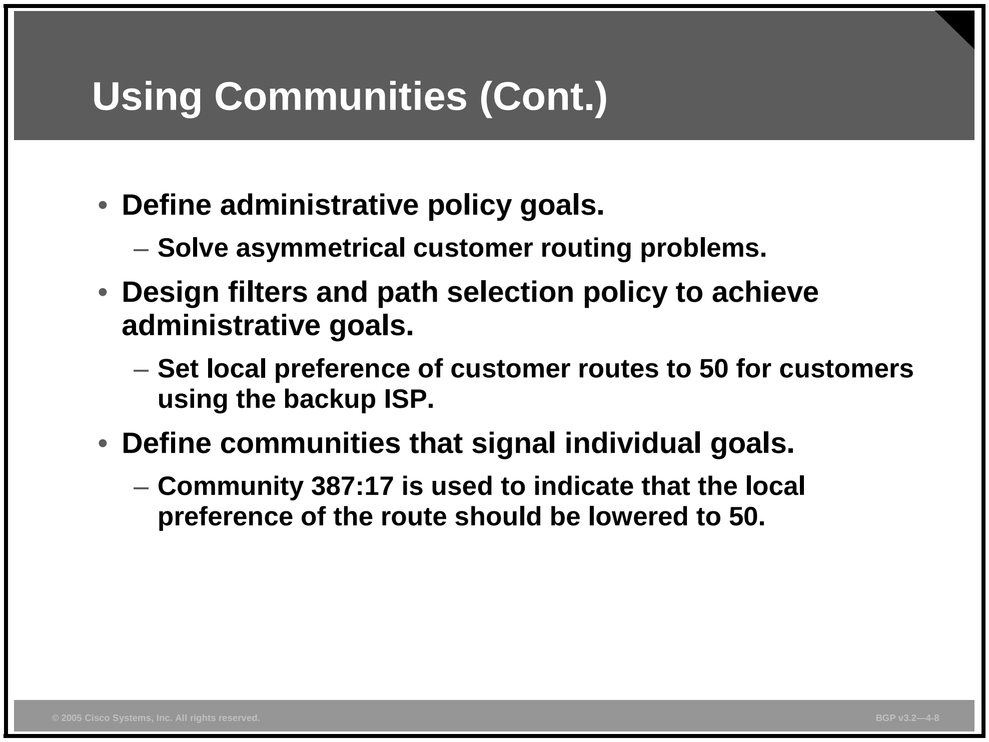
Step 4 Apply communities on incoming updates from neighboring autonomous systems or tell the neighbors to set the communities themselves.

Step 5 Enable community distribution throughout your AS to allow community propagation.

Step 6 Match communities with route-maps and route filters, change BGP attributes, or influence the route selection process based on the communities that are attached to the BGP routes.

Example: Using Communities

This example shows how you can define goals and assign communities to them.



Using Communities (Cont.)

- **Define administrative policy goals.**
  - **Solve asymmetrical customer routing problems.**
- **Design filters and path selection policy to achieve administrative goals.**
  - **Set local preference of customer routes to 50 for customers using the backup ISP.**
- **Define communities that signal individual goals.**
  - **Community 387:17 is used to indicate that the local preference of the route should be lowered to 50.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-8

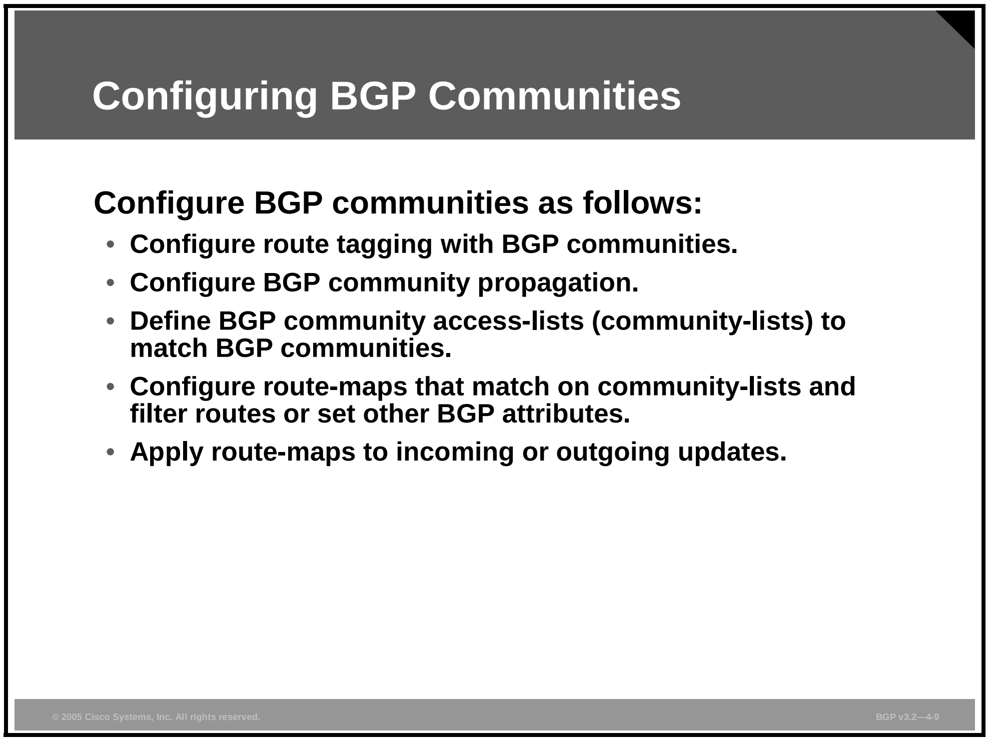
This table lists the goals and the community values.

Goal	Community Value
Set local preference of 50.	387:17
Set local preference of 150.	387:18
Prepend the AS path once when sending the network to external neighbors.	387:21
Prepend the AS path twice when sending the network to external neighbors.	387:22
Prepend the AS path three times when sending the network to external neighbors.	387:23

All customers of the service provider should know this list so that they can use the BGP communities without having to discuss their use with the service provider.

Configuring BGP Communities

This topic lists the activities that are required to successfully deploy BGP communities in a BGP-based network.



Configuring BGP Communities

Configure BGP communities as follows:

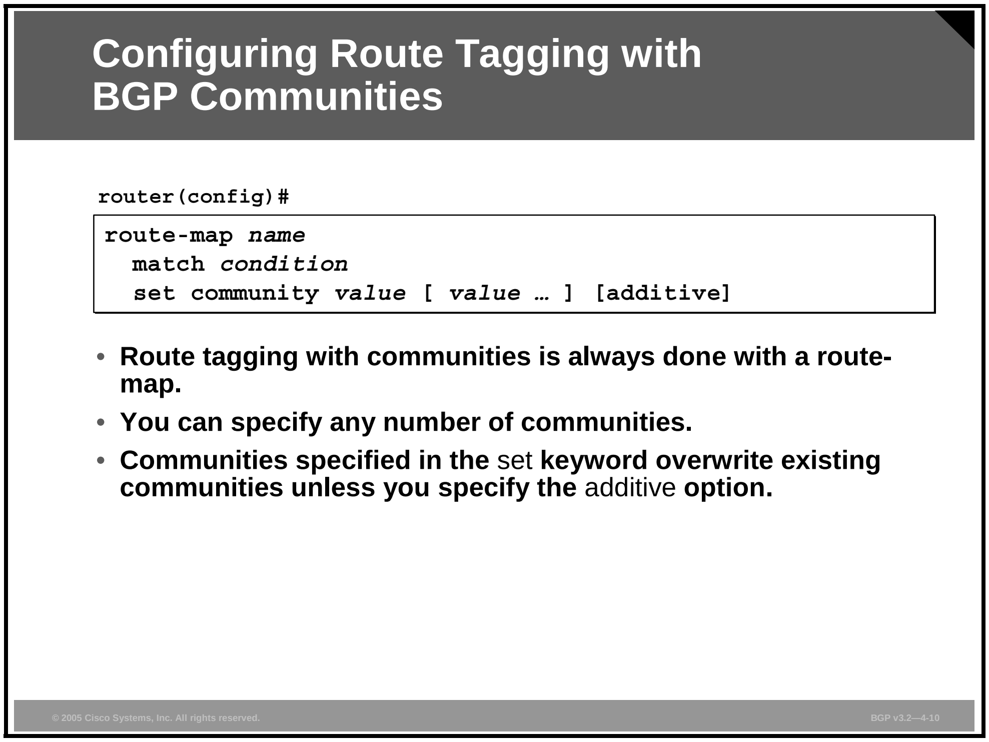
- **Configure route tagging with BGP communities.**
- **Configure BGP community propagation.**
- **Define BGP community access-lists (community-lists) to match BGP communities.**
- **Configure route-maps that match on community-lists and filter routes or set other BGP attributes.**
- **Apply route-maps to incoming or outgoing updates.**

Configuration activities when you are using communities include the following:

- Setting communities, which requires a route-map.
- Enabling community propagation per neighbor for all internal neighbors. If communities are sent to external neighbors, you must enable community propagation for external neighbors.
- Creating community-lists to be used within route-maps to match on community values.
- Creating route-maps where community-lists are used to match on community values. You can then use route-maps to filter based on community values or to set other parameters or attributes (for example, local preference, MED, or AS-path prepending).
- Applying route-maps to incoming or outgoing updates.

Configuring Route Tagging with BGP Communities

This topic lists the Cisco IOS commands that are required to configure route tagging with BGP communities.



Configuring Route Tagging with BGP Communities

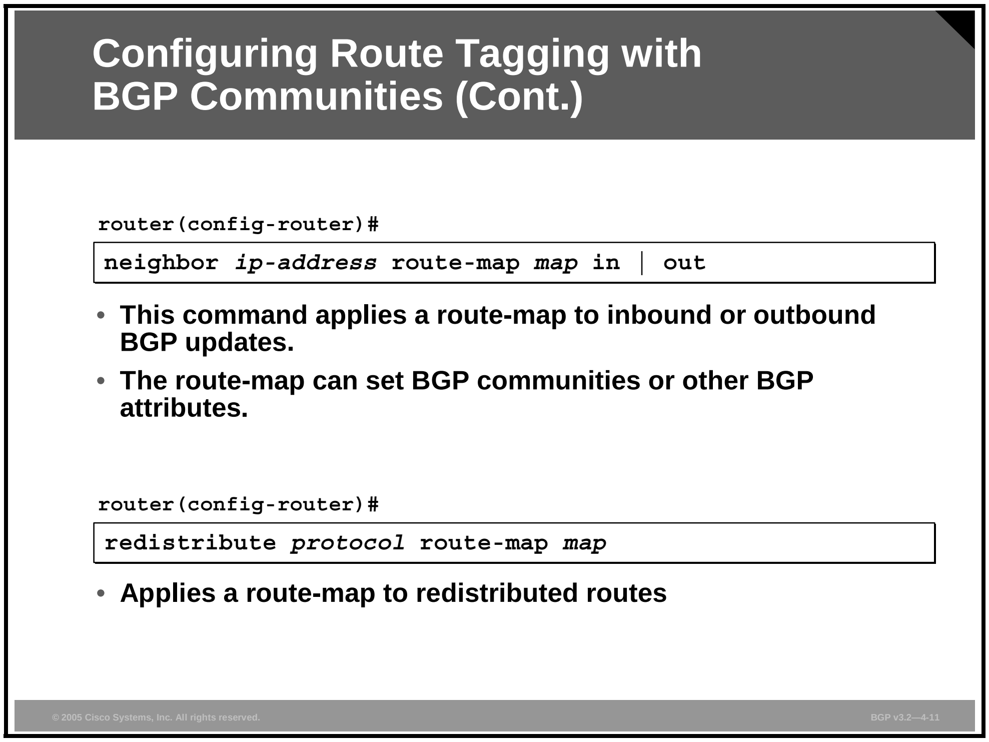
```
router(config)#
```

```
route-map name
  match condition
  set community value [ value … ] [additive]
```

- **Route tagging with communities is always done with a route-map.**
- **You can specify any number of communities.**
- **Communities specified in the** set **keyword overwrite existing communities unless you specify the** additive **option.**

BGP v3.2—4-10

In a route-map configuration mode, you should use the **set community** command to attach a community attribute (or a set of communities) to a route. You can attach up to 32 communities to a single route with one route-map set statement. If the keyword **additive** is used, the original communities are preserved and the router simply appends the new communities to the route. Omitting the **additive** keyword results in the overwriting of any original community attributes.



Configuring Route Tagging with BGP Communities (Cont.)

router(config-router)#

```
neighbor ip-address route-map map in | out
```

- **This command applies a route-map to inbound or outbound BGP updates.**
- **The route-map can set BGP communities or other BGP attributes.**

router(config-router)#

```
redistribute protocol route-map map
```

- **Applies a route-map to redistributed routes**

BGP v3.2—4-11

You can apply a route-map to incoming or outgoing updates. You can also use it with redistribution from another routing protocol.

Note A route-map is a filtering mechanism that has an “implicit deny” for all networks that are not matched in any route-map statement. If a route-map is not intended to filter routes, then you should add another route-map statement at the end to permit all remaining networks without changing it (no **match** and no **set** commands are used within that route-map statement).

Originally, Cisco IOS software accepted and displayed BGP community values as a single 32-bit value in a digital format. Newer Cisco IOS versions support the new format, where you can set or view a community as two colon-separated 16-bit numbers.

The global command **ip bgp-community new-format** is recommended on all routers whenever communities contain the AS number.

After being converted, configuration files with communities in the new *as:nn* format are not compatible with older versions of Cisco IOS software. This example shows the difference in appearance between the old and new formats:

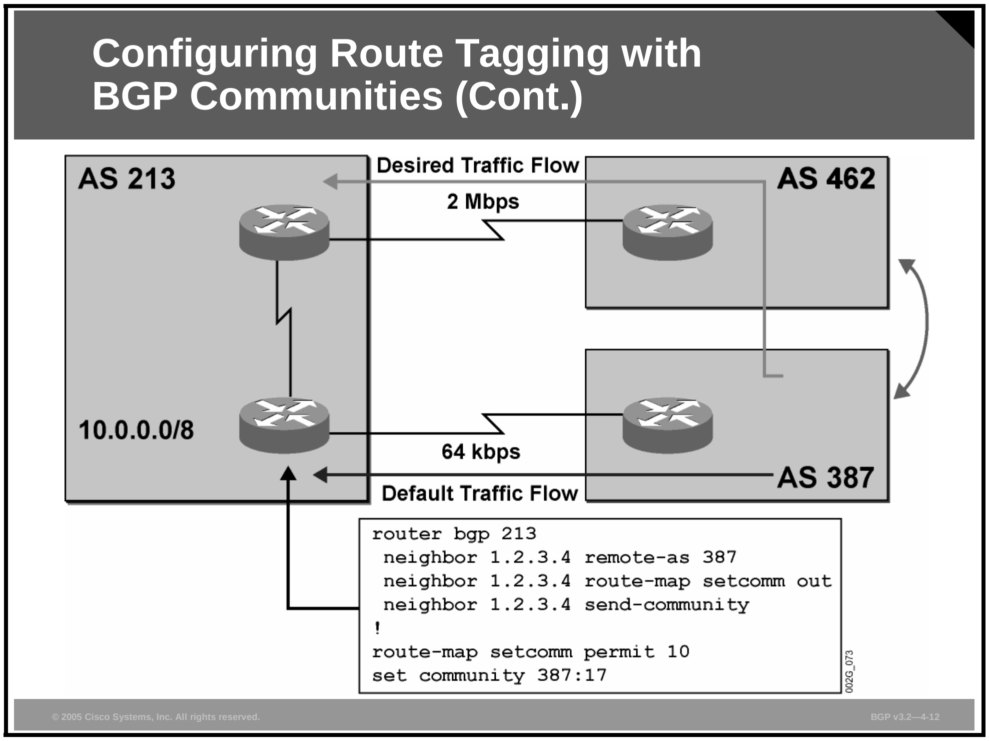
```
router# show ip bgp 6.0.0.0
Community: 6553620
```

After the **ip bgp-community new-format** command:

```
router# show ip bgp 6.0.0.0
Community: 100:20
```

Example: Configuring Route Tagging with BGP Communities

In this example, a border router in AS 213 applies a community value of 387:17 to all networks that are sent to neighboring AS 387.



Configuring Route Tagging with BGP Communities (Cont.)

```
router bgp 213
 neighbor 1.2.3.4 remote-as 387
 neighbor 1.2.3.4 route-map setcomm out
 neighbor 1.2.3.4 send-community
!
route-map setcomm permit 10
set community 387:17
```

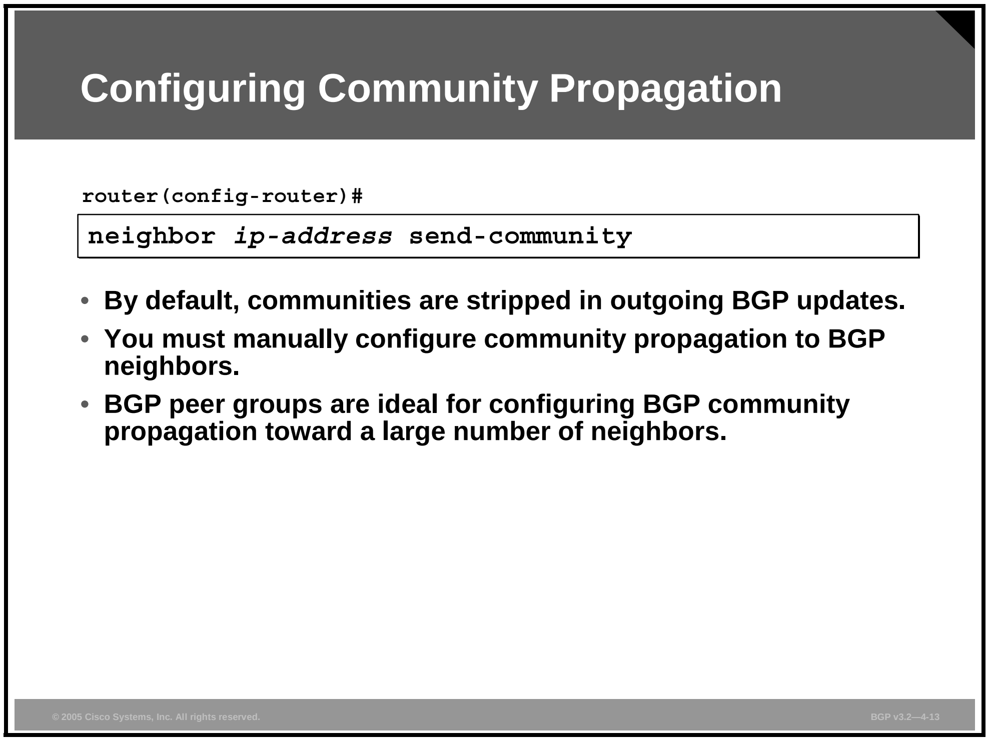
Another route-map entry is not needed because the first statement permits all networks (no **match** command means “match all”).

If it is more desirable to set communities on specific routes, you can use a standard access-list to match against, with the **match ip address** command in the route-map.

In a later example, networks with community 387:17 will have the local preference changed to a value of 50 within AS 387 to force AS 387 to prefer the other path that carries the default local preference of 100.

Configuring Community Propagation

This topic describes the Cisco IOS command that is required to enable BGP community propagation to BGP neighbors.



Configuring Community Propagation

```
router(config-router)#
neighbor ip-address send-community
```

- **By default, communities are stripped in outgoing BGP updates.**
- **You must manually configure community propagation to BGP neighbors.**
- **BGP peer groups are ideal for configuring BGP community propagation toward a large number of neighbors.**

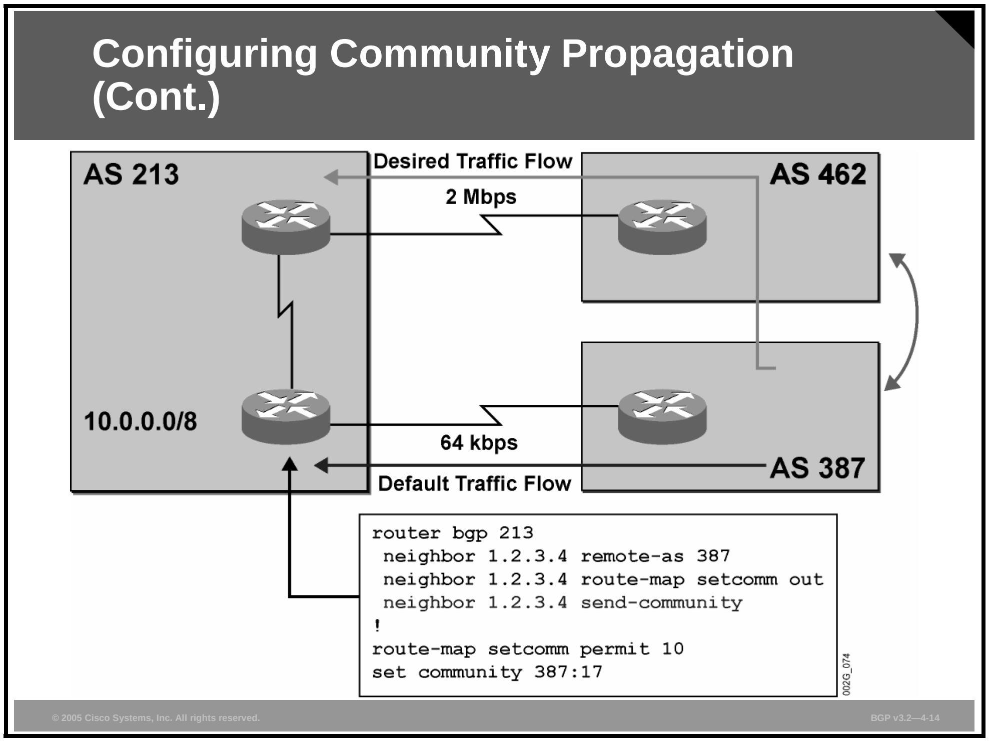
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-13

A command that is commonly forgotten by network administrators when configuring BGP communities is the **neighbor** *ip-address* **send-community** command. This command is needed to propagate community attributes to BGP neighbors. Even if you use an outgoing route-map to set communities, by default, the router will strip out any community values that are attached to outgoing BGP updates if you have not configured this command for the specific BGP neighbor.

You can also apply this command to a peer group.

Example: Configuring Community Propagation

This example illustrates community propagation.



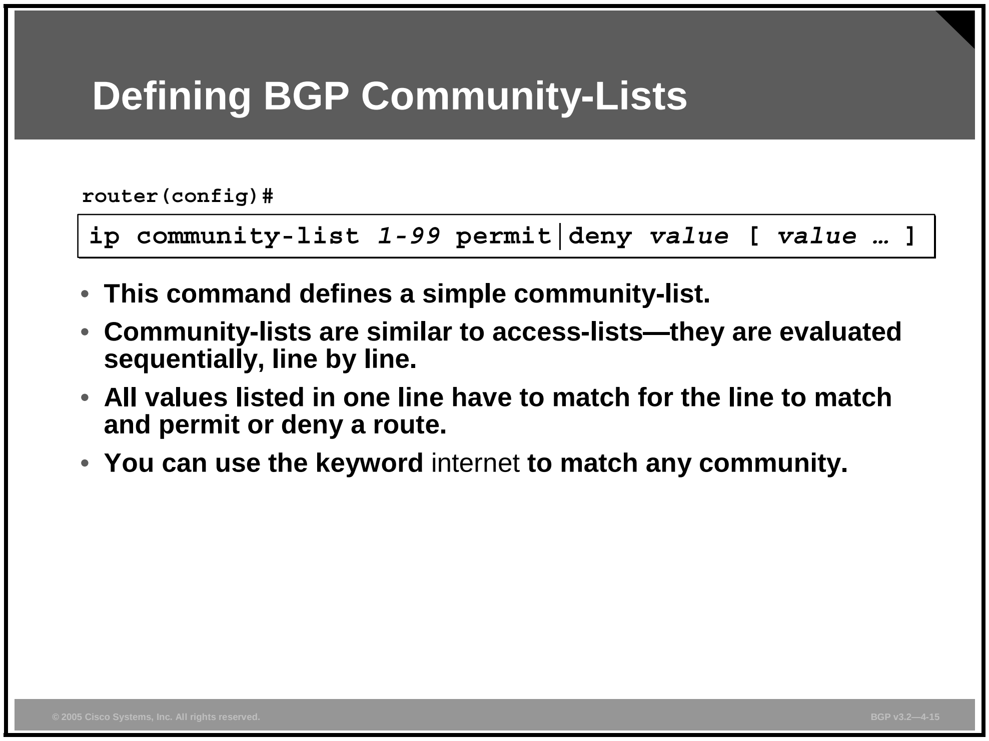
Configuring Community Propagation (Cont.)

```
router bgp 213
 neighbor 1.2.3.4 remote-as 387
 neighbor 1.2.3.4 route-map setcomm out
 neighbor 1.2.3.4 send-community
!
route-map setcomm permit 10
set community 387:17
```

The configuration example that was discussed earlier in this lesson must include the **send-community** command to enable community propagation from AS 213 to AS 387.

Defining BGP Community-Lists

This topic lists the Cisco IOS commands that are required to match routes based on attached BGP communities using community-lists.



Defining BGP Community-Lists

```
router(config)#
ip community-list 1-99 permit|deny value [ value … ]
```

- **This command defines a simple community-list.**
- **Community-lists are similar to access-lists—they are evaluated sequentially, line by line.**
- **All values listed in one line have to match for the line to match and permit or deny a route.**
- **You can use the keyword** internet **to match any community.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-15

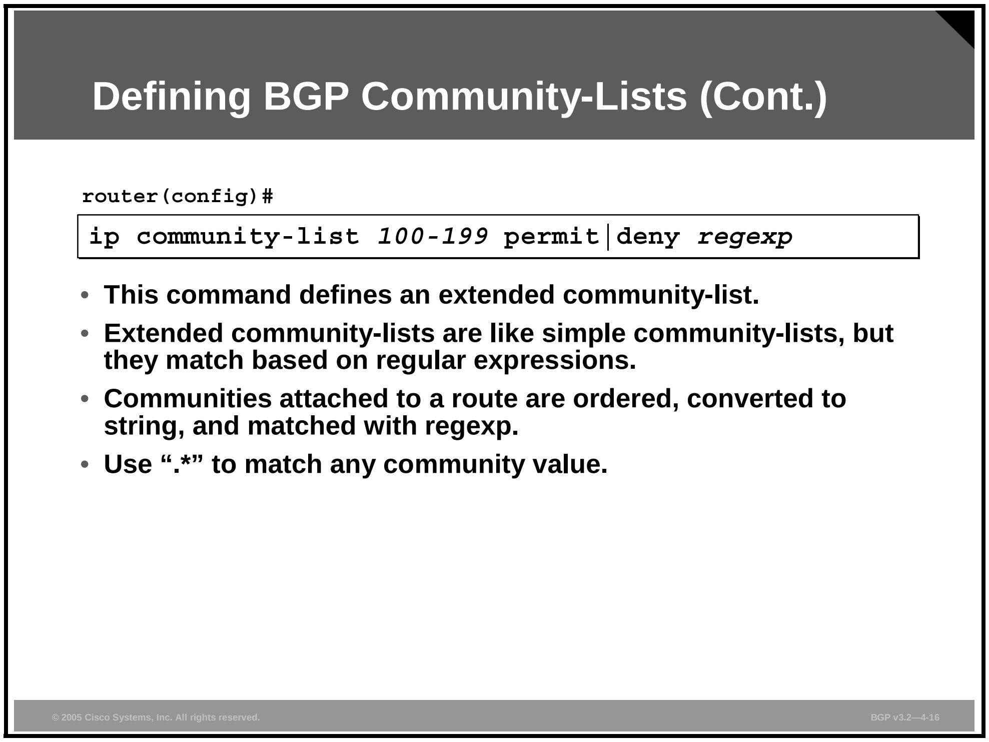
You can use a standard community access-list to find community attributes in routing updates. A standard community-list is defined by its assigned list number, which can range from 1 to 99.

Community-lists are similar to standard IP access-lists in the following ways:

- The router evaluates the lines in the community-list sequentially.
- If no line matches communities that are attached to a BGP route, the route is implicitly denied.

Standard community-lists are different from standard IP access-lists in the following ways:

- The keyword **internet** should be used to permit any community value.
- If more values are listed in a single line, they all have to be in an update to produce a match.



Defining BGP Community-Lists (Cont.)

```
router(config)#
ip community-list 100-199 permit|deny regexp
```

- This command defines an extended community-list.
- Extended community-lists are like simple community-lists, but they match based on regular expressions.
- Communities attached to a route are ordered, converted to string, and matched with regexp.
- Use “.*” to match any community value.

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—4-16

An expanded community-list is defined by its assigned list number, which can range from 100 to 199. Regular expressions are used to match community attributes. When a router processes a list of communities that are attached to a network update, they are converted into an ordered string of characters. This example shows how the process is accomplished:

1. The original list of communities in an update:

   “10.0.0.0/8, NH=1.1.1.1, origin=I, AS-path=20 30 40, community=10:101, community=10:201, community=10:105, community=10:205”

2. A string of characters containing an ordered list of community values:

   “_10:101_10:105_10:201_10_205_” (“_” represents a space)

3. A regular expression:

   “permit _10:.0[1-5]_” (“_” represents an underscore that matches spaces)

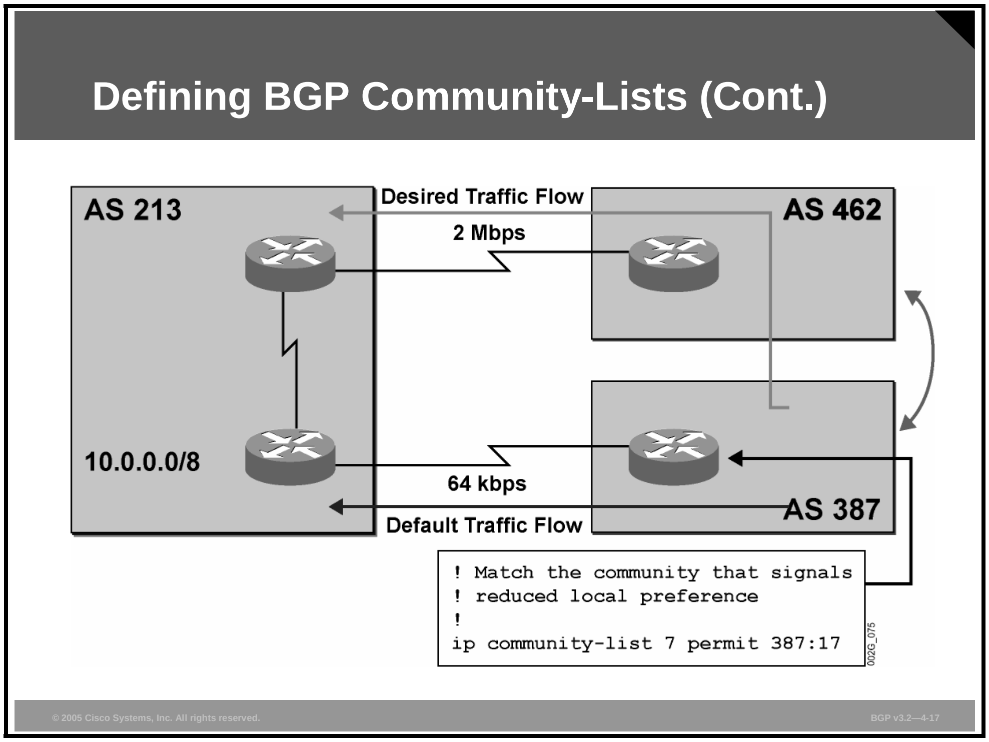
4. The result:

   This regular expression permits the route because it permits all routes with communities where the first 16 bits carry the AS number 10 and the second 16 bits contain 0 as the second digit and a number between 1 and 5 as the third digit; the first digit can be anything (as indicated by the “.”).

   Use regular expression “.*” to permit any community.

Example: Defining BGP Community-Lists

This example shows a portion of the configuration of the router in AS 387.



Defining BGP Community-Lists (Cont.)

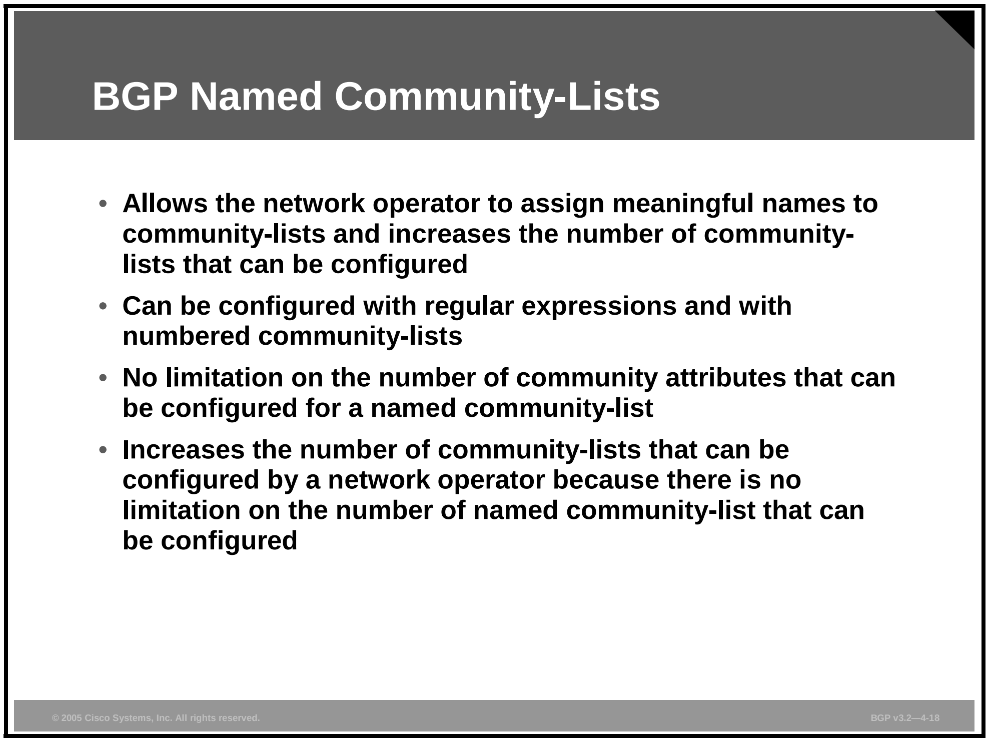
```
! Match the community that signals
! reduced local preference
!
ip community-list 7 permit 387:17
```

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-17

The access-list has been configured to match communities that were previously set by the router in AS 213.

BGP Named Community-Lists

This topic describes the function of the BGP Named Community Lists feature.



BGP Named Community-Lists

- **Allows the network operator to assign meaningful names to community-lists and increases the number of community-lists that can be configured**
- **Can be configured with regular expressions and with numbered community-lists**
- **No limitation on the number of community attributes that can be configured for a named community-list**
- **Increases the number of community-lists that can be configured by a network operator because there is no limitation on the number of named community-list that can be configured**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-18

The BGP Named Community Lists feature introduces a new type of community-list called the named community-list. A named community-list can be configured with regular expressions and with numbered community-lists. The BGP Named Community Lists feature allows the network operator to assign meaningful names to community-lists. All rules of numbered communities apply to named community-lists except that there is no limitation on the number of community attributes that can be configured for a named community-list. Although both standard and expanded community-lists have a limitation of 100 community groups that can be configured within each type of list, a named community-list does not have this limitation.

ip-community-list

To create a numbered or named community-list for BGP and to control access to it, use the **ip community-list** command in global configuration mode.

- **ip community-list** {*standard-list-number* | *expanded-list-number [regular-expression]* | {**standard | expanded**} *community-list-name*} {**permit** | **deny**} *community-number* | *regular-expressio*n

To delete the community-list, use the **no** form of this command.

- **no ip community-list** *standard-list-number* | *expanded-list-number* | *community-list-name*

Syntax Description

Parameter	Description
standard-list-number	Specifies a standard community-list number from 1 to 99 that identifies one or more permit or deny groups of communities.
expanded-list-number	Specifies an expanded community-list number from 100 to 199 that identifies one or more permit or deny groups of communities.
standard	Configures a standard named community-list.
expanded	Configures an expanded named community-list.
community-list-name	The community-list name.
permit	Permits access for a matching condition.
deny	Denies access for a matching condition.
community-number	Community number configured by a **set community** command. Valid value is one of the following: • A number from 1 to 4294967200. You can specify a single number or multiple numbers separated by a space. • **internet**—The Internet community. • **no-export**—Routes with this community are sent to peers in other subautonomous systems within a confederation. Do not advertise this route to an External BGP (EBGP) peer. External systems are those outside the confederation. If there is no confederation, an external system is any EBGP peer. • **local-as**—Send this route to peers in other subautonomous systems within the local confederation. Do not advertise this route to an external system. • **no-advertise**—Do not advertise this route to any peer (internal or external).

match community

To match a BGP community, use the **match community** command in route-map configuration mode.

- **match community** *standard-list-number* | *expanded-list-number* | *community-list-name* [**exact**]

To remove the **match community** command from the configuration file and restore the system to its default condition where the software removes the BGP community list entry, use the **no** form of this command.

- **no match community** *standard-list-number* | *expanded-list-number* | *community-list-name* [**exact**]

Syntax Description

Parameter	Description
standard-list-number	Specifies a standard community-list number from 1 to 99 that identifies one or more permit or deny groups of communities.
expanded-list-number	Specifies an expanded community-list number from 100 to 199 that identifies one or more permit or deny groups of communities.
community-list-name	The community-list name.
exact	(Optional) Indicates that an exact match is required. All of the communities and only those communities specified must be present.

set comm-list delete

To remove communities from the community attribute of an inbound or outbound update, use the **set comm-list delete** command in route-map configuration mode.

- **set comm-list** *community-list-number* | *community-list-name* **delete**

To negate a previous **set comm-list delete** command, use the **no** form of this command.

- **no set comm-list** *community-list-number* | *community-list-name* **delete**

Syntax Description

Parameter	Description
community-list-number	A standard or expanded community-list number
community-list-name	A standard or expanded community-list name

Named Community List Examples

The following example creates a standard community-list that permits all routes except the routes with the communities 5 and 10 or 10 and 15:

```
Router(config)# ip community-list 1 deny 5 10
Router(config)# ip community-list 1 deny 10 15
Router(config)# ip community-list 1 permit internet
```

The following example creates a standard community-list that permits all routes within the local AS:

```
Router(config)# ip community-list 1 permit local-as
```

The following example creates a standard named community-list with the name COMMUNITY_A that permits all routes within the local AS and denies all routes with the internet community attribute:

```
Router(config)# ip community-list standard COMMUNITY_A permit local-AS
Router(config)# ip community-list standard COMMUNITY_A deny internet
```

The following example creates an expanded named community-list with the name COMMUNITY_B that will not advertise routes to EBGP peers:

```
Router(config)# ip community-list expanded COMMUNITY_B permit no-
export
```

The following example creates a named community-list with the name COMMUNITY_C that will not advertise this route to any EBGP or EBGP peers:

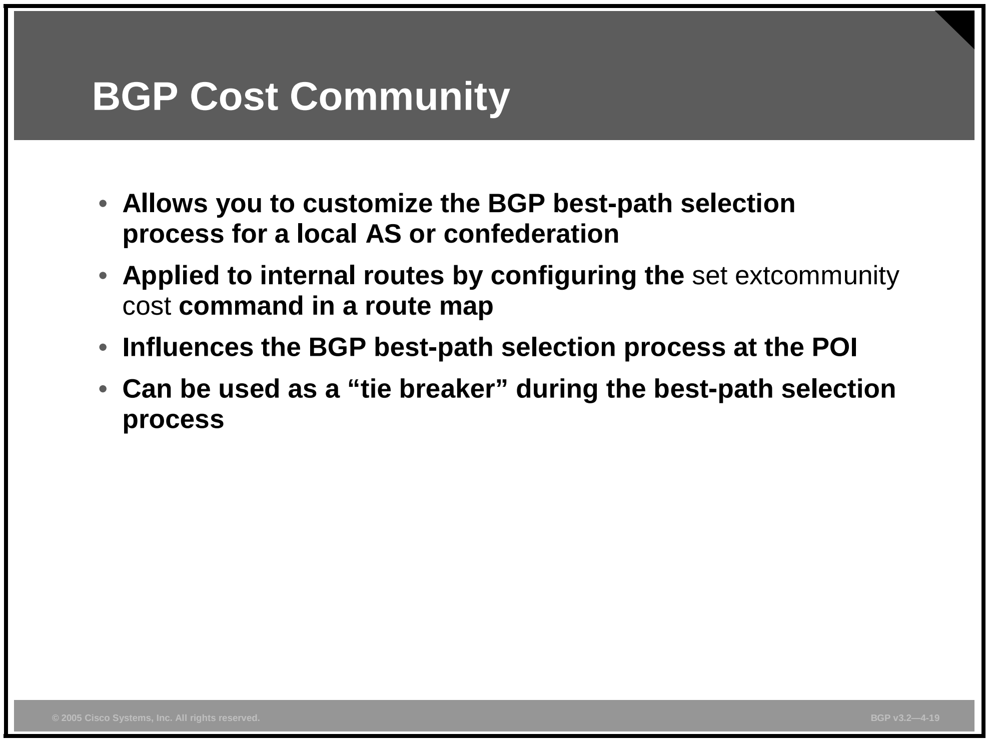
```
Router(config)# ip community-list expanded COMMUNITY_C permit no-
advertise
```

The following example uses a regular expression. The example creates a filter that will deny all communities that contain a number:

```
Router(config)# ip community-list 100 deny [0-9]*
```

BGP Cost Community

This topic describes the function of the BGP Cost Community feature.



BGP Cost Community

- **Allows you to customize the BGP best-path selection process for a local AS or confederation**
- **Applied to internal routes by configuring the** set extcommunity cost **command in a route map**
- **Influences the BGP best-path selection process at the POI**
- **Can be used as a "tie breaker" during the best-path selection process**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-19

The cost community is a nontransitive extended community attribute that is passed to Internal BGP (IBGP) and confederation peers but not to EBGP peers. The configuration of the BGP Cost Community feature allows you to customize the BGP best-path selection process for a local AS or confederation by assigning cost values to specific routes.

The cost community attribute is applied to internal routes by configuring the **set extcommunity cost** command in a route map.

set extcommunity cost

To create a set clause to apply the cost community attribute to routes that pass through a route map, use the **set extcommunity cost** command in route-map configuration mode.

- **set extcommunity cost** [**igp**] *community-id cost-value*

To delete the cost community set clause, use the **no** form of this command.

- **no set extcommunity cost** [**igp**] *community-id cost-value*

Syntax Description

Parameter	Description
igp	The IGP point of insertion (POI). The configuration of this keyword forces the cost community to be evaluated after the IGP distance to the next hop has been compared.
community-id	The ID for the configured extended community. The range is from 0 to 255.
cost-value	The configured cost that is set for matching paths in the route map. The range is from 0 to 4294967295.

The cost community set clause is configured with a cost community ID number (0 to 255) and cost number (0 to 4294967295). The cost number value determines the preference for the path. The path with the lowest cost community number is preferred. Paths that are not specifically configured with the cost community attribute are assigned a default cost number value of 2147483647 (the midpoint between 0 and 4294967295) and evaluated by the best-path selection process accordingly. When two paths have been configured with the same cost number value, the path selection process prefers the path with the lowest cost community ID. The cost extended community attribute is propagated to IBGP peers when extended community exchange is enabled with the **neighbor send-community** command.

The cost community attribute influences the BGP best-path selection process at the point of insertion (POI). By default, the POI follows the IGP metric comparison. When BGP receives multiple paths to the same destination, it uses the best-path selection process to determine which path is the best path. BGP automatically makes the decision and installs the best path into the routing table. The POI allows you to assign a preference to a specific path when multiple equal-cost paths are available. If the POI is not valid for local best-path selection, the cost community attribute is silently ignored.

Multiple paths can be configured with the cost community attribute for the same POI. The path with the lowest cost community ID is considered first. In other words, all of the cost community paths for a specific POI are considered, starting with the one with the lowest cost community ID. Paths that do not contain the cost community (for the POI and community ID being evaluated) are assigned the default community cost value (2147483647). If the cost community values are equal, then cost community comparison proceeds to the next-lowest community ID for this POI.

Applying the cost community attribute at the POI allows you to assign a value to a path originated or learned by a peer in any part of the local AS or confederation. The cost community can be used as a “tie breaker” during the best path selection process. Multiple instances of the cost community can be configured for separate equal-cost paths within the same AS or confederation. For example, a lower cost community value can be applied to a specific exit path in a network with multiple equal-cost exit points, and the specific exit path will be preferred by the BGP best-path selection process.

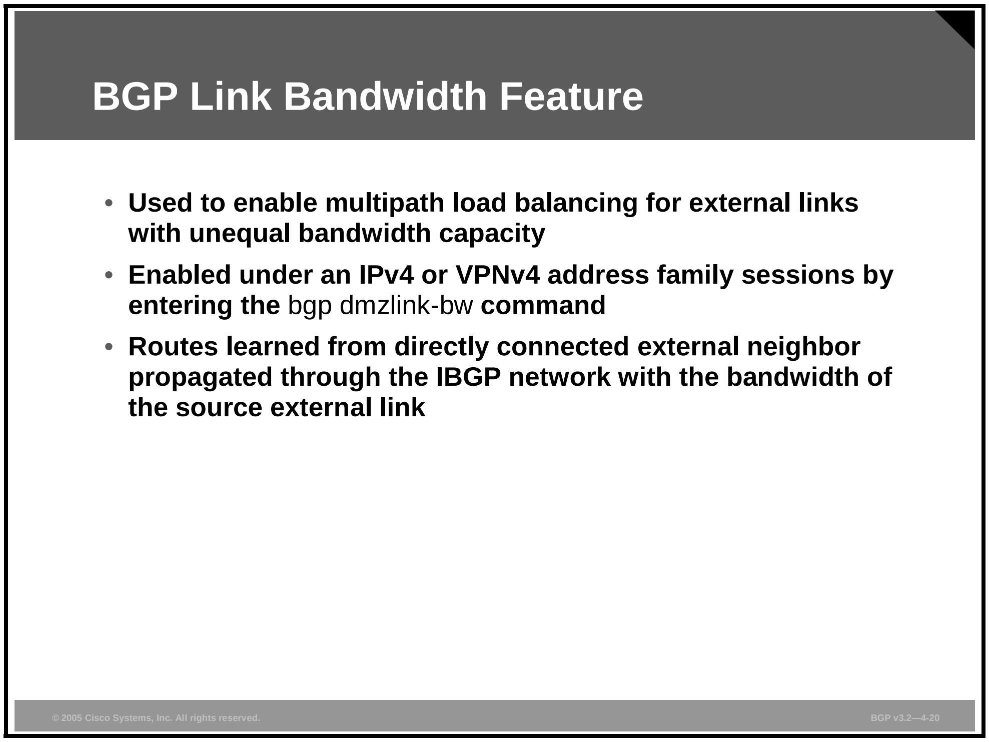
BGP Cost Community Configuration Example

The following example configuration shows the configuration of the **set extcommunity cost** command. This example applies the cost community ID of 1 and cost community value of 100 to routes that are permitted by the route-map. This configuration will cause the best-path selection process to prefer this route over other equal-cost paths that were not permitted by this route map sequence.

```
Router(config)# router bgp 50000
Router(config-router)# neighbor 10.0.0.1 remote-as 50000
Router(config-router)# neighbor 10.0.0.1 update-source Loopback 0
Router(config-router)# address-family ipv4
Router(config-router-af)# neighbor 10.0.0.1 activate
Router(config-router-af)# neighbor 10.0.0.1 route-map COST1 in
Router(config-router-af)# neighbor 10.0.0.1 send-community both
Router(config-router-af)# exit
Router(config)# route-map COST1 permit 10
Router(config-route-map)# match ip-address 1
Router(config-route-map)# set extcommunity cost 1 100
```

BGP Link Bandwidth

This topic describes the function of the BGP Link Bandwidth feature.



BGP Link Bandwidth Feature

- **Used to enable multipath load balancing for external links with unequal bandwidth capacity**
- **Enabled under an IPv4 or VPNv4 address family sessions by entering the** bgp dmzlink-bw **command**
- **Routes learned from directly connected external neighbor propagated through the IBGP network with the bandwidth of the source external link**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-20

The BGP Link Bandwidth feature is used to enable multipath load balancing for external links with unequal bandwidth capacity. This feature is enabled under an IP version 4 (IPv4) or Virtual Private Network version 4 (VPNv4) address-family session by entering the **bgp dmzlink-bw** command. This feature supports IBGP, EBGP multipath load balancing, and Enhanced Interior Gateway Routing Protocol (EIGRP) multipath load balancing in Multiprotocol Label Switching (MPLS) Virtual Private Networks (VPNs). When this feature is enabled, routes learned from directly connected external neighbors are propagated through the IBGP network with the bandwidth of the source external link.

The BGP Link Bandwidth feature allows BGP to be configured to send traffic over multiple IBGP or EBGP learned paths where the traffic that is sent is proportional to the bandwidth of the links that are used to exit the AS. The configuration of this feature can be used with EBGP and IBGP multipath features to enable unequal-cost load balancing over multiple links. Unequal-cost load balancing over links with unequal bandwidth was not possible in BGP before the BGP Link Bandwidth feature was introduced.

The link bandwidth extended community attribute indicates the preference of an AS exit link in terms of bandwidth. This extended community is applied to external links between directly connected EBGP peers by entering the **neighbor dmzlink-bw** command. The link bandwidth extended community attribute is propagated to IBGP peers when extended community exchange is enabled with the **neighbor send-community** command.

bgp dmzlink-bw

To configure BGP to distribute traffic proportionally over external links with unequal bandwidth when multipath load balancing is enabled, use the **bgp dmzlink-bw** command in address family configuration mode.

- **bgp dmzlink-bw**

To disable traffic distribution proportional to the link bandwidth, use the **no** form of this command.

- **no bgp dmzlink-bw**

This command has no keywords or arguments.

neighbor dmzlink-bw

To configure BGP to advertise the bandwidth of links that are used to exit an AS, use the **neighbor dmzlink-bw** command in address family configuration mode.

- **neighbor** *ip-address* **dmzlink-bw**

To disable link bandwidth advertisement, use the **no** form of this command.

- **no neighbor** *ip-address* **dmzlink-bw**

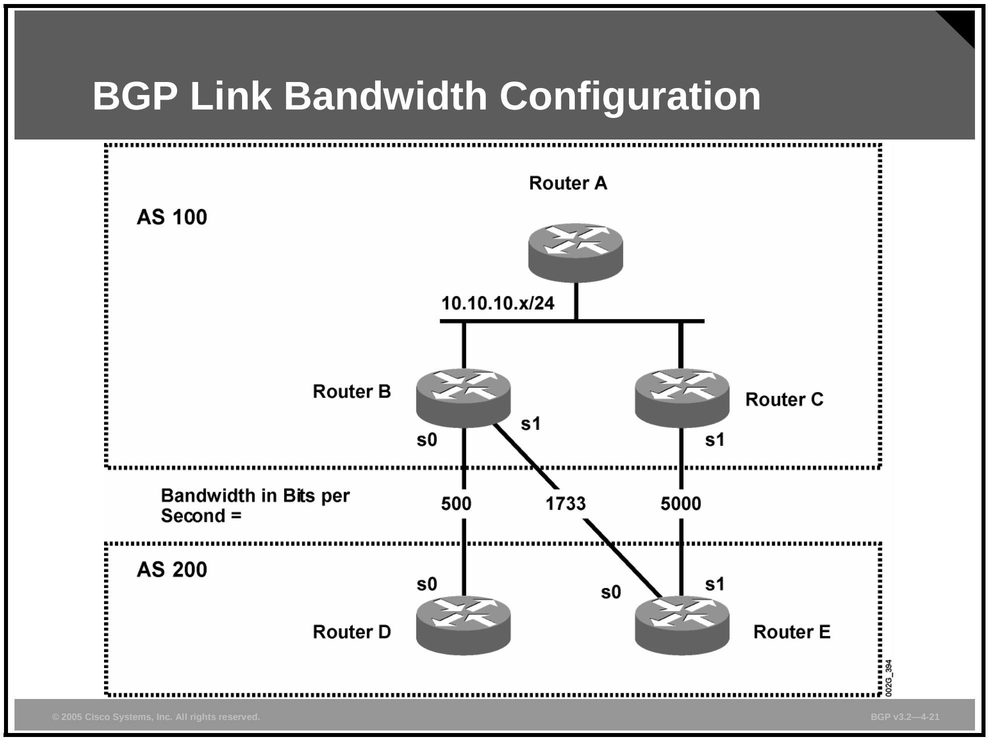
Syntax Description

Parameter	Description
ip-address	The IP address that identifies the external interface

The link bandwidth extended community attribute is a 4-byte value that is configured for a link on the demilitarized zone (DMZ) interface that connects two single-hop EBGP peers. The link bandwidth extended community attribute is used as a traffic-sharing value relative to other paths while forwarding traffic. Two paths are designated as equal for load balancing if the weight, local preference, AS-path length, MED, and Interior Gateway Protocol (IGP) costs are the same.

Example: BGP Link Bandwidth Configuration

In this example, the BGP Link Bandwidth feature is configured so that BGP will distribute traffic proportionally to the bandwidth of each external link.



BGP Link Bandwidth Configuration

The figure shows two external autonomous systems connected by three links, with each AS carrying a different amount of bandwidth (unequal-cost links). Multipath load balancing is enabled, and traffic is balanced proportionally.

Example: BGP Link Bandwidth Configuration

Router A Configuration

In the following example, Router A is configured to support IBGP multipath load balancing and to exchange the BGP extended community attribute with IBGP neighbors:

```
Router A(config)# router bgp 100
Router A(config-router)# neighbor 10.10.10.2 remote-as 100
Router A(config-router)# neighbor 10.10.10.2 update-source Loopback 0
Router A(config-router)# neighbor 10.10.10.3 remote-as 100
Router A(config-router)# neighbor 10.10.10.3 update-source Loopback 0
Router A(config-router)# address-family ipv4
Router A(config-router)# bgp dmzlink-bw
Router A(config-router-af)# neighbor 10.10.10.2 activate
Router A(config-router-af)# neighbor 10.10.10.2 send-community both
Router A(config-router-af)# neighbor 10.10.10.3 activate
Router A(config-router-af)# neighbor 10.10.10.3 send-community both
Router A(config-router-af)# maximum-paths ibgp 6
```

Router B Configuration

In the following example, Router B is configured to support multipath load balancing, to distribute Router D and Router E link traffic proportionally to the bandwidth of each link, and to advertise the bandwidth of these links to IBGP neighbors as an extended community:

```
Router B(config)# router bgp 100
Router B(config-router)# neighbor 10.10.10.1 remote-as 100
Router B(config-router)# neighbor 10.10.10.1 update-source Loopback 0
Router B(config-router)# neighbor 10.10.10.3 remote-as 100
Router B(config-router)# neighbor 10.10.10.3 update-source Loopback 0
Router B(config-router)# neighbor 172.16.1.1 remote-as 200
Router B(config-router)# neighbor 172.16.1.1 ebgp-multihop 1
Router B(config-router)# neighbor 172.16.2.2 remote-as 200
Router B(config-router)# neighbor 172.16.2.2 ebgp-multihop 1
Router B(config-router)# address-family ipv4
Router B(config-router-af)# bgp dmzlink-bw
Router B(config-router-af)# neighbor 10.10.10.1 activate
Router B(config-router-af)# neighbor 10.10.10.1 next-hop-self
Router B(config-router-af)# neighbor 10.10.10.1 send-community both
Router B(config-router-af)# neighbor 10.10.10.3 activate
Router B(config-router-af)# neighbor 10.10.10.3 next-hop-self
Router B(config-router-af)# neighbor 10.10.10.3 send-community both
Router B(config-router-af)# neighbor 172.16.1.1 activate
Router B(config-router-af)# neighbor 172.16.1.1 dmzlink-bw
Router B(config-router-af)# neighbor 172.16.2.2 activate
Router B(config-router-af)# neighbor 172.16.2.2 dmzlink-bw
Router B(config-router-af)# maximum-paths ibgp 6
Router B(config-router-af)# maximum-paths 6
```

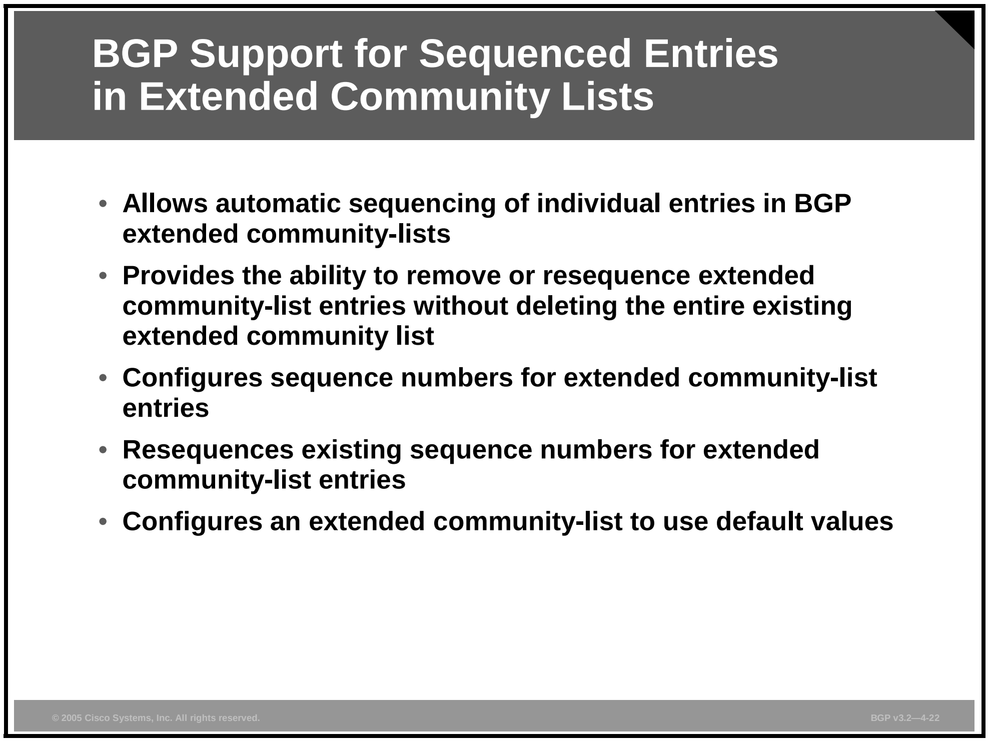
Router C Configuration

In the following example, Router C is configured to support multipath load balancing and to advertise the bandwidth of the link with Router E to IBGP neighbors as an extended community:

```
Router C(config)# router bgp 100
Router C(config-router)# neighbor 10.10.10.1 remote-as 100
Router C(config-router)# neighbor 10.10.10.1 update-source Loopback 0
Router C(config-router)# neighbor 10.10.10.2 remote-as 100
Router C(config-router)# neighbor 10.10.10.2 update-source Loopback 0
Router C(config-router)# neighbor 172.16.3.30 remote-as 200
Router C(config-router)# neighbor 172.16.3.30 ebgp-multihop 1
Router C(config-router)# address-family ipv4
Router C(config-router-af)# bgp dmzlink-bw
Router C(config-router-af)# neighbor 10.10.10.1 activate
Router C(config-router-af)# neighbor 10.10.10.1 send-community both
Router C(config-router-af)# neighbor 10.10.10.1 next-hop-self
Router C(config-router-af)# neighbor 10.10.10.2 activate
Router C(config-router-af)# neighbor 10.10.10.2 send-community both
Router C(config-router-af)# neighbor 10.10.10.2 next-hop-self
Router C(config-router-af)# neighbor 172.16.3.3 activate
Router C(config-router-af)# neighbor 172.16.3.3 dmzlink-bw
Router C(config-router-af)# maximum-paths ibgp 6
Router C(config-router-af)# maximum-paths 6
```

BGP Support for Sequenced Entries in Extended Community Lists

This topic describes the BGP Support for Sequenced Entries in Extended Community Lists feature.



BGP Support for Sequenced Entries in Extended Community Lists

- **Allows automatic sequencing of individual entries in BGP extended community-lists**
- **Provides the ability to remove or resequence extended community-list entries without deleting the entire existing extended community list**
- **Configures sequence numbers for extended community-list entries**
- **Resequences existing sequence numbers for extended community-list entries**
- **Configures an extended community-list to use default values**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-22

This feature allows automatic sequencing of individual entries in BGP extended community-lists. This feature also provides the ability to remove or resequence extended community-list entries without deleting the entire existing extended community-list.

Both named and numbered extended community-lists can be configured in IP extended community-list configuration mode. To enter IP extended community-list configuration mode, issue the **ip extcommunity-list** command with either the **expanded** or **standard** keyword followed by the extended community-list name. This configuration mode supports all of the functions that are available in global configuration mode. In addition, you can perform the following operations:

- Configure sequence numbers for extended community-list entries
- Resequence existing sequence numbers for extended community-list entries
- Configure an extended community-list to use default values

ip extcommunity-list

To create an extended community-list and control access to it, use the **ip extcommunity-list** command in global configuration mode.

- **ip extcommunity-list** *expanded-list-number* | **expanded** *list-name* {**permit** | **deny**} [*regular-expression*] | *standard-list-number* | **standard** *list-name* {**permit** | **deny**} [**rt** *extcom-value*] [**soo** *extcom-value*]

To delete the entire community-list, use the **no** form of this command.

- **no ip extcommunity-list** *expanded-list-number* | **expanded** *list-name* | *standard-list-number* | **standard** *list-name*

Syntax Description

Parameter	Description
expanded-list-number	An expanded list number from 100 to 500 that identifies one or more permit or deny groups of extended communities.
standard-list-number	A standard list number from 1 to 99 that identifies one or more permit or deny groups of extended communities.
expanded *list-name*	Creates an expanded named extended community-list and enters IP extended community-list configuration mode.
standard *list-name*	Creates a standard named extended community-list and enters IP extended community-list configuration mode.
permit	Permits access for a matching condition.
deny	Denies access for a matching condition.
regular-expression	(Optional) An input string pattern to match against.
rt	(Optional) Specifies the route target (RT) extended community attribute. The **rt** keyword can be configured only with standard extended community-lists and not expanded community-lists.
soo	(Optional) Specifies the site of origin (SOO) extended community attribute. The **soo** keyword can be configured only with standard extended community-lists and not expanded community-lists.
extcom-value	Specifies the RT or SOO extended community value. The value can be one of the following combinations: ■ autonomous-system-number : network-number ■ ip-address : network-number The colon is used to separate the AS number and network number or IP address and network number.
sequence-number	(Optional) The sequence number of a named or numbered extended community-list. This value can be a number from 1 to 2147483647.

Parameter	Description
default	(Optional) Sets a keyword or argument to default behavior or value.
exit	(Optional) Exits IP extended community-list configuration mode.
resequence	(Optional) Changes the sequences of extended community-list entries to the default sequence numbering or to the specified sequence numbering. Extended community entries are sequenced by 10-number increments by default.
starting-sequence	(Optional) Specifies the number for the first entry in an extended community-list.
sequence-increment	(Optional) Specifies the increment range for each subsequent extended community-list entry.

Sequenced Extended Community-List Entry Configuration: Example

The following example creates and configures a named extended community-list that will permit routes only from RT 64512:10, 65000:20, 64535:30 and SOO 65535:40. All other routes are implicitly denied.

```
Router(config)# ip extcommunity-list standard NAMED_LIST

Router(config-extcom-list)# 1 permit rt 64512:10

Router(config-extcom-list)# 2 permit rt 65000:20

Router(config-extcom-list)# 3 permit rt 64535:30

Router(config-extcom-list)# 4 permit soo 65535:40

Router(config-extcom-list)# end
```

Resequenced Extended Community-List Entry Configuration: Example

The following example resequences the extended community-list entries in the named community-list. The first entry is resequenced to the number 50 and the range for each subsequent entry to follow by 100 (for example, 150, 250, 350, and so on):

```
Router(config)# ip extcommunity-list standard NAMED_LIST

Router(config-extcom-list)# resequence 50 100

Router(config-extcom-list)# end
```

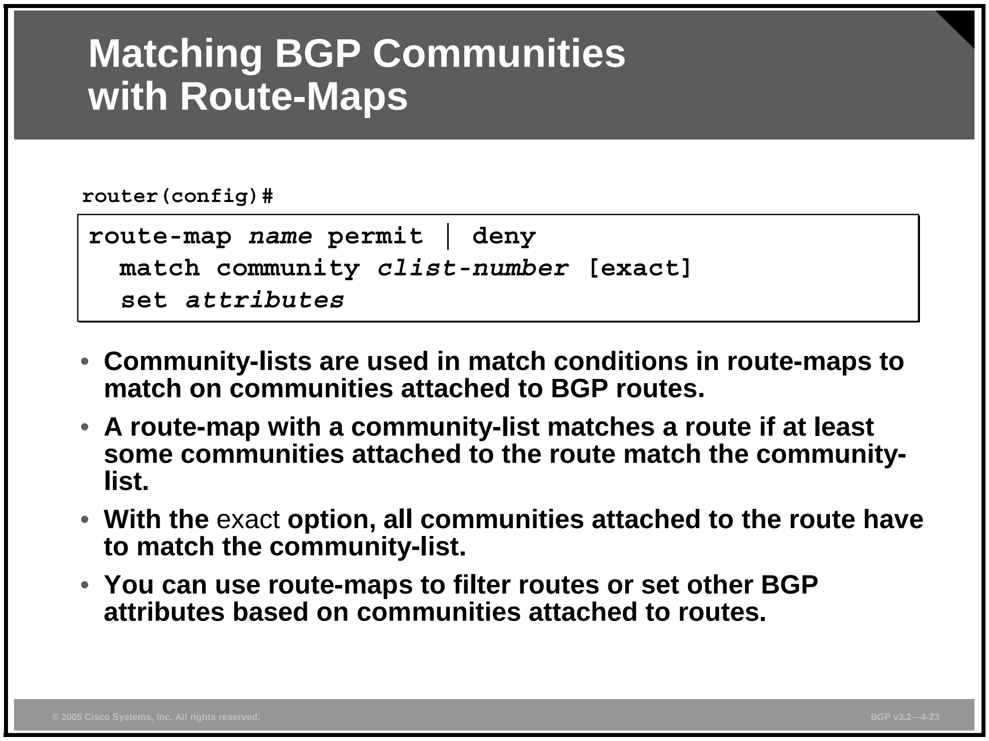
Sequenced Extended Community-List Entry Verification: Example

The following example uses the **show ip extcommunity-list** EXEC command to display routes that are permitted by the named extended community-list. This example also shows the configuration from the first example after it has been resequence with user-defined values.

```
Router> show ip extcommunity-list

Standard extended community-list NAMED_LIST

     50 permit RT:64512:10

     150 permit RT:64512:10

     250 permit RT:64512:10

     350 permit RT:64512:10
```

Matching BGP Communities with Route-Maps

This topic discusses the command that is required to match routes based on attached BGP communities using route-maps.



Matching BGP Communities with Route-Maps

```
router(config)#
```

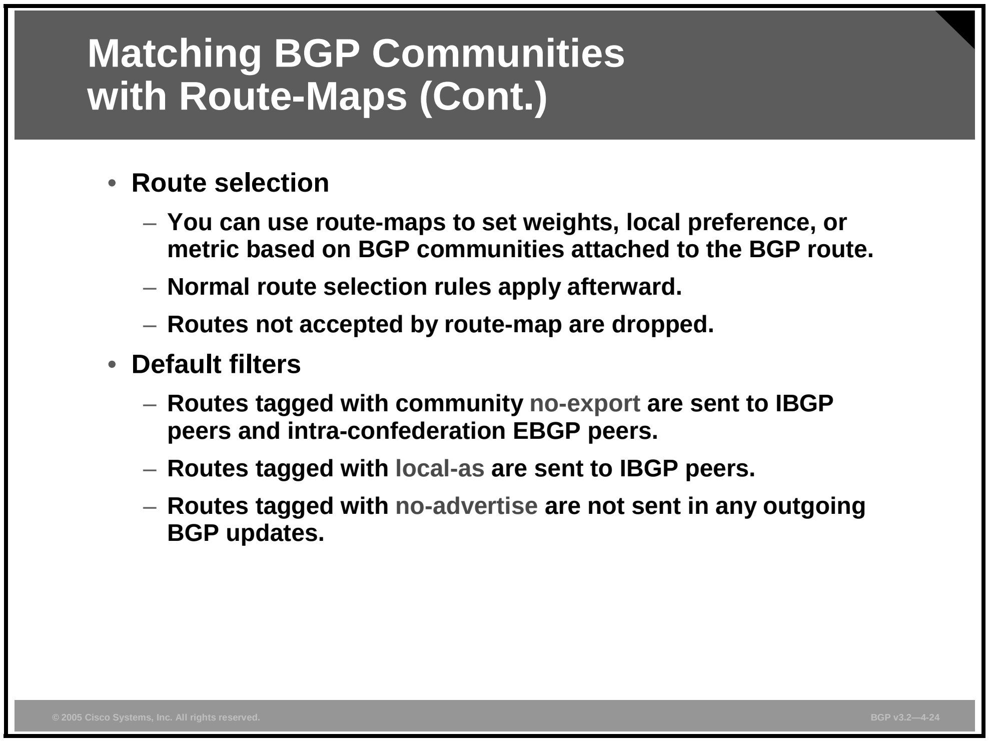
```
route-map name permit | deny
  match community clist-number [exact]
  set attributes
```

- **Community-lists are used in match conditions in route-maps to match on communities attached to BGP routes.**
- **A route-map with a community-list matches a route if at least some communities attached to the route match the community-list.**
- **With the** exact **option, all communities attached to the route have to match the community-list.**
- **You can use route-maps to filter routes or set other BGP attributes based on communities attached to routes.**

© 2005 Cisco Systems, Inc. All rights reserved.   BGP v3.2—4-23

Network administrators use route-maps to match networks that carry a subset of communities that are permitted by the community-list. Other parameters or attributes can then be set based on community values. If you use the keyword **exact**, all communities that are attached to a BGP route have to be matched by the community-list.

You can use a route-map to filter or modify BGP routing updates. Any BGP-related **set** commands can be used to set BGP parameters and attributes (that is, weight, local preference, and MED).



Matching BGP Communities with Route-Maps (Cont.)

- **Route selection**
  - **You can use route-maps to set weights, local preference, or metric based on BGP communities attached to the BGP route.**
  - **Normal route selection rules apply afterward.**
  - **Routes not accepted by route-map are dropped.**
- **Default filters**
  - **Routes tagged with community no-export are sent to IBGP peers and intra-confederation EBGP peers.**
  - **Routes tagged with local-as are sent to IBGP peers.**
  - **Routes tagged with no-advertise are not sent in any outgoing BGP updates.**

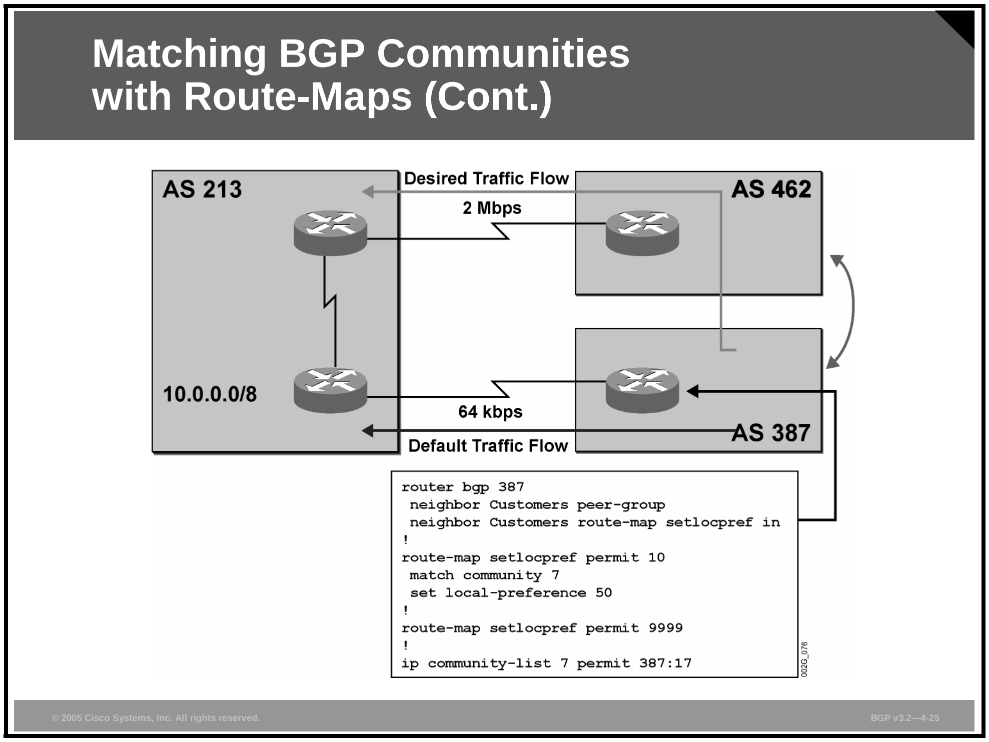
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—4-24

As mentioned before, there are some predefined community values that cause routers to automatically filter routing updates:

- **no-advertise:** If a router receives an update carrying this community, the router will not forward it to any neighbor.
- **local-as:** This community has a similar meaning to **no-export**, but it keeps a route within the local subautonomous system. The route is not propagated to intra-confederation external neighbors or to any other external neighbors.
- **no-export:** If a router receives an update carrying this community, the router will not propagate it to any external neighbors except to intra-confederation external neighbors.
- **internet:** This value advertises this route to the Internet community, to which all routers belong.

Example: Matching BGP Communities with Route-Maps

This example shows a configuration that translates community 387:17 into local preference 50.



Matching BGP Communities with Route-Maps (Cont.)

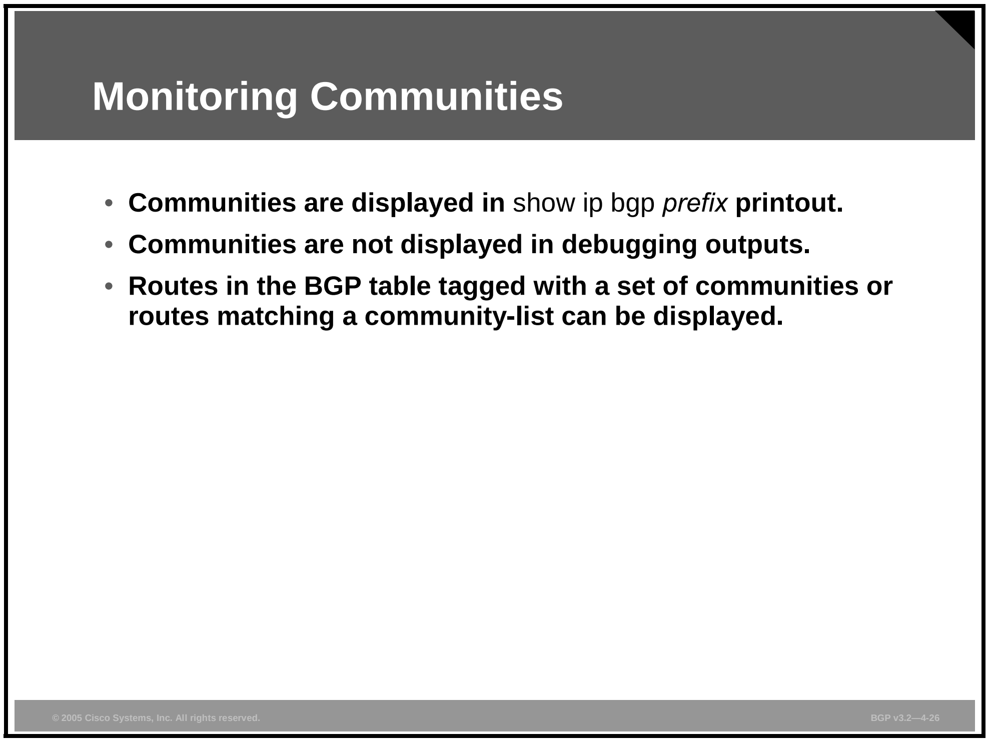
```
router bgp 387
 neighbor Customers peer-group
 neighbor Customers route-map setlocpref in
!
route-map setlocpref permit 10
 match community 7
 set local-preference 50
!
route-map setlocpref permit 9999
!
ip community-list 7 permit 387:17
```

All updates that are received from neighboring AS 213 are processed by the route-map, which uses a community-list to find community 387:17. If the community-list matches one of the community attributes, the **set** command is executed and the route is permitted. If the route does not contain the right community, the route is simply permitted by route-map statement 9999 without changing anything in the update.

The result is that AS 387 prefers other paths to AS 213 because they have a default local preference of 100.

Monitoring Communities

This topic lists the commands that are required to monitor BGP communities.

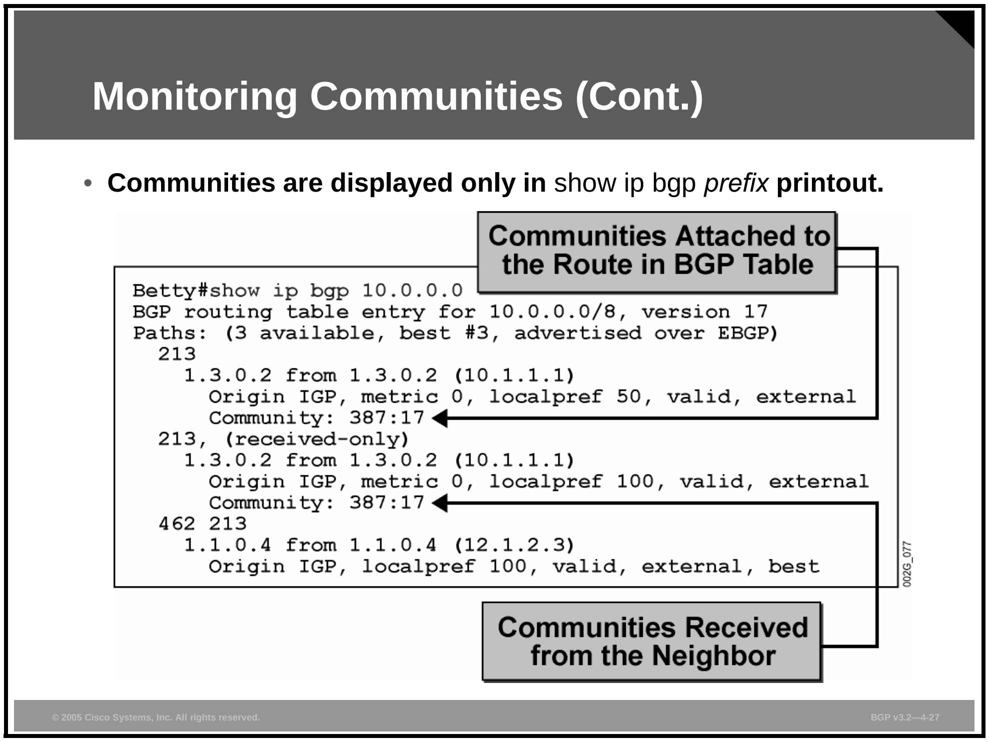


Monitoring Communities

- **Communities are displayed in** show ip bgp *prefix* **printout.**
- **Communities are not displayed in debugging outputs.**
- **Routes in the BGP table tagged with a set of communities or routes matching a community-list can be displayed.**

Because a community is an attribute that can appear more than once in a single update, the **show ip bgp** command does not show it. You can view communities only if you use the **show ip bgp** *prefix* command.

If you use the **show ip bgp community-list** command, all networks that are permitted by the community-list are listed.



Monitoring Communities (Cont.)

- **Communities are displayed only in** show ip bgp *prefix* **printout.**

```
Betty#show ip bgp 10.0.0.0
BGP routing table entry for 10.0.0.0/8, version 17
Paths: (3 available, best #3, advertised over EBGP)
  213
    1.3.0.2 from 1.3.0.2 (10.1.1.1)
      Origin IGP, metric 0, localpref 50, valid, external
      Community: 387:17
  213, (received-only)
    1.3.0.2 from 1.3.0.2 (10.1.1.1)
      Origin IGP, metric 0, localpref 100, valid, external
      Community: 387:17
  462 213
    1.1.0.4 from 1.1.0.4 (12.1.2.3)
      Origin IGP, localpref 100, valid, external, best
```

Communities Attached to the Route in BGP Table
Communities Received from the Neighbor
002G_077

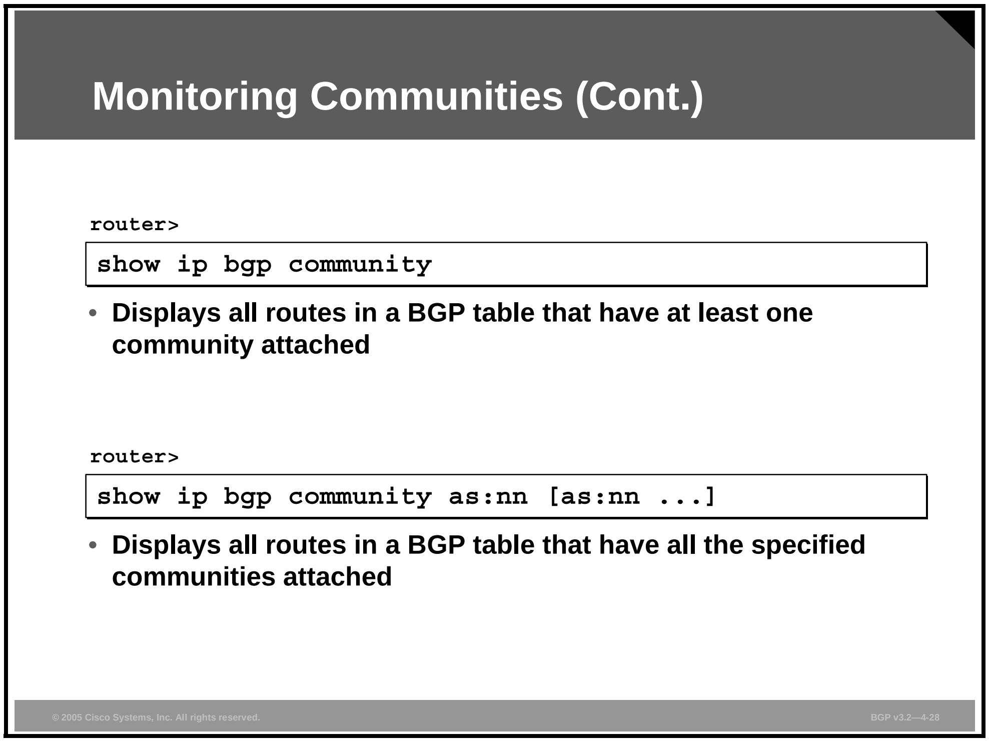
BGP v3.2—4-27

This example shows the output of the **show ip bgp** *prefix* command where inbound soft reconfiguration was enabled on one of the neighbors. The original update contained one single community attribute (387:17), which can be seen from the second path marked with “received-only.” This update was then processed by an inbound route-map, which matched the community 387:17 and changed the local preference of the received route to 50.

Another use of the **show** command is to filter the output of the **show ip bgp** command.

If the keyword **community** is included, all networks that have at least one community attribute are displayed.

If the keyword **community** is followed by one or more community values, only the networks that carry all those communities are displayed.



Monitoring Communities (Cont.)

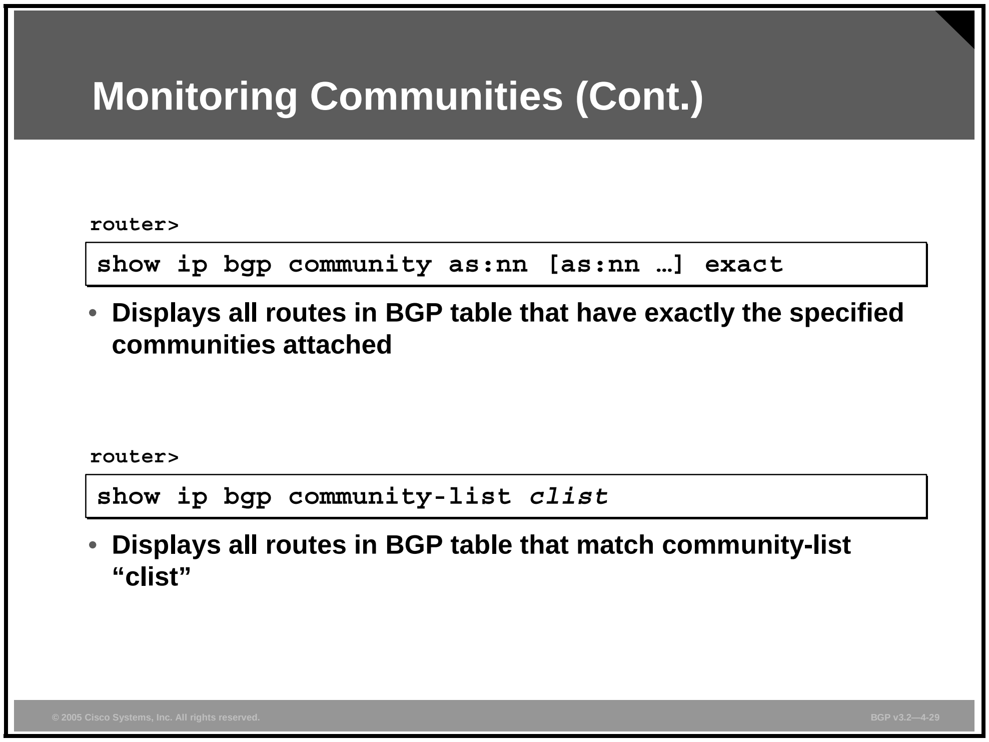
```
router>
show ip bgp community
```

- **Displays all routes in a BGP table that have at least one community attached**

```
router>
show ip bgp community as:nn [as:nn ...]
```

- **Displays all routes in a BGP table that have all the specified communities attached**

BGP v3.2—4-28



Monitoring Communities (Cont.)

```
router>
show ip bgp community as:nn [as:nn …] exact
```

- **Displays all routes in BGP table that have exactly the specified communities attached**

```
router>
show ip bgp community-list clist
```

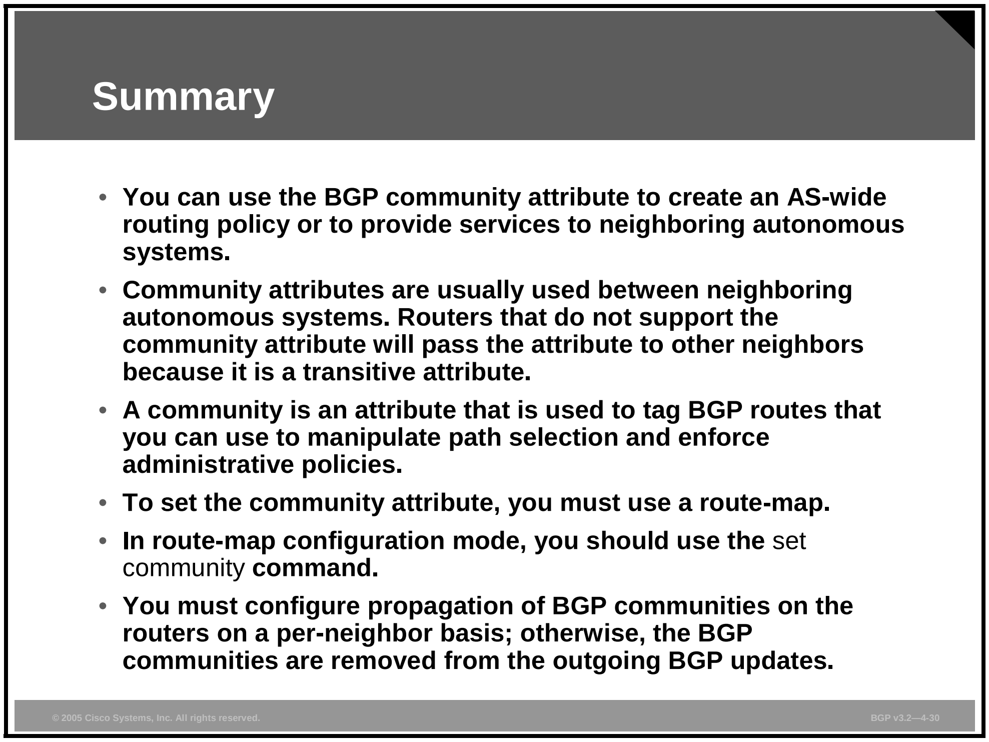
- **Displays all routes in BGP table that match community-list "clist"**

If the keyword **exact** is added at the end, only the networks that match exactly are displayed.

You can also use a community-list to filter the output of the **show ip bgp** command.

Summary

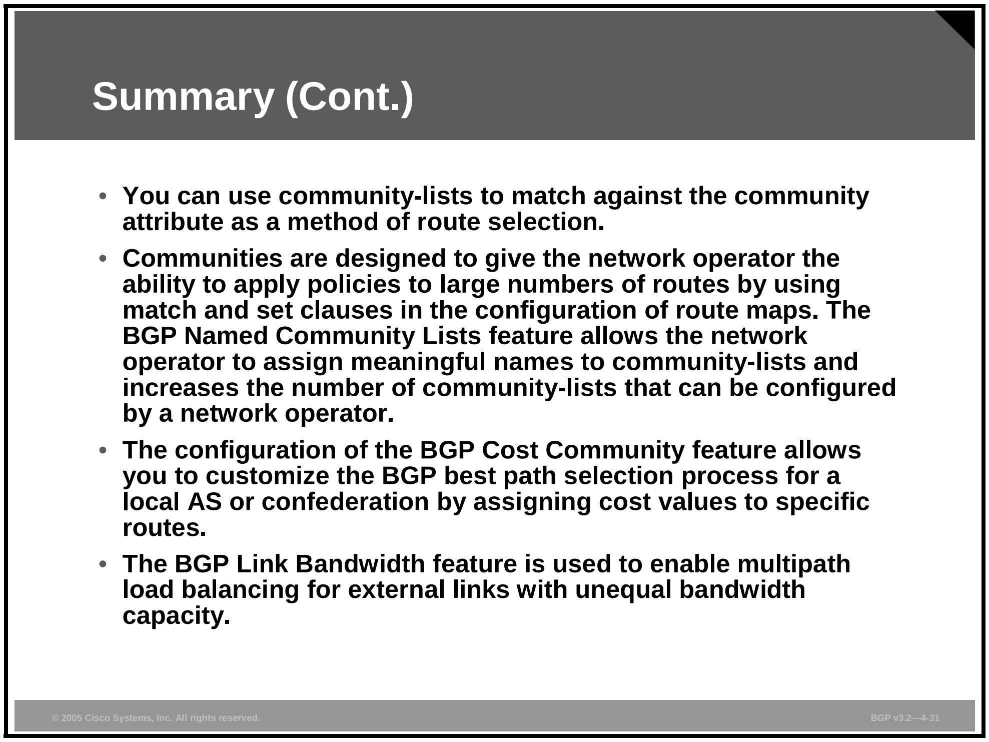
This topic summarizes the key points discussed in this lesson.



Summary

- You can use the BGP community attribute to create an AS-wide routing policy or to provide services to neighboring autonomous systems.
- Community attributes are usually used between neighboring autonomous systems. Routers that do not support the community attribute will pass the attribute to other neighbors because it is a transitive attribute.
- A community is an attribute that is used to tag BGP routes that you can use to manipulate path selection and enforce administrative policies.
- To set the community attribute, you must use a route-map.
- In route-map configuration mode, you should use the set community command.
- You must configure propagation of BGP communities on the routers on a per-neighbor basis; otherwise, the BGP communities are removed from the outgoing BGP updates.

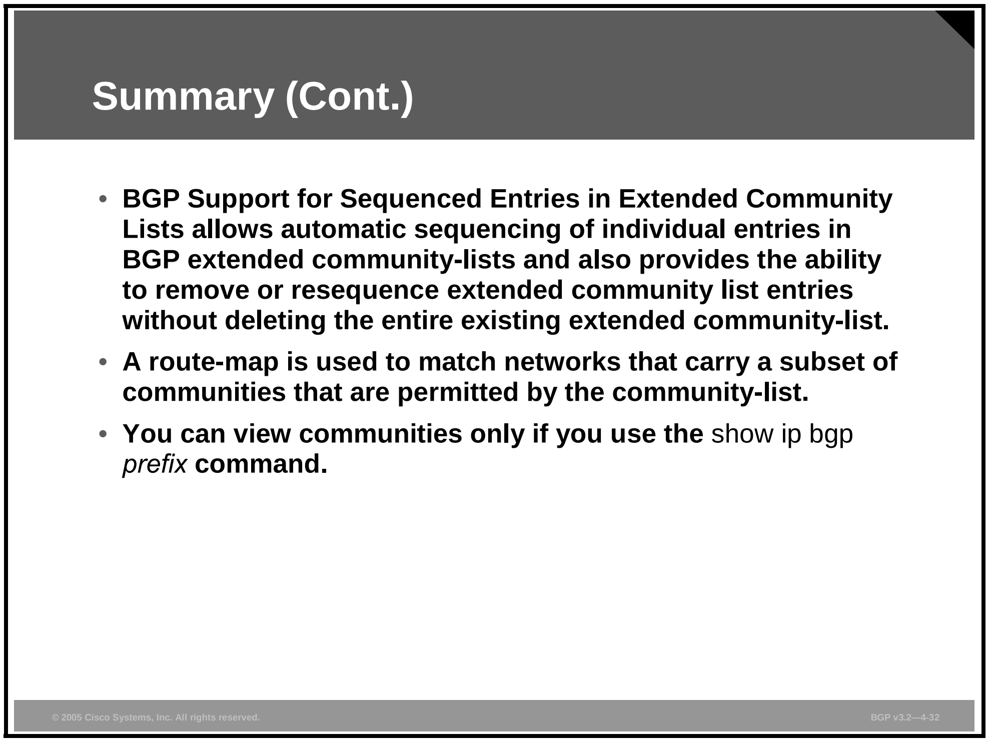
BGP v3.2—4-30



Summary (Cont.)

- You can use community-lists to match against the community attribute as a method of route selection.
- Communities are designed to give the network operator the ability to apply policies to large numbers of routes by using match and set clauses in the configuration of route maps. The BGP Named Community Lists feature allows the network operator to assign meaningful names to community-lists and increases the number of community-lists that can be configured by a network operator.
- The configuration of the BGP Cost Community feature allows you to customize the BGP best path selection process for a local AS or confederation by assigning cost values to specific routes.
- The BGP Link Bandwidth feature is used to enable multipath load balancing for external links with unequal bandwidth capacity.

BGP v3.2—4-31

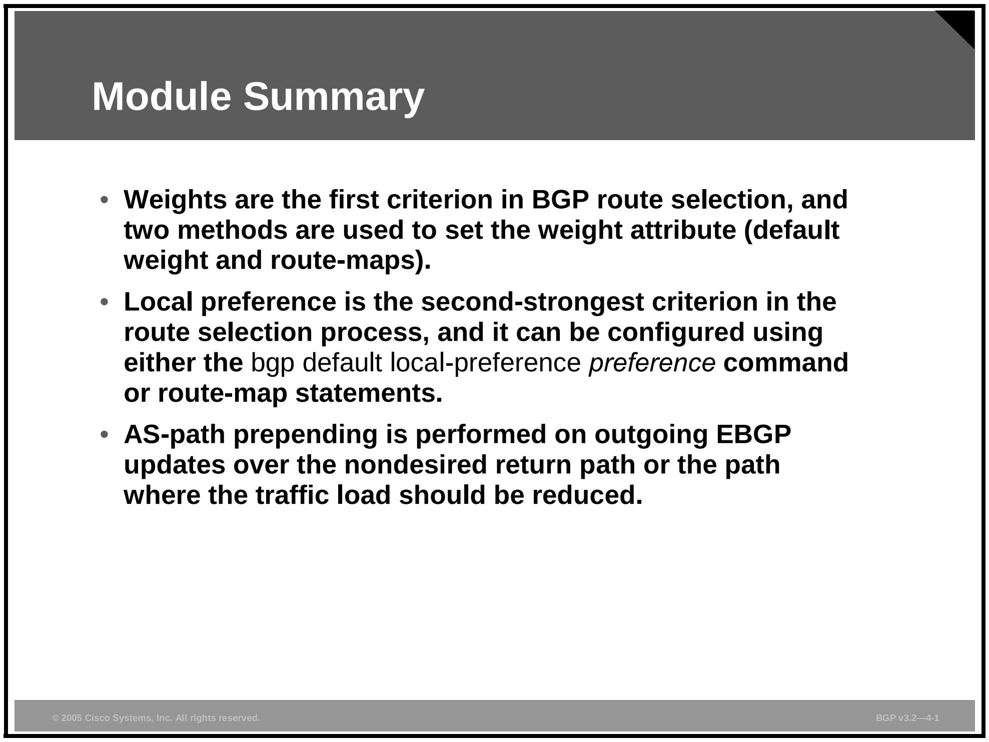


Summary (Cont.)

- **BGP Support for Sequenced Entries in Extended Community Lists allows automatic sequencing of individual entries in BGP extended community-lists and also provides the ability to remove or resequence extended community list entries without deleting the entire existing extended community-list.**
- **A route-map is used to match networks that carry a subset of communities that are permitted by the community-list.**
- **You can view communities only if you use the** show ip bgp *prefix* **command.**

Module Summary

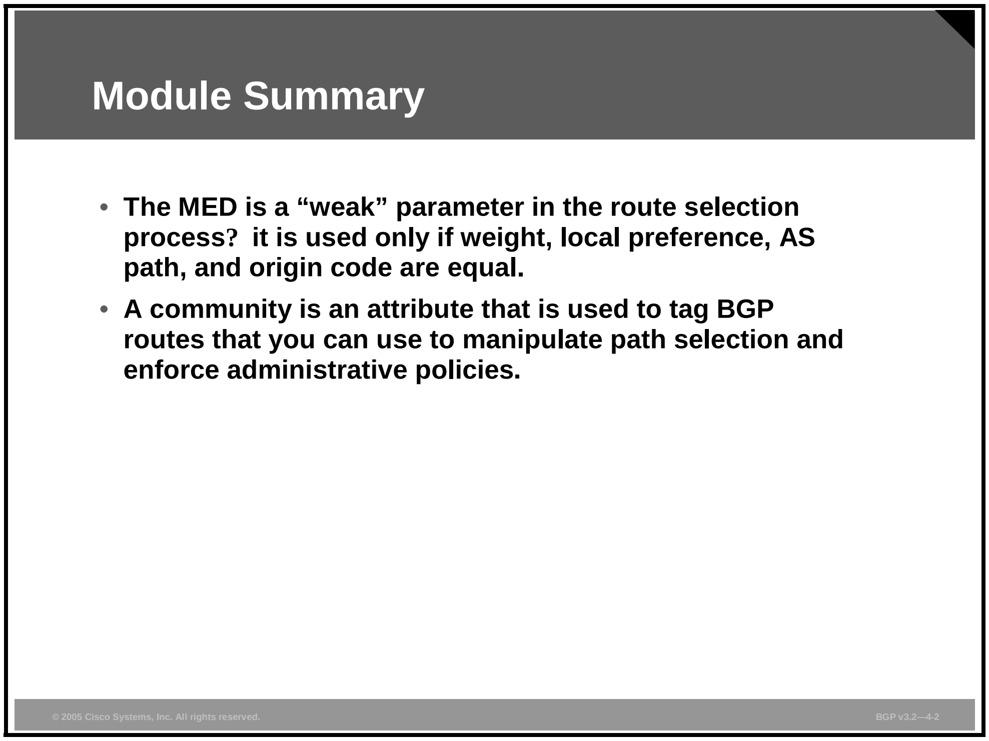
This topic summarizes the key points discussed in this module.



Module Summary

- **Weights are the first criterion in BGP route selection, and two methods are used to set the weight attribute (default weight and route-maps).**
- **Local preference is the second-strongest criterion in the route selection process, and it can be configured using either the** bgp default local-preference *preference* **command or route-map statements.**
- **AS-path prepending is performed on outgoing EBGP updates over the nondesired return path or the path where the traffic load should be reduced.**

BGP v3.2—4-1



Module Summary

- **The MED is a “weak” parameter in the route selection process? it is used only if weight, local preference, AS path, and origin code are equal.**
- **A community is an attribute that is used to tag BGP routes that you can use to manipulate path selection and enforce administrative policies.**

BGP v3.2—4-2

This module discussed attributes influencing route selection. The first lesson described how to influence BGP route selection by setting the weight attribute of incoming BGP routes. The second lesson discussed how to influence BGP route selection by setting the BGP local preference attribute of incoming BGP routes. The third lesson described AS-path prepending and the Cisco IOS commands required to properly configure and monitor AS-path configurations. The fourth lesson explained how to influence BGP route selection by setting the BGP MED attribute of outgoing BGP routes. Finally, the fifth lesson described how to influence BGP route selection by setting the BGP community attribute on outgoing BGP routes and discussed BGP communities and their use to facilitate proper return path selection.

References

For additional information, refer to these resources:

- Cisco Systems, Inc. *Border Gateway Protocol*. http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/bgp.htm.
- Cisco Systems, Inc. *Configuring BGP*. http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgcr/fipr_c/ipcprt2/1cfbgp.htm#xtocid15.
- Cisco Systems, Inc. *Using the Border Gateway Protocol for Interdomain Routing*. http://www.cisco.com/warp/public/459/14.html.
- Cisco Systems, Inc. *BGP Case Studies*. http://www.cisco.com/warp/public/459/bgp-toc.html.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) What is the difference between local preference and weight? (Source: Influencing BGP Route Selection with Weights)

A) Local preference has a higher priority in BGP path selection.
B) Local preference is used AS-wide while weight is local to a single router.
C) Local preference is local only to a specific BGP-speaking router.
D) Local preference is used to influence incoming path selection.

Q2) What is the default weight for routes that are received from a BGP neighbor? (Source: Influencing BGP Route Selection with Weights)

A) 0
B) 100
C) 32768
D) depends on the Cisco IOS release

Q3) When are the weights that are configured on a neighbor enforced? (Source: Influencing BGP Route Selection with Weights)

A) Before the new weights can take effect, the BGP process on the router must be removed and reconfigured.
B) The router must first be rebooted for the new weights to take effect.
C) The new weights will be applied after the BGP update interval of 30 minutes expires.
D) The new weight configuration is applied to all routes that are received following the configuration change.

Q4) How could you implement a primary/backup ISP routing policy by using weights? (Source: Influencing BGP Route Selection with Weights)

A) assign higher weights to all routes that are received from the backup ISP
B) assign lower weights to all routes that are received from the backup ISP
C) assign higher weights to all routes that are received from the primary ISP
D) assign lower weights to all routes that are received from the primary ISP

Q5) When you are using route-maps to modify weights, what happens by default to a route that does not match any of the route-map statements? (Source: Influencing BGP Route Selection with Weights)

A) The route is accepted with the weight attribute unmodified.
B) The route is discarded.
C) The route is inserted into the BGP table but not into the IP routing table.
D) An error is displayed on the router console and in router debugs.

Q6) Which method of influencing route selection with weights is the last to be applied on an incoming interface? (Source: Influencing BGP Route Selection with Weights)

A) prefix-list
B) route-map
C) filter-list weight
D) default weight

Q7) Which two of the following statements are correct regarding BGP route selection? (Choose two.) (Source: Influencing BGP Route Selection with Weights)

A) If two routes have the same weight attribute, the route with the lowest local preference is chosen.
B) The route with the highest weight is always chosen first.
C) The weight attribute is global within an AS.
D) The weight attribute is only local to the local router.
E) The weight value is propagated by all routers.

Q8) What is a key difference between the local preference and weight attributes? (Source: Setting BGP Local Preference)

A) Local preference is local to the route on which it is configured.
B) Local preference is local to the AS within which it has been configured.
C) Local preference is local to the BGP administrative domain.
D) Local preference is global to a BGP domain.

Q9) Which two of the following statements about the influence of local preference on BGP route selection is accurate? (Choose two.) (Source: Setting BGP Local Preference)

A) When you set local preference, you can view it on neighboring routers, but you must reset it.
B) You can use local preference to ensure AS-wide route selection policy.
C) Local preference is used to select routes with unequal weights.
D) Local preference is the second-strongest criterion in the route selection process.

Q10) Which Cisco IOS command is used to change the default value of local preference? (Source: Setting BGP Local Preference)

A) **set local-preference**
B) **bgp default local-preference**
C) **show ip bgp**
D) **show ip bgp** *prefix*

Q11) Which Cisco IOS command is used to configure BGP local preference with route-map statements? (Source: Setting BGP Local Preference)

A) **set local-preference**
B) **bgp default local-preference**
C) **show ip bgp**
D) **show ip bgp** *prefix*

Q12) Which Cisco IOS command is necessary to display the locally applied BGP value? (Source: Setting BGP Local Preference)

A) **show bgp preference detail**
B) **show ip bgp**
C) **show ip bgp detail**
D) **show ip bgp** *prefix*

Q13) What is AS-path prepending? (Source: Using AS-Path Prepending)

A) when a router, sending a BGP update, adds the AS number of the router from which it received the route to the AS-path attribute
B) when a router, sending a BGP update, adds the AS number of the router to which it is sending the route to the AS-path attribute
C) when a router, sending a BGP update, adds its AS number to the AS-path attribute multiple times
D) when a router uses the AS-path attribute in route selection

Q14) The AS path will be the route selection criterion that is used when which of the following is true? (Source: Using AS-Path Prepending)

A) It is the first criterion that is used in BGP route selection.
B) It is used when there is no difference in weight, local preference, or route origination.
C) It is used when the MED is identical on the candidate routes.
D) The weight, local preference, MED, and origin attributes must be identical before the AS-path attribute is used for route selection.

Q15) Which command do you use to manipulate the AS-path attribute? (Source: Using AS-Path Prepending)

A) the global configuration command **set as-path prepend** *as-number*
B) the router configuration command **set as-path prepend** *as-number*
C) **set as-path prepend** *as-number* in a route-map
D) the interface global command **set as-path prepend** *as-number*

Q16) The following configuration is from a router in AS 347, which is advertising network 11.0.0.0/8 to an EBGP neighbor 2.0.0.2 in AS 529:

```
route-map addAS permit 10
  set as-path prepend 347 347 347

router bgp 347
  neighbor 2.0.0.2 remote-as 529
  neighbor 2.0.0.2 route-map addAS out
```

What are the contents of the AS-path attribute for route 11.0.0.0/8 on a router that is residing in AS 529? (Source: Using AS-Path Prepending)

A) 347 347 347
B) 347 347 347 347
C) 529 347 347 347
D) 529 347 347 347 347

Q17) Why do network administrators need to use AS-path prepending? (Source: Using AS-Path Prepending)

A) AS-path prepending allows a customer to potentially influence return path route selection.
B) AS-path prepending is used on a customer router to control outgoing route updates.
C) Service providers use AS-path prepending to control incoming updates from a customer AS.
D) AS-path prepending is used between service providers that are connected to the same customer AS to determine which will be the primary link to the customer.

Q18) How does AS-path prepending affect a router? (Source: Using AS-Path Prepending)

- A) AS-path prepending is simply a term that is used to describe when a router uses the AS-path attribute in route selection and hence does not affect router resources.
- B) The longer the AS-path attribute attached to BGP updates, the more router memory requirements increase.
- C) AS-path prepending does not impact the router because Cisco IOS software recognizes that AS-path prepending is in use and stores a single AS number with a pointer to the number of AS-path prepends.
- D) AS-path prepending causes the router to operate in process-switching mode because the BGP update must be stored, manipulated, and then rewritten to accommodate the new AS-path attribute.

Q19) Which two of the following are characteristics of the function of the BGP Hide Local-Autonomous System feature? (Choose two.) (Source: Using AS-Path Prepending)

- A) allows you to transparently change the AS number for the entire BGP network
- B) ensures that routes can be propagated throughout the AS
- C) allows customization of the AS number for EBGP peer groupings through the **set as-path** command
- D) changes the AS number for all IBGP peers at the same time

Q20) What is the typical application of the MED attribute? (Source: Understanding BGP Multi-Exit Discriminators)

- A) to influence path selection out of an originating AS
- B) to provide a strong metric to select the best path when multiple routes exist
- C) to have a BGP attribute traversing many autonomous systems while influencing path selection
- D) to influence the return path of traffic back into an AS

Q21) What are three BGP attributes that are compared before the MED? (Choose three.) (Source: Understanding BGP Multi-Exit Discriminators)

- A) largest weight
- B) originated routes
- C) AS-path length
- D) lowest IP address

Q22) Which two statements about the Cisco IOS command that is required to configure changes to the default BGP MED on a Cisco IOS router are accurate? (Choose two.) (Source: Understanding BGP Multi-Exit Discriminators)

- A) The MED is a mandatory attribute.
- B) Using the **default-metric** command in BGP configuration mode will cause one redistributed network to have the specified MED value.
- C) There is no MED attribute that is attached to a route by default.
- D) To set the default metric value (MED) for BGP routes, use the **default-metric** command.

Q23) Which two statements about the Cisco IOS commands that are required to configure changes to the BGP MED attribute with route-map statements are accurate? (Choose two.) (Source: Understanding BGP Multi-Exit Discriminators)

A) The **set metric** command is used within route-map configuration mode to set the MED attribute.
B) The **neighbor** *address* **route-map** *name* **in | out** command applies a route-map to incoming updates from all neighbors.
C) Per-neighbor MED is configured by using a route-map with a match condition.
D) You can use a route-map to set the MED on incoming or outgoing updates.

Q24) Which two of the following statements about the Cisco IOS commands that are required to configure advanced MED features on Cisco routers are accurate? (Choose two.) (Source: Understanding BGP Multi-Exit Discriminators)

A) The **bgp bestpath med confed** command allows routers to compare paths learned from confederation peers.
B) When you enable a deterministic MED comparison, you allow a router to compare MED values after it considers BGP route type (external or internal) and IGP metric to the next-hop address.
C) Use the **bgp always-compare-med** command to force the router to compare the MED even if the paths come from different autonomous systems.
D) Cisco IOS software, on the other hand, regards a missing MED attribute as having a value of 1.

Q25) If you configure inbound soft reconfiguration with a route-map and issue the **show ip bgp** *prefix* command, which value of the MED attribute is displayed? (Source: Understanding BGP Multi-Exit Discriminators)

A) Only the original route (no MED) is displayed.
B) Both the original route and the modified route are displayed.
C) Only the modified route is displayed.
D) The MED attribute is not displayed with the **show ip bgp** *prefix* command.

Q26) Which of the following statements about the Cisco IOS commands that are required to troubleshoot BGP MED configurations on a Cisco router is accurate? (Source: Understanding BGP Multi-Exit Discriminators)

A) To see the original MED, you need to enable hard reconfiguration on the router.
B) The command **show ip bgp neighbor** *address* **received-routes** displays the original updates before any filters or route-maps have filtered or changed them.
C) If hard reconfiguration is enabled, the original updates to the MED attribute are available by using the **show ip bgp** *prefix* command.
D) Issuing the **show ip route** command will display the MED value.

Q27) What are two reasons why it is not feasible to use the MED to influence return path selection when multiple autonomous systems are involved? (Choose two.) (Source: Addressing BGP Communities)

A) The MED attribute is designed to influence outbound path selection only.
B) The AS-path attribute would be used for path selection regardless of any configured MED value.
C) The weight attribute will always be used, given that it is first in the BGP route selection process.
D) The MED cannot be propagated across several autonomous systems.

Q28) Does the community attribute have any influence on BGP path selection? (Source: Addressing BGP Communities)

A) No, communities are simply tags that are applied to BGP routes.
B) No, communities are nontransitive attributes.
C) Yes, BGP paths are selected based on the value in the community tag.
D) Yes, the community attribute is part of the BGP route selection process.

Q29) Match the steps with the actions describing how BGP communities can facilitate proper return path selection. (Source: Addressing BGP Communities)

A) Step 1
B) Step 2
C) Step 3
D) Step 4
E) Step 5
F) Step 6

_____ 1. Define the filters and route selection policy that will achieve the required goals.

_____ 2. Define administrative policy goals that you need to implement.

_____ 3. Apply communities on incoming updates from neighboring autonomous systems or tell the neighbors to set the communities themselves.

_____ 4. Assign a community value to each goal.

_____ 5. Match communities with route-maps and route filters, change BGP attributes, or influence the route selection process based on the communities that are attached to the BGP routes.

_____ 6. Enable community distribution throughout your AS to allow community propagation.

Q30) How many community tags can be attached to a single BGP route? (Source: Addressing BGP Communities)

A) 1
B) 32
C) 255
D) depends on the number that is configured with the **ip bgp community** command

Q31) Which three of the following are activities that are required to successfully deploy BGP communities in a BGP-based network? (Choose three.) (Source: Addressing BGP Communities)

A) setting communities, which requires a route-map
B) creating community-lists to be used within route-maps to match on community values
C) enabling community propagation per neighbor for designated internal neighbors
D) creating route-maps where community-lists are used to match on community values

Q32) Which two of the following statements about the Cisco IOS commands that are required to configure route tagging with BGP communities are accurate? (Choose two.) (Source: Addressing BGP Communities)

A) You can attach up to 35 communities to a single route with one route-map set statement.
B) Omitting the **additive** keyword from the **set community** command results in overwriting any original community attributes.
C) The global command **ip bgp-community new-format** is recommended on all routers whenever communities contain the AS number.
D) You cannot use a route-map with redistribution from another routing protocol.

Q33) Which of the following statements about the Cisco IOS command that is required to enable BGP community propagation to BGP neighbors is accurate? (Source: Addressing BGP Communities)

A) Community propagation to BGP neighbors is automatically configured.
B) It is necessary to manually strip communities in outgoing BGP updates.
C) The **neighbor** *ip-address* **send-community** command cannot be applied to a peer group.
D) The **neighbor** *ip-address* **send-community** command is needed to propagate community attributes to BGP neighbors.

Q34) Which match criteria are specified in a standard BGP community-list? (Source: Addressing BGP Communities)

A) destination IP addresses
B) regular expressions
C) community attribute values
D) AS numbers

Q35) What is the result of tagging a route with the no-export community? (Source: Addressing BGP Communities)

A) The route will not be advertised within the local AS.
B) The upstream AS will not be allowed to export the route.
C) The route cannot be exported to another routing protocol.
D) The router will not propagate the route to any external neighbors except to intra-confederation external neighbors.

Q36) Match the functions to the Cisco IOS commands that monitor BGP communities. (Source: Addressing BGP Communities)

A) **show ip bgp community**
B) **show ip bgp community as:nn [as:nn …]**
C) **show ip bgp community as:nn [as:nn …] exact**
D) **show ip bgp community-list** *clist*

_____ 1. displays all routes in a BGP table that have all the specified communities attached

_____ 2. displays all routes in a BGP table that have at least one community attached

_____ 3. displays all routes in BGP table that match community-list *clist*

_____ 4. displays all routes in BGP table that have exactly the specified communities attached

Q37) Which two of the following statements about the function of the BGP Link Bandwidth feature are accurate? (Choose two.) (Source: Addressing BGP Communities)

A) The BGP Link Bandwidth feature is used to enable multipath load balancing for external links with unequal bandwidth capacity.
B) When the BGP Link Bandwidth feature is enabled, routes learned from directly connected external neighbor are propagated through the IBGP network with the bandwidth of the source external link.
C) The configuration of the BGP Link Bandwidth feature can be used only with IBGP multipath features to enable unequal-cost load balancing over multiple links.
D) To configure BGP to advertise the bandwidth of links that are used to exit an AS use the **bgp dmzlink-bw** command.

Q38) Which three of the following functions of BGP named community-lists are accurate? (Choose three.) (Source: Addressing BGP Communities)

A) allows the network operator to assign meaningful names to community-lists
B) sets limits on the number of named community-list that can be configured
C) cannot be configured with regular expressions and with numbered community-lists
D) increases the number of community-lists that can be configured by a network operator
E) applies policies to large numbers of routes by using match and set clauses in the configuration of route maps
F) sets limit of 200 community groups that can be configured within each type of list

Q39) Which two of the following statements about the BGP Cost Community feature are accurate? (Choose two.) (Source: Addressing BGP Communities)

A) The path with the highest cost community number is preferred.
B) The cost community attribute influences the BGP best path selection process at the POI.
C) The cost community can be used as a "tie breaker" during the best-path selection process.
D) If the cost community values are equal, then cost community comparison proceeds to the next highest community ID for this POI.

Q40) Which three of the following are functions of the BGP Support for Sequenced Entries in Extended Community Lists? (Choose three.) (Source: Addressing BGP Communities)

A) allows automatic sequencing of individual entries in BGP extended community-lists
B) provides the ability to remove or resequence extended community-list entries without deleting the entire existing extended community-list
C) configures an extended community-list to use custom values
D) removes or resequences extended community-list entries while deleting the entire existing extended community-list
E) is activated by the **ip extcommunity-list** command in global configuration mode
F) configures sequence numbers for standard community-list entries

Module Self-Check Answer Key

Q1)	B
Q2)	A
Q3)	D
Q4)	C
Q5)	B
Q6)	B
Q7)	B, D
Q8)	B
Q9)	B, D
Q10)	B
Q11)	A
Q12)	D
Q13)	C
Q14)	B
Q15)	C
Q16)	B
Q17)	A
Q18)	B
Q19)	A, B
Q20)	D
Q21)	A, B, C
Q22)	C, D
Q23)	A, D
Q24)	A, C
Q25)	B
Q26)	B
Q27)	B, D
Q28)	A
Q29)	1-B, 2-A 3-D, 4-C 5-E 6-F
Q30)	B
Q31)	A, B, D
Q32)	B, C
Q33)	D
Q34)	C
Q35)	D
Q36)	1-B 2-A 3-D 4-C
Q37)	A, B
Q38)	A, D, E
Q39)	B, C
Q40)	A, B, E

Module 5

Customer-to-Provider Connectivity with BGP

Overview

Today, companies use the Internet for a variety of reasons, including increasing employee productivity, increasing sales, increasing customer satisfaction, and reducing cycle time. A key component in connecting companies to the Internet is the service provider. Depending on business goals, application requirements, and administrative policies, a company will use different methods to connect to a service provider—or even multiple service providers.

This module discusses the different requirements for connectivity between customers and service providers. Included in this module is a discussion of physical connection methods, redundancy, load balancing, and technical requirements such as addressing and autonomous system (AS) numbering. In addition, this module details the configuration requirements to connect a customer to a single service provider by using static routes and the Border Gateway Protocol (BGP). Also provided in this module are the configuration requirements to connect a customer to multiple service providers by using BGP.

Module Objectives

Upon completing this module, you will be able to configure the service provider network to behave as a transit AS in a typical implementation with multiple BGP connections to other autonomous systems. This ability includes being able to meet these objectives:

- Describe the requirements to connect customer networks to the Internet in a service provider environment
- Implement customer connectivity by using static routing in a service provider network
- Implement customer connectivity using BGP in a customer implementation in which you must support multiple connections to a single ISP
- Implement customer connectivity using BGP in a customer scenario in which you must support connections to multiple ISPs

Lesson 1

Understanding Customer-to-Provider Connectivity Requirements

Overview

Customers connect to the Internet by using service providers to enable applications such as intranet connectivity with Virtual Private Networks (VPNs), extranet connectivity with suppliers, and other Internet applications. When planning network connectivity to an Internet service provider (ISP), network designers must give careful consideration to the various aspects of the connectivity, including physical connection types, the redundancy provided by the connection method that is chosen, IP addressing requirements, and autonomous system (AS) numbering considerations, if the network design is going to meet both the business and technical requirements of the applications that are planned for the network.

This lesson discusses solutions for connecting customer networks to service providers. Also included in this lesson is a discussion of customer network redundancy requirements, routing requirements, IP addressing requirements, and autonomous system (AS) numbering requirements.

Objectives

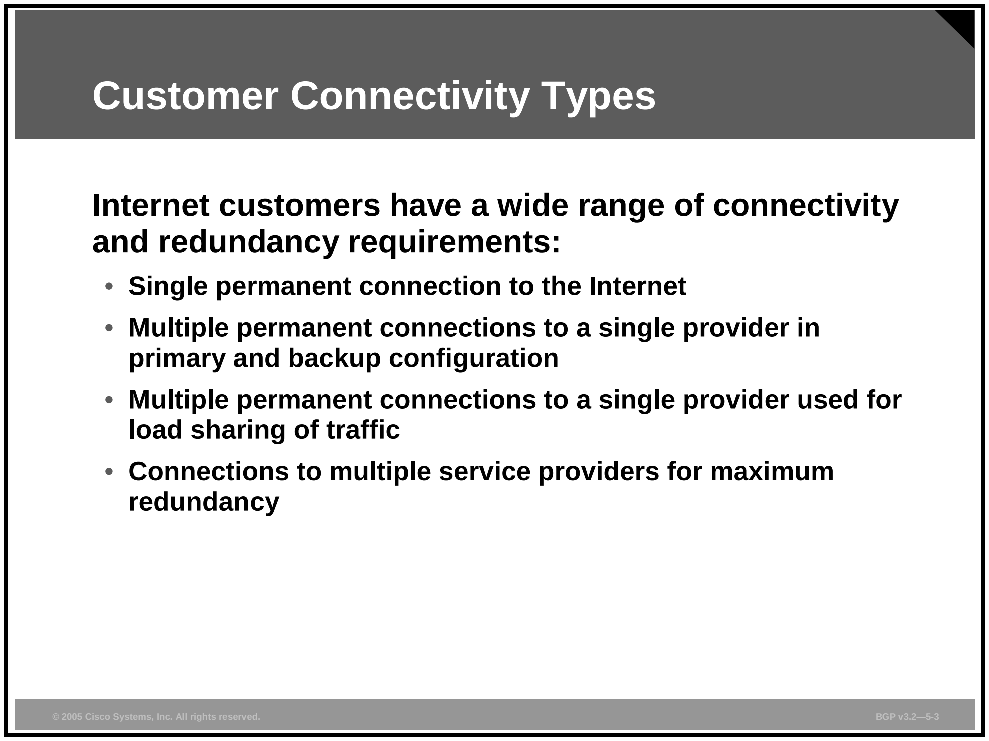
Upon completing this lesson, you will be able to describe the requirements to connect customer networks to the Internet in a service provider environment. This ability includes being able to meet these objectives:

- Identify various physical connections that are used by customers to connect to a service provider
- Describe the levels of redundancy that are provided by each physical connection type that is used by customers to connect to a service provider
- Identify various routing schemes that are used by customers to connect to a service provider

- Describe routing schemes that are appropriate for each physical connection type that is used by customers to connect to a service provider
- Describe the addressing schemes that are used by customers to connect to a service provider
- Describe the AS numbering schemes that are used by customers to connect to a service provider

Customer Connectivity Types

This topic identifies the various physical connections that are used by customers to connect to a service provider.



Customer Connectivity Types

Internet customers have a wide range of connectivity and redundancy requirements:

- **Single permanent connection to the Internet**
- **Multiple permanent connections to a single provider in primary and backup configuration**
- **Multiple permanent connections to a single provider used for load sharing of traffic**
- **Connections to multiple service providers for maximum redundancy**

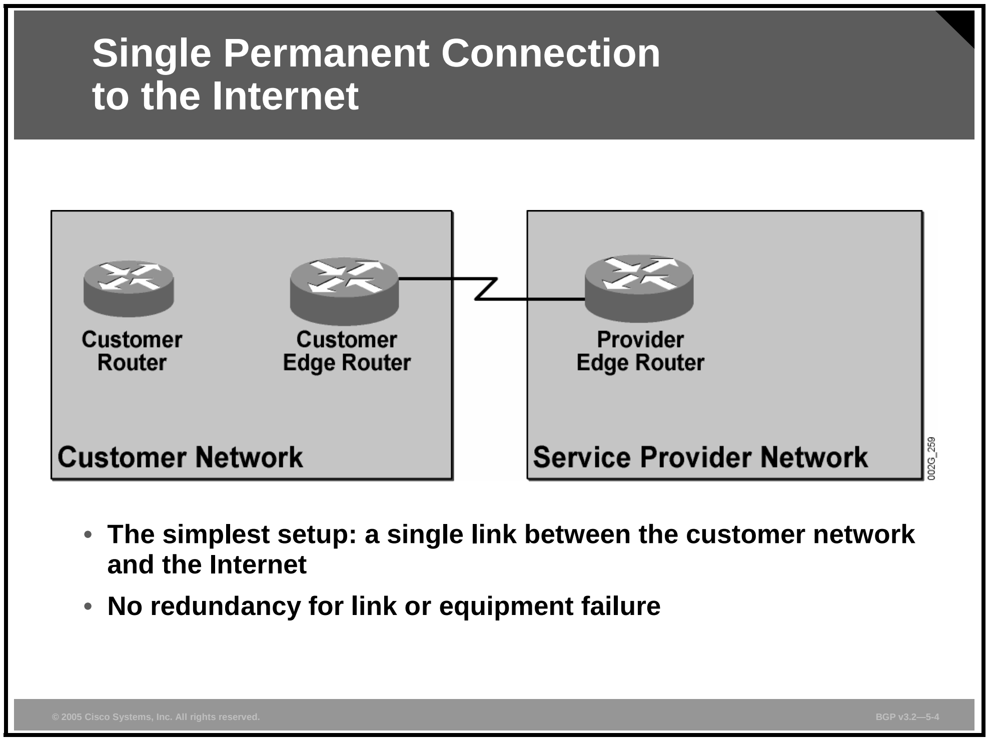
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-3

Service provider customers have different requirements for their Internet connectivity. These different requirements result in different solutions:

- A single permanent connection to one ISP. This solution meets the requirements for the vast majority of customers.
- Multiple permanent connections in which one of the lines is primary and the other line is used for backup only. This setup also provides redundancy on the links. Compared to a dial-up backup, a permanent backup link is preferred for various reasons, such as the severe bandwidth limitations on dial-up lines and the time that is required to establish a dial-up connection.
- Multiple permanent connections to one ISP, which is used for load sharing of traffic. This solution gives redundancy on the links but also provides additional bandwidth.
- Permanent connections to more than one ISP. This solution provides the highest level of redundancy, because it not only can cope with link-level failures but also with failures within the network of a service provider.

Redundancy in Customer Connections

This topic describes the levels of redundancy that are provided by each physical connection type that is used by customers to connect to a service provider.



Single Permanent Connection to the Internet

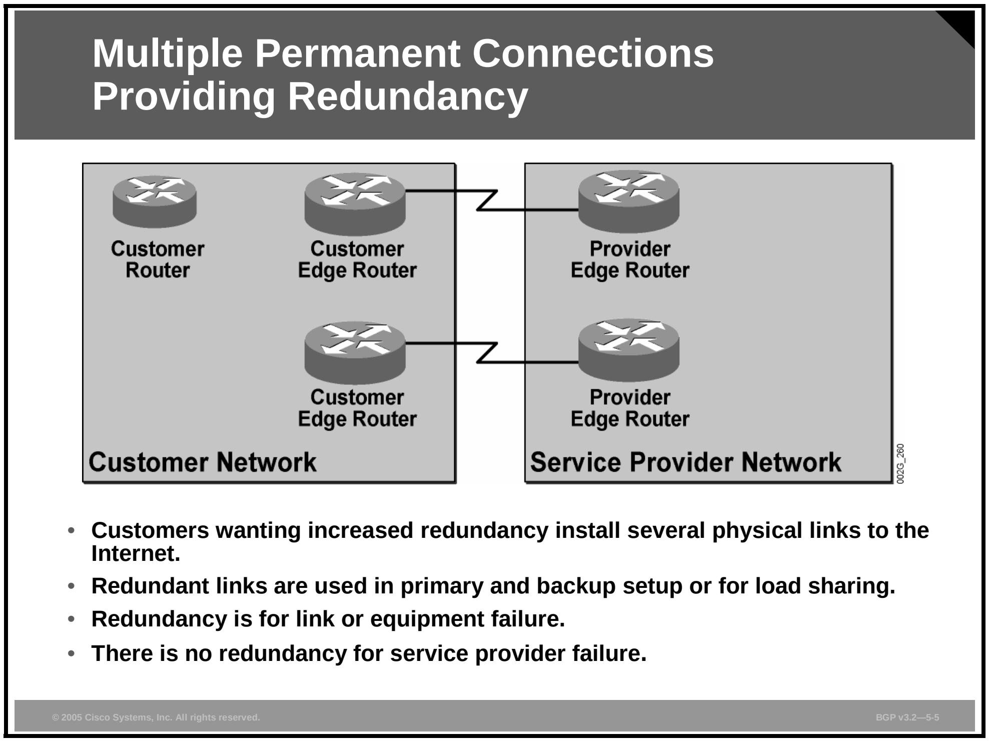
- **The simplest setup: a single link between the customer network and the Internet**
- **No redundancy for link or equipment failure**

BGP v3.2—5-4

A single permanent connection to one ISP is the most common setup. This setup is also the simplest to implement.

The customer network has an edge router. This router is connected to one of the edge routers of the ISP. The connection is permanent and could be a leased line, a Frame Relay or ATM permanent virtual circuit (PVC), a LAN segment, or something equivalent.

There is no redundancy in this solution. Any failure on the permanent link or either of the two edge routers causes a complete outage of the service. Serious failures within the ISP network that affect all customers of this ISP also affect the customer in this example.



Multiple Permanent Connections Providing Redundancy

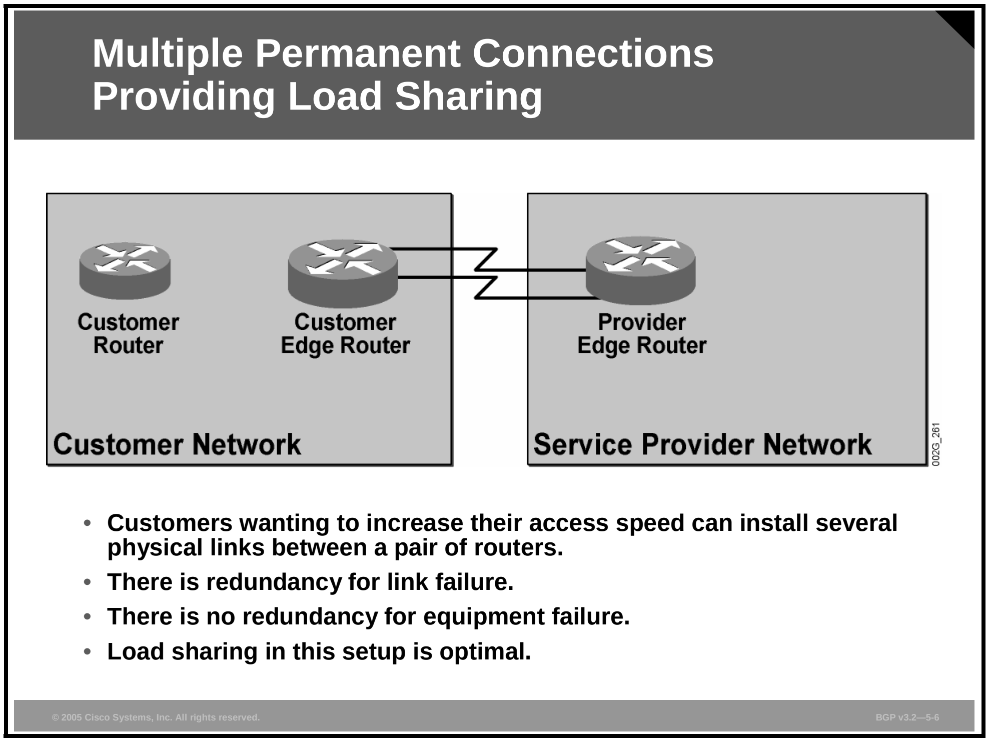
- Customers wanting increased redundancy install several physical links to the Internet.
- Redundant links are used in primary and backup setup or for load sharing.
- Redundancy is for link or equipment failure.
- There is no redundancy for service provider failure.

In this setup, one customer edge router connects to one ISP edge router. A different customer edge router is used to connect to another ISP edge router. If one of these routers fails, only one of the connections breaks down; the other connection is still available.

In some cases, the two links may be implemented between the customer and the provider for load sharing and in other cases strictly for backup purposes. Backup PVCs in Frame Relay or ATM networks can sometimes be very cost-efficient, provided that these PVCs carry only a very small volume of traffic and that the primary path is available.

When load sharing between both links is a desired network characteristic, the distribution of the load over the links is more complicated than when both links terminate in the same router.

Again, because the customer is connected to a single ISP, serious ISP network failures that affect all customers of this ISP will also affect the customer in this scenario, regardless of the backup link.



Multiple Permanent Connections Providing Load Sharing

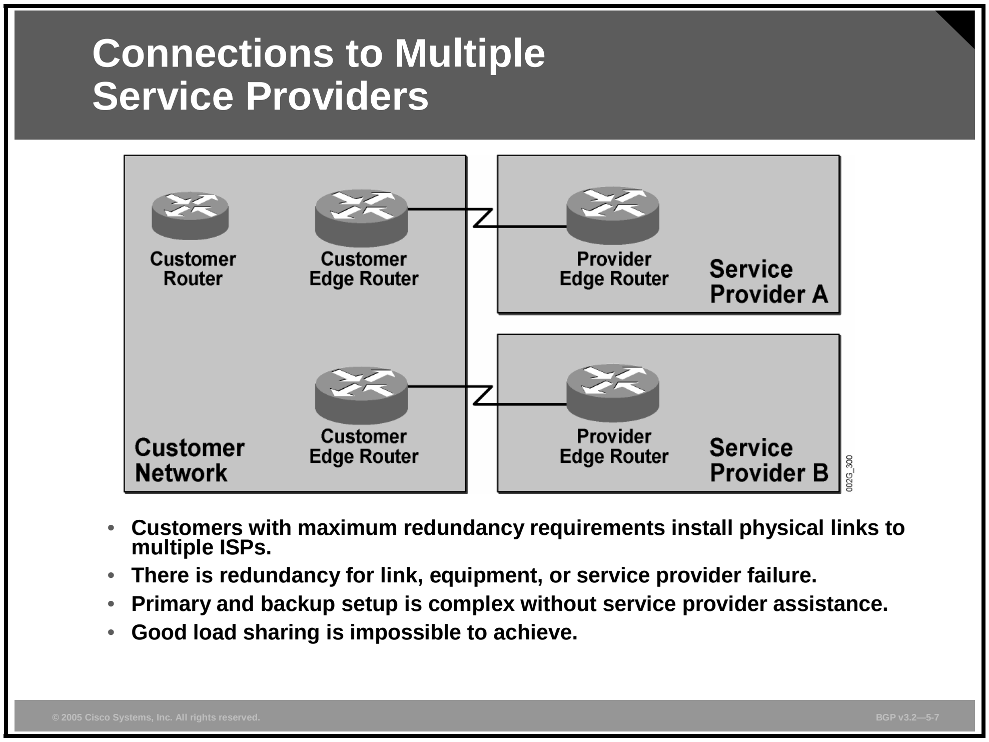
- **Customers wanting to increase their access speed can install several physical links between a pair of routers.**
- **There is redundancy for link failure.**
- **There is no redundancy for equipment failure.**
- **Load sharing in this setup is optimal.**

BGP v3.2—5-6

In the example, a single router in the customer network is connected to a single router in the ISP network. The redundancy is limited to the link level because router failures are not covered. Using two parallel links between two routers allows for an optimal distribution of load over the links.

Depending on the switching path that is used in the customer and the ISP routers, load sharing can be performed based on the destination address only (fast switching), on source-destination address pairs (default behavior for Cisco Express Forwarding [CEF]), or on a packet-by-packet basis (process switching or CEF).

As in the previous examples, serious ISP network failures that affect all customers of this ISP will also affect this customer, regardless of the link backup.



Connections to Multiple Service Providers

- Customers with maximum redundancy requirements install physical links to multiple ISPs.
- There is redundancy for link, equipment, or service provider failure.
- Primary and backup setup is complex without service provider assistance.
- Good load sharing is impossible to achieve.

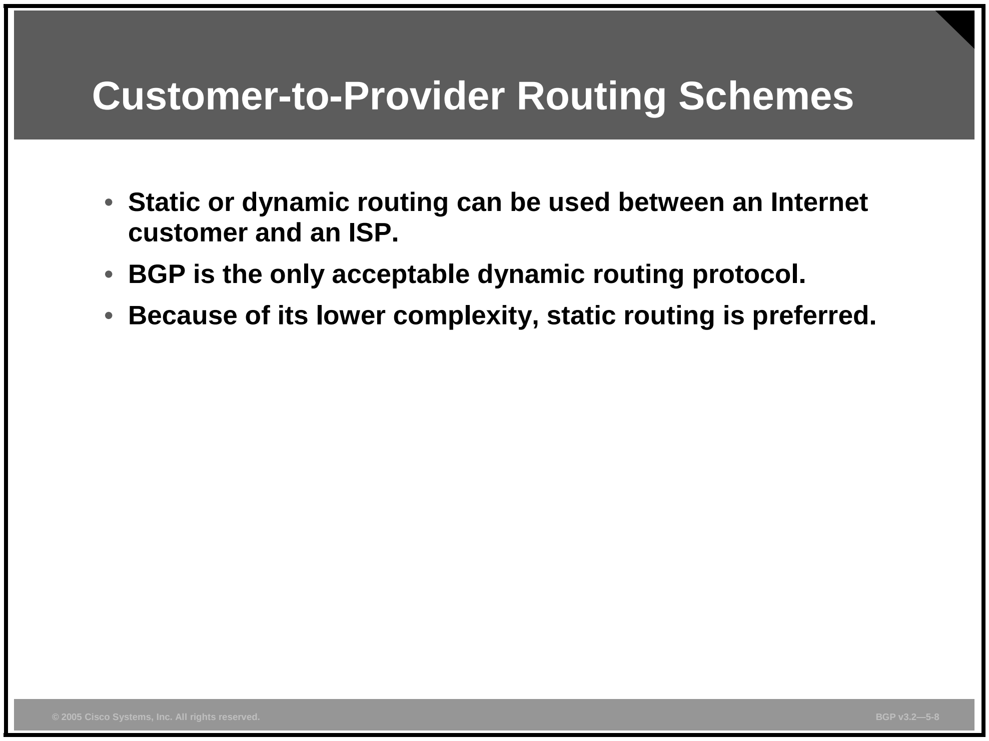
In the example, two edge routers in the customer network have one permanent connection each to different ISPs. Link failures and router failures are covered by the redundancy in exactly the same way as in the previous example in which the two customer routers are connected to two different routers in one ISP network. However, because the two connections in this example go to two different ISPs, the redundancy also covers problems within one ISP network.

The two links may in some cases be implemented by the customer for load sharing and in other cases be used strictly for backup purposes. Controlling load distribution over the links is more complicated in this example. Avoiding any load on the backup link may require assistance from the ISP to which the backup link is connected.

Load sharing between the links in this setup can never be optimal. Equal distribution of the return traffic load from the Internet over the two separate links cannot be achieved. Distribution of the load of outgoing traffic is done based on destination addresses. Slowly adjusting the appropriate router configuration parameters and observing the link traffic load changes that result can enable you to reach an acceptable distribution of router traffic between the two links.

Customer-to-Provider Routing Schemes

This topic identifies various routing schemes that customers use to connect to a service provider.



Customer-to-Provider Routing Schemes

- Static or dynamic routing can be used between an Internet customer and an ISP.
- BGP is the only acceptable dynamic routing protocol.
- Because of its lower complexity, static routing is preferred.

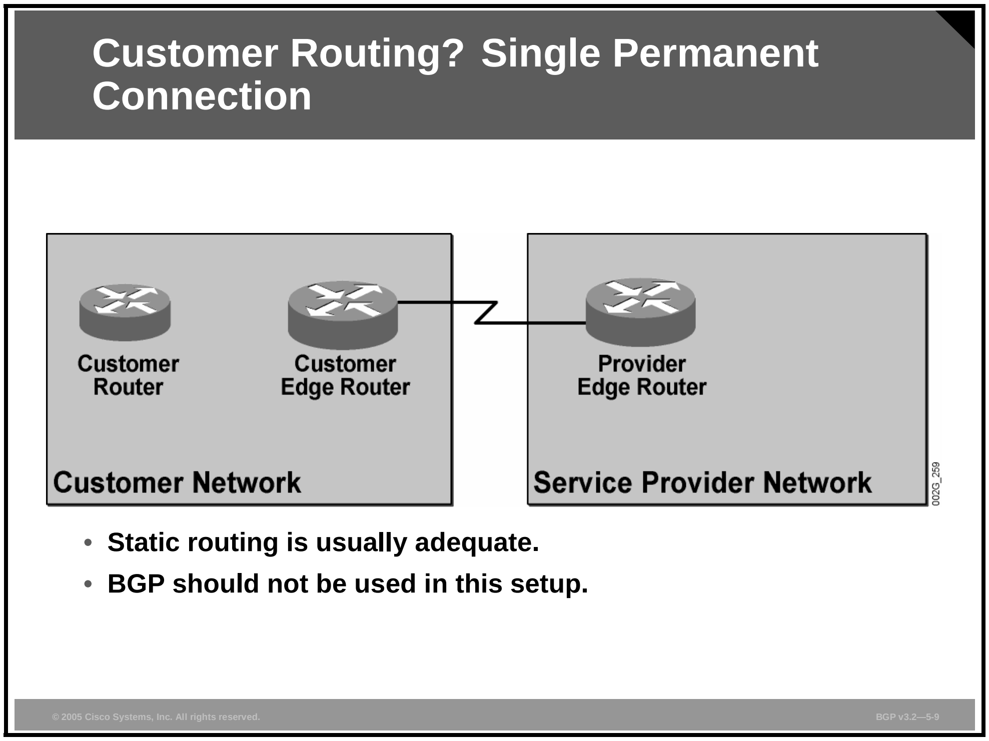
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-8

Different solutions for connecting a customer network to the network of an ISP require different methods of routing information exchange:

- **Static routing:** Static routing is preferred because of its lower complexity. In a normal case, the customer network must have a default route to the ISP network and the ISP network must have a route to the IP prefixes that the customer has in its network. As always, static routing provides very low, if any, redundancy.
- **Dynamic routing:** Dynamic routing provides redundancy. The customer and the ISP networks must be configured to exchange a common routing protocol. BGP is the only choice because of the large volumes of routing information, the inherent security mechanisms of BGP, and the ability of BGP to handle routing policies.

Customer Routing

This topic describes routing schemes that are appropriate for each physical connection type that is used by customers to connect to a service provider.

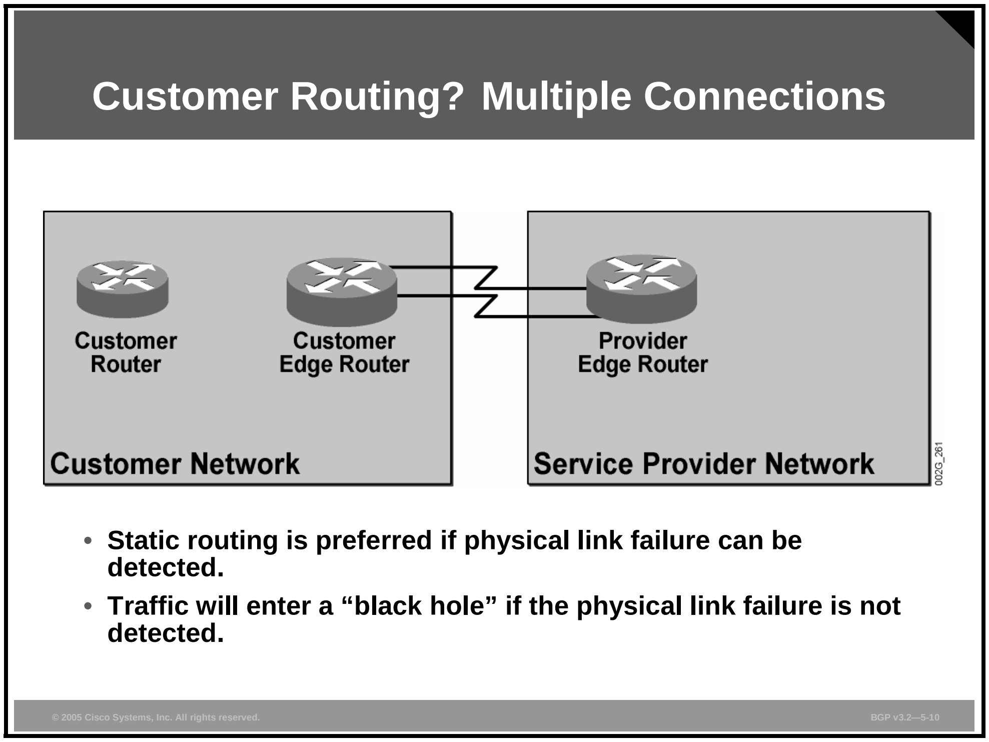


Customer Routing? Single Permanent Connection

- Static routing is usually adequate.
- BGP should not be used in this setup.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-9

When the customer has a single permanent connection to the Internet, static routing is usually adequate. The physical topology does not provide any redundancy, and it is therefore unnecessary to add the complexity of dynamic routing. Keep the network simple by avoiding the use of BGP in this case.



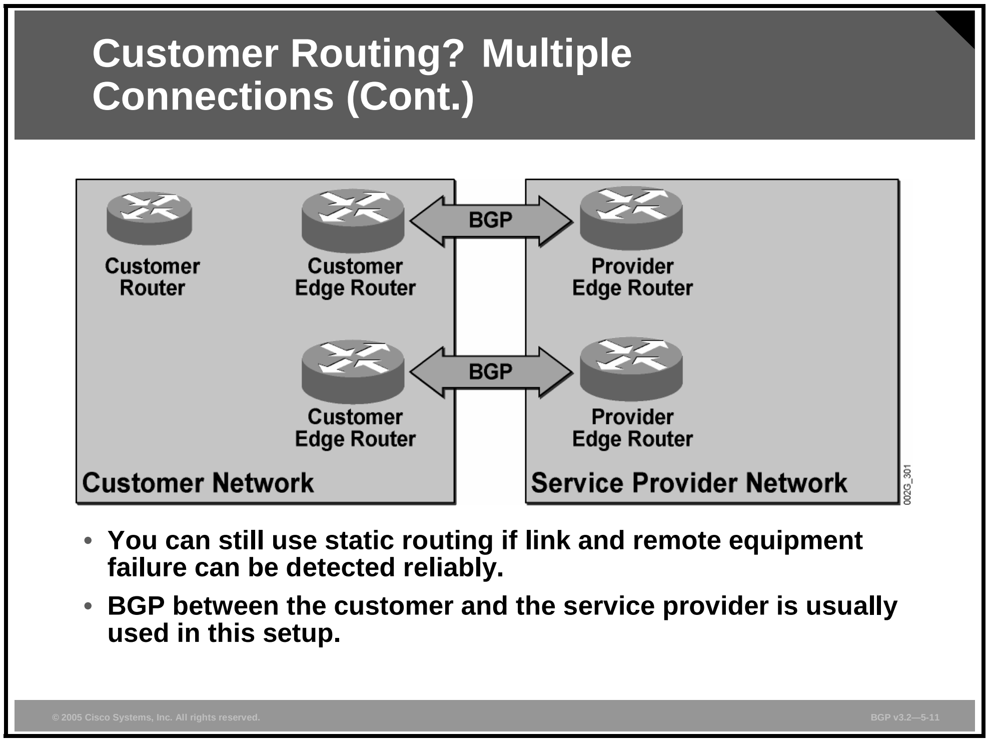
Customer Routing? Multiple Connections

- **Static routing is preferred if physical link failure can be detected.**
- **Traffic will enter a “black hole” if the physical link failure is not detected.**

Multiple permanent connections between a single router on the customer network and a single router on the service provider network should be configured with static routing, provided that link failure can be detected by link-level procedures.

With this type of connection, two static routes are configured on each network, pointing to both links between the customer and the ISP. If either of the links fails, the link-level procedures should detect this failure and place the interface in a down state. In this case, the static route is invalid and is not used for forwarding packets. The router will subsequently forward all packets over the remaining link.

If the link-level procedures cannot detect a link failure, the static route pointing out over the failed link is still valid. The router continues to use this static route to send some of the traffic out on the failed interface. This situation effectively creates a “black hole” for some of the traffic.



Customer Routing? Multiple Connections (Cont.)

- **You can still use static routing if link and remote equipment failure can be detected reliably.**
- **BGP between the customer and the service provider is usually used in this setup.**

You can also use static routing for multiple permanent connections between two different routers on the customer network to two different routers on the service provider network if the failures can be detected by the link-level procedures. When one of the connections is lost, the link-level procedure detects this loss and places the interface in a down state. Because the interface is in the down state, the static route that points out of the down interface becomes invalid. As a result, the router stops the redistribution of the static route into BGP.

However, customers that require the use of multiple connections and multiple routers very often do not rely on the link-level procedures. These customers require a routing protocol such as BGP to detect the failures. Because BGP uses handshaking and reliable transfer, it always detects a failed link or failed remote router.

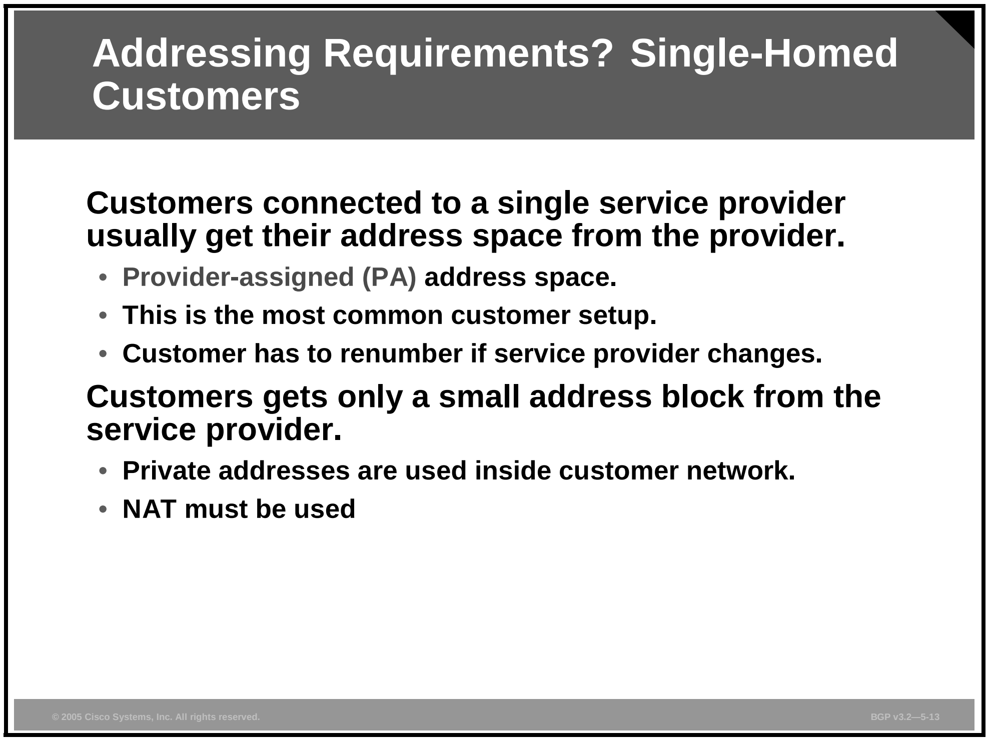
Multiple permanent connections to more than one ISP always require the use of dynamic routing with BGP. The customers that require this type of connection do not just want to protect the network connectivity from link failures or remote router failures, they also want to protect their network connectivity from serious problems in the network of an ISP.

Monitoring the link status cannot detect a problem inside one of the ISP networks. If the link is still up and the ISP edge router is still up, the link-level procedures do not indicate any problems. However, the ISP network may suffer from severe problems. An ISP network can be partitioned or disconnected from the rest of the Internet without having any problems with the edge router and the access line to the customer network.

The only way to detect this situation is to use BGP with both ISPs and receive full Internet routing from both of them. When one of the ISPs has problems, the edge router, being the BGP neighbor of the customer, withdraws the routes that it can no longer reach. This action means that the customer routers know which Internet routes that each ISP can reach at the moment.

Addressing Requirements

This topic describes the various addressing schemes that customers use to connect to a service provider.



Addressing Requirements? Single-Homed Customers

Customers connected to a single service provider usually get their address space from the provider.

- **Provider-assigned (PA) address space.**
- **This is the most common customer setup.**
- **Customer has to renumber if service provider changes.**

Customers gets only a small address block from the service provider.

- **Private addresses are used inside customer network.**
- **NAT must be used**

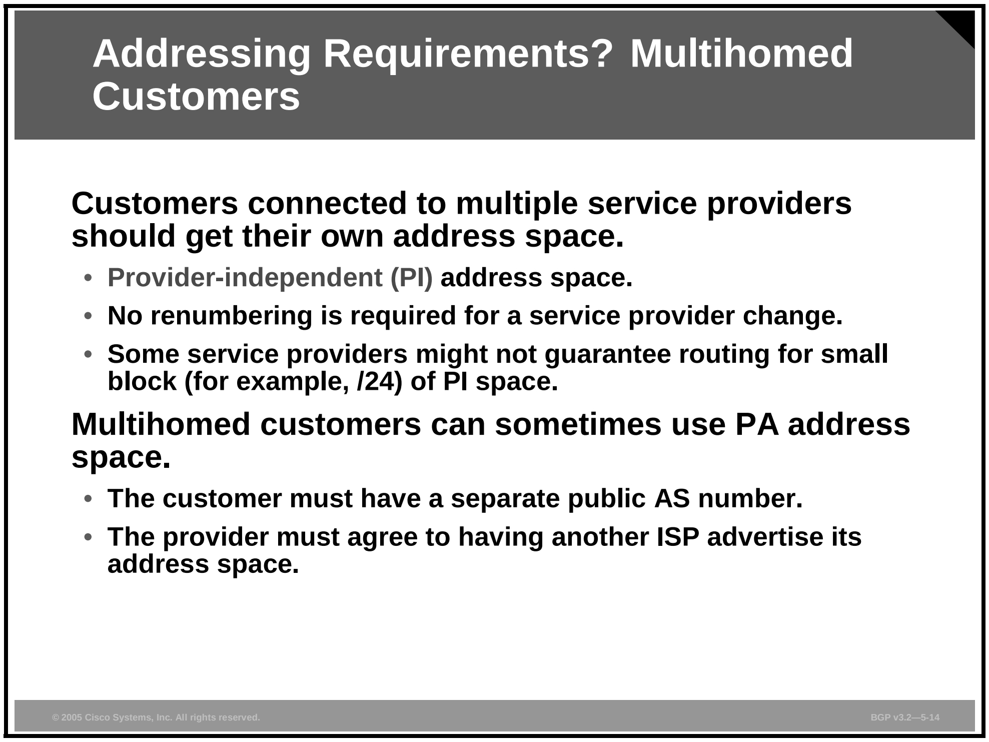
BGP v3.2—5-13

Customers that are connected to a single ISP usually get their address space assigned by the ISP. An ISP is usually assigned a large address space to delegate to its customers. Because all customers of one ISP get their addresses from one address space or a few address spaces, it is very likely that the ISP is able to aggregate the customer addresses before sending the routes to the rest of the Internet.

Most customers are connected to a single ISP, which means that they are using provider-assigned (PA) addresses. If the customer should decide to change its service provider, the customer must return its PA addresses to the old ISP and receive a new assignment of PA addresses from the new ISP. Otherwise, the ISPs are no longer able to perform efficient address aggregation.

The consequence for the customer is that the customer has to renumber its network when it changes its service provider.

Some customers decide to use private addresses within their network and do Network Address Translation (NAT) at the connection point to the ISP. This setup means that customers require only a very small portion of public addresses from the ISP. In addition to conserving address space for the benefit of the Internet as a whole, this setup also means that when the customer decides to change its service provider, addresses are renumbered only at the NAT point. The rest of the customer network does not need to be renumbered.



Addressing Requirements? Multihomed Customers

Customers connected to multiple service providers should get their own address space.

- **Provider-independent (PI) address space.**
- **No renumbering is required for a service provider change.**
- **Some service providers might not guarantee routing for small block (for example, /24) of PI space.**

Multihomed customers can sometimes use PA address space.

- **The customer must have a separate public AS number.**
- **The provider must agree to having another ISP advertise its address space.**

BGP v3.2—5-14

Customers that are connected to more than one ISP should, if possible, assign their own address space and not have addresses that are delegated from any of their ISPs. Such assigned addresses are called provider-independent (PI) addresses.

A customer using PI addresses can change its service provider without renumbering its network. The address space is not in any way bound to a particular provider. This arrangement means that no ISP can aggregate the customer routes before sending them to the rest of the Internet. The routes propagate through the Internet with the prefix lengths given.

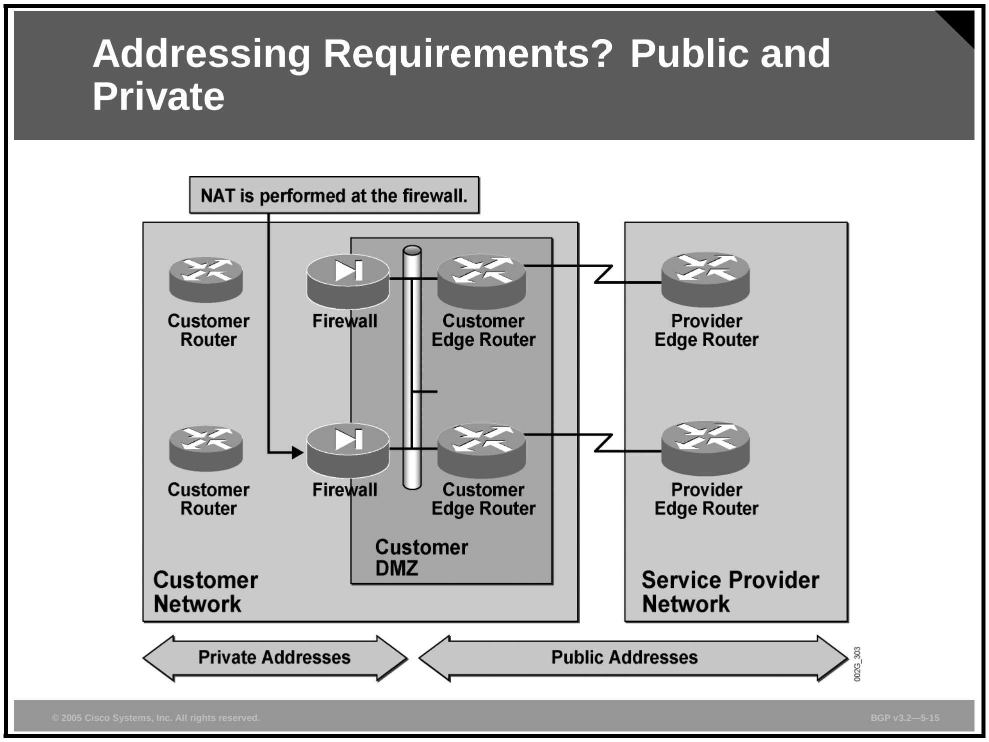
Some large ISPs filter out routes with long prefixes. ISPs do not want to populate their routing tables with a large number of explicit routes that should have been aggregated into a route summary before they were sent to them. As a result, the customer announcing small blocks of PI addresses, which cannot be aggregated, may not be reachable from all parts of the Internet. A larger block of PI addresses solves the problem.

A multihomed customer can in some cases use PA addresses. The address space must be assigned from one of the ISPs. When the customer announces the block of PA addresses to both ISPs, both should propagate the addresses to the rest of the Internet. The provider that assigned the address space should also announce the larger block of addresses, of which the customer is announcing a subset.

Other ISPs now receive two alternate explicit routes and an overlapping route summary. Filtering out explicit routes is more likely at this time because the other ISPs recognize these as routes that can be aggregated. If the other ISPs filter out the more explicit routes, the customer is still reachable as long as both providers are announcing the overlapping route summary.

Example: Addressing Requirements

In this example, the customer uses private addresses inside its own network.



Addressing Requirements? Public and Private

BGP v3.2—5-15

Only a very small network segment, called the customer demilitarized zone (DMZ), has been assigned public addresses.

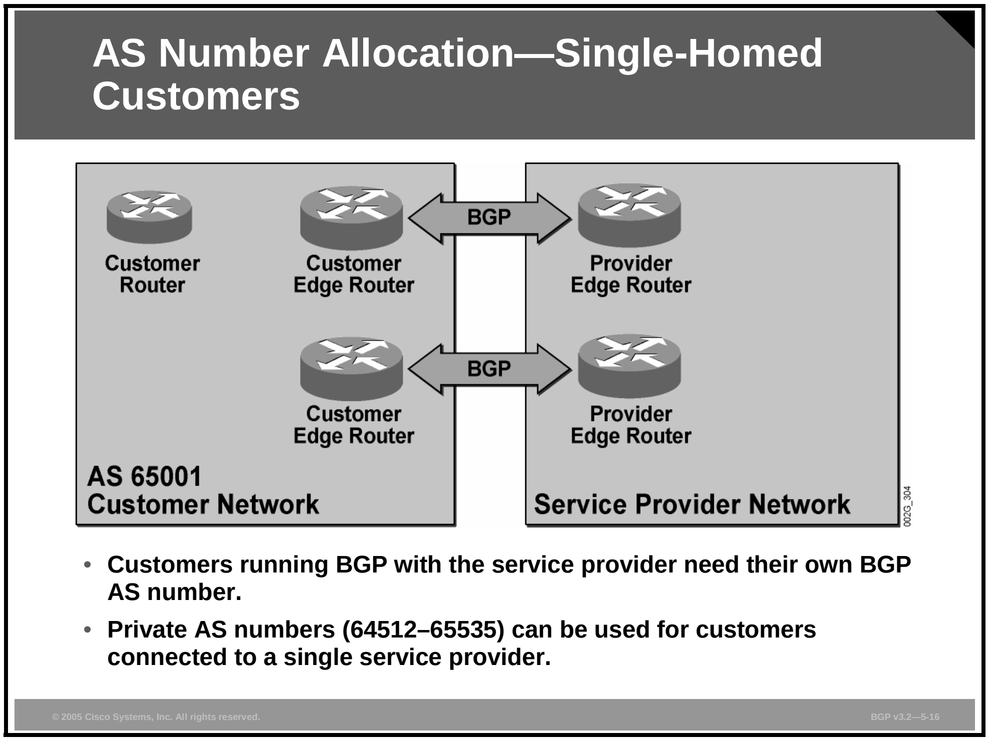
The customer network is connected to the customer DMZ using two alternate firewalls with both firewalls doing NAT. All packets leaving the customer network have their addresses translated to a public address belonging to the DMZ subnet. The reverse translation is made in the reverse traffic direction.

In this case, the customer requires only a very small block of public addresses. These addresses can be PA addresses. If the customer decides to change its service provider, renumbering is not a problem because only a few devices need to be reconfigured by the customer.

Care must be taken so that traffic flows symmetrically through the firewalls. Otherwise, NAT does not work. The easiest way to achieve this symmetry is to allow only one firewall be active at a time.

AS Number Allocation

This topic describes various AS numbering schemes that customers use to connect to a service provider.



AS Number Allocation—Single-Homed Customers

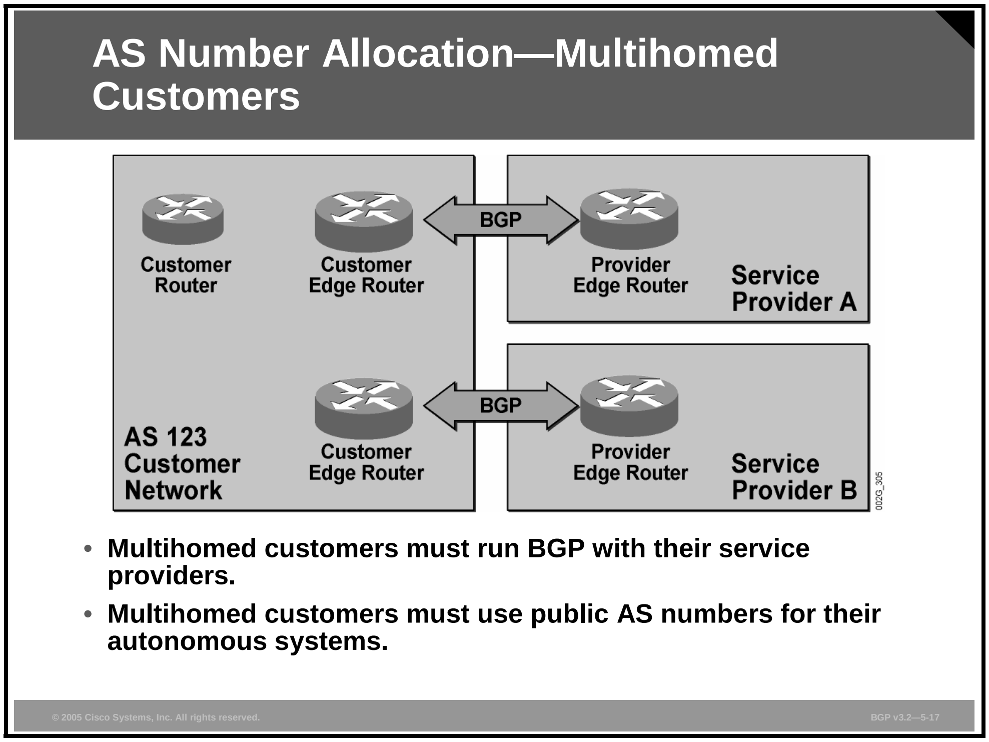
- **Customers running BGP with the service provider need their own BGP AS number.**
- **Private AS numbers (64512–65535) can be used for customers connected to a single service provider.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-16

BGP requires the use of AS numbers. When BGP is configured, the AS number is mandatory information. Public AS numbers are a scarce resource, however. Customers should use public AS numbers only when they are required. A customer that uses BGP to exchange routing information with only one ISP does not require a public AS number. This customer can use a private AS number.

An ISP network that is running BGP with some of its customers must determine whether a public or a private AS number is required for each customer. When the customer can use a private AS number, the ISP must allocate one from the range of private AS numbers (64512 to 65535). The ISP must make sure not to assign any of the private AS numbers to more than one customer.

When the ISP receives BGP routes from the customer, the ISP routers see the private AS number in the AS path and treat the private number as any other AS number. However, before the ISP propagates any of these routes to the rest of the Internet, it must remove the private AS numbers from the AS path, because the same AS number may be in use by someone else. After the private AS number is removed, the route appears as belonging to the public AS of the ISP.



AS Number Allocation—Multihomed Customers

- **Multihomed customers must run BGP with their service providers.**
- **Multihomed customers must use public AS numbers for their autonomous systems.**

BGP v3.2—5-17

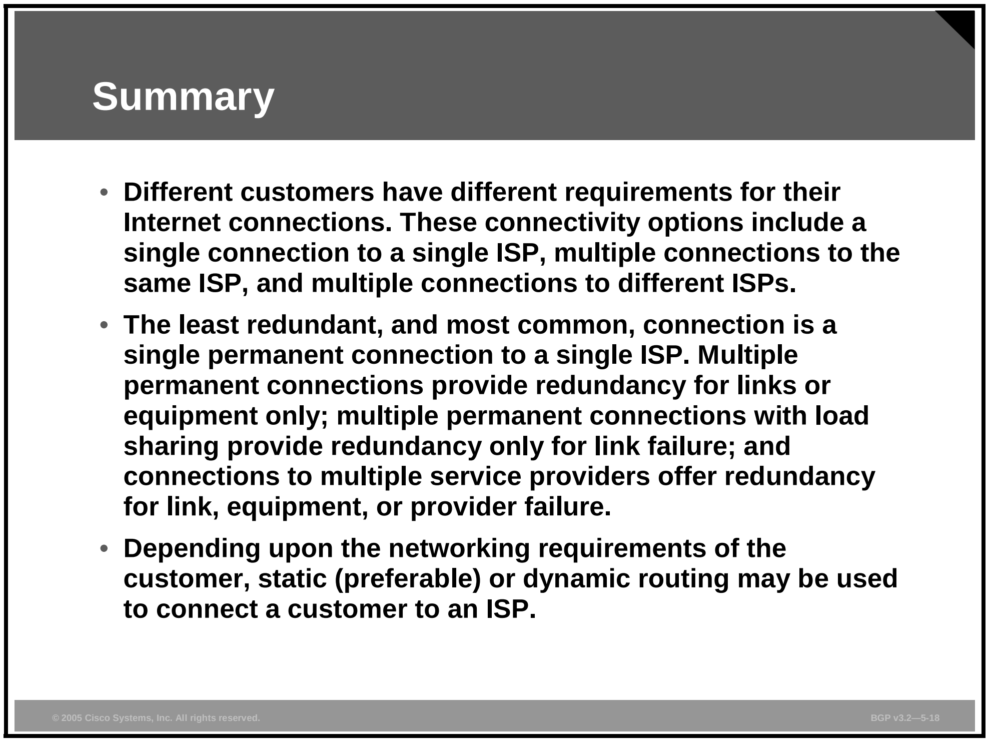
A multihomed customer requires a public AS number and must run BGP with both of its ISPs. The customer should not use a private AS number because both ISPs must propagate the customer routes to the rest of the Internet. If the customer does use a private AS number, and both ISPs remove the number before sending it to the rest of the Internet, then the customer routes will appear to be local in the public AS of both ISPs. To make BGP work correctly, multihomed customers need to avoid this situation.

Note With the help of the AS number translation feature, private AS numbers can also be used for multihomed customers, but this type of configuration is not encouraged.

Multihomed customers are correctly connected to the Internet by assigning a public AS number to the customer network. This public AS number appears in the AS path and should be propagated by the service provider to the rest of the Internet. The customer network is now reachable by the rest of the Internet through both providers. The route with the shortest AS path is used by Internet endpoints as the best route to the customer network.

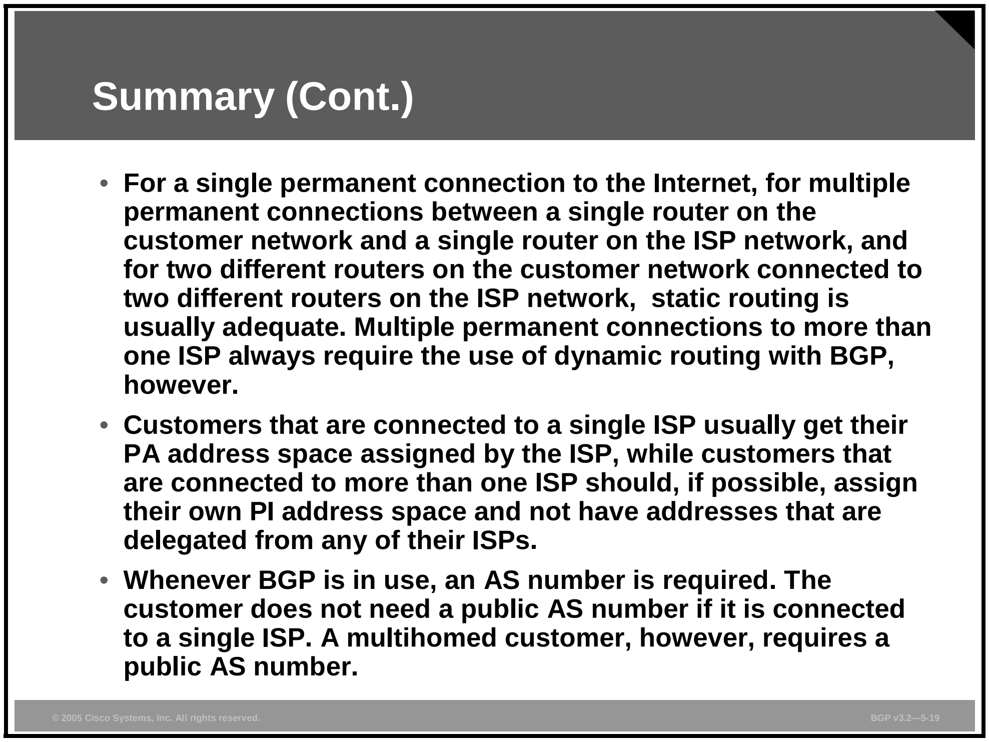
Summary

This topic summarizes the key points discussed in this lesson.



Summary

- Different customers have different requirements for their Internet connections. These connectivity options include a single connection to a single ISP, multiple connections to the same ISP, and multiple connections to different ISPs.
- The least redundant, and most common, connection is a single permanent connection to a single ISP. Multiple permanent connections provide redundancy for links or equipment only; multiple permanent connections with load sharing provide redundancy only for link failure; and connections to multiple service providers offer redundancy for link, equipment, or provider failure.
- Depending upon the networking requirements of the customer, static (preferable) or dynamic routing may be used to connect a customer to an ISP.



Summary (Cont.)

- For a single permanent connection to the Internet, for multiple permanent connections between a single router on the customer network and a single router on the ISP network, and for two different routers on the customer network connected to two different routers on the ISP network, static routing is usually adequate. Multiple permanent connections to more than one ISP always require the use of dynamic routing with BGP, however.
- Customers that are connected to a single ISP usually get their PA address space assigned by the ISP, while customers that are connected to more than one ISP should, if possible, assign their own PI address space and not have addresses that are delegated from any of their ISPs.
- Whenever BGP is in use, an AS number is required. The customer does not need a public AS number if it is connected to a single ISP. A multihomed customer, however, requires a public AS number.

Lesson 2

Implementing Customer Connectivity Using Static Routing

Overview

When a customer can connect to the Internet through either a single connection to a service provider or multiple connections to the same Internet service provider (ISP), static routing is the simplest routing approach to implement between customer and provider. When network administrators are implementing customer-to-provider connectivity with static routes, knowledge of static routing implementation guidelines will aid in successfully deploying static routing network configurations.

This lesson discusses static routing as a solution for connecting customer networks to service providers. Also included in this lesson is a discussion of when static routing should be used between a customer and a provider and information on how to configure static routing in nonredundant, backup, and load-sharing configurations.

Objectives

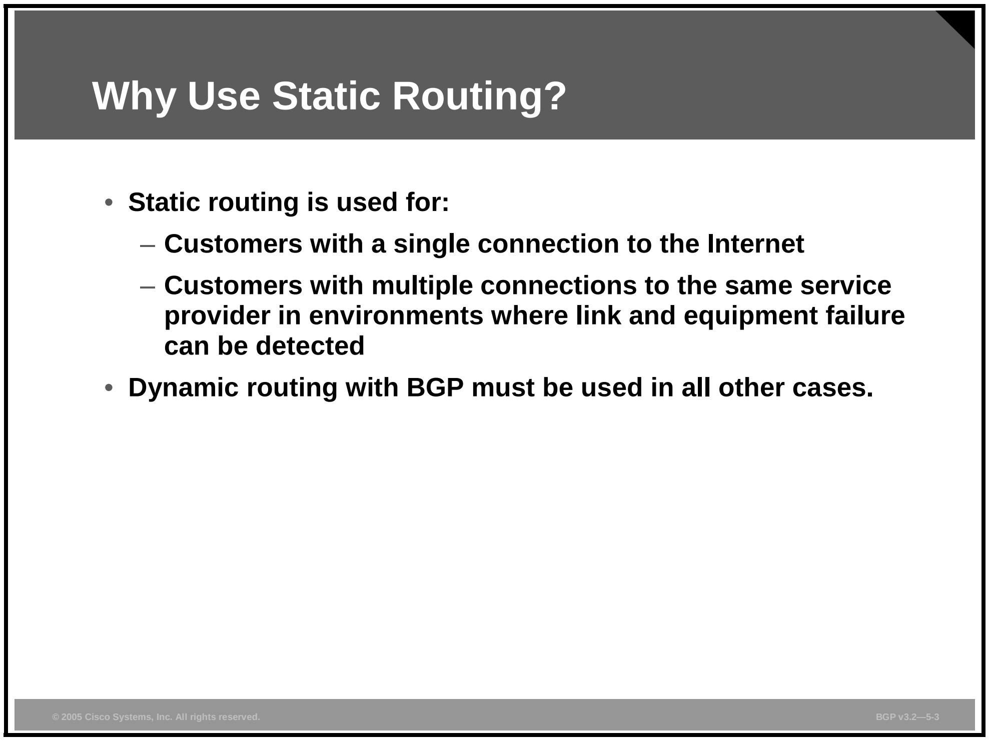
Upon completing this lesson, you will be able to implement customer connectivity by using static routing in a service provider network. This ability includes being able to meet these objectives:

- Identify when to use static routing between a customer and a service provider in a BGP environment
- Describe the characteristics of static routing between a customer and a service provider in a BGP environment
- Identify design considerations for propagating static routes in a service provider network
- Configure static route propagation in a BGP environment with different service levels
- Configure a typical backup setup that uses static routing between a customer and a service provider in a BGP environment

- Describe the limitations of floating static routes when they are used in typical backup static routing scenarios and the corrective actions to overcome these limitations
- Describe the characteristics of load sharing when you are configuring static routing between a customer and a service provider

Why Use Static Routing?

This topic identifies when to use static routing between a customer and a service provider in a Border Gateway Protocol (BGP) environment.



Why Use Static Routing?

- **Static routing is used for:**
  - **Customers with a single connection to the Internet**
  - **Customers with multiple connections to the same service provider in environments where link and equipment failure can be detected**
- **Dynamic routing with BGP must be used in all other cases.**

Static routing is the best solution to implement when there is no redundancy in the network topology. A single connection between the customer network and the service provider network does not provide any redundancy. If the link goes down, the connection is lost regardless of which routing protocol is configured in the customer or provider network. When there are redundant connections between the customer network and the network of a single service provider, static routing can be used under specific circumstances.

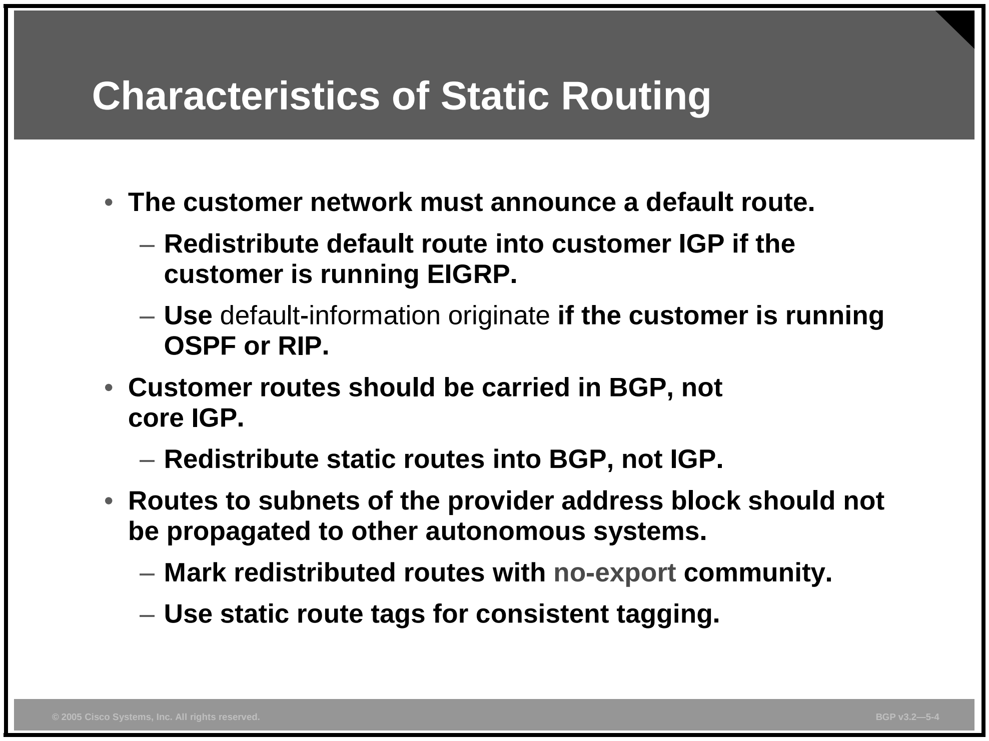
A static default route must be conditionally announced by the customer edge routers that are using an Interior Gateway Protocol (IGP). If the link to one of the customer edge routers goes down, then the router must be able to detect the failure and invalidate the static default route. Announcement of this router as a default gateway that is using an IGP must now cease. Likewise, on the service provider edge routers, the static routes that are pointing to the customer networks must be invalidated if the link between them goes down, and redistribution to BGP is therefore stopped.

If link-level procedures cannot detect a link failure, the interface remains in the up state. The static routes are not invalidated, and packets are forwarded into a “black hole.” In such cases, because the router cannot detect a failure at the link level, BGP must be used between the customer and the provider.

BGP must also be used between the customer and the service provider networks when the customer is multihomed. This is the case regardless of which link failure detection mechanisms are in use.

Characteristics of Static Routing

This topic describes the characteristics of static routing between a customer and a service provider in a BGP environment.



Characteristics of Static Routing

- **The customer network must announce a default route.**
  - **Redistribute default route into customer IGP if the customer is running EIGRP.**
  - **Use** default-information originate **if the customer is running OSPF or RIP.**
- **Customer routes should be carried in BGP, not core IGP.**
  - **Redistribute static routes into BGP, not IGP.**
- **Routes to subnets of the provider address block should not be propagated to other autonomous systems.**
  - **Mark redistributed routes with no-export community.**
  - **Use static route tags for consistent tagging.**

BGP v3.2—5-4

When static routing is implemented between the customer network and the ISP network, the edge router of the customer must announce itself as a default gateway or a gateway of last resort. This procedure must be done using the IGP within the customer network because different routers within the customer network must be able to select the best route to the exit point of the network.

Different IGPs use different methods of announcing a router as a gateway of last resort. Enhanced Interior Gateway Routing Protocol (EIGRP) uses the concept of default network, while Open Shortest Path First (OSPF) and Routing Information Protocol (RIP) send reachability information about network 0.0.0.0/0. In either case, the network operators of the customer network are responsible for configuring their network to use the customer edge router as a gateway of last resort.

When static routing is used between the customer and the provider, the edge router of the provider must propagate a static route that points to the customer network, to all other routers within the ISP network, and also to the rest of the Internet. The network operators in the ISP network propagate the route using a configuration command to start redistributing the routes into BGP.

Customer routes should not be redistributed into the IGP of the ISP network. Care should be taken that the IGP of the ISP network does not carry too many routes. Redistributing customer routes into the IGP could potentially cause poor performance and might eventually cause a complete shutdown of IGP routing at the service provider.

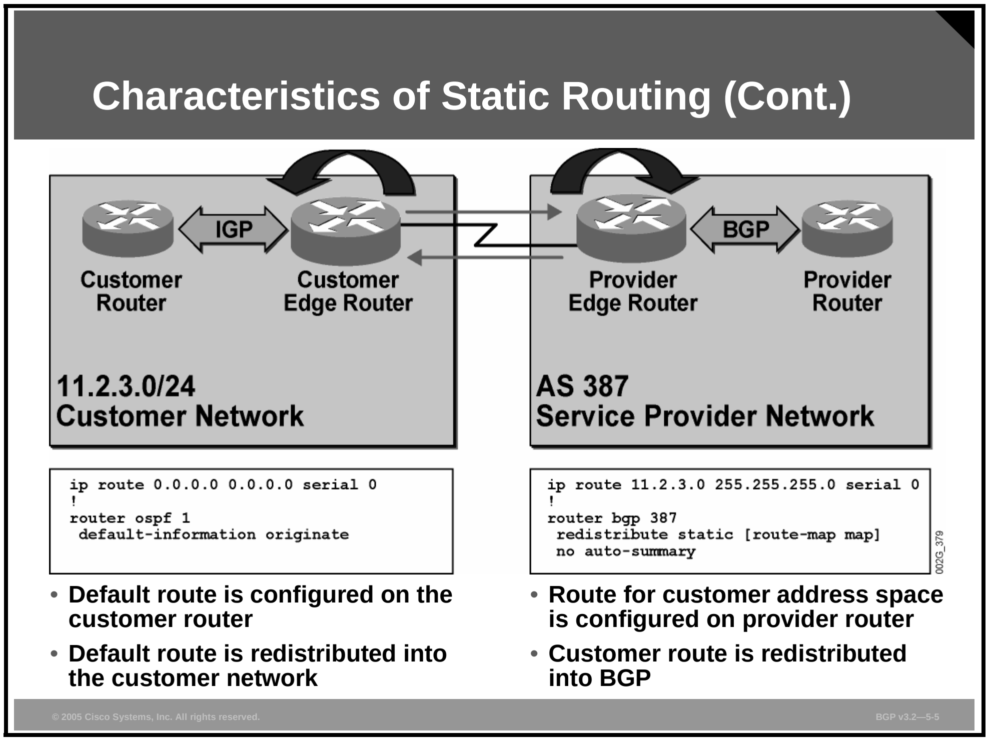
If a customer uses provider-assigned (PA) addresses and the ISP announces a large block of addresses for which the network of this customer is only a small portion of the block, then the routes of this customer should not be propagated by the service provider to the rest of the Internet. Instead, the rest of the Internet should receive only an announcement containing the larger block of addresses.

An easy way of achieving this setup is to use communities within the ISP network. Any customer route that should not be announced to the rest of the Internet is marked using the no-export community. To ensure that the BGP communities are propagated, at least over all Internal Border Gateway Protocol (IBGP) sessions, the network operators of the ISP network must configure a send-community option for all IBGP neighbors. The edge routers of the ISP network then see the no-export community and filter those routes out before sending the update to External Border Gateway Protocol (EBGP) neighbors.

Communities are set using route-maps. A route-map can select routes based on various attributes. One of these attributes is the route tag. Through configuration, a route tag can be assigned by the router to specific static routes. This option means that the network operators of the ISP network can invent a scheme of tagging where all static routes that should not be propagated to other autonomous systems are assigned a specific tag. Then a route-map can select all routes with that tag and assign them the no-export community.

Example: Characteristics of Static Routing

In the figure, the customer network is connected to the Internet by using a single permanent connection to a single service provider.



Characteristics of Static Routing (Cont.)

```
ip route 0.0.0.0 0.0.0.0 serial 0
!
router ospf 1
 default-information originate
```

```
ip route 11.2.3.0 255.255.255.0 serial 0
!
router bgp 387
 redistribute static [route-map map]
 no auto-summary
```

- **Default route is configured on the customer router**
- **Default route is redistributed into the customer network**

- **Route for customer address space is configured on provider router**
- **Customer route is redistributed into BGP**

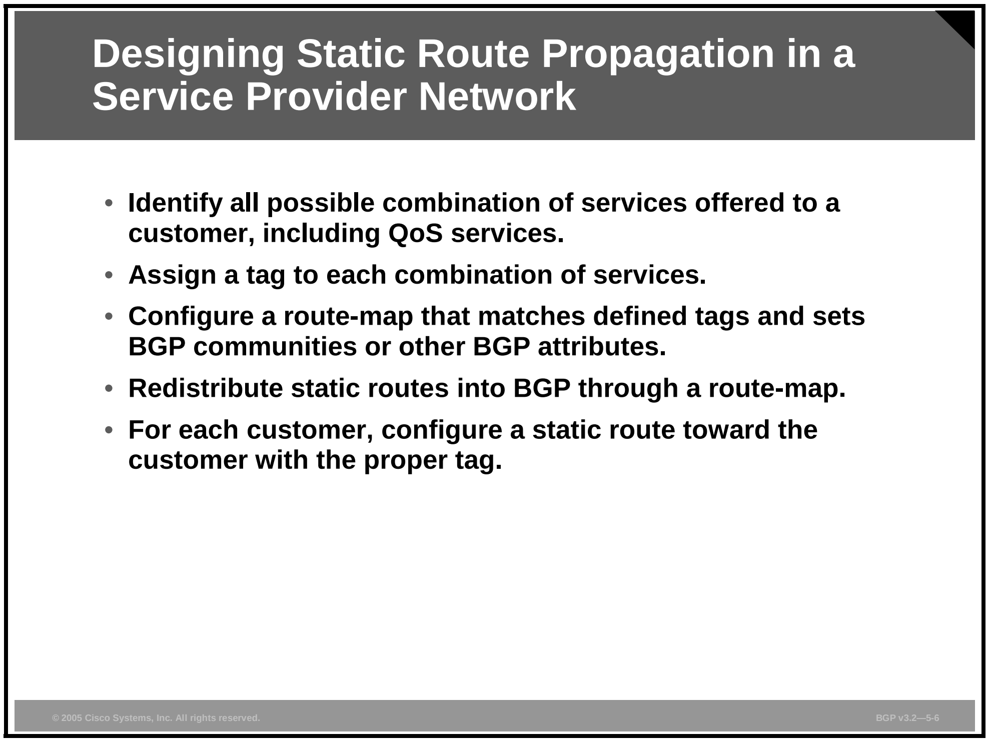
In this case, a routing protocol does not add any redundancy and would only add complexity.

The customer edge router has a static default route pointing to the interface serial 0. If the serial interface goes down, the route becomes invalid. The **default-information originate** command is configured in the OSPF process on the customer router; therefore, the router announces a default route into OSPF as long as it has a valid default route itself.

The service provider edge router also has a static route, declaring the customer IP network number as reachable over the serial 0 interface. It also becomes invalid if the interface goes into the down state. The ISP edge router must forward this information to all other ISP routers and to the rest of the Internet. This action is accomplished by redistributing the static route into BGP. As long as the static route is valid, BGP announces it. To the rest of the Internet, the customer network appears as reachable within the autonomous system (AS) of the ISP. As far as the rest of the Internet is concerned, the customer is a part of the service provider AS.

Designing Static Route Propagation in a Service Provider Network

This topic identifies the design considerations for propagating static routes in a service provider network.



Designing Static Route Propagation in a Service Provider Network

- **Identify all possible combination of services offered to a customer, including QoS services.**
- **Assign a tag to each combination of services.**
- **Configure a route-map that matches defined tags and sets BGP communities or other BGP attributes.**
- **Redistribute static routes into BGP through a route-map.**
- **For each customer, configure a static route toward the customer with the proper tag.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-6

You can easily extend the principle of using tags when you are configuring static routes, and of assigning different communities based on those tags, to implement a more complex routing policy. To propagate static routes in a service provider network, complete these steps:

Step 1 Identify all different service levels that are offered to customers and then all the different combinations of these service levels.

Step 2 Assign each combination its own tag value and its own community.

Step 3 Configure a route-map, which selects routes with each of the assigned tags and sets the corresponding community value. Because the processing of a route-map stops when the match clause of a statement is met, each route should be assigned a single combination of communities only. Therefore, you must take great care to assign a tag and a community combination to each combination of services that are provided.

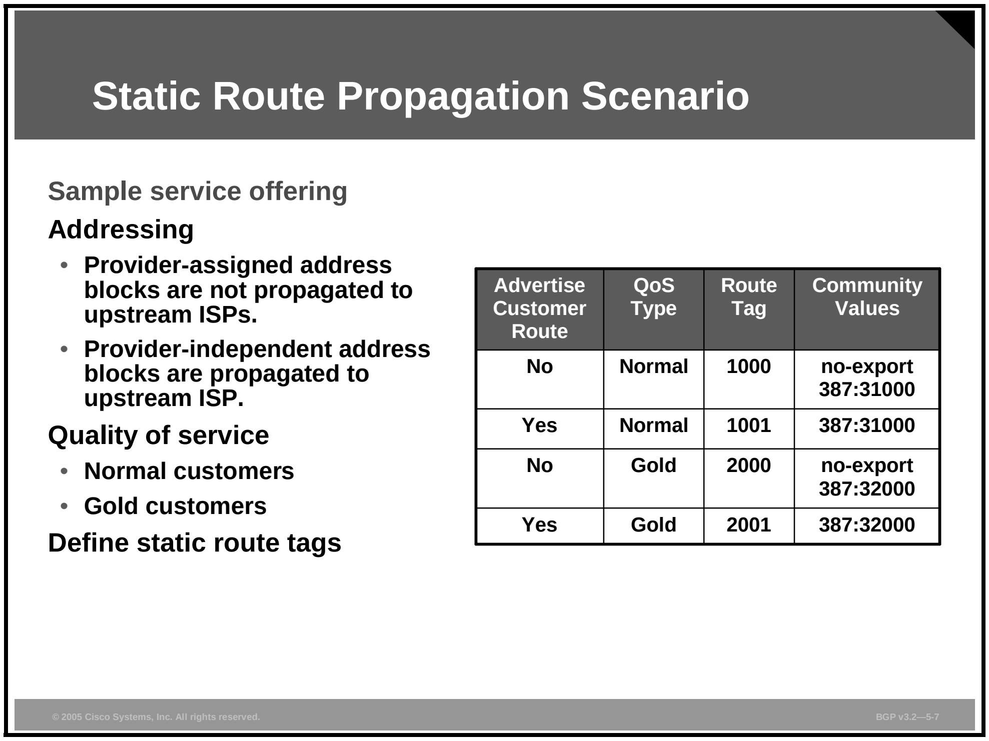
Note When the provider edge routers redistribute static routes into BGP, these routes must pass through the route-map. BGP assigns the correct community depending on the tag values that are given on the configuration line for each of the static routes.

Step 4 Finally, configure static routes. Before you configure a static route for a specific customer, you must identify the combination of the services that are provided to this customer. Then you must look up the corresponding tag value. After you have configured the route, you must assign the tag.

With this routing policy, every static route to a customer network is assigned a tag and the redistributed BGP route is assigned a corresponding community. The BGP communities that are attached to the routes signal to other routers in the ISP network which particular service combination you should use.

Example: Static Route Propagation

This example shows a scenario with varied service levels in which static route propagation is configured in a BGP environment.



Static Route Propagation Scenario

Sample service offering

Addressing

- **Provider-assigned address blocks are not propagated to upstream ISPs.**
- **Provider-independent address blocks are propagated to upstream ISP.**

Quality of service

- **Normal customers**
- **Gold customers**

Define static route tags

Advertise Customer Route	QoS Type	Route Tag	Community Values
No	Normal	1000	no-export 387:31000
Yes	Normal	1001	387:31000
No	Gold	2000	no-export 387:32000
Yes	Gold	2001	387:32000

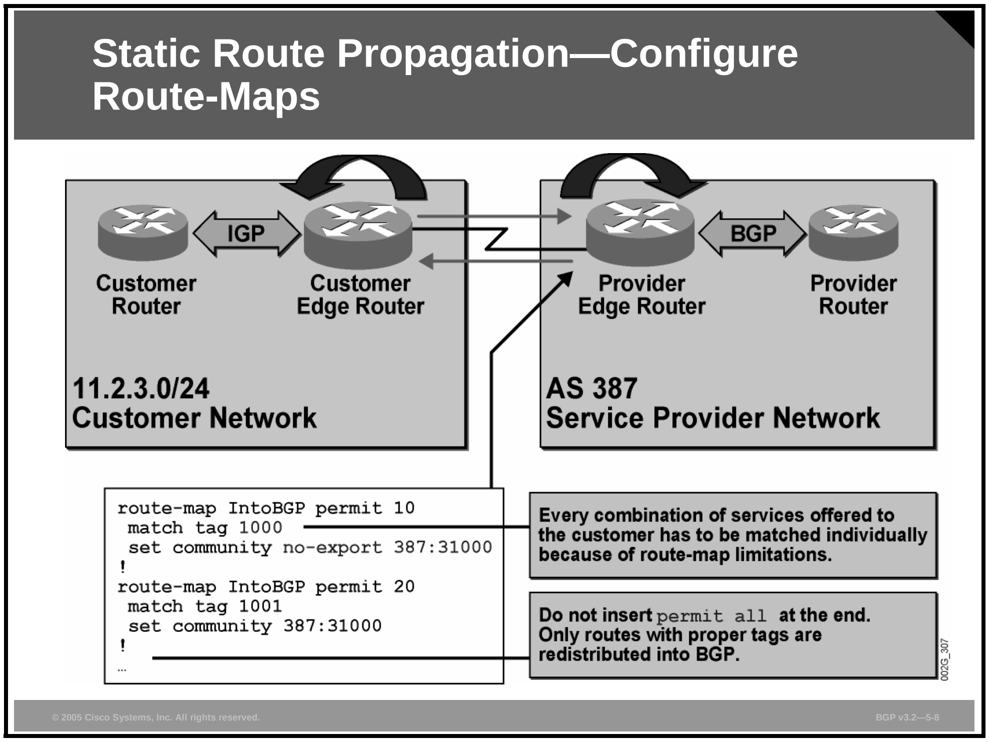
BGP v3.2—5-7

In this scenario, the service provider offers two different service levels to its customers: Normal and Gold. Customers are also assigned IP address blocks. Some customers have PA addresses, which the ISP does not announce as explicit routes. The large route summary block announced by the ISP covers these customers. Other customers use provider-independent (PI) addresses that must be explicitly announced to the Internet by the service provider.

Because there are two different quality of service (QoS) services, Normal and Gold, and because there are both PA and PI addresses, the total number of combinations to cover the network policy is four:

- Normal QoS routes that are assigned by the ISP and should not be explicitly announced
- Normal QoS routes that are PI routes and should be explicitly announced
- Gold QoS routes that are assigned by the ISP and should not be explicitly announced
- Gold QoS routes that are PI routes and should be explicitly announced

Each of these four combinations receives its own tag value and its own community combination.



Static Route Propagation—Configure Route-Maps

```
route-map IntoBGP permit 10
 match tag 1000
 set community no-export 387:31000
!
route-map IntoBGP permit 20
 match tag 1001
 set community 387:31000
!
...
```

Every combination of services offered to the customer has to be matched individually because of route-map limitations.

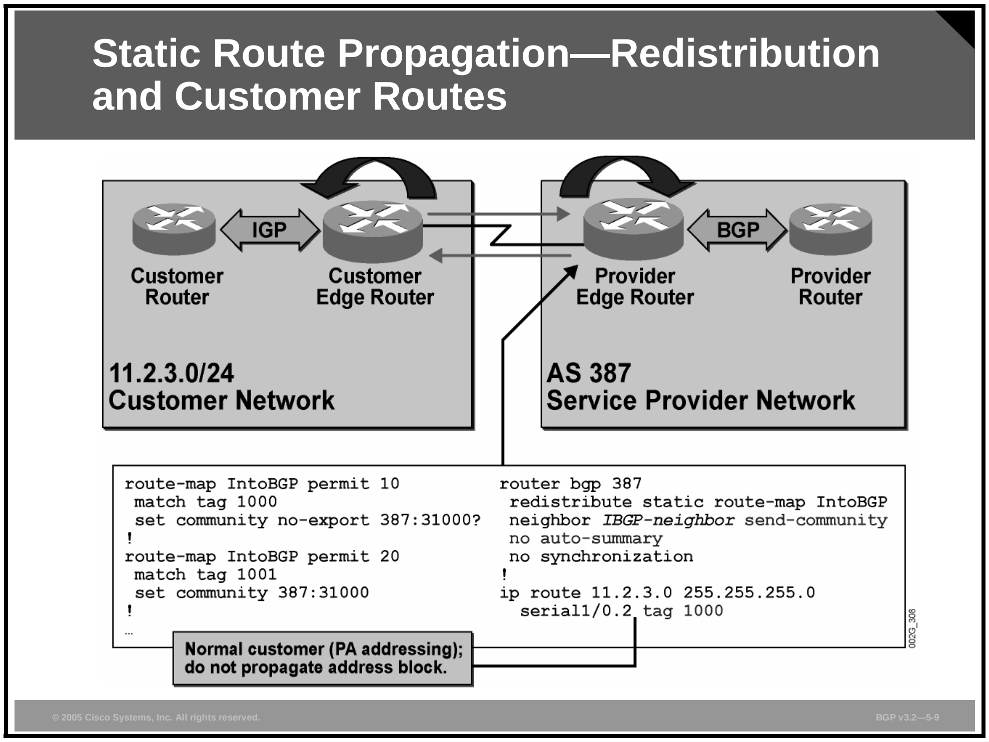
Do not insert `permit all` **at the end. Only routes with proper tags are redistributed into BGP.**

BGP v3.2—5-8

Network operators configure a route-map in the ISP edge router that has the static routes to the customer network. Redistribution of the configured customer static routes into BGP is also performed at the ISP edge router.

Because a route-map can match an individual route in a single route-map statement only, a single tag value, representing each combination of services, must be assigned to the static routes by the router. When a route is matched, the interpretation of the route-map for that individual route stops. The route-map has one statement for each combination, and each statement matches a tag value and assigns the corresponding community combination for that tag.

The route-map is applied during the redistribution of customer static routes into BGP at the provider edge router. Because the route-map has no “permit any” statement at the end, the static routes that are not assigned any of the tags being used are not redistributed. The route-map filters these routes out, forcing the network operators to make a tag assignment to all customer routes. Furthermore, the route-map filtering can help catch administrator configuration entry errors, thus giving all customers the service combination that they are entitled to.



Static Route Propagation—Redistribution and Customer Routes

```
route-map IntoBGP permit 10
 match tag 1000
 set community no-export 387:31000?
!
route-map IntoBGP permit 20
 match tag 1001
 set community 387:31000
!
…
```

```
router bgp 387
 redistribute static route-map IntoBGP
 neighbor IBGP-neighbor send-community
 no auto-summary
 no synchronization
!
ip route 11.2.3.0 255.255.255.0
  serial1/0.2 tag 1000
```

Normal customer (PA addressing); do not propagate address block.

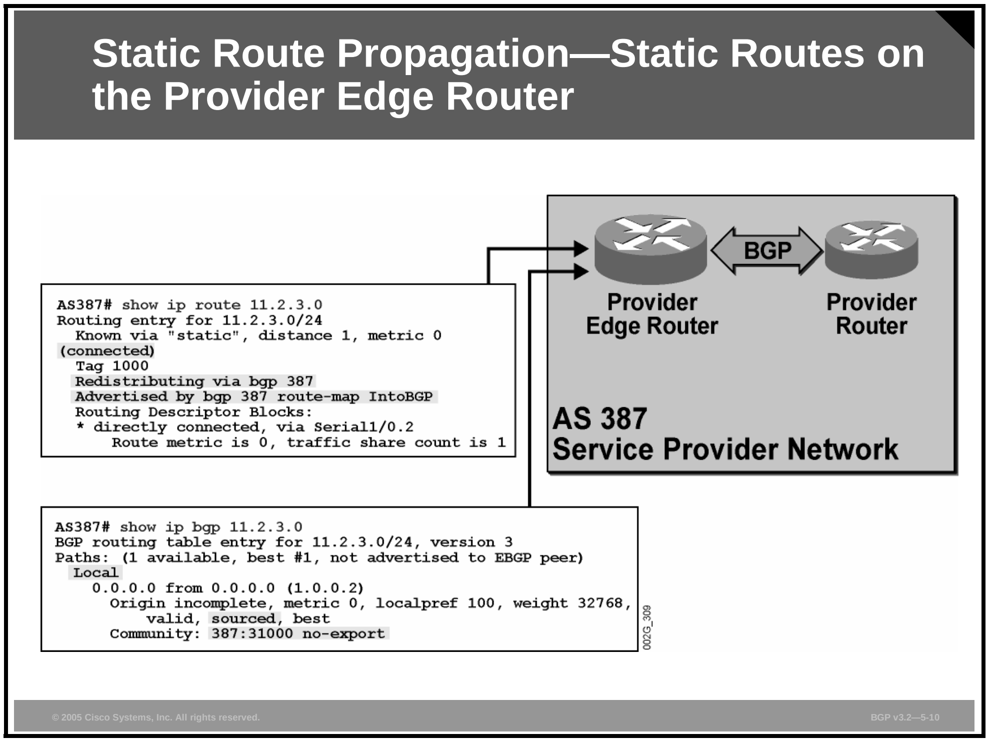
The figure shows how the service provider edge router uses the route-map named “IntoBGP” when redistributing static routes into BGP. Because the route-map assigns community values that will be used by other routers within the ISP network, network operators must configure all IBGP neighbors with the **send-community** qualifier.

Use the **no auto-summary** BGP configuration command to avoid having the subnet 11.2.3.0/24 automatically summarized into 11.0.0.0/8.

When connecting customers, the network operators identify which service combination to use for this particular customer. The three services associated with this particular customer are as follows:

- Apply normal QoS
- Use a PA network number
- Do not enable the provider to explicitly announce customer routes

A static route to the customer is configured and assigned the appropriate tag value of 1000, which represents the specified services that are assigned to the customer.



Static Route Propagation—Static Routes on the Provider Edge Router

```
AS387# show ip route 11.2.3.0
Routing entry for 11.2.3.0/24
  Known via "static", distance 1, metric 0
(connected)
  Tag 1000
  Redistributing via bgp 387
  Advertised by bgp 387 route-map IntoBGP
  Routing Descriptor Blocks:
  * directly connected, via Serial1/0.2
      Route metric is 0, traffic share count is 1
```

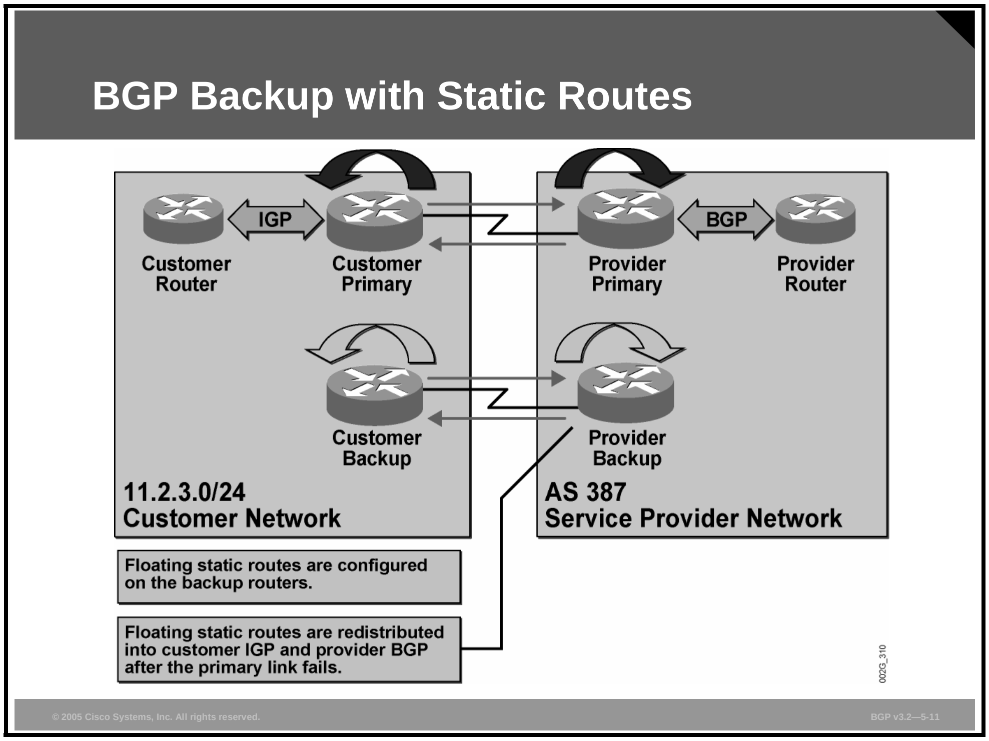
```
AS387# show ip bgp 11.2.3.0
BGP routing table entry for 11.2.3.0/24, version 3
Paths: (1 available, best #1, not advertised to EBGP peer)
  Local
    0.0.0.0 from 0.0.0.0 (1.0.0.2)
      Origin incomplete, metric 0, localpref 100, weight 32768,
          valid, sourced, best
      Community: 387:31000 no-export
```

The **show ip route** command displays information from the routing table about subnet 11.2.3.0/24. The route is learned by static configuration and is redistributed via BGP. The router, through the use of a statically assigned tag, has assigned a tag value of 1000 to the customer route, and the route must pass through the route-map into BGP before being inserted into the BGP table.

The **show ip bgp** command displays information from the BGP table about subnet 11.2.3.0/24. The route is local within this AS and is sourced by this router. The BGP communities 387:31000 and no-export have been assigned by the router to the redistributed customer route by using the provider-defined route-map prior to inserting the customer route into the BGP table.

BGP Backup with Static Routes

This topic explains how to configure a typical backup setup that uses static routing between a customer and a service provider in a BGP environment.



BGP Backup with Static Routes

Floating static routes are configured on the backup routers.

Floating static routes are redistributed into customer IGP and provider BGP after the primary link fails.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-11

This example illustrates a case where the customer network has two connections to a single service provider. One connection between the customer network and the ISP is the primary connection, and the other connection is used for backup purposes only. If link-level procedures can detect link failures and a failure in the remote router, then static routing can be used instead of a dynamic routing protocol between the customer and provider networks.

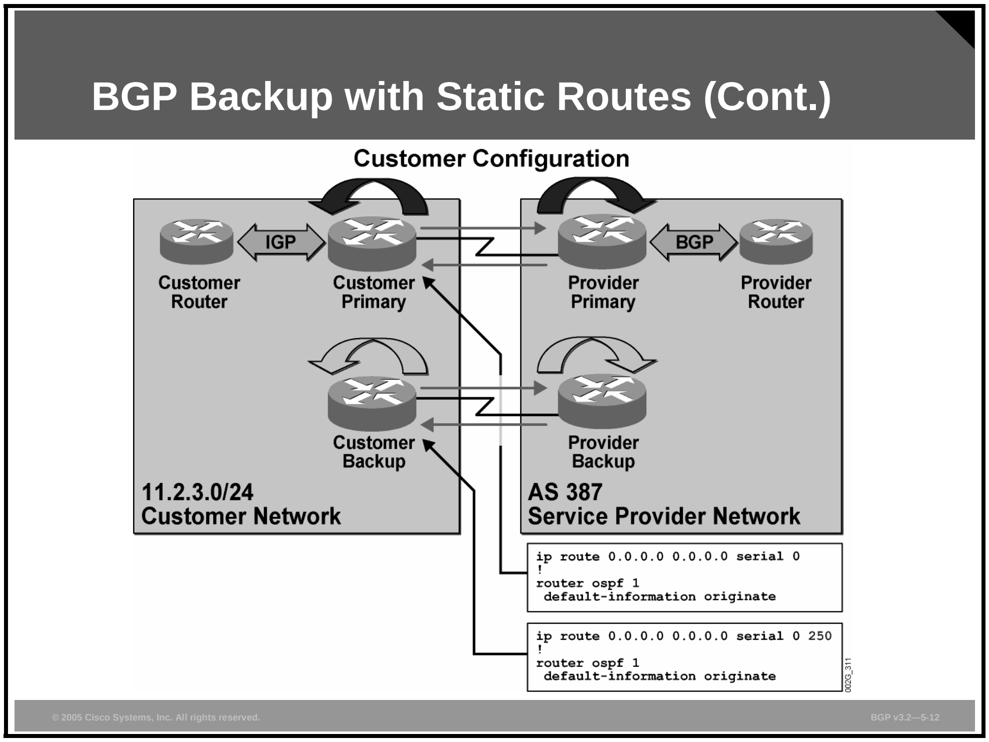
As in the previous example, where no backup link is available, the primary edge router of the customer has a static default route toward the ISP and the primary edge router of the ISP has static routes toward the customer. The customer router redistributes the static default route into its IGP. The ISP router redistributes the static routes into BGP.

If the primary link goes down, the link-level procedures set the interface to the down state, causing the static routes pointing out through the interface to be invalid and removing the routes from the routing table. When the interface changes back to the up state, the static route will reappear in the routing table.

Redistribution of routes into any routing protocol is conditioned by the appearance of the route in the routing table. Thus, if the interface goes down, the router removes the static route from its routing table, and the route is withdrawn from the routing protocol. When the static route reappears, the redistribution process inserts it into the routing protocol again.

The backup edge router of the customer also uses a default static route toward the ISP, via the backup link. The backup edge router is also redistributing the default route into the IGP. However, the static route that is used is a floating static route, which is assigned a high administrative distance (AD), higher than the AD of the customer IGP. As long as the primary link works, the IGP provides the customer backup edge router with the primary default route. Because of the higher AD, the backup static default route is not installed into the backup router routing table. Because the static route is not in the routing table, it is not redistributed. If the primary link fails, the IGP no longer feeds the backup edge router with a default route. The backup static default route is the only remaining default route. Therefore, the router will install the floating default route into its routing table and subsequently redistribute it into the IGP.

The backup edge router of the ISP can also use floating static routes, which are redistributed into the ISP BGP process.



BGP Backup with Static Routes (Cont.)

```
ip route 0.0.0.0 0.0.0.0 serial 0
!
router ospf 1
 default-information originate
```

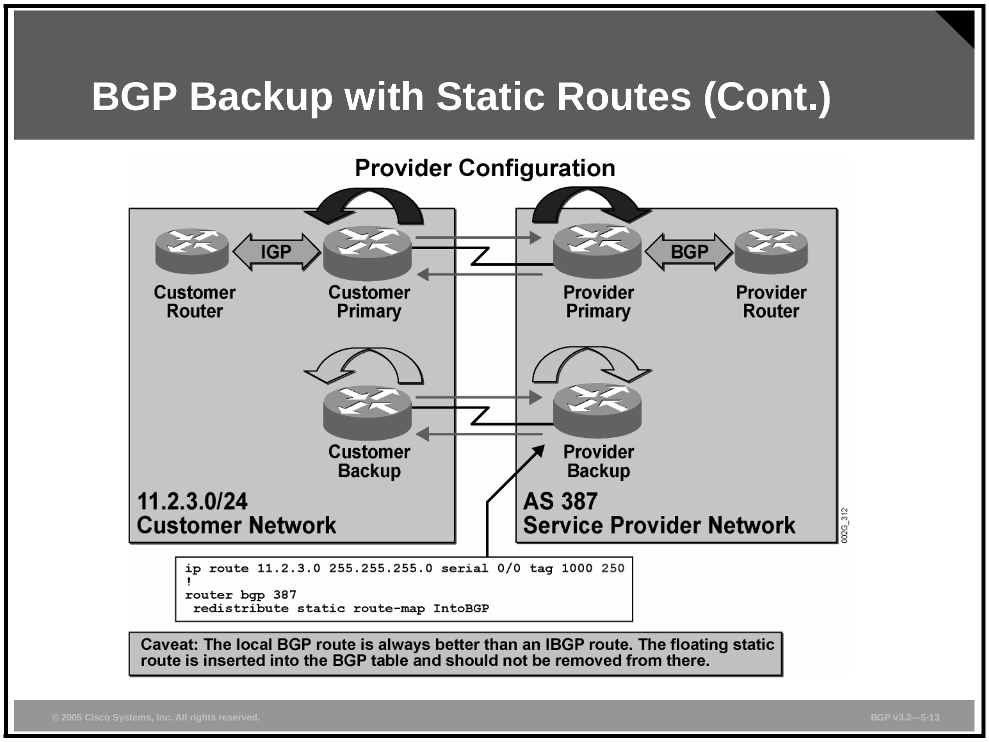
```
ip route 0.0.0.0 0.0.0.0 serial 0 250
!
router ospf 1
 default-information originate
```

In the figure, the customer network and the ISP network are connected using leased lines with High-Level Data Link Control (HDLC) encapsulation. Both the primary and the backup edge routers in the customer network have a static default route toward the serial interface leading to the ISP. Both routers also do redistribution of the default route into the OSPF protocol, which is being used as an IGP within the customer network.

However, the static default route in the backup edge router is configured with an AD value set to 250. This AD value is higher than the AD values of any routing protocol. This configuration means that as long as the backup router receives the default route by OSPF, the static default route is not used.

When the primary link goes down, the static default route in the primary router is not valid. The OSPF protocol stops announcing the default route, because the **default-information originate** command makes OSPF contingent on the availability of that static default route in the routing table before announcement.

The backup router now installs its static default route in the routing table. The conditions for announcing the default route by OSPF are met and the rest of the customer routers see the backup router as the gateway of last resort.



BGP Backup with Static Routes (Cont.)

```
ip route 11.2.3.0 255.255.255.0 serial 0/0 tag 1000 250
!
router bgp 387
 redistribute static route-map IntoBGP
```

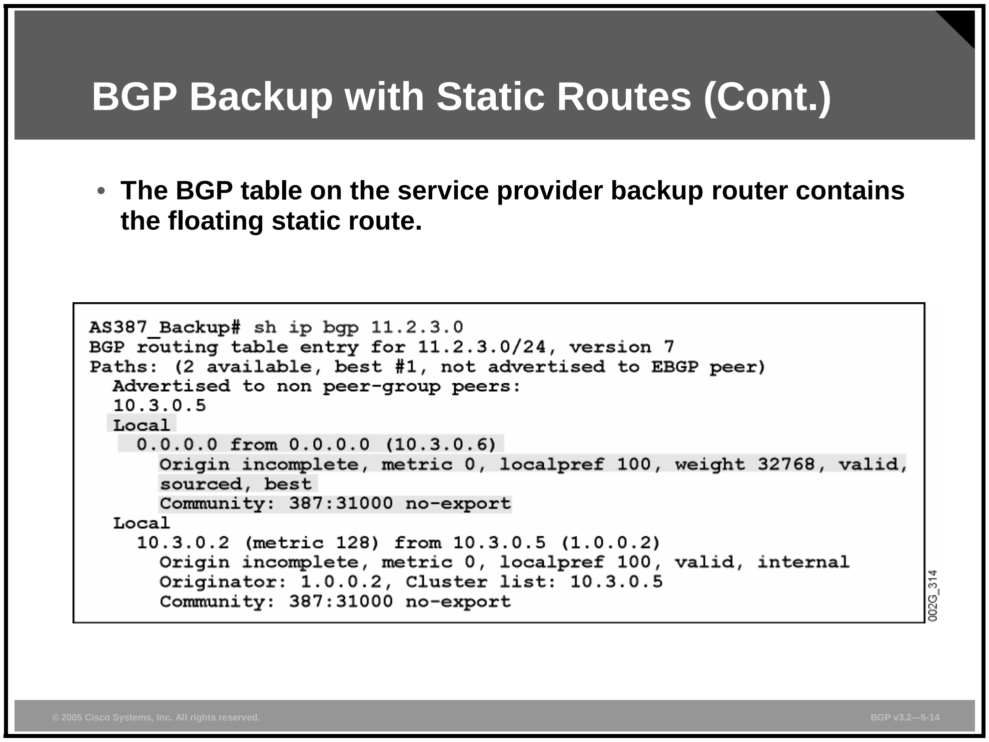
Caveat: The local BGP route is always better than an IBGP route. The floating static route is inserted into the BGP table and should not be removed from there.

When floating static routes are configured on the provider edge routers, they are also redistributed into BGP. This configuration makes things a little bit more complicated.

The network operator configures a floating static route to the customer subnet 11.2.3.0/24. In the provider edge router, the floating static route is assigned the same tag value as the tag value being used in the primary router. The route-map IntoBGP is the same as in the primary router and provides the routes to the customer network with the same communities (the same QoS level and indication whether to explicitly announce route prefix to the rest of the Internet).

The floating static route is configured with an AD value of 250. This value is higher than any routing protocol. When the backup edge router of the ISP no longer receives any routing protocol information about the customer networks, the router will automatically install the floating static route and subsequently redistribute it into BGP.

Based on BGP route selection rules, the redistributed floating static route will always remain the preferred path if additional BGP configuration is not performed on the provider edge router. This preference means that regardless of whether the primary link comes back, the backup router selects the locally sourced route as the best route. Therefore, the backup router continues to announce a path toward the customer network. The backup link does not go back to the Idle state.



BGP Backup with Static Routes (Cont.)

- **The BGP table on the service provider backup router contains the floating static route.**

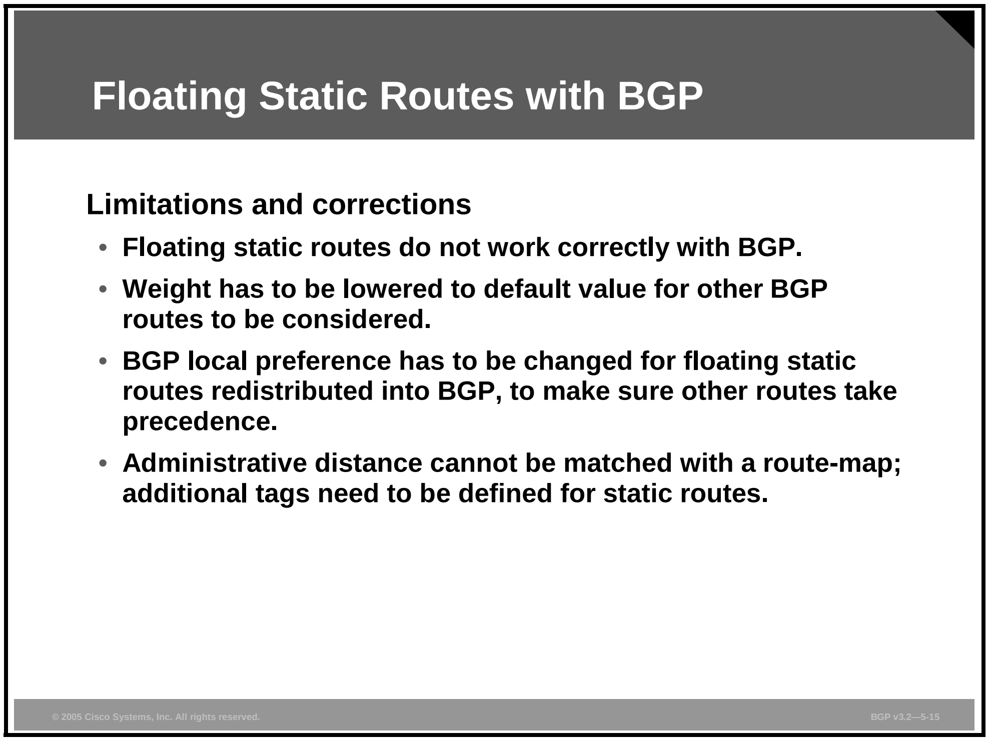
```
AS387_Backup# sh ip bgp 11.2.3.0
BGP routing table entry for 11.2.3.0/24, version 7
Paths: (2 available, best #1, not advertised to EBGP peer)
  Advertised to non peer-group peers:
  10.3.0.5
  Local
    0.0.0.0 from 0.0.0.0 (10.3.0.6)
      Origin incomplete, metric 0, localpref 100, weight 32768, valid,
      sourced, best
      Community: 387:31000 no-export
  Local
    10.3.0.2 (metric 128) from 10.3.0.5 (1.0.0.2)
      Origin incomplete, metric 0, localpref 100, valid, internal
      Originator: 1.0.0.2, Cluster list: 10.3.0.5
      Community: 387:31000 no-export
```

In this example, the **show ip bgp** command is used in the backup edge router of the provider to display the information about the customer network 11.2.3.0/24. The primary link has come back, so the backup router now sees two alternate routes. The first route is the route that the router itself has redistributed into BGP using the floating static route. This route is locally sourced by this AS and has been assigned a weight value of 32768. The second route is the one that has been received by IBGP from the primary edge router. This AS also sources this route, but no weight value is assigned.

The BGP route selection algorithm selects the route with weight value 32768 as the best. As a result, the route that was received from the primary edge router is not a candidate to be installed in the routing table and never competes with the floating static route. The floating static route stays in the routing table, and redistribution of the route continues until the backup link goes down and the route becomes invalid.

Floating Static Routes with BGP

This topic describes the limitations of floating static routes when used in typical backup static routing scenarios and the corrective actions to overcome these limitations.



Floating Static Routes with BGP

Limitations and corrections

- **Floating static routes do not work correctly with BGP.**
- **Weight has to be lowered to default value for other BGP routes to be considered.**
- **BGP local preference has to be changed for floating static routes redistributed into BGP, to make sure other routes take precedence.**
- **Administrative distance cannot be matched with a route-map; additional tags need to be defined for static routes.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-15

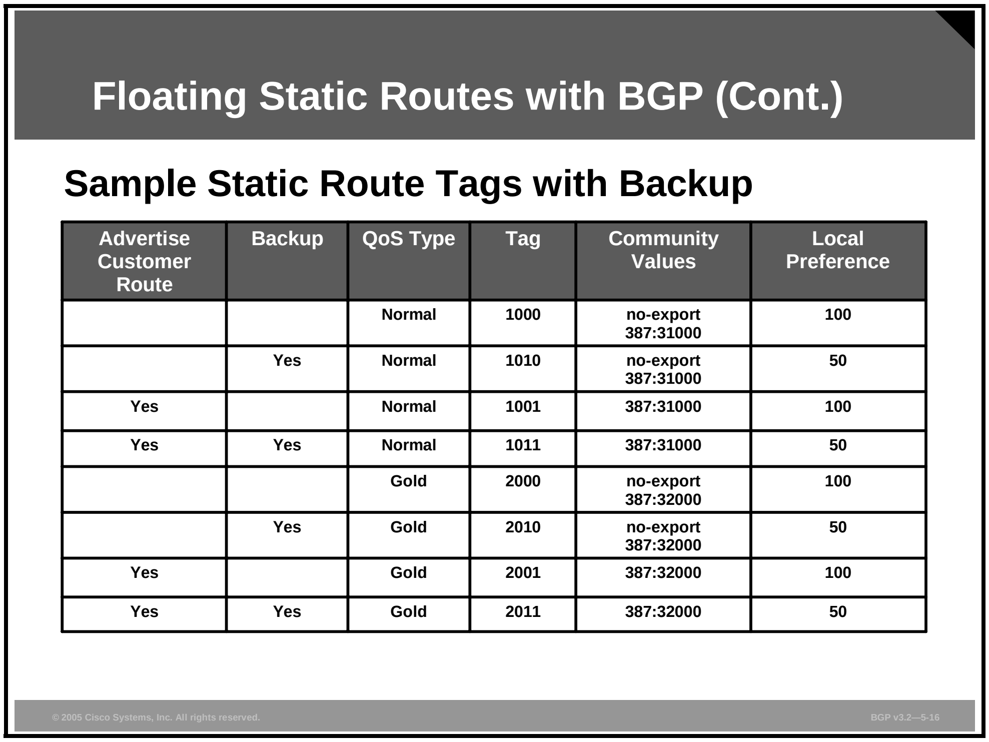
Unfortunately, floating static routes do not work correctly with BGP. After they are inserted, the floating static route is never removed from the routing table even if the primary link comes back.

Whenever you use floating static routes in combination with redistribution into BGP, you will need to take additional configuration steps to ensure that the BGP route selection algorithm selects the primary route as the best BGP route when it reappears:

- When a router redistributes a floating static route into BGP, the weight value assigned to the floating static route must be reduced. Otherwise, the floating static route will always be selected as the best BGP route after the first failure of the primary link occurs.
- Local preference values must be also be assigned by the router to the floating static route so that the floating static route has a lower local preference than the primary route. This assignment ensures that the primary route is selected as the best BGP route after it comes back.

These two requirements must be specified on the provider edge router in the route-map IntoBGP that is used for the redistribution. The route-map must select the floating static routes and set weight and local preference. However, a route-map cannot do matching based on the AD value that has been assigned to a static route. Some other means are required to make it possible for the route-map to distinguish between normal static routes that should have normal weight and local preference and the floating static ones that should have their values modified.

The solution is to create additional tag values for this set of static routes. The tag value must not only reflect the QoS level and whether to announce the route, but the tag value must also indicate if it is a primary route or a backup route.



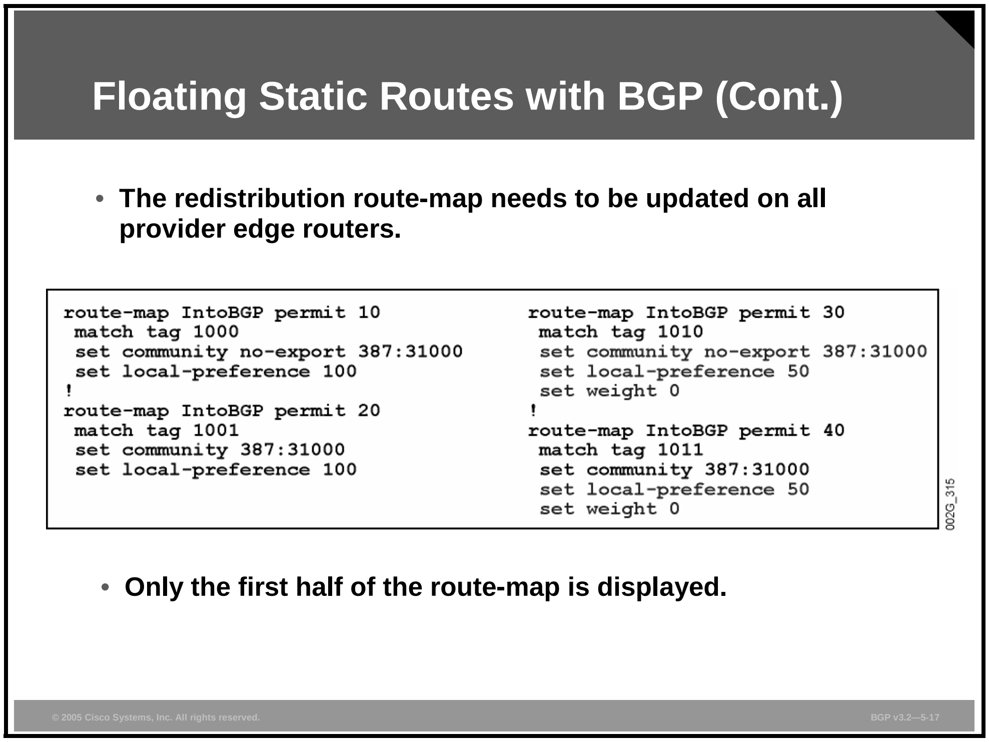
Floating Static Routes with BGP (Cont.)

Sample Static Route Tags with Backup

Advertise Customer Route	Backup	QoS Type	Tag	Community Values	Local Preference
		Normal	1000	no-export 387:31000	100
	Yes	Normal	1010	no-export 387:31000	50
Yes		Normal	1001	387:31000	100
Yes	Yes	Normal	1011	387:31000	50
		Gold	2000	no-export 387:32000	100
	Yes	Gold	2010	no-export 387:32000	50
Yes		Gold	2001	387:32000	100
Yes	Yes	Gold	2011	387:32000	50

Eight tag values have currently been identified. Each tag value indicates a specific combination of explicit route propagation (backup or primary) and QoS level.

When network operators configure static routes in the provider edge router, they must consider which of the combinations that they should use for the route. The route-map that they use when redistributing the static routes into BGP must be configured to recognize all eight combinations and to set the appropriate weight and community and local preference values.



Floating Static Routes with BGP (Cont.)

- **The redistribution route-map needs to be updated on all provider edge routers.**

```
route-map IntoBGP permit 10
 match tag 1000
 set community no-export 387:31000
 set local-preference 100
!
route-map IntoBGP permit 20
 match tag 1001
 set community 387:31000
 set local-preference 100
route-map IntoBGP permit 30
 match tag 1010
 set community no-export 387:31000
 set local-preference 50
 set weight 0
!
route-map IntoBGP permit 40
 match tag 1011
 set community 387:31000
 set local-preference 50
 set weight 0
```

002G_315

- **Only the first half of the route-map is displayed.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-17

The configuration output in the figure displays the first half of the route-map IntoBGP. The output shows how four of the eight different tags are identified by match clauses. For each of the tag values, the route-map sets the community, the local preference, and, in some cases, the weight.

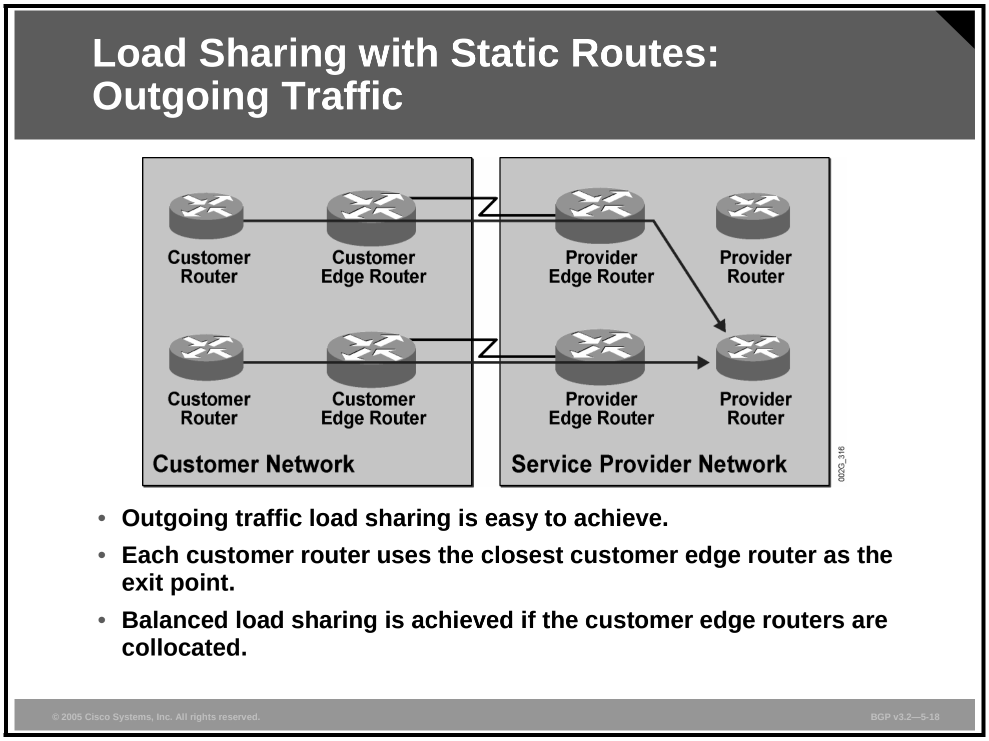
Because the displayed half of the route-map deals only with the four tags that indicate QoS Normal, all statements in the configuration display have set the BGP community attribute to 387:31000. The part of the route-map that is not shown deals with the four tags that indicate QoS Gold, which would be configured to set the BGP community attribute to 387:32000.

Tag values of 1000, 1010, 2000, and 2010 indicate that the route should not be explicitly propagated. The routes that should not be explicitly advertised by the provider to the rest of the Internet are assigned the no-export community by the route-map.

Tag values 1010, 1011, 2010, and 2011 all indicate that the route is a backup route. Those tags have their weight value set to 0 and their local preference value set to 50. These settings ensure that on the return of a failed primary route, the provider edge router will select the primary route as its best path and remove the backup floating static route from its route table.

Load Sharing with Static Routes

This topic describes the characteristics of load sharing when you are configuring static routing between a customer and a service provider.



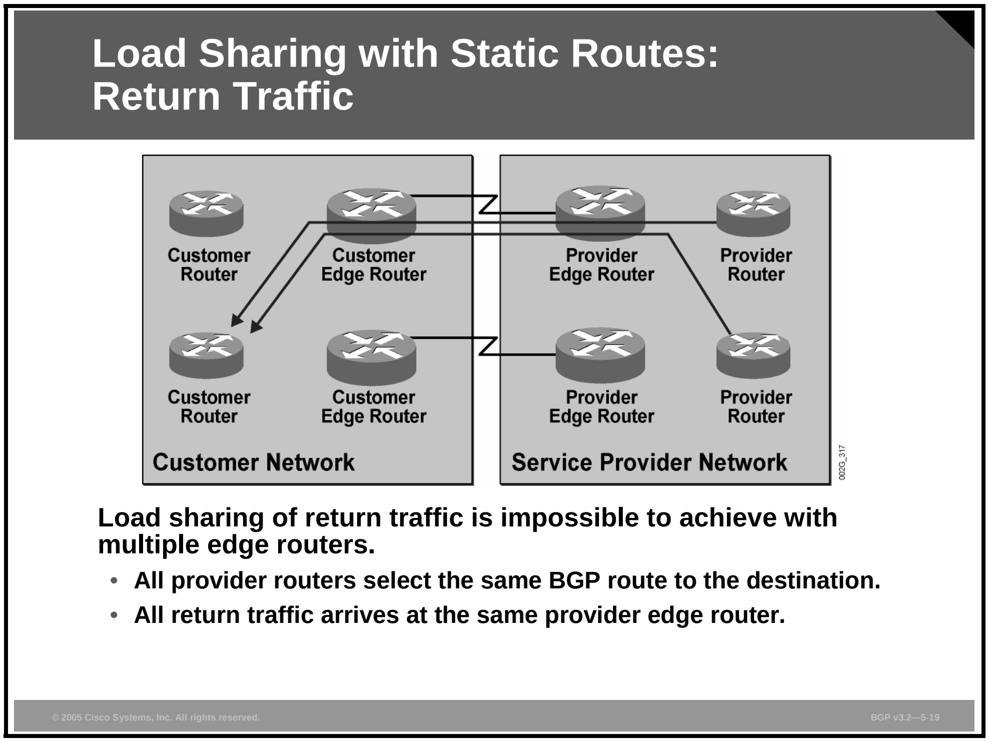
Load Sharing with Static Routes: Outgoing Traffic

- **Outgoing traffic load sharing is easy to achieve.**
- **Each customer router uses the closest customer edge router as the exit point.**
- **Balanced load sharing is achieved if the customer edge routers are collocated.**

Load sharing of outgoing customer traffic is accomplished by configuring a standard default static route in both customer edge routers. Each static route is valid as long as the serial link in each router is up. When both static routes are valid, both customer edge routers announce the default route into the customer network.

The remaining routers in the customer network see two candidate gateways of last resort. These remaining routers choose the closest one, with respect to the IGP metric. The part of the network that is closer to the uppermost exit point uses that exit point for all outgoing traffic. The other part of the network uses the other (lower) exit point.

If both exit points are collocated, they are equally distant from each of the other routers in the customer network. Each router within the customer network therefore uses load sharing of traffic sent out both exit points.



Load Sharing with Static Routes: Return Traffic

Load sharing of return traffic is impossible to achieve with multiple edge routers.

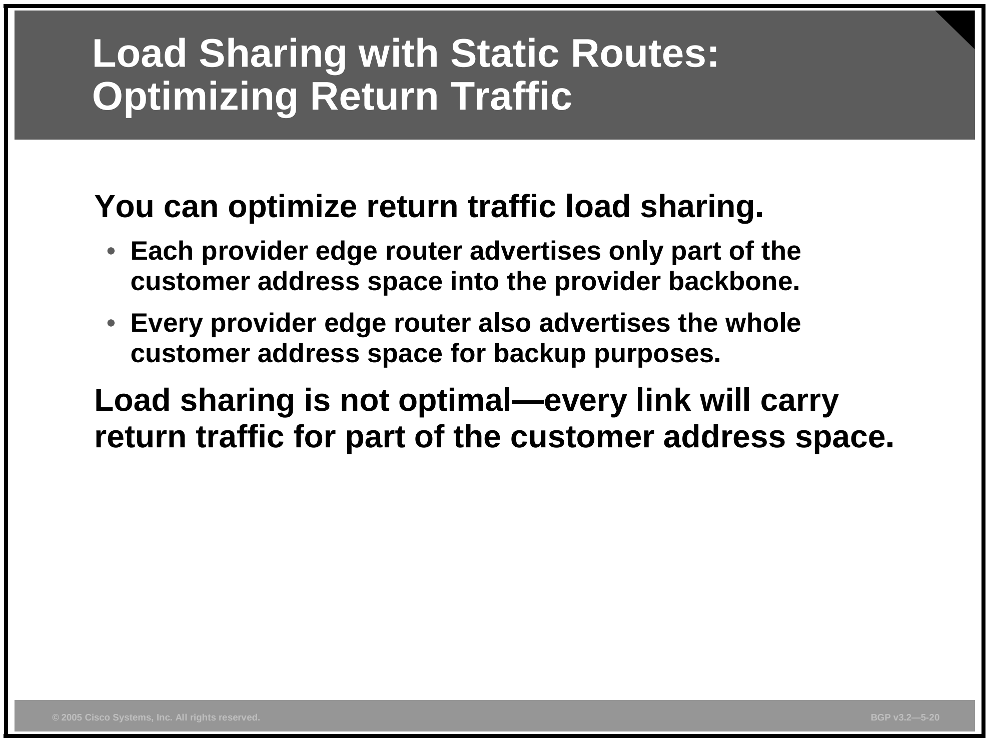
- **All provider routers select the same BGP route to the destination.**
- **All return traffic arrives at the same provider edge router.**

BGP v3.2—5-19

The provider routers receive routes toward the customer network via BGP. BGP in its default behavior selects a single route as the best route, allowing no load sharing. The provider routers that receive the same BGP route from two edge routers will always select the closer edge router (if all other BGP attributes are equal, the IBGP route with the closer next hop is selected). The part of the ISP network that is closer to the uppermost connection uses that connection. The other part of the ISP network uses the other (lower) connection.

If both connection points are collocated, all provider routers select the same IBGP route based on router-ID (because the IGP metrics are always equal) and all the return traffic is sent over a single link toward the customer network, resulting in no load sharing.

Note	Since Cisco IOS Software Release 12.2, the IBGP multipath load-sharing feature enables the BGP-speaking router to select multiple IBGP paths as the best paths to a destination. The best paths or multipaths are then installed in the IP routing table of the router.



Load Sharing with Static Routes: Optimizing Return Traffic

You can optimize return traffic load sharing.

- **Each provider edge router advertises only part of the customer address space into the provider backbone.**
- **Every provider edge router also advertises the whole customer address space for backup purposes.**

Load sharing is not optimal—every link will carry return traffic for part of the customer address space.

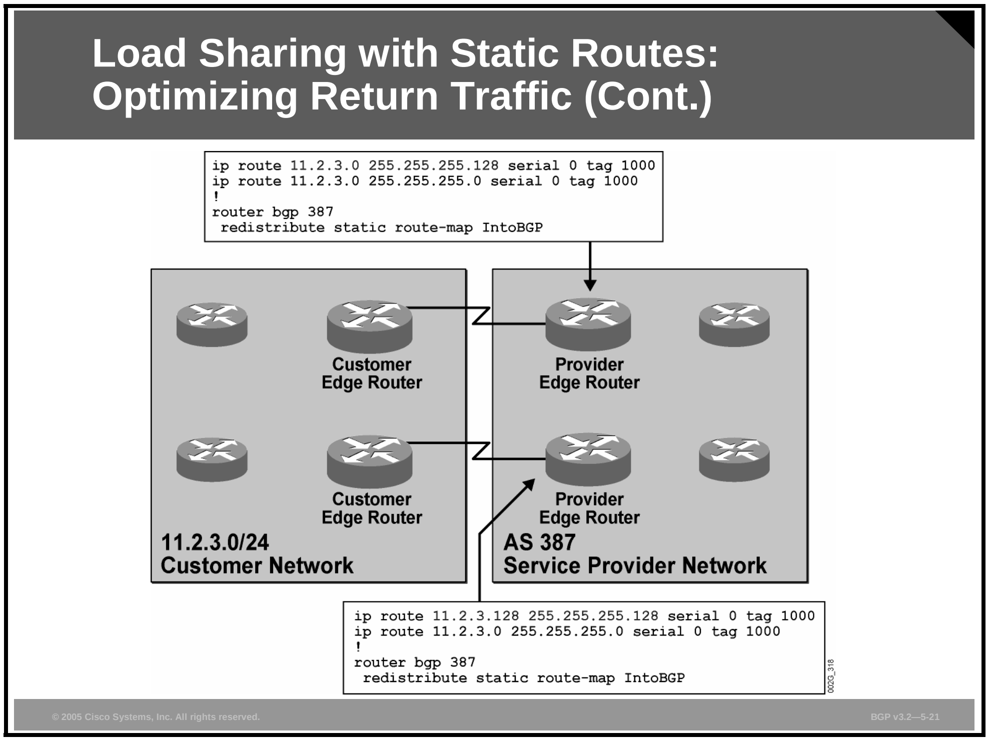
To obtain better control of the return traffic load, the customer address space must be advertised to the provider edge routers using multiple, more explicit routes. The upper edge router could advertise half the address space, and the lower edge router could advertise the other half. For backup reasons, they also should both advertise the entire address space as a larger route summary.

As long as both paths are available, the traffic from the ISP to the customer uses the most explicit route. In this case, two explicit routes are used to send traffic representing one half of the address space over one link and traffic representing the other half of the address space over the other link.

Load sharing in this way does not result in an equal load on the links but rather a statistically based distribution of the traffic load over the links.

Example: Load Sharing with Static Routes

In the example, the customer address space 11.2.3.0/24 is partitioned into two smaller blocks: 11.2.3.0/25 and 11.2.3.128/25.



Load Sharing with Static Routes: Optimizing Return Traffic (Cont.)

```
ip route 11.2.3.0 255.255.255.128 serial 0 tag 1000
ip route 11.2.3.0 255.255.255.0 serial 0 tag 1000
!
router bgp 387
 redistribute static route-map IntoBGP
```

```
ip route 11.2.3.128 255.255.255.128 serial 0 tag 1000
ip route 11.2.3.0 255.255.255.0 serial 0 tag 1000
!
router bgp 387
 redistribute static route-map IntoBGP
```

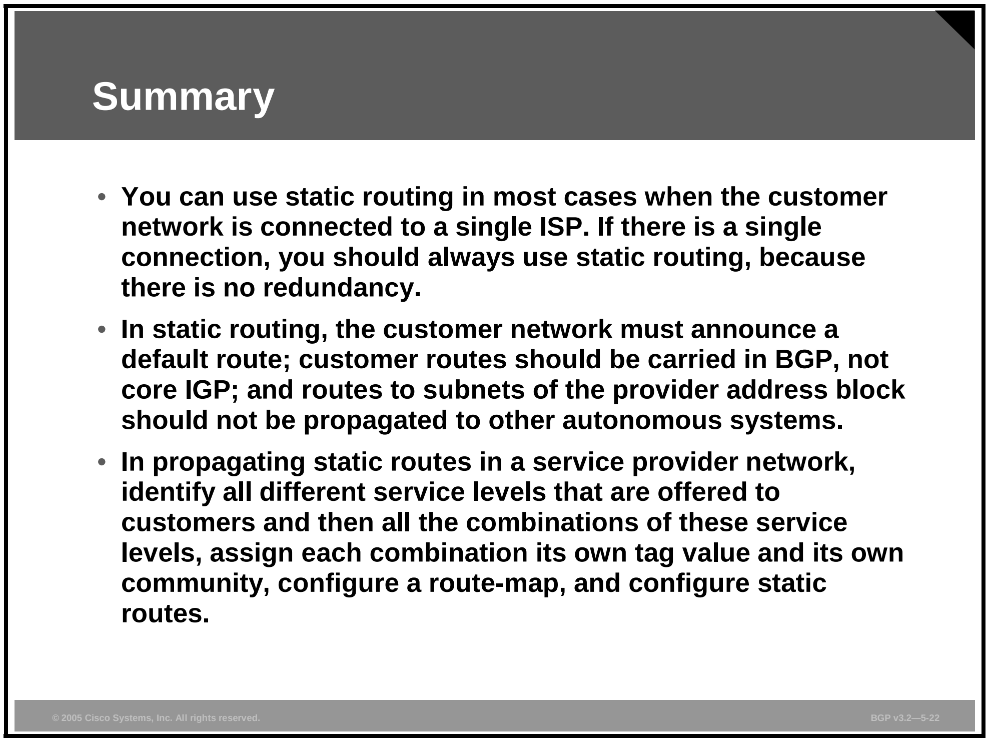
Customer Edge Router
Provider Edge Router
Customer Edge Router
Provider Edge Router
11.2.3.0/24 Customer Network
AS 387 Service Provider Network
002G_318
© 2005 Cisco Systems, Inc. All rights reserved.
BGP v3.2—5-21

The upper provider edge router advertises the route to 11.2.3.0/25, and the lower router advertises the route to 11.2.3.128/25. Both edge routers also advertise the entire address space 11.2.3.0/24.

The routers in the ISP network direct traffic with destination addresses in the 11.2.3.0/25 range to the upper connection point. Traffic to destinations in the 11.2.3.128/25 range is directed to the lower connection point.

Summary

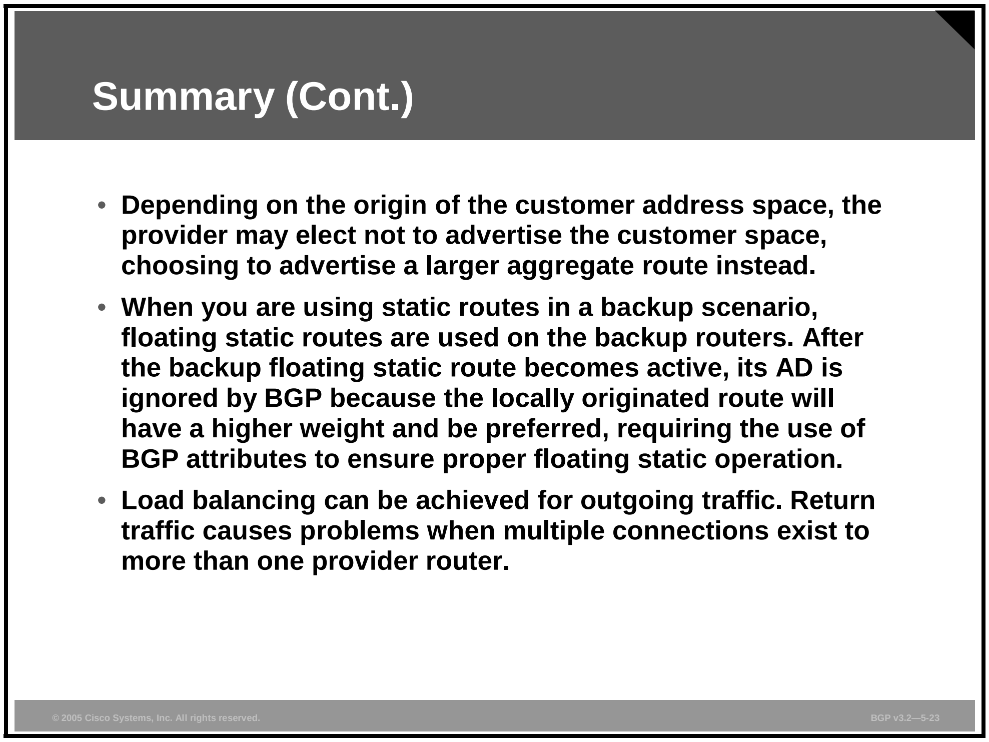
This topic summarizes the key points discussed in this lesson.



Summary

- You can use static routing in most cases when the customer network is connected to a single ISP. If there is a single connection, you should always use static routing, because there is no redundancy.
- In static routing, the customer network must announce a default route; customer routes should be carried in BGP, not core IGP; and routes to subnets of the provider address block should not be propagated to other autonomous systems.
- In propagating static routes in a service provider network, identify all different service levels that are offered to customers and then all the combinations of these service levels, assign each combination its own tag value and its own community, configure a route-map, and configure static routes.

BGP v3.2—5-22



Summary (Cont.)

- Depending on the origin of the customer address space, the provider may elect not to advertise the customer space, choosing to advertise a larger aggregate route instead.
- When you are using static routes in a backup scenario, floating static routes are used on the backup routers. After the backup floating static route becomes active, its AD is ignored by BGP because the locally originated route will have a higher weight and be preferred, requiring the use of BGP attributes to ensure proper floating static operation.
- Load balancing can be achieved for outgoing traffic. Return traffic causes problems when multiple connections exist to more than one provider router.

BGP v3.2—5-23

Lesson 3

Connecting a Multihomed Customer to a Single Service Provider

Overview

When multiple connections to the same service provider are the only means that a customer has of connecting to the Internet, it is important that the connections be correctly configured to ensure proper interaction between the customer and service provider network. It is also important to understand how to configure routing protocols so that customer backup or load-balancing requirements are met.

This lesson discusses the use of multiple connections between a customer and a single Internet service provider (ISP) for backup and load-sharing purposes. Included in this lesson is a discussion of how to configure a customer network and a provider network to accommodate multiple connections between them. Also discussed in this lesson are topics specific to networks with multiple connections between a customer and a single provider, such as private autonomous system (AS) number removal and configuration of a network to support either backup links or load sharing (balancing).

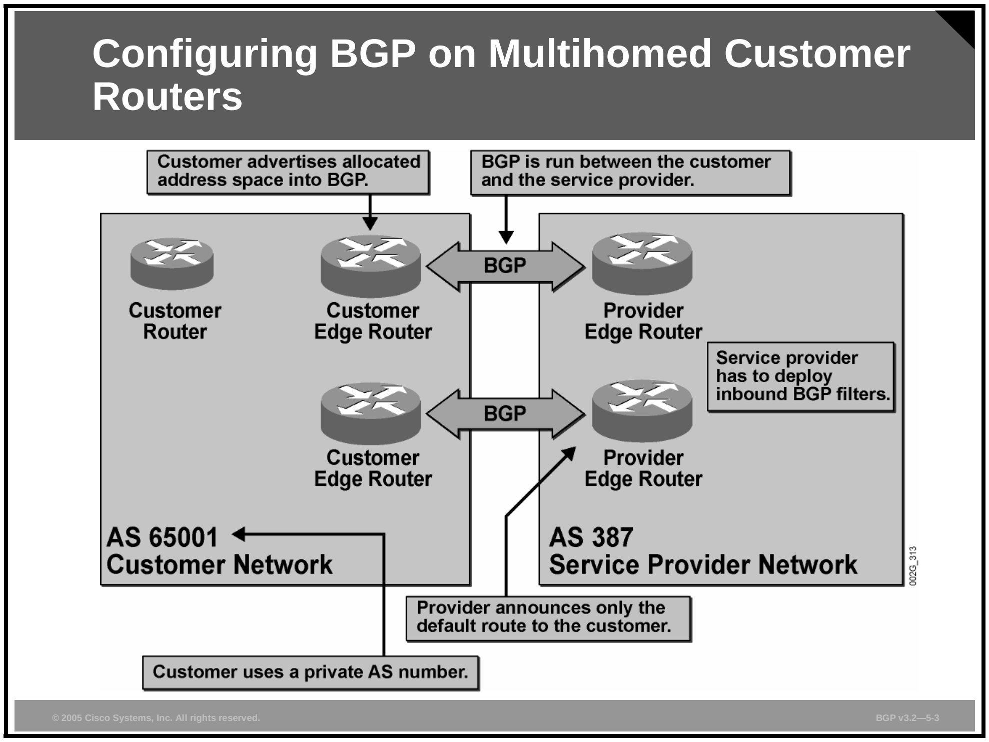
Objectives

Upon completing this lesson, you will be able to implement customer connectivity using BGP in a customer implementation in which you must support multiple connections to a single ISP. This ability includes being able to meet these objectives:

- Configure BGP on a customer network to establish routing between a multihomed customer and a single service provider
- Configure conditional advertising of a customer address space when you are using BGP to establish routing between a multihomed customer and a single service provider
- Configure BGP on a service provider network to establish routing between a multihomed customer and a single service provider
- Disable the propagation of private AS numbers to EBGP peers in a service provider network where a multihomed customer is advertising private numbers in the AS path
- Describe the BGP Support for Dual-AS Configuration for Network AS Migrations feature
- Configure a typical backup setup between a multihomed customer and a single service provider in a BGP environment
- Describe how you can implement load sharing between a multihomed customer and a single service provider in a BGP environment
- Identify the Cisco IOS command that is required to configure load sharing between a multihomed customer and a single service provider using BGP multipath
- Configure load sharing between a multihomed customer and a single service provider using EBGP multihop

Configuring BGP on Multihomed Customer Routers

This topic describes how to configure Border Gateway Protocol (BGP) on a customer network to establish routing between a multihomed customer and a single service provider.



Configuring BGP on Multihomed Customer Routers

Customer advertises allocated address space into BGP.

BGP is run between the customer and the service provider.

Service provider has to deploy inbound BGP filters.

Provider announces only the default route to the customer.

Customer uses a private AS number.

BGP v3.2—5-3

In the example, the customer network is connected to a service provider network using multiple permanent links. BGP is used to exchange routing information between the customer and the provider.

Selecting BGP as the routing protocol between the customer and provider network ensures that a link failure or the failure of a remote router is detected. In this scenario, the customer does not require the use of a public AS number or full Internet routing. Instead, a private AS number is assigned to the customer network, and the ISP sends a default route to the customer through BGP.

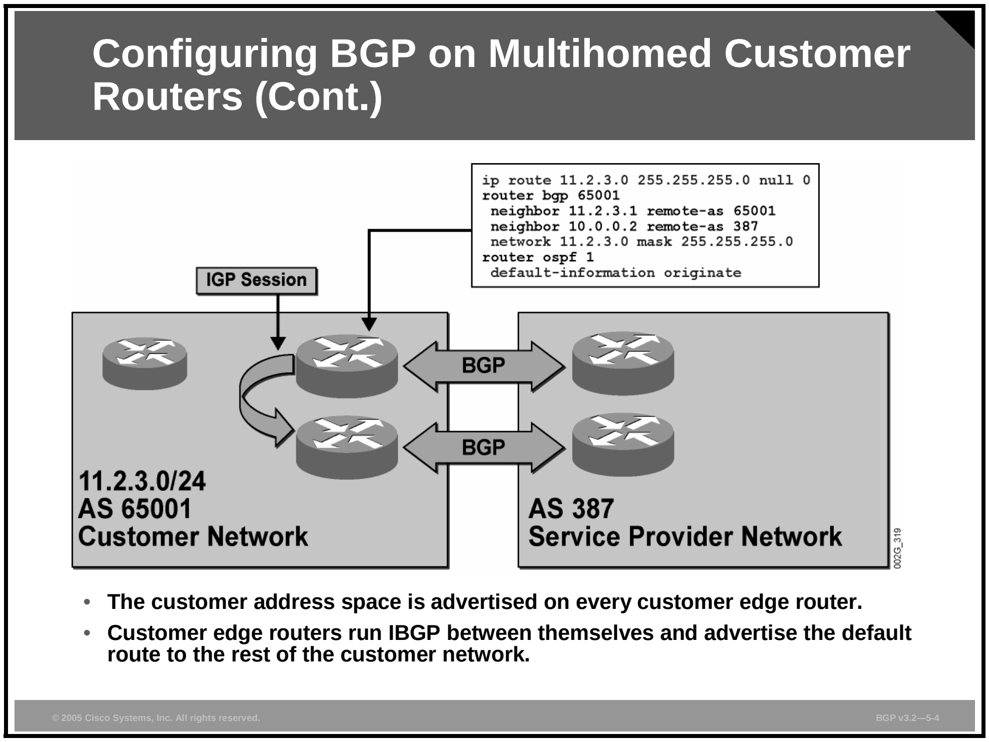
The big difference in this case, as compared to a network scenario where static routes and redistribution are used, is that routers within the private AS of the customer now advertise the customer routes via BGP. Thus, the customer is responsible for announcing its own address space. The ISP receives routes from the customer and conditionally propagates them (similar to static routing). If the customer uses provider-assigned (PA) address space and the ISP can summarize the address space, it will not propagate the explicit routes from the customer to the Internet. The private AS number in the AS-path attribute must be removed before the ISP can propagate any of the customer routes.

Because the customer is now creating BGP routes that are received by the ISP, any error made by the customer can influence routing operation within the ISP network and, if propagated, within the Internet as a whole. Announcing a route to a network to which the customer has not been assigned may cause routing problems. There is always a risk that such routing problems can occur in a service provider network. However, the risk is much greater when the customer, whose network administrators usually have less experience with BGP, enters the configuration.

To reduce the risk of erroneous route advertising, the ISP should always filter any BGP information that it has received from the customer network. The ISP should reject routes to networks that are not expected to be in the customer AS. Routes that contain an AS path with unexpected AS numbers should also be rejected.

Example: Configuring BGP on Multihomed Customer

In the figure, the customer has been assigned the private AS number 65001.



Configuring BGP on Multihomed Customer Routers (Cont.)

```
ip route 11.2.3.0 255.255.255.0 null 0
router bgp 65001
 neighbor 11.2.3.1 remote-as 65001
 neighbor 10.0.0.2 remote-as 387
 network 11.2.3.0 mask 255.255.255.0
router ospf 1
 default-information originate
```

- **The customer address space is advertised on every customer edge router.**
- **Customer edge routers run IBGP between themselves and advertise the default route to the rest of the customer network.**

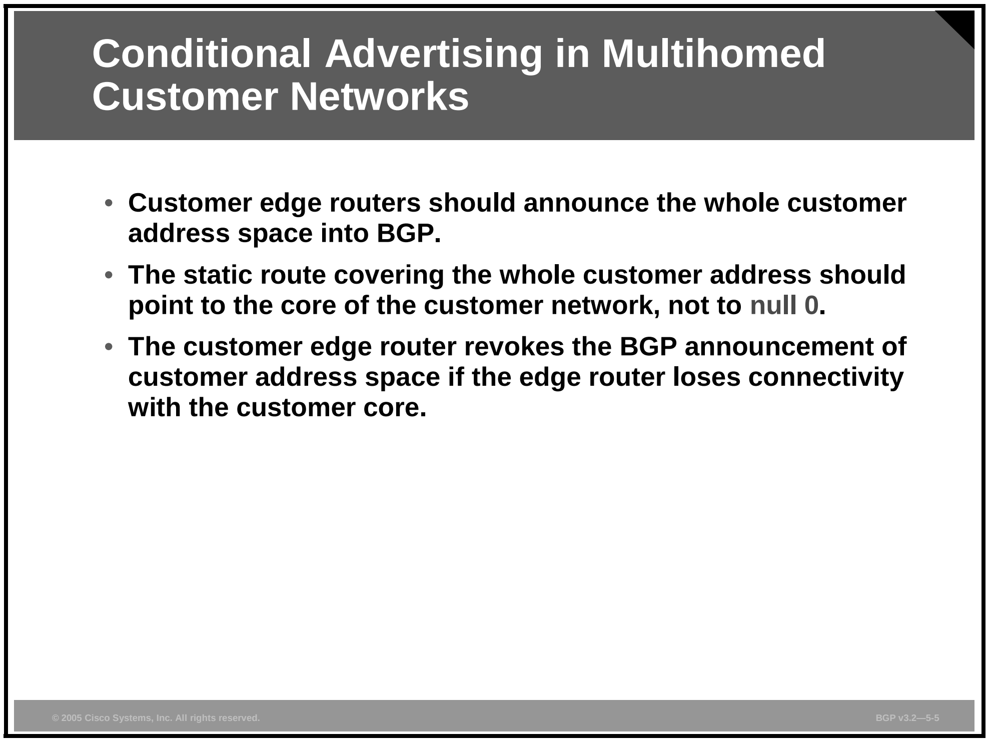
BGP v3.2—5-4

Both customer edge routers are configured to run BGP and should advertise all of the customer networks with the **network** command. If only one router advertises the network, a single point of failure has been introduced. The two customer edge routers must also run IBGP between them to make common decisions regarding BGP routing information.

Each customer edge router has an External Border Gateway Protocol (EBGP) session with the ISP router on the other side of the link. Over that EBGP session, the ISP announces only a default route to the customer AS. When EBGP receives the default route, it installs it in the routing table and redistributes it into the Interior Gateway Protocol (IGP)—in this case, Open Shortest Path First (OSPF)—of the customer.

Conditional Advertising in Multihomed Customer Networks

This topic describes how to configure conditional advertising of a customer address space when you are using BGP.



Conditional Advertising in Multihomed Customer Networks

- **Customer edge routers should announce the whole customer address space into BGP.**
- **The static route covering the whole customer address should point to the core of the customer network, not to null 0.**
- **The customer edge router revokes the BGP announcement of customer address space if the edge router loses connectivity with the customer core.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-5

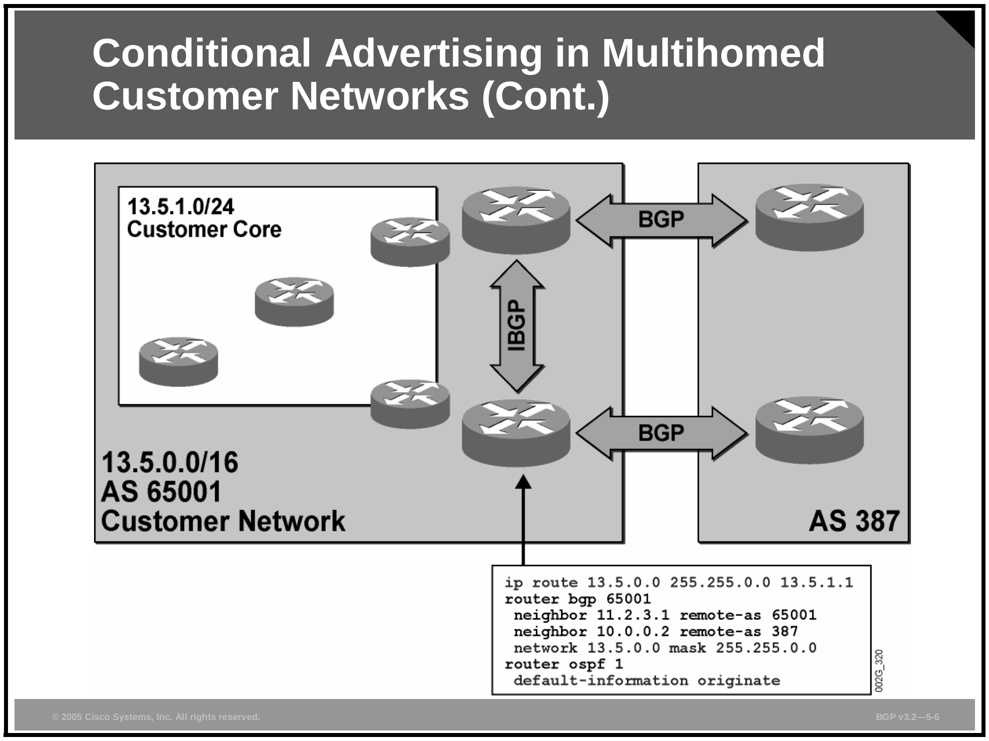
As a rule of thumb, the customer should announce addresses as large as possible (the larger the address space that can be aggregated, the better). The BGP advertisement is configured on the customer edge routers using the **network** command. Route advertisement is conditioned by the appearance of a corresponding network or subnet in the routing table of the edge router. If the network or subnet is manually entered into the routing table by configuring a static route to null 0, the condition is always true because the static route is always there, and the BGP advertisement is always performed.

If the customer edge router loses connectivity to the rest of the customer network but is still connected to the ISP network, BGP advertisement must cease. In this case, BGP advertisement can be stopped if BGP advertisements are bound to the reachability status of a specific subnet in the core of the customer network, according to the customer IGP.

The problem with using a static route to null 0 is that it conditions the network statement in the BGP configuration so that BGP always advertises the route. If the customer edge router loses connectivity with the rest of the customer network, the router continues to advertise the entire customer address space. The ISP network receives a valid route from the customer edge router. Traffic is sent to this router, but because the router has lost connectivity with the rest of the network, the traffic is dropped (routed to the null 0 interface using the static route).

Example: Conditional Advertising in Multihomed Customer Networks

In this example, the customer network uses the address space 13.5.0.0/16.



Conditional Advertising in Multihomed Customer Networks (Cont.)

```
ip route 13.5.0.0 255.255.0.0 13.5.1.1
router bgp 65001
 neighbor 11.2.3.1 remote-as 65001
 neighbor 10.0.0.2 remote-as 387
 network 13.5.0.0 mask 255.255.0.0
router ospf 1
 default-information originate
```

The address space is further subnetted at the customer site. One of the subnets (subnet 13.5.1.0/24) is identified as being a central part of the customer core network.

The customer edge routers participate in the IGP routing of the customer. This participation means that these routers have information about which of the subnets within the address space 13.5.0.0/16 are currently reachable. If these subnets are available, there is an explicit route to each of them. If any of the subnets go down, or if the path toward them goes down, the route to that subnet is removed from the routing table.

The BGP advertisement in each of the customer edge routers is configured to advertise the full address space that is used by the customer. When this route is advertised by the customer edge routers, the ISP network, and thus the rest of the Internet, see the complete address space of the customer as one single route, 13.5.0.0/16.

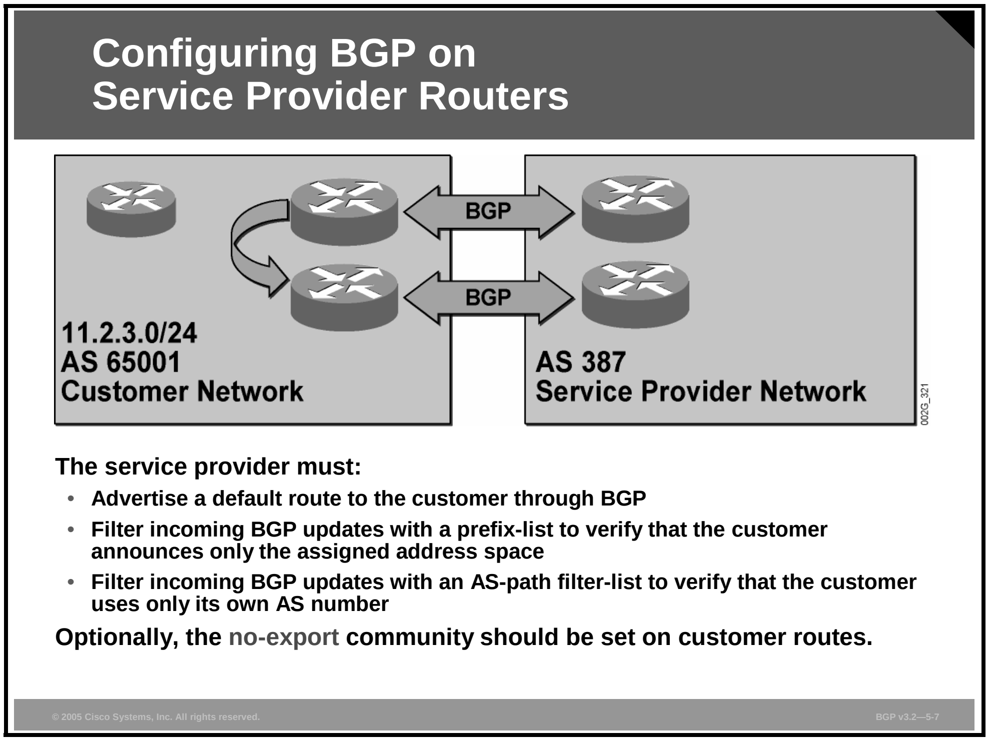
Advertisement of the customer address space by BGP is conditioned by the appearance of the static route, IP route 13.5.0.0 255.255.0.0 13.5.1.1. If the static route is valid, then the BGP route 13.5.0.0/16 is advertised. The static route is a recursive route, which means that the router takes another look in the routing table for the address 13.5.1.1 before determining what to do with the static route. The idea is that 13.5.1.1 is reachable via the IGP. The subnet 13.5.1.0/24 is announced by the IGP. If this subnet is reachable by the edge router, then the static route to 13.5.0.0/16 is valid. If there is no route to 13.5.1.1, then the static route is invalid.

Note The condition, whether or not to advertise the entire customer address space 13.5.0.0/16, is controlled by the IGP reachability of a single subnet, 13.5.1.0/24.

The IGP configuration also includes origination of the default route by both edge routers.

Configuring BGP on Service Provider Routers

This topic describes how to configure BGP on a service provider network to establish routing between a multihomed customer and a single service provider.



Configuring BGP on Service Provider Routers

The service provider must:

- **Advertise a default route to the customer through BGP**
- **Filter incoming BGP updates with a prefix-list to verify that the customer announces only the assigned address space**
- **Filter incoming BGP updates with an AS-path filter-list to verify that the customer uses only its own AS number**

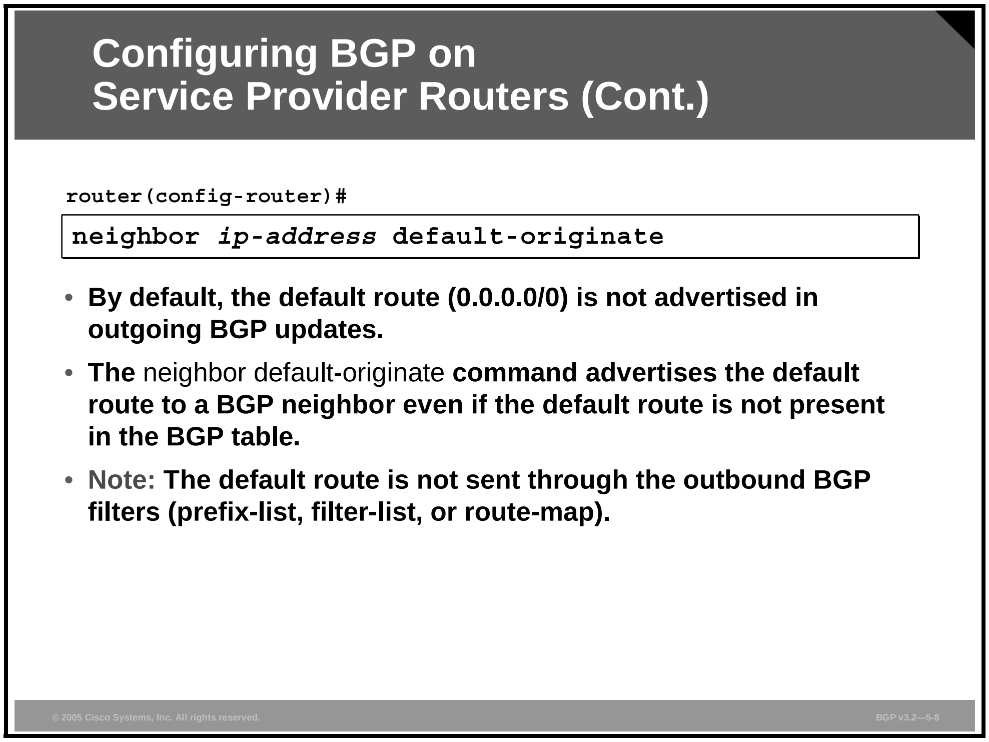
Optionally, the no-export community should be set on customer routes.

In the ISP network, the two edge routers must have BGP sessions configured for the customer. There is no point in feeding the full Internet routing table to the customer, because the table contains the same set of routes for both links and the customer always uses the ISP for all traffic toward the Internet. Injection of a default route in the customer network would accomplish the same task.

The customer is responsible for its own advertisements. Because customers are much less likely to be experienced in BGP configuration than the ISP, they are more likely to make errors. Therefore, the ISP must protect itself and the rest of the Internet from those errors.

The service provider should use a prefix-list that allows only customer-assigned routes and denies any other route to ensure that private address space or any other illegal networks that are erroneously announced by the customer never reach the ISP BGP table. Filtering based on the AS path also provides some protection from customer configuration errors. Only routes that originated within the customer AS are allowed. A filter-list performs this check.

If the customer address space is PA address space and it represents only a small part of a larger block that is announced by the ISP, the explicit BGP routes that are received from the customer need not be advertised to the rest of the Internet. The ISP can announce the big block, attracting any traffic toward any subnet within the block. After the traffic enters the ISP network, the more explicit routes to the customer network are available and used. In this case, the provider edge router can tag the BGP routes that are received from the customer with the no-export well-known community, restricting them from being sent by the ISP to any other AS.



Configuring BGP on Service Provider Routers (Cont.)

```
router(config-router)#
neighbor ip-address default-originate
```

- **By default, the default route (0.0.0.0/0) is not advertised in outgoing BGP updates.**
- **The** neighbor default-originate **command advertises the default route to a BGP neighbor even if the default route is not present in the BGP table.**
- Note: **The default route is not sent through the outbound BGP filters (prefix-list, filter-list, or route-map).**

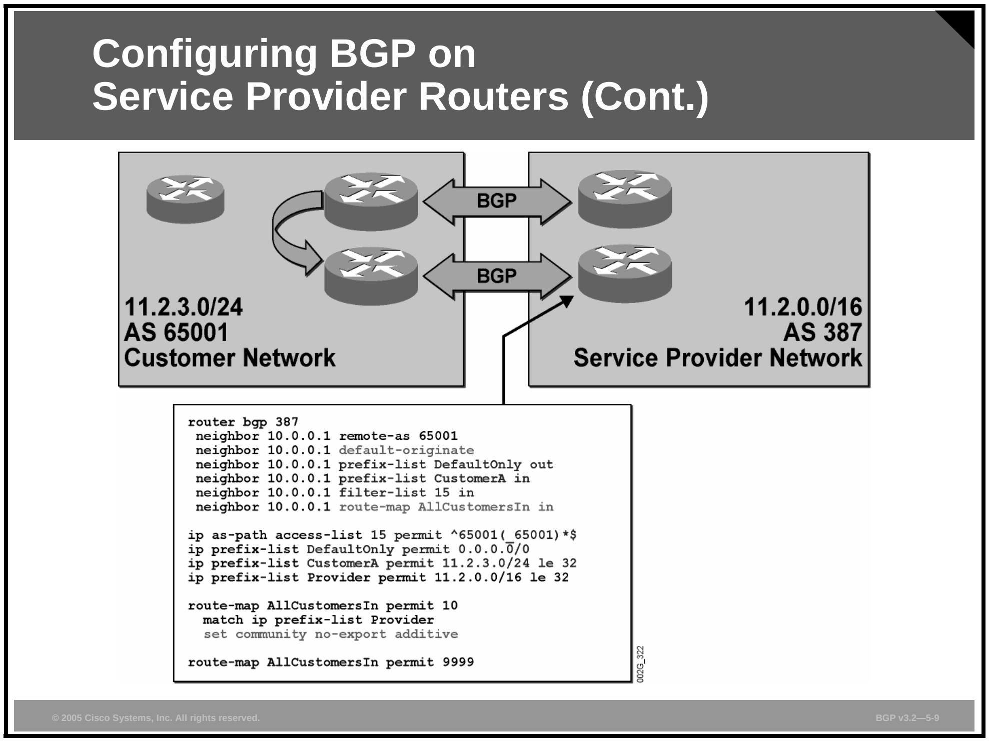
BGP v3.2—5-8

The default route, 0.0.0.0/0, is not advertised in outgoing BGP updates unless it is explicitly configured. The **neighbor default-originate** router configuration command is used to initiate the advertisement of the default route to a neighbor.

No checking is done by BGP before the default route is advertised. The default route does not need to be present in the BGP table before it is advertised using this command. The default route is also sent without being filtered by any outgoing prefix-lists, filter-lists, or route-maps.

Example: Configuring BGP on Service Provider Routers

This example shows the configuration of an ISP edge router.



Configuring BGP on Service Provider Routers (Cont.)

```
router bgp 387
 neighbor 10.0.0.1 remote-as 65001
 neighbor 10.0.0.1 default-originate
 neighbor 10.0.0.1 prefix-list DefaultOnly out
 neighbor 10.0.0.1 prefix-list CustomerA in
 neighbor 10.0.0.1 filter-list 15 in
 neighbor 10.0.0.1 route-map AllCustomersIn in

ip as-path access-list 15 permit ^65001(_65001)*$
ip prefix-list DefaultOnly permit 0.0.0.0/0
ip prefix-list CustomerA permit 11.2.3.0/24 le 32
ip prefix-list Provider permit 11.2.0.0/16 le 32

route-map AllCustomersIn permit 10
  match ip prefix-list Provider
  set community no-export additive

route-map AllCustomersIn permit 9999
```

The customer is assigned the private AS number 65001. The BGP session is opened with the customer IP address, 10.0.0.1.

The ISP sends the default route only to the customer. This route is configured using first the **default-originate** command and then the prefix-list DefaultOnly.

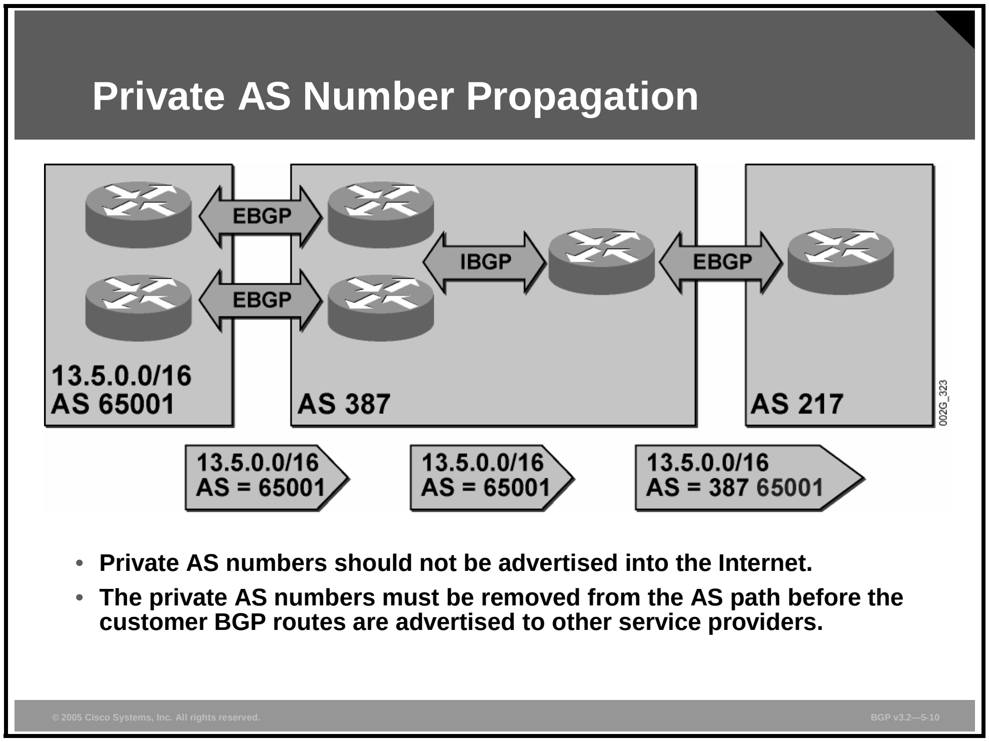
Received routes from the customer must first pass the prefix-list CustomerA. There is one dedicated prefix-list for each individual customer; the prefix-list permits only the routes that the customer is allowed to announce. If the routes are allowed by the prefix-list, they must also pass the filter-list named “15 in.” In this case, the filter allows the private AS of the customer in any number of repetitions, as long as it is the only AS number in the path. This filter-list allows for AS-path prepending configurations on the customer side. If the received route is allowed by both the prefix-list and the filter-list, then the route-map AllCustomersIn is applied.

The route-map is a general route-map that is used for all customers. It checks every route that is received, via the prefix-list Provider, and if the route is within the big block of PA address space that the ISP announces to the rest of the Internet, the customer route is marked with the no-export community. This mark means that the route is used within the ISP AS only and is not sent to the rest of the Internet.

Routes that are received from the customer and are allowed by the prefix-list and filter-list, but do not fall within the PA address space, are allowed by the route-map and are not changed in any way. The ISP propagates these routes to the rest of the Internet.

Removing Private AS Numbers

This topic describes how to disable the propagation of private AS numbers to EBGP peers in a service provider network in which a multihomed customer is advertising private numbers in the AS path.



Private AS Number Propagation

- Private AS numbers should not be advertised into the Internet.
- The private AS numbers must be removed from the AS path before the customer BGP routes are advertised to other service providers.

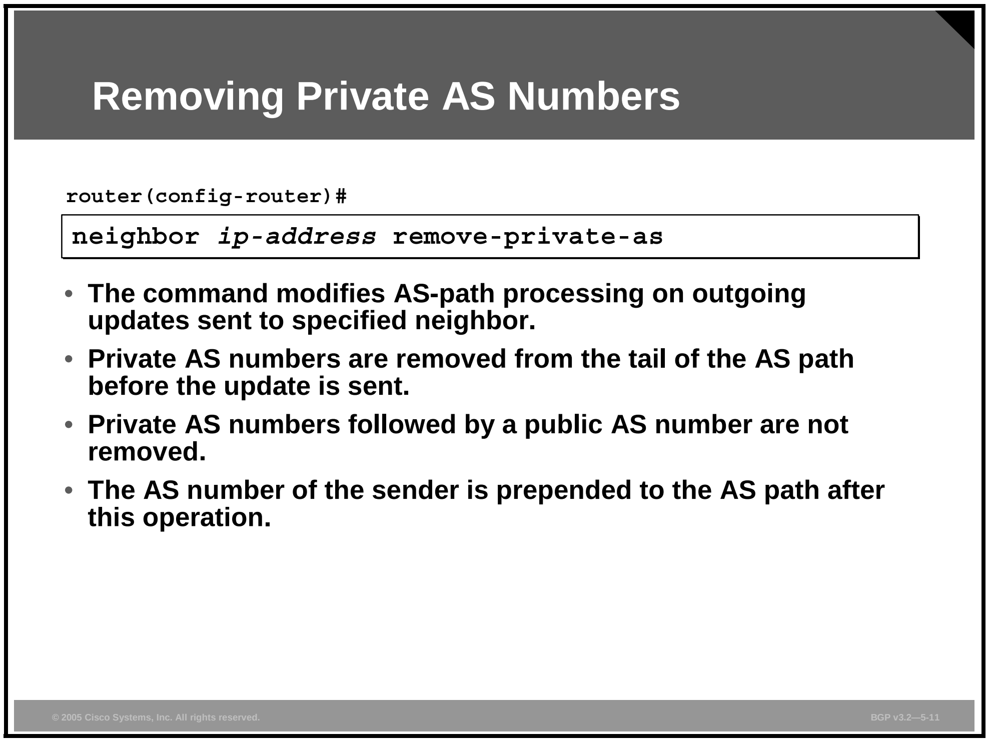
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-10

Routes that are received by the ISP from the customer are propagated to the rest of the Internet only if they are part of the provider-independent (PI) address space.

When the ISP receives BGP routes from the customer, the AS-path attribute of the received routes contains only the AS number of the customer. If the customer uses AS-path prepending, there may be several repetitions of the customer AS number in the AS path. If customer routes are propagated by the service provider to the Internet, the AS number of the customer will be present in the AS path unless it is explicitly removed.

Note If the customer has been assigned a private AS number, this AS number must never be advertised by any router to the rest of the Internet.

Removal of a private AS number from the AS path is accomplished by using the **remove-private-as** command on the ISP EBGP sessions with the rest of the Internet. In the figure, removal of the private AS number takes place on the EBGP session between AS 387 and AS 217.



Removing Private AS Numbers

```
router(config-router)#
neighbor ip-address remove-private-as
```

- The command modifies AS-path processing on outgoing updates sent to specified neighbor.
- Private AS numbers are removed from the tail of the AS path before the update is sent.
- Private AS numbers followed by a public AS number are not removed.
- The AS number of the sender is prepended to the AS path after this operation.

BGP v3.2—5-11

neighbor remove-private-as

To remove private AS numbers from the AS path (a list of AS numbers that a route passes through to reach a BGP peer) in outbound routing updates, use the **neighbor remove-private-as** router configuration command.

- **neighbor** {*ip-address* | *peer-group-name*} **remove-private-as**

To disable this function, use the **no** form of this command.

- **no neighbor** {*ip-address* | *peer-group-name*} **remove-private-as**

Syntax Description

Parameter	Description
ip-address	IP address of the BGP-speaking neighbor
peer-group-name	Name of a BGP peer group

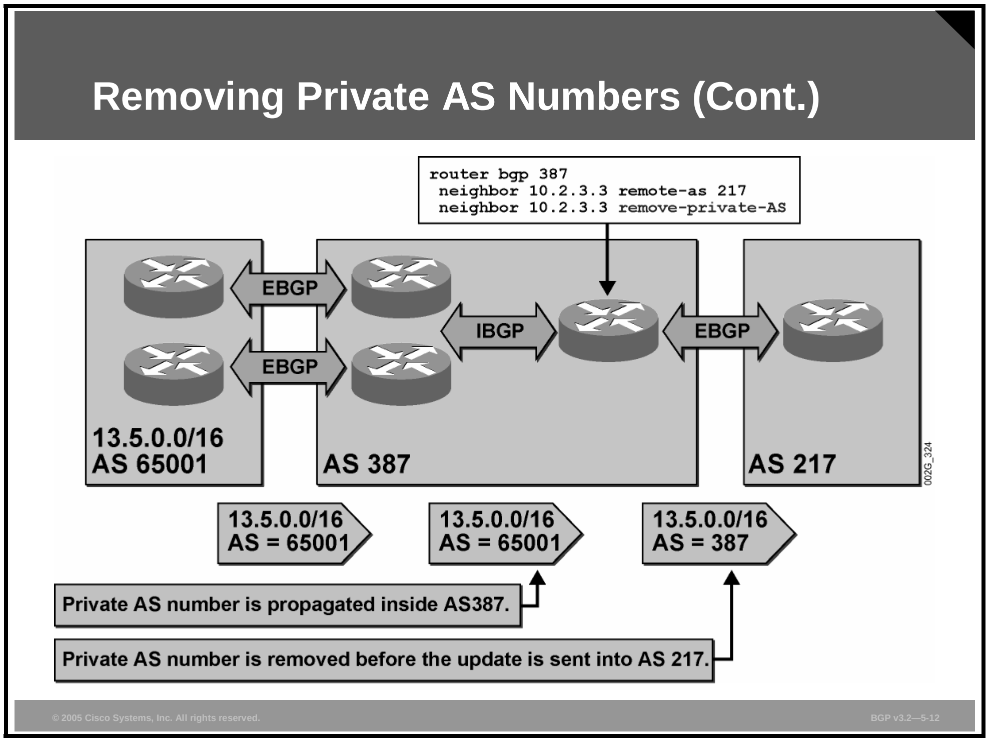
Use this command on the service provider egress routers. Before any of the customer routes of the ISP are advertised by the service provider to the rest of the Internet, the AS numbers in the range 64512 to 65535 must be removed. The command removes those AS numbers if they are at the tail end of the AS path.

Caution Private AS numbers followed by public AS numbers are not removed because the command's visibility is only on the last (tail end) AS number.

The AS number of the ISP is automatically prepended to the AS-path attribute after the **remove-private-as** operation has completed. This situation means that the AS number of the ISP has not already been prepended to the AS-path attribute when the tail of the AS path is checked for private AS numbers.

Example: Removing Private AS Numbers

In this example, the service provider AS (387) receives routes from the customer.



Removing Private AS Numbers (Cont.)

```
router bgp 387
 neighbor 10.2.3.3 remote-as 217
 neighbor 10.2.3.3 remove-private-AS
```

Private AS number is propagated inside AS387.

Private AS number is removed before the update is sent into AS 217.

BGP v3.2—5-12

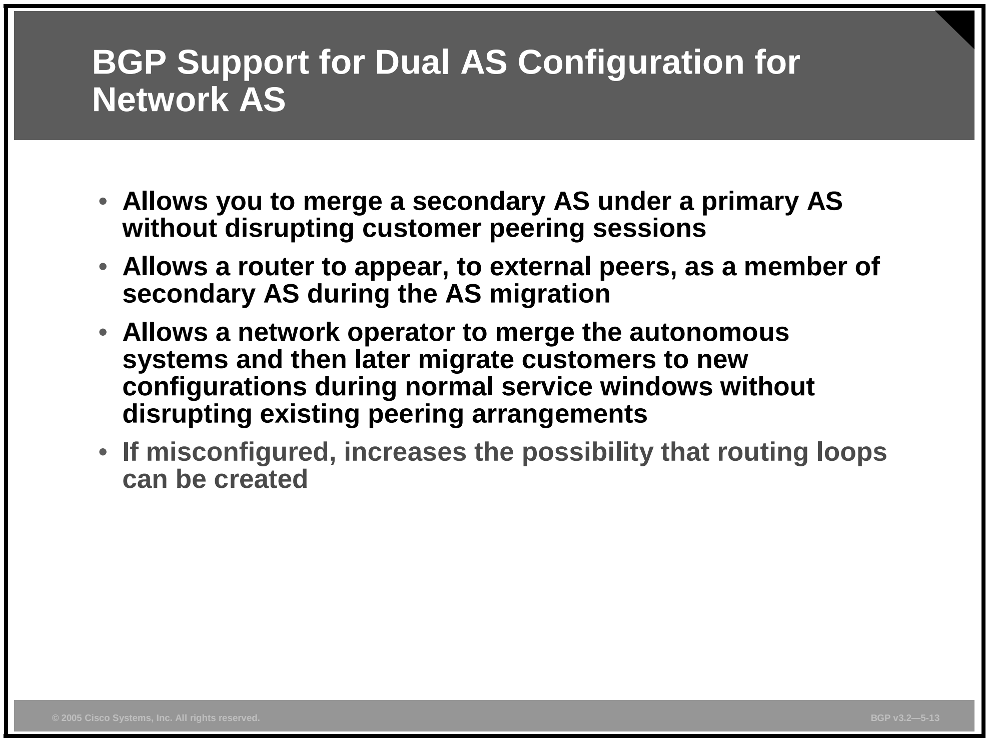
The customer is assigned the private AS number 65001 by the ISP; therefore, routes that are received by the provider have an AS path containing only AS 65001. This information should be kept and used within the ISP network and should never be propagated to the rest of the Internet (AS 217 in this example).

The edge router in AS 387 has been configured to remove private AS numbers on EBGP routes toward AS 217. If private AS numbers appear in the tail end of the AS path (before AS 387 is added), they are removed.

This configuration must be applied to all egress routers in AS 387 that serve EBGP neighbors leading to other ISPs. No private AS number may be present in an AS path of a route that is propagated to a network using a public AS number.

BGP Support for Dual AS Configuration for Network AS Migrations

This topic describes the BGP Support for Dual AS Configuration for Network AS Migrations feature.



BGP Support for Dual AS Configuration for Network AS

- **Allows you to merge a secondary AS under a primary AS without disrupting customer peering sessions**
- **Allows a router to appear, to external peers, as a member of secondary AS during the AS migration**
- **Allows a network operator to merge the autonomous systems and then later migrate customers to new configurations during normal service windows without disrupting existing peering arrangements**
- **If misconfigured, increases the possibility that routing loops can be created**

BGP v3.2—5-13

AS migration can be necessary when a telecommunications provider or ISP purchases another network. It is desirable for the provider to be able to integrate the second AS without disrupting existing customer peering arrangements. The amount of configuration required in the customer networks can make this a cumbersome task that is difficult to complete without disrupting service, however.

The BGP Support for Dual AS Configuration for Network AS Migrations feature allows you to merge a secondary AS under a primary AS without disrupting customer peering sessions. The configuration of this feature is transparent to customer networks. This feature allows a router to appear, to external peers, as a member of secondary AS during the AS migration. It also allows the network operator to merge the autonomous systems and then later migrate customers to new configurations during normal service windows without disrupting existing peering arrangements.

The **neighbor local-as** command is used to customize the AS-path attribute by adding and removing AS numbers for routes received from EBGP neighbors.

neighbor local-as

To customize the AS-path attribute for routes received from an EBGP neighbor, use the **neighbor local-as** command in address-family or router configuration mode.

- **neighbor** *ip-address* **local-as** [*as-number* [**no-prepend** [**replace-as** [**dual-as**]]]]

To disable AS-path attribute customization, use the **no** form of this command.

- **no neighbor** *ip-address* **local-as**

Syntax Description

Parameter	Description
ip-address	Specifies the IP address of the EBGP neighbor.
as-number	Specifies an AS number to prepend to the AS-path attribute. The range of values for this argument is any valid AS number from 1 to 65535.
no-prepend	(Optional) Does not prepend the local AS number to any routes received from the EBGP neighbor.
replace-as	(Optional) Prepends only the local AS number to the AS-path attribute. The AS number from the local BGP routing process is not prepended.
dual-as	(Optional) Configures the EBGP neighbor to establish a peering session using the real AS number (from the local BGP routing process) or by using the AS number configured with the **ip-address** argument (**local-as**)

Note AS-path customization increases the possibility that routing loops can be created if it is misconfigured. The larger the number of customer peerings, the greater the risk. You can minimize this possibility by applying policies on the ingress interfaces to block the AS number that is in transition or routes that have no **local-as** configuration.

Caution BGP prepends the AS number from each BGP network that a route traverses to maintain network reachability information and to prevent routing loops. This feature should be configured only for the AS migration and should be deconfigured after the transition has been completed. This procedure should be attempted only by an experienced network operator, because routing loops can be created with improper configuration.

Dual-AS Configuration: Example

The following examples show how this feature is used to merge two autonomous systems without interrupting peering arrangements with the customer network. The **neighbor local-as** command is configured to allow Router1 to maintain peering sessions through AS 100 and AS 200. Router2 is a customer router that runs a BGP routing process in AS 300 and is configured to peer with AS 200.

AS 100 (provider network):

```
Router1(config)# interface Serial3/0

Router1(config-int)# ip address 10.3.3.11 255.255.255.0

Router1(config-int)# !

Router1(config)# router bgp 100

Router1(config-router)# no synchronization

Router1(config-router)# bgp router-id 100.0.0.11

Router1(config-router)# neighbor 10.3.3.33 remote-as 300

Router1(config-router)# neighbor 10.3.3.33 local-as 200 no-prepend
replace-as dual-as
```

AS 200 (provider network):

```
Router1(config)# interface Serial3/0

Router1(config-int)# ip address 10.3.3.11 255.255.255.0

Router1(config-int)# !

Router1(config)# router bgp 200

Router1(config-router)# bgp router-id 100.0.0.11

Router1(config-router)# neighbor 10.3.3.33 remote-as 300
```

AS 300 (customer network):

```
Router2(config)# interface Serial3/0

Router2(config-int)# ip address 10.3.3.33 255.255.255.0

Router2(config-int)# !

Router2(config)# router bgp 300

Router2(config-router)# bgp router-id 100.0.0.3

Router2(config-router)# neighbor 10.3.3.11 remote-as 200
```

After the transition is complete, the configuration on Router3 can be updated to peer with AS 100 during a normal maintenance window or during other scheduled downtime.

```
Router2(config-router)# neighbor 10.3.3.11 remote-as 100
```

Dual-AS Confederation Configuration: Example

The following example can be used in place of the Router1 configuration in the previous example. The only difference between these configurations is that in this example Router1 is configured to be part of a confederation.

```
Router1(config)# interface Serial3/0

Router1(config-int)# ip address 10.3.3.11 255.255.255.0

Router1(config-int)# !

Router1(config)# router bgp 65534

Router1(config-router)# no synchronization

Router1(config-router)# bgp confederation identifier 100

Router1(config-router)# bgp router-id 100.0.0.11

Router1(config-router)# neighbor 10.3.3.33 remote-as 300

Router1(config-router)# neighbor 10.3.3.33 local-as 200 no-prepend
replace-as dual-as
```

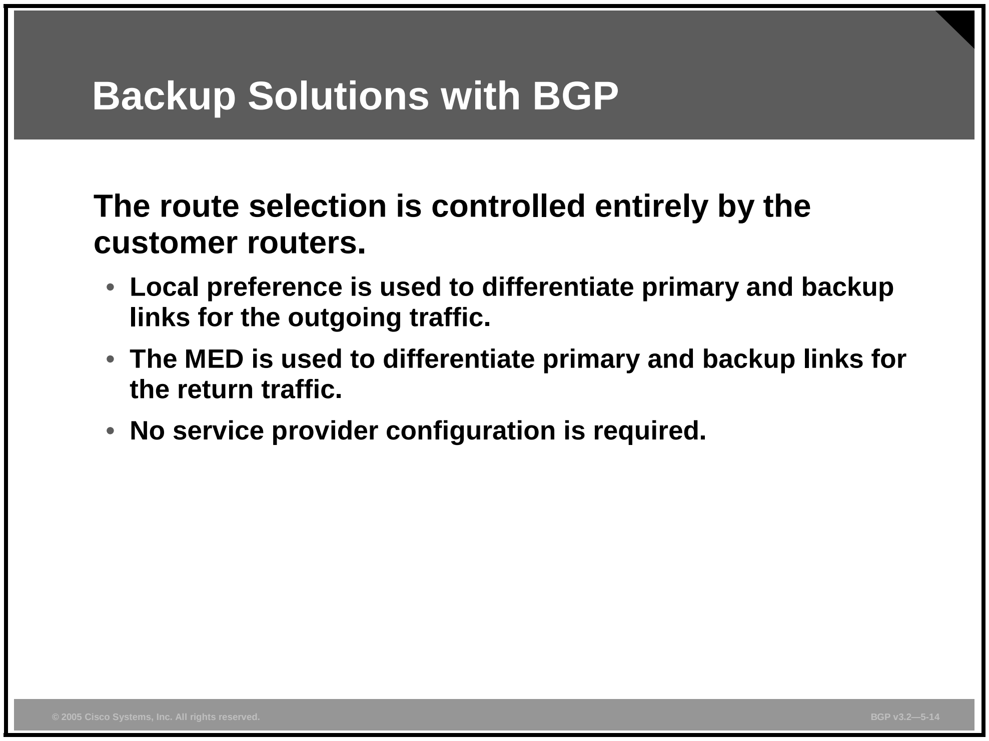
Replace-AS Configuration: Example

The following example strips private AS 64512 from outbound routing updates for the 10.3.3.33 neighbor and replaces it with AS 300:

```
Router(config)# router bgp 64512

Router(config-router)# neighbor 10.3.3.33 local-as 300 no-prepend
replace-as
```

Backup Solutions with BGP

This topic explains how to configure a typical backup setup between a multihomed customer and a service provider in a BGP environment.



Backup Solutions with BGP

The route selection is controlled entirely by the customer routers.

- **Local preference is used to differentiate primary and backup links for the outgoing traffic.**
- **The MED is used to differentiate primary and backup links for the return traffic.**
- **No service provider configuration is required.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-14

When a customer uses BGP on multiple links between its network and the ISP network, the customer is solely responsible for controlling how it uses the links. The customer can choose to use its links in a primary/backup scenario or in a load-sharing scenario.

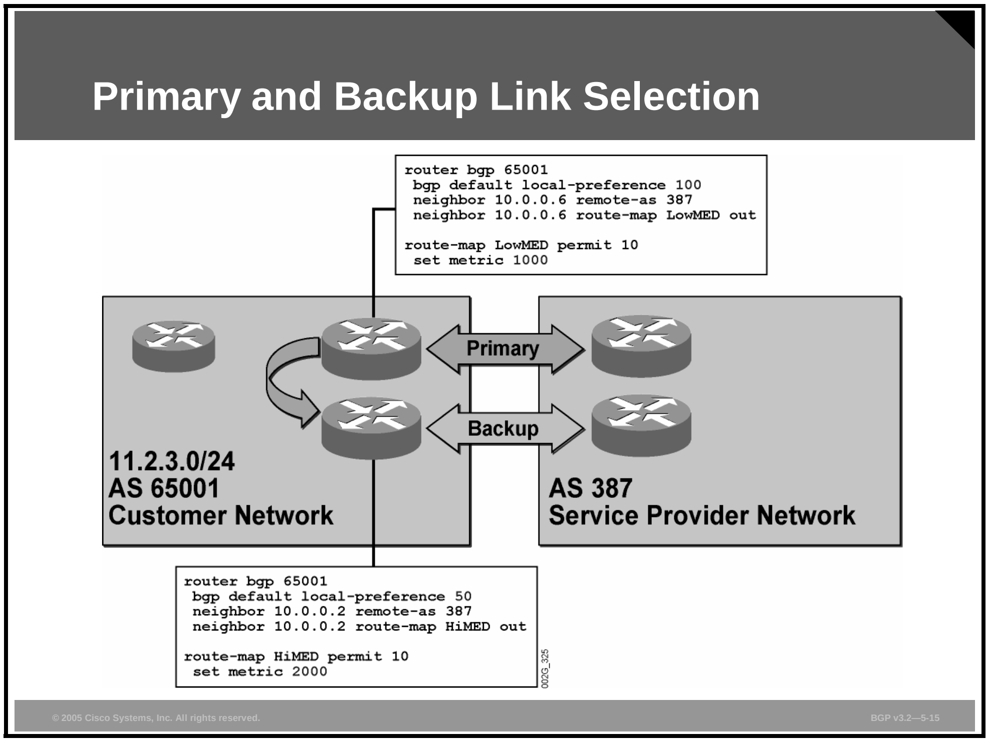
If one link is primary, then the other should be used for backup only. The customer can use the local preference configuration to direct all outgoing traffic over the primary link.

Incoming traffic to the customer is controlled by using either AS-path prepending or the multi-exit discriminator (MED). Because the customer has multiple connections to the same AS, the MED is the ideal attribute to use. When the customer announces its routes to the ISP, a bad (high) MED value on the backup link and a good (low) value on the primary link are set.

The MED and AS-path length are checked by the receiving EBGP peer only if the weight and local preference attributes have not been configured. In this case, the ISP should not use any of these configuration options. The ISP should rely solely on the attributes that it has received from the customer.

Example: Primary/Backup Link Selection

In the figure, the customer is connected to the ISP over two permanent connections.



Primary and Backup Link Selection

```
router bgp 65001
 bgp default local-preference 100
 neighbor 10.0.0.6 remote-as 387
 neighbor 10.0.0.6 route-map LowMED out

route-map LowMED permit 10
 set metric 1000
```

```
router bgp 65001
 bgp default local-preference 50
 neighbor 10.0.0.2 remote-as 387
 neighbor 10.0.0.2 route-map HiMED out

route-map HiMED permit 10
 set metric 2000
```

The customer uses the upper connection as the primary connection and the lower connection as the backup.

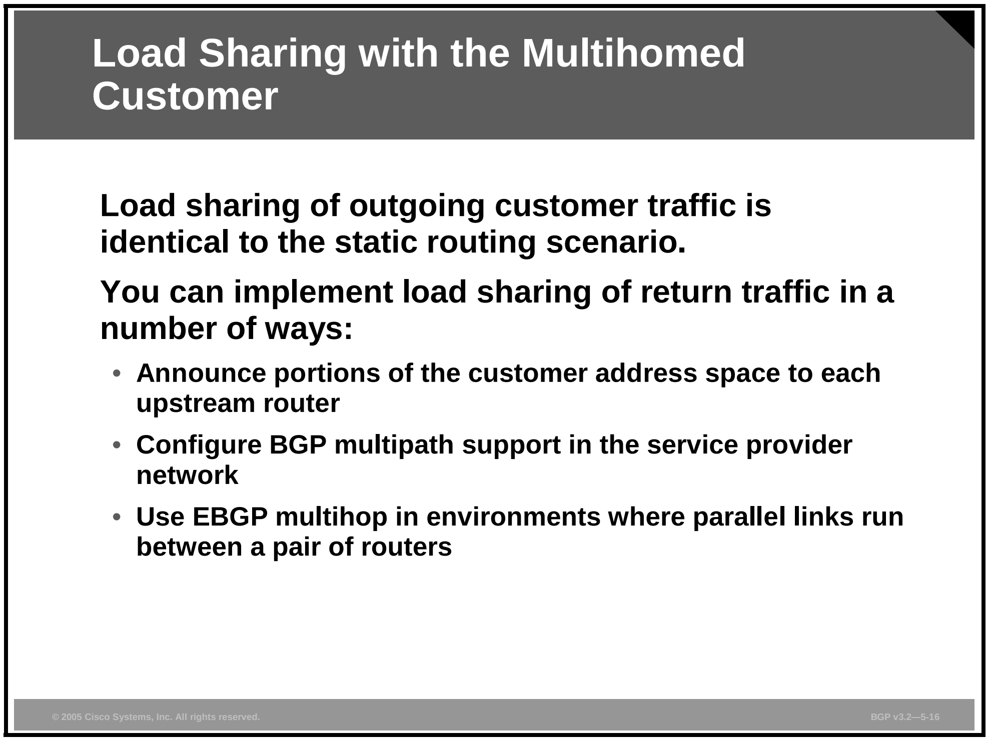
The BGP configuration on the ISP side is transparent. This transparency means that no particular preference is configured to use the upper or lower connection. The ISP relies on the attribute values that are received from the customer.

The primary edge router on the customer side is configured to set local preference to the value 100 on all EBGP routes that are received. The backup edge router sets the local preference attribute to a value of 50. This configuration means that the outgoing traffic toward any destination that is announced by the ISP is primarily sent over the upper link.

Incoming traffic to the customer is directed to the primary link by using the MED. In the primary edge router of the customer, all routes that are sent to the ISP have their MED attribute set to the value 1000 by the route-map named “LowMED out.” In the backup edge router of the customer, all routes that are sent to the ISP have their MED attribute set to the value 2000 by the route-map named “HiMED out.” Because the ISP receives the routes with all other attributes set to the same values, the MED values direct traffic for the customer to the primary link.

Load Sharing with the Multihomed Customer

This topic describes how you can implement load sharing between a multihomed customer and a service provider in a BGP environment.



Load Sharing with the Multihomed Customer

Load sharing of outgoing customer traffic is identical to the static routing scenario.

You can implement load sharing of return traffic in a number of ways:

- **Announce portions of the customer address space to each upstream router**
- **Configure BGP multipath support in the service provider network**
- **Use EBGP multihop in environments where parallel links run between a pair of routers**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-16

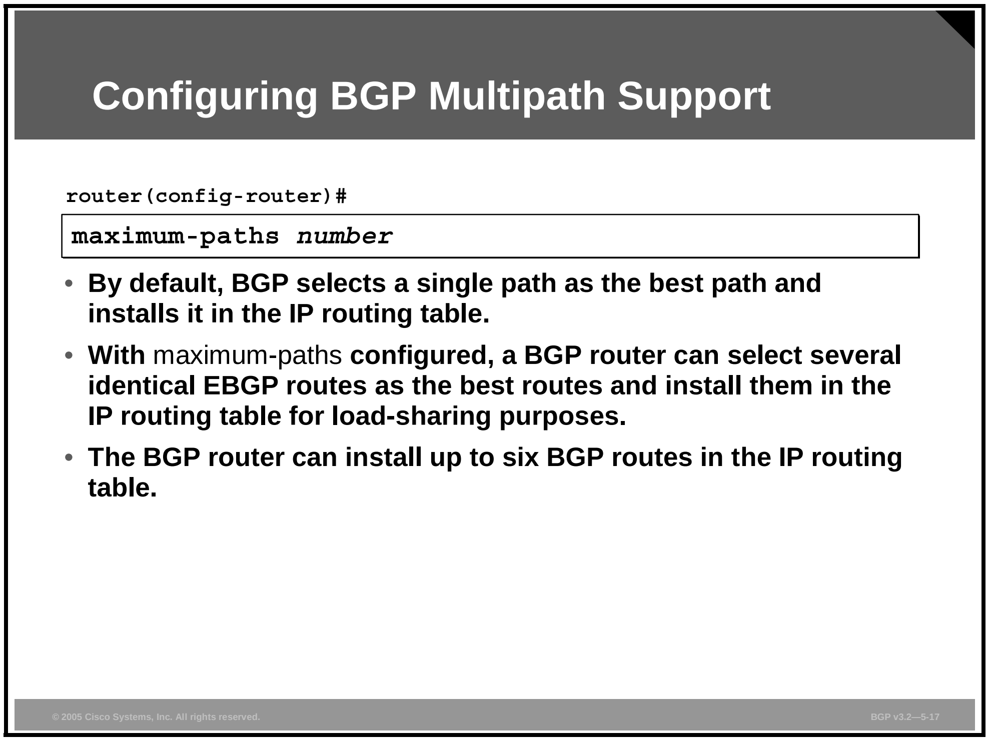
Load sharing of outgoing traffic from the customer network is identical to the static routing scenario. The customer IGP is configured to send information about a gateway of last resort. There is no difference whether the edge router gets its default by static routing or by incoming EBGP updates.

Load sharing of the return traffic coming back to the customer network from the ISP can be implemented in a number of ways:

- The customer can divide its address space into several announcements. The customer edge router can send each announcement over one of its EBGP sessions with the ISP. For backup purposes, the customer should advertise the entire address space over all of its EBGP sessions. The ISP now uses the most explicit route rule, and as long as both links are up, traffic with destinations within one part of the customer address space is routed over one of the links and traffic to the other part is routed over the other link.
- If the customer announces equivalent routes over both links, the ISP routers use the closest connection with respect to the IGP of the ISP. If an ISP router has an equivalent distance to both connection points, the use of the **maximum-paths** (BGP multipath) option causes load sharing.
- If the multiple links between the customer and the ISP network terminate in one single router on the customer side and one single router on the ISP side, the two routers must establish their EBGP session from loopback interface to loopback interface. Static or dynamic routing is required for one router to get information on how to reach the loopback interface of the other router. The use of the **ebgp-multihop** option is also required because the address of the neighbor is not directly connected.

Load Sharing with BGP Multipath

This topic presents the Cisco IOS command that is required to configure load sharing between a multihomed customer and a service provider through the use of BGP multipath.



Configuring BGP Multipath Support

```
router(config-router)#
maximum-paths number
```

- **By default, BGP selects a single path as the best path and installs it in the IP routing table.**
- **With** maximum-paths **configured, a BGP router can select several identical EBGP routes as the best routes and install them in the IP routing table for load-sharing purposes.**
- **The BGP router can install up to six BGP routes in the IP routing table.**

BGP v3.2—5-17

By default, BGP route selection rules select one, and only one, route as the best. If there are two identical routes, the tiebreaker is either the most stable route or the router-ID of the peer router that is advertising the route. However, when the **maximum-paths** router configuration command is used, the BGP route selection process will select more than one route as best if they are identical. The routes are all installed in the routing table, and load sharing takes place.

maximum-paths

To control the maximum number of parallel routes that an IP routing protocol can support, use the **maximum-paths** command in address family or router configuration mode.

- **maximum-paths** *number*

To restore the default value, use the **no** form of this command.

- **no maximum-paths**

Syntax Description

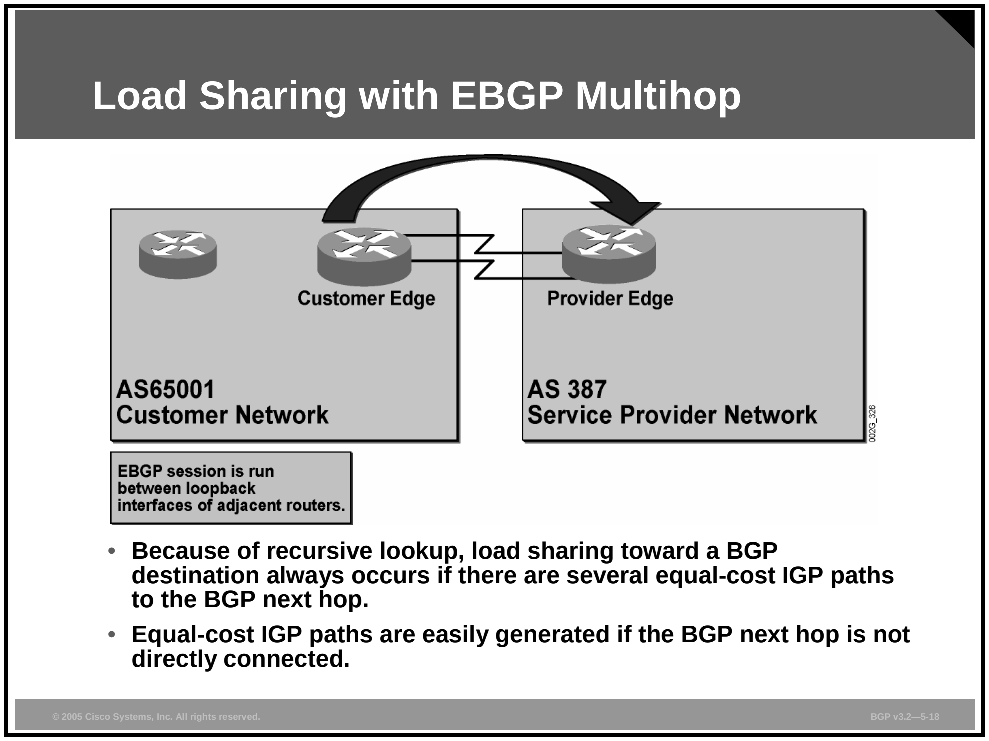
Parameter	Description
number	Maximum number of parallel routes that an IP routing protocol installs in a routing table, in the range of 1 to 6

Note Load sharing between alternative BGP routes is achieved only if the EBGP routes are identical according to all BGP route selection rules and **maximum-paths** is configured with a value larger than 1.

A BGP router can install up to six BGP routes in the IP routing table. The actual type of load sharing (per-session or per-packet) that occurs between the routes depends on the switching mode that is used.

Load Sharing with EBGP Multihop

This topic describes how to configure load sharing between a multihomed customer and a service provider through the use of EBGP multihop.



Load Sharing with EBGP Multihop

EBGP session is run between loopback interfaces of adjacent routers.

- **Because of recursive lookup, load sharing toward a BGP destination always occurs if there are several equal-cost IGP paths to the BGP next hop.**
- **Equal-cost IGP paths are easily generated if the BGP next hop is not directly connected.**

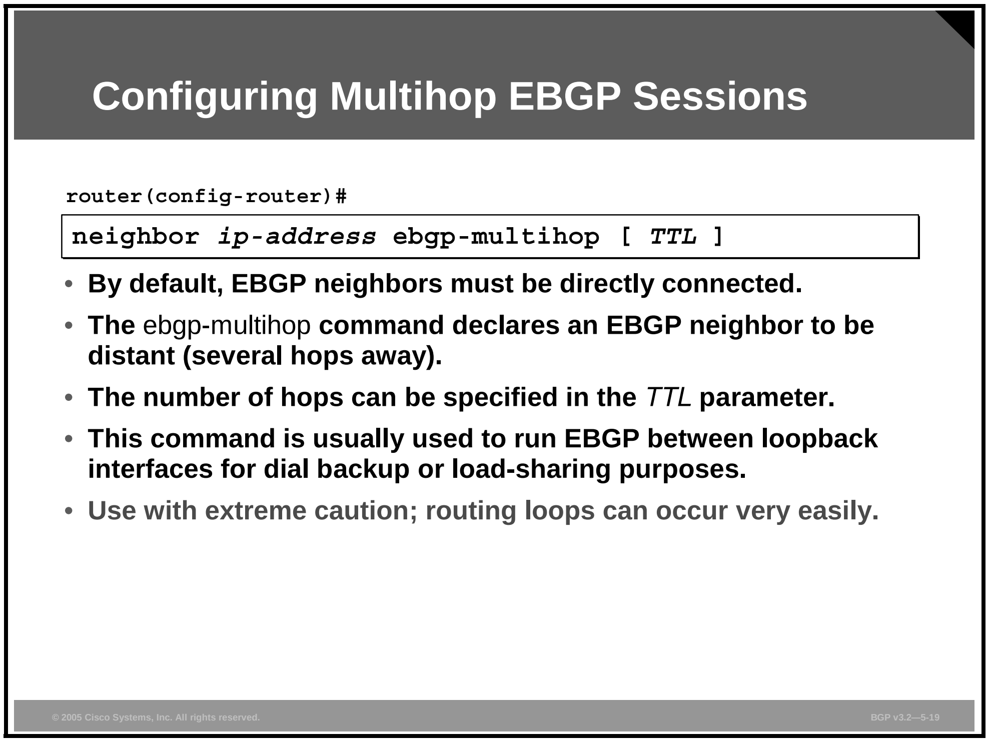
BGP v3.2—5-18

When two adjacent routers have multiple links between them, you can configure the EBGP session from loopback interface to loopback interface. In this case, you must use the **ebgp-multihop** option to make the BGP session go into the active state. There must be static or dynamic routing in use to provide both routers with information on how to reach the loopback interfaces of each other. Otherwise, their EBGP session does not complete establishment.

Routing to the loopback interface of the neighboring router is required to establish the EBGP session and is also used in the recursive lookup when the routes are installed by the router in its routing table. The two routes to the loopback interface of the neighboring router should be equivalent for load sharing to occur.

After configuration, one single EBGP session is established between the two routers. This session is used to exchange the routing information. There is only one BGP route to each destination, and it has a next hop that refers to the loopback interface of the other router.

Before installing a route to a specific destination in its routing table, a router will perform a recursive lookup to resolve the next hop. In this case, the recursive lookup will result in finding two alternative routes. The router will install the BGP route to the final destination twice in the routing table (Forwarding Information Base [FIB]). The first time, the route is installed with one of the resolved next-hop addresses, and the second time with the other resolved next-hop address. Because multiple equal-cost paths exist, the router can load-share over the two paths, depending on the switching mode.



Configuring Multihop EBGP Sessions

```
router(config-router)#
```

```
neighbor ip-address ebgp-multihop [ TTL ]
```

- **By default, EBGP neighbors must be directly connected.**
- **The** ebgp-multihop **command declares an EBGP neighbor to be distant (several hops away).**
- **The number of hops can be specified in the** *TTL* **parameter.**
- **This command is usually used to run EBGP between loopback interfaces for dial backup or load-sharing purposes.**
- **Use with extreme caution; routing loops can occur very easily.**

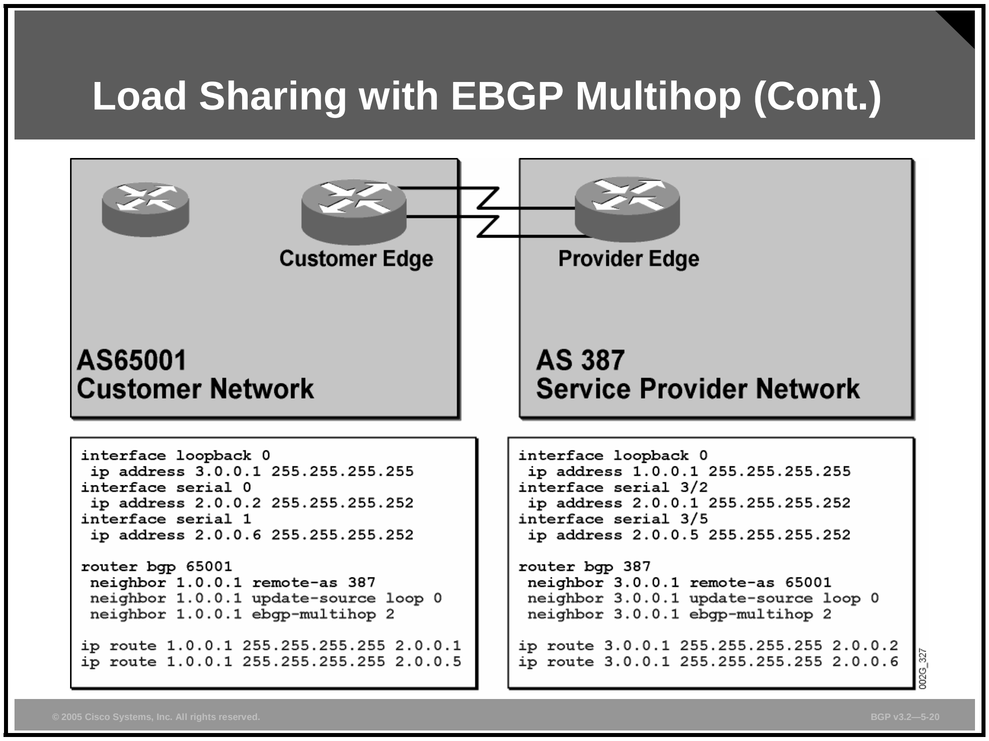
By default, EBGP neighbors must be directly connected. Cisco IOS software verifies that an EBGP neighbor is reachable as directly connected over one of the router interfaces before the session goes into the active state. For an EBGP session, IP packets that carry the TCP segments with BGP information are also sent using a Time to Live (TTL) value set to the value 1. This value means that they cannot be routed.

The **ebgp-multihop** neighbor configuration command changes this behavior. Although the neighbor is several hops away, the session goes into the Active state, and packets start to be exchanged. The TTL value of the IP packets is set to a value larger than 1. If no value is specified on the command line, 255 is used.

Use the **ebgp-multihop** command when you are establishing EBGP sessions between loopback interfaces for load-sharing purposes. You must take great care when using **ebgp-multihop**, because proper packet forwarding relies on all the intermediate routers along the path to the EBGP peer to make the correct forwarding decision. If the intermediate routers have a correct path to the EBGP peer but a wrong path to the final destination, the packet may get into a routing loop.

Example: Load Sharing with EBGP Multihop

In the figure, the customer network and the ISP network are connected using two parallel links between a single router on the customer side and a single router on the ISP side.



Load Sharing with EBGP Multihop (Cont.)

```
interface loopback 0
 ip address 3.0.0.1 255.255.255.255
interface serial 0
 ip address 2.0.0.2 255.255.255.252
interface serial 1
 ip address 2.0.0.6 255.255.255.252

router bgp 65001
 neighbor 1.0.0.1 remote-as 387
 neighbor 1.0.0.1 update-source loop 0
 neighbor 1.0.0.1 ebgp-multihop 2

ip route 1.0.0.1 255.255.255.255 2.0.0.1
ip route 1.0.0.1 255.255.255.255 2.0.0.5
```

```
interface loopback 0
 ip address 1.0.0.1 255.255.255.255
interface serial 3/2
 ip address 2.0.0.1 255.255.255.252
interface serial 3/5
 ip address 2.0.0.5 255.255.255.252

router bgp 387
 neighbor 3.0.0.1 remote-as 65001
 neighbor 3.0.0.1 update-source loop 0
 neighbor 3.0.0.1 ebgp-multihop 2

ip route 3.0.0.1 255.255.255.255 2.0.0.2
ip route 3.0.0.1 255.255.255.255 2.0.0.6
```

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-20

In this case, only one EBGP session is configured between the customer and provider routers. The session should be established from the loopback interface in one router to the loopback interface in the other.

Each of the two edge routers has two static host routes that point to the loopback interface on the other router. The EBGP session is established from loopback to loopback using **ebgp-multihop**.

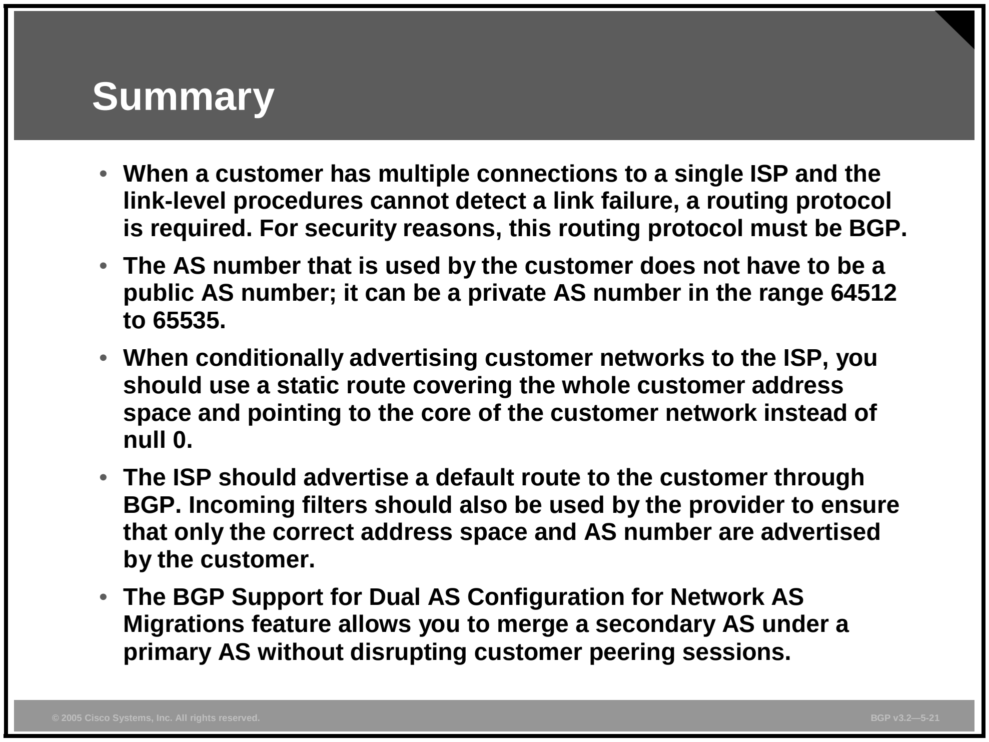
The customer receives an EBGP route from the ISP with the next hop set to 1.0.0.1. The customer edge router performs a recursive lookup and finds that it can reach 1.0.0.1 via 2.0.0.1 and via 2.0.0.5. These two routes are equivalent. Therefore, the route to the final destination is installed in the routing table of the customer router using both paths.

Depending on the switching mode in use, load sharing is done per packet, per destination, or per source and destination pair.

In this example, link-level procedures ensure that if one of the links goes down, the corresponding static link goes down. All BGP routes in the routing table that rely on the static route to the link that went down are invalidated. However, the BGP routes in the routing table that rely on the remaining link are still valid and used.

Summary

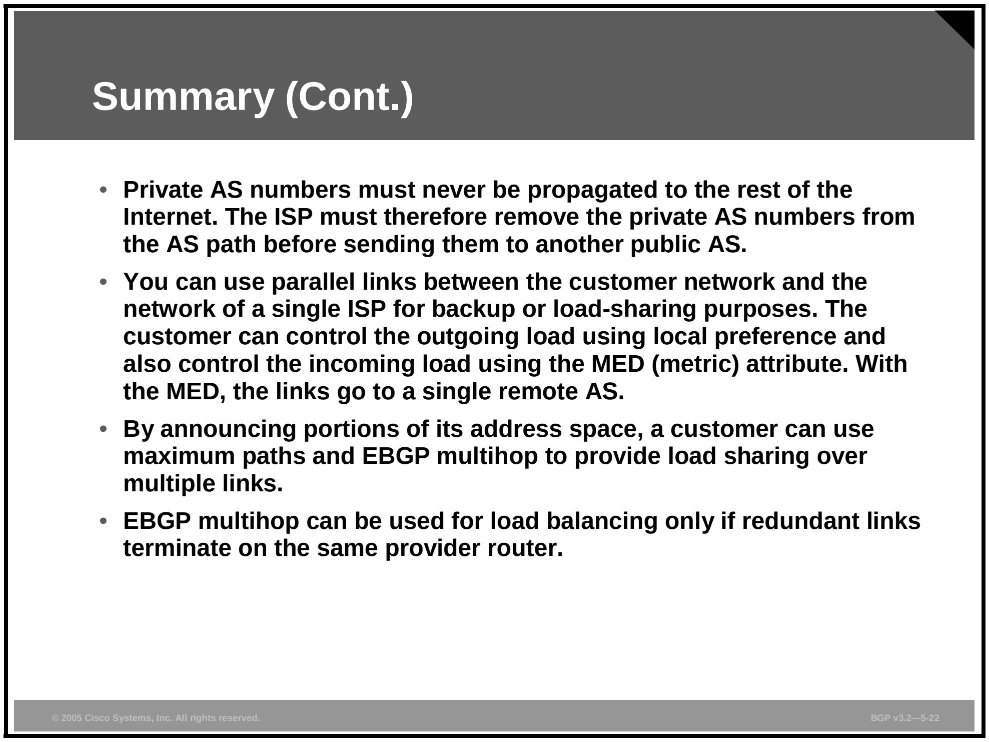
This topic summarizes the key points discussed in this lesson.



Summary

- When a customer has multiple connections to a single ISP and the link-level procedures cannot detect a link failure, a routing protocol is required. For security reasons, this routing protocol must be BGP.
- The AS number that is used by the customer does not have to be a public AS number; it can be a private AS number in the range 64512 to 65535.
- When conditionally advertising customer networks to the ISP, you should use a static route covering the whole customer address space and pointing to the core of the customer network instead of null 0.
- The ISP should advertise a default route to the customer through BGP. Incoming filters should also be used by the provider to ensure that only the correct address space and AS number are advertised by the customer.
- The BGP Support for Dual AS Configuration for Network AS Migrations feature allows you to merge a secondary AS under a primary AS without disrupting customer peering sessions.

BGP v3.2—5-21



Summary (Cont.)

- Private AS numbers must never be propagated to the rest of the Internet. The ISP must therefore remove the private AS numbers from the AS path before sending them to another public AS.
- You can use parallel links between the customer network and the network of a single ISP for backup or load-sharing purposes. The customer can control the outgoing load using local preference and also control the incoming load using the MED (metric) attribute. With the MED, the links go to a single remote AS.
- By announcing portions of its address space, a customer can use maximum paths and EBGP multihop to provide load sharing over multiple links.
- EBGP multihop can be used for load balancing only if redundant links terminate on the same provider router.

BGP v3.2—5-22

Lesson 4

Connecting a Multihomed Customer to Multiple Service Providers

Overview

When a customer requires the maximum redundancy in its network design, it should implement a multihomed strategy that uses multiple service providers. This configuration requires specific considerations to be implemented properly. Addressing and autonomous system (AS) number selection are important considerations that affect the implementation of the network. It is also important to understand how to configure routing protocols so that customer backup or load-sharing requirements are met.

This lesson discusses using multiple connections between a customer and multiple service providers for backup and load-sharing purposes. Also included in this lesson is a discussion of the Border Gateway Protocol (BGP) characteristics that are used to configure customer and provider networks to accommodate the multiple connections between them. This lesson also discusses topics specific to networks with multiple connections between a customer and multiple providers such as address selection, private AS number translation, and configuration of the network to support either backup links or load sharing.

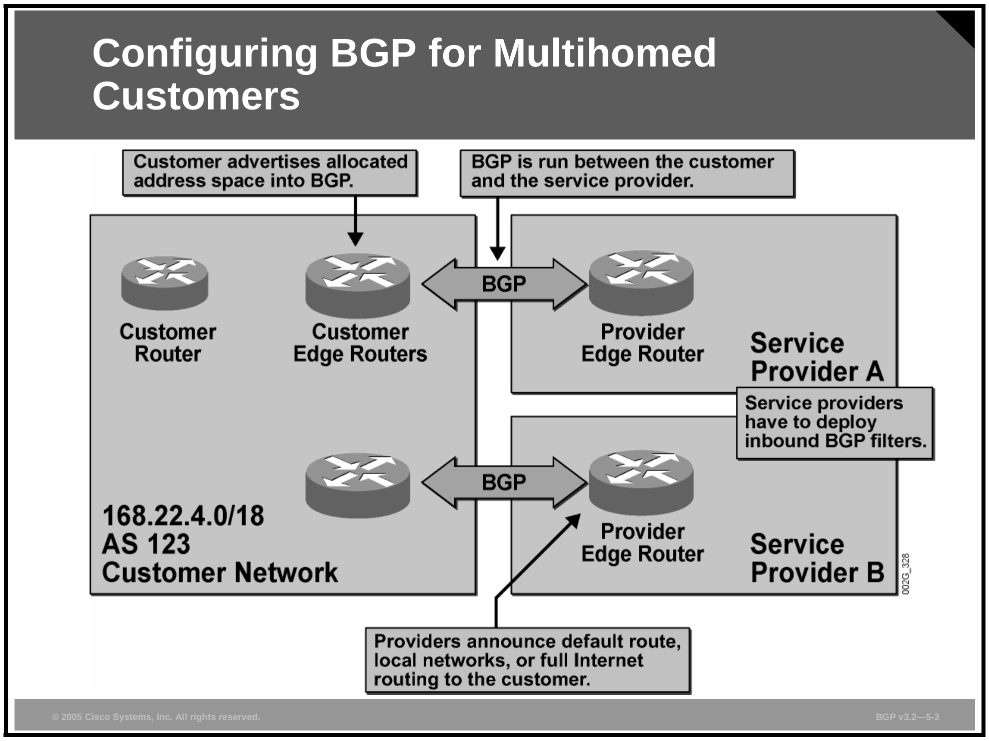
Objectives

Upon completing this lesson, you will be able to implement customer connectivity using BGP in a customer scenario where you must support connections to multiple ISPs. This ability includes being able to meet these objectives:

- Describe BGP configuration characteristics that are used to establish routing between a multihomed customer and multiple service providers
- Describe addressing strategies that are available to a multihomed customer that is connected to multiple service providers
- Describe AS numbering strategies that are available to a multihomed customer that is connected to multiple service providers
- Describe the operation of AS number translation
- Describe how you can implement a typical backup setup between a multihomed customer and multiple service providers in a BGP environment
- Describe the use of BGP attributes to influence inbound link selection in customer networks that are multihomed to multiple service providers
- Describe how you can implement load sharing between a multihomed customer and multiple service providers in a BGP environment

Configuring BGP for Multihomed Customers

This topic describes the different characteristics of a BGP configuration that is used to establish routing between a multihomed customer and multiple service providers.



Configuring BGP for Multihomed Customers

Customer advertises allocated address space into BGP.

BGP is run between the customer and the service provider.

Service providers have to deploy inbound BGP filters.

Providers announce default route, local networks, or full Internet routing to the customer.

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—5-3

The highest level of resilience to network failures is achieved in network designs that connect the customer network to two different service provider networks. Customers use this option when the requirement for resilient Internet connectivity is very high. This requirement also involves duplication of equipment to make the customer network fully redundant.

BGP must be used between the customer and both service providers, because static routing will not work in this type of network. It is not enough to detect link failures or a failure in the remote router by link-level procedures. Failures that occur beyond the directly connected router must also be detected, and the only means of detecting these failures is by using a routing protocol. The only routing protocol that is suited for the Internet environment is BGP. Correctly configured, BGP takes care of rerouting in the following situations:

- Link failure between the customer network and the network of one of the Internet service providers (ISPs).
- Edge router failure on either the customer or the ISP side
- Link failure or router failure within the customer network that causes the customer edge router to lose connectivity with the customer network core. This situation requires correct configuration of route advertisement as described in an earlier lesson.
- Link failure or router failure within the ISP network that causes the ISP edge router to lose connectivity with the rest of the Internet

Multihomed customers have multiple permanent links to different ISPs. The links should terminate in different edge routers in the customer network. Otherwise, one of the major advantages, resilience to router failure, is lost.

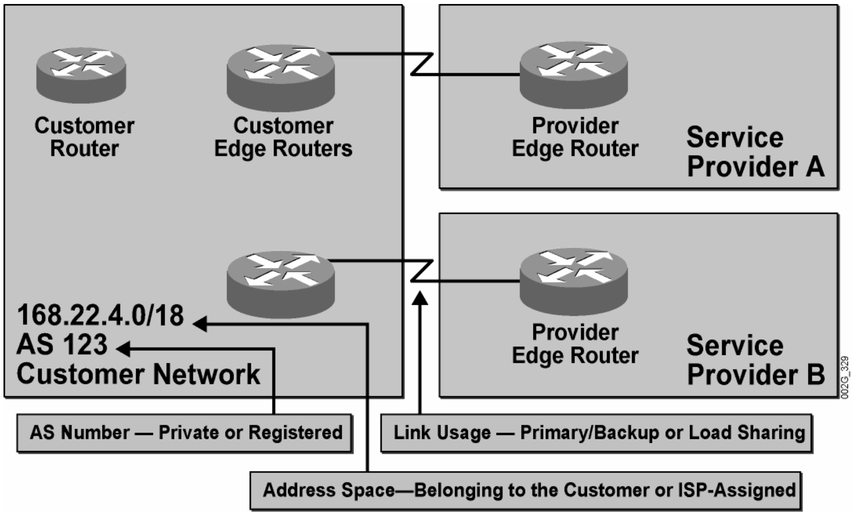
Multihomed customers should use BGP with both ISPs. The customer should advertise its address space to both providers. Route advertisement should be configured in both customer edge routers. The advertisement should be conditioned at the edge routers by the appropriate route policies leading toward the core of the customer network. This setup is analogous to that configured when you are connecting a multihomed customer network to a single provider.

The customer should take care not to move any routing information between the two ISPs. It must use outgoing filters to prevent any route that is received from one of the ISPs from being propagated to the other. Otherwise, the customer network appears as a transit network between the two ISPs.

Both ISPs must apply filters on the incoming BGP information from the customer to protect themselves and the rest of the Internet from errors in the BGP configuration of the customer. Each of the service providers must accept routes from the customer that indicate networks within the customer address space only. AS-path filter-lists should be implemented on the provider edge routers to allow incoming routes only if they have the correct AS-path attribute value. If the incoming filters on the ISP edge router accept customer routes, then the service provider should propagate those routes to the rest of the Internet.

Both ISPs must provide the customer with at least some BGP routes. Depending on customer requirements, the volume of BGP routes that are provided by the ISP could range anywhere from the default route only to the full Internet routing table.

Configuring BGP for Multihomed Customers (Cont.)



AS Number — Private or Registered

Link Usage — Primary/Backup or Load Sharing

Address Space—Belonging to the Customer or ISP-Assigned

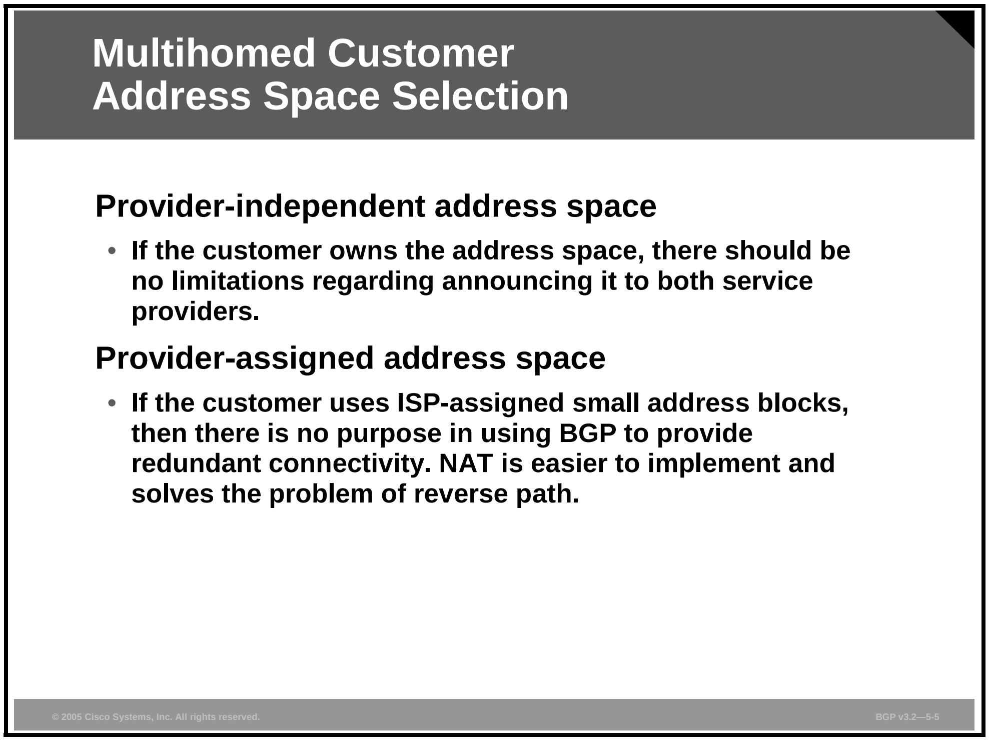
BGP v3.2—5-4

Before configuring the multihomed network, you need to consider the following questions:

- Should any of the links be used as primary and the others as backup?
- Should both links share the load?
- What address space is the customer using? Is the customer address space provider-assigned (PA) or provider-independent (PI)?
- What AS number is the customer using? (Is the customer using a public or a private AS number?)

Multihomed Customer Address Space Selection

This topic describes the various addressing strategies that are available to a multihomed customer that is connected to multiple service providers.



Multihomed Customer Address Space Selection

Provider-independent address space

- **If the customer owns the address space, there should be no limitations regarding announcing it to both service providers.**

Provider-assigned address space

- **If the customer uses ISP-assigned small address blocks, then there is no purpose in using BGP to provide redundant connectivity. NAT is easier to implement and solves the problem of reverse path.**

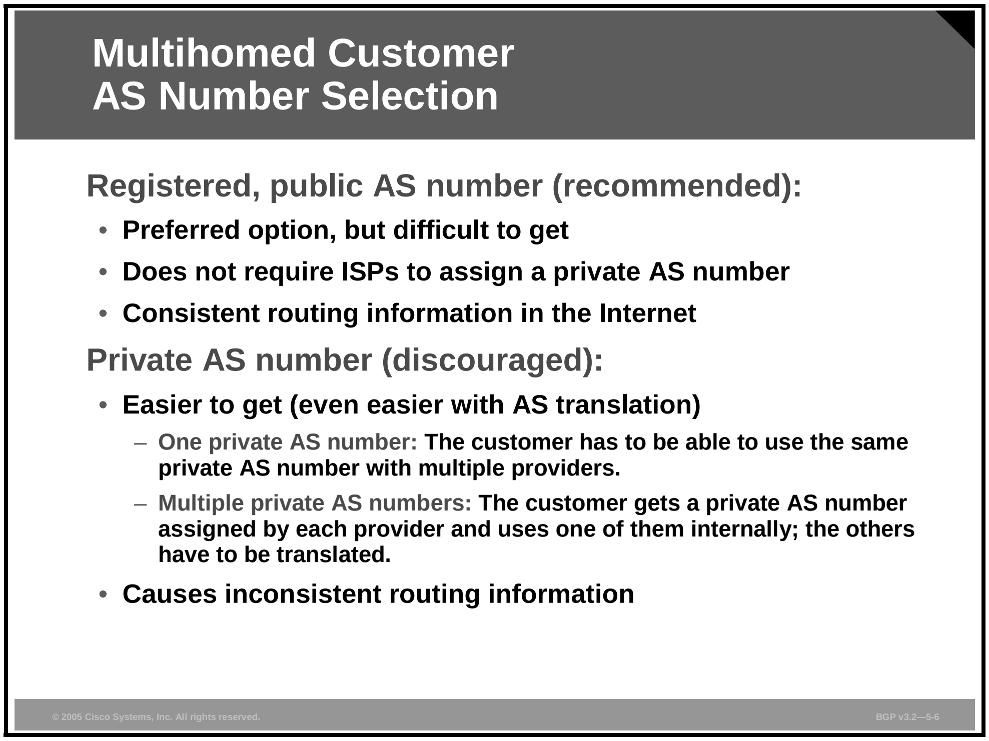
BGP v3.2—5-5

If the customer has its own address space, it should announce it to both service providers. Both providers are responsible for propagating the customer routes to the rest of the Internet without doing any summarization.

However, if the customer uses a small block of addresses that is assigned by one of the ISPs, an alternative design, not involving BGP, is to use two different PA address spaces and do Network Address Translation (NAT). With NAT, the router translates traffic going out over one of its connections to one of the PA addresses. If traffic goes out the other way, the addresses are translated to an address from the address space of the other provider.

Multihomed Customer AS Number Selection

This topic describes the various AS numbering strategies that are available to a multihomed customer that is connected to multiple service providers.



Multihomed Customer AS Number Selection

Registered, public AS number (recommended):

- **Preferred option, but difficult to get**
- **Does not require ISPs to assign a private AS number**
- **Consistent routing information in the Internet**

Private AS number (discouraged):

- **Easier to get (even easier with AS translation)**
  - **One private AS number: The customer has to be able to use the same private AS number with multiple providers.**
  - **Multiple private AS numbers: The customer gets a private AS number assigned by each provider and uses one of them internally; the others have to be translated.**
- **Causes inconsistent routing information**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-6

The use of BGP requires an AS number. The preferred option is to use a registered, public AS number. However, registered AS numbers are assigned only to those who really need them because public AS numbers are a scarce resource. A customer with BGP sessions to multiple ISPs must use a registered, public AS number. A customer that is connected to only one ISP does not require a public AS number. In that case, a private AS number in the range 64512 to 65535 is sufficient.

Whenever the customer has a public AS number assigned, there are no conflicts in the BGP setup, because the number is guaranteed to be unique within the Internet. Route announcements are made by both the customer and service provider without tampering with the AS path. As a result, consistent AS-path information is propagated by the service provider to the rest of the Internet.

In cases where the customer does not have a public AS number, it must use a private AS number. Because private AS numbers are not propagated to the Internet, several network administrators can, independently of each other, make this assignment. In this case, AS numbers are reused, which conserves AS number space. A service provider normally assigns private AS numbers to its customers. This arrangement ensures that unique private AS numbers are used among the customers of a single ISP.

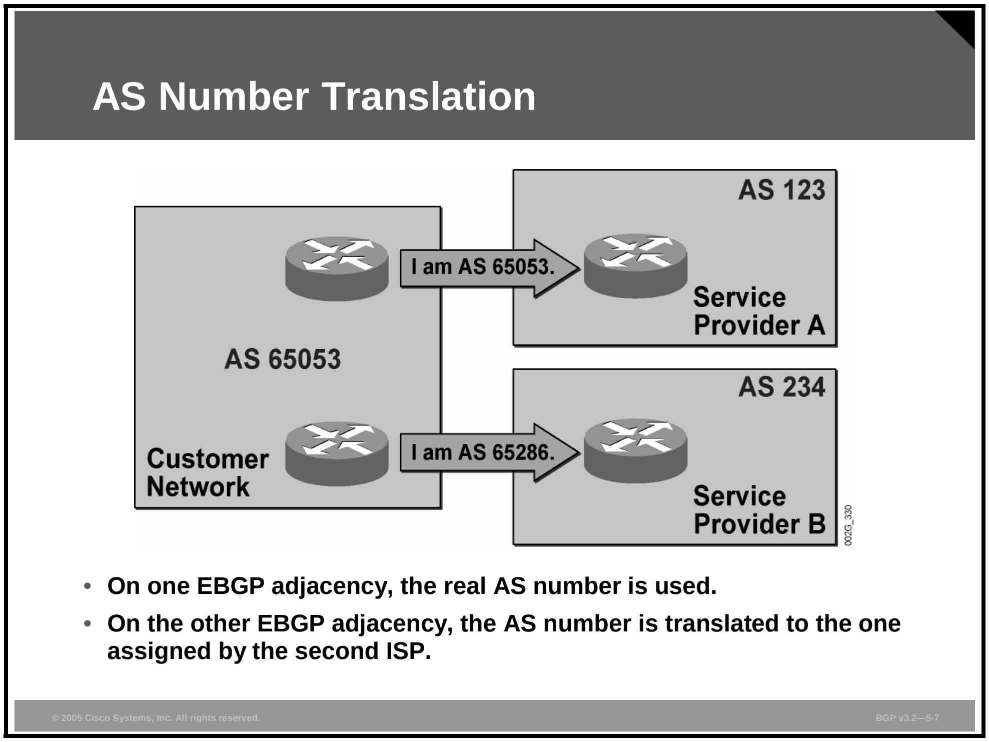
When a customer is going to be multihomed and the private AS number that has already been assigned by one of the ISPs comes in conflict with an AS number that has been assigned by the other ISP, the customer needs to consider renumbering the customer AS. If the two service providers can reach a common agreement on which private AS number that the multihomed customer should use, renumbering is a solution. If no common agreement can be made or if renumbering, for some reason, is not an option, AS translation must be configured on the customer network.

No router should ever propagate private AS numbers to the rest of the Internet. An ISP can keep track of which private AS numbers that it has assigned to its customers and avoid reuse or conflicts within that scope. However, as soon as the scope is widened to include other ISPs, conflicts will happen. Each ISP, therefore, removes private AS numbers from the AS path before sending routes outside its own AS.

When the routes with the private AS numbers removed are propagated to the rest of the Internet, the AS path looks as though the routes were originated within the public AS of the ISP. All information about the private AS lying behind the public AS is lost. In the case of a multihomed customer, the customer routes are, in the first step, propagated into each of the autonomous systems of its ISPs. In the next step, the routes have the private AS number removed as the routes are propagated to the rest of the Internet. Now the customer routes appear to be originating in the autonomous systems of both ISPs. To an outside observer, there is now an AS-path inconsistency because it looks as though the same route belongs to different autonomous systems.

AS Number Translation

This topic describes the operation of AS number translation.



AS Number Translation

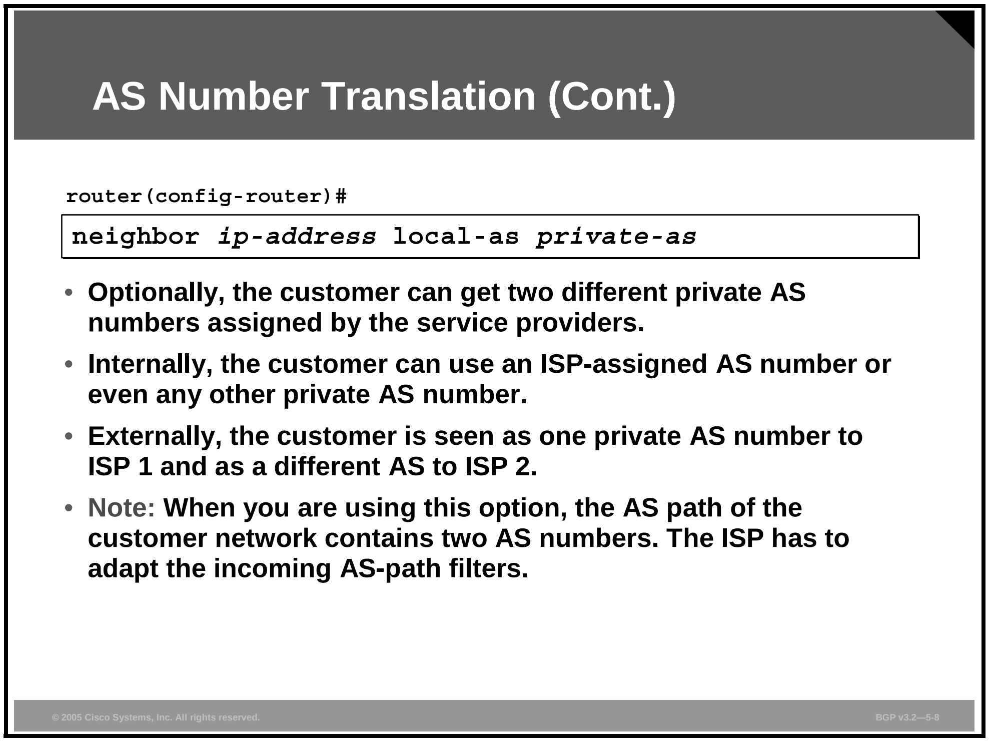
- On one EBGP adjacency, the real AS number is used.
- On the other EBGP adjacency, the AS number is translated to the one assigned by the second ISP.

The figure shows a case where a customer is multihomed but forced to use two private AS numbers (for example, because of the scarcity of public AS numbers).

In the figure, service provider A has assigned the private AS number 65053 to the customer. Service provider B did not agree to use this private AS number when connecting to the customer. Instead, service provider B has assigned the private AS number 65286.

The customer now has two different private AS numbers: 65053 and 65286. The customer decides to use 65053 internally. All router BGP configuration lines have 65053 as the AS number. The customer uses AS number 65286 only when establishing the External Border Gateway Protocol (EBGP) session to AS 234.

In the example, service provider A (AS 123) has an EBGP session to the customer where the AS number 65053 is used at the customer end. Service provider B (AS 234) has an EBGP session to the customer where the AS number 65286 is used at the customer end. Translation between these two private AS numbers takes place in the customer edge router as part of the EBGP session to AS 234.



AS Number Translation (Cont.)

```
router(config-router)#
neighbor ip-address local-as private-as
```

- Optionally, the customer can get two different private AS numbers assigned by the service providers.
- Internally, the customer can use an ISP-assigned AS number or even any other private AS number.
- Externally, the customer is seen as one private AS number to ISP 1 and as a different AS to ISP 2.
- Note: When you are using this option, the AS path of the customer network contains two AS numbers. The ISP has to adapt the incoming AS-path filters.

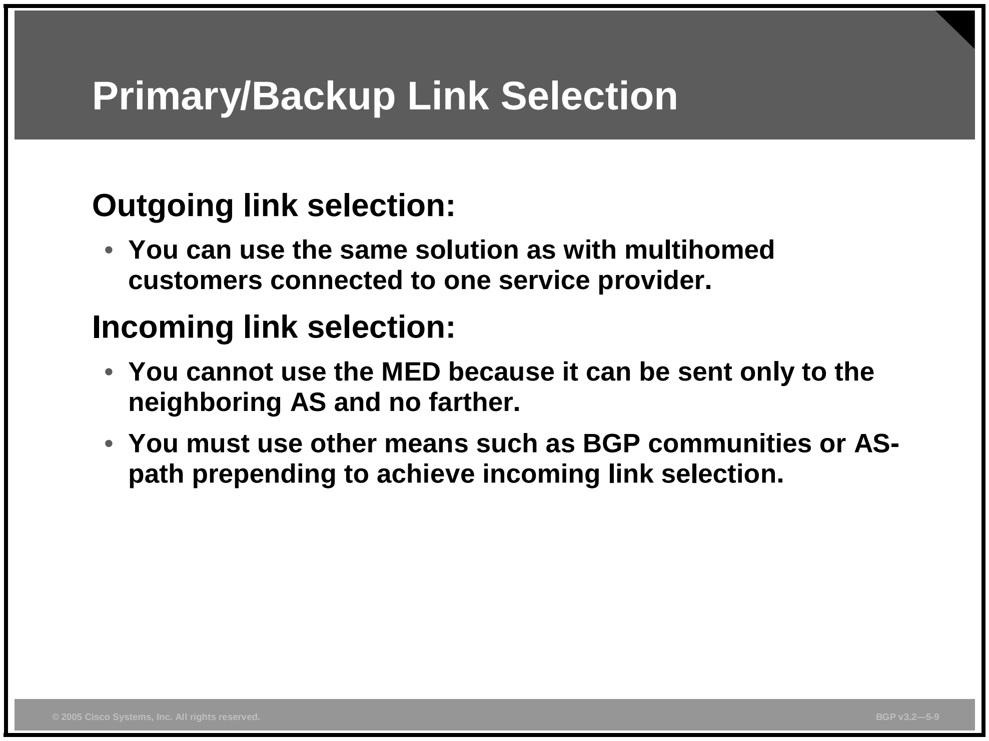
The **neighbor** *ip-address* **local-as** *private-as* router configuration command is used to indicate the AS number that the local router uses as its local AS number in the BGP Open message. The remote router is assumed to have an EBGP session to the indicated local AS.

Internally, the customer network uses another private AS number. When routes are sent to the neighbor, the internal AS number is automatically prepended in the AS path first, and then the specified local AS number is prepended as well. As a consequence, the ISP receives the routes with an AS path with both AS numbers in it. The ISP has to adapt its incoming filter-lists as a result of this situation.

Note Some service providers might be unwilling to change their AS-path input filters, leaving the customer no other option than to use a public AS number or to connect to a single ISP with a private AS number.

Primary/Backup Link Selection

This topic describes how you can implement a typical backup setup between a multihomed customer and multiple service providers in a BGP environment.



Primary/Backup Link Selection

Outgoing link selection:

- **You can use the same solution as with multihomed customers connected to one service provider.**

Incoming link selection:

- **You cannot use the MED because it can be sent only to the neighboring AS and no farther.**
- **You must use other means such as BGP communities or AS-path prepending to achieve incoming link selection.**

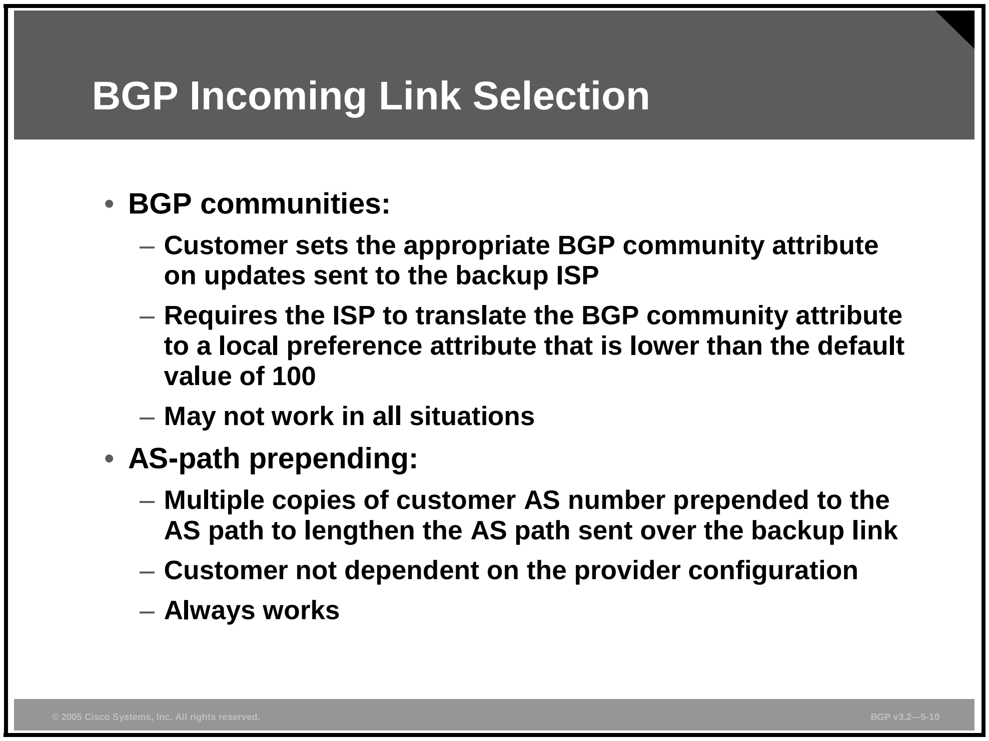
When using BGP on multiple links between a customer and several service provider networks, the customer is solely responsible for controlling the use of the links between them for outgoing traffic. The customer chooses whether to use these links in a primary/backup or a load-sharing configuration.

If one link is primary and the other is used for backup purposes only, the customer can use the local preference attribute in the configuration to direct all outgoing traffic over the primary link. This configuration is no different than the configuration that is used for customers with multiple connections running BGP to a single service provider.

Controlling the load distribution of incoming traffic over multiple links is more difficult in the multihomed scenario when links to multiple service providers are used. You cannot use the multi-exit discriminator (MED) when the customer connects to multiple providers because the updates are sent to two different autonomous systems. Recall that the MED is used only when you compare routes that are received from a single directly connected AS over two parallel links. Therefore, route selection decisions will most likely use the AS-path attribute and prefer the route with the shortest AS-path length.

BGP Incoming Link Selection

This topic describes the use of BGP attributes to influence inbound link selection in customer networks that are multihomed to multiple service providers.



BGP Incoming Link Selection

- **BGP communities:**
  - **Customer sets the appropriate BGP community attribute on updates sent to the backup ISP**
  - **Requires the ISP to translate the BGP community attribute to a local preference attribute that is lower than the default value of 100**
  - **May not work in all situations**
- **AS-path prepending:**
  - **Multiple copies of customer AS number prepended to the AS path to lengthen the AS path sent over the backup link**
  - **Customer not dependent on the provider configuration**
  - **Always works**

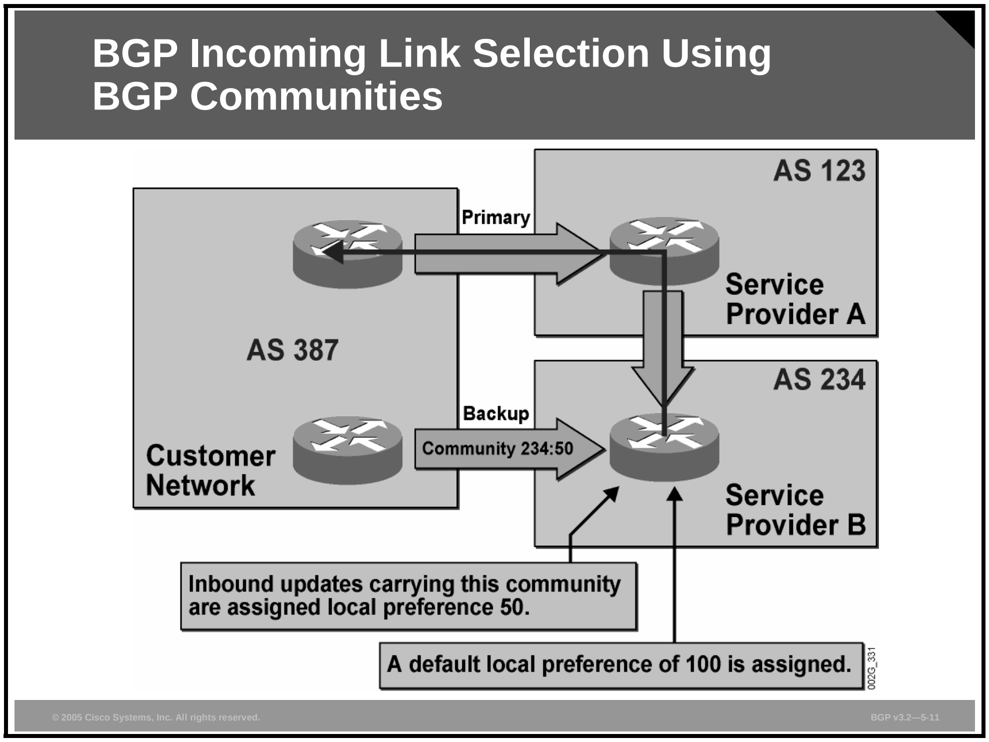
BGP v3.2—5-10

To remove incoming traffic from the backup link, the customer must influence route selection in the backup AS. The backup ISP must be forced to prefer the primary path to reach the customer network, although this choice means selecting a route with a longer AS path.

One way to influence route selection is to use local preference in the network of the backup ISP. Using local preference creates an administrative scalability issue if each customer requires its use, because the ISP must maintain the configuration.

One scalable way of setting local preference in an ISP network is to use communities. The customer sets a well-known community value on the routes that are sent to the backup ISP. The ISP recognizes the community and sets the local preference for these routes. This solution is available only if the ISP has implemented and announced the use of communities. If communities and a local preference setting are used, route selection occurs only if there are alternative routes to compare.

Another way of influencing route selection in the backup ISP is to do AS-path prepending before sending the advertisement to the backup ISP. When the customer sends routes over the backup link, multiple copies of its own AS number are prepended to the AS path of each route. The backup ISP receives the routes and makes normal route-selection decisions. No special weight or local preference settings are used; the BGP route selection is based exclusively on the AS-path length. No special configuration is required in the service provider network.



BGP Incoming Link Selection Using BGP Communities

Inbound updates carrying this community are assigned local preference 50.

A default local preference of 100 is assigned.

In the example here, the backup service provider B (AS 234) has defined the meaning of community 234:50. When AS 234 receives routes with this community, the local preference is set to 50.

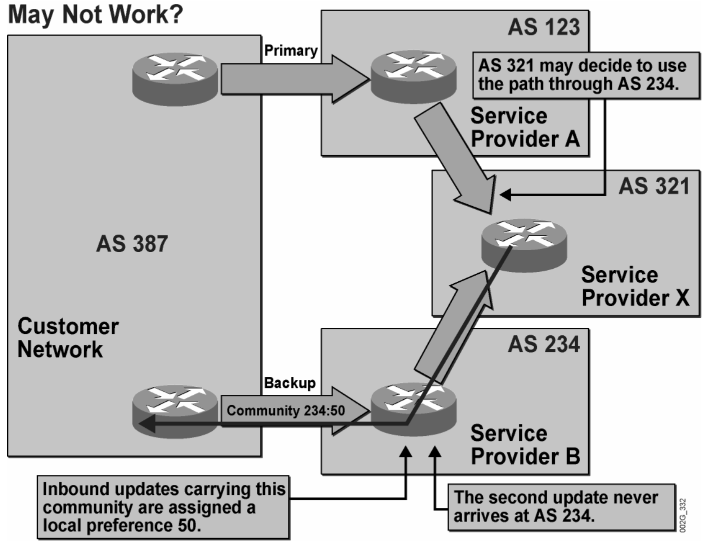
The customer AS 387 is advertising the route over the primary link without any communities. It is received by AS 123 and propagated to AS 234. When AS 234 receives the route via AS 123, there is no community set. AS 234 therefore assigns the default local preference value of 100.

The customer is also advertising the route over the backup link. However, in this case, the route has the community 234:50 set. When AS 234 receives this route, it recognizes the community, and the local preference value is set to 50.

Route selection is now performed in AS 234. The route that has been received via AS 123 is preferred based on the local preference values.

BGP Incoming Link Selection Using BGP Communities (Cont.)

May Not Work?



AS 321 may decide to use the path through AS 234.

Inbound updates carrying this community are assigned a local preference 50.

The second update never arrives at AS 234.

BGP v3.2—5-12

Even when the use of communities is correctly configured, the desired load distribution may not always be achieved. As this example shows, AS 234 does not always receive the primary route, although nothing is wrong in the network.

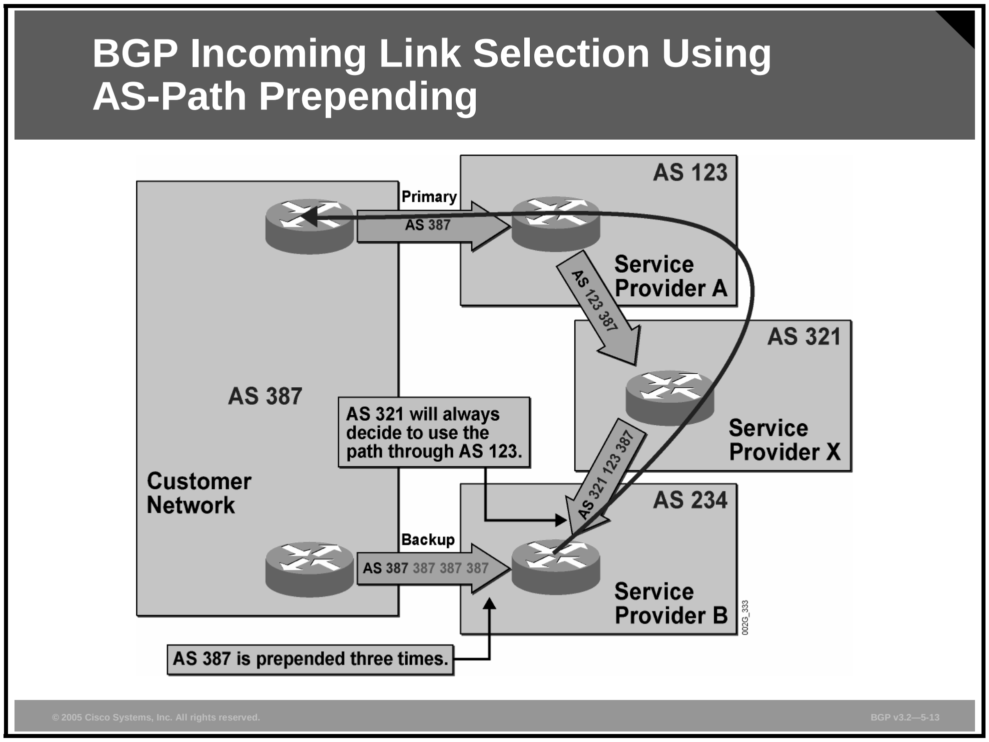
The customer AS 387 sends routes with community 234:50 over the backup link to AS 234. AS 234 receives the routes and sets the local preference to 50. If AS 234 over some period of time selects the directly connected path to AS 387 as the best, it propagates the route to AS 321. As the route is propagated over the EBGP session between AS 234 and AS 321, the local preference value that is used within AS 234 is lost.

AS 321 does not have any use for the community 234:50 because this community is defined and implemented only within AS 234. Potentially, the community value can also be stripped off during BGP route propagation.

Customer AS 387 also sends the routes over the primary link to AS 123. The routes are propagated to AS 321, which now sees two alternative routes to the destination networks within AS 387. Neither weight nor local preference is used by the routers in AS 321 as criteria for reaching AS 387. Both alternatives have equal AS-path lengths.

The route selection decision that will be made in AS 321 is hard to predict, but the outcome definitely influences the route selection decision that was made in AS 234. If AS 321 prefers the route to the customer network via AS 234 for any reason, then the second-best alternative via AS 123 and the primary link is never propagated to AS 234.

In this case, AS 234 never sees the primary path and has to stick to the backup link and announce the route to AS 321. The network has reached a steady state when the traffic uses the backup link although the primary link is available.



BGP Incoming Link Selection Using AS-Path Prepending

AS 321 will always decide to use the path through AS 123.

AS 387 is prepended three times.

BGP v3.2—5-13

In this example, the customer AS 387 is performing AS-path prepending on the backup link. Three copies of the customer AS number (387) are prepended to the AS path. As the route goes out over the EBGP session, BGP prepends the local AS number to the AS-path attribute. AS 234 receives routes from AS 387 over the backup link with an AS-path length of four (the original AS 387 plus three prepended copies that the customer edge router applied to the AS-path attribute).

The customer advertises networks without AS-path prepending over the primary link. AS 123 receives routes with an AS-path length of one and propagates these routes to AS 321, which then receives them with an AS-path length of two.

If, for a short period of time, AS 321 received the customer routes via AS 234, the AS-path length of those routes would have been five. In that case, AS 321 selects the route from AS 123 as the best and propagates it to AS 234.

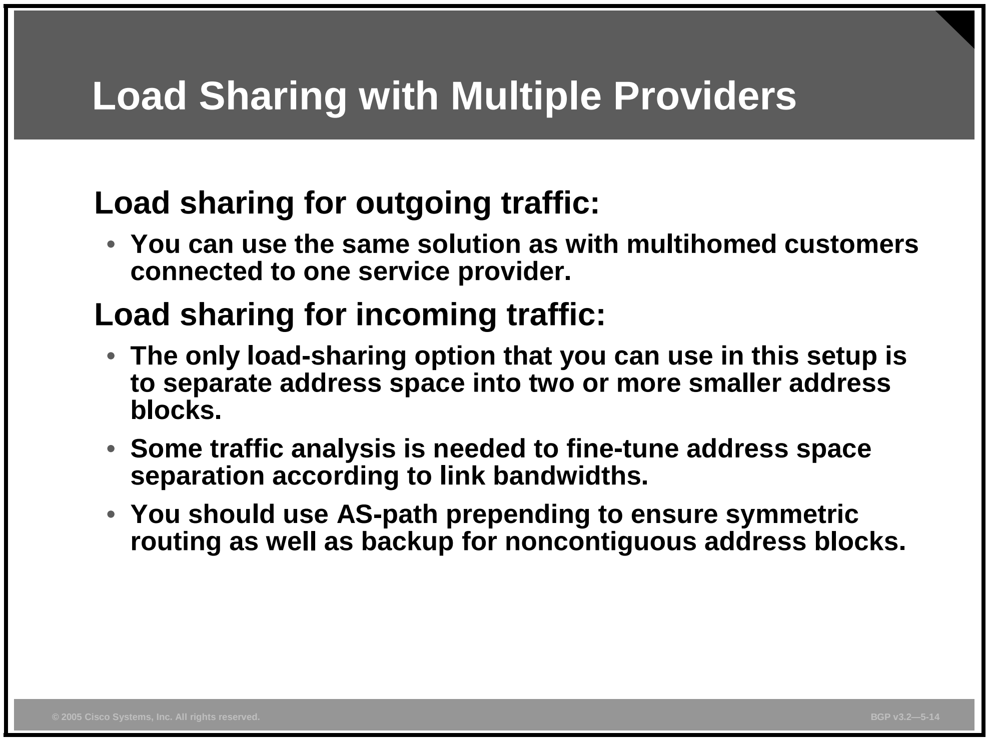
AS 234 now sees both alternatives. The customer routes that have been received directly from the customer have an AS-path length of four. The routes that have been received via AS 321 have an AS-path length of three. Because no weight or local preference is configured in this example, AS 234 selects the route via AS 321 as the best.

The desired result, to have all traffic enter the customer network via the primary link, is now achieved.

Note If the backup ISP is implementing incoming AS-path filters for this customer with the length of the AS path equal to one, the ISP has to change the configuration of the AS-path filter for the customer. The ISP can either create a new filter, allowing multiple copies of the customer AS number only for this customer, or use regular expression variables to create a common filter for all customers that belong to one peer group.

Load Sharing with Multiple Providers

This topic describes how you can implement load sharing between a multihomed customer and multiple service providers in a BGP environment.



Load Sharing with Multiple Providers

Load sharing for outgoing traffic:

- **You can use the same solution as with multihomed customers connected to one service provider.**

Load sharing for incoming traffic:

- **The only load-sharing option that you can use in this setup is to separate address space into two or more smaller address blocks.**
- **Some traffic analysis is needed to fine-tune address space separation according to link bandwidths.**
- **You should use AS-path prepending to ensure symmetric routing as well as backup for noncontiguous address blocks.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-14

Load sharing over links to two different ISPs can be compared to doing load sharing over two parallel links to a single ISP. The only difference is that there is only one option that is available to control incoming traffic. Controlling load distribution of the outgoing traffic is configured in exactly the same way as when a multihomed customer connects to a single service provider.

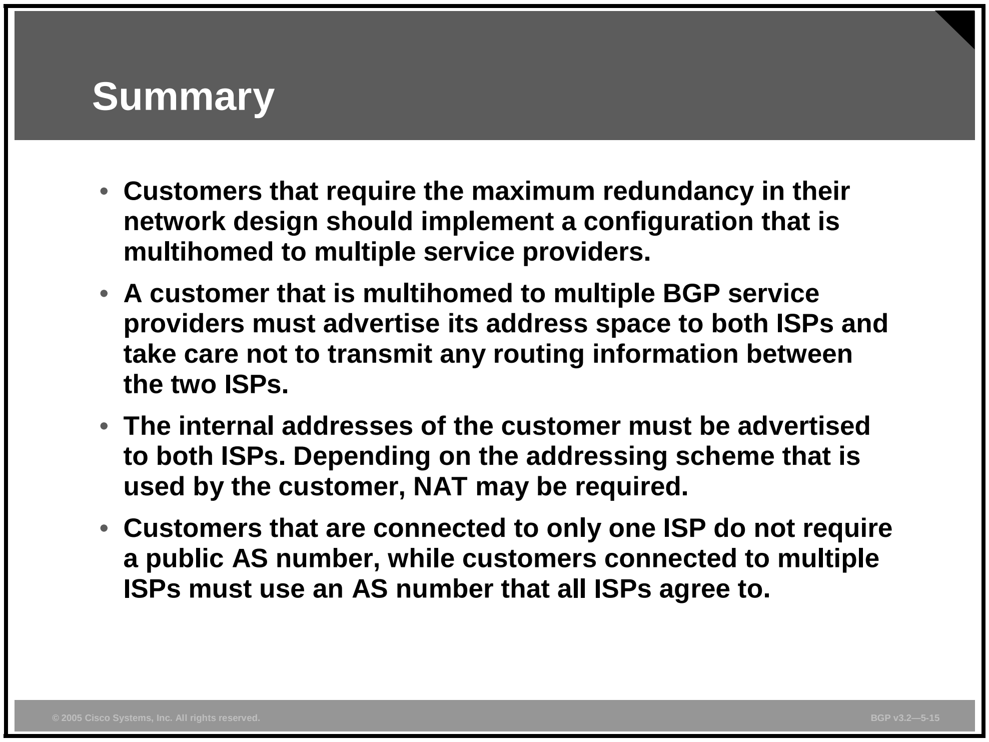
The customer can control the load distribution of incoming traffic based on traffic destination. The customer divides its address space into several announcements. One announcement is sent to one of the ISPs. Another announcement is sent to the other ISP. For backup purposes, the customer announces the entire address space to both ISPs. The ISPs now use the most-explicit route rule, and as long as both links are up, traffic with destinations within one part of the customer address space is routed over one of the links and traffic to the other part is routed over the other link.

It is very difficult to predict the volume of traffic that will be directed to one part of the customer address space and the volume that will be directed to the other part. You should monitor the results of changing route updates by watching the load on the links before and after implementing the change. If the load distribution is not satisfactory, you can further modify the division of the address space. You must then check the load on the links again and further fine-tune the configuration.

A customer may decide to use both the division of address space into several advertisements and AS-path prepending together. Some part of the customer address space may be advertised by the customer network with a longer AS path over one of the links to fine-tune the load. Also, there may be cases where there are noncontiguous subnets that cannot be divided because the prefixes would be too long. These subnets are evenly distributed between the links in a primary/backup configuration.

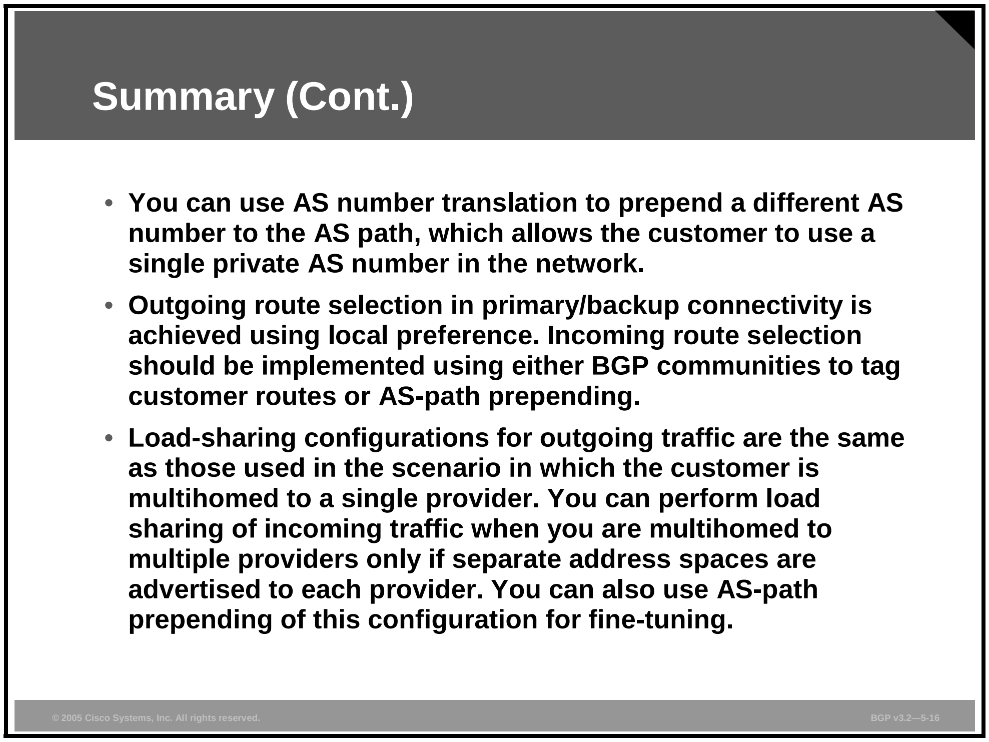
Summary

This topic summarizes the key points discussed in this lesson.



Summary

- Customers that require the maximum redundancy in their network design should implement a configuration that is multihomed to multiple service providers.
- A customer that is multihomed to multiple BGP service providers must advertise its address space to both ISPs and take care not to transmit any routing information between the two ISPs.
- The internal addresses of the customer must be advertised to both ISPs. Depending on the addressing scheme that is used by the customer, NAT may be required.
- Customers that are connected to only one ISP do not require a public AS number, while customers connected to multiple ISPs must use an AS number that all ISPs agree to.

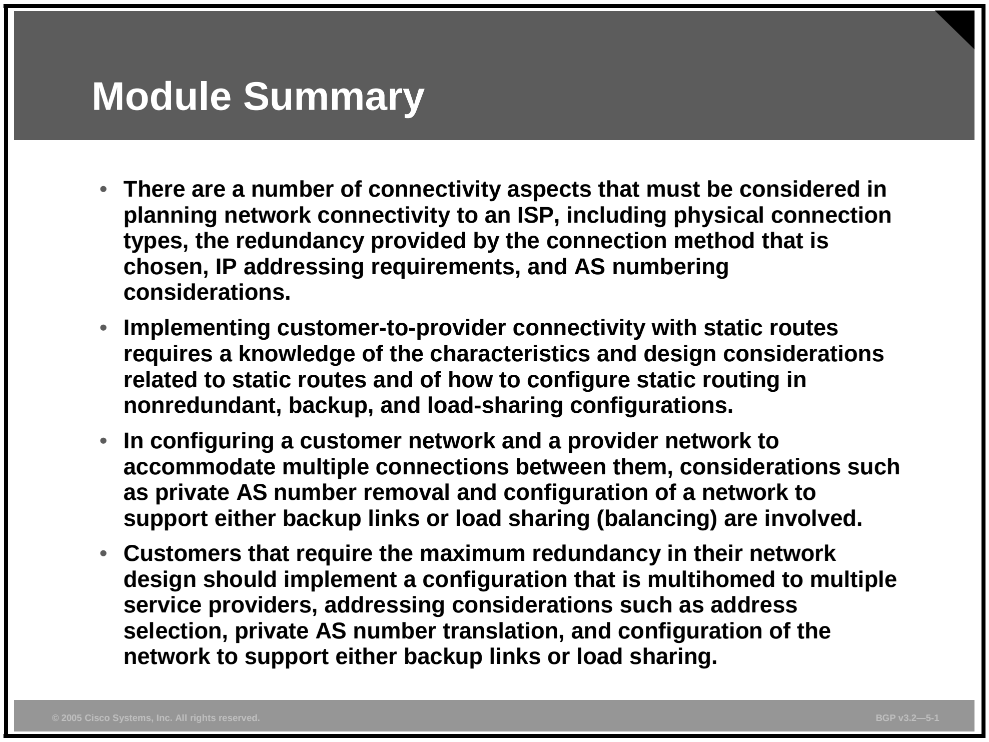


Summary (Cont.)

- You can use AS number translation to prepend a different AS number to the AS path, which allows the customer to use a single private AS number in the network.
- Outgoing route selection in primary/backup connectivity is achieved using local preference. Incoming route selection should be implemented using either BGP communities to tag customer routes or AS-path prepending.
- Load-sharing configurations for outgoing traffic are the same as those used in the scenario in which the customer is multihomed to a single provider. You can perform load sharing of incoming traffic when you are multihomed to multiple providers only if separate address spaces are advertised to each provider. You can also use AS-path prepending of this configuration for fine-tuning.

Module Summary

This topic summarizes the key points discussed in this module.



Module Summary

- **There are a number of connectivity aspects that must be considered in planning network connectivity to an ISP, including physical connection types, the redundancy provided by the connection method that is chosen, IP addressing requirements, and AS numbering considerations.**
- **Implementing customer-to-provider connectivity with static routes requires a knowledge of the characteristics and design considerations related to static routes and of how to configure static routing in nonredundant, backup, and load-sharing configurations.**
- **In configuring a customer network and a provider network to accommodate multiple connections between them, considerations such as private AS number removal and configuration of a network to support either backup links or load sharing (balancing) are involved.**
- **Customers that require the maximum redundancy in their network design should implement a configuration that is multihomed to multiple service providers, addressing considerations such as address selection, private AS number translation, and configuration of the network to support either backup links or load sharing.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—5-1

This module explained the different requirements for connectivity between customers and service providers. The first lesson provided a connectivity overview that included physical connection methods, redundancy, load balancing, and technical requirements. The second lesson described how to implement customer connectivity by using static routing in a service provider network. The third lesson addressed the implementation of customer connectivity using BGP in a customer scenario in which multiple connections to a single ISP were supported. The final lesson focused on the implementation of customer connectivity using BGP in a customer scenario in which connections to multiple ISPs were supported.

References

For additional information, refer to these resources:

- Cisco Systems, Inc. *The Easy Guide to Selecting an Internet Service Provider*. http://www.cisco.com/warp/public/cc/so/cuso/smso/crn/ezgd_pl.htm.
- Cisco Systems, Inc. *How NAT Works*. http://www.cisco.com/en/US/tech/tk648/tk361/technologies_tech_note09186a0080094831.shtml.
- Cisco Systems, Inc. *Sample Configurations for Load Sharing with BGP in Single and Multihomed Environments: Sample Configurations*. http://www.cisco.com/warp/public/459/40.html.
- Cisco Systems, Inc. *Sample Configuration for BGP with Two Different Service Providers (Multihoming)*. http://www.cisco.com/warp/public/459/27.html.
- Doyle, Jeff. *Routing TCP/IP, Volume 1*. Cisco Press, 1998, ISBN 1-57870-041-8.
- Cisco Systems, Inc. *Removing Private Autonomous System Numbers in BGP*. http://www.cisco.com/warp/public/459/32.html.
- Cisco Systems, Inc. *How BGP Routers Use the Multi-Exit Discriminator for Best Path Selection*. http://www.cisco.com/warp/public/459/37.html.
- Cisco Systems, Inc. *Configuring the BGP Local-AS Feature*. http://www.cisco.com/warp/public/459/39.html.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) If a customer requires additional bandwidth and redundancy, which approach is preferred? (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) a single permanent connection to one ISP
B) permanent connections to more than one ISP
C) dial-up connections to more than one ISP
D) multiple permanent connections to one ISP

Q2) Which type of redundancy do multiple permanent connections that provide load-sharing configuration display? (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) link
B) equipment
C) service provider
D) routing protocol

Q3) In a customer-to-provider routing scheme, which method of routing is preferred because of its lower complexity? (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) policy-based routing
B) dynamic routing
C) content routing
D) static routing

Q4) Why is it that with multiple permanent connections to more than one ISP, the use of dynamic routing with BGP is required? (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) When one of the connections is lost, the link level detects this loss and places the interface in a down state.
B) Monitoring of the link status cannot detect a problem inside one of the ISP networks.
C) Static routes detect problems inside one of the ISP networks.
D) It is not required, and static routing may be used.

Q5) What can be done when a customer is assigned only a very small subnet of public addresses? (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) purchase more addresses as required
B) use NAT
C) add a service provider
D) add links to the same service provider

Q6) What are two different addressing schemes that customers use to connect to a service provider? (Choose two.) (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) provider-independent
B) customer-independent
C) provider-assigned
D) customer-assigned

Q7) Which two of the following criteria are required for a customer to be multihomed to multiple ISPs? (Choose two.) (Source: Understanding Customer-to-Provider Connectivity Requirements)

A) The customer must have a public AS number.
B) The customer must have a private AS number.
C) The customer must run BGP with both of its ISPs.
D) The customer must run BGP with one ISP and may use static routing with the other.

Q8) What are two requirements for being able to use static routing as part of installing redundant connections between the customer network and a single service provider network? (Choose two.) (Source: Implementing Customer Connectivity Using Static Routing)

A) The router must be able to detect a link failure.
B) The default route must be announced using the customer IGP.
C) If one link goes down, the interface must remain in an up state.
D) The customer IGP must continue to advertise the static default route.

Q9) A customer route that should not be announced to the rest of the Internet is marked using what? (Source: Implementing Customer Connectivity Using Static Routing)

A) a route tag
B) the export community
C) the no-export community
D) the public address filter

Q10) When you are designing static route propagation in a service provider network, which three steps must you take? (Choose three.) (Source: Implementing Customer Connectivity Using Static Routing)

A) assign a tag to each combination of services
B) configure a community that matches defined tags
C) redistribute static routes into BGP through a route-map
D) identify all possible combinations of services that are offered to a customer

Q11) What does a route-map assign that will be used by other routers within a network? (Source: Implementing Customer Connectivity Using Static Routing)

A) a tag
B) community values
C) public addressing
D) QoS

Q12) Which three key pieces of information can you derive from the following router command output? (Choose three.) (Source: Implementing Customer Connectivity Using Static Routing)

```
AS387_Backup# sh ip bgp 11.2.3.0
BGP routing table entry for 11.2.3.0/24, version 7
Paths: (2 available, best #1, not advertised to EBGP peer)
  Advertised to non peer-group peers:
  10.3.0.5
  Local
    0.0.0.0 from 0.0.0.0 (10.3.0.6)
      Origin incomplete, metric 0, localpref 100, weight 32768, valid,
      sourced, best
      Community: 387:31000 no-export
  Local
    10.3.0.2 (metric 128) from 10.3.0.5 (1.0.0.2)
      Origin incomplete, metric 0, localpref 100, valid, internal
      Originator: 1.0.0.2, Cluster list: 10.3.0.5
      Community: 387:31000 no-export
```

A) The primary link has come back up, so the backup router now sees two alternate routes.

B) The primary link has not come back up, but the backup router still sees two alternate routes.

C) The first route is the route that the router itself has redistributed into BGP using the floating static route. This route is locally sourced by the AS and has been assigned a weight value of 32768.

D) The second route is the one that has been received by IBGP from the primary edge router. The AS also sources this route, but no weight value is assigned.

Q13) Which two things can you do to overcome the problems that occur when a floating static route is redistributed into BGP? (Choose two.) (Source: Implementing Customer Connectivity Using Static Routing)

A) You must raise the weight value.

B) You must lower the weight value.

C) You must set the AD at a higher value than all other routes.

D) You must assign local preference values, giving the floating static route a lower local preference value than the primary route.

Q14) What are three characteristics of using static routes during load sharing of outgoing traffic? (Choose three.) (Source: Implementing Customer Connectivity Using Static Routing)

A) Outgoing traffic load sharing is easy to achieve.

B) Each customer router uses the closest customer edge router as the exit point.

C) Balanced load sharing is achieved if the customer edge routers are collocated.

D) Local preference values must be assigned, giving the floating static route a lower local preference value than the primary route.

Q15) What are three responsibilities of the customer when the customer is multihomed to a single service provider? (Choose three.) (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) Customer edge routers must run IBGP between them.

B) The customer must advertise a default route.

C) The customer must conditionally advertise its assigned address space into BGP.

D) The customer edge routers must run EBGP with the provider.

Q16) Given the following router command output, which method has been used to influence return traffic in a primary/backup link implementation for this multihomed customer? (Source: Connecting a Multihomed Customer to a Single Service Provider)

```
Provider# show ip bgp

BGP table version is 5, local router ID is 10.0.33.34
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
*  10.10.20.0/24    192.168.63.3          1000             0 100 100 i
*>                  192.168.64.4          2000             0 400 100 i
*> 30.30.30.0/24    192.168.63.3             0             0 100 I
*> 40.40.40.0/24    192.168.64.4             0             0 400 I
```

A) MED
B) local preference
C) weight
D) AS-path prepending

Q17) What are three responsibilities of the provider router when supporting a multihomed customer? (Choose three.) (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) The provider must advertise a default route to the customer through BGP.
B) The provider must filter customer routes to verify that proper addressing is used.
C) The provider must remove the private AS number, if it is in use by the customer.
D) The provider must configure new AS-path filters to allow AS-path prepending; otherwise, a primary/backup link cannot be established.

Q18) What will occur if private AS numbers are advertised to the Internet? (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) The Internet will not be able to route packets.
B) Internet routers could drop routes based on BGP loop-prevention mechanisms.
C) Customer load balancing will not function.
D) Customer configurations for the primary/backup link using AS-path prepending will not function.

Q19) Which two BGP configurations are required to properly implement a backup solution for a multihomed customer that is connected to a single provider? (Choose two.) (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) The customer should set local preference to influence outgoing route selection.
B) The customer should set the weight attribute to influence outgoing path selection.
C) The customer should set the MED on each route to influence return path selection.
D) The customer should configure AS-path prepending to ensure proper outgoing path selection.

Q20) A customer router has been configured with maximum paths set to a value of 4. Given the following router command output, over how many links will the router need to perform load balancing? (Source: Connecting a Multihomed Customer to a Single Service Provider)

```
router# show ip bgp

BGP table version is 5, local router ID is 10.0.33.34
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
*  10.10.20.0/24    192.168.63.3                           0 300 100 100 i
*>                  192.168.64.4                           0 400 100 i
*                   192.168.65.5                           0 500 100 i
*> 30.30.30.0/24    192.168.63.3             0             0 300 i
*> 40.40.40.0/24    192.168.64.4             0             0 400 i
```

A) The router will use only the path marked as “best” by BGP.

B) The router will perform load balancing over two paths to reach network 10.10.20.0/24.

C) The router will perform load balancing over three paths to reach network 10.10.20.0/24.

D) There is not enough information to determine the correct answer.

Q21) Which three methods can you use to provide load sharing over network links between a multihomed customer and a single provider? (Choose three.) (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) advertising of split addressing space to the provider

B) configuring **ebgp-multihop** between the customer and the provider

C) using the BGP **maximum-paths** command to perform load balancing over parallel links

D) configuring multiple static routes that point to the provider

Q22) Why is it not required to configure maximum paths under the BGP routing process when load balancing is being performed because the **ebgp-multihop** command has been configured? (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) By default, BGP will perform load balancing over up to four paths, configurable up to six.

B) The static route or IGP process is responsible for load balancing in this configuration.

C) Configuring multihop enables maximum paths equal to the TTL setting of the **neighbor ebgp-multihop** command.

D) Configuring **ebgp-multipath** is a required component of **ebgp-multihop** load balancing.

Q23) Which two of the following characteristics accurately describe the BGP Support for Dual AS Configuration for Network AS Migrations feature? (Choose two.) (Source: Connecting a Multihomed Customer to a Single Service Provider)

A) allows you to merge a secondary AS under a primary AS without disrupting customer peering sessions
B) allows a router to appear, to external peers, as a member of primary AS during the AS migration
C) allows a router to appear, to external peers, as a member of secondary AS during the AS migration
D) eliminates the possibility that routing loops can be created

Q24) A multihomed customer is using AS number 65550 internally. The customer is connected to two different providers. Provider 1 (in AS 222) has assigned the customer an AS number of AS 65101. Provider 2 (in AS 333) has assigned the customer an AS number of AS 65201. Given that the customer will use AS number translation for its internal AS, what is the AS-path attribute (attached to routes that originated in the customer network) that will be displayed on a router in the network of Provider 2? (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) 65550 i
B) 65201 i
C) 65201 65550 i
D) 333 65201 i

Q25) Which three methods can you use to provide load sharing over network links between a multihomed customer and multiple providers? (Choose three.) (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) advertising of split addressing space to the provider
B) configuring of multiple static routes that point to the provider
C) using the BGP **maximum-paths** command to perform load sharing over parallel links
D) AS-path prepending to fine-tune the load-sharing configuration

Q26) What are three BGP configuration characteristics of a multihomed customer that is connected to multiple providers? (Choose three.) (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) The customer announces assigned addressing to its providers through BGP
B) The customer announces a default route to its network through BGP.
C) The provider announces a default route, local routes, or full Internet routing to the customer via BGP.
D) The customer configures outbound filters to prevent its network from becoming a transit area.

Q27) A multihomed customer is using AS number 1024 and is connected to two different providers (Provider 1: AS 222 and Provider 2: AS 333). The customer has configured the MED to ensure a proper return path so that Provider 1 is the primary provider and Provider 2 is the backup provider. Unfortunately, return traffic continues to use the backup link. What is a possible cause of this problem? (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) The backup provider is ignoring the MED attribute on received routes.
B) The MED attribute cannot be sent to the backup provider because it is local to AS 1024 only.
C) The customer has not set the proper BGP communities to allow the primary and backup providers to correctly set the MED attribute.
D) The MED cannot be used in this scenario, because it will not be advertised to providers upstream of Provider 2.

Q28) What are three important considerations for customers that wish to connect to multiple providers? (Choose three.) (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) The customer has to consider whether to use PA or PI address space.
B) The customer has to decide whether to use static routes or BGP to connect to upstream providers.
C) The customer has to decide whether to use a public AS number or a private AS number scheme.
D) The customer has to decide whether to perform load sharing or use a primary/backup implementation over redundant links.

Q29) Which AS number selection is the best possible choice for a customer that is multihomed to multiple providers? (Source: Connecting a Multihomed Customer to Multiple Service Providers)

A) a single public AS number
B) a single private AS number
C) two private AS numbers that are used in conjunction with AS number translation
D) multiple private AS numbers, one used internally by the customer and the others used in conjunction with AS number translation for each provider

Q30) Given the following router command output, which two methods have been configured to influence return traffic in a primary/backup link for this multihomed customer? (Choose two.) (Source: Connecting a Multihomed Customer to Multiple Service Providers)

```
Provider# show ip bgp

BGP table version is 5, local router ID is 10.0.33.34
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
*  10.10.20.0/24    192.168.63.3          2000             0 100 100
100 i
*>                  192.168.64.4                           0 300 100 i
*> 30.30.30.0/24    192.168.63.3             0             0 300 100 i
*> 40.40.40.0/24    192.168.64.4             0             0 100 i
```

A) MED

B) local preference

C) split address advertisement

D) AS-path prepending

Module Self-Check Answer Key

Q1)	D
Q2)	A
Q3)	D
Q4)	B
Q5)	B
Q6)	A, C
Q7)	A, C
Q8)	A, B
Q9)	C
Q10)	A, C, D
Q11)	B
Q12)	A, C, D
Q13)	B, D
Q14)	A, B, C
Q15)	A, C, D
Q16)	A
Q17)	A, B, C
Q18)	B
Q19)	A, C
Q20)	B
Q21)	A, B, C
Q22)	B
Q23)	A, C
Q24)	C
Q25)	A, C, D
Q26)	A, C, D
Q27)	D
Q28)	A, C, D
Q29)	A
Q30)	C, D

Module 6

Scaling Service Provider Networks

Overview

In standard Border Gateway Protocol (BGP) implementations, all BGP routers within an autonomous system (AS) must be fully meshed so that all external routing information can be distributed among the other routers that reside within the AS. Therefore, within an AS, all routers must establish TCP sessions with all other BGP routers. As the AS grows, scalability challenges arise because of an ever-increasing number of TCP sessions and demands for router CPU and memory resources.

This module discusses network scalability concerns that are common to large, complex service provider networks. The module also discusses BGP route reflectors and confederations as scalability mechanisms that allow network designers to steer away from BGP full-mesh requirements and improve network scalability by reducing the number of TCP sessions that are required within an AS. Also discussed in this module are the Cisco IOS commands that are needed to configure and monitor BGP route reflectors and confederations.

Module Objectives

Upon completing this module, you will be able to enable route reflection and confederations as possible solutions to BGP scaling issues in a typical service provider network with multiple BGP connections to other autonomous systems. This ability includes being able to meet these objectives:

- Describe common routing scalability issues in service provider networks
- Describe the function of route reflectors in a BGP environment
- Describe the function of hierarchical route reflectors, based upon established route reflector design rules
- Configure proper operation of route reflectors to modify IBGP split-horizon rules in an existing IBGP network
- Describe the function of confederations in a BGP environment
- Configure proper operation of confederations to modify IBGP AS-path processing in an existing IBGP network

Lesson 1

Scaling IGP and BGP in Service Provider Networks

Overview

Properly scaling IP addressing, Interior Gateway Protocols (IGPs), and Border Gateway Protocol (BGP) is a common area of concern to all service providers and can be the difference between a successful and a problematic BGP implementation. Because service provider networks are complex and must meet the administrative policy and routing demands of the internal network, different customers, and other providers, proper scaling is crucial to the success of the network. Interactions between IGPs and the BGP, specifically when network administrators are supporting internal routing, customer connectivity, and transit traffic (and the administrative policies that match), can be quite complex. Furthermore, the large number of prefixes that are required to support full Internet routing requires administrators to fully characterize IGP and BGP interactions for internal networks and customers alike.

This lesson discusses network scalability concerns common to large, complex service provider networks. Included in this lesson is a description of a typical Internet service provider (ISP) network and discussion of the propagation of internal and customer routing information, scaling considerations for IGPs and BGP, and scaling of IP addressing in service provider networks.

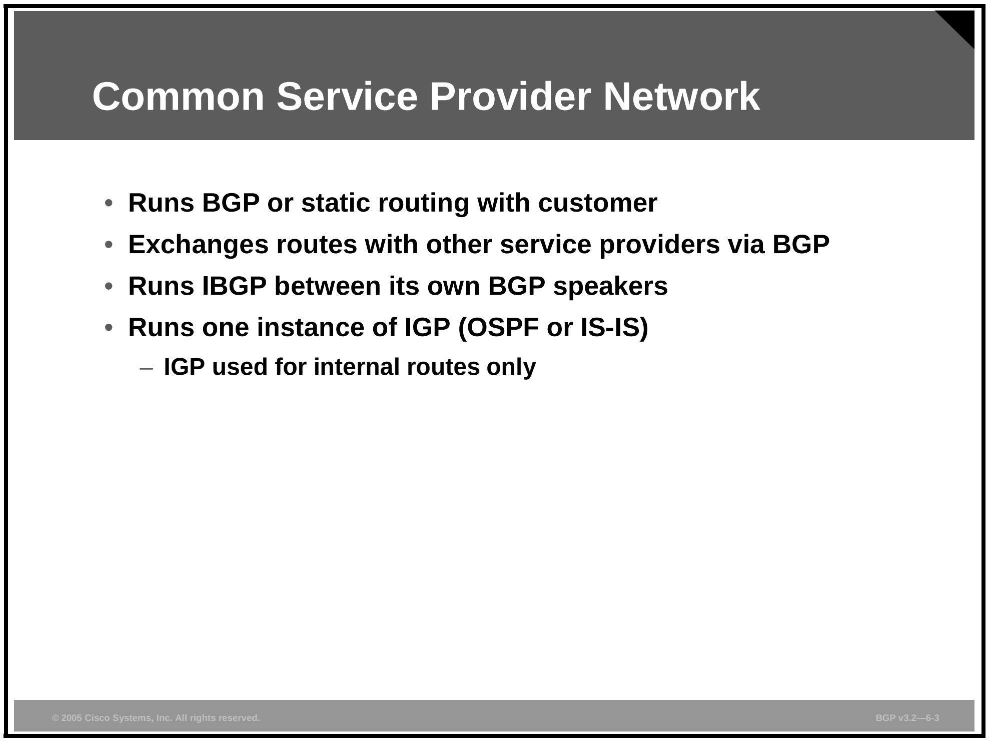
Objectives

Upon completing this lesson, you will be able to describe common routing scalability issues in service provider networks. This ability includes being able to meet these objectives:

- Describe the basic structure of service provider networks
- Describe the propagation of internal and customer routes in service provider networks
- Describe proper scaling of IGPs and BGP in service provider networks
- Describe scaling issues that are relevant to IP addressing in ISP networks
- Describe the function of BGP policy accounting in relation to BGP scaling

Common Service Provider Network

This topic describes the basic structure of service provider networks.



Common Service Provider Network

- **Runs BGP or static routing with customer**
- **Exchanges routes with other service providers via BGP**
- **Runs IBGP between its own BGP speakers**
- **Runs one instance of IGP (OSPF or IS-IS)**
  - **IGP used for internal routes only**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-3

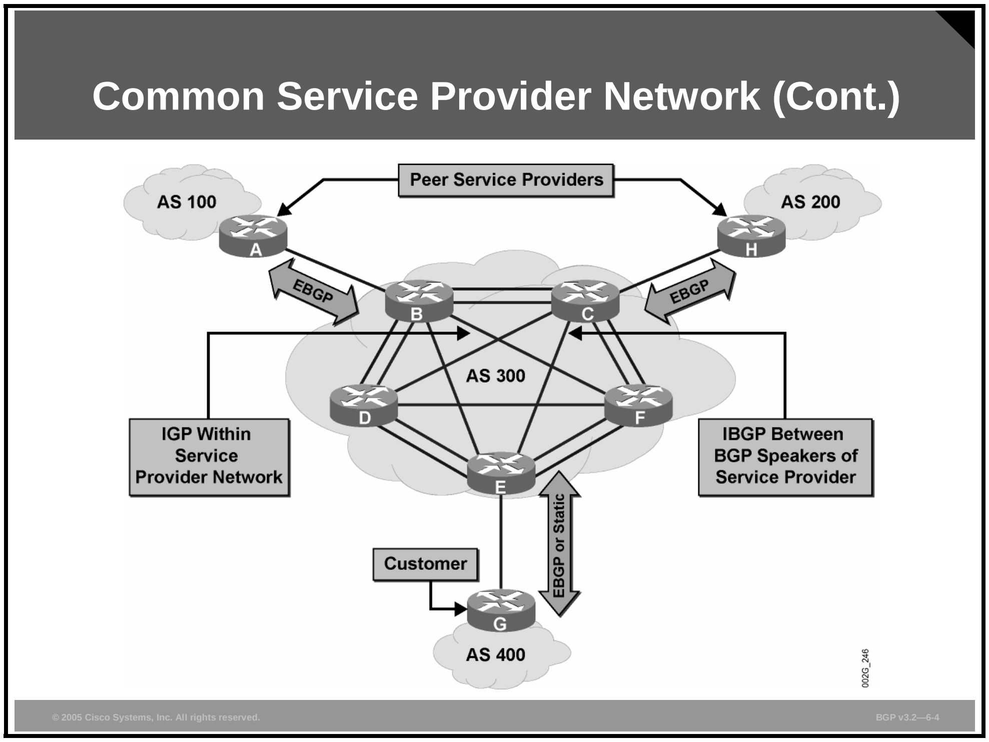
The common service provider network runs External Border Gateway Protocol (EBGP) or static routing with customers. EBGP is always used as the routing protocol between different service providers.

Internal Border Gateway Protocol (IBGP) is required in the provider network because all EBGP-speaking routers in an autonomous system (AS) must exchange external routes via IBGP. Also, non-EBGP speakers are required to take part in the IBGP exchange if they are in a transit path and forward packets based on destination IP addresses.

The service provider network also runs an IGP. The protocols of choice are Open Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (IS-IS). The IGP is used for two purposes:

- Provides IP connectivity between all IBGP speakers so that TCP sessions for IBGP can be established between BGP-speaking routers
- Provides optimal routing to the BGP next-hop address

A single IGP should be used within the entire AS. This setup facilitates effective packet forwarding from the ingress router to egress routers. The IGP is configured to carry internal routes only, including internal links and loopback addresses of the routers. For performance and scalability reasons, no customer routes or external routes should be injected into the IGP.



Common Service Provider Network (Cont.)

Peer Service Providers

IGP Within Service Provider Network

IBGP Between BGP Speakers of Service Provider

Customer

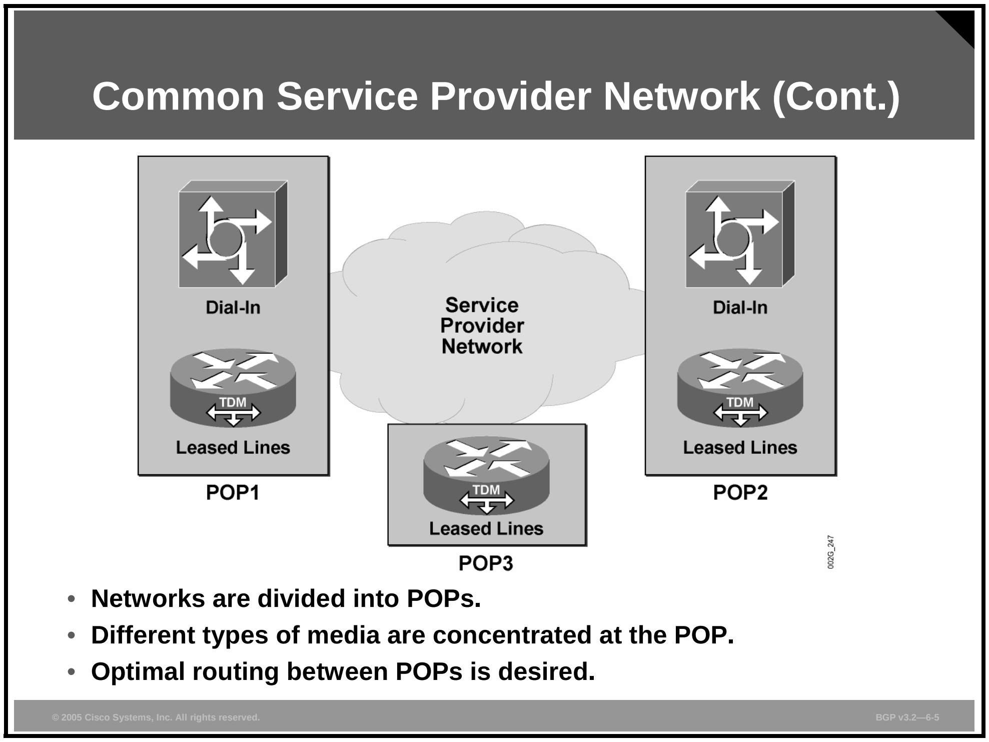
BGP v3.2—6-4

The typical service provider network consists of a network core that connects various edge devices. Some of the edge devices connect customers; others connect to other service providers.

The edge devices that connect to other service providers use EBGP to exchange routing information. The edge devices that connect customers use either static routing or EBGP.

Unless Multiprotocol Label Switching (MPLS) is configured on the service provider backbone, routers in a transit path are also required to have full routing information. Therefore, these routers take part in the IBGP routing exchange.

An IGP is also required within the service provider network. The IGP is used to carry internal routes, including the loopback interface addresses of IBGP-speaking routers. The IGP provides reachability information to establish IBGP sessions and to perform the recursive routing lookup for the BGP next hop.



Common Service Provider Network (Cont.)

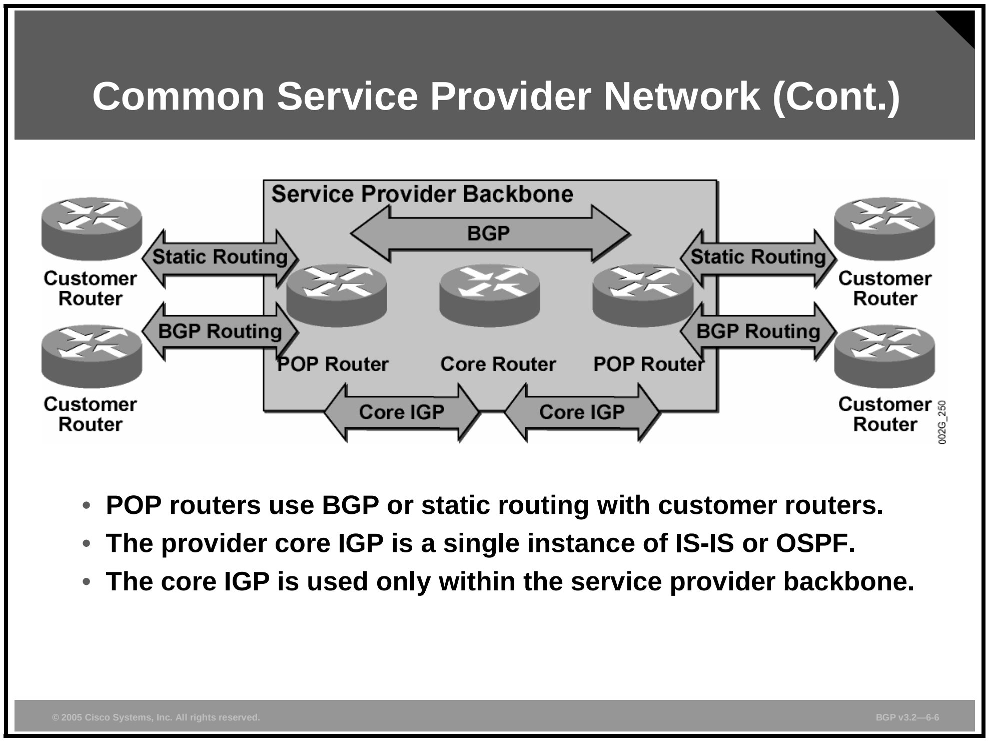
- Networks are divided into POPs.
- Different types of media are concentrated at the POP.
- Optimal routing between POPs is desired.

BGP v3.2—6-5

Service provider devices that connect access links to customers are physically located in groups that are called points of presence (POPs). In general, the POP is a group of routers where access links are terminated. The edge routers that peer with other service providers can in this sense be considered a POP.

Service providers use different types of access links with different types of customers and usually mix access links in the same POP. Some customers use leased lines, others use xDSL, and still others use dial-in access or any other access that the provider can support.

POP routers connect to the network core using a layer of concentration routers at the POP. The network core forwards packets between POPs, various customer access points, or peering points with other service providers. Optimal routing between POPs is a desirable feature.



Common Service Provider Network (Cont.)

- **POP routers use BGP or static routing with customer routers.**
- **The provider core IGP is a single instance of IS-IS or OSPF.**
- **The core IGP is used only within the service provider backbone.**

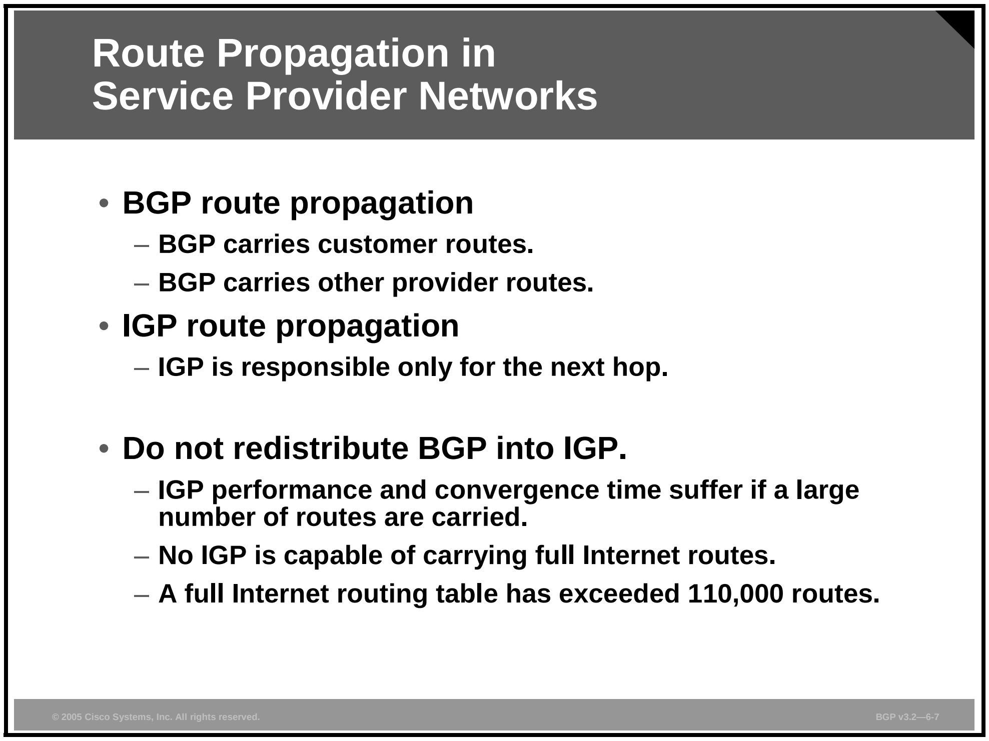
Customer access lines terminate in the POP edge routers. In many cases, the POP edge routers use static routing to customer networks. The POP edge routers advertise static routes to the rest of the service provider network and to other autonomous systems using BGP.

Service providers use BGP routing with the customer when redundancy requires the use of a routing protocol.

The service provider backbone typically uses a single instance of either IS-IS or OSPF as its IGP. The IGP is used within the provider backbone only. The provider backbone exchanges no IGP routing information with customer routers or with routers in other autonomous systems.

Route Propagation in Service Provider Networks

This topic describes the propagation of internal and customer routes in service provider networks.



Route Propagation in Service Provider Networks

- **BGP route propagation**
  - **BGP carries customer routes.**
  - **BGP carries other provider routes.**
- **IGP route propagation**
  - **IGP is responsible only for the next hop.**
- **Do not redistribute BGP into IGP.**
  - **IGP performance and convergence time suffer if a large number of routes are carried.**
  - **No IGP is capable of carrying full Internet routes.**
  - **A full Internet routing table has exceeded 110,000 routes.**

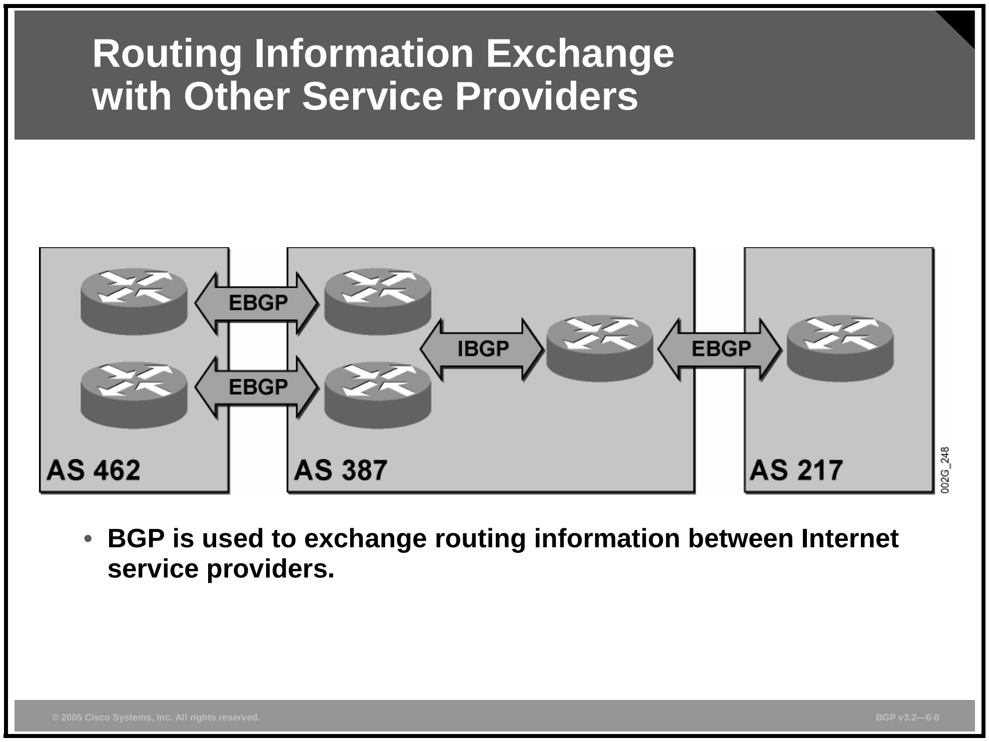
BGP v3.2—6-7

It is important to avoid sending any unnecessary routing information in the IGP. The IGP performs best if it carries as few routes as possible. Optimally, the IGP should contain only information about BGP next hops and routes that are internal to the service provider network, enabling the establishment of IBGP sessions.

All other routing information should be carried in BGP, which is designed to scale for large volumes of routing information. Customer routes and the routes from other service providers should be carried in BGP. These routes should not be propagated from BGP into the provider IGP.

IGP performance and convergence time suffer if the IGP carries a larger number of routes. The design goal should be to minimize the volume of routing information that is carried by the IGP. Naturally, the number of route flaps is also reduced as the number of routes is reduced.

BGP scales to a much larger volume of routing information because of the inherent qualities of the design of BGP. Potentially, the BGP routers of the service provider can receive the full Internet routing table, which has exceeded 110,000 routes. You should therefore never redistribute the routing information that has been received by BGP into the IGP, because no IGP is capable of carrying several tens of thousands of routes.



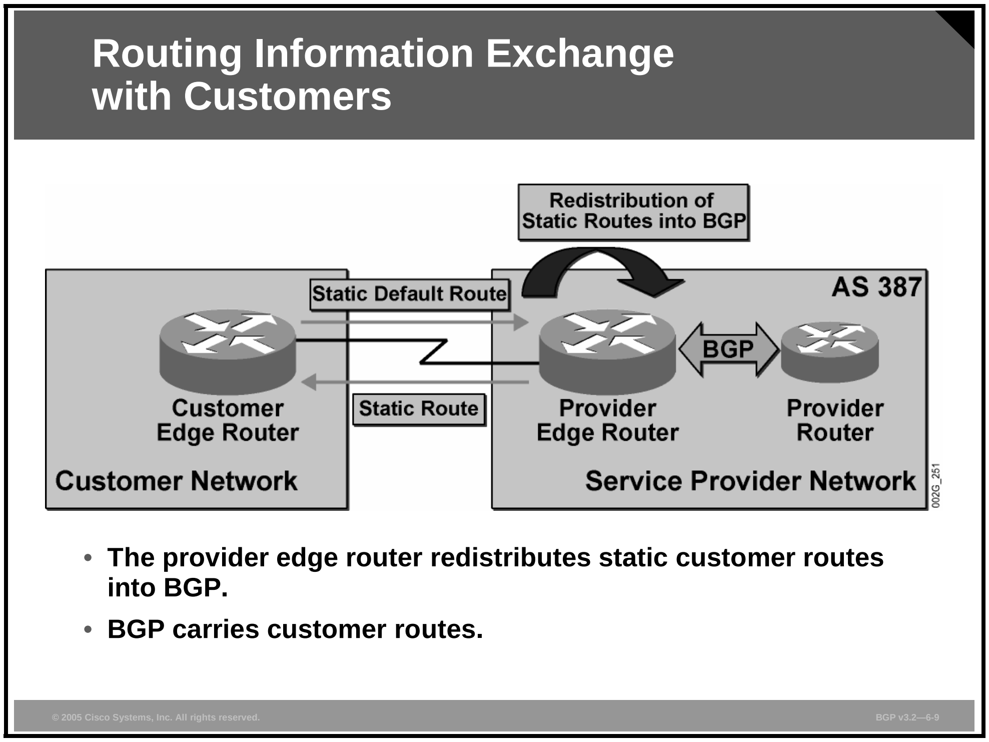
Routing Information Exchange with Other Service Providers

- **BGP is used to exchange routing information between Internet service providers.**

BGP v3.2—6-8

Provider edge routers use BGP to exchange routing information with other service provider networks for redundancy and scalability reasons.

Static routing with other service providers is generally not a viable solution due to the dynamic routing requirements of the service provider environment. Routing information is received at provider edge routers using EBGP and then propagated using IBGP to the rest of the service provider network. At another edge router, the routing information is further propagated to a different service provider using EBGP with other autonomous systems.

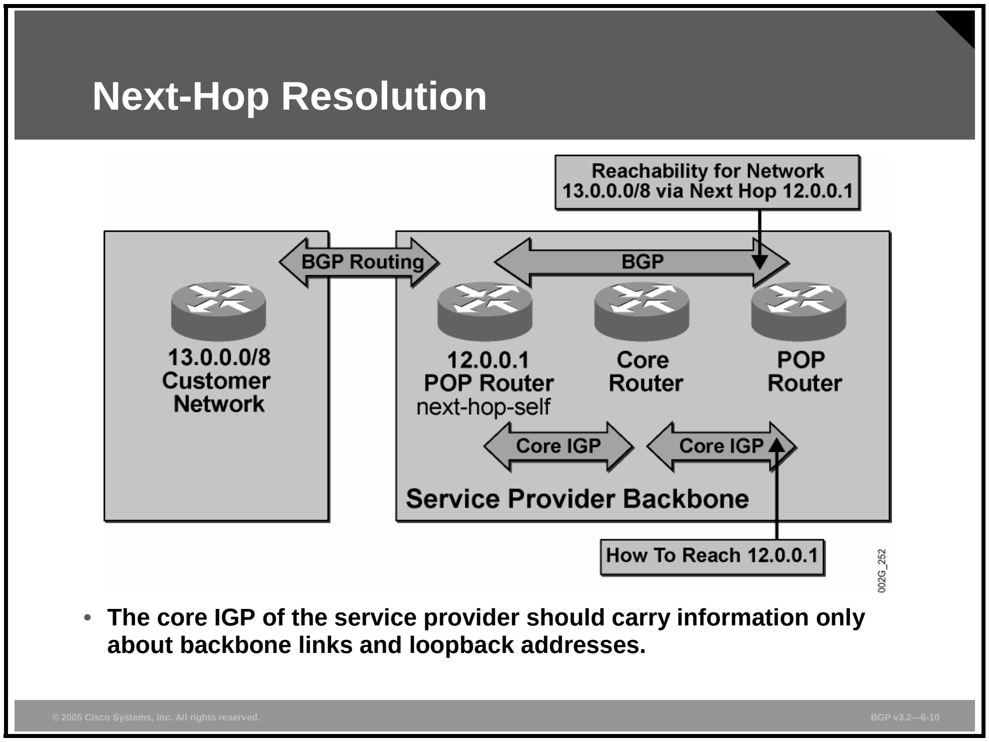


Routing Information Exchange with Customers

- **The provider edge router redistributes static customer routes into BGP.**
- **BGP carries customer routes.**

The provider edge router typically uses static routing to reach customer networks. In this case, the customer typically configures a static default route that points to the edge router of the service provider.

The provider edge router redistributes customer static routes into BGP. The service provider network then uses BGP to propagate the information to the rest of the service provider network using IBGP. The service provider also advertises customer routing information to other autonomous systems using BGP.



Next-Hop Resolution

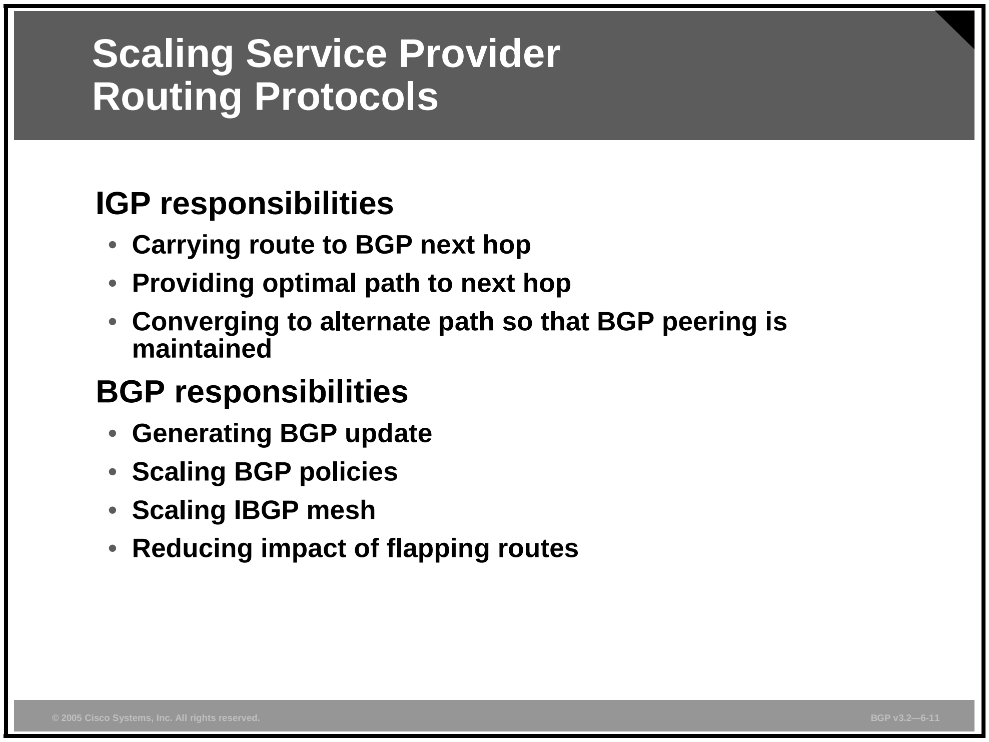
- **The core IGP of the service provider should carry information only about backbone links and loopback addresses.**

The IGP used in the service provider core should carry information only about backbone links and loopback addresses. The service provider should use BGP to carry all other information.

Use the BGP **next-hop-self** command when BGP routing is exchanged with the customer or other service providers. Using the **next-hop-self** command results in the BGP next hop being set to the loopback address of the service provider edge router and not to the access link address of the customer. The IGP can then be relieved of the burden of carrying information about the access link. The benefit of not carrying customer link information is that a flapping access link will not disturb the service provider IGP.

Scaling Service Provider Routing Protocols

This topic describes proper scaling of IGPs and BGP in service provider networks.



Scaling Service Provider Routing Protocols

IGP responsibilities

- **Carrying route to BGP next hop**
- **Providing optimal path to next hop**
- **Converging to alternate path so that BGP peering is maintained**

BGP responsibilities

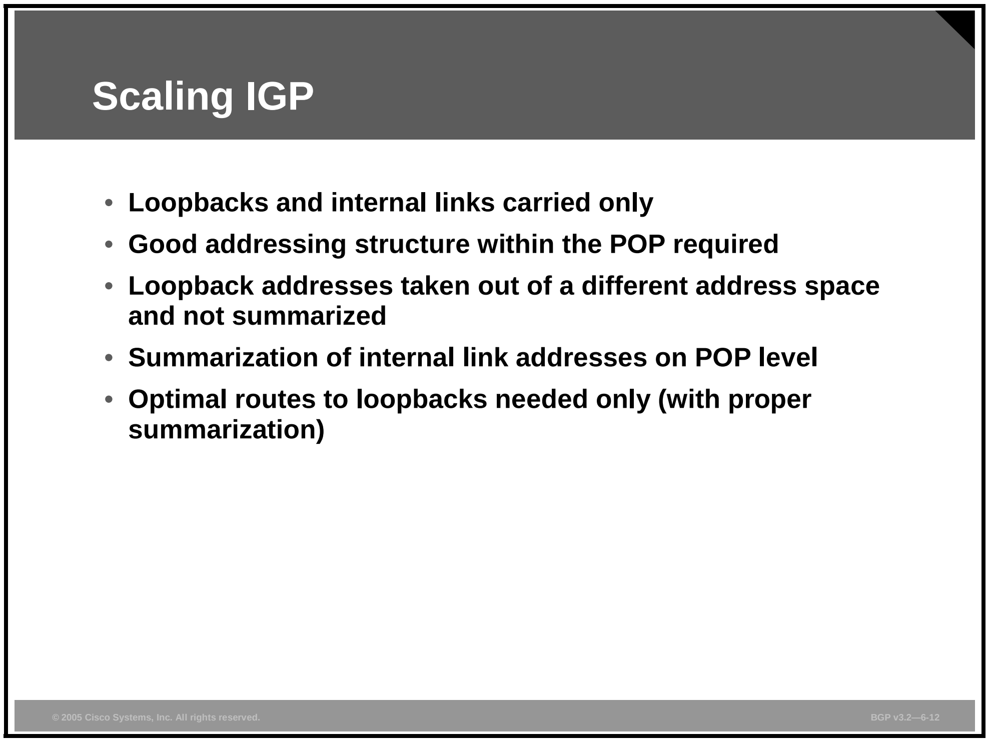
- **Generating BGP update**
- **Scaling BGP policies**
- **Scaling IBGP mesh**
- **Reducing impact of flapping routes**

The IGP is responsible for the following:

- Carrying routes to the BGP next hops to facilitate recursive routing
- Providing an optimal path to the next hop, thereby optimizing packet flow toward all BGP destinations
- Converging to an alternate path in the case of lost links or routers in a redundant network (which should be quick so that BGP sessions are not lost)

The BGP is responsible for the following:

- Generating BGP updates about reachable and unreachable networks
- Implementing and scaling the BGP routing policy, which can be quite cumbersome in large service provider networks with many EBGP-speaking routers
- Implementing and scaling IBGP sessions between all BGP-speaking routers in the AS
- Reducing the impact of individual flapping routes through route summarization



Scaling IGP

- Loopbacks and internal links carried only
- Good addressing structure within the POP required
- Loopback addresses taken out of a different address space and not summarized
- Summarization of internal link addresses on POP level
- Optimal routes to loopbacks needed only (with proper summarization)

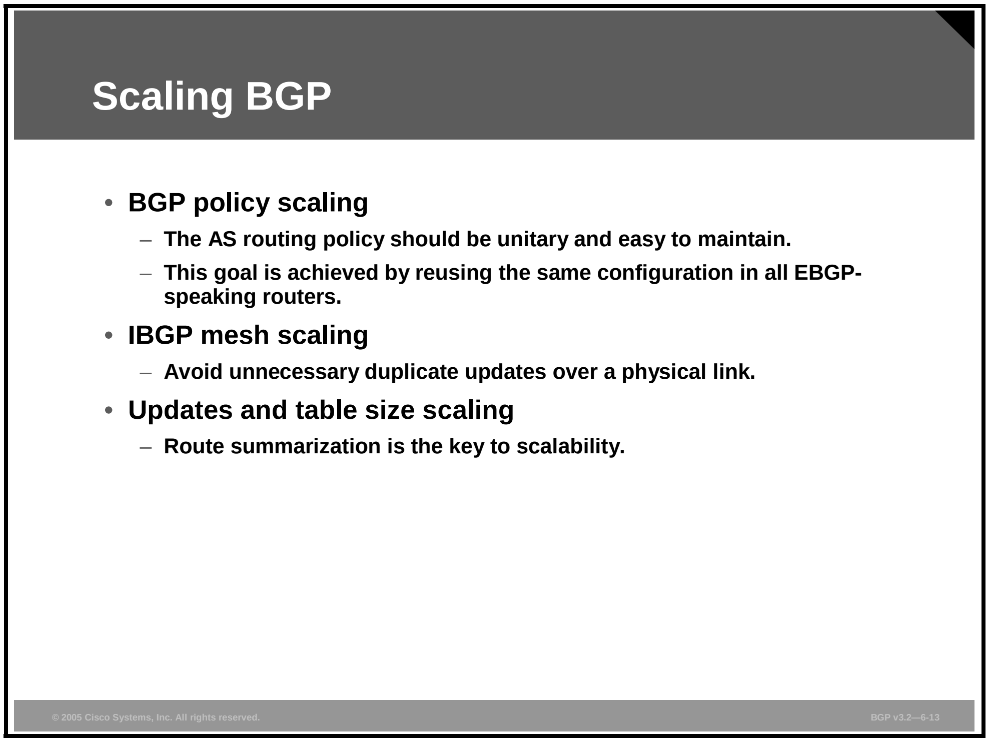
BGP v3.2—6-12

In scaling an IGP, it is important to limit the number of routes that are carried by the IGP. Optimally, the IGP carries only loopback interfaces and internal links.

The number of routes that are carried by the IGP can be even further reduced with route summarization. However, care must be taken because loopback addresses should never be summarized. Route summarization always introduces the risk of suboptimal routing and should be carefully planned, because it is important that recursive routing lookup always use optimal routing to the next hop. Also, in an MPLS environment, a label switched path (LSP) must be unbroken between edge routers, and summarizing loopback interfaces will break the LSP.

Internal links can always be summarized because they are not used as BGP next-hop addresses. To facilitate proper route summarization, internal links and loopback interfaces on a router should be assigned addresses from two different address spaces. Also, the internal links of a router should be assigned addresses depending upon which POP the routers belong to.

If implemented correctly, all internal router links in one POP can be summarized at the POP level and injected into the core as a single route. But, all router loopback addresses within the POP are still propagated into the core as individual host routes, giving optimal routing to all loopback interfaces.



Scaling BGP

- BGP policy scaling
  - The AS routing policy should be unitary and easy to maintain.
  - This goal is achieved by reusing the same configuration in all EBGP-speaking routers.
- IBGP mesh scaling
  - Avoid unnecessary duplicate updates over a physical link.
- Updates and table size scaling
  - Route summarization is the key to scalability.

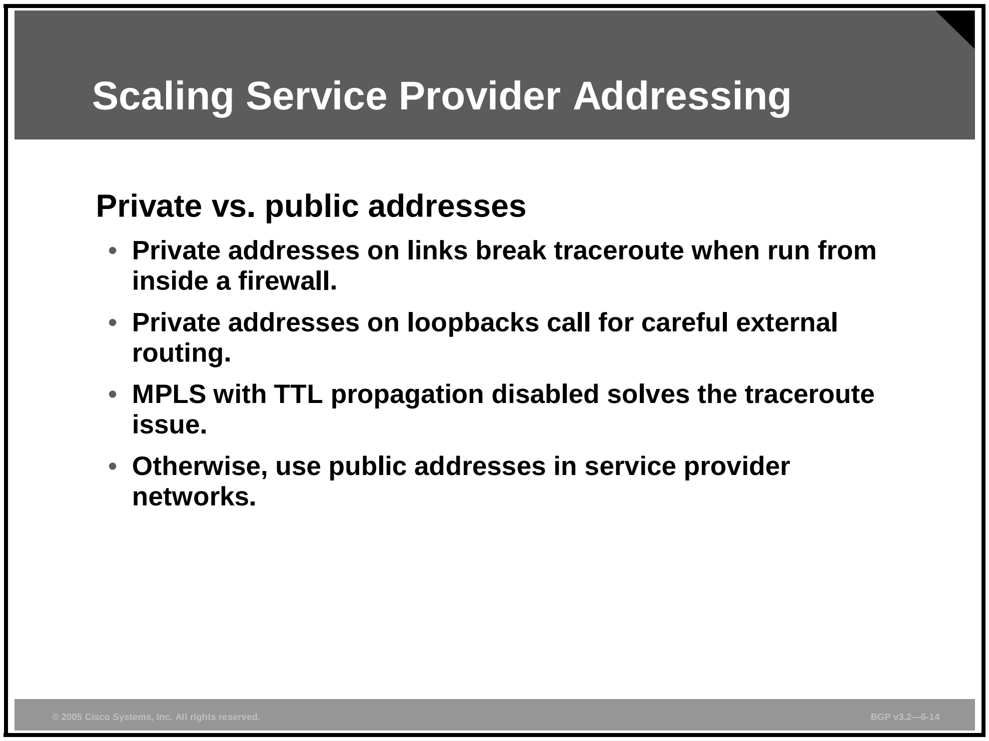
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-13

The task of scaling BGP actually involves three different and independent scaling tasks:

- **BGP policy scaling:** The AS routing policy should be unitary and easy to maintain. Different edge routers of the same AS should not use different policies and thereby advertise different routes to neighboring autonomous systems. Regardless of which router is currently active, the same routing policy should be in place. Administratively, replication of the same routing policies requires the same configuration lines in several edge routers.
- **IBGP mesh scaling:** All BGP-speaking routers must be updated with consistent IBGP information. In the traditional BGP approach, ensuring consistent routing information was achieved by establishing a full mesh of IBGP sessions between all routers within the AS. An IBGP full mesh is certainly not scalable, and several tools are now available to achieve the same results without the full mesh.
- **Updates and table size scaling:** The number of routes in the routing table and the number of updates that are sent and received represent the third scaling task. Route summarization is the key to this scalability.

Scaling Service Provider Addressing

This topic describes the scaling issues that are relevant to IP addressing in service provider networks.



Scaling Service Provider Addressing

Private vs. public addresses

- **Private addresses on links break traceroute when run from inside a firewall.**
- **Private addresses on loopbacks call for careful external routing.**
- **MPLS with TTL propagation disabled solves the traceroute issue.**
- **Otherwise, use public addresses in service provider networks.**

BGP v3.2—6-14

Using private addresses in a service provider network has some drawbacks. Private addresses on the provider internal links will cause trouble for the traceroute application. When the **traceroute** command is executed from a router inside a customer network that resides inside a firewall, the Internet Control Message Protocol (ICMP) replies that are generated by the provider router will have the source IP address assigned using the outgoing interface. If this is a private address, the customer firewall will most likely filter the packet because of address-spoofing detection rules. Even if the packet were allowed to enter the customer network, Domain Name System (DNS) reverse lookups would either fail or result in confusing printouts.

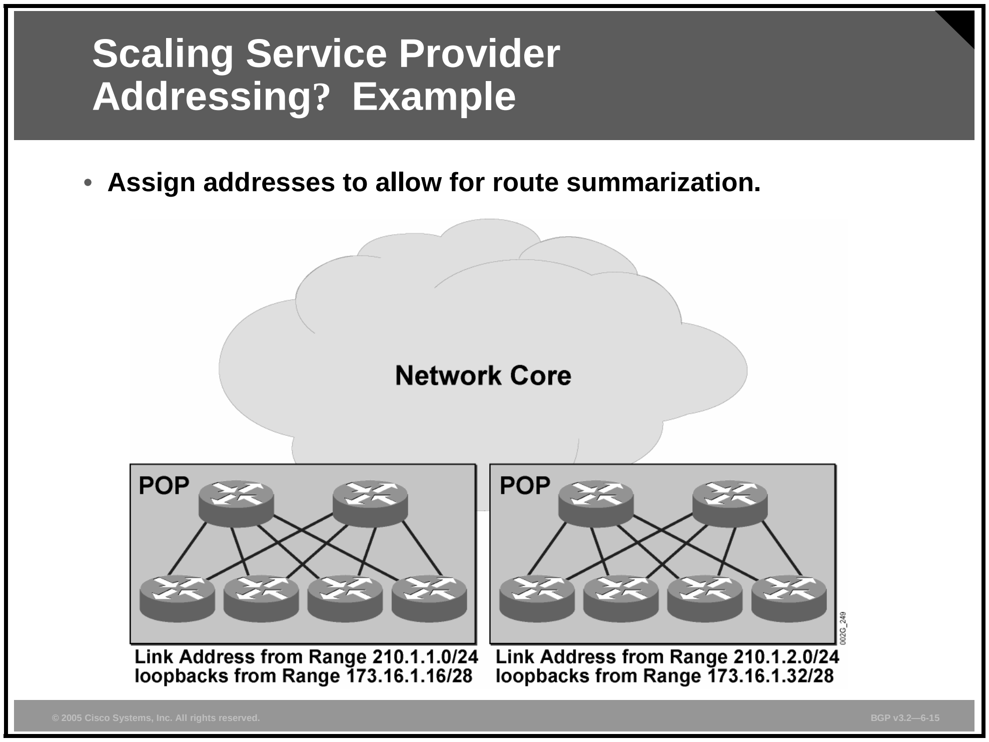
Using MPLS without Time to Live (TTL) propagation in the service provider network can easily overcome the traceroute problem with private addresses. If these functions are used, the provider network will appear as a single hop to the traceroute application. The intermediate routers will be invisible and thus can use private addresses.

Using private addresses on the service provider router loopback interfaces is possible. However, you must take care not to advertise any private addresses to any other AS.

A rule of safety is to prevent the announcement of any private addresses by using prefix-lists that are applied on outgoing updates to external neighbors. The same prefix-list mechanism can also be used on the provider edge routers to prevent accepting private addresses from any other AS if the other AS, by mistake, announces private addresses.

Example: Scaling Service Provider Addressing

This example illustrates assigning addresses to allow for route summarization.



Scaling Service Provider Addressing? Example

- Assign addresses to allow for route summarization.

Link Address from Range 210.1.1.0/24
loopbacks from Range 173.16.1.16/28

Link Address from Range 210.1.2.0/24
loopbacks from Range 173.16.1.32/28

BGP v3.2—6-15

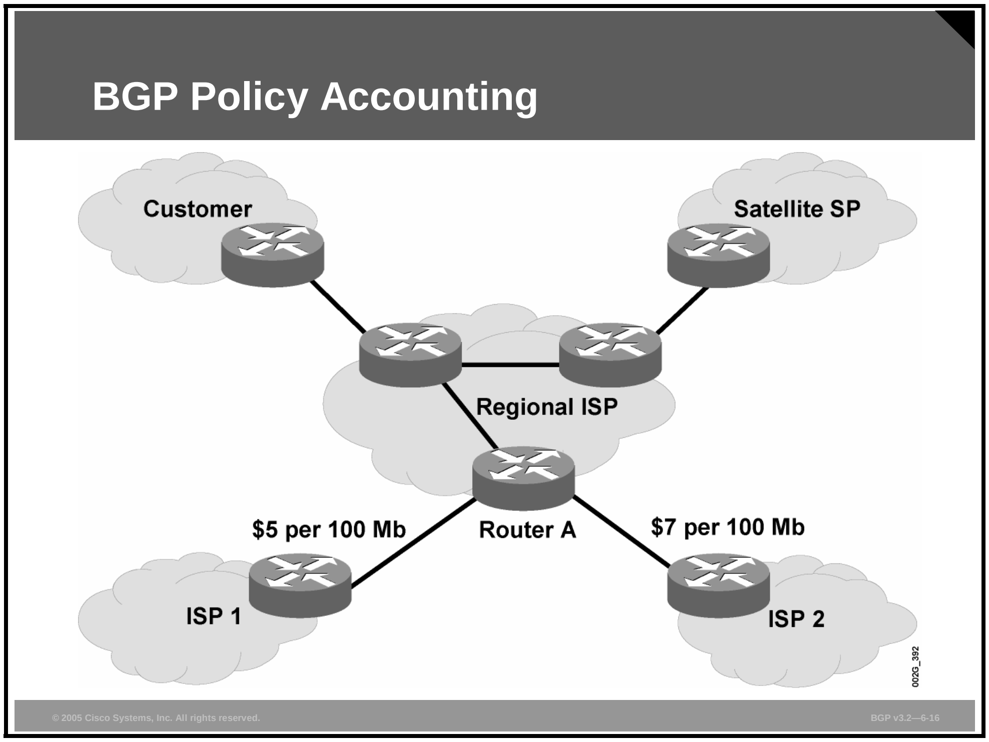
In the figure, the left POP has been allocated two different address spaces. The address space 210.1.1.0/24 has been allocated to assign addresses to internal links within the POP. The address space 173.16.1.16/28 is used to assign addresses to loopback interfaces on routers within the POP.

Likewise, the right POP has assigned 210.1.2.0/24 to internal links and 173.16.1.32/28 to be used with loopback interfaces.

The two POPs connect to the core, and, as they do, both summarize the range for their internal links while they avoid summarizing the addresses that are assigned to the loopback interfaces of the POP routers.

BGP Policy Accounting

This topic describes the function of BGP policy accounting in relation to BGP scaling.



BGP Policy Accounting

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-16

As network administrators learn to manage and scale larger and larger networks, they must also be able to account for the usage of a growing customer base. How can you ensure that customers are being charged correctly for their network utilization? How can you ensure that they are receiving the services for which they have contracted? BGP policy accounting addresses these concerns.

BGP policy accounting using AS numbers can be used to improve the design of network circuit peering and transit agreements between ISPs.

BGP policy accounting measures and classifies IP traffic that is sent to, or received from, different peers. Policy accounting is enabled on an input interface, and counters based on parameters such as community-list, AS number, or AS path are assigned to identify the IP traffic. Using BGP policy accounting, you can account for traffic according to the route that it traverses. Service providers can identify and account for all traffic by customer and can bill accordingly. In the figure, BGP policy accounting can be implemented in Router A to measure packet and byte volumes in AS buckets. Customers are billed appropriately for traffic that is routed from a domestic, international, or satellite source.

Using the BGP **table-map** command, prefixes added to the routing table are classified by BGP attribute, AS number, or AS path. Packet and byte counters are incremented per input interface. A Cisco IOS policy-based classifier maps the traffic into one of eight possible buckets, representing different traffic classes.

Implementing BGP policy accounting on an edge router can highlight potential design improvements for peering and transit agreements.

bgp-policy

To enable BGP policy accounting or policy propagation on an interface, use the **bgp-policy** command in interface configuration mode.

- **bgp-policy** {**accounting** | **ip-prec-map**}

To disable BGP policy propagation or policy accounting, use the **no** form of this command.

- **no bgp-policy** {**accounting** | **ip-prec-map**}

Syntax Description

Parameter	Description
accounting	Accounting policy based on community-lists, AS numbers, or AS paths
ip-prec-map	Quality of service (QoS) policy based on the IP precedence

set traffic-index

To indicate where to output packets that pass a match clause of a route map for BGP policy accounting, use the **set traffic-index** command in route-map configuration mode.

- **set traffic-index** *bucket-number*

To delete an entry, use the **no** form of this command.

- **no set traffic-index** *bucket-number*

Syntax Description

Parameter	Description
bucket-number	Number, in the range from 1 to 8, representing a bucket into which packet and byte statistics are collected for a specific traffic classification

The BGP Policy Accounting Output Interface Accounting feature introduces several extensions to enable BGP policy accounting (PA) on an output interface and to include accounting based on a source address for both input and output traffic on an interface. Counters based on parameters such as community-list, AS number, or AS path are assigned to identify the IP traffic.

Specifying the Match Criteria for BGP Policy Accounting: Example

In the following example, BGP communities are specified in community-lists, and a route-map named “set_bucket” is configured to match each of the community-lists to a specific accounting bucket using the **set traffic-index** command:

```
ip community-list 30 permit 100:190
ip community-list 40 permit 100:198
ip community-list 50 permit 100:197
ip community-list 60 permit 100:296
!
route-map set_bucket permit 10
 match community-list 30
 set traffic-index 2
!
route-map set_bucket permit 20
 match community-list 40
 set traffic-index 3
!
route-map set_bucket permit 30
 match community-list 50
 set traffic-index 4
!
route-map set_bucket permit 40
 match community-list 60
 set traffic-index 5
```

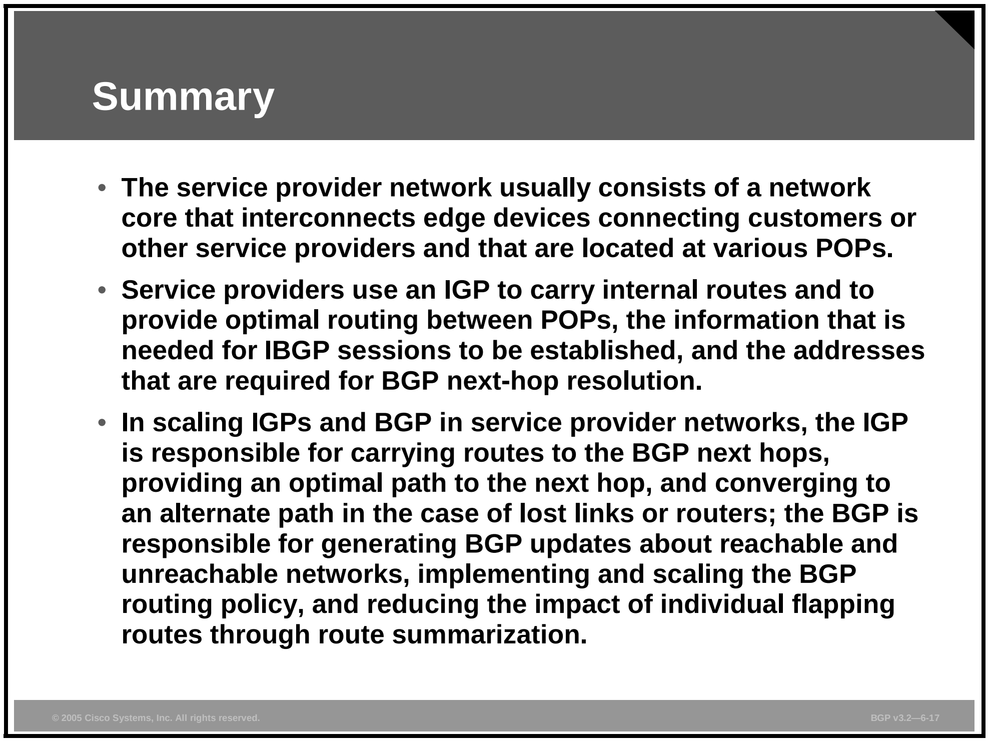
Classifying the IP Traffic and Enabling BGP Policy Accounting: Example

In the following example, BGP policy accounting is enabled on packet over SONET (POS) interface 7/0. The policy accounting criteria are based on the source address of the input traffic, and the **table-map** command is used to modify the bucket number when the IP routing table is updated with routes learned from BGP.

```
router bgp 65000
 table-map set_bucket
 network 10.15.1.0 mask 255.255.255.0
 neighbor 10.14.1.1 remote-as 65100
!
ip classless
ip bgp-community new-format
!
interface POS7/0
 ip address 10.15.1.2 255.255.255.0
 bgp-policy accounting input source
 no keepalive
 crc 32
 clock source internal
```

Summary

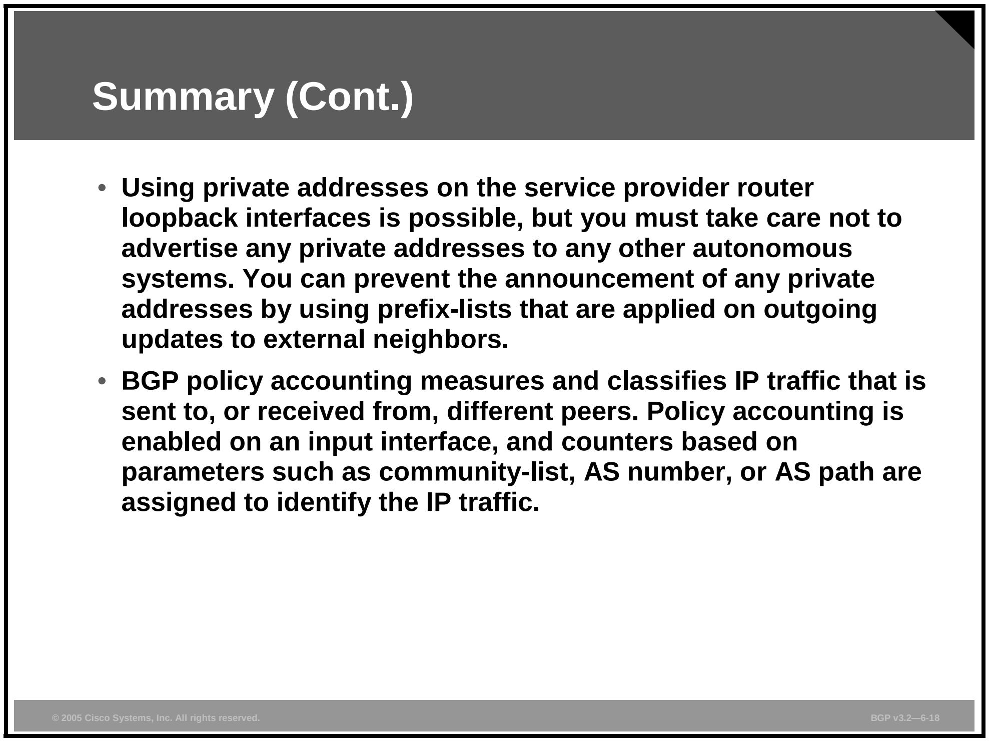
This topic summarizes the key points discussed in this lesson.



Summary

- The service provider network usually consists of a network core that interconnects edge devices connecting customers or other service providers and that are located at various POPs.
- Service providers use an IGP to carry internal routes and to provide optimal routing between POPs, the information that is needed for IBGP sessions to be established, and the addresses that are required for BGP next-hop resolution.
- In scaling IGPs and BGP in service provider networks, the IGP is responsible for carrying routes to the BGP next hops, providing an optimal path to the next hop, and converging to an alternate path in the case of lost links or routers; the BGP is responsible for generating BGP updates about reachable and unreachable networks, implementing and scaling the BGP routing policy, and reducing the impact of individual flapping routes through route summarization.

BGP v3.2—6-17



Summary (Cont.)

- Using private addresses on the service provider router loopback interfaces is possible, but you must take care not to advertise any private addresses to any other autonomous systems. You can prevent the announcement of any private addresses by using prefix-lists that are applied on outgoing updates to external neighbors.
- BGP policy accounting measures and classifies IP traffic that is sent to, or received from, different peers. Policy accounting is enabled on an input interface, and counters based on parameters such as community-list, AS number, or AS path are assigned to identify the IP traffic.

BGP v3.2—6-18

Lesson 2

Introducing Route Reflectors

Overview

Large Border Gateway Protocol (BGP) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors and confederations. Route reflectors enable BGP routing information to be distributed in a fashion that does not require a physical fully meshed network. Network overhead is reduced by decreasing the number of TCP connections that are required to distribute routing information and by lessening router CPU and memory requirements.

This lesson introduces BGP route reflectors by explaining why they improve BGP scalability. Modified split-horizon rules, applied when you are using route reflectors, are also discussed. The lesson concludes by describing the various redundancy mechanisms that are used with route reflectors, including route reflector clusters.

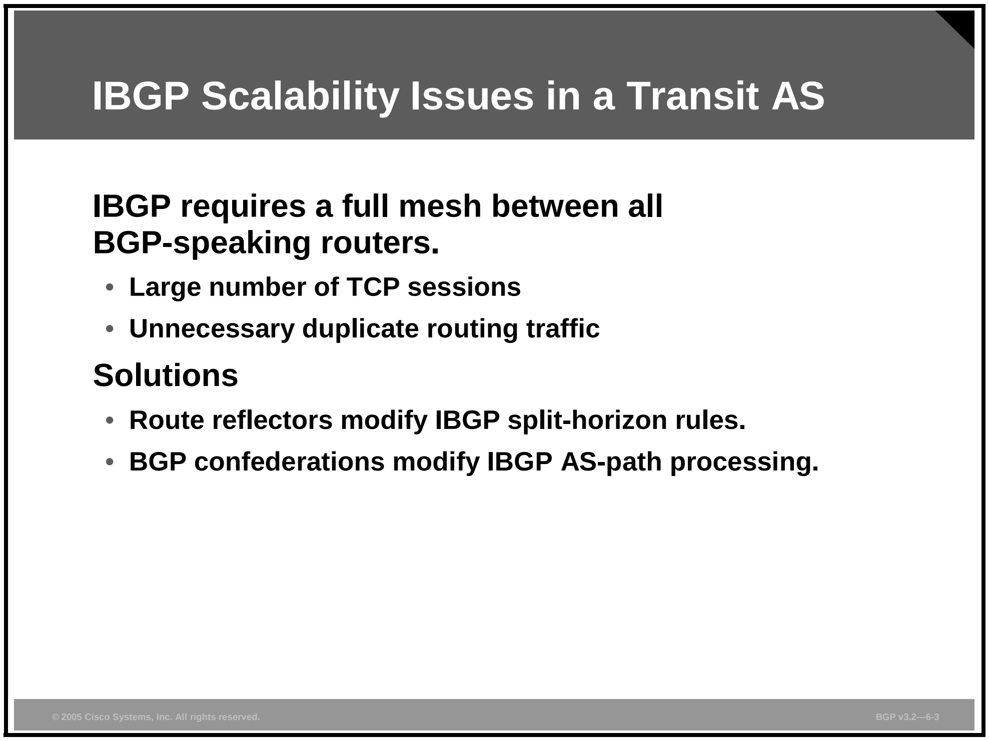
Objectives

Upon completing this lesson, you will be able to describe the function of route reflectors in a BGP environment. This ability includes being able to meet these objectives:

- Explain the need for BGP route reflectors in BGP transit backbones
- Explain how route reflectors modify traditional IBGP split-horizon rules
- Explain the benefits of deploying redundant route reflectors
- Explain how route reflector clusters prevent loops in the deployment of route reflectors in redundant configurations
- Describe additional route reflector mechanisms that have been designed to prevent routing loops

IBGP Scalability Issues in a Transit AS

This topic explains the need for BGP route reflectors by describing the scalability issues of BGP transit backbones.



IBGP Scalability Issues in a Transit AS

IBGP requires a full mesh between all BGP-speaking routers.

- **Large number of TCP sessions**
- **Unnecessary duplicate routing traffic**

Solutions

- **Route reflectors modify IBGP split-horizon rules.**
- **BGP confederations modify IBGP AS-path processing.**

BGP v3.2—6-3

Classic Internal Border Gateway Protocol (IBGP) split-horizon rules specify that updates that are received on an External Border Gateway Protocol (EBGP) session should be forwarded on all IBGP and EBGP sessions, but updates that are received on an IBGP session should be forwarded only to all EBGP sessions. This rule requires a BGP boundary router to be able to send routing updates to all other BGP-speaking routers in its own autonomous system (AS) directly through a separate IBGP session to each of them.

The primary reason for the IBGP split-horizon rule is to avoid routing information loops within the AS. If the information that is received through an IBGP session is forwarded on other IBGP sessions, the information might come back to the originator and be forwarded again in a never-ending loop. The originator would not detect the loop because no BGP attributes are changed on IBGP sessions.

The general design rule in classic IBGP is to have a full mesh of IBGP sessions. But a full mesh of IBGP sessions between *n* number of routers would require $(n * (n - 1)) / 2$ IBGP sessions. For example, a router with an AS that contains 10 routers would require $(10 * (10 - 1)) / 2 = 45$ IBGP sessions. Imagine the number of sessions (and the associated router configuration) that would be required for a single AS containing 500 routers.

Every IBGP session uses a single TCP session to another IBGP peer. An update that must be sent to all IBGP peers must be sent on each of the individual TCP sessions. If a router is attached to the rest of the network over just a single link, this single link has to carry all TCP/IP packets for all IBGP sessions. This requirement results in multiplication of the update over the single link.

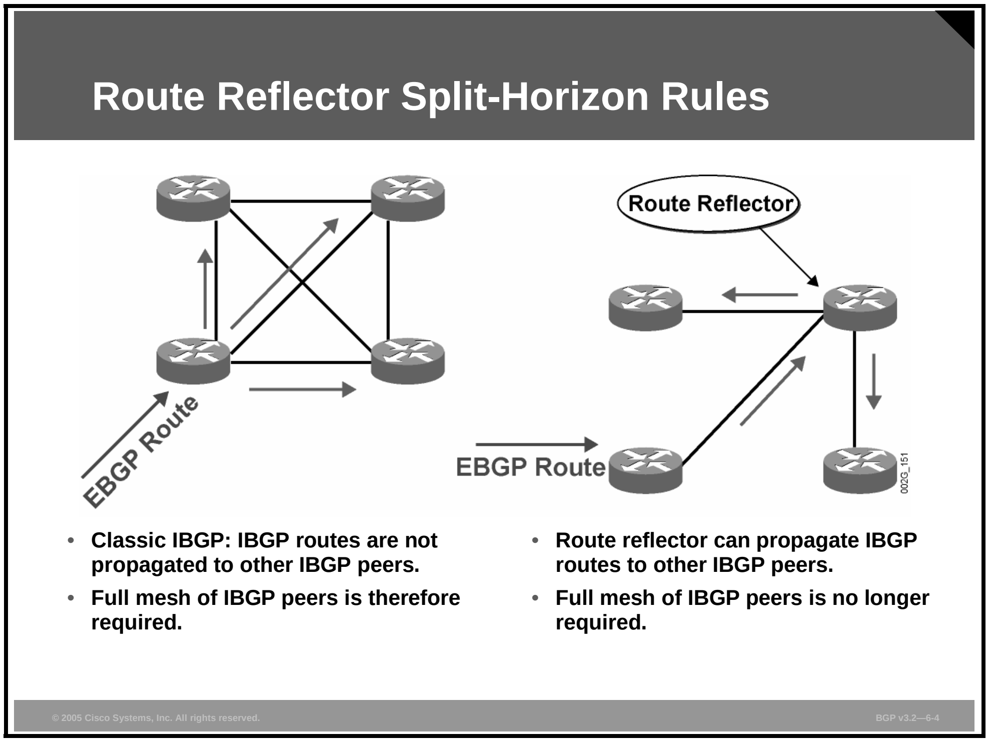
Route reflectors are BGP scalability mechanisms that enable routing information to be redistributed to all routers within an AS while eliminating the need for a fully meshed topology within the AS. This feature reduces the number of TCP sessions that must be maintained, lowering network overhead and CPU and memory resource requirements.

Two different solutions are available to achieve greater scalability when you are faced with the full-mesh rules of IBGP autonomous systems:

- Route reflectors modify the classic IBGP split-horizon rule and allow a particular router to forward incoming IBGP updates to an outgoing IBGP session under certain conditions. This router becomes a concentration router, or a route reflector.
- BGP confederations (covered in a separate lesson) introduce the concept of a number of smaller autonomous systems within the original AS. The small autonomous systems exchange BGP updates between them using intra-confederation EBGP sessions.

Route Reflector Split-Horizon Rules

This topic explains how route reflectors modify traditional IBGP split-horizon rules.



Route Reflector Split-Horizon Rules

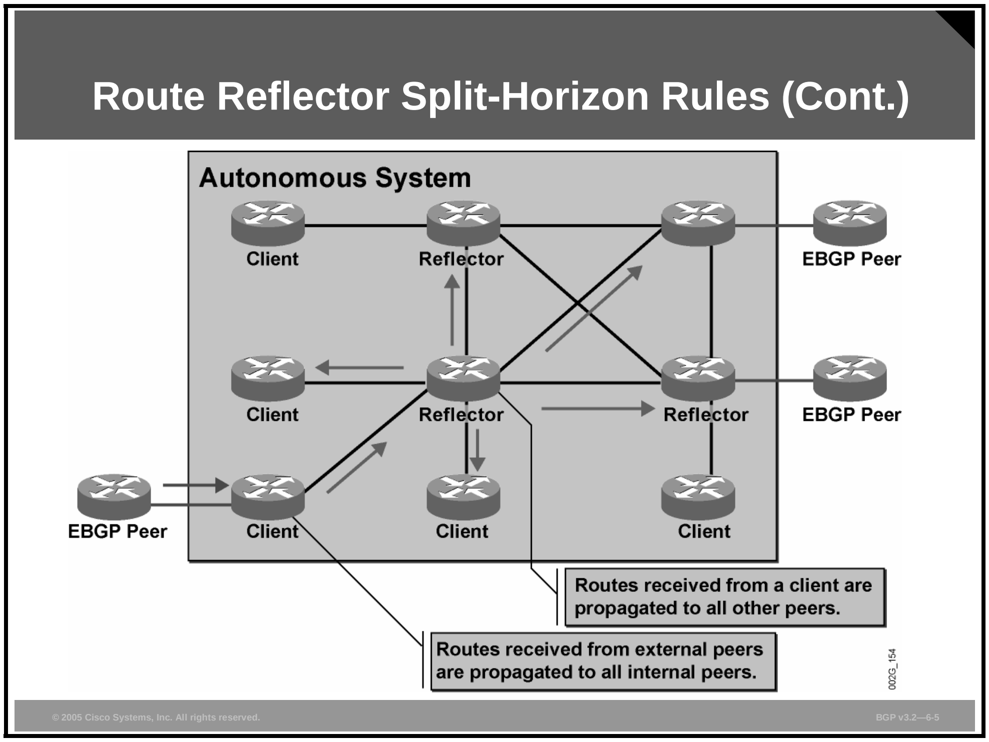
- Classic IBGP: IBGP routes are not propagated to other IBGP peers.
- Full mesh of IBGP peers is therefore required.

- Route reflector can propagate IBGP routes to other IBGP peers.
- Full mesh of IBGP peers is no longer required.

In classic IBGP, the BGP boundary router needs to forward the route that is received from an EBGP peer to every other router within its own AS using a dedicated IBGP session for each one. Also, the BGP boundary router forwards routes that are sourced by a router in the same way. To allow every router to update every other router, a full mesh of IBGP sessions is required.

The IBGP route reflector design relaxes the need for a full mesh. The router configured as a route reflector, under certain conditions, will relay updates that are received through an IBGP session to another IBGP session. This capability requires modifications of the classic IBGP split-horizon rules.

The route reflector concept introduces processing overhead on the concentration router and, if it is configured incorrectly, can cause routing loops and instability.



Route Reflector Split-Horizon Rules (Cont.)

Routes received from a client are propagated to all other peers.

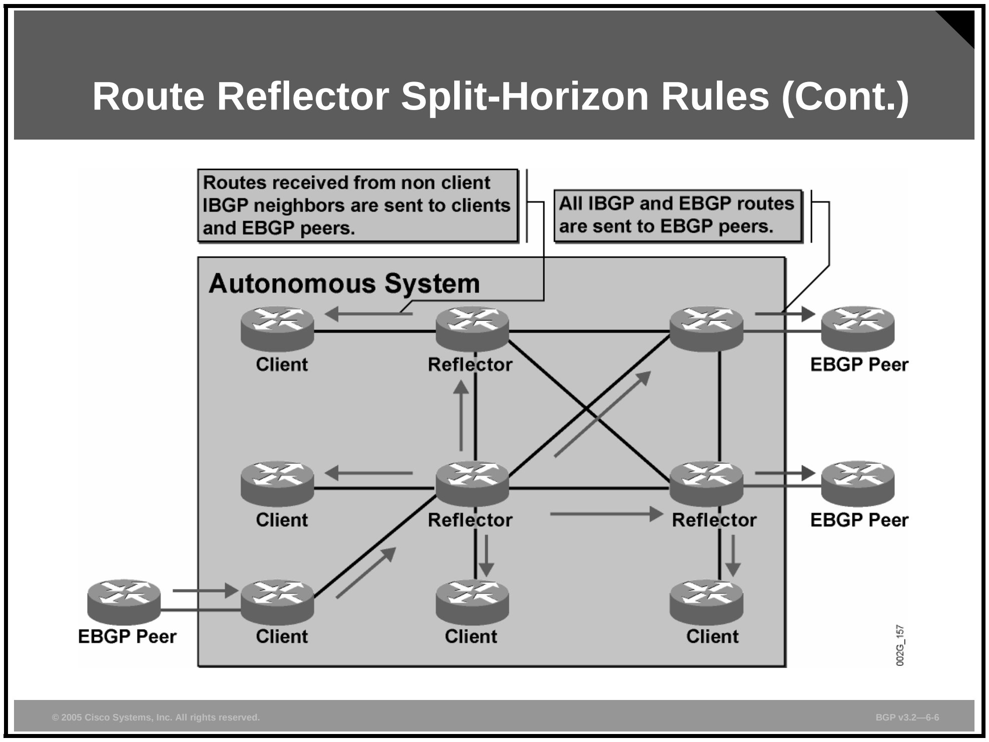
Routes received from external peers are propagated to all internal peers.

When you implement a route-reflector-based IBGP network, the BGP routers are divided into route reflectors (which implement modified split-horizon rules) and clients (which are behaving like traditional IBGP routers).

Route reflector clients are excluded from the full mesh. They can have any number of EBGP sessions but may have only one IBGP session, the session with their route reflector. Clients conform to the classic IBGP split-horizon rules and forward a received route from EBGP on their IBGP neighbor sessions. But the route reflector conforms to the route reflector split-horizon rules and recognizes that it has an IBGP session to a client. When the IBGP update is received from the client, the route reflector forwards the update to other IBGP neighbors, therefore alleviating the IBGP full-mesh requirement for its clients.

Similarly, when the route reflector receives an IBGP update from a neighbor that is not its client, it forwards the update to all of its clients.

Forwarding of an IBGP update in a route reflector does not change the next-hop attribute or any other common BGP attribute. This feature means that the client will use the optimum route by means of recursive routing, regardless of the way that it has received the BGP route.



Route Reflector Split-Horizon Rules (Cont.)

Routes received from non client IBGP neighbors are sent to clients and EBGP peers.

All IBGP and EBGP routes are sent to EBGP peers.

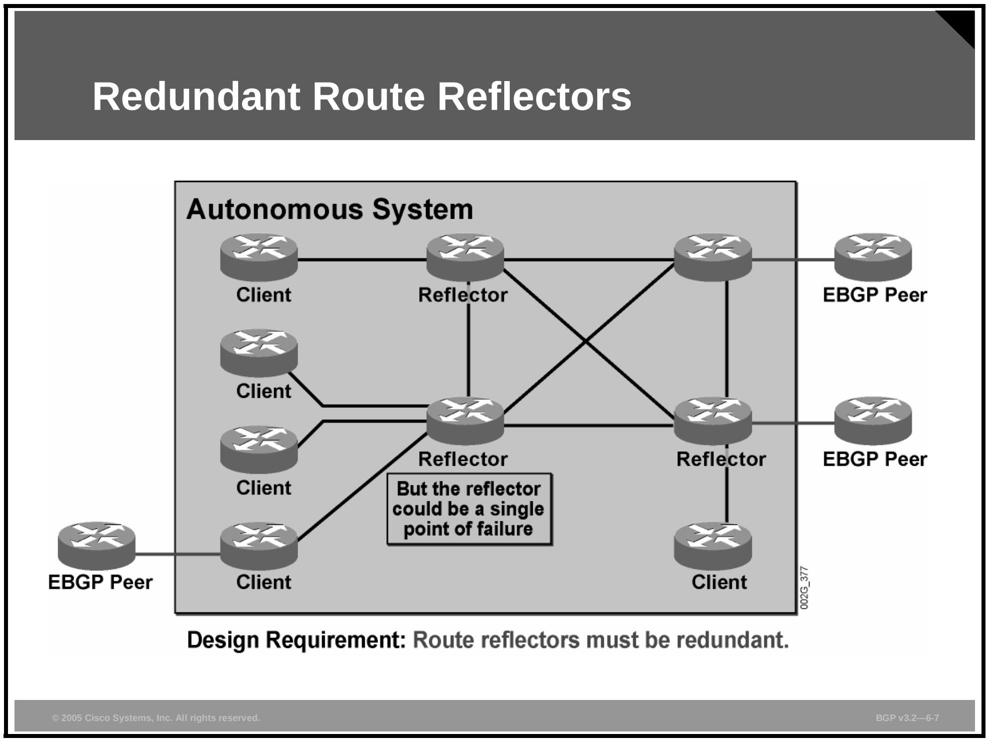
The figure shows how an AS with nine routers running BGP reduces the required number of IBGP TCP sessions from 36 to 11 by using route reflectors.

The table presents detailed IBGP split-horizon rules as modified by the introduction of BGP route reflectors. For purposes of definition, a “route reflector” is a BGP speaker that can advertise IBGP learned routers to another IBGP peer and, hence, can reflect routes. IBGP peers of the route reflector fall under two categories: “clients” and “nonclients.” The route reflector and its clients form a “cluster.” All IBGP peers of the route reflector that are not part of the cluster are nonclients. A “classic” IBGP router is a router that does not support route reflector functionality.

Type of Router	Incoming Update From	Is Forwarded To
Classic	EBGP peer	All peers (IBGP and EBGP)
	IBGP peer	EBGP peers
Route reflector	EBGP peer	All peers (IBGP and EBGP)
	Nonclient IBGP peer	EBGP peers and clients
	Client IBGP peer	All peers but the sender
Client	EBGP peer	All peers (IBGP and EBGP)
	IBGP peer	EBGP peers

Redundant Route Reflectors

This topic explains the benefits of deploying redundant route reflectors.



Redundant Route Reflectors

But the reflector could be a single point of failure

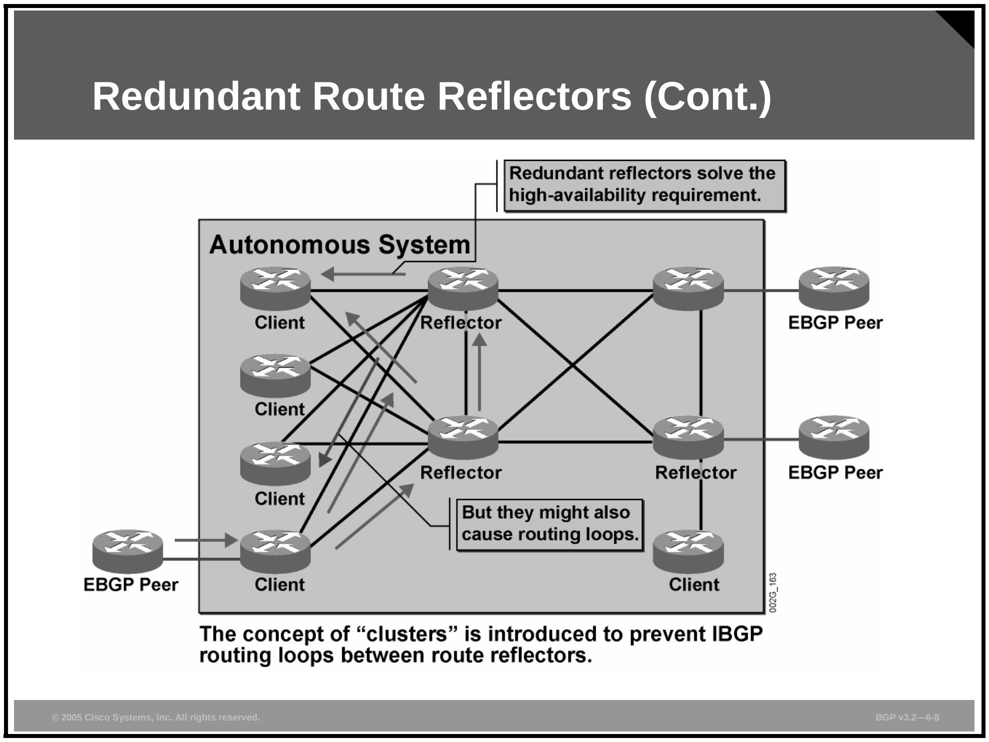
Design Requirement: Route reflectors must be redundant.

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—6-7

Clients may have any number of EBGP peers but may have IBGP sessions only with their route reflector or reflectors. If the reflector fails, its client can no longer send BGP updates to, or receive them from, the rest of the AS. The route reflector is, therefore, a single point of failure.

To avoid introducing a single point of failure into the network, the route reflector functionality must be as redundant as the physical network. If a client will still be physically attached to the network after its route reflector has failed, the client should have a redundant route reflector. Thus, in all highly available networks, route reflectors must be redundant.



Redundant Route Reflectors (Cont.)

Redundant reflectors solve the high-availability requirement.

But they might also cause routing loops.

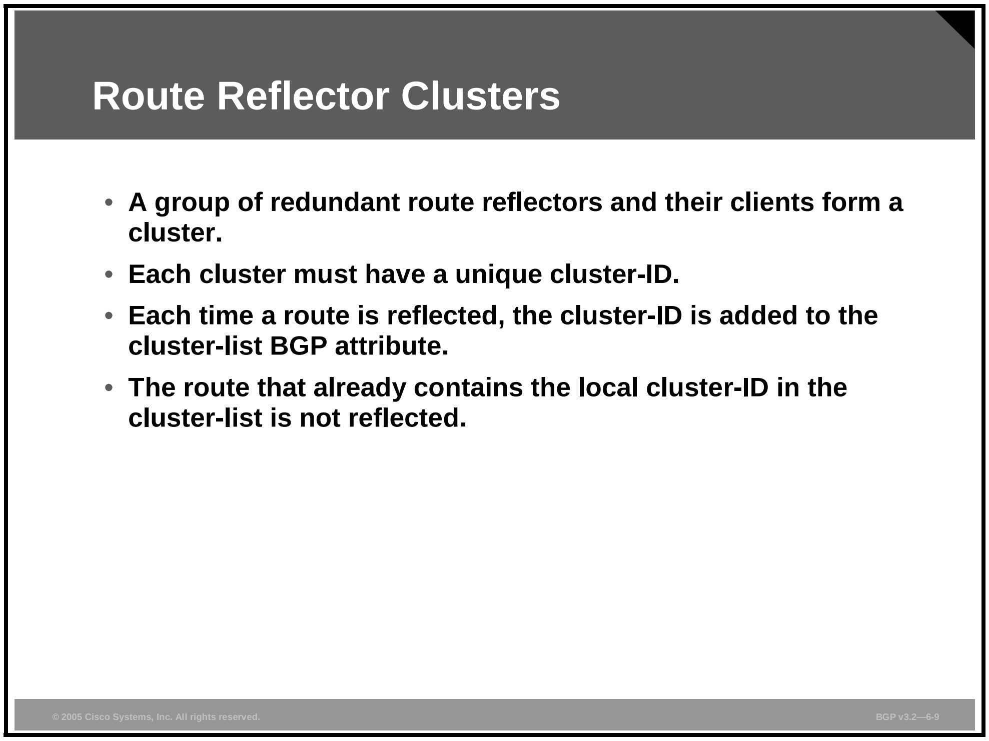
The concept of “clusters” is introduced to prevent IBGP routing loops between route reflectors.

BGP v3.2—6-8

A client may have IBGP sessions to more than one route reflector to avoid a single point of failure. Each client will receive the same route from both of its reflectors. Both route reflectors will receive the same IBGP update from their client, and they will both reflect the update to the rest of the clients. Additionally, both route reflectors will get updated from the full mesh and reflect those updates to their clients. As a result, each client will get two copies of all routes. Under certain circumstances (particularly when you use weights on IBGP sessions to influence BGP route selection), improper route reflection can result in an IBGP routing loop that is impossible to detect. Additional BGP attributes are thus necessary to prevent these routing loops.

Route Reflector Clusters

This topic explains how route reflector clusters prevent loops in the deployment of route reflectors in redundant configurations.



Route Reflector Clusters

- **A group of redundant route reflectors and their clients form a cluster.**
- **Each cluster must have a unique cluster-ID.**
- **Each time a route is reflected, the cluster-ID is added to the cluster-list BGP attribute.**
- **The route that already contains the local cluster-ID in the cluster-list is not reflected.**

A router that is acting as a route reflector client does not require any specific configuration. It simply has fewer IBGP sessions than it would have if it were part of the full mesh. But improperly configuring the client to also be a reflector could easily cause a loop. An IBGP route coming in from one of the real reflectors to the client could be forwarded by the client, erroneously acting as reflector, to the other reflector.

Route reflector clusters prevent IBGP routing loops in redundant route reflector designs.

The role of the network designer is to properly identify which route reflectors and their clients will form a cluster. The designer assigns to the cluster a cluster-ID number that is unique within the AS.

Note The cluster-ID number must be configured in the route reflectors. The clients should not be configured with this information.

A route reflector router can reflect routes only within a single cluster. A route reflector can, however, participate in another cluster but only as a client. A client can function as a client only to a route reflector belonging to the same cluster.

When a route is reflected, the reflector creates the cluster-list attribute and attaches it to the route if it does not already exist. It then sets its cluster-ID number in the cluster-list or adds its cluster-ID number to an already existing cluster-list attribute. If the route, for any reason, is ever reflected back to the same reflector, it will recognize its cluster-ID number in the cluster-list and not forward it again. The first route reflector that reflects the route also sets an additional BGP attribute, called “originator-ID,” and adds it to the BGP router-ID of its client.

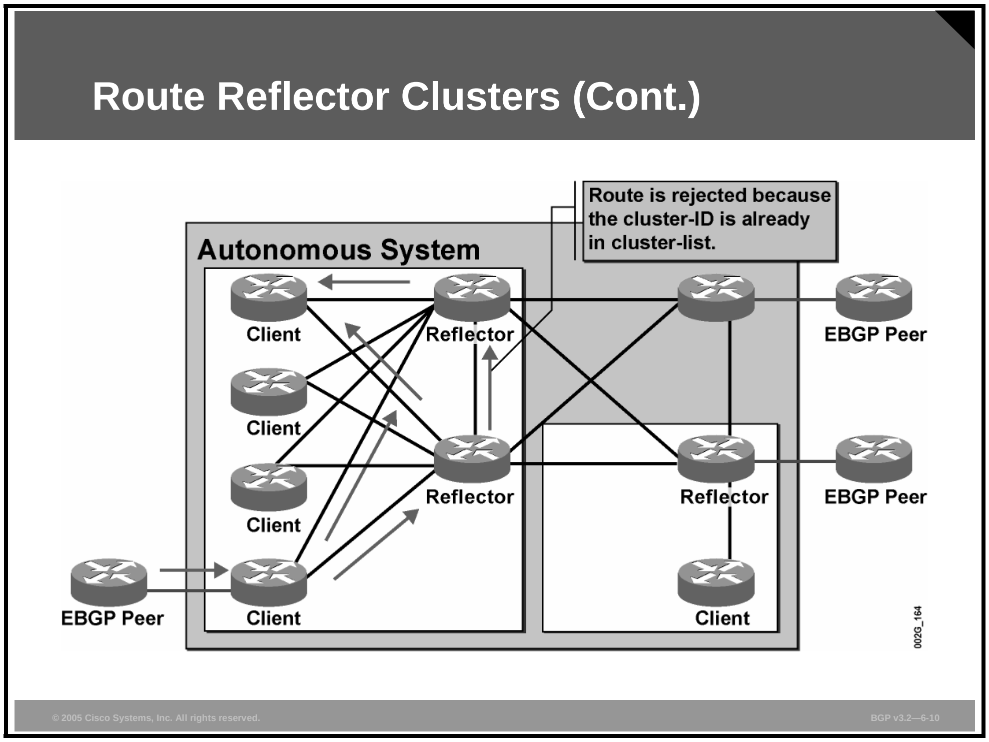
Note The cluster-list and originator-ID attributes are nontransitive optional BGP attributes, allowing routers that do not support route reflector functionality to coexist with route reflectors and their clients in the same AS.

Based on cluster-list and originator-ID attributes, routers can implement two loop-prevention mechanisms:

- Any router that receives an IBGP update with the originator-ID attribute set to its own BGP router-ID will ignore that update.
- Any route reflector that receives an IBGP update with its cluster-ID already in the cluster-list will ignore that update.

Example: Route Reflector Clusters

The figure shows a cluster with redundant route reflectors.



Route Reflector Clusters (Cont.)

Route is rejected because the cluster-ID is already in cluster-list.

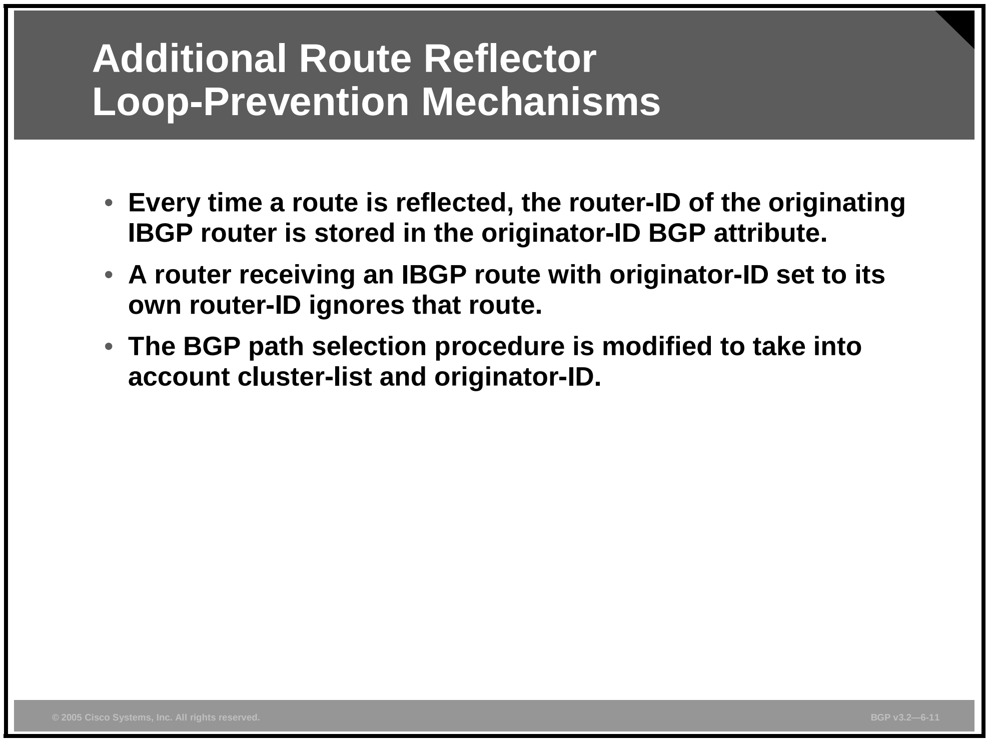
© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—6-10

The client in the cluster forwards the received EBGP update to both reflectors. The route reflectors forward the update into the IBGP full mesh. This behavior means that they send the update to each other as well. But when a route reflector receives a BGP update from another route reflector, it recognizes their common cluster-ID number in the cluster-list attribute. Therefore, the newly received route update is ignored.

Additional Route Reflector Loop-Prevention Mechanisms

This topic describes additional route reflector mechanisms designed to prevent routing loops.



Additional Route Reflector Loop-Prevention Mechanisms

- **Every time a route is reflected, the router-ID of the originating IBGP router is stored in the originator-ID BGP attribute.**
- **A router receiving an IBGP route with originator-ID set to its own router-ID ignores that route.**
- **The BGP path selection procedure is modified to take into account cluster-list and originator-ID.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-11

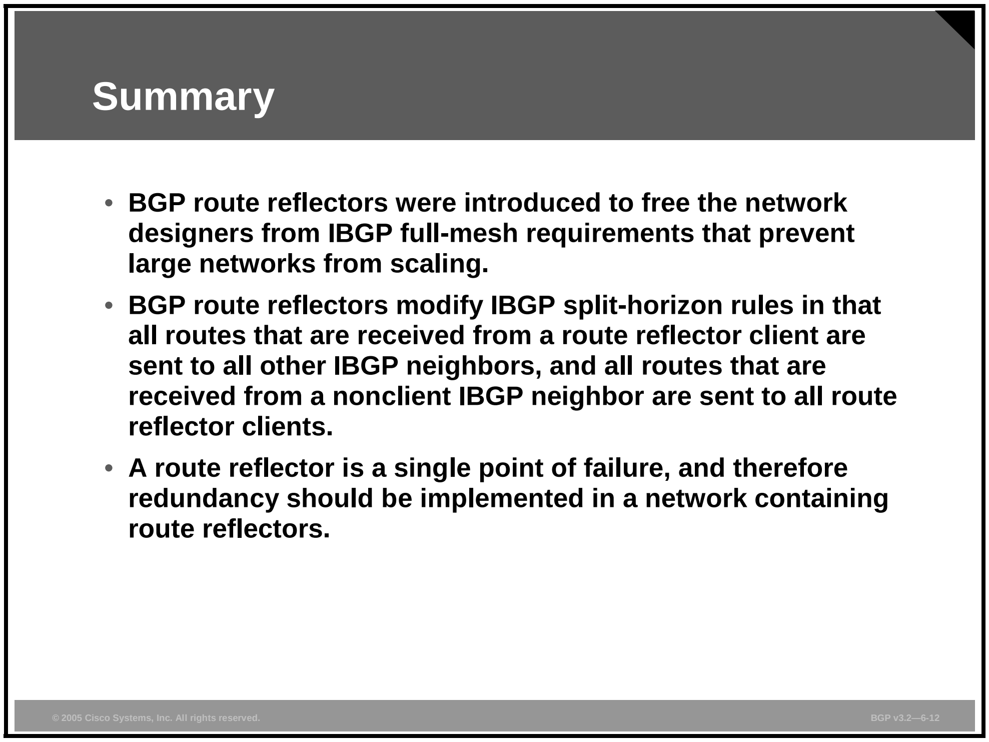
When a route is reflected, the route reflector sets the originator-ID BGP attribute (nontransitive optional BGP attribute) to the router-ID of the peer from which it received the route. Any router that receives a route with its own router-ID in the originator-ID attribute silently ignores that route.

BGP path selection rules have been modified to select the best route in scenarios where a router might receive reflected and nonreflected routes or several reflected routes:

- The traditional BGP path selection parameters—such as weight, local preference, origin, and multi-exit discriminator (MED)—are compared first.
- If these parameters are equal, the routes that are received from EBGP neighbors are preferred over routes that are received from IBGP neighbors.
- When a router receives two IBGP routes, the nonreflected routes (routes with no originator-ID attribute) are preferred over reflected routes.
- The reflected routes with shorter cluster-lists are preferred over routes with longer cluster-lists.
- If the additional route-reflector-oriented selection criteria do not yield a decision, the rest of the traditional BGP path selection rules are followed.

Summary

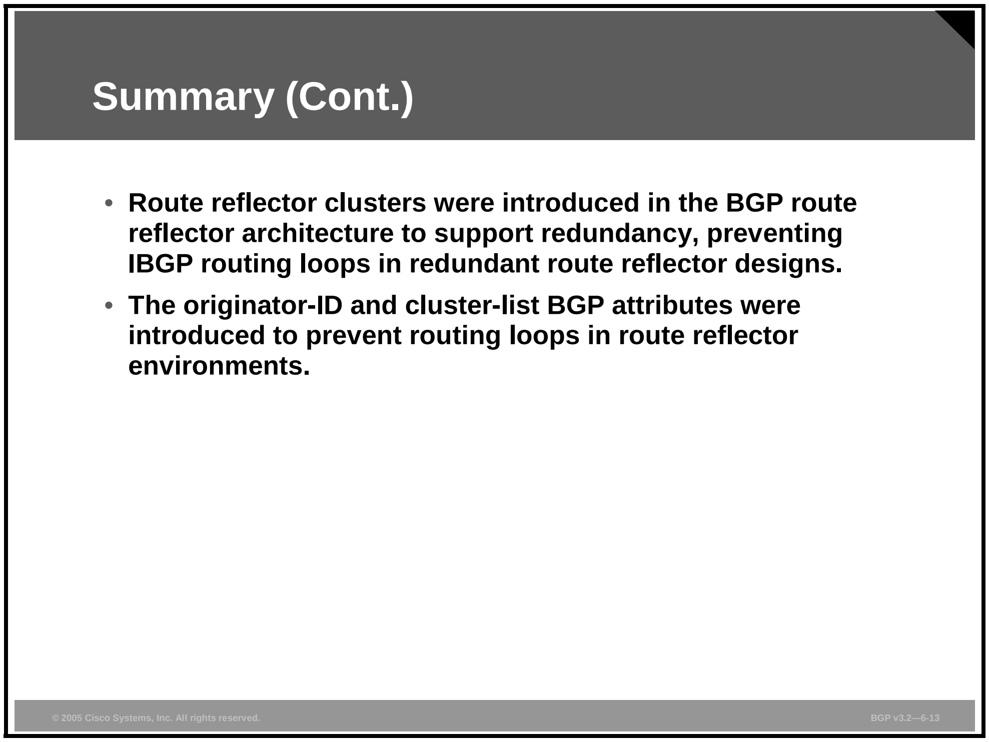
This topic summarizes the key points discussed in this lesson.



Summary

- BGP route reflectors were introduced to free the network designers from IBGP full-mesh requirements that prevent large networks from scaling.
- BGP route reflectors modify IBGP split-horizon rules in that all routes that are received from a route reflector client are sent to all other IBGP neighbors, and all routes that are received from a nonclient IBGP neighbor are sent to all route reflector clients.
- A route reflector is a single point of failure, and therefore redundancy should be implemented in a network containing route reflectors.

BGP v3.2—6-12



Summary (Cont.)

- Route reflector clusters were introduced in the BGP route reflector architecture to support redundancy, preventing IBGP routing loops in redundant route reflector designs.
- The originator-ID and cluster-list BGP attributes were introduced to prevent routing loops in route reflector environments.

BGP v3.2—6-13

Lesson 3

Designing Networks with Route Reflectors

Overview

Large Border Gateway Protocol (BGP) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors and confederations. Route reflectors are a BGP scalability mechanism that enables routing information to be redistributed to all routers within an autonomous system (AS) while eliminating the need for a fully meshed topology within the AS. Properly implementing these features requires careful network design within the AS.

This lesson introduces the network design rules that network designers should follow when implementing a network with BGP route reflectors. It also lists the potential issues that can arise if the network design rules are not adhered to. The lesson concludes by describing the concept of hierarchical route reflectors.

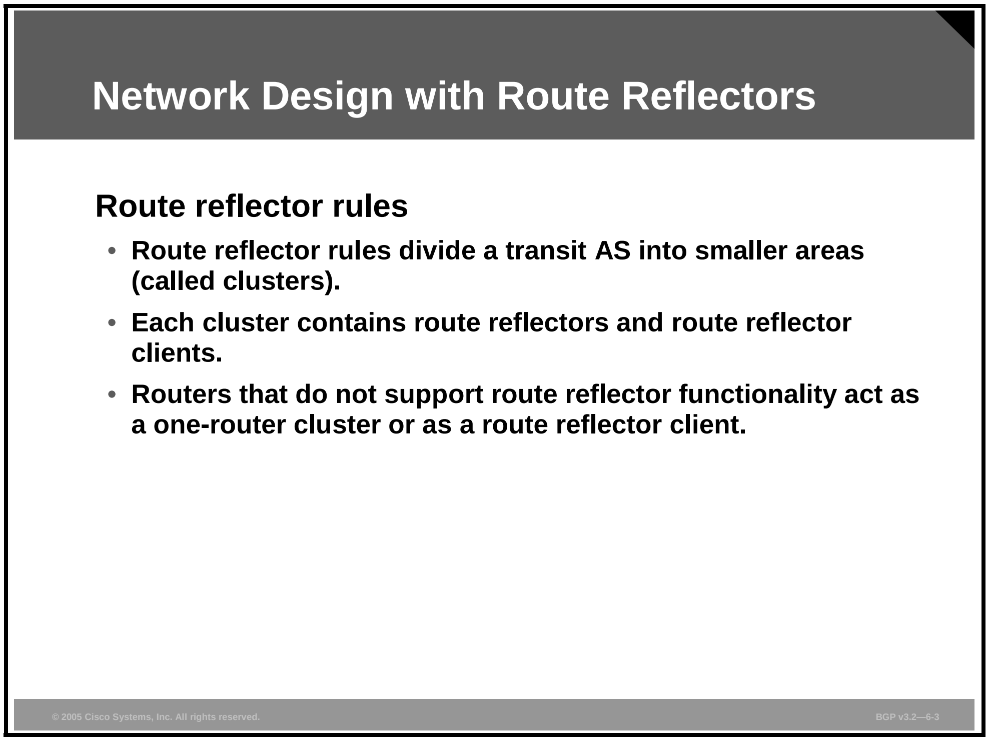
Objectives

Upon completing this lesson, you will be able to describe the function of hierarchical route reflectors, based on established route reflector design rules. This ability includes being able to meet these objectives:

- List the network design rules for implementing BGP route reflectors
- List the potential issues that can arise if you do not follow the route reflector network design rules
- Explain the function of hierarchical route reflectors

Network Design with Route Reflectors

This topic lists the network design rules for implementing BGP route reflectors and Internal Border Gateway Protocol (IBGP) sessions.



Network Design with Route Reflectors

Route reflector rules

- **Route reflector rules divide a transit AS into smaller areas (called clusters).**
- **Each cluster contains route reflectors and route reflector clients.**
- **Routers that do not support route reflector functionality act as a one-router cluster or as a route reflector client.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-3

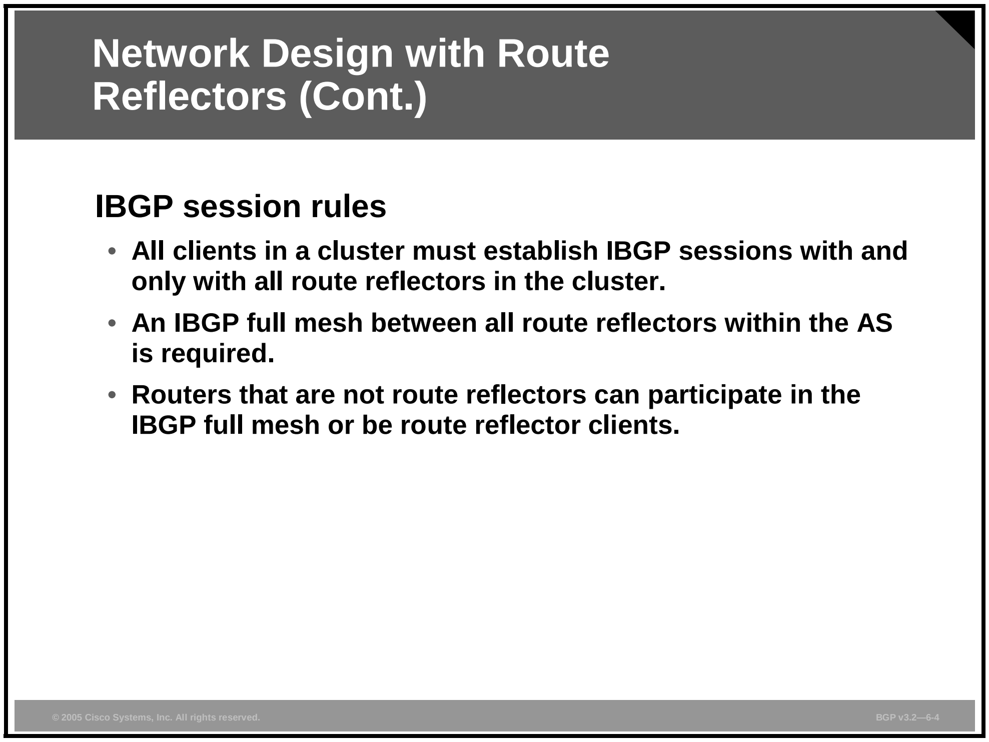
The physical topology of the network could serve as a guide to route reflector design.

Implementing route reflectors within the transit AS will create smaller areas (or groups) of routers. These smaller groupings of routers are called clusters. A cluster consists of route reflector routers, either redundant or nonredundant, and the client routers that are connected to them.

In designing the implementation of route reflectors within a transit AS, identify a group of peripheral routers that are physically connected to the same backbone router or routers. Consider the peripheral routers as clients and the backbone routers as route reflectors. Then, consider this group of routers together to form a cluster.

Note Additional design examples and rules are available in the module "BGP Transit Autonomous Systems."

Only the routers that are configured as route reflectors require a Cisco IOS software version with route reflector functionality. A router lacking this functionality in its installed Cisco IOS software can function as a client or be a part of the full mesh. Normally, this situation is not a concern because route reflector functionality has been incorporated in Cisco IOS software since Release 11.1.



Network Design with Route Reflectors (Cont.)

IBGP session rules

- **All clients in a cluster must establish IBGP sessions with and only with all route reflectors in the cluster.**
- **An IBGP full mesh between all route reflectors within the AS is required.**
- **Routers that are not route reflectors can participate in the IBGP full mesh or be route reflector clients.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-4

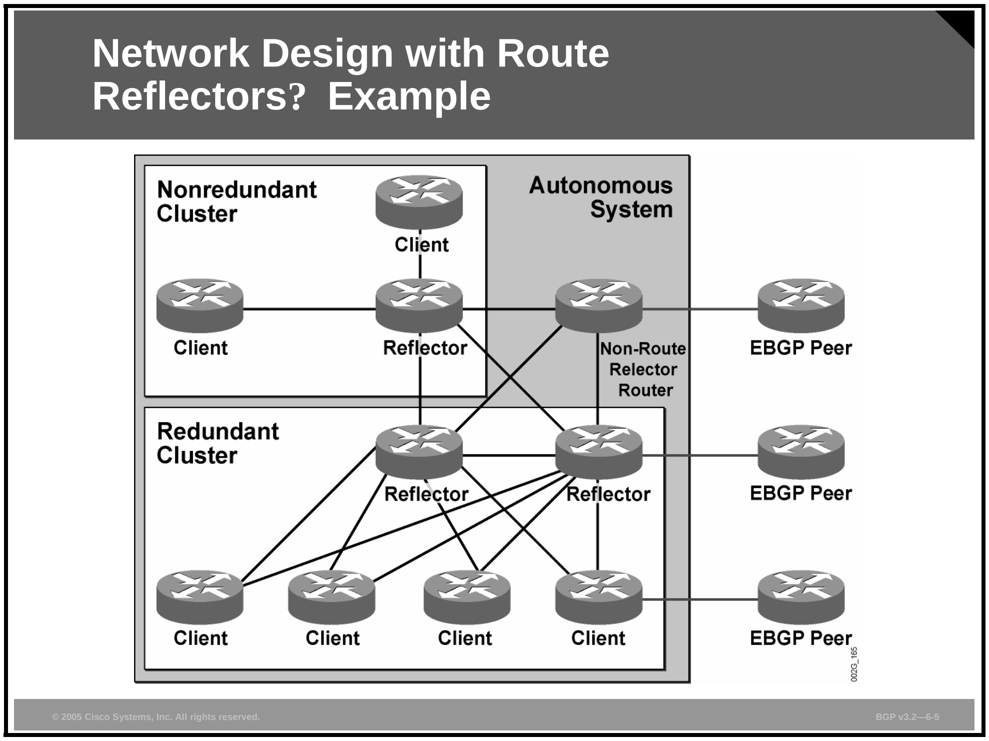
The principal goal for designing networks with BGP route reflectors is to reduce the size of the full mesh of IBGP sessions by excluding some routers from the mesh. The routers that are excluded from the full mesh, the clients, have to send their IBGP information to, and receive it from, at least one router that belongs to the full mesh, the route reflector. Thus, the full mesh is still there, but it is smaller, and all route reflectors have to be part of it.

All clients in a cluster should have IBGP sessions with all their route reflectors and their route reflectors only. If a client does not have sessions with all the reflectors in the cluster, the redundancy is violated. If a client has IBGP sessions to routers other than the route reflectors, unnecessary routing traffic is generated.

Both clients and other routers, those that are not route reflectors, obey the classic IBGP split-horizon rules. Thus, non-route-reflector routers are either clients to a reflector or are participating directly in the full mesh.

Example: Network Design with Route Reflectors

In this example, the routers that serve as route reflectors and the non-route-reflector router have IBGP sessions in a full mesh.



Network Design with Route Reflectors? Example

BGP v3.2—6-5

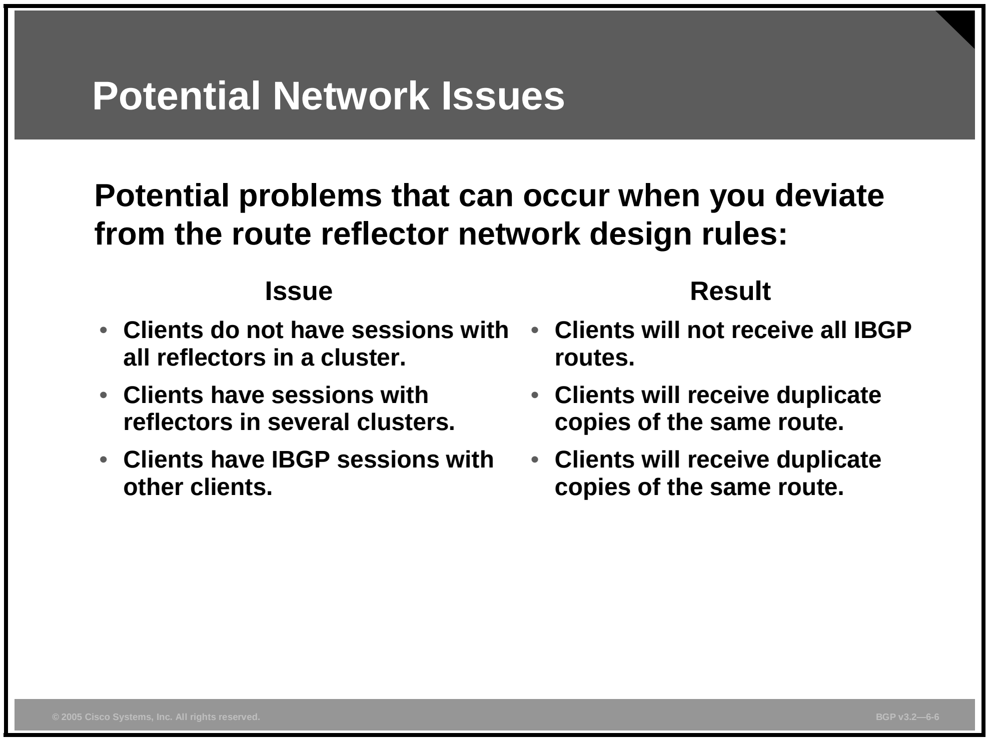
In the area called the “redundant cluster,” the four client routers and the two route reflector routers make up the cluster. Each of the four client routers has an IBGP session with the two route reflectors and only with those two route reflectors.

In the nonredundant area, each of the two client routers has a single physical connection to a route reflector router. These three routers form a nonredundant cluster. The router designated as the route reflector in the cluster is already a single point of failure in this physical design because a failure of this router will prevent the clients in the cluster from reaching the rest of the network. Therefore, there is no new single point of failure that is introduced when the router is configured as the only route reflector in this cluster. Each of the two clients has a single IBGP session to the route reflector.

The other router shown is not configured as a route reflector nor is it a client to any other route reflector. This other router serves as an example of where a non-route-reflector router participates in the full mesh.

Potential Network Issues

This topic lists the potential issues that can arise if the route reflector network design rules as explained in the previous topic are not followed.



Potential Network Issues

Potential problems that can occur when you deviate from the route reflector network design rules:

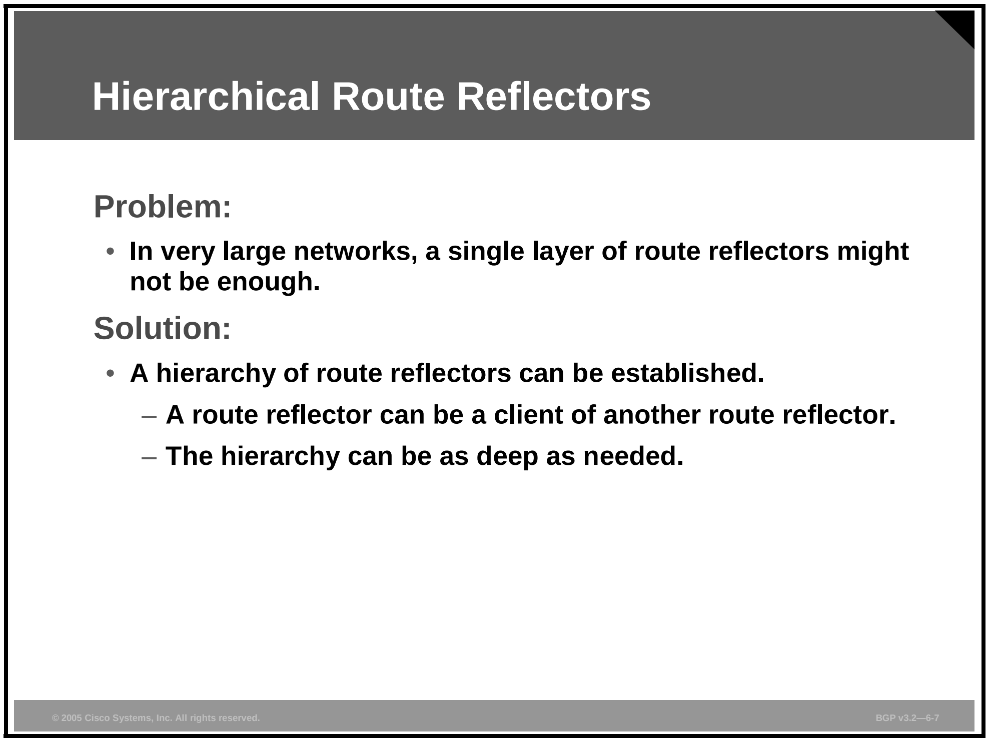
Issue	Result
Clients do not have sessions with all reflectors in a cluster.	**Clients will not receive all IBGP routes.**
Clients have sessions with reflectors in several clusters.	**Clients will receive duplicate copies of the same route.**
Clients have IBGP sessions with other clients.	**Clients will receive duplicate copies of the same route.**

Two nontransitive optional BGP attributes, originator-ID and cluster-list, are both used to prevent fatal loops of information. The use of these two attributes makes a network fairly insensitive to poor configuration. However, for optimal performance, you must have an optimal configuration. Here are some of the problems that could occur if you deviate from route reflector network design rules:

- If route reflectors are not connected with IBGP sessions in a full mesh, some clusters will not have all the routes.
- If a client has IBGP sessions with some route reflectors in a cluster, but not with all of them, the client might miss some BGP routes.
- If a client has IBGP sessions to route reflectors that belong to different clusters, the BGP update from the client will be forwarded by the client into the full mesh with different cluster-IDs in the cluster-list attribute. When the BGP update enters the mesh, it will reach the other route reflector, which will, unnecessarily, accept the route as valid and forward it into its cluster. This situation, in turn, causes unnecessary duplication of updates to the clients.
- If a client has IBGP sessions to other clients in the same cluster, those clients will receive unnecessary duplications of updates.

Hierarchical Route Reflectors

This topic explains the function of hierarchical route reflectors.



Hierarchical Route Reflectors

Problem:

- **In very large networks, a single layer of route reflectors might not be enough.**

Solution:

- **A hierarchy of route reflectors can be established.**
  - **A route reflector can be a client of another route reflector.**
  - **The hierarchy can be as deep as needed.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-7

Network designers can build route reflector clusters in hierarchies. With hierarchies, a router serving as a route reflector in one cluster can act as a client in another cluster.

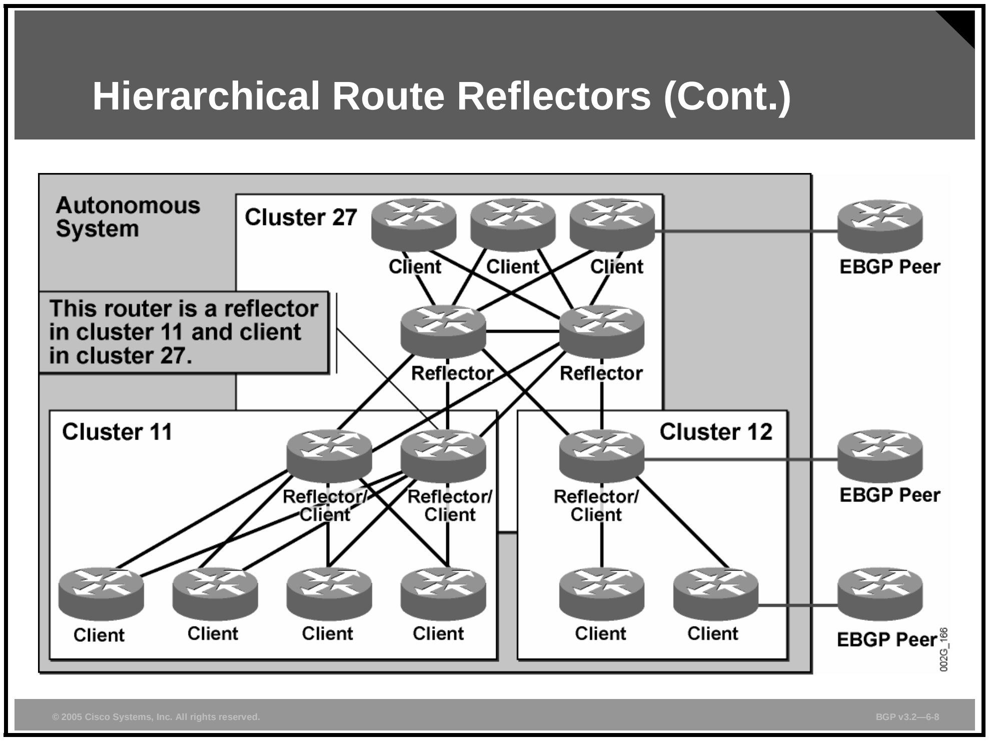
Clients are not configured to be route reflector clients; they simply have fewer IBGP sessions. However, a network designer must configure a route reflector. In configuring an IBGP session on a route reflector, the designer must configure the session to reach a client in order for the route reflector IBGP split-horizon rules to start working. All other IBGP sessions that are configured on the route reflector are a part of the full mesh. Also, the designer must configure the cluster-ID on the route reflector.

A router that is configured to be a route reflector will still have ordinary IBGP sessions that are part of the full mesh. If these sessions are reduced in number and only a few remain, and the remaining ones reach a second level of route reflectors, a hierarchy of route reflectors is created.

When a designer builds a first level of clusters, the remaining full mesh is smaller than when all routers belonged to it. But if it is large enough, the designer can build an additional level of route reflectors.

Example: Hierarchical Route Reflectors

In this example, the first level of route reflector clusters was built by creating cluster 11 and cluster 12.



Hierarchical Route Reflectors (Cont.)

This router is a reflector in cluster 11 and client in cluster 27.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-8

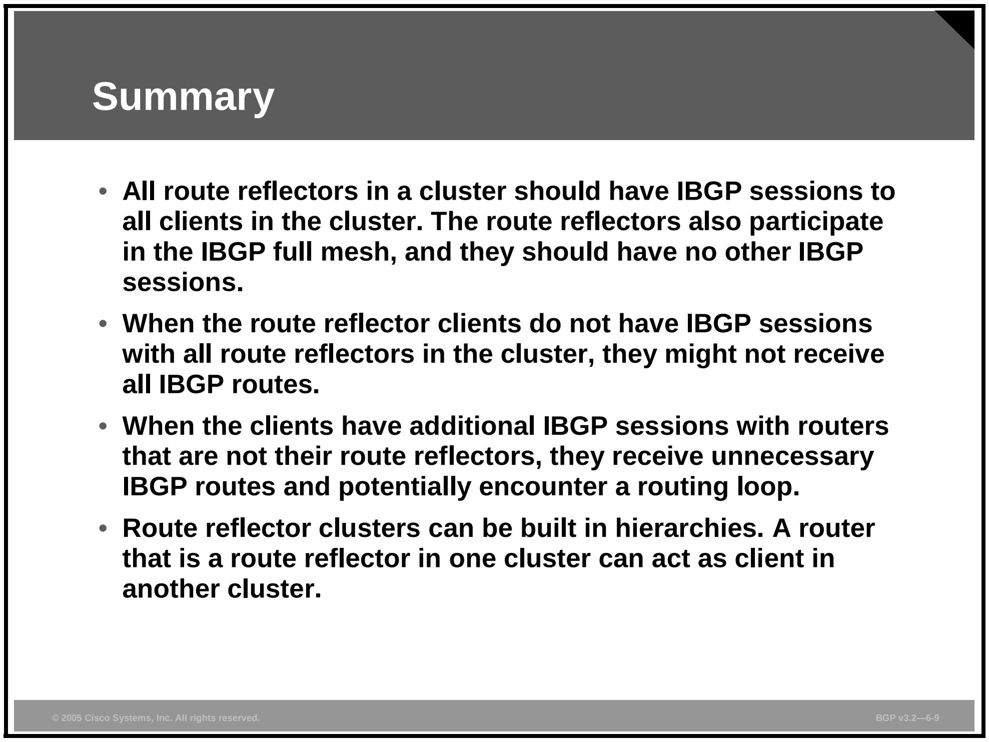
This first step reduced the original full mesh of 14 routers to a full mesh of 8 routers.

A second level of route reflector clusters was built by creating cluster 27. This second step further reduced the full mesh of eight routers to a full mesh consisting of only two routers. Only the two route reflectors in cluster 27 should be connected in a full mesh.

When a client in the lowest level receives an External Border Gateway Protocol (EBGP) update, it will forward it on all configured IBGP sessions to a route reflector. The route reflector recognizes BGP updates that are received from configured clients and will forward these updates to all other clients that use normal IBGP sessions. The update, sent on a normal IBGP session, will be a second-level client update to the second-level route reflector. The second-level route reflector will recognize that the update was received from a client and will forward it to all other clients and into the full mesh.

Summary

This topic summarizes the key points discussed in this lesson.



Summary

- All route reflectors in a cluster should have IBGP sessions to all clients in the cluster. The route reflectors also participate in the IBGP full mesh, and they should have no other IBGP sessions.
- When the route reflector clients do not have IBGP sessions with all route reflectors in the cluster, they might not receive all IBGP routes.
- When the clients have additional IBGP sessions with routers that are not their route reflectors, they receive unnecessary IBGP routes and potentially encounter a routing loop.
- Route reflector clusters can be built in hierarchies. A router that is a route reflector in one cluster can act as client in another cluster.

Lesson 4

Configuring and Monitoring Route Reflectors

Overview

Large Border Gateway Protocol (BGP) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors and confederations. Route reflectors enable BGP routing information to be distributed in a fashion that does not require a physical full-mesh network. Implementing such a network requires knowledge of the steps to properly migrate and configure route reflectors and the commands that are used to verify the operation of a configured network.

This lesson introduces the steps that are required to successfully migrate an existing autonomous system (AS) to BGP route reflectors. It also lists the Cisco IOS commands that are required to configure and monitor route reflectors.

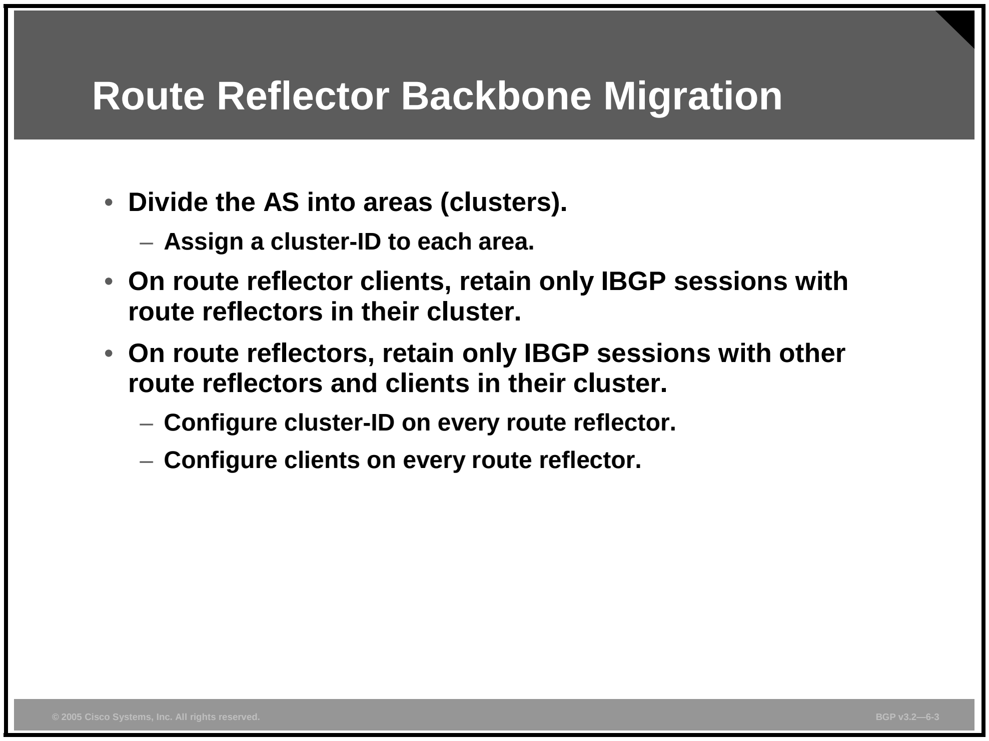
Objectives

Upon completing this lesson, you will be able to configure proper operation of route reflectors to modify IBGP split-horizon rules in an existing IBGP network. This ability includes being able to meet these objectives:

- List the steps to migrate an existing IBGP backbone to a backbone with route reflectors
- Identify the configuration changes and related Cisco IOS commands that are required to configure route reflectors on a BGP backbone
- Identify the Cisco IOS commands that are required to monitor a BGP backbone that contains route reflectors

Route Reflector Backbone Migration

This topic lists the steps that are required to successfully migrate an existing Internal Border Gateway Protocol (IBGP) backbone to a backbone with route reflectors.



Route Reflector Backbone Migration

- **Divide the AS into areas (clusters).**
  - **Assign a cluster-ID to each area.**
- **On route reflector clients, retain only IBGP sessions with route reflectors in their cluster.**
- **On route reflectors, retain only IBGP sessions with other route reflectors and clients in their cluster.**
  - **Configure cluster-ID on every route reflector.**
  - **Configure clients on every route reflector.**

The physical topology of the AS serves as a guide to designing clusters. You should introduce no additional single points of failure when you are deploying route reflectors. If the physical topology is redundant, a good practice is to have redundant route reflectors. If the physical topology is not redundant, introducing a nonredundant cluster does not add a single point of failure because the network was already nonredundant.

The following planning and preparation steps are required before you migrate from a full mesh of IBGP sessions to a route reflector design:

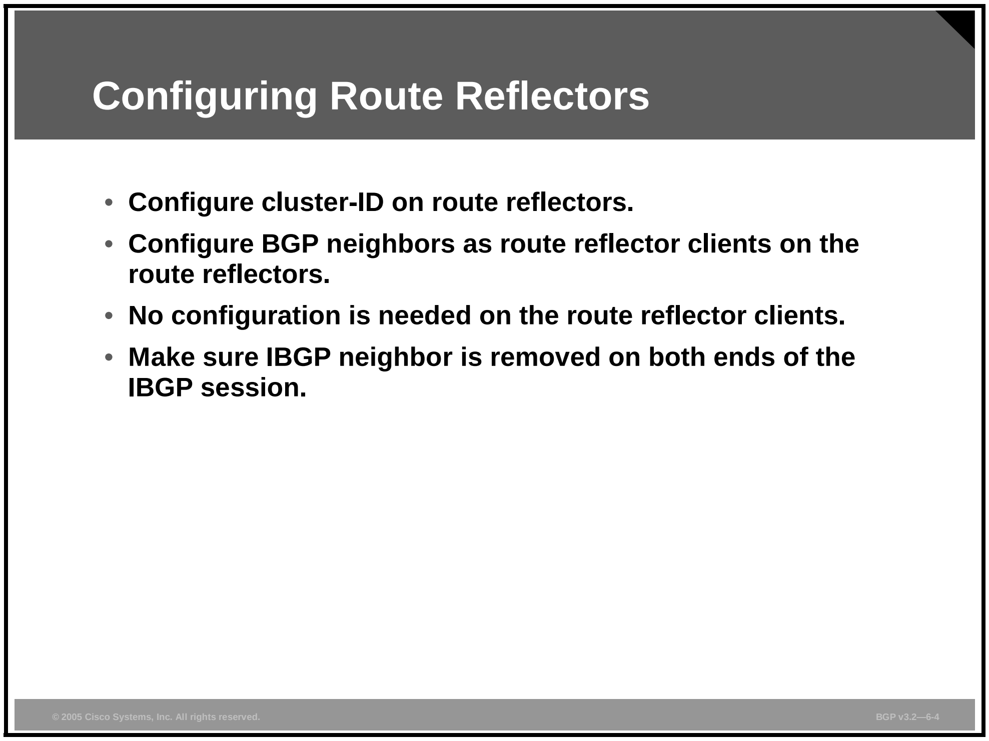
Step 1 Identify a group of peripheral routers that are physically connected to the same set of backbone routers. Consider the peripheral routers as clients and the backbone routers as route reflectors. Let the routers form a cluster. Make sure that no router belongs to two different clusters, because this setup would represent an illegal configuration.

Step 2 Create a numbering plan that indicates how numbers are assigned to the clusters in the network. The plan must make sure to uniquely identify each of the clusters within the AS. Clusters are not seen from outside the AS, so the plan does not need to be coordinated with any other AS. To ease troubleshooting, it is recommended that numbers lower than 256 be used, because cluster-IDs are displayed in IP address format.

Note The default value of a cluster-ID is the BGP router-ID of the route reflector. If you decide to implement nonredundant clusters, you do not have to plan the cluster-ID numbers, because the BGP router-IDs should be unique.

Configuring Route Reflectors

This topic lists the configuration changes and related Cisco IOS commands that are required to configure BGP route reflectors.

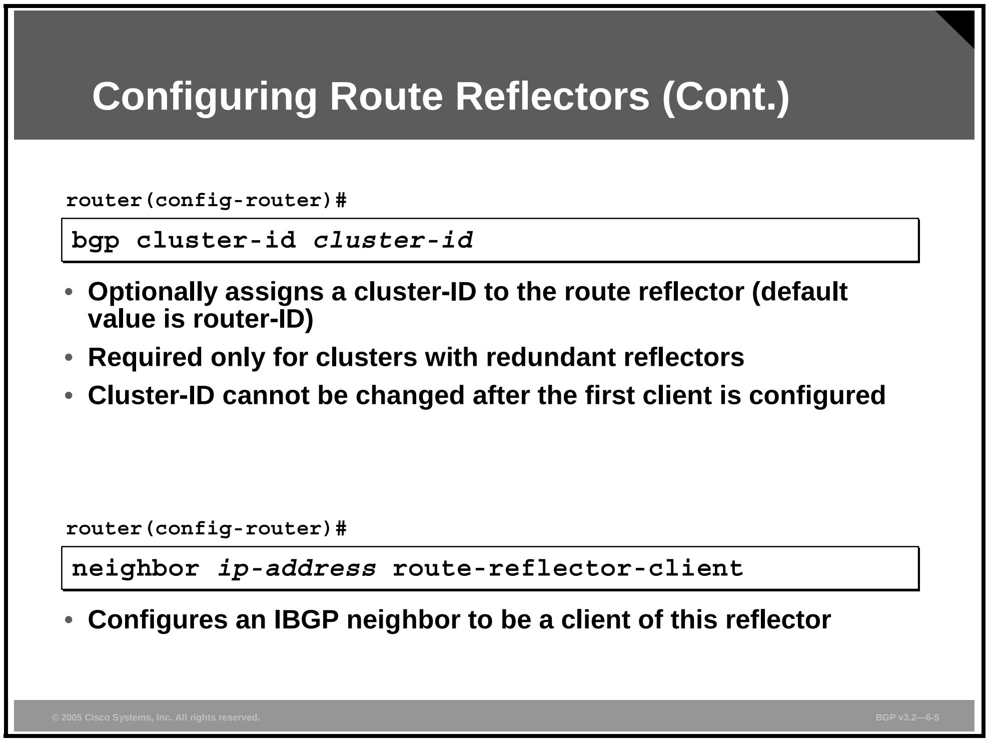


Configuring Route Reflectors

- **Configure cluster-ID on route reflectors.**
- **Configure BGP neighbors as route reflector clients on the route reflectors.**
- **No configuration is needed on the route reflector clients.**
- **Make sure IBGP neighbor is removed on both ends of the IBGP session.**

As part of the planning and preparation that is necessary to migrate from a full mesh of IBGP sessions to a route reflector design, you need to make the following configuration changes:

- Configure the proper cluster-ID value on the route reflectors.
- Configure the route reflector with information about which IBGP neighbor sessions are reaching their clients.
- In the clients, remove all IBGP sessions to neighbors that are not route reflectors in the client cluster.
- Make sure that the IBGP neighbor is removed on both ends of the IBGP session.



Configuring Route Reflectors (Cont.)

```
router(config-router)#
bgp cluster-id cluster-id
```

- Optionally assigns a cluster-ID to the route reflector (default value is router-ID)
- Required only for clusters with redundant reflectors
- Cluster-ID cannot be changed after the first client is configured

```
router(config-router)#
neighbor ip-address route-reflector-client
```

- Configures an IBGP neighbor to be a client of this reflector

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-5

bgp cluster-id

Use the **bgp cluster-id** command to configure the cluster-ID if the BGP cluster has redundant route reflectors.

- **bgp cluster-id** *cluster-id*

To remove the cluster-ID, use the **no** form of this command.

- **no bgp cluster-id** *cluster-id*

Syntax Description

Parameter	Description
cluster-id	Cluster-ID of the router acting as a route reflector. The cluster-ID is a maximum of 4 bytes.

neighbor route-reflector-client

This command used to configure the router as a BGP route reflector and configure the specified neighbor as its client. When all the clients are disabled, the local router is no longer a route reflector.

- **neighbor** *ip-address* **route-reflector-client**

To indicate that the neighbor is not a client, use the **no** form of this command.

- **no neighbor** *ip-address* **route-reflector-client**

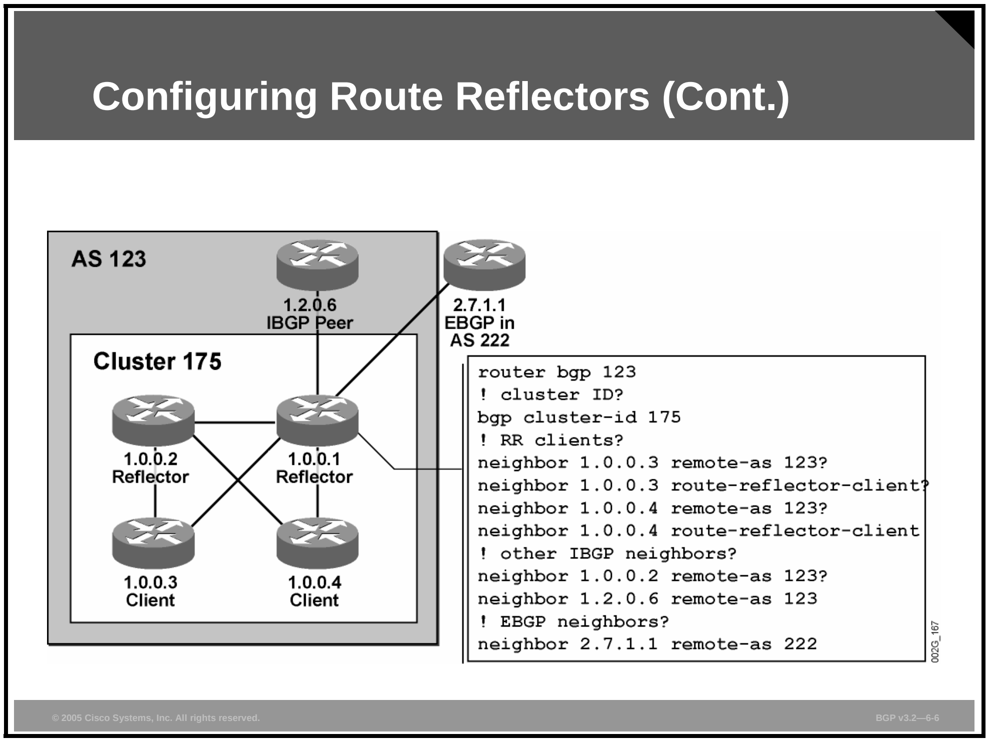
Syntax Description

Parameter	Description
ip-address	Neighbor IP address

By default, there is no route reflector in the AS.

Example: Configuring Route Reflectors

In this example, AS 123 has been divided into clusters with route reflectors.



Configuring Route Reflectors (Cont.)

```
router bgp 123
! cluster ID?
bgp cluster-id 175
! RR clients?
neighbor 1.0.0.3 remote-as 123?
neighbor 1.0.0.3 route-reflector-client?
neighbor 1.0.0.4 remote-as 123?
neighbor 1.0.0.4 route-reflector-client
! other IBGP neighbors?
neighbor 1.0.0.2 remote-as 123?
neighbor 1.2.0.6 remote-as 123
! EBGP neighbors?
neighbor 2.7.1.1 remote-as 222
```

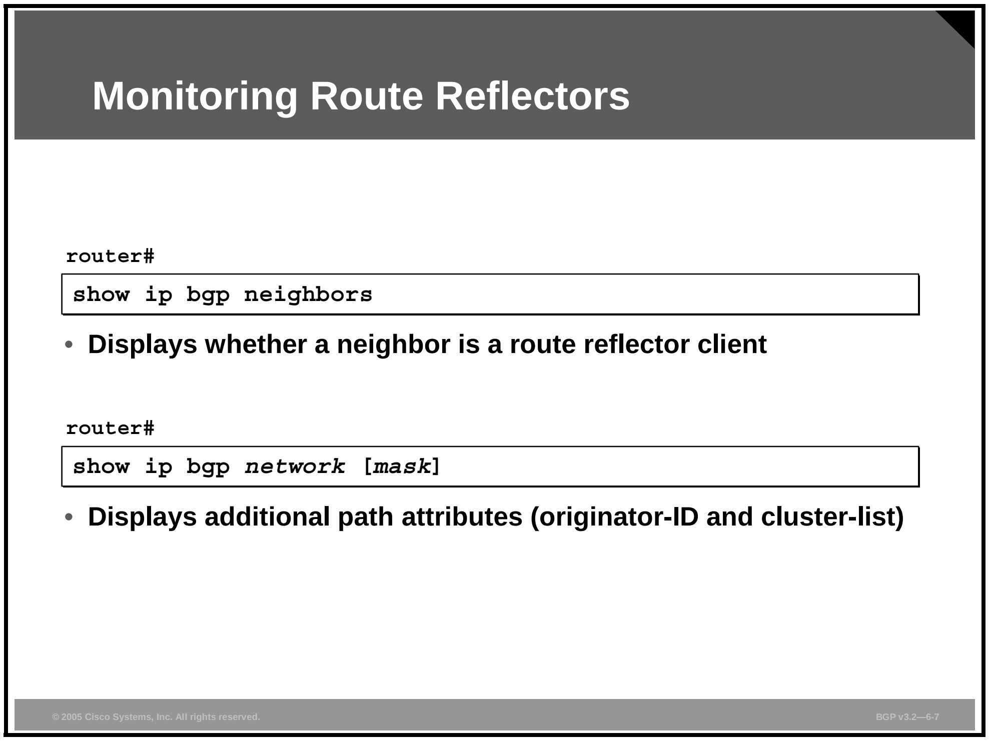
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-6

The routers with router-ID 1.0.0.1 and 1.0.0.2 are route reflectors in a cluster that has been assigned cluster-ID 175. The routers with router-ID 1.0.0.3 and 1.0.0.4 are clients to these two route reflectors.

The figure shows a portion of the configuration in router 1.0.0.1. The cluster-ID is assigned to the router under the **router bgp** process definition of the router configuration. After the router has been assigned, the route reflector client configuration is added under the **router bgp** process for the two neighbors that identify the two sessions reaching clients.

Monitoring Route Reflectors

This topic lists the Cisco IOS commands that are required to monitor route reflector configurations.



Monitoring Route Reflectors

```
router#
show ip bgp neighbors
```

- **Displays whether a neighbor is a route reflector client**

```
router#
show ip bgp network [mask]
```

- **Displays additional path attributes (originator-ID and cluster-list)**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-7

show ip bgp neighbors

To display information about the TCP and BGP connections to neighbors, use the **show ip bgp neighbors** EXEC command.

- **show ip bgp neighbors** [*address*] [**received-routes** | **routes** | **advertised-routes** | {**paths** *regular-expression*} | **dampened-routes**]

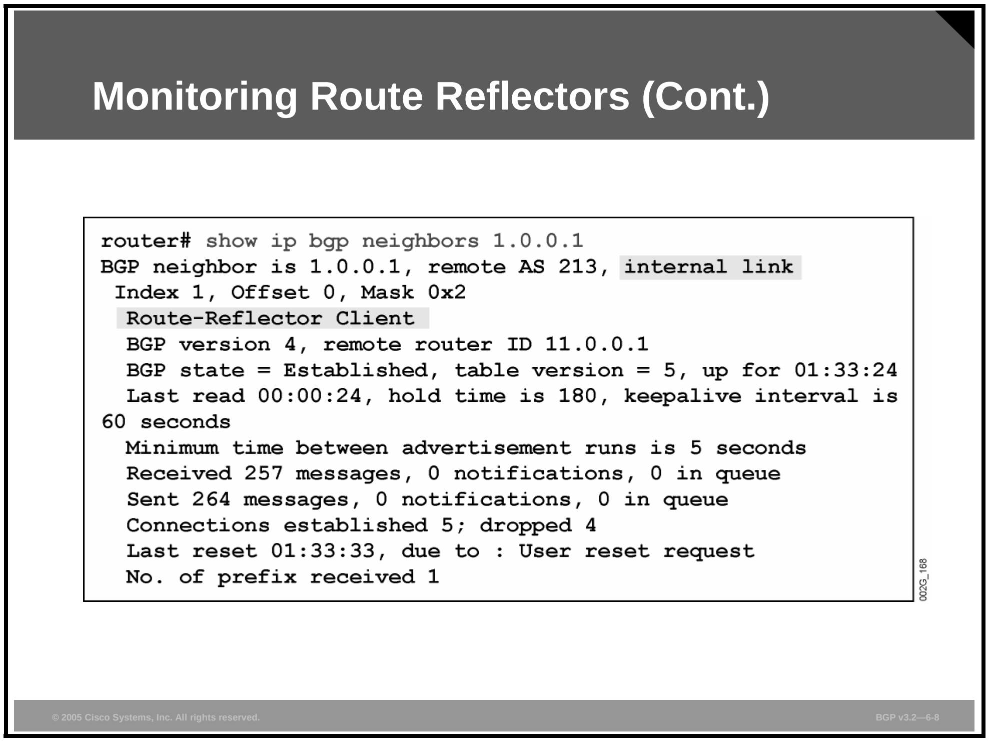
In this case, the **show ip bgp neighbors** command is used on the router not to see routes or paths that have been received but to see the status of the neighbor session, so no other qualifiers than the optional IP address are given.

show ip bgp

To display entries in the BGP routing table, use the **show ip bgp** EXEC command.

- **show ip bgp** [*network*] [*network-mask*] [**longer-prefixes**]

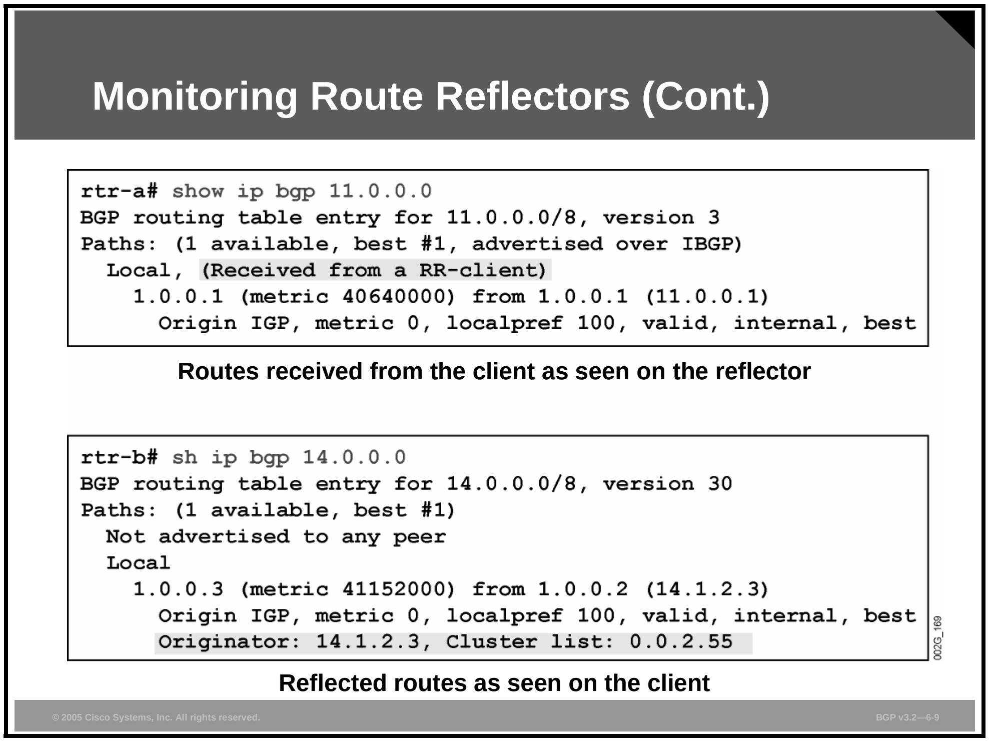
When details are displayed for a specific route entry in the BGP table, the cluster-list and originator-ID attributes are also shown.



Monitoring Route Reflectors (Cont.)

```
router# show ip bgp neighbors 1.0.0.1
BGP neighbor is 1.0.0.1, remote AS 213, internal link
 Index 1, Offset 0, Mask 0x2
  Route-Reflector Client
  BGP version 4, remote router ID 11.0.0.1
  BGP state = Established, table version = 5, up for 01:33:24
  Last read 00:00:24, hold time is 180, keepalive interval is
60 seconds
  Minimum time between advertisement runs is 5 seconds
  Received 257 messages, 0 notifications, 0 in queue
  Sent 264 messages, 0 notifications, 0 in queue
  Connections established 5; dropped 4
  Last reset 01:33:33, due to : User reset request
  No. of prefix received 1
```

The **show ip bgp neighbors** command, issued on the route reflector router, indicates that the neighbor is a route reflector client.



Monitoring Route Reflectors (Cont.)

```
rtr-a# show ip bgp 11.0.0.0
BGP routing table entry for 11.0.0.0/8, version 3
Paths: (1 available, best #1, advertised over IBGP)
  Local, (Received from a RR-client)
    1.0.0.1 (metric 40640000) from 1.0.0.1 (11.0.0.1)
      Origin IGP, metric 0, localpref 100, valid, internal, best
```

Routes received from the client as seen on the reflector

```
rtr-b# sh ip bgp 14.0.0.0
BGP routing table entry for 14.0.0.0/8, version 30
Paths: (1 available, best #1)
  Not advertised to any peer
  Local
    1.0.0.3 (metric 41152000) from 1.0.0.2 (14.1.2.3)
      Origin IGP, metric 0, localpref 100, valid, internal, best
      Originator: 14.1.2.3, Cluster list: 0.0.2.55
```

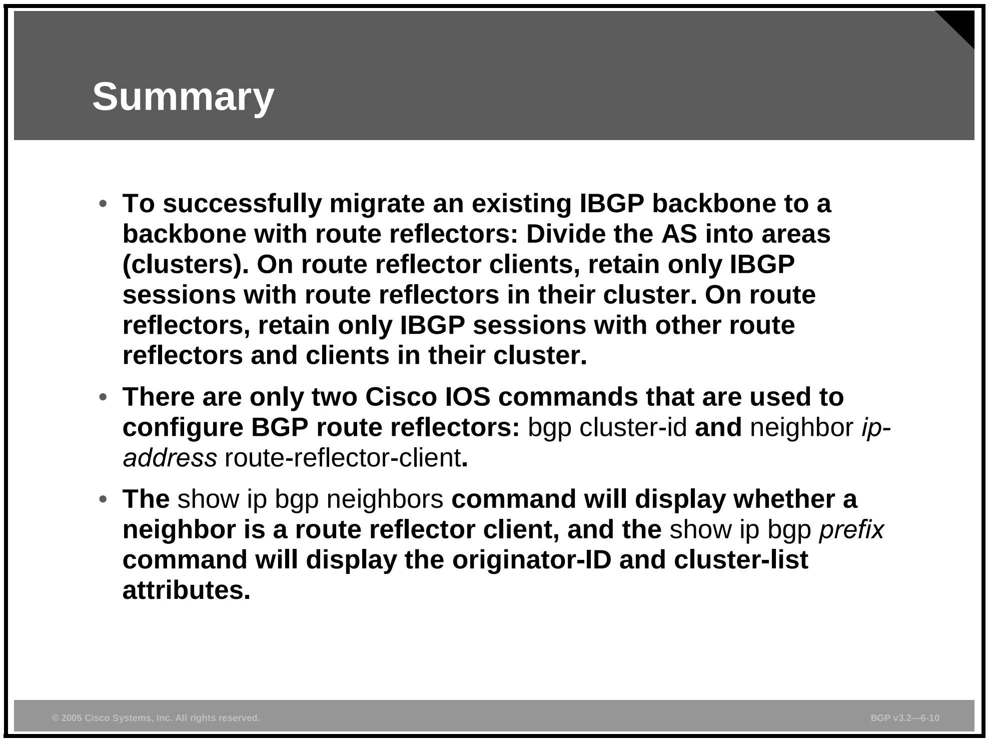
Reflected routes as seen on the client

The first example of a **show ip bgp** command is issued on a route reflector router. It shows that this particular entry in the BGP table was received from a route reflector client.

The second example shows an entry in the BGP table that at some point was reflected from a route reflector. The reflecting router has added the originator-ID and cluster-list attributes to the route.

Summary

This topic summarizes the key points discussed in this lesson.



Summary

- **To successfully migrate an existing IBGP backbone to a backbone with route reflectors: Divide the AS into areas (clusters). On route reflector clients, retain only IBGP sessions with route reflectors in their cluster. On route reflectors, retain only IBGP sessions with other route reflectors and clients in their cluster.**
- **There are only two Cisco IOS commands that are used to configure BGP route reflectors:** bgp cluster-id **and** neighbor *ip-address* route-reflector-client**.**
- **The** show ip bgp neighbors **command will display whether a neighbor is a route reflector client, and the** show ip bgp *prefix* **command will display the originator-ID and cluster-list attributes.**

Lesson 5

Introducing Confederations

Overview

Large Border Gateway Protocol (BGP) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors and confederations. Routers within an autonomous system (AS) are typically configured in a full mesh. Confederations and route reflectors are BGP scalability mechanisms that enable routing information to be redistributed to all routers within an AS while eliminating the need for a fully meshed topology within the AS. These features reduce the number of TCP sessions that must be maintained, which lowers network overhead and CPU and memory requirements. Confederations can serve as an alternative or a complement to route reflectors, enabling network administrators to break up an AS into a set of logical subautonomous systems.

This lesson introduces BGP confederations by explaining why confederations are used to improve BGP scalability. This lesson also discusses AS-path propagation and processing in an AS that contains confederations.

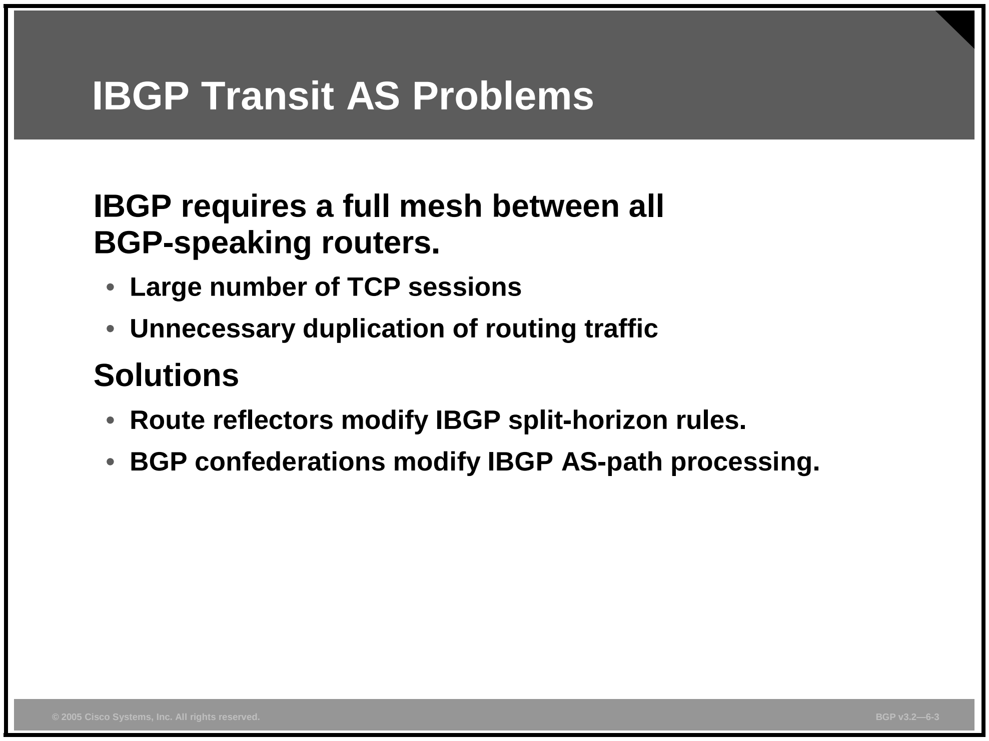
Objectives

Upon completing this lesson, you will be able to describe the function of confederations in a BGP environment. This ability includes being able to meet these objectives:

- Describe the IBGP full-mesh requirement when you are using a transit AS in relation to the potential issues that this requirement can cause
- Explain how you can use BGP confederations to split an AS into a series of smaller autonomous systems
- Describe AS-path propagation in BGP confederations
- Explain AS-path attribute processing in an AS that contains BGP confederations
- Explain the properties of intra-confederation EBGP sessions

IBGP Transit AS Problems

This topic describes the Internal Border Gateway Protocol (IBGP) full-mesh requirement when you are using a transit AS and the potential issues that this requirement can cause.



IBGP Transit AS Problems

IBGP requires a full mesh between all BGP-speaking routers.

- **Large number of TCP sessions**
- **Unnecessary duplication of routing traffic**

Solutions

- **Route reflectors modify IBGP split-horizon rules.**
- **BGP confederations modify IBGP AS-path processing.**

BGP v3.2—6-3

Classic IBGP split-horizon rules specify that BGP updates that are received on an External Border Gateway Protocol (EBGP) session should be forwarded on all IBGP and EBGP sessions, but BGP updates that are received on an IBGP session should be forwarded on all EBGP sessions only. This rule requires a boundary router to be able to update all other BGP-speaking routers in its own AS directly via an IBGP session that is established to each of them.

IBGP split-horizon rules avoid routing information loops within the AS. If IBGP information were forwarded to another IBGP peer router, the information might come back to the originator and be forwarded again in a never-ending loop. The originator would not detect the loop because no BGP attributes are changed when an update is sent through an IBGP session.

The general design rule in classic IBGP is to have a full mesh of IBGP sessions between all BGP-speaking routers inside an AS. However, a full mesh of IBGP sessions between *n* number of routers would require (*n* * (*n* – 1)) / 2 IBGP sessions. For example, an AS with 10 routers would require (10 * (10 – 1)) / 2 = 45 IBGP sessions.

Every BGP session between two routers is established through a separate TCP session to the BGP peer. An update that must be sent to all IBGP peers must be sent separately on each of the TCP sessions. If a router is attached to the rest of the network over just a single link, this single link has to carry all TCP/IP packets for all IBGP sessions. This requirement results in duplication of BGP updates over the single link.

Two solutions are available to achieve greater scalability when you are faced with the full-mesh rules of IBGP autonomous systems:

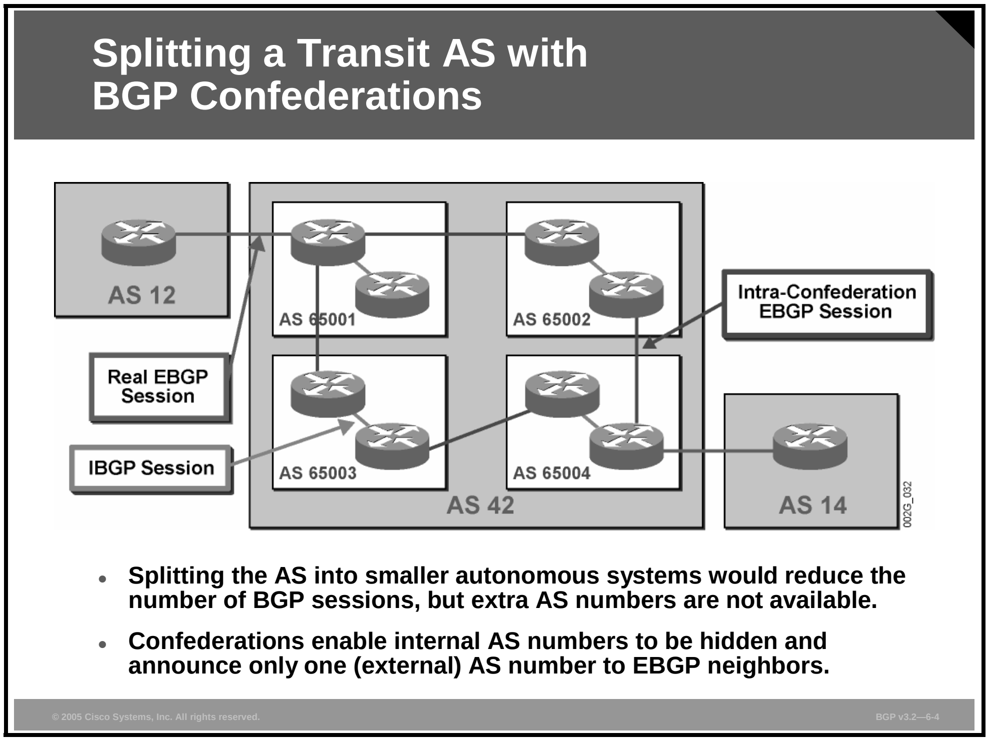
- **Route reflectors:** Route reflectors modify the classic IBGP split-horizon rules and allow a particular router to forward incoming IBGP updates to an outgoing IBGP session under certain conditions. This router becomes a concentration router, or a route reflector.
- **BGP confederations:** BGP confederations introduce the concept of a number of smaller autonomous systems within the original AS. The small autonomous systems exchange BGP updates between them using intra-confederation EBGP sessions.

Example: IBGP Transit AS Problems

A large service provider backbone, acting as a transit AS, contains 150 routers. Because the classic IBGP design rule requires a full mesh, the AS would require 11,175 sessions (n * [n – 1] / 2). This number is impractical, so you should use other solutions, such as BGP confederations, to reduce the full-mesh requirements of the network.

Splitting a Transit AS with BGP Confederations

This topic describes how you can use BGP confederations to split an AS into a series of smaller autonomous systems.



Splitting a Transit AS with BGP Confederations

- **Splitting the AS into smaller autonomous systems would reduce the number of BGP sessions, but extra AS numbers are not available.**
- **Confederations enable internal AS numbers to be hidden and announce only one (external) AS number to EBGP neighbors.**

A large number of routers in a large transit AS would traditionally introduce a complex full-mesh structure of IBGP sessions. By splitting the AS into a number of small autonomous systems, you can provide each one of the small systems with a fairly simple IBGP structure. Interconnections between these autonomous systems could then be made using EBGP, which allows for arbitrary topologies.

Splitting an AS into smaller autonomous systems requires a large number of official AS numbers, which are a scarce resource.

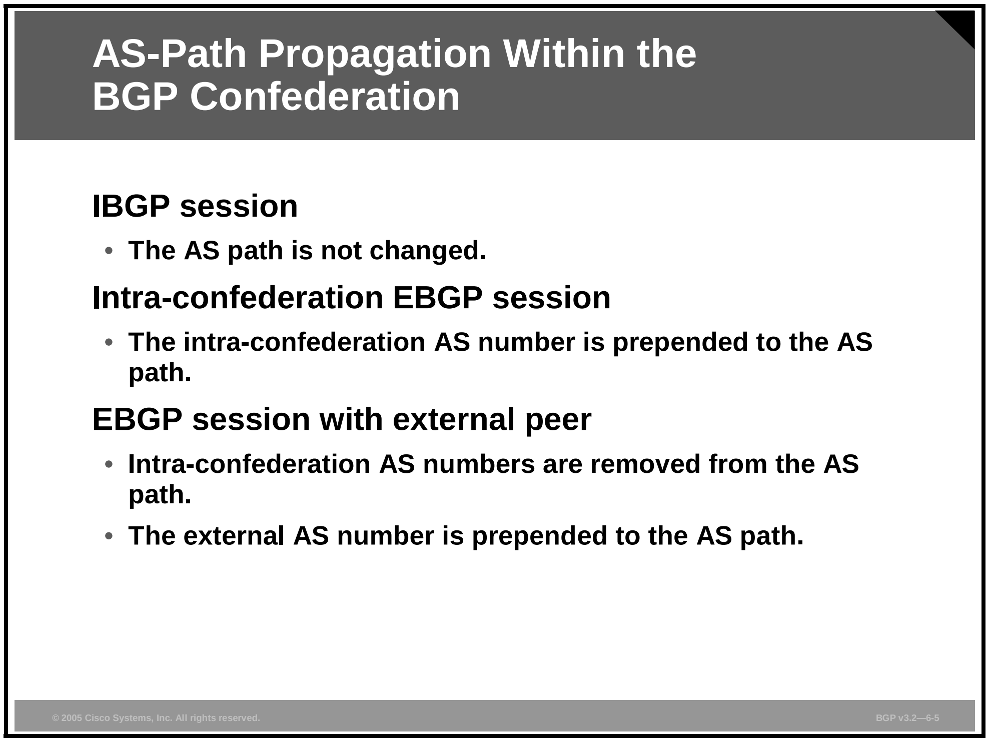
However, by introducing the BGP confederation, you can enable a large AS to be partitioned into a number of smaller autonomous systems (called “member autonomous systems”) where each is internal to the larger AS. The AS numbers of each member-AS that is used within the confederation are never visible from outside the confederation itself. This invisibility allows private AS numbers (in the range 64512 to 65535) to be assigned to autonomous systems inside a confederation to identify a member-AS, without the need to coordinate AS number assignments with an official AS delegation authority.

Within a member-AS, the classic IBGP rules apply. Therefore, all BGP routers inside the member-AS must still maintain a full mesh of BGP sessions.

Between member autonomous systems inside a confederation, EBGP sessions are established. These EBGP sessions behave slightly differently from classic EBGP sessions and are therefore named intra-confederation EBGP sessions to differentiate them from true EBGP sessions.

AS-Path Propagation Within the BGP Confederation

This topic describes how the AS-path attribute is propagated inside and outside of the BGP confederation.



AS-Path Propagation Within the BGP Confederation

IBGP session

- **The AS path is not changed.**

Intra-confederation EBGP session

- **The intra-confederation AS number is prepended to the AS path.**

EBGP session with external peer

- **Intra-confederation AS numbers are removed from the AS path.**
- **The external AS number is prepended to the AS path.**

BGP v3.2—6-5

The mandatory well-known BGP attribute AS-path is modified on EBGP and intra-confederation EBGP sessions. The sender prepends its own AS number to the AS path whenever an EBGP update is sent. When a BGP update traverses the Internet, every AS it passes through is recorded in the AS path. If the update, for any reason, comes back to an AS in which it has already been, the receiving router recognizes its own AS in the AS path and silently ignores the update. This mechanism prevents information loops and allows arbitrary topology when you are interconnecting autonomous systems.

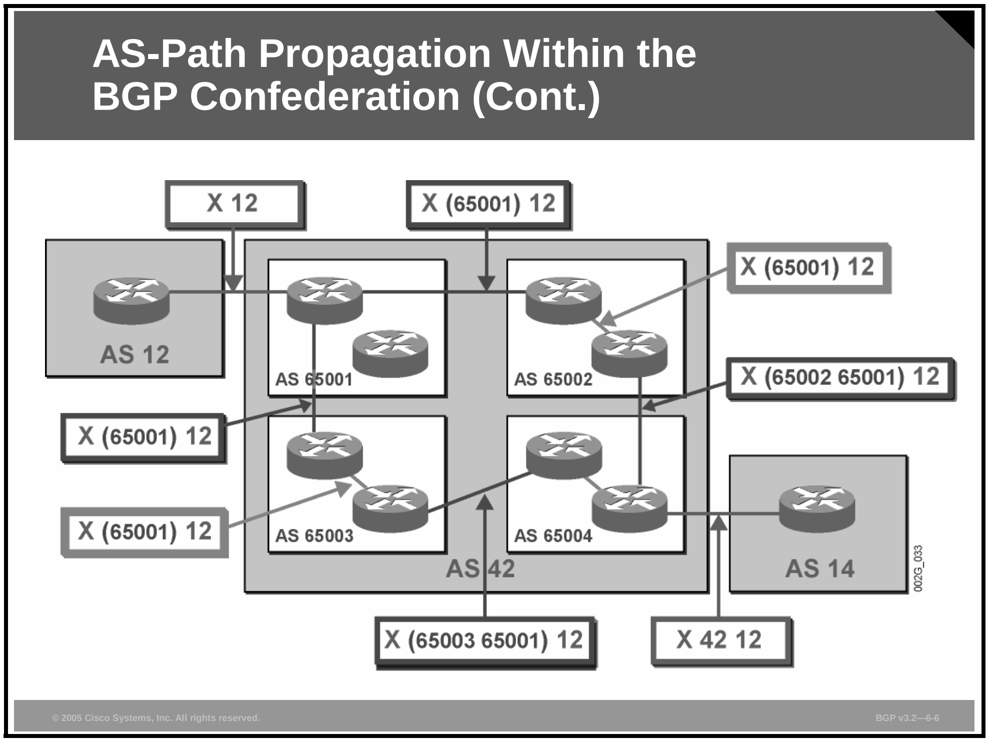
IBGP sessions do not modify the AS-path attribute, so the topology within each AS is limited to the full mesh, and the propagation of BGP updates across multiple IBGP sessions is prohibited.

When a router sends a BGP update over an intra-confederation EBGP session, it prepends the member-AS number to the AS path. This information is maintained by the routers within the confederation and prevents routing information loops inside the confederation.

When a router sends a BGP update over a true EBGP session to an AS outside of the confederation, it removes the part of the AS path describing the member-AS numbers and prepends the official AS number to the AS path. As a result, the confederation appears as one single AS to the outside world.

Example: AS-Path Propagation Within the BGP Confederation

The figure illustrates how the AS-path attribute is processed within a BGP confederation.



AS-Path Propagation Within the BGP Confederation (Cont.)

BGP v3.2—6-6

Network X originates inside AS 12 and is announced by the edge router in AS 12 over a true EBGP session to the ingress router in the confederation. The edge router in AS 12 determines that the edge router that it is communicating with resides in AS 42. AS 12 prepends its assigned AS number in the AS path (which was previously empty, because network X originated in AS 12). When the EBGP update arrives at the confederation member-AS 65001, the AS-path attribute has been set to 12.

The member-AS 65001 has intra-confederation EBGP sessions to member-AS 65002 and to member-AS 65003. The router in AS 65001 prepends its own AS number to the AS path. When doing this, it signals that this part of the AS path describes the intra-confederation AS path. When printed out, the intra-confederation part of the AS path is displayed within parentheses. Therefore, when member AS 65001 sends the update to member-AS 65002 and member-AS 65003, the route to network X has an AS-path attribute set to (65001) 12.

Within a member-AS, the router sends the update using classic IBGP. The router does not modify the AS path when transmitting it over an IBGP session.

Member-AS 65002 and member-AS 65003 both prepend their AS number, so member-AS 65004 receives the update about the route to X via two different paths, one with AS path (65002 65001) 12 and the other with AS path (65003 65001) 12.

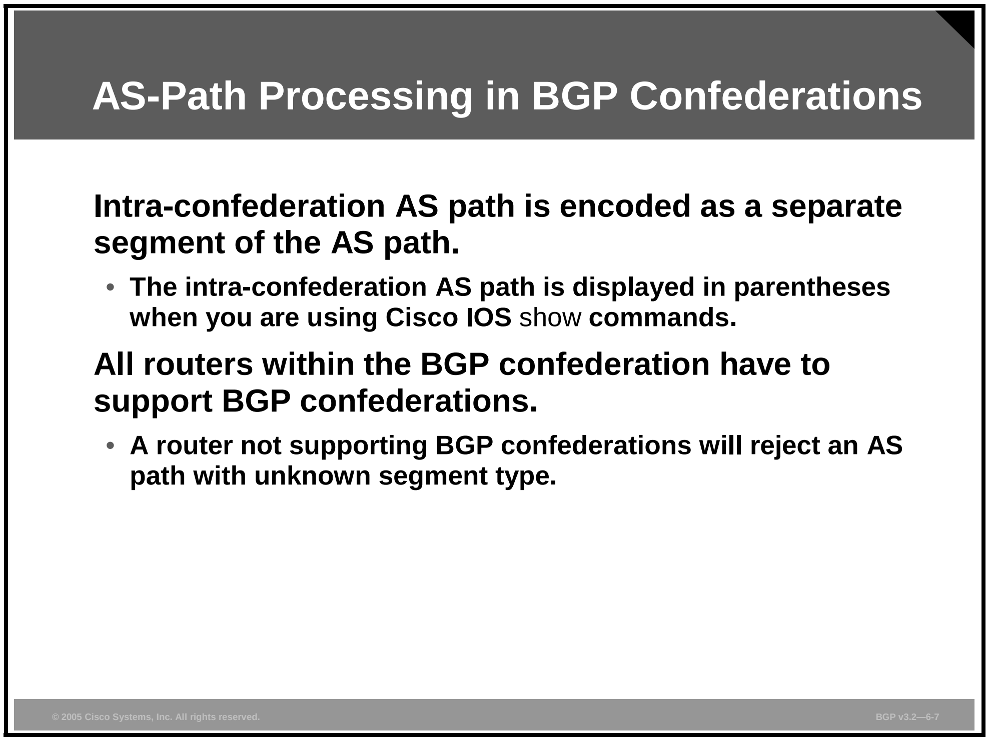
Member-AS 65004 selects one of the alternatives as the best BGP route. It then forwards this update on the intra-confederation EBGP session. This update could introduce a loop, but if the update were ever to be forwarded all the way back up to member-AS 65001, the loop would be detected and member-AS 65001 would silently ignore the update.

Member-AS 65004 also forwards the update about network X on a true EBGP session to AS 14. When it does, it removes all the parenthesized information in the AS path and replaces it with the official AS number of the confederation, 42. AS 14 thus receives the update about network X with an AS path that is set to 42 12.

The routers in AS 14 select the best path based on the length of the AS path if no other policy is configured. AS14 will see the route to X with an AS-path length of two. When AS 14 forwards packets that are destined for network X into the confederation (AS 42), member-AS 65004 must make a forwarding decision. This decision process inside the confederation will use both confederate AS paths for loop avoidance but not for choosing the shortest AS path within the confederation. The multiple step BGP path-selection process will treat both AS paths as equal and have to use the other attributes to select the preferred path. All else being equal, the BGP decision process chooses normal EBGP routes over confederation EBGP routes and confederation EBGP over IBGP routes.

AS-Path Processing in BGP Confederations

This topic explains how the BGP AS-path attribute is processed within an AS that contains a BGP confederation.



AS-Path Processing in BGP Confederations

Intra-confederation AS path is encoded as a separate segment of the AS path.

- **The intra-confederation AS path is displayed in parentheses when you are using Cisco IOS** show **commands.**

All routers within the BGP confederation have to support BGP confederations.

- **A router not supporting BGP confederations will reject an AS path with unknown segment type.**

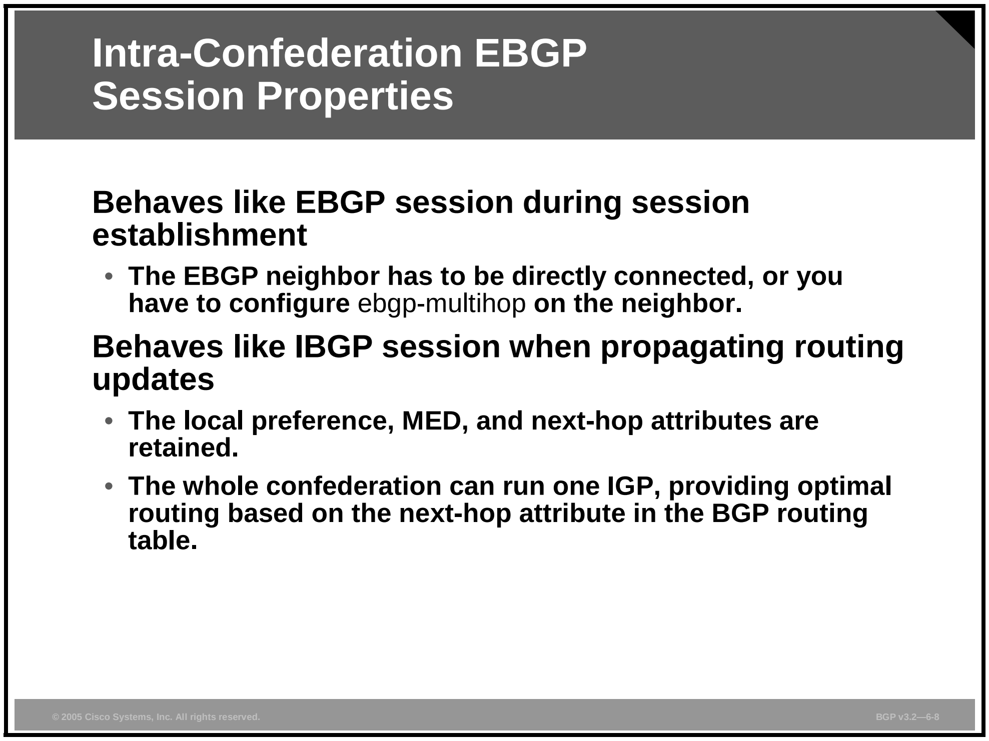
BGP v3.2—6-7

When BGP routing updates are sent by BGP-speaking routers over a BGP session, the BGP attributes are encoded in a binary structure. The AS-path attribute, which is printed out and displayed as a text string, is actually a type, length, value (TLV) binary field, which is composed of several segments. The intra-confederation part of the AS path is encoded by the intra-confederation router as a separate segment of the AS path with a new type code. This segment of the AS path contains a sequence of AS numbers that encode the member autonomous systems that the BGP update has traversed.

Because this segment is an extension to the original interpretation of the mandatory well-known BGP attribute AS path, a BGP implementation that does not support BGP confederations will not understand the intra-confederation part of the AS path. If a router receives a BGP update with a mandatory well-known attribute that the router cannot interpret, it will send a notification to the neighbor that sent the offending update and terminate the session. A router that does not support BGP confederations, therefore, cannot operate inside a BGP confederation.

Intra-Confederation EBGP Session Properties

This topic describes EBGP sessions between different autonomous systems that are contained within the confederation.



Intra-Confederation EBGP Session Properties

Behaves like EBGP session during session establishment

- **The EBGP neighbor has to be directly connected, or you have to configure** ebgp-multihop **on the neighbor.**

Behaves like IBGP session when propagating routing updates

- **The local preference, MED, and next-hop attributes are retained.**
- **The whole confederation can run one IGP, providing optimal routing based on the next-hop attribute in the BGP routing table.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-8

Intra-confederation EBGP sessions, while having EBGP-like properties (for example, updating the AS-path attribute when propagating BGP routes), still run inside a real AS and therefore have to share some properties with IBGP sessions to achieve the same end results. Similar to IBGP sessions, the BGP attributes of local preference, multi-exit discriminator (MED), and next-hop are not changed in updates that are propagated across intra-confederation EBGP sessions. All routers in all member autonomous systems inside the confederation consequently use the same next-hop address when they are doing recursive routing. Because all intra-confederation routers use the same next-hop address, the entire confederation should use the same Interior Gateway Protocol (IGP) to resolve the BGP next-hop address. The IGP information should not be limited by the member-AS boundary.

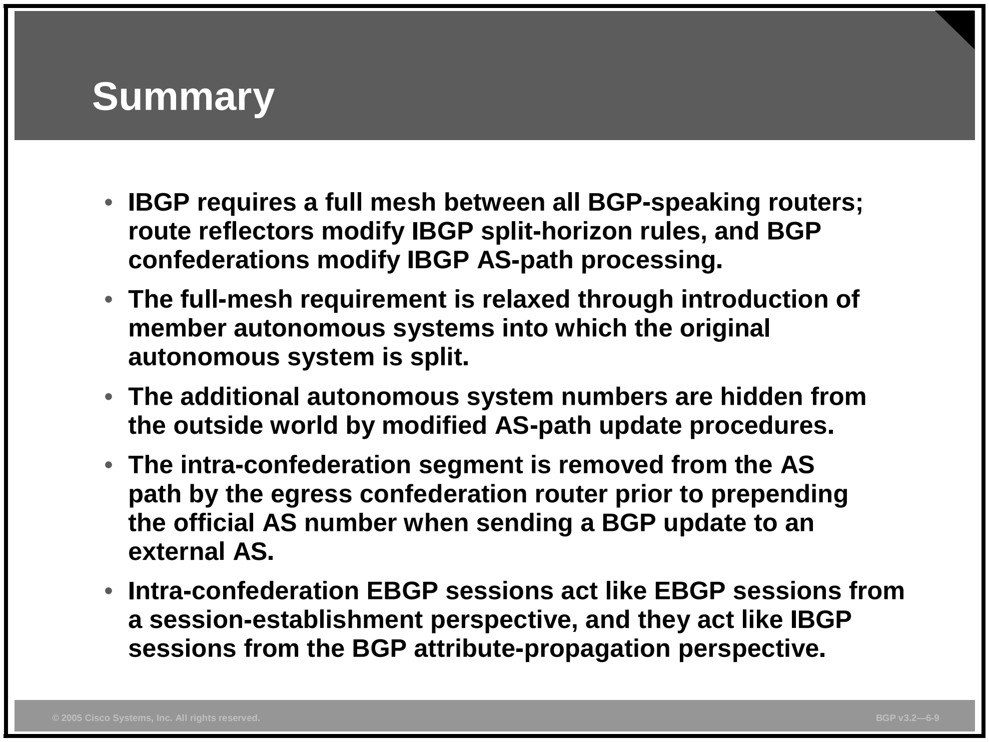
On the other hand, intra-confederation EBGP sessions behave exactly like EBGP sessions when they are established. EBGP sessions are normally opened between directly connected interfaces. However, because all routers within the confederation run the same IGP and exchange internal routing information, there is no problem for them to open multihop sessions. Resiliency of BGP sessions and consequent stability of BGP routing are introduced into the network if the intra-confederation EBGP sessions are established between loopback interfaces, just like IBGP sessions normally are.

When intra-confederation EBGP sessions are opened between loopback interfaces, the **ebgp-multihop** qualifier must be given to the session. Otherwise the EBGP session will never leave the Idle state.

Actually, the intra-confederation EBGP sessions could be established between intra-confederation routers in an arbitrary topology, not necessarily following the physical topology. The next hop of the route will always contain the IP address of a BGP router outside of the confederation, and packet forwarding will follow the optimal path, because recursive routing will rely on the IGP to reach the BGP next hop. The intra-confederation EBGP sessions are merely used to distribute the BGP updates to all member autonomous systems. To avoid unnecessary duplication of routing updates, network designers should take great care when designing the topology of the intra-confederation EBGP sessions.

Summary

This topic summarizes the key points discussed in this lesson.



Summary

- IBGP requires a full mesh between all BGP-speaking routers; route reflectors modify IBGP split-horizon rules, and BGP confederations modify IBGP AS-path processing.
- The full-mesh requirement is relaxed through introduction of member autonomous systems into which the original autonomous system is split.
- The additional autonomous system numbers are hidden from the outside world by modified AS-path update procedures.
- The intra-confederation segment is removed from the AS path by the egress confederation router prior to prepending the official AS number when sending a BGP update to an external AS.
- Intra-confederation EBGP sessions act like EBGP sessions from a session-establishment perspective, and they act like IBGP sessions from the BGP attribute-propagation perspective.

Lesson 6

Configuring and Monitoring Confederations

Overview

Large Border Gateway Protocol (BGP) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors and confederations. Confederations enable BGP routing information to be distributed in a fashion that does not require a physical full-mesh network. Implementing such a network requires knowledge of the steps to properly migrate and configure BGP confederations and the commands that are used to verify the operation of the configured network.

This lesson introduces the steps that are required to successfully migrate an existing autonomous system (AS) to BGP confederations. It also presents the Cisco IOS commands that are required to configure and monitor confederations.

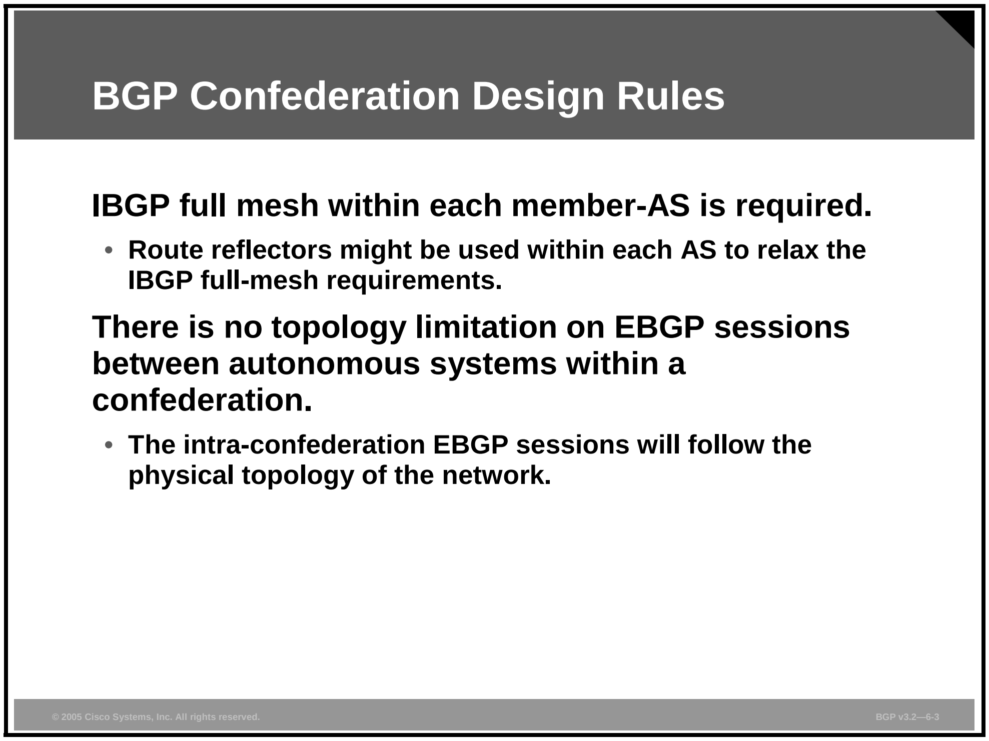
Objectives

Upon completing this lesson, you will be able to configure proper operation of confederations to modify IBGP AS-path processing in an existing IBGP network. This ability includes being able to meet these objectives:

- Describe the basic design rules that network designers should follow when planning a transit AS for BGP confederations
- Explain how to plan a BGP backbone for a configuration that includes BGP confederations
- Explain how to configure BGP confederations on a BGP backbone
- Identify the Cisco IOS commands that are required to monitor a BGP backbone that contains BGP confederations

BGP Confederation Design Rules

This topic describes the basic design rules that network designers should follow when planning a transit AS for BGP confederations.



BGP Confederation Design Rules

IBGP full mesh within each member-AS is required.

- **Route reflectors might be used within each AS to relax the IBGP full-mesh requirements.**

There is no topology limitation on EBGP sessions between autonomous systems within a confederation.

- **The intra-confederation EBGP sessions will follow the physical topology of the network.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-3

When designing BGP confederations, keep in mind two basic design rules:

- There are no restrictions on intra-confederation EBGP sessions.
- A full mesh of IBGP sessions is still needed inside every member-AS.

BGP confederations do not modify Internal Border Gateway Protocol (IBGP) behavior, and, therefore, the classic IBGP split-horizon rules still apply. As a result, a full mesh of IBGP sessions between all routers in the member-AS is required. The basic idea of BGP confederations is to make the member-AS smaller than the original AS so that the full mesh will not be too complex. Route reflector functionality is also available within a member-AS to reduce the complexity of IBGP sessions if needed.

In theory, the member autonomous systems may be interconnected by intra-confederation External Border Gateway Protocol (EBGP) sessions in an arbitrary topology. However, a structure that is too complicated introduces unnecessary duplication of information. There is no split-horizon function on EBGP sessions, meaning that in a redundant configuration, where a BGP update may be forwarded by a router over two different member autonomous systems to reach a third member-AS, the receiving member-AS will get both copies of the update.

Experience shows that a centralized confederation design leads to the best behavior. Centralized design means that all member autonomous systems will exchange information with each other via a central member-AS backbone.

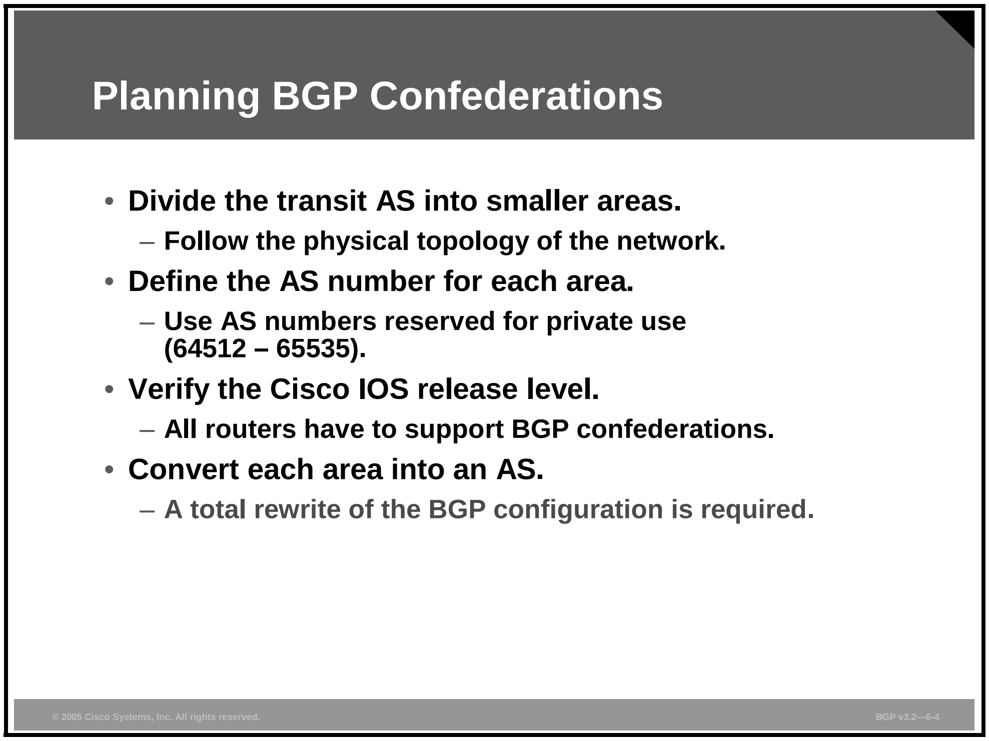
On the other hand, too few intra-confederation EBGP sessions may introduce single points of failure. If two member autonomous systems are redundantly connected on the physical level but have only a single intra-confederation EBGP session between them, a single point of failure is introduced.

Good network design should never introduce additional single points of failure. If the physical topology is redundant, then the intra-confederation EBGP sessions should be redundant as well.

In a hierarchical network topology, the network core could serve as a central member-AS backbone. The more peripheral parts of the network could be divided into several member autonomous systems that are all connected to the central member-AS.

Planning BGP Confederations

This topic explains how to plan a BGP backbone for a configuration that includes BGP confederations.



Planning BGP Confederations

- Divide the transit AS into smaller areas.
  - Follow the physical topology of the network.
- Define the AS number for each area.
  - Use AS numbers reserved for private use (64512 – 65535).
- Verify the Cisco IOS release level.
  - All routers have to support BGP confederations.
- Convert each area into an AS.
  - A total rewrite of the BGP configuration is required.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-4

The physical topology of the AS serves as a guide to design the confederation. You should introduce no additional single points of failure when implementing the confederation. If the physical topology is redundant, good practice is to have redundant intra-confederation EBGP sessions. If the physical topology is not redundant, introducing a nonredundant set of sessions does not add a single point of failure—the network was already nonredundant.

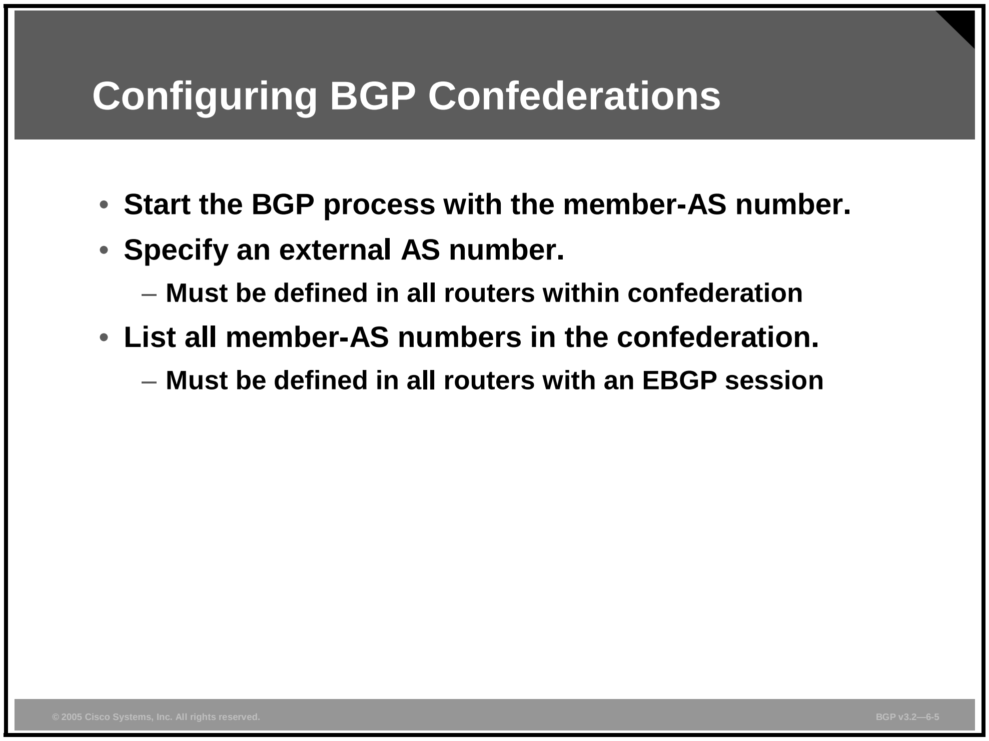
You need to make the following preparations before migrating from a full mesh of IBGP sessions to a confederation design:

- Identify a group of core routers that can serve as a central member-AS.
- Identify several groups of more peripheral routers where, within each group, routers are well connected. Let each group be its own member-AS.
- Make a plan for assigning AS numbers (64512 to 65535) to your member-AS. The plan must uniquely identify each member-AS within the confederation. No member-AS is seen from outside the confederation, so you do not need to coordinate the plan with any other AS.
- Make sure that no router is lacking support (the correct Cisco IOS release level) for a BGP confederation. If any router lacks support for this feature, it will break the network.
- Remove the original BGP configuration with the original official AS number.

Note	Support for the BGP confederation feature is included in Cisco IOS software releases from Release 10.3.

Configuring BGP Confederations

This topic explains how to configure BGP confederations on a BGP backbone. An example of a BGP confederations configuration is also included in this topic.



Configuring BGP Confederations

- **Start the BGP process with the member-AS number.**
- **Specify an external AS number.**
  - **Must be defined in all routers within confederation**
- **List all member-AS numbers in the confederation.**
  - **Must be defined in all routers with an EBGP session**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-5

After you have done the preparation that is necessary to migrate from a full mesh of IBGP sessions to a confederation design, you need to complete the following configuration steps:

Step 1 Make a new BGP configuration that uses the internal member-AS number according to the AS number plan for the confederation.

Step 2 Specify the original official AS number as the identifier of the confederation. This information will be used by egress confederation routers whenever you are communicating with other external autonomous systems.

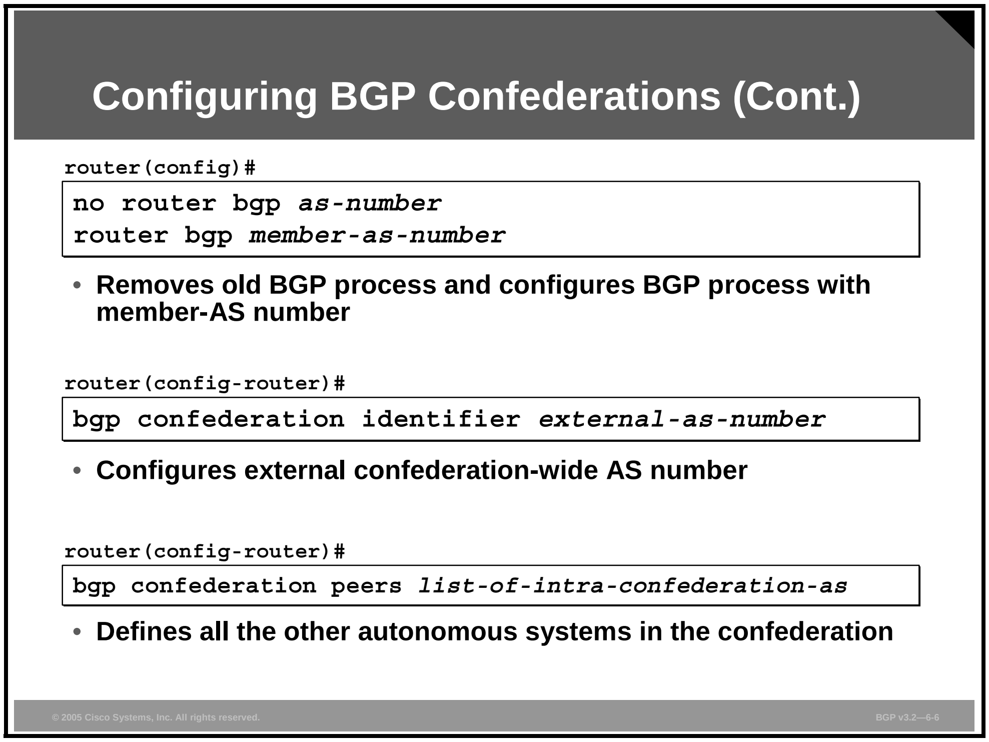
Step 3 Specify a list of the member-AS numbers being used. The router uses this information to distinguish between intra-confederation EBGP behavior and true EBGP behavior.

Step 4 Configure all the IBGP sessions in the full mesh within the member-AS.

Step 5 Configure intra-confederation EBGP sessions between each member-AS in a way that introduces no additional single point of failure.

Step 6 Configure true EBGP sessions with external autonomous systems.

Note Removing the original BGP configuration and creating a new BGP configuration will always cause interruption in network availability. Migration to BGP confederation has to be a well-planned process.



Configuring BGP Confederations (Cont.)

```
router(config)#
no router bgp as-number
router bgp member-as-number
```

- Removes old BGP process and configures BGP process with member-AS number

```
router(config-router)#
bgp confederation identifier external-as-number
```

- Configures external confederation-wide AS number

```
router(config-router)#
bgp confederation peers list-of-intra-confederation-as
```

- Defines all the other autonomous systems in the confederation

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—6-6

router bgp

To configure the BGP routing process, use the **router bgp** global configuration command.

- **router bgp** *as-number*

To remove a routing process, use the **no** form of this command.

- **no router bgp** *as-number*

Syntax Description

Parameter	Description
as-number	Number of an AS that identifies the router to other BGP routers and tags the routing information that is passed along

Note The AS number that is specified when you are configuring a BGP process inside a confederation is the intra-confederation (member) AS number.

bgp confederation identifier

To specify a BGP confederation identifier, use the **bgp confederation identifier** router configuration command.

- **bgp confederation identifier** *external-as-number*

To remove the confederation identifier, use the **no** form of this command.

- **no bgp confederation identifier** *external-as-number*

Syntax Description

Parameter	Description
external-as-number	Public AS number that the confederation is using externally

bgp confederation peers

To configure the autonomous systems that belong to the confederation, use the **bgp confederation peers** router configuration command.

- **bgp confederation peers** *as-number* [*as-number*]

To remove an AS from the confederation, use the **no** form of this command.

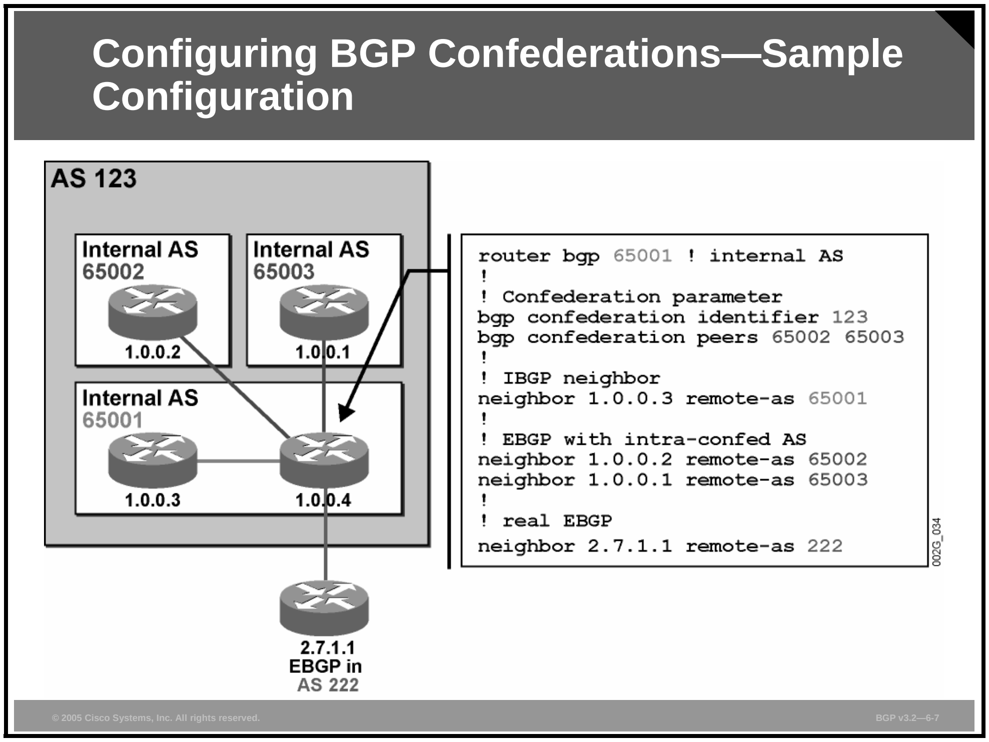
- **no bgp confederation peers** *as-number* [*as-number*]

Syntax Description

Parameter	Description
as-number	AS numbers of member autonomous systems inside the confederation This list is used by the router to distinguish an intra-confederation EBGP session from a real EBGP session with an external AS.

Example: Configuring BGP Confederations

This example illustrates the use of confederation-related BGP configuration commands.



Configuring BGP Confederations—Sample Configuration

```
router bgp 65001 ! internal AS
!
! Confederation parameter
bgp confederation identifier 123
bgp confederation peers 65002 65003
!
! IBGP neighbor
neighbor 1.0.0.3 remote-as 65001
!
! EBGP with intra-confed AS
neighbor 1.0.0.2 remote-as 65002
neighbor 1.0.0.1 remote-as 65003
!
! real EBGP
neighbor 2.7.1.1 remote-as 222
```

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—6-7

Note The example shows only a portion of the configuration of router 1.0.0.4. The configuration displayed is not a complete configuration—only those parts that are relevant to BGP confederations are displayed.

AS 123 is transformed into a BGP confederation. The member-AS 65001 serves as a central member-AS. Member autonomous systems 65002 and 65003 are connected to AS 65001. From the outside, the confederation looks like a single AS, still identified by AS number 123.

The internal member-AS number is specified in the **router bgp** configuration command. This number is the AS number that will be prepended to the AS-path attribute by the router when updates are forwarded across intra-confederation EBGP sessions.

The original, official AS number is given as the BGP confederation identifier. This number is the AS number that will replace the internal member-AS information when egress confederation routers forward updates across EBGP sessions with external autonomous systems.

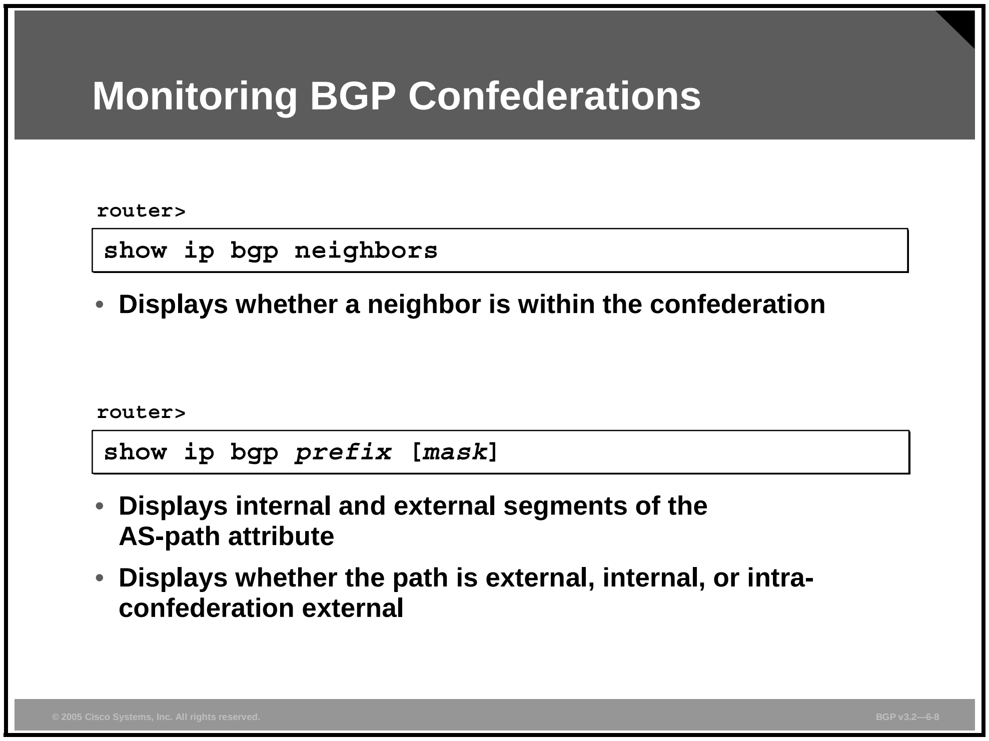
The **bgp confederation peers** configuration step allows the router to identify the type of session to be opened with the BGP peers that are configured through the neighbor statement.

Neighbor 1.0.0.3 is in the same member-AS, so its session is an IBGP session. Both neighbor 1.0.0.2 (AS 65002) and 1.0.0.1 (AS 65003) belong to different member autonomous systems (the AS number appears in the list of BGP confederation peers), so their sessions are intra-confederation EBGP sessions.

Neighbor 2.7.1.1 belongs to AS 222, which is not listed as a member-AS in the **bgp confederation peers** command, so its session is a true EBGP session.

Monitoring BGP Confederations

This topic identifies the basic commands that are used to monitor a BGP backbone that you have configured with BGP confederations.



Monitoring BGP Confederations

```
router>
show ip bgp neighbors
```

- **Displays whether a neighbor is within the confederation**

```
router>
show ip bgp prefix [mask]
```

- **Displays internal and external segments of the AS-path attribute**
- **Displays whether the path is external, internal, or intra-confederation external**

BGP v3.2—6-8

show ip bgp neighbors

To display information about the TCP and BGP connections to neighbors, use the **show ip bgp neighbors** EXEC command.

- **show ip bgp neighbors** [*address*] [**received-routes** | **routes** | **advertised-routes** | {**paths** *regular-expression*} | **dampened-routes**]

Syntax Description

Parameter	Description
address	Address of a specific neighbor about which you wish to display information. If you omit this argument, all neighbors are displayed.
received-routes	Displays all received routes (both accepted and rejected) from the specified neighbor.
routes	Displays all routes that are received and accepted. This is a subset of the output from the **received-routes** keyword.
advertised-routes	Displays all the routes that the router has advertised to the neighbor.
paths *regexp*	Regular expression that is used to match the paths that are received.
dampened-routes	Displays the dampened routes to the neighbor at the IP address that is specified.

If the **show ip bgp neighbors** command is executed on a router without any keywords, the resulting information that is displayed does not show routes or paths that are received by the router, but instead shows the status of its neighbor sessions.

show ip bgp

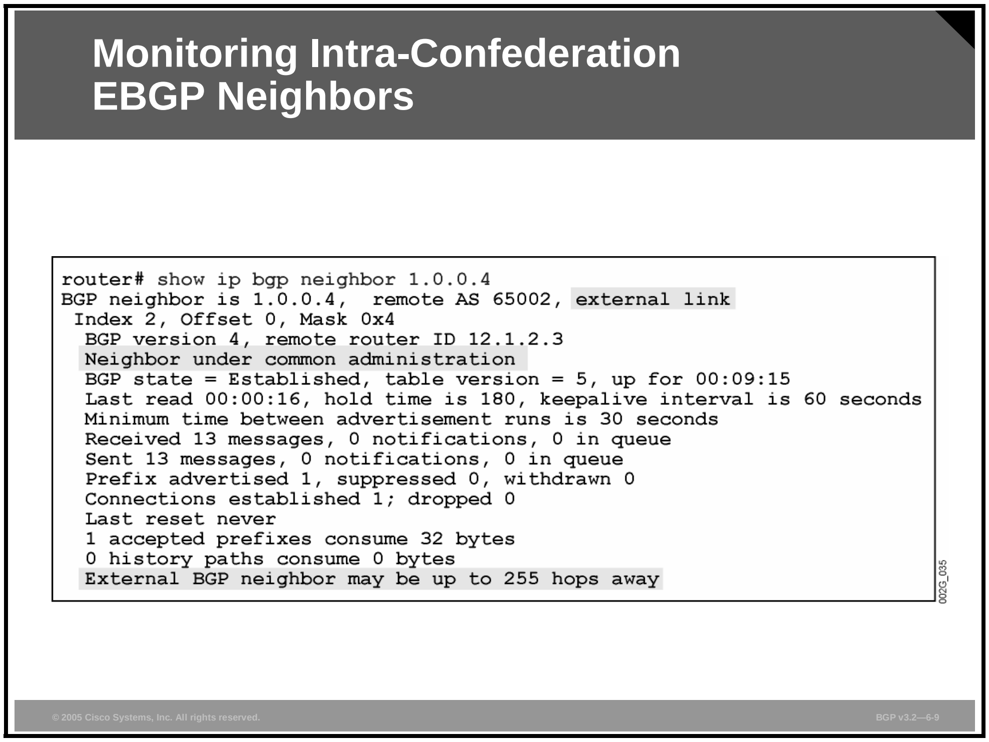
To display entries in the BGP routing table, use the **show ip bgp** EXEC command.

- **show ip bgp** [*network*] [*network-mask*] [**longer-prefixes**]

When details are displayed for a specific route entry in the BGP table, the next hop and AS path are displayed along with information that indicates whether a BGP update was received over an intra-confederation EBGP session or a regular EBGP session.

Syntax Description

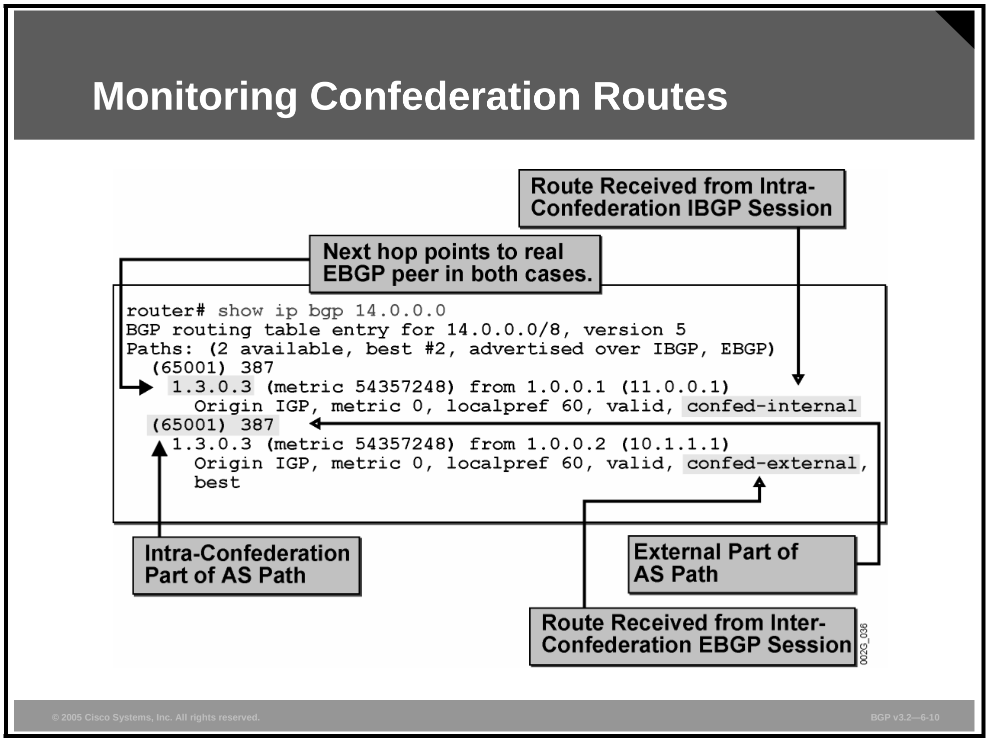
Parameter	Description
network	Network number, which is entered to display a particular network in the BGP routing table
network-mask	Displays all BGP routes that match the address and mask pair
longer-prefixes	Displays the *network* route and its more specific routes



Monitoring Intra-Confederation EBGP Neighbors

```
router# show ip bgp neighbor 1.0.0.4
BGP neighbor is 1.0.0.4,  remote AS 65002, external link
 Index 2, Offset 0, Mask 0x4
  BGP version 4, remote router ID 12.1.2.3
  Neighbor under common administration
  BGP state = Established, table version = 5, up for 00:09:15
  Last read 00:00:16, hold time is 180, keepalive interval is 60 seconds
  Minimum time between advertisement runs is 30 seconds
  Received 13 messages, 0 notifications, 0 in queue
  Sent 13 messages, 0 notifications, 0 in queue
  Prefix advertised 1, suppressed 0, withdrawn 0
  Connections established 1; dropped 0
  Last reset never
  1 accepted prefixes consume 32 bytes
  0 history paths consume 0 bytes
  External BGP neighbor may be up to 255 hops away
```

In the figure, the **show ip bgp neighbors** command has been executed on a router within a confederation. As a result, information about the intra-confederation EBGP session is displayed. The session is an external link (indicating an EBGP session) under common administration (indicating an intra-confederation EBGP session).



Monitoring Confederation Routes

Route Received from Intra-Confederation IBGP Session

Next hop points to real EBGP peer in both cases.

```
router# show ip bgp 14.0.0.0
BGP routing table entry for 14.0.0.0/8, version 5
Paths: (2 available, best #2, advertised over IBGP, EBGP)
  (65001) 387
    1.3.0.3 (metric 54357248) from 1.0.0.1 (11.0.0.1)
      Origin IGP, metric 0, localpref 60, valid, confed-internal
  (65001) 387
    1.3.0.3 (metric 54357248) from 1.0.0.2 (10.1.1.1)
      Origin IGP, metric 0, localpref 60, valid, confed-external,
      best
```

Intra-Confederation Part of AS Path

External Part of AS Path

Route Received from Inter-Confederation EBGP Session

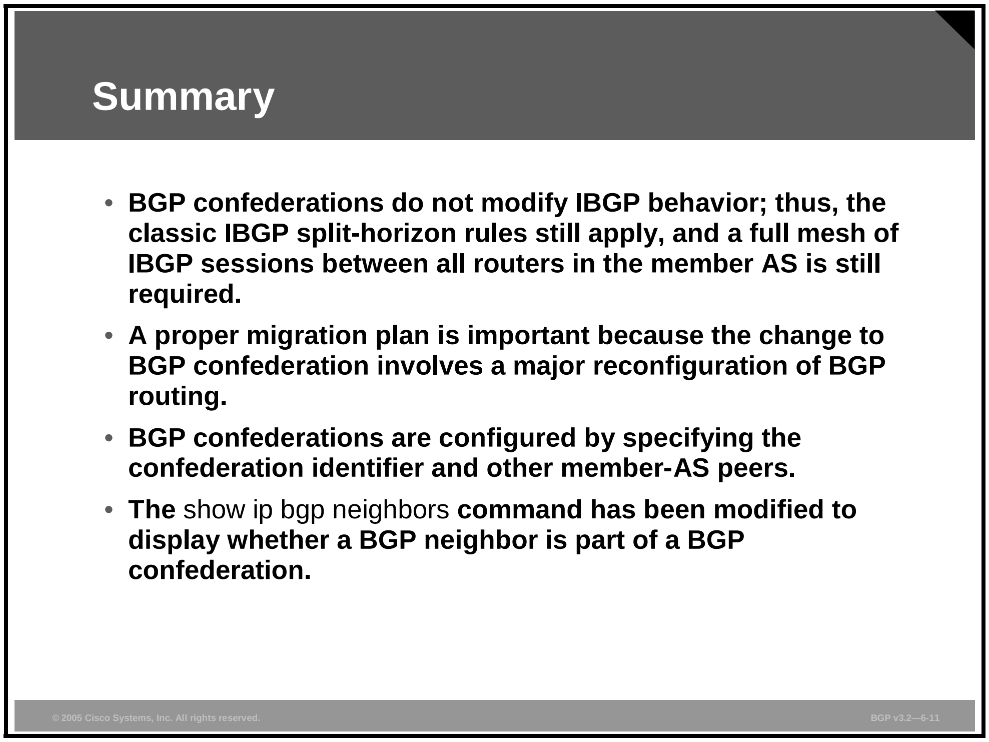
In this example, the **show ip bgp** command is executed on a router within the confederation to display information about the class A network 14.0.0.0.

The command response indicates that the router has received information about the network 14.0.0.0 on two different BGP sessions. One of the sessions is an intra-confederation EBGP session, and the other session is an IBGP session. Both updates have the same next-hop address, which was set by the true EBGP peer that originally sent the update into the confederation. The next hop is resolved by recursive routing, and, therefore, the forwarding decision will be the same regardless of which BGP entries are actually used. The second IP address is the address of the neighbor, which is followed by the router-ID of that neighbor (enclosed in parentheses).

The AS path is the same for both entries. It contains a parenthesized number, (65001). This is the part of the AS path that describes the intra-confederation AS path. The part of the AS path that follows is the external part, 387. This number reveals that the confederation has a true EBGP session with the official AS (AS 387), from which an update about network 14.0.0.0 was received. The update was forwarded to the router using IBGP within member-AS 65001. The router in the local member-AS on which this command was executed has two different intra-confederation EBGP sessions with member-AS 65001. So, the update about network 14.0.0.0 has entered the local AS via two different paths.

Summary

This topic summarizes the key points discussed in this lesson.



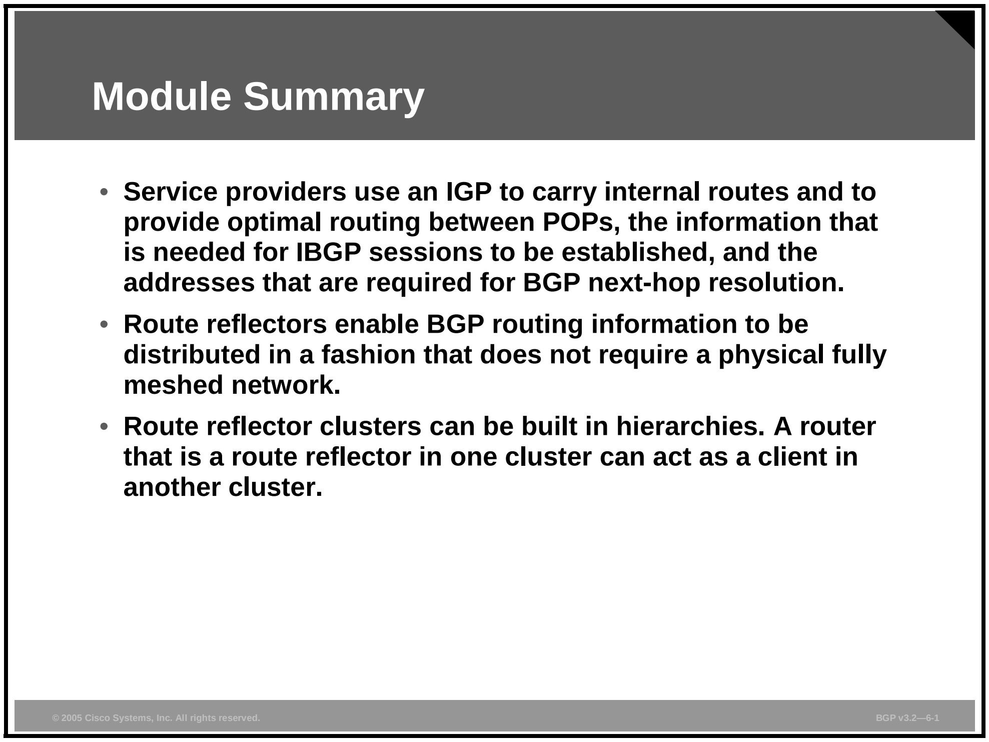
Summary

- **BGP confederations do not modify IBGP behavior; thus, the classic IBGP split-horizon rules still apply, and a full mesh of IBGP sessions between all routers in the member AS is still required.**
- **A proper migration plan is important because the change to BGP confederation involves a major reconfiguration of BGP routing.**
- **BGP confederations are configured by specifying the confederation identifier and other member-AS peers.**
- **The** show ip bgp neighbors **command has been modified to display whether a BGP neighbor is part of a BGP confederation.**

BGP v3.2—6-11

Module Summary

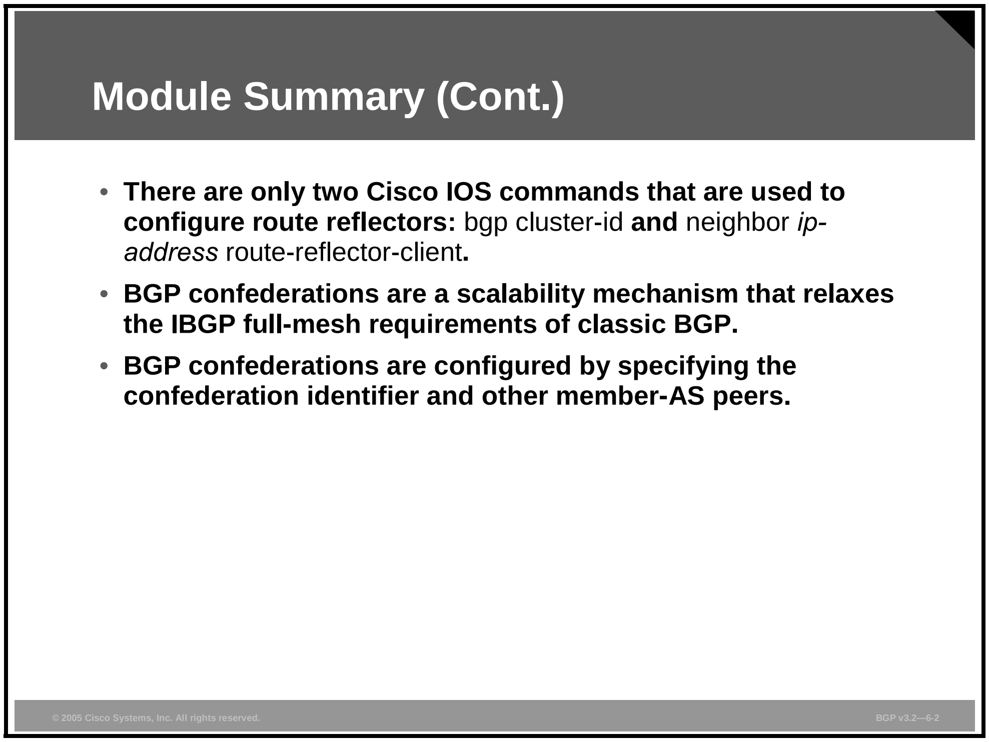
This topic summarizes the key points discussed in this module.



Module Summary

- **Service providers use an IGP to carry internal routes and to provide optimal routing between POPs, the information that is needed for IBGP sessions to be established, and the addresses that are required for BGP next-hop resolution.**
- **Route reflectors enable BGP routing information to be distributed in a fashion that does not require a physical fully meshed network.**
- **Route reflector clusters can be built in hierarchies. A router that is a route reflector in one cluster can act as a client in another cluster.**

BGP v3.2—6-1



Module Summary (Cont.)

- **There are only two Cisco IOS commands that are used to configure route reflectors:** bgp cluster-id **and** neighbor *ip-address* route-reflector-client**.**
- **BGP confederations are a scalability mechanism that relaxes the IBGP full-mesh requirements of classic BGP.**
- **BGP confederations are configured by specifying the confederation identifier and other member-AS peers.**

BGP v3.2—6-2

This module discussed network scalability concerns that are common in large service provider networks. The first lesson outlined the service provider network and described the scaling of IGP and BGP. The next three lessons explained route reflectors—specifically, operating them, the concept of hierarchical route reflectors, and configuring them. The last two lessons described the operation of confederations and how to configure them.

References

For additional information, refer to these resources:

- Cisco Systems, Inc. *Designing Large-Scale IP Internetworks*. http://www.cisco.com/univercd/cc/td/doc/cisintwk/idg4/nd2003.htm.
- Cisco Systems, Inc. *BGP Case Studies*. “BGP Case Studies 4” http://www.cisco.com/warp/public/459/bgp-toc.html#routereflectors.
- Cisco Systems, Inc. *Configuring BGP*. http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122cgcr/fipr_c/ipcprt2/1cfbgp.htm#xtocid45.
- Cisco Systems, Inc. *Using the Border Gateway Protocol for Interdomain Routing*. http://www.cisco.com/univercd/cc/td/doc/cisintwk/ics/icsbgp4.htm.
- Traina, Paul. *Autonomous System Confederations for BGP*. http://www.ietf.org/rfc/rfc1965.txt?number=1965.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) Which three characteristics are common to typical service provider networks? (Choose three.) (Source: Scaling IGP and BGP in Service Provider Networks)

A) The provider network uses two IGPs, one for customer routes and one for internal provider routes.
B) Service providers exchange routes with other providers using BGP.
C) Service providers run IBGP within their network in addition to their IGP requirements.
D) Service providers typically use either static routes or EBGP with their customers.

Q2) What is the typical role of an IGP within a service provider network? (Source: Scaling IGP and BGP in Service Provider Networks)

A) The IGP carries customer routes for redistribution into BGP at the provider edge.
B) The IGP advertises a default route to customers of the service provider.
C) The IGP resolves next-hop IP addresses.
D) The IGP carries BGP routes across the provider network.

Q3) Why should you avoid the use of private IP addressing in service provider networks? (Source: Scaling IGP and BGP in Service Provider Networks)

A) Private addressing can prevent customer network troubleshooting utilities such as traceroute from functioning correctly.
B) Private IP addressing is not allowed on the Internet and will not function in a service provider network.
C) Private IP addressing prevents the service provider from properly summarizing customer routes if it is also using private address space.
D) Private IP addressing prevents service provider applications such as MPLS from operating properly in an Internet-supporting environment.

Q4) Which three requirements are key to properly scaling BGP in a service provider environment? (Choose three.) (Source: Scaling IGP and BGP in Service Provider Networks)

A) IBGP full-mesh scaling tools to reduce duplicate traffic within the AS
B) summarization of customer routes to reduce the number of prefixes that are carried
C) improvement in BGP convergence time by using the IGP for route propagation within the provider AS
D) proper scaling of the AS-wide routing policy to ease administration and maintenance requirements

Q5) Which two of the following statements about the function of BGP policy accounting are accurate? (Choose two.) (Source: Scaling IGP and BGP in Service Provider Networks)

A) BGP policy accounting is enabled on an output interface.
B) BGP policy accounting using AS numbers can be used to improve the design of network circuit peering and transit agreements between ISPs.
C) In BGP policy accounting, counters based on parameters such as community-list, AS number, or AS path are assigned to identify the IP traffic.
D) A Cisco IOS policy-based classifier maps the traffic into one of five possible buckets, representing different traffic classes.

Q6) What is the main problem that is solved by implementing BGP route reflectors? (Source: Introducing Route Reflectors)

A) the large number of routes that are carried in the IGP when BGP is deployed
B) the ability of BGP to scale a single AS in a large network
C) the need for a homogeneous method of applying policies to routes that are carried through an AS
D) the ability to support service-level parameters with greater ease

Q7) How does a route reflector modify the IBGP split-horizon rule? (Source: Introducing Route Reflectors)

A) forwards EBGP updates to all peers (IBGP and EBGP)
B) treats all neighbors as EBGP peers, which eliminates the IBGP mesh requirements
C) forwards IBGP updates from clients to other IBGP neighbors
D) appends the cluster-ID to the AS path, which allows peers to be treated as EBGP neighbors

Q8) Why are redundant route reflectors mandatory in any high-availability network design? (Source: Introducing Route Reflectors)

A) All neighbors peer with the route reflector, and a large number of neighbors can make the route reflector router unstable.
B) EBGP peers can inject BGP updates into the AS only through the route reflector.
C) Route reflectors maintain more routing information, which makes them more prone to congestion and failure.
D) Clients can form IBGP relationships only with the route reflector.

Q9) What is the main reason for implementing redundant route reflectors with clusters? (Source: Introducing Route Reflectors)

A) to eliminate routing loops in redundant configurations
B) to limit the number of neighbor sessions with each route reflector
C) to provide another scalability mechanism targeted at removing the IBGP full-mesh requirement
D) to enhance security within the AS

Q10) How does the originator-ID attribute assist in the elimination of routing loops that are caused by redundant route reflector designs? (Source: Introducing Route Reflectors)

A) If the originator-ID matches the router-ID of the reflector, local preference is set on the route to make it a backup.
B) The originator-ID attribute is set to the cluster-ID to ensure that a route traverses the AS only one time.
C) A router that receives a route in which the originator-ID matches its router-ID will ignore that route.
D) The originator-ID allows the router to know if the route originated locally or from an external source so that administrative distance rules for the route can be verified.

Q11) What can occur if a client has IBGP neighbor relationships with other routers that are not configured as route reflectors? (Source: Designing Networks with Route Reflectors)

A) This is an invalid configuration.
B) The client will notify the route reflector and be promoted to a route reflector as well.
C) Routing black holes can occur and cause lost traffic inside the AS.
D) Unnecessary routing traffic will be generated.

Q12) Which potential problem can occur if a client does not have an IBGP session with all route reflectors in a cluster? (Source: Designing Networks with Route Reflectors)

A) This is an invalid configuration.
B) The client might not receive all BGP routes.
C) EBGP routes that are received by the client will not be distributed properly throughout the AS.
D) Duplicate routing traffic will be sent to the client.

Q13) Which problem are hierarchical route reflectors designed to solve? (Source: Designing Networks with Route Reflectors)

A) lack of a consistent application of security and routing policies throughout the AS
B) scalability of autonomous systems in very large routing domains
C) routing loops caused by redundant cluster configurations
D) administrative overhead when you are implementing router reflector network designs

Q14) Which two BGP parameters do you have to configure on a route reflector? (Choose two.) (Source: Configuring and Monitoring Route Reflectors)

A) cluster-ID
B) originator-ID
C) cluster-list
D) route reflector clients

Q15) What are three migration steps that are required to convert from a fully meshed IBGP AS to an AS that is based on route reflectors? (Choose three.) (Source: Configuring and Monitoring Route Reflectors)

A) remove unnecessary IBGP sessions
B) configure the clients on the route reflectors
C) configure IBGP sessions between route reflector clients
D) configure the cluster-ID on the route reflectors

Q16) Which command should you use to identify route reflector clients without inspecting the router configuration? (Source: Configuring and Monitoring Route Reflectors)

A) **show ip bgp** *prefix*
B) **show ip bgp neighbors**
C) **show ip bgp clients**
D) **show ip bgp summary**

Q17) What is the main problem that is solved by implementing BGP confederations? (Source: Introducing Confederations)

A) the large number of routes that are carried in the IGP when BGP is deployed
B) the ability for BGP to scale a single AS in a large network
C) the need for a homogeneous method of applying policies to routes that are carried through an AS
D) the ability to support service level parameters with greater ease

Q18) Although confederations eliminate the need for a fully meshed topology within the AS, where does the BGP full-mesh requirement still apply? (Source: Introducing Confederations)

A) to all EBGP neighbor sessions
B) inside each member AS
C) between the member autonomous systems in the confederation
D) no longer applies when confederations are used in an AS

Q19) How does an IBGP router that receives the AS-path attribute in a BGP update determine whether the route has crossed a member-AS within a confederation? (Source: Introducing Confederations)

A) The router can determine this fact by the presence of the confederation bit in the flag field of the BGP update.
B) The router can determine this fact because the AS-path attribute will contain only the AS number of the ingress EBGP peer.
C) The member-AS numbers will be indicated by the presence of parentheses surrounding the AS number entry.
D) The IBGP router cannot determine whether the route has crossed a member-AS because the AS number of each AS boundary that has been crossed is appended to the AS-path attribute.

Q20) How does an EBGP router that receives the AS-path attribute in a BGP update determine whether the route has crossed a member-AS within a confederation? (Source: Introducing Confederations)

A) The router can determine this fact by the presence of the confederation bit in the flag field of the BGP update.
B) The router can determine this fact because the AS-path attribute will contain only the AS number of the ingress EBGP peer.
C) The member-AS numbers will be indicated by the presence of parentheses surrounding the AS number entry.
D) The EBGP router cannot determine whether the route has crossed a member-AS because the member-AS entries are removed from the AS path prior to exiting the confederation.

Q21) Why is it impossible for a router that does not support BGP confederations to operate inside an AS that is configured as a confederation? (Source: Introducing Confederations)

A) The router will believe that the AS path is longer than the actual AS path and route incorrectly.
B) The router will be unable to interpret the AS-path attribute and terminate its BGP session with a peer.
C) The router will automatically convert the intra-confederation AS numbers to the external AS number of the confederation, causing an AS number mismatch.
D) The router will process BGP updates as normal because it has to be aware only of the member-AS to which it belongs, causing incorrect routing information to propagate through the AS.

Q22) Which three IBGP properties are retained within the confederation even though EBGP sessions between member autonomous systems are formed? (Choose three.) (Source: Introducing Confederations)

A) local preference
B) MED
C) weight
D) next-hop

Q23) How can you reduce the IBGP full mesh within a confederation AS? (Source: Configuring and Monitoring Confederations)

A) You cannot reduce the full mesh because all IBGP peers must be fully meshed within the member-AS.
B) Implementing router reflectors inside a member-AS can reduce the IBGP full-mesh requirement.
C) You can nest a confederation within a confederation to remove the IBGP full-mesh requirement.
D) Because confederations are used, there is no requirement for an IBGP full mesh within each member-AS.

Q24) Which two BGP parameters do you need to specify on every router within a confederation? (Choose two.) (Source: Configuring and Monitoring Confederations)

A) a list of all AS numbers in the confederation
B) a list of all true EBGP sessions
C) the official AS number (as the identifier of the confederation)
D) the correct MD5 authentication password in each peer

Q25) What does the BGP confederation identifier define? (Source: Configuring and Monitoring Confederations)

A) the AS number of the confederation external peer
B) the public AS number that the confederation is using externally
C) the AS number of the member-AS
D) the MD5 authentication password for the confederation

Q26) How will the **show ip bgp** command display the intra-confederation segment of the AS path? (Source: Configuring and Monitoring Confederations)

A) as a regular entry in the AS-path attribute
B) as a separate AS-path list independent of the AS-path attribute
C) as an entry in the AS path enclosed by parentheses
D) not displayed because it is not a part of the EBGP AS path

Module Self-Check Answer Key

Q1)	B, C, D
Q2)	C
Q3)	A
Q4)	A, B, D
Q5)	B, C
Q6)	B
Q7)	C
Q8)	D
Q9)	A
Q10)	C
Q11)	D
Q12)	B
Q13)	B
Q14)	A, D
Q15)	A, B, D
Q16)	B
Q17)	B
Q18)	B
Q19)	C
Q20)	D
Q21)	B
Q22)	A, B, D
Q23)	B
Q24)	A, C
Q25)	B
Q26)	C

Module 7

Optimizing BGP Scalability

Overview

The Border Gateway Protocol (BGP) is designed for reliability and scalability. As such, it has become the de facto standard protocol that is used to carry the more than 110,000 prefixes in the Internet today. Likewise, BGP has a tremendous amount of flexibility with regard to administrative policy controls, route selection, performance tuning, and scalability features. This module introduces advanced BGP configuration tools that are designed to improve BGP scalability and performance. Tools that are discussed in this module include convergence time reduction features, limiting the number of prefixes, peer groups, and route dampening.

Module Objectives

Upon completing this module, you will be able to use available BGP tools and features to optimize the scalability of the BGP routing protocol in a typical BGP network. This ability includes being able to meet these objectives:

- Configure Cisco IOS performance improvements to reduce BGP convergence time
- Configure BGP to limit the number of prefixes that are received from a neighbor
- Use BGP peer groups to share common configuration parameters between multiple BGP peers
- Use route dampening to minimize the impact of unstable routes

Lesson 1

Improving BGP Convergence

Overview

As the number of routes in the Internet increases, demands on router CPU and memory resources will increase. Border Gateway Protocol (BGP) processing affects both router resources and network convergence time. It is important that network convergence be as fast as possible to ensure accurate routing information between domains. It is also important that router resources be optimized whenever possible. Cisco IOS performance improvements for BGP are designed to aid network administrators in achieving these goals.

This lesson introduces various Cisco IOS performance improvements that have been designed to reduce BGP convergence time. Included in this lesson are discussions of convergence, BGP routing processes, and the effects of BGP routing processes on router CPU resources. The lesson also discusses the commands that are required to configure and monitor BGP for various Cisco IOS performance improvements in the areas of path maximum transmission unit (PMTU) discovery, input hold queue, BGP scan time, and BGP advertisement interval.

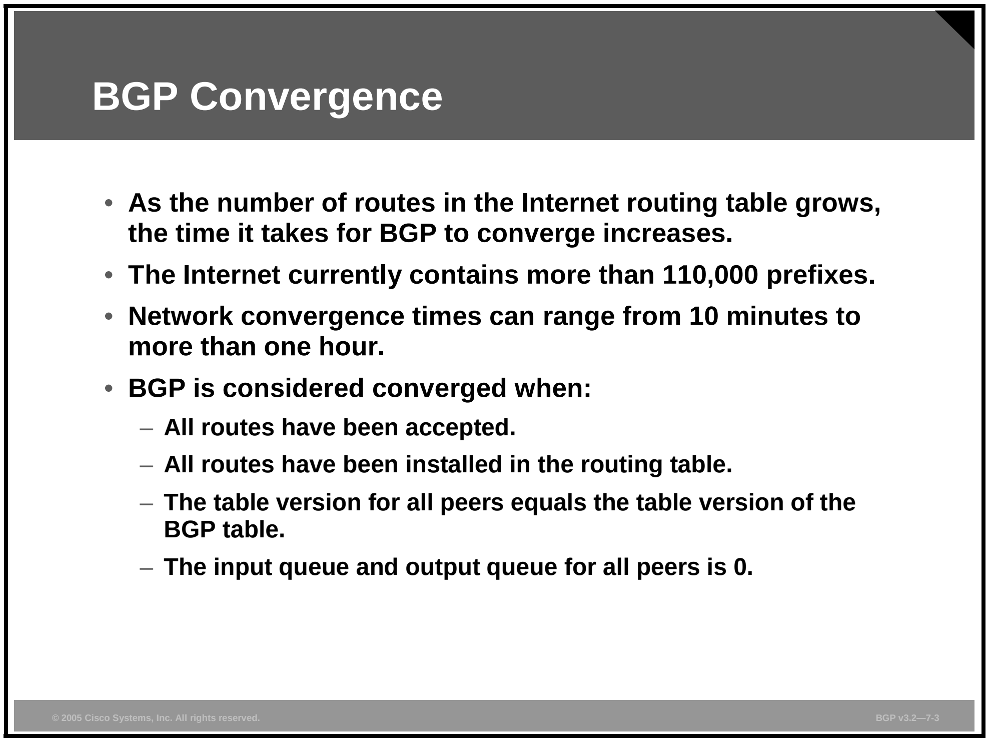
Objectives

Upon completing this lesson, you will be able to configure Cisco IOS performance improvements to reduce BGP convergence time. This ability includes being able to meet these objectives:

- Describe convergence in BGP networks
- Describe the BGP router processes and their functions
- Describe the effects of BGP processes on router CPU resources
- Identify Cisco IOS performance improvements to reduce BGP convergence time
- Identify the Cisco IOS commands required to configure and monitor PMTU discovery
- Identify the Cisco IOS commands required to configure and monitor the input queue depth on a router interface
- Identify the Cisco IOS commands required to configure and monitor BGP scan time
- Identify the Cisco IOS commands required to configure and monitor the BGP advertisement interval
- Describe the function of the BGP Nonstop Forwarding Awareness feature

BGP Convergence

This topic describes the concept of convergence in BGP networks.



BGP Convergence

- As the number of routes in the Internet routing table grows, the time it takes for BGP to converge increases.
- The Internet currently contains more than 110,000 prefixes.
- Network convergence times can range from 10 minutes to more than one hour.
- BGP is considered converged when:
  - All routes have been accepted.
  - All routes have been installed in the routing table.
  - The table version for all peers equals the table version of the BGP table.
  - The input queue and output queue for all peers is 0.

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—7-3

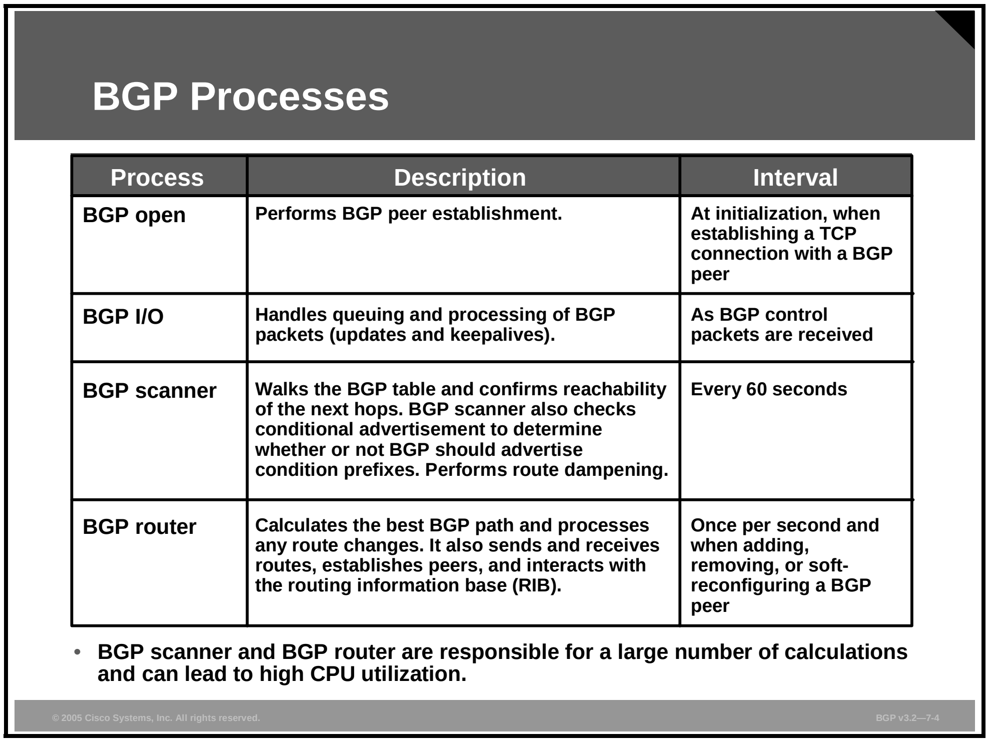
As the number of routes in the Internet routing table grows, service providers and large enterprise customers are experiencing a dramatic increase in the amount of time that BGP takes to converge. Networks that once converged in 10 or 15 minutes may now take up to an hour in some cases, and even longer in extreme situations. In general, convergence is defined as the process of bringing all routing tables to a state of consistency. The BGP routing protocol is considered converged when the following conditions are true:

- All routes have been accepted.
- All routes have been installed in the routing table.
- The table version for all peers equals the table version of the BGP table.
- The input queue and output queue for all peers is 0.

Convergence time is an important consideration in a network, because nonconverged networks can cause routing loops, packet delays, and even packet loss as a result of black holes.

BGP Processes

This topic describes the different BGP router processes and their functions.



BGP Processes

Process	Description	Interval
BGP open	**Performs BGP peer establishment.**	**At initialization, when establishing a TCP connection with a BGP peer**
BGP I/O	**Handles queuing and processing of BGP packets (updates and keepalives).**	**As BGP control packets are received**
BGP scanner	**Walks the BGP table and confirms reachability of the next hops. BGP scanner also checks conditional advertisement to determine whether or not BGP should advertise condition prefixes. Performs route dampening.**	**Every 60 seconds**
BGP router	**Calculates the best BGP path and processes any route changes. It also sends and receives routes, establishes peers, and interacts with the routing information base (RIB).**	**Once per second and when adding, removing, or soft-reconfiguring a BGP peer**

- **BGP scanner and BGP router are responsible for a large number of calculations and can lead to high CPU utilization.**

BGP v3.2—7-4

In general, a Cisco IOS process consists of the individual threads and associated data that perform router tasks, such as system maintenance, packet switching, and implementing routing protocols.

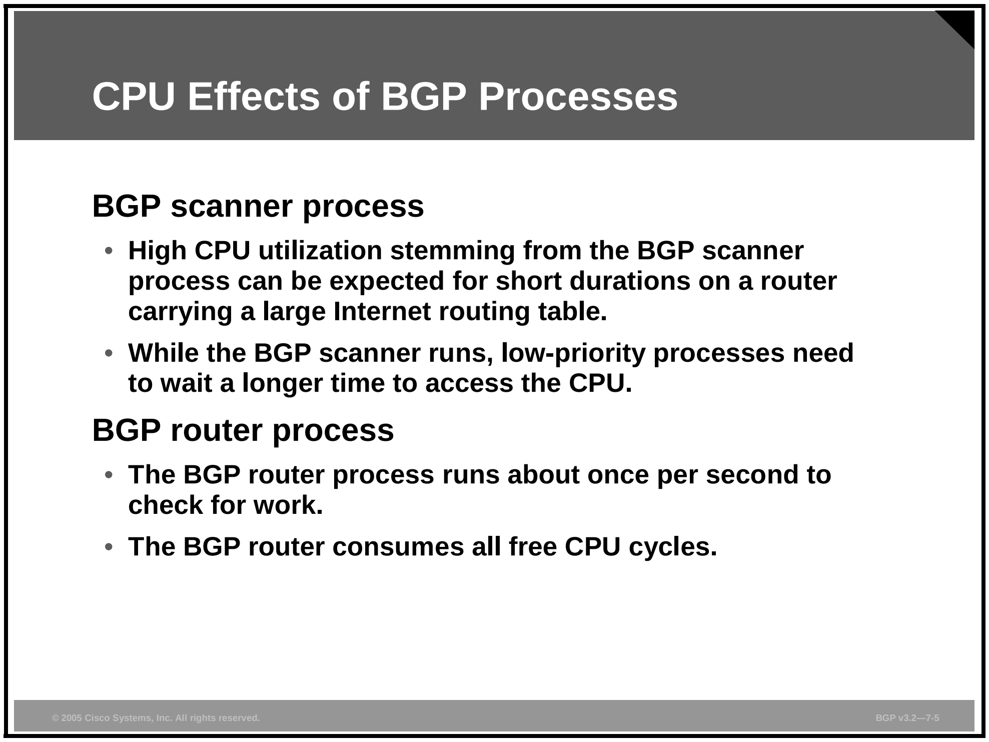
Note A thread is an information placeholder that allows a single process to be halted (interrupted) on the router so that the CPU can service another process. The information that is contained within the thread allows the interrupted process to restart exactly where it left off when the CPU is ready to continue to service that process thread.

Several Cisco IOS processes that are executed on the router enable BGP to run. You can use the **show process cpu | include BGP** command to see the volume of CPU resources that are consumed (utilization) because of running BGP processes.

The figure lists the function of each of the BGP router processes and how often each process is executed on the router. It shows that each process runs at different times, depending on the tasks that are handled by the specific process. Because BGP scanner and BGP router are responsible for a large number of calculations, you may notice high CPU utilization during the running of either one of these processes.

CPU Effects of BGP Processes

This topic describes how running BGP router processes affects router CPU resources.



CPU Effects of BGP Processes

BGP scanner process

- High CPU utilization stemming from the BGP scanner process can be expected for short durations on a router carrying a large Internet routing table.
- While the BGP scanner runs, low-priority processes need to wait a longer time to access the CPU.

BGP router process

- The BGP router process runs about once per second to check for work.
- The BGP router consumes all free CPU cycles.

On routers that carry a large Internet routing table, high CPU utilization stemming from the BGP scanner process can be expected for short periods of time. Once per minute, the BGP scanner "walks" (scans) the BGP routing table and performs important maintenance tasks. These tasks include checking the next hop that is referenced in the BGP table of the router and verifying that the next-hop devices can be reached. Thus, a large BGP table takes an equally large amount of time to be walked and validated.

The BGP scanner walks the BGP routing table to update any data structures and walks the table for route redistribution purposes. In this context, the routing table is also known as the routing information base (RIB), which the router outputs when the **show ip route** command is executed. Both tables are stored separately in the router memory and can be very large, thus consuming CPU and memory resources.

Because the BGP scanner runs through the entire BGP table, the duration of the high CPU utilization condition that is caused by the BGP scanner process varies with the number of neighbors and the number of routes that are learned per neighbor.

While the BGP scanner runs, low-priority processes need to wait a longer time to access the CPU. One low-priority process controls Internet Control Message Protocol (ICMP) packets such as pings. Packets that are destined to or have originated from the router may experience higher than expected latency because the ICMP process must wait behind the BGP scanner. The BGP scanner process runs for some time, is suspended, then ICMP runs, ICMP is suspended, the BGP scanner runs, and so on. In contrast, pings sent through a router should be switched via Cisco Express Forwarding (CEF) and should not experience any additional latency. When you are troubleshooting periodic spikes in latency, compare forwarding times for packets that are forwarded through a router versus packets that are processed directly by the CPU on the router.

The BGP router process runs about once per second to check for work. BGP convergence defines the duration between the time when the first BGP peer is established and the point at which BGP is converged. To ensure the shortest possible convergence times, the BGP router consumes all free CPU cycles. However, after it starts, it relinquishes (or suspends) the CPU intermittently.

Example: CPU Effects of BGP Processes

Convergence time is a direct measurement of how long the BGP router process runs on the CPU, not the total time that the process is actually running. This example investigates the high CPU utilization condition during BGP convergence as BGP exchanges prefixes with two External Border Gateway Protocol (EBGP) peers.

Step 1 Capture a baseline for normal CPU utilization before starting the test.

```
router# show process cpu

CPU utilization for five seconds: 0%/0%; one minute: 4%; five minutes: 5%
```

Step 2 After the test starts, the CPU reaches 100 percent utilization. The **show process cpu** command shows that the high CPU condition is caused by the BGP router, denoted by 139 (the Cisco IOS process ID for the BGP router) in the following output:

```
router# show process cpu

CPU utilization for five seconds: 100%/0%; one minute: 99%; five minutes:

81%[output omitted] 139     6795740   1020252   6660 88.34% 91.63% 74.01%   0
BGP Router
```

Step 3 Monitor the router by capturing multiple outputs of the **show ip bgp summary** and **show process cpu** commands during the event. The **show ip bgp summary** command captures the state of the BGP neighbors.

```
router# show ip bgp summary

Neighbor    V    AS  MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down State/PfxRcd

10.1.1.1    4  64512  309453  157389    19981    0  253 22:06:44 111633

172.16.1.1  4  65101  188934    1047    40081   41    0 00:07:51 58430
```

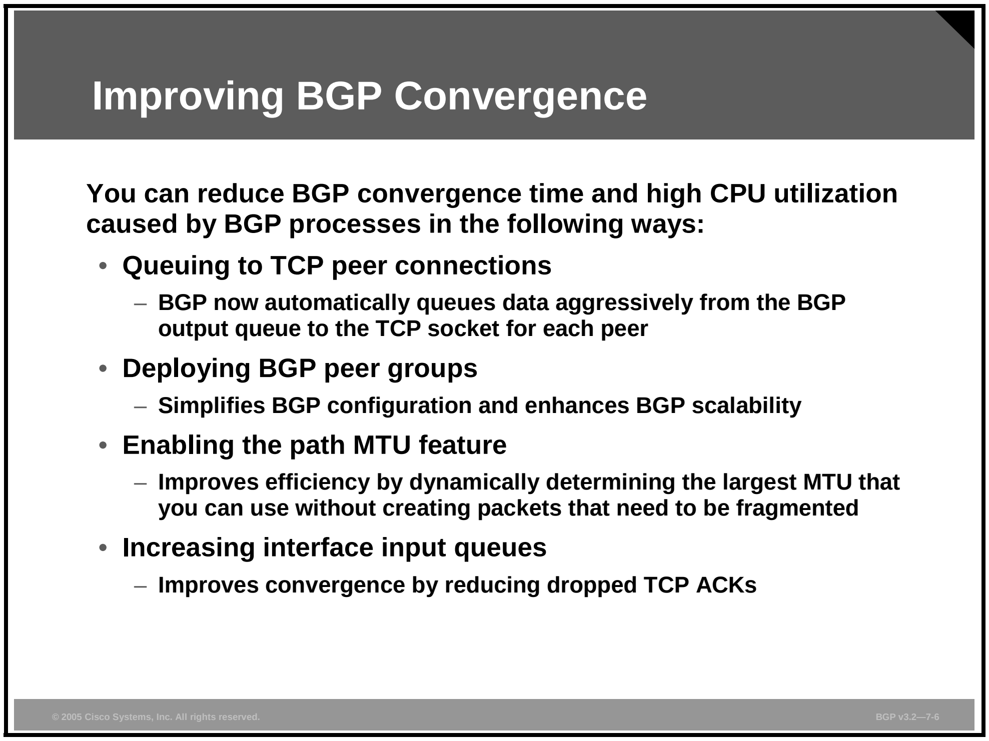
Step 4 When the router completes prefix exchange with its BGP peers, the CPU utilization rates should return to normal levels. The computed 1-minute and 5-minute averages will settle back down as well but may show higher than normal levels for a longer period than the 5-second rate.

```
router# show process cpu

CPU utilization for five seconds: 3%/0%; one minute: 82%; five minutes: 91%
```

Step 5 Using the output from the **show** commands will allow you to compute the BGP convergence time. In particular, the Up/Down column of the **show ip bgp summary** command is compared to the start and stop times of the high CPU utilization condition. Typically, BGP convergence can take several minutes when routers exchange a large Internet routing table.

Improving BGP Convergence

This topic identifies Cisco IOS performance improvements that reduce BGP convergence times.



Improving BGP Convergence

You can reduce BGP convergence time and high CPU utilization caused by BGP processes in the following ways:

- **Queuing to TCP peer connections**
  - **BGP now automatically queues data aggressively from the BGP output queue to the TCP socket for each peer**
- **Deploying BGP peer groups**
  - **Simplifies BGP configuration and enhances BGP scalability**
- **Enabling the path MTU feature**
  - **Improves efficiency by dynamically determining the largest MTU that you can use without creating packets that need to be fragmented**
- **Increasing interface input queues**
  - **Improves convergence by reducing dropped TCP ACKs**

BGP convergence can often be an issue in networks requiring quick propagation of routing information. Cisco IOS software provides the following performance-improvement features, which have been designed to reduce BGP convergence time and the high CPU utilization that is caused by a running BGP process:

- **Queuing to TCP peer connections:** Instead of queuing data once per second, BGP now queues data aggressively from the BGP output queue to the TCP socket for each peer until the output queues have drained completely. Because BGP now sends at a faster rate, it converges more quickly.

- **BGP peer groups:** The major benefit of specifying a BGP peer group is that it reduces the volume of system resources (CPU and memory) that are used in BGP update generation. Peer groups also simplify BGP configuration because many repetitive configuration elements (such as filters) are applied by the router only once (to the peer group) instead of applying them to each neighbor.

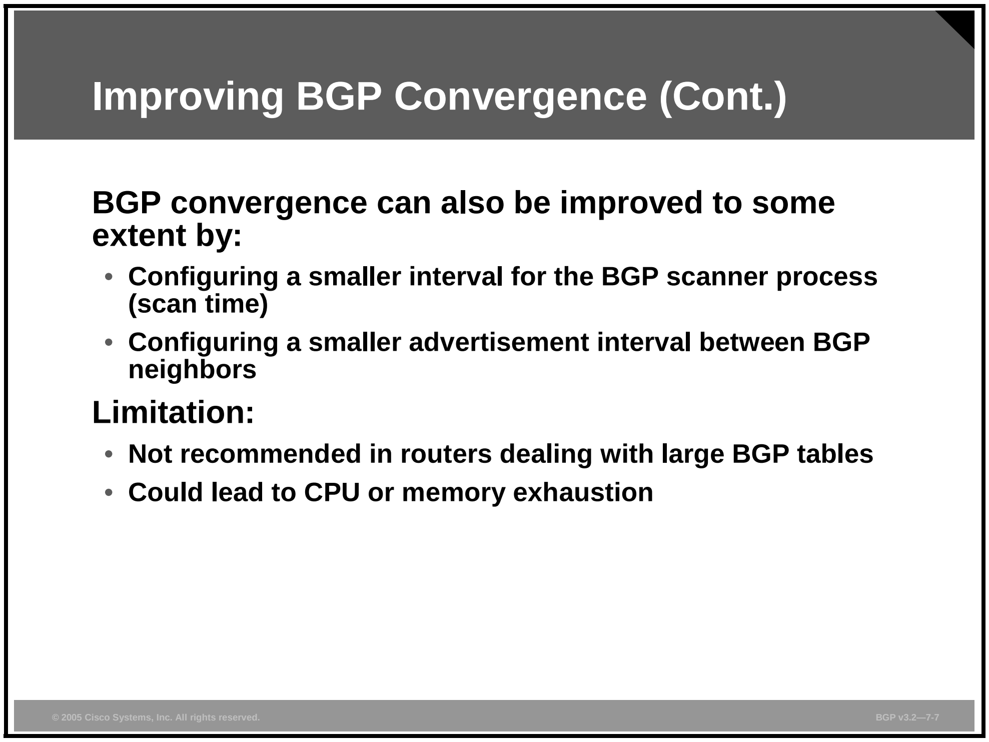
  Because peer groups allow the routing table to be checked only once and allow updates to be replicated to all other in-sync peer group members (depending on the number of peer group members, the number of prefixes in the table, and the number of prefixes that are advertised), they can significantly reduce router resource requirements.

- **PMTU feature:** All TCP sessions are bounded by a limit on the number of bytes that a single packet can transport. This limit, known as the Maximum Segment Size (MSS), is 536 bytes by default. In other words, TCP breaks up packets in a transmit queue into 536-byte chunks before passing packets down to the IP layer. The advantage of a 536-byte MSS is that packets are not likely to be fragmented at an IP device along the path to the destination, because most links use an MTU of at least 1500 bytes. The disadvantage is that smaller packets increase the amount of bandwidth that is used for transport overhead.

  Because BGP builds a TCP connection to all peers, a 536-byte MSS affects BGP convergence times. The solution is to enable the PMTU feature by means of the **ip tcp path-mtu-discovery** command. You can use this feature to dynamically determine how large the MSS value can be without creating packets that need to be fragmented. PMTU allows TCP to determine the smallest MTU size among all links in a TCP session. TCP then uses this MTU value, minus room for the IP and TCP headers, as the MSS for the session.

- **Increase interface input queues:** If BGP is advertising thousands of routes to many neighbors, TCP must transmit thousands of packets. BGP peers receive these packets and send TCP acknowledgments (ACKs) to the advertising BGP speaker, causing the BGP speaker to receive a flood of TCP ACKs in a short period of time. If the ACKs arrive at a rate that is too high for the router CPU, packets back up in inbound interface queues.

  By default, router interfaces use an input queue size of 75 packets. In addition, special control packets such as BGP updates use a special queue with Selective Packet Discard (SPD). This special queue holds 100 packets. During BGP convergence, TCP ACKs can quickly fill the 175 spots of input buffering, causing newly arriving packets to be dropped. On routers with 15 or more BGP peers that also exchange the full Internet routing table, more than 10,000 drops per interface per minute may be seen. Increasing the interface input queue depth using the **hold-queue in** command helps reduce the number of dropped TCP ACKs, reducing the amount of work that BGP must do to converge.



Improving BGP Convergence (Cont.)

BGP convergence can also be improved to some extent by:

- **Configuring a smaller interval for the BGP scanner process (scan time)**
- **Configuring a smaller advertisement interval between BGP neighbors**

Limitation:

- **Not recommended in routers dealing with large BGP tables**
- **Could lead to CPU or memory exhaustion**

BGP v3.2—7-7

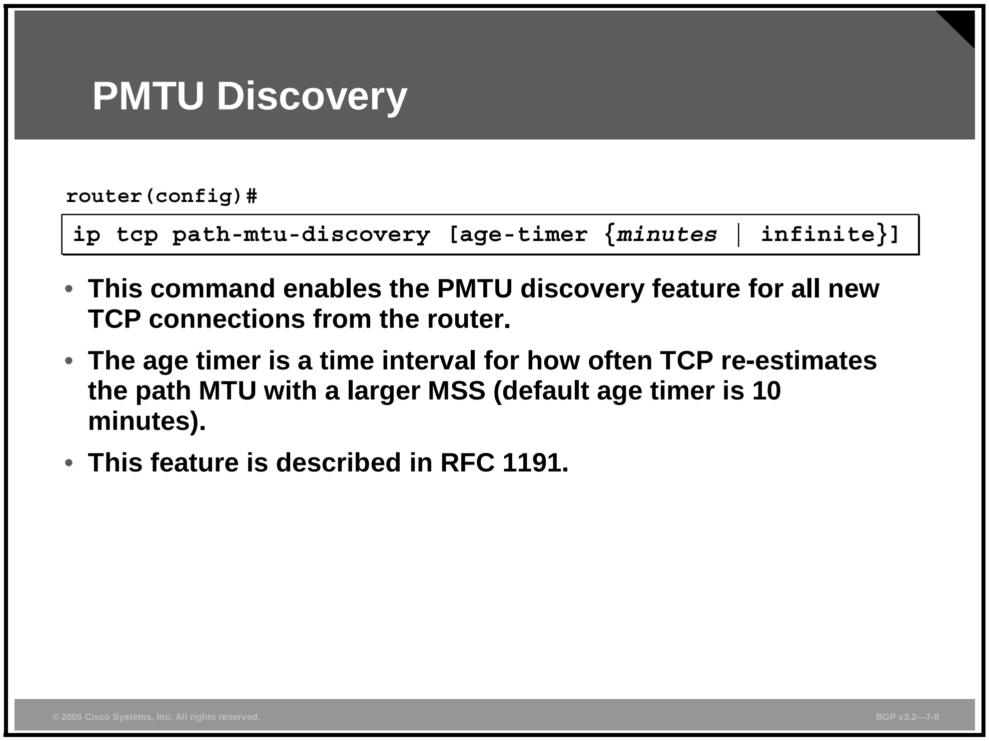
Network administrators also need to improve BGP convergence in certain scenarios; for example, in networks using the conditional advertisement feature. There are two additional BGP parameters that they can use to influence BGP convergence speed:

- **Scan time:** Controlling the BGP scanner process, responsible for verifying information in the BGP table
- **Advertisement interval:** Controlling the rate at which successive advertisements are sent to a BGP neighbor

Network administrators must take care when configuring these two parameters. Setting the values too low for a specific network environment could lead to a significant consumption of router resources. The larger the BGP tables and the more unstable the BGP network, the greater the danger of exhausting the resources of a router.

PMTU Discovery

This topic identifies the Cisco IOS commands that are required to configure and monitor PMTU discovery.



PMTU Discovery

```
router(config)#
ip tcp path-mtu-discovery [age-timer {minutes | infinite}]
```

- **This command enables the PMTU discovery feature for all new TCP connections from the router.**
- **The age timer is a time interval for how often TCP re-estimates the path MTU with a larger MSS (default age timer is 10 minutes).**
- **This feature is described in RFC 1191.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-8

PMTU discovery is a method for maximizing the use of available bandwidth in the network between the endpoints of a TCP connection. It is described in RFC 1191. Existing connections are not affected when this feature is turned on or off.

Customers using TCP connections to move bulk data between systems on distinct subnets would benefit most by enabling this feature.

The age timer is a time interval for how often TCP re-estimates the PMTU with a larger MSS. The default value of the age timer is 10 minutes, but it can be manually configured up to 30 minutes or disabled (set to infinite). If the MSS that is used for the connection is smaller than the peer connection can handle, the router will attempt to use a larger MSS each time that the age timer expires. The discovery process is stopped when either the sent MSS is as large as the peer negotiated or the user has disabled the timer on the router. You can turn off the age timer by setting it to "infinite."

ip tcp path-mtu-discovery

To enable the PMTU discovery feature for all new TCP connections from the router, use the **ip tcp path-mtu-discovery** global configuration command.

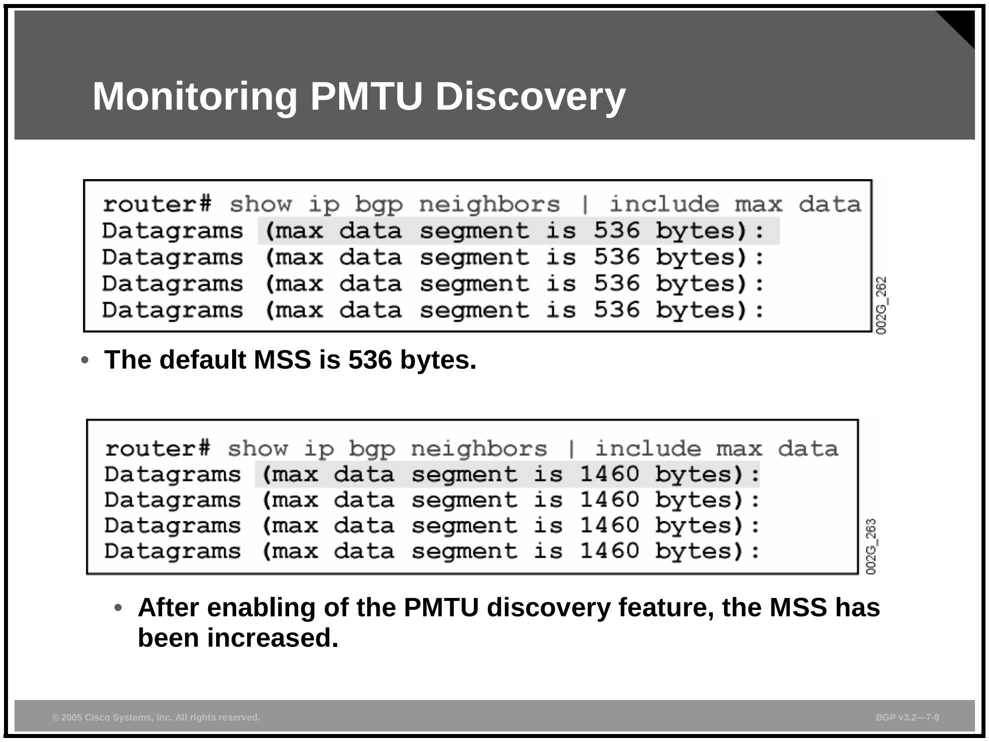
- **ip tcp path-mtu-discovery** [**age-timer** {*minutes* | **infinite**}]

To disable the function, use the **no** form of this command.

- **no ip tcp path-mtu-discovery** [**age-timer** {*minutes* | **infinite**}]

Syntax Description

Parameter	Description
age-timer *minutes*	(Optional) Time interval (in minutes) after which TCP re-estimates the PMTU with a larger MSS. The maximum interval is 30 minutes; the default is 10 minutes.
age-timer infinite	(Optional) Turns off the age timer.



Monitoring PMTU Discovery

```
router# show ip bgp neighbors | include max data
Datagrams (max data segment is 536 bytes):
Datagrams (max data segment is 536 bytes):
Datagrams (max data segment is 536 bytes):
Datagrams (max data segment is 536 bytes):
```

- **The default MSS is 536 bytes.**

```
router# show ip bgp neighbors | include max data
Datagrams (max data segment is 1460 bytes):
Datagrams (max data segment is 1460 bytes):
Datagrams (max data segment is 1460 bytes):
Datagrams (max data segment is 1460 bytes):
```

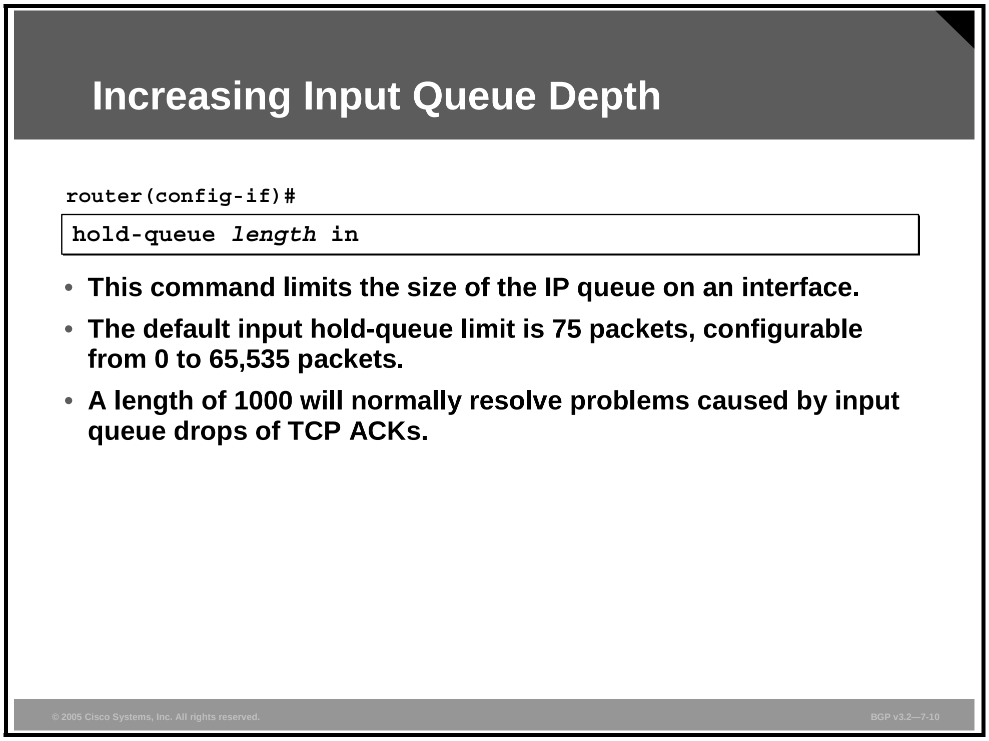
- **After enabling of the PMTU discovery feature, the MSS has been increased.**

By default, the MSS is 536 bytes. As shown in the figure, the **show ip bgp neighbors** | **include max data** command can be used to verify the size of the MSS before the PMTU discovery feature is enabled on the router.

After using the **ip tcp path-mtu-discovery** command to enable PMTU discovery, the router dynamically determines how large the MSS can be without creating IP packets that require fragmentation. At the bottom of figure, the output shows that the PMTU feature has been enabled and the **show ip bgp neighbors** | **include max data** command has been used to determine that the PMTU discovery feature has set the MSS to 1460 bytes.

Increasing Input Queue Depth

This topic identifies the Cisco IOS commands that are required to configure and monitor the input queue depth on a router interface.



Increasing Input Queue Depth

```
router(config-if)#
hold-queue length in
```

- This command limits the size of the IP queue on an interface.
- The default input hold-queue limit is 75 packets, configurable from 0 to 65,535 packets.
- A length of 1000 will normally resolve problems caused by input queue drops of TCP ACKs.

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-10

Each interface owns an input queue into which incoming packets are placed to await processing by the router. Frequently, the rate at which incoming packets are placed in the input queue exceeds the rate at which the router can process the packets. Each input queue has a size that indicates the maximum number of packets that can be placed in the queue. After the input queue becomes full, the interface drops any new incoming packets.

hold-queue

To specify the size of the IP input or output queue on an interface, use the **hold-queue** command in interface configuration mode.

- **hold-queue** *length* {**in** | **out**}

To restore the default values for an interface, use the **no** form of this command with the appropriate keyword.

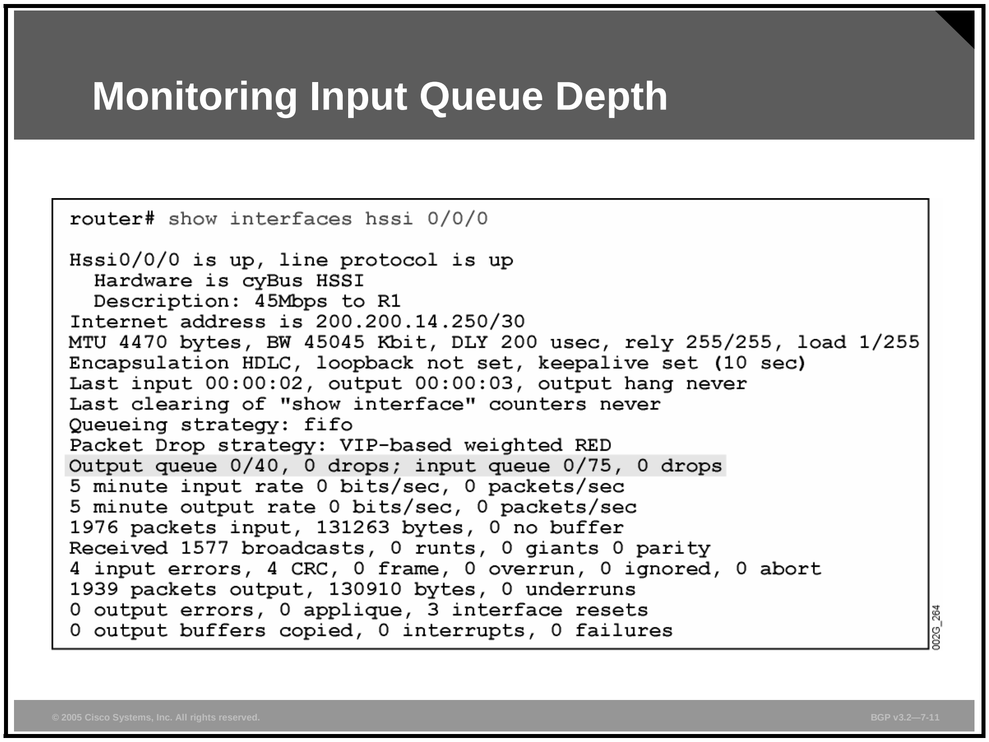
- **no hold-queue** *length* {**in** | **out**}

Syntax Description

Parameter	Description
length	Integer that specifies the maximum number of packets in the queue. The range of allowed values is 0 to 65535.
in	Specifies the input queue. The default is 75 packets. For asynchronous interfaces, the default is 10 packets. These limits prevent a malfunctioning interface from consuming an excessive amount of memory.
out	Specifies the output queue. The default is 40 packets. For asynchronous interfaces, the default is 10 packets. These limits prevent a malfunctioning interface from consuming an excessive amount of memory.

Caution Increasing the hold queue can have detrimental effects on network routing and response times. For protocols that use SEQ or ACK packets to determine round-trip times, do not increase the output queue. Dropping packets instead informs hosts to slow down transmissions to match available bandwidth. This approach is generally better than having duplicate copies of the same packet within the network (which can happen with large hold queues).

Note The Cisco 12000 Series now uses a default SPD headroom value of 1000. It retains the default input queue size of 75. Use the **show spd** command to view these special input queues.



Monitoring Input Queue Depth

```
router# show interfaces hssi 0/0/0

Hssi0/0/0 is up, line protocol is up
  Hardware is cyBus HSSI
  Description: 45Mbps to R1
Internet address is 200.200.14.250/30
MTU 4470 bytes, BW 45045 Kbit, DLY 200 usec, rely 255/255, load 1/255
Encapsulation HDLC, loopback not set, keepalive set (10 sec)
Last input 00:00:02, output 00:00:03, output hang never
Last clearing of "show interface" counters never
Queueing strategy: fifo
Packet Drop strategy: VIP-based weighted RED
Output queue 0/40, 0 drops; input queue 0/75, 0 drops
5 minute input rate 0 bits/sec, 0 packets/sec
5 minute output rate 0 bits/sec, 0 packets/sec
1976 packets input, 131263 bytes, 0 no buffer
Received 1577 broadcasts, 0 runts, 0 giants 0 parity
4 input errors, 4 CRC, 0 frame, 0 overrun, 0 ignored, 0 abort
1939 packets output, 130910 bytes, 0 underruns
0 output errors, 0 applique, 3 interface resets
0 output buffers copied, 0 interrupts, 0 failures
```

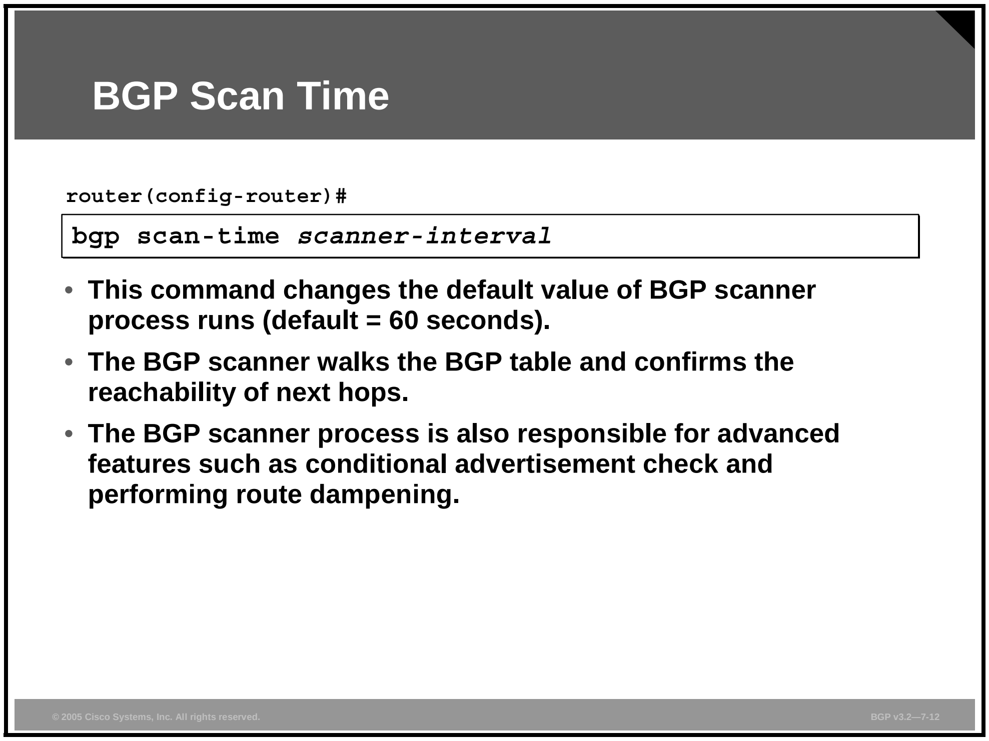
In the figure, the **show interface** {*interface-identifier*} command displays the current input queue levels and the number of incoming packets dropped.

The input queue *x*/*y* counter displays the current number of packets in the input queue *x* and the current size of the input queue *y*. The drops counter indicates the number of incoming packets that have been dropped.

If the current number of packets in the input queue is consistently at or greater than 80 percent of the current size of the input queue, the size of the input queue may require tuning to accommodate the rate of incoming packets. Even if the current number of packets in the input queue never seems to approach the size of the input queue, bursts of packets may still be overflowing the queue. If the drops counter is increasing at a high rate, the size of the input queue may require tuning to accommodate the bursts.

BGP Scan Time

This topic identifies the Cisco IOS commands that are required to configure and monitor BGP scan time.



BGP Scan Time

```
router(config-router)#
bgp scan-time scanner-interval
```

- **This command changes the default value of BGP scanner process runs (default = 60 seconds).**
- **The BGP scanner walks the BGP table and confirms the reachability of next hops.**
- **The BGP scanner process is also responsible for advanced features such as conditional advertisement check and performing route dampening.**

Network administrators use the **bgp scan-time** command to configure the time interval for repetitions of the BGP scanner process.

The BGP scanner process walks (scans) the BGP table and confirms the reachability of next hops. A change of this status triggers a new BGP route selection for the network. Changes are then propagated by the router to established BGP neighbors. Increasing the BGP scanner process frequency will make the router find a changed status more quickly, but it will also consume more CPU resources.

The BGP scanner process is also responsible for some advanced BGP features. It checks the conditional advertisement to determine whether or not BGP should advertise conditional prefixes or perform route dampening.

bgp scan-time

To configure a nondefault value of the scanning interval for BGP routing information, use the **bgp scan-time** command.

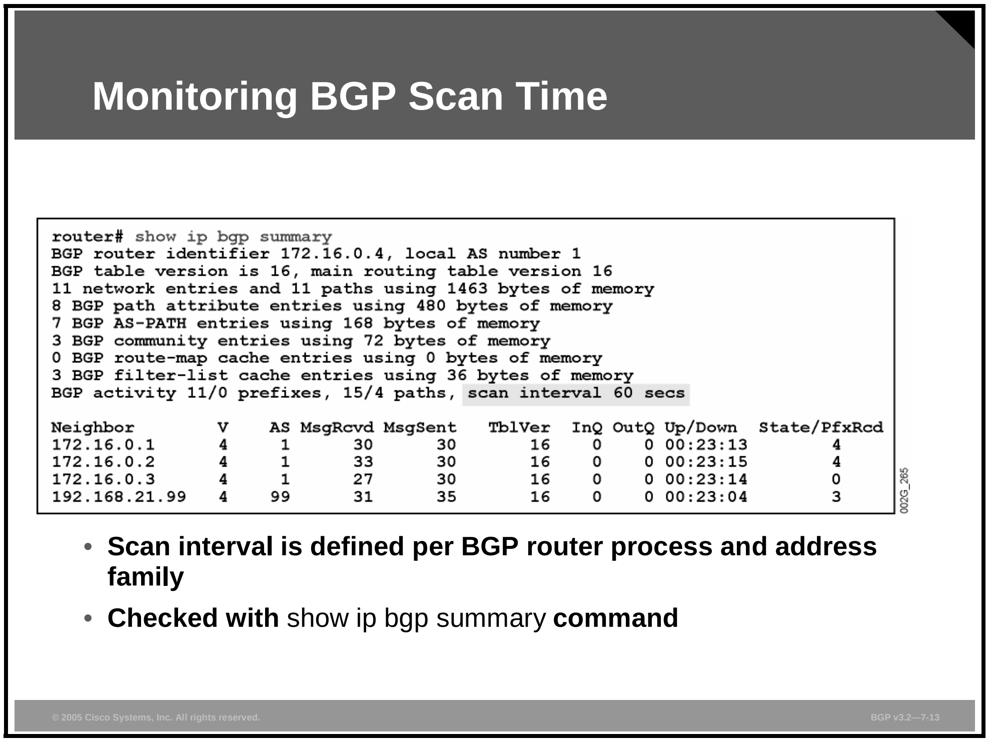
- **bgp scan-time** [**import**] *scanner-interval*

To disable the scanning interval of a router, use the **no** form of this command.

- **no bgp scan-time**

Syntax Description

Parameter	Description
import	(Optional) Configures import processing of Virtual Private Network version 4 (VPNv4) unicast routing information from BGP routers into routing tables.
scanner-interval	Specifies the scanning interval of BGP routing information. Valid values that are used for selecting the desired scanning interval are from 5 to 60 seconds. By default, the scanning interval is 60 seconds.



Monitoring BGP Scan Time

```
router# show ip bgp summary
BGP router identifier 172.16.0.4, local AS number 1
BGP table version is 16, main routing table version 16
11 network entries and 11 paths using 1463 bytes of memory
8 BGP path attribute entries using 480 bytes of memory
7 BGP AS-PATH entries using 168 bytes of memory
3 BGP community entries using 72 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
3 BGP filter-list cache entries using 36 bytes of memory
BGP activity 11/0 prefixes, 15/4 paths, scan interval 60 secs

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
172.16.0.1      4     1      30      30       16    0    0 00:23:13        4
172.16.0.2      4     1      33      30       16    0    0 00:23:15        4
172.16.0.3      4     1      27      30       16    0    0 00:23:14        0
192.168.21.99   4    99      31      35       16    0    0 00:23:04        3
```

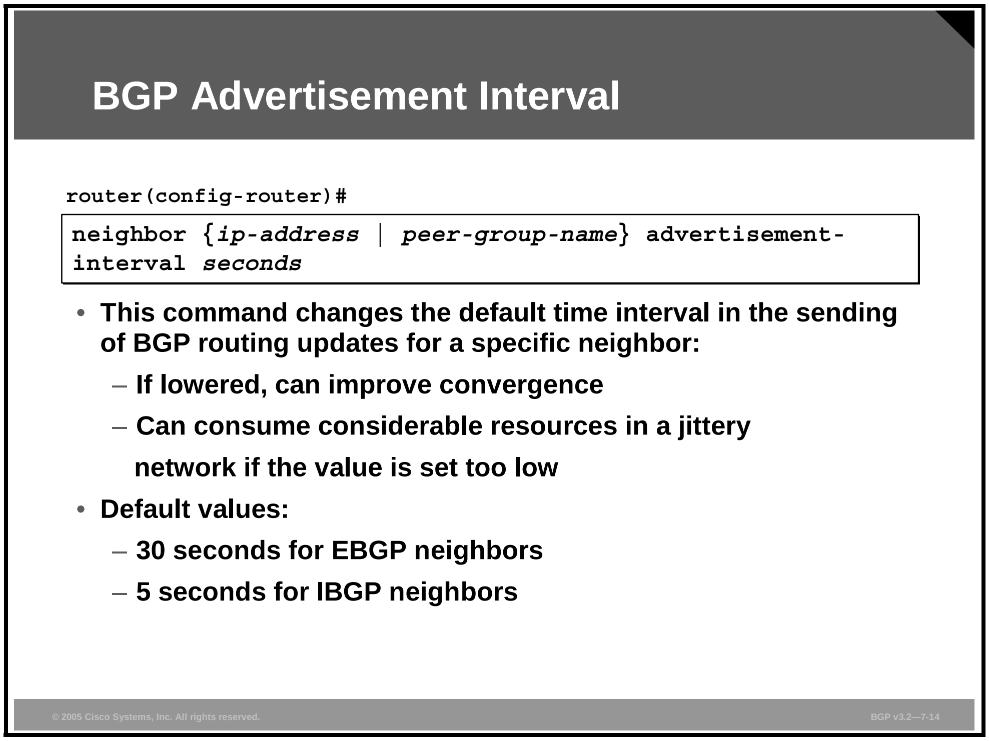
- **Scan interval is defined per BGP router process and address family**
- **Checked with** show ip bgp summary **command**

You can check the configured BGP scan interval using the **show ip bgp summary** command.

The configured BGP scan interval will apply to the entire BGP routing protocol process.

BGP Advertisement Interval

This topic identifies the Cisco IOS commands that are required to configure and monitor the BGP advertisement interval.



BGP Advertisement Interval

```
router(config-router)#
```

```
neighbor {ip-address | peer-group-name} advertisement-
interval seconds
```

- **This command changes the default time interval in the sending of BGP routing updates for a specific neighbor:**
  - **If lowered, can improve convergence**
  - **Can consume considerable resources in a jittery network if the value is set too low**
- **Default values:**
  - **30 seconds for EBGP neighbors**
  - **5 seconds for IBGP neighbors**

With the help of the **neighbor advertisement-interval** command, network administrators can modify the default advertisement interval for a specific BGP peer. BGP advertisements are rate-limited by the router that is using the advertisement interval timer (defined as the parameter MinRouteAdvertisementInterval in RFC 1771). When a BGP-speaking router sends a route update to a neighbor for a specific destination, it is not allowed to send another update to the neighbor about the same destination until a period of time equal to the advertisement interval has elapsed. In that way, the advertisement interval timer acts as a form of rate limiting on a per-destination basis, even though the value of the advertisement interval is configured for each neighbor.

It is important to note that during the time that the router is waiting for the advertisement interval timer to expire, the router can still receive and process route updates from BGP neighbors. Using the **neighbor advertisement-interval** command does not rate-limit BGP route selection (inbound updates and subsequent processing) but only the rate of outgoing route advertisements.

When faster propagation of successive BGP updates (which are batched and rate-limited) is required, the network administrator can lower the default value of the advertisement interval, thus improving convergence.

In routers that handle large BGP tables and less stable networks, lowering the advertisement interval could potentially lead to consuming large portions of router resources.

neighbor advertisement-interval

To set the minimum interval in the sending of BGP routing updates, use the **neighbor advertisement-interval** router configuration command.

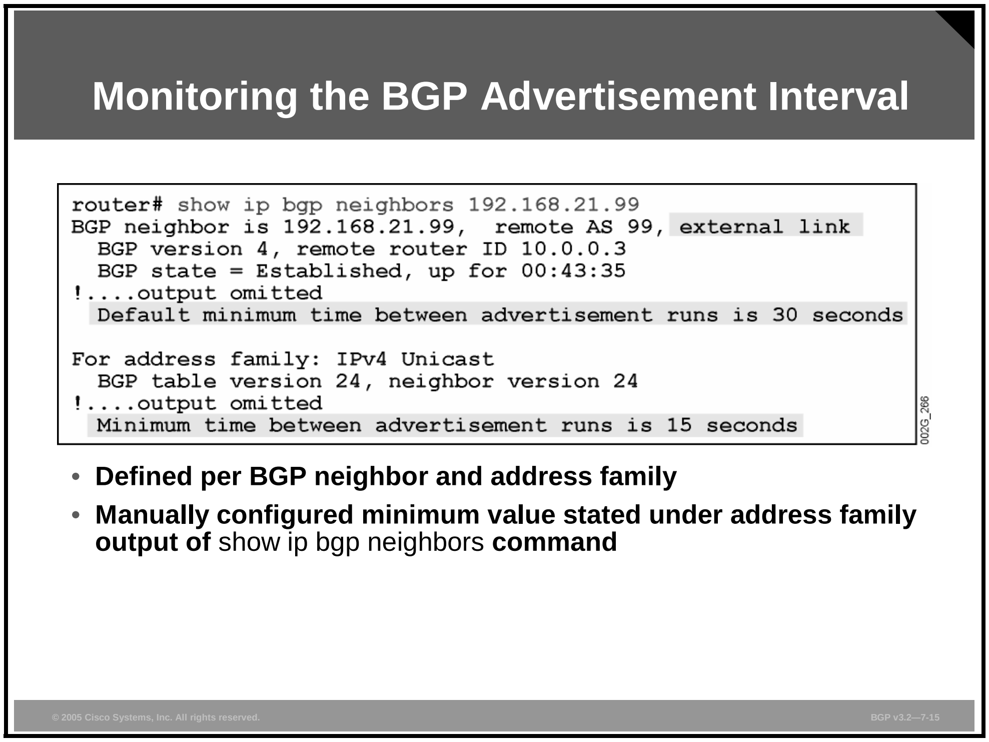
- **neighbor** {*ip-address* | *peer-group-name*} **advertisement-interval** *seconds*

To remove an entry, use the **no** form of this command.

- **no neighbor** {*ip-address* | *peer-group-name*} **advertisement-interval** *seconds*

Syntax Description

Parameter	Description
ip-address	Neighbor IP address.
peer-group-name	Name of a BGP peer group. If a BGP peer group is specified by using the *peer-group-name* argument, all members of the peer group will inherit the characteristic that is configured with this command.
seconds	Time in seconds. Integer from 0 to 600. The default is 30 seconds for external peers and 5 seconds for internal peers.



Monitoring the BGP Advertisement Interval

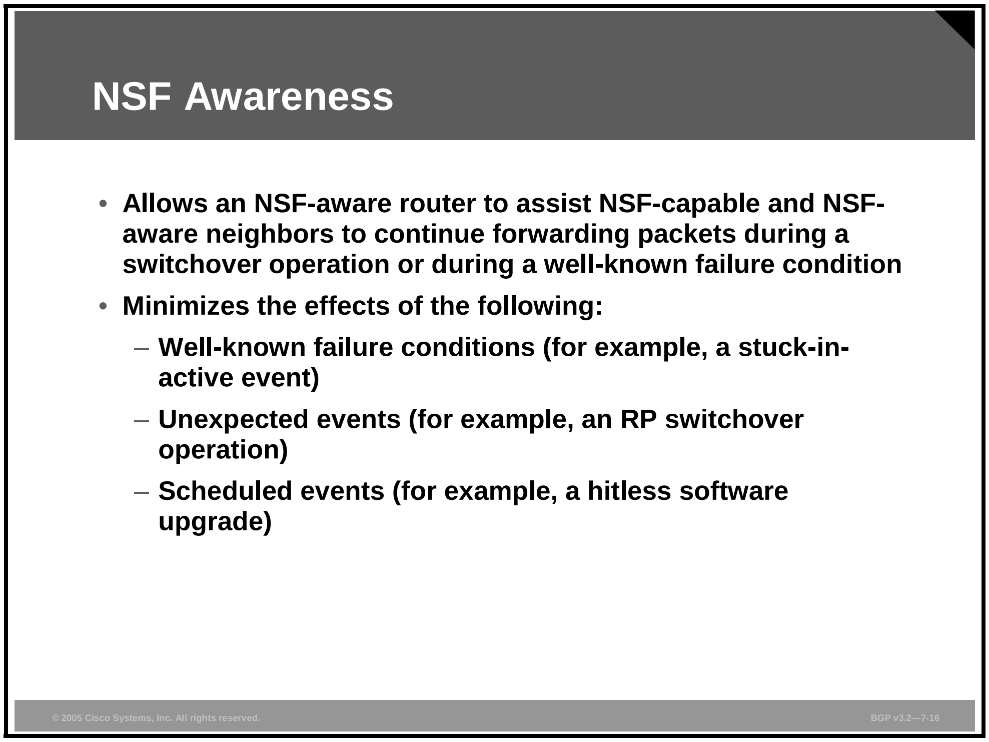
```
router# show ip bgp neighbors 192.168.21.99
BGP neighbor is 192.168.21.99,  remote AS 99, external link
  BGP version 4, remote router ID 10.0.0.3
  BGP state = Established, up for 00:43:35
!....output omitted
  Default minimum time between advertisement runs is 30 seconds

For address family: IPv4 Unicast
  BGP table version 24, neighbor version 24
!....output omitted
  Minimum time between advertisement runs is 15 seconds
```

- **Defined per BGP neighbor and address family**
- **Manually configured minimum value stated under address family output of** show ip bgp neighbors **command**

You can examine the currently configured BGP advertisement interval with the **show ip bgp neighbors** command. The advertisement interval is defined for a specific neighbor in a specific BGP address family. Actual values of the advertisement interval are therefore stated under the specific address-family portion of the neighbor output. The default timer for EBGP is 30 seconds. After the parameter was changed in the example shown in the figure, the timer switched to 15 seconds.

BGP Nonstop Forwarding Awareness

This topic describes the function of the BGP Nonstop Forwarding Awareness feature.



NSF Awareness

- **Allows an NSF-aware router to assist NSF-capable and NSF-aware neighbors to continue forwarding packets during a switchover operation or during a well-known failure condition**
- **Minimizes the effects of the following:**
  - **Well-known failure conditions (for example, a stuck-in-active event)**
  - **Unexpected events (for example, an RP switchover operation)**
  - **Scheduled events (for example, a hitless software upgrade)**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-16

Cisco Nonstop Forwarding (NSF) awareness allows an NSF-aware router to assist NSF-capable and NSF-aware neighbors to continue forwarding packets during a switchover operation or during a well-known failure condition. The Enhanced Interior Gateway Routing Protocol (EIGRP) Nonstop Forwarding Awareness feature allows an NSF-aware router that is running EIGRP to forward packets along routes that are already known for a router that is performing a switchover operation or is in a well-known failure mode. This capability allows the EIGRP peers of the failing router to retain the routing information that is advertised by the failing router and continue to use this information until the failed router has returned to normal operating behavior and is able to exchange routing information. The peering session is maintained throughout the entire NSF operation.

The deployment of EIGRP NSF awareness can minimize the effects of the following:

- Well-known failure conditions (for example, a stuck-in-active event)
- Unexpected events (for example, a route processor [RP] switchover operation)
- Scheduled events (for example, a hitless software upgrade)

EIGRP NSF awareness is enabled by default, and its operation is transparent to the network operator and EIGRP peers that do not support NSF capabilities.

NSF-aware routers are completely compatible with non-NSF-aware or -capable neighbors in an EIGRP network. A non-NSF-aware neighbor will ignore NSF capabilities and reset the adjacency when they are received.

The NSF-capable router will drop any queries that are received while converging to minimize the number of transient routes that are sent to neighbors. But the NSF-capable router will still acknowledge these queries to prevent these neighbors from resetting adjacency.

Note An NSF-aware router must be up and completely converged with the network before it can assist an NSF-capable router in an NSF restart operation.

The route-hold timer sets the maximum period of time that the NSF-aware router will hold known routes for an NSF-capable neighbor during a switchover operation or a well-known failure condition. The route-hold timer is configurable so that you can tune network performance and avoid undesired effects, such as “black holing” routes if the switchover operation takes too much time. When this timer expires, the NSF-aware router scans the topology table and discards any stale routes, allowing EIGRP peers to find alternate routes instead of waiting during a long switchover operation.

timers nsf route-hold

To set the route-hold timer to determine how long an NSF-aware router that is running EIGRP will hold routes for an inactive peer, use the **timers nsf route-hold** command in router configuration mode.

- **timers nsf route-hold** *seconds*

To return the route-hold timer to the default value, use the **no** form of this command.

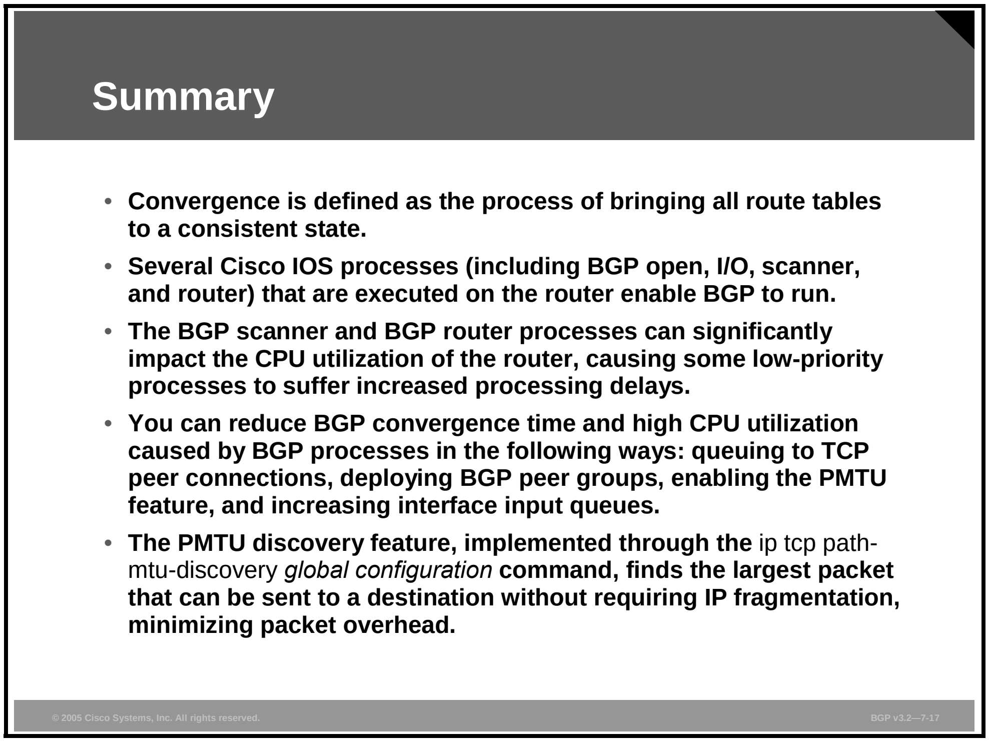
- no **timers nsf route-hold**

Syntax Description

Parameter	Description
`seconds`	The time, in seconds, that EIGRP will hold routes for an inactive peer. The configurable time range is from 20 to 300 seconds.

Summary

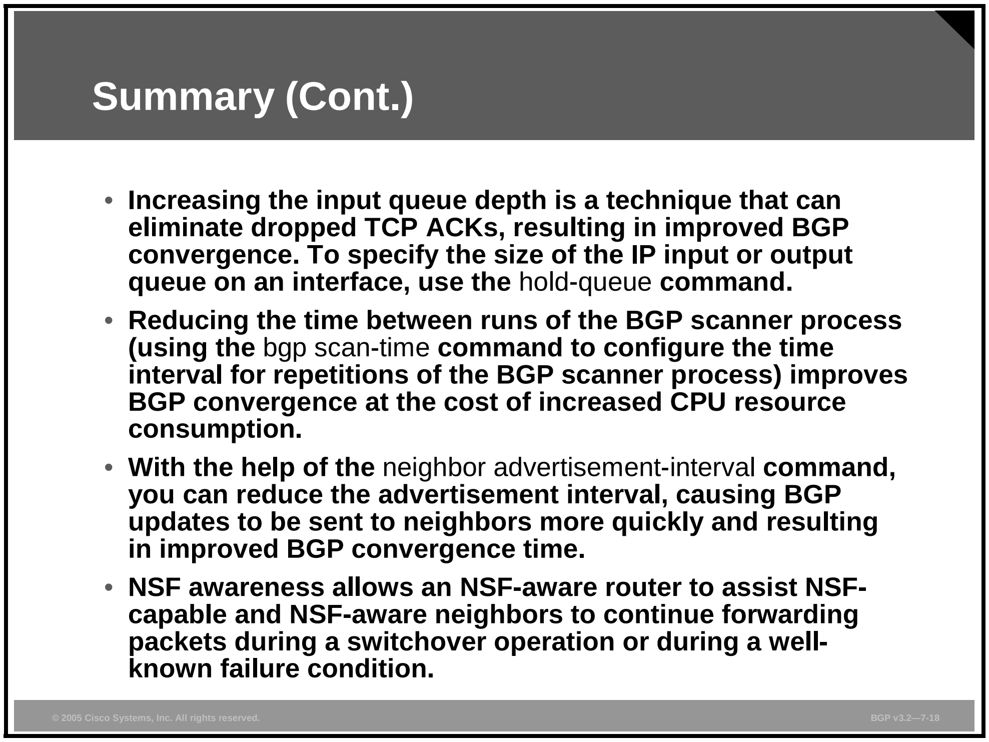
This topic summarizes the key points discussed in this lesson.



Summary

- Convergence is defined as the process of bringing all route tables to a consistent state.
- Several Cisco IOS processes (including BGP open, I/O, scanner, and router) that are executed on the router enable BGP to run.
- The BGP scanner and BGP router processes can significantly impact the CPU utilization of the router, causing some low-priority processes to suffer increased processing delays.
- You can reduce BGP convergence time and high CPU utilization caused by BGP processes in the following ways: queuing to TCP peer connections, deploying BGP peer groups, enabling the PMTU feature, and increasing interface input queues.
- The PMTU discovery feature, implemented through the ip tcp path-mtu-discovery *global configuration* command, finds the largest packet that can be sent to a destination without requiring IP fragmentation, minimizing packet overhead.

BGP v3.2—7-17



Summary (Cont.)

- Increasing the input queue depth is a technique that can eliminate dropped TCP ACKs, resulting in improved BGP convergence. To specify the size of the IP input or output queue on an interface, use the hold-queue command.
- Reducing the time between runs of the BGP scanner process (using the bgp scan-time command to configure the time interval for repetitions of the BGP scanner process) improves BGP convergence at the cost of increased CPU resource consumption.
- With the help of the neighbor advertisement-interval command, you can reduce the advertisement interval, causing BGP updates to be sent to neighbors more quickly and resulting in improved BGP convergence time.
- NSF awareness allows an NSF-aware router to assist NSF-capable and NSF-aware neighbors to continue forwarding packets during a switchover operation or during a well-known failure condition.

BGP v3.2—7-18

Lesson 2

Limiting the Number of Prefixes Received from a BGP Neighbor

Overview

There are currently more than 110,000 prefixes on the Internet. There are many circumstances in which network administrators do not need or desire their routers to carry full Internet routing. Furthermore, there is a need to provide protective controls on customer-facing routers to ensure that a configuration error does not cause the accidental advertisement of prefixes from autonomous systems that did not originate them.

Border Gateway Protocol (BGP) is designed for reliability and scalability. As such, it has a tremendous amount of flexibility regarding administrative policy controls, route selection, and performance tuning and scalability features. This lesson introduces an advanced BGP configuration tool that has been designed to improve BGP scalability and performance by reducing the number of prefixes that a router receives from a BGP neighbor. Also discussed in this lesson is the need for prefix limiting and how to configure and monitor the maximum-prefix function.

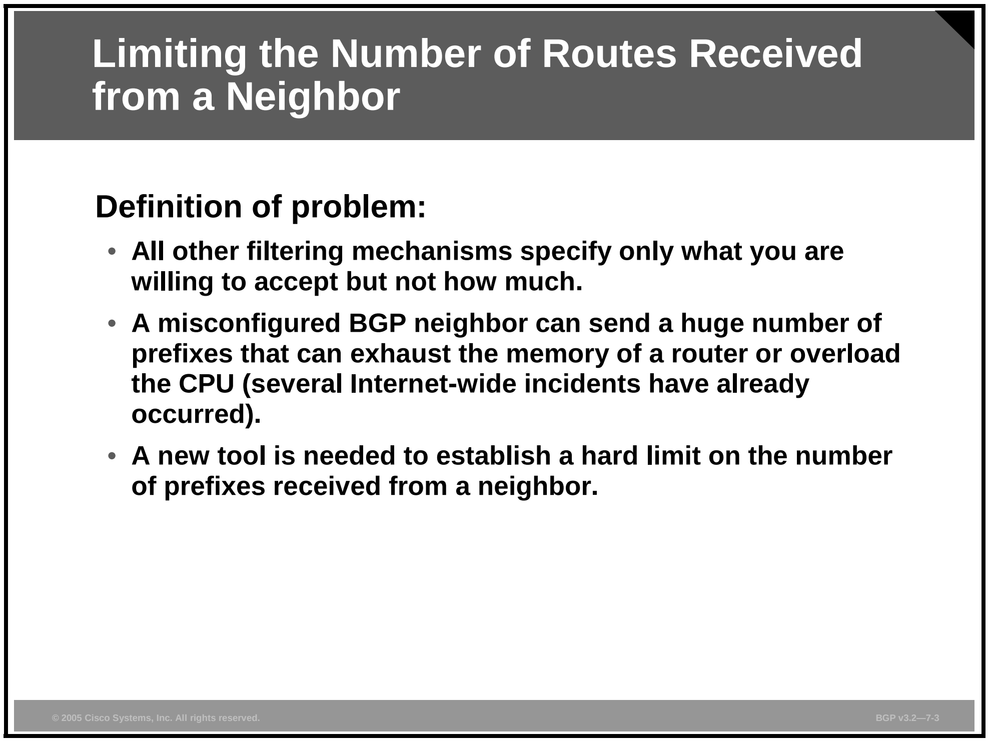
Objectives

Upon completing this lesson, you will be able to configure BGP to limit the number of prefixes that are received from a neighbor. This ability includes being able to meet these objectives:

- Describe the need for limiting the number of routes that are received from a BGP neighbor
- Identify the Cisco IOS command that is required to configure the BGP maximum-prefix function
- Identify the Cisco IOS command that is required to monitor the BGP maximum-prefix function

Limiting the Number of Routes Received from a Neighbor

This topic describes the need for limiting the number of routes (prefixes) that are received from a BGP neighbor.



Limiting the Number of Routes Received from a Neighbor

Definition of problem:

- **All other filtering mechanisms specify only what you are willing to accept but not how much.**
- **A misconfigured BGP neighbor can send a huge number of prefixes that can exhaust the memory of a router or overload the CPU (several Internet-wide incidents have already occurred).**
- **A new tool is needed to establish a hard limit on the number of prefixes received from a neighbor.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-3

Incoming filters and route-maps indicate the BGP attribute values that a route should have to be accepted. Route filters can be applied that match routes based on the network number or the BGP attributes that are attached to a route. The most commonly applied filter is one that matches the contents of the autonomous system (AS)-path attribute.

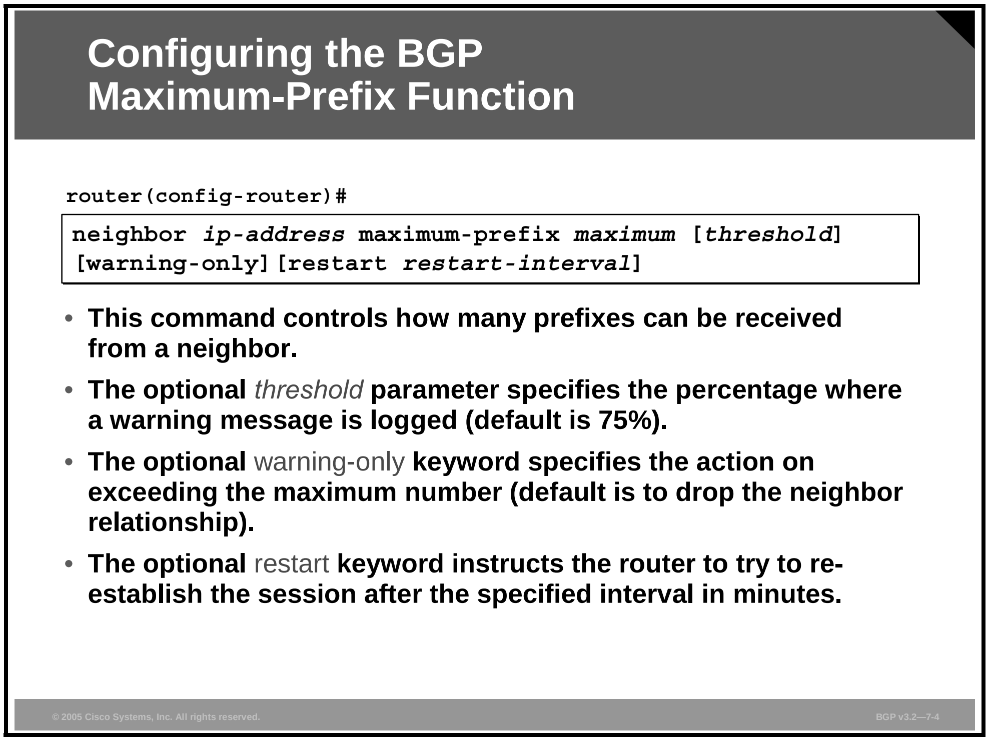
An Internet service provider (ISP) with a multihomed customer may use filters to ensure that the routes that are received from the customer originate within the AS of the customer. Using an AS-path access-list is one method of achieving this goal.

An improperly configured filter in a customer router may accidentally cause a large number of Internet routes to be received by the customer. Even worse, a faulty configuration may cause prefixes to be advertised by the customer as though the routes originated inside the customer AS. This situation would result in a BGP table in the ISP router that lists many of the possible destination networks on the Internet as reachable in the customer AS (with the AS path containing only a single entry, the customer AS). The BGP route selection in the ISP network would select those routes as the best (based on the AS-path length) and direct much of the provider traffic intended for the Internet to the customer network.

An AS-path filter in the ISP router would not prevent this accident. The routes that are sent by the customer have the anticipated AS-path value. A prefix-list that distinctly identifies and permits each of the network numbers that the customer may advertise would have prevented the accident. But such a prefix-list is hard to maintain.

Configuring the BGP Maximum-Prefix Function

This topic identifies the Cisco IOS command that is used to configure the maximum-prefix function in BGP.



Configuring the BGP Maximum-Prefix Function

```
router(config-router)#
neighbor ip-address maximum-prefix maximum [threshold]
[warning-only][restart restart-interval]
```

- **This command controls how many prefixes can be received from a neighbor.**
- **The optional** *threshold* **parameter specifies the percentage where a warning message is logged (default is 75%).**
- **The optional** warning-only **keyword specifies the action on exceeding the maximum number (default is to drop the neighbor relationship).**
- **The optional** restart **keyword instructs the router to try to re-establish the session after the specified interval in minutes.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-4

A scalable solution to the need for limiting the number of routes (prefixes) that are received from a BGP neighbor is to use a new tool that limits the number of routes that are received from a specific neighbor.

neighbor maximum-prefix

To control how many prefixes that a BGP router can receive from a neighbor, use the **neighbor maximum-prefix** router configuration command.

- **neighbor** {*ip-address* | *peer-group-name*} **maximum-prefix** *maximum* [*threshold*] [**warning-only**] [**restart** *restart-interval*]

To disable this function, use the **no** form of this command.

- **no neighbor** {*ip-address* | *peer-group-name*} **maximum-prefix** *maximum*

Syntax Description

Parameter	Description
ip-address	IP address of the neighbor.
peer-group-name	Name of a BGP peer group.
maximum	Maximum number of prefixes that are allowed from this neighbor.
threshold	(Optional) Integer specifying at what percentage of maximum the router starts to generate a warning message. The range is 1 to 100 percent; the default is 75 percent.
warning-only	(Optional) Allows the router to generate a log message when the maximum is exceeded, instead of terminating the peering.
restart	(Optional) Configures the router to automatically re-establish a peering session that has been disabled because the maximum-prefix limit has been exceeded. The configurable range of the restart interval is from 1 to 65,535 minutes.

This command allows you to configure a maximum number of prefixes that a BGP router is allowed to receive from a peer. It adds another mechanism (in addition to distribute-lists, filter-lists, and route-maps) to control prefixes that are received from a peer.

When the number of received prefixes exceeds the maximum number that is configured, the router terminates the peering (by default). However, if the **warning-only** keyword is configured, the router sends a log message but continues peering with the sender. If the peer is terminated, the peer session remains down until the **clear ip bgp** command is issued on the router, unless you have included the **restart** keyword in the configuration.

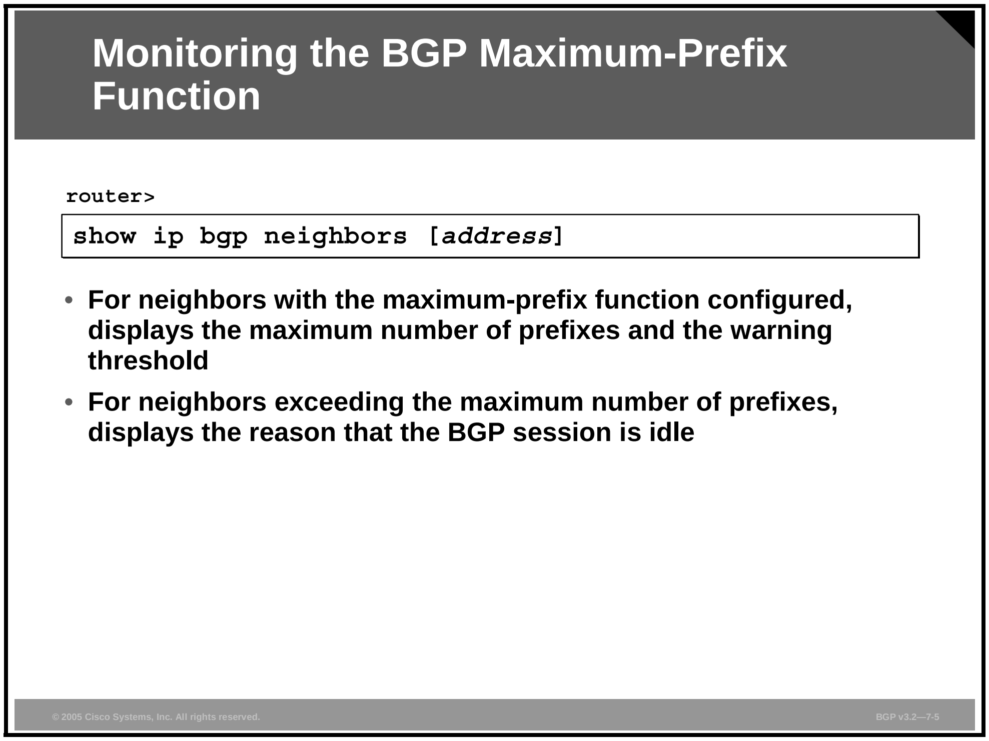
Note You can use the **bgp dampening** command to configure the dampening of a flapping route or interface when a peer is sending too many prefixes and causing network instability. You should need the **restart** command only when you are troubleshooting or tuning a router that is sending an excessive number of prefixes.

The BGP Restart Session After Max-Prefix Limit feature enhances the capabilities of the **neighbor maximum-prefix** command with the introduction of the **restart** keyword. This enhancement allows the network operator to configure the time interval after which a peering session is re-established by a router when the number of prefixes that have been received from a peer has exceeded the maximum prefix limit. The **restart** keyword has a configurable timer argument that is specified in minutes. The time range of the timer argument is from 1 to 65535.

This feature attempts to re-establish a disabled peering session at the time interval that is configured by the network operator. However, the configuration of the restart timer alone cannot change or correct a peer that is sending an excessive number of prefixes. The network operator will need to reconfigure the maximum-prefix limit or reduce the number of prefixes that are sent from the peer. A peer that is configured to send too many prefixes can cause instability in the network, where an excessive number of prefixes are rapidly advertised and withdrawn. In this case, the **warning-only** keyword can be configured to disable the restart capability while the network operator corrects the underlying problem.

Monitoring the BGP Maximum-Prefix Function

This topic identifies the Cisco IOS command that is used to monitor the operation of the maximum-prefix function in BGP.



Monitoring the BGP Maximum-Prefix Function

```
router>
show ip bgp neighbors [address]
```

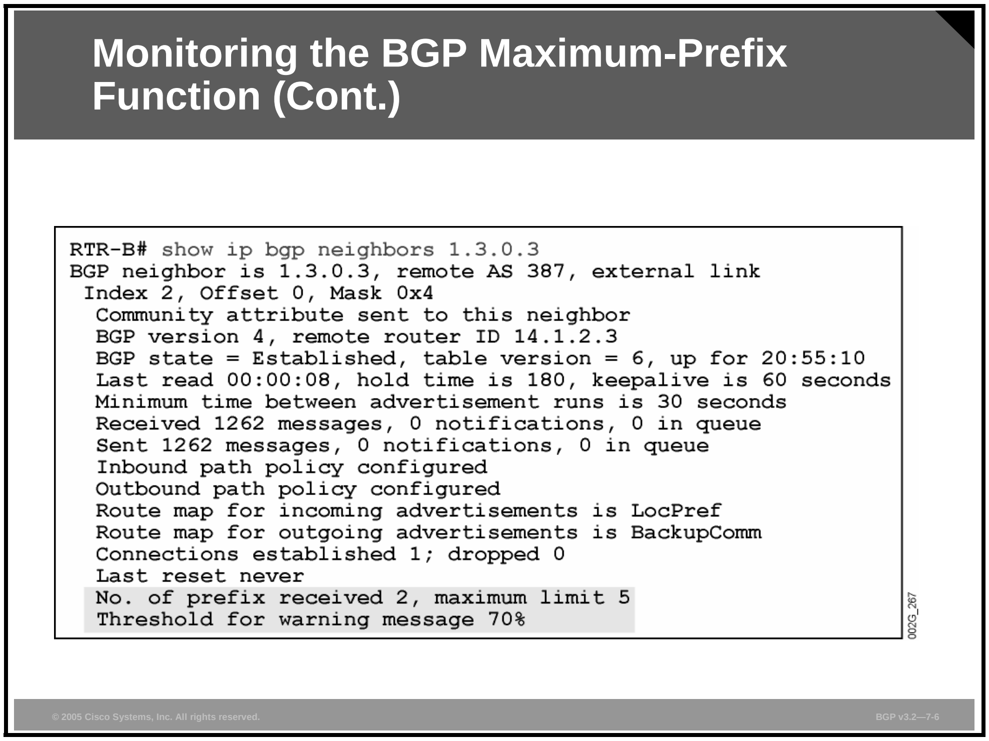
- For neighbors with the maximum-prefix function configured, displays the maximum number of prefixes and the warning threshold
- For neighbors exceeding the maximum number of prefixes, displays the reason that the BGP session is idle

BGP v3.2—7-5

Network administrators use the **show ip bgp neighbors** command to monitor the status of BGP neighbors. Among other things, the command displays information about how many prefixes a BGP router has received from a neighbor and if any limits have been configured.

Example: Monitoring the BGP

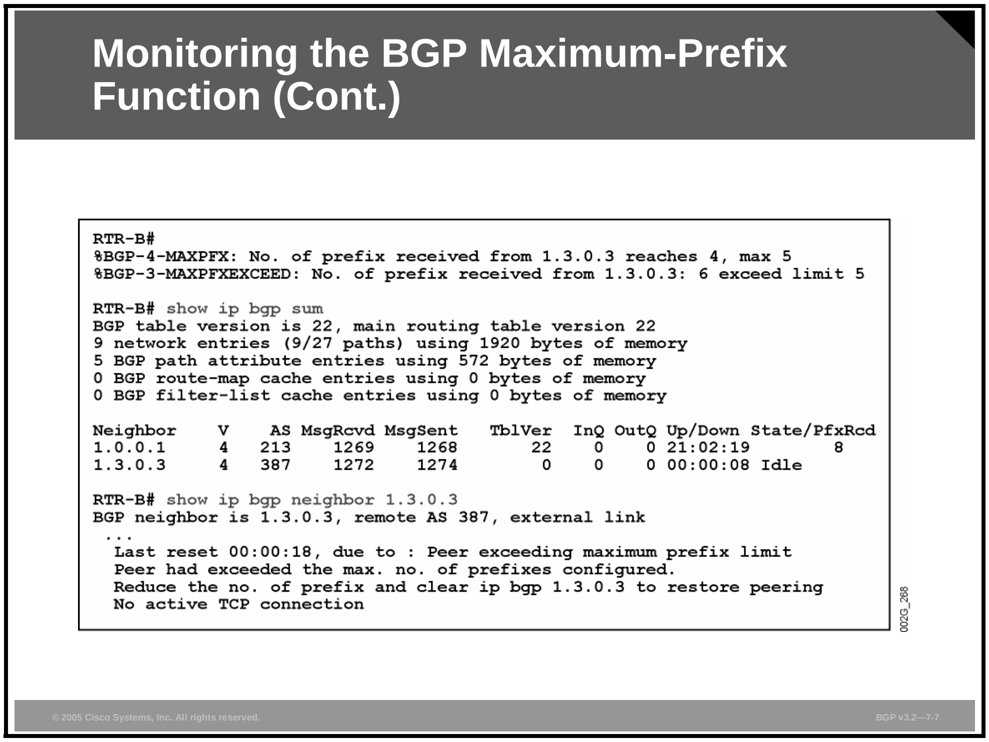
In this example, the neighbor with IP address 1.3.0.3 has been configured with a prefix limit of five prefixes.



Monitoring the BGP Maximum-Prefix Function (Cont.)

```
RTR-B# show ip bgp neighbors 1.3.0.3
BGP neighbor is 1.3.0.3, remote AS 387, external link
 Index 2, Offset 0, Mask 0x4
  Community attribute sent to this neighbor
  BGP version 4, remote router ID 14.1.2.3
  BGP state = Established, table version = 6, up for 20:55:10
  Last read 00:00:08, hold time is 180, keepalive is 60 seconds
  Minimum time between advertisement runs is 30 seconds
  Received 1262 messages, 0 notifications, 0 in queue
  Sent 1262 messages, 0 notifications, 0 in queue
  Inbound path policy configured
  Outbound path policy configured
  Route map for incoming advertisements is LocPref
  Route map for outgoing advertisements is BackupComm
  Connections established 1; dropped 0
  Last reset never
  No. of prefix received 2, maximum limit 5
  Threshold for warning message 70%
```

Currently, the BGP router has received two prefixes, which is under the limit. The threshold for the warning message is set to 70 percent, meaning that after receiving four prefixes from the 1.3.0.3 neighbor, the BGP router will generate a warning message.



Monitoring the BGP Maximum-Prefix Function (Cont.)

```
RTR-B#
%BGP-4-MAXPFX: No. of prefix received from 1.3.0.3 reaches 4, max 5
%BGP-3-MAXPFXEXCEED: No. of prefix received from 1.3.0.3: 6 exceed limit 5

RTR-B# show ip bgp sum
BGP table version is 22, main routing table version 22
9 network entries (9/27 paths) using 1920 bytes of memory
5 BGP path attribute entries using 572 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory

Neighbor    V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
1.0.0.1     4   213    1269    1268       22    0    0 21:02:19        8
1.3.0.3     4   387    1272    1274        0    0    0 00:00:08 Idle

RTR-B# show ip bgp neighbor 1.3.0.3
BGP neighbor is 1.3.0.3, remote AS 387, external link
 ...
  Last reset 00:00:18, due to : Peer exceeding maximum prefix limit
  Peer had exceeded the max. no. of prefixes configured.
  Reduce the no. of prefix and clear ip bgp 1.3.0.3 to restore peering
  No active TCP connection
```

002G_268

BGP v3.2—7-7

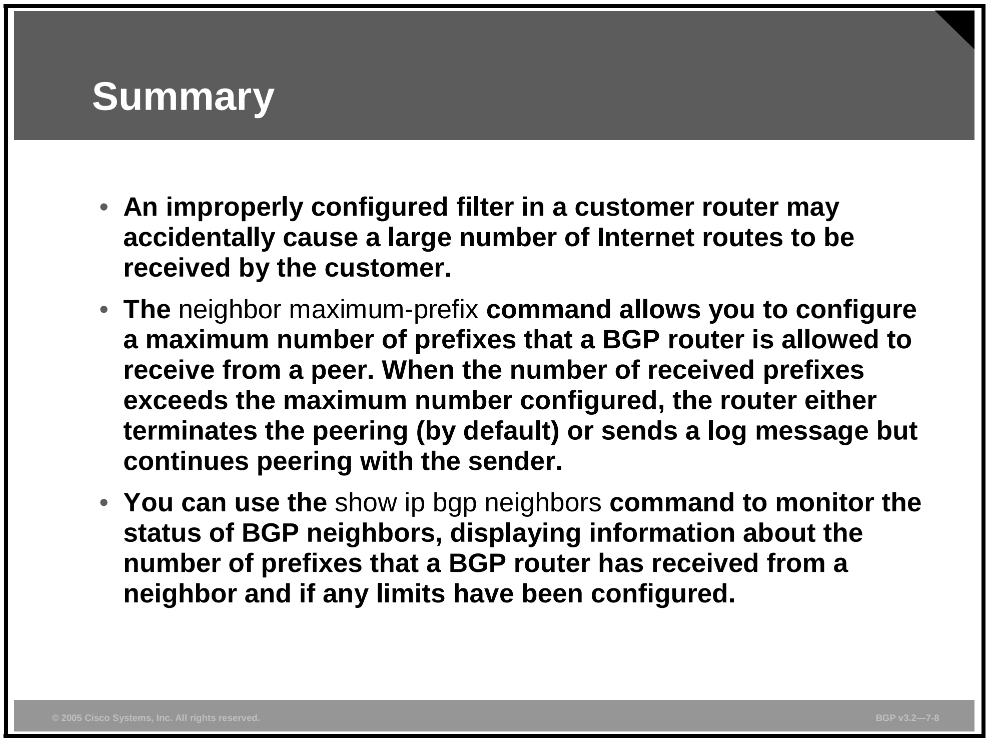
The logging outputs, in the example here, show that a BGP neighbor is close to exceeding the configured maximum-prefix limit. The total number of received prefixes has reached four, which is over the threshold to generate a warning message. The warning is displayed on the console and optionally sent to a syslog server.

The logging output then shows that two more prefixes have been received. The total number is now six, which is above the configured limit. The BGP session is therefore terminated. The output of the **show ip bgp neighbors** command shows that the reason for resetting the session is that the peer exceeded the configured maximum number of prefixes. As a result of the session being reset, the BGP session will remain in the Idle state.

To force the neighbor from the Idle state into the Active state and to re-establish the BGP session, the network administrator must issue the **clear ip bgp ip-address** command for the neighbor (except if the network administrator has specified the **restart** keyword in the configuration, in which case the router tries to re-establish the BGP session automatically after the expiration of the configured restart timeout interval).

Summary

This topic summarizes the key points discussed in this lesson.



Summary

- **An improperly configured filter in a customer router may accidentally cause a large number of Internet routes to be received by the customer.**
- **The** neighbor maximum-prefix **command allows you to configure a maximum number of prefixes that a BGP router is allowed to receive from a peer. When the number of received prefixes exceeds the maximum number configured, the router either terminates the peering (by default) or sends a log message but continues peering with the sender.**
- **You can use the** show ip bgp neighbors **command to monitor the status of BGP neighbors, displaying information about the number of prefixes that a BGP router has received from a neighbor and if any limits have been configured.**

Lesson 3

Implementing BGP Peer Groups

Overview

Scaling routers to meet the demands of full Internet routing and associated administrative policies requires protocols like Border Gateway Protocol (BGP) with embedded scalability mechanisms. In environments where network administrators must configure a large number of BGP peers, peer groups are a scalability tool that reduces both administrative overhead and router resource requirements.

Typical service provider networks usually contain BGP-speaking routers that consist of many neighbors that are configured with the same administrative policies (such as outbound route-maps, distribute-lists, filter-lists, update source, and so on). Network administrators can group together neighbors with the same update policies into peer groups to simplify configuration and, more importantly, to make BGP updates more efficient. This lesson introduces peer groups as a BGP scalability mechanism. The lesson also discusses the commands that are required to properly configure and monitor BGP peer groups.

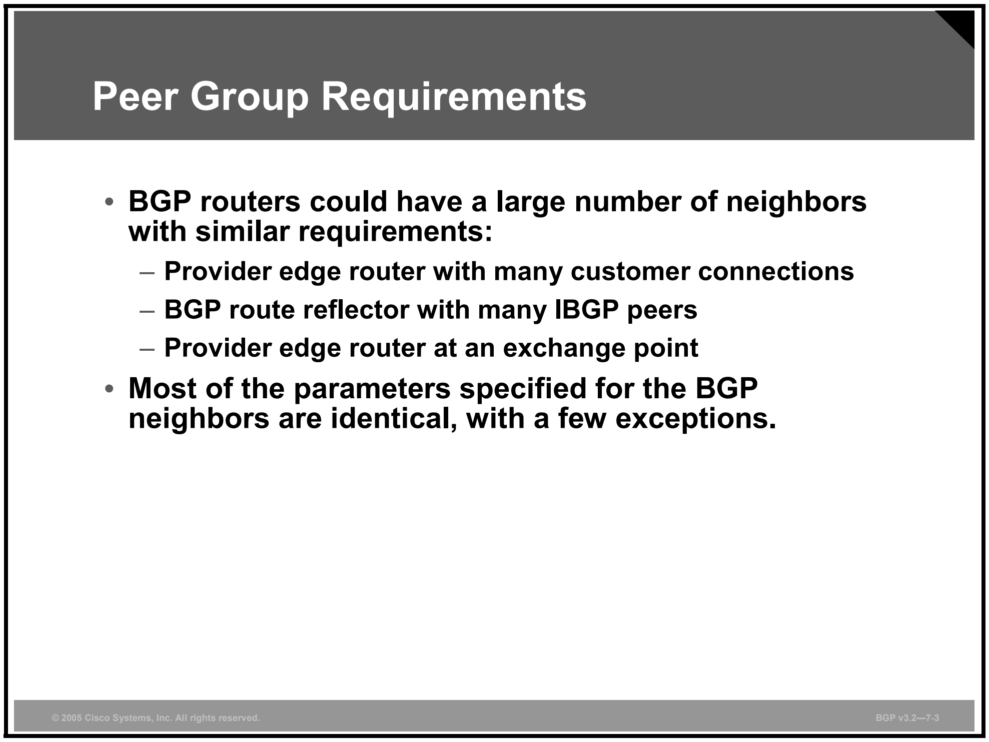
Objectives

Upon completing this lesson, you will be able to use BGP peer groups to share common configuration parameters between multiple BGP peers. This ability includes being able to meet these objectives:

- Describe the need for BGP peer groups
- Describe the performance benefits of using BGP peer groups
- Describe the limitations of BGP peer groups
- Describe the characteristics of BGP peer groups when they are implemented in Cisco IOS software
- Describe the function of the BGP Dynamic Update Peer-Groups feature
- Describe the function of the BGP Configuration Using Peer Templates feature
- Identify the Cisco IOS commands that are required to configure BGP peer groups
- Identify the Cisco IOS commands that are required to monitor BGP peer groups

Peer Group Requirements

This topic describes the need for BGP peer groups.



Peer Group Requirements

- **BGP routers could have a large number of neighbors with similar requirements:**
  - **Provider edge router with many customer connections**
  - **BGP route reflector with many IBGP peers**
  - **Provider edge router at an exchange point**
- **Most of the parameters specified for the BGP neighbors are identical, with a few exceptions.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-3

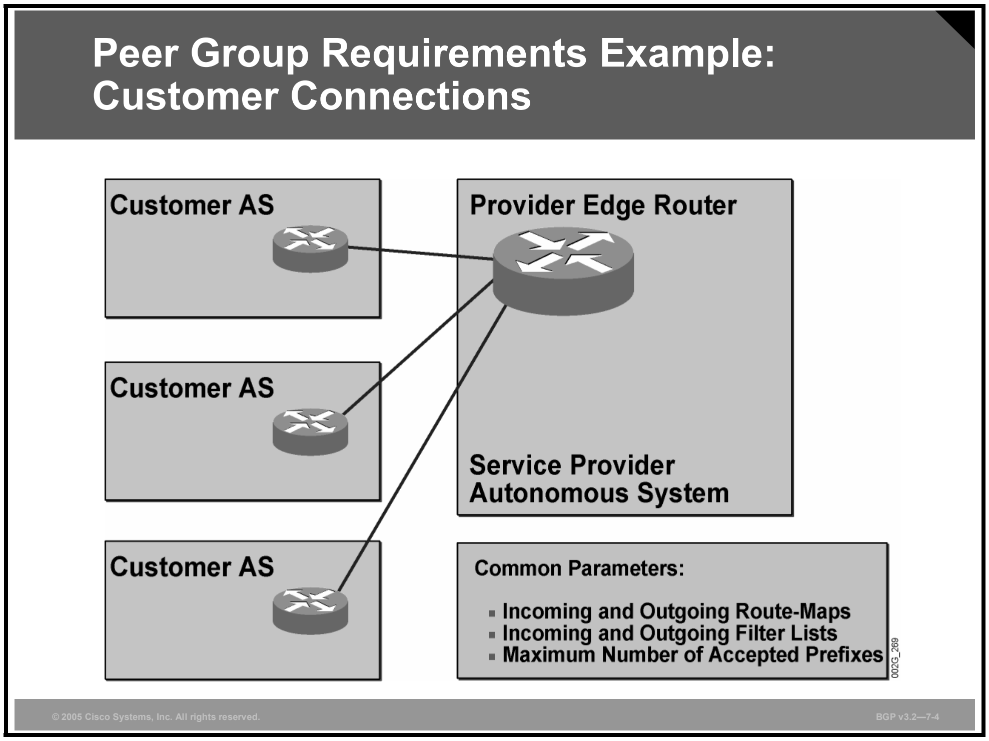
In many cases, a network administrator must configure a single router with a large number of neighbors, each neighbor having parameters similar to the others. This situation may cause time-consuming configuration work, because the network administrator has to configure almost the same filter-list, route-map, and prefix-list for all of the neighbors. A service provider network with an edge router having a large number of customers attached to it, where each customer requires its own BGP session, may find that all of the BGP sessions to its customer routers have almost identical configurations.

Likewise, Internal Border Gateway Protocol (IBGP) sessions are almost always identically configured. If a full mesh is deployed within an autonomous system (AS), a large number of peer configurations might exist. Recall that an AS containing only 15 routers will require ([15 * 14] / 2) = 105 neighbor sessions to meet the full-mesh requirement of BGP. Configuring 105 neighbors with duplicate parameters leads to a tremendous amount of redundant configuration.

To ease the burden of configuring a large number of neighbors with identical or similar parameters (for example, route-maps, filter-lists, or prefix-lists), the concept of peer groups was introduced. The network administrator can configure a template, or peer group. The administrator configures the peer group with all the BGP parameters that are to be applied to many BGP peers. Actual BGP neighbors are bound to the peer group, and the network administrator applies the peer group configuration on each of the BGP sessions. BGP neighbors of a single router can be divided into several groups, each group having its own BGP parameters. Actual neighbors are then bound to the appropriate group, resulting in an optimum BGP configuration.

Example: Peer Group Requirements–Customer Connections

This example illustrates a service provider network with a group of customer autonomous systems that can be treated in the same (or a very similar) way.



Peer Group Requirements Example: Customer Connections

Customer AS

Customer AS

Customer AS

Provider Edge Router

Service Provider Autonomous System

Common Parameters:

- Incoming and Outgoing Route-Maps
- Incoming and Outgoing Filter Lists
- Maximum Number of Accepted Prefixes

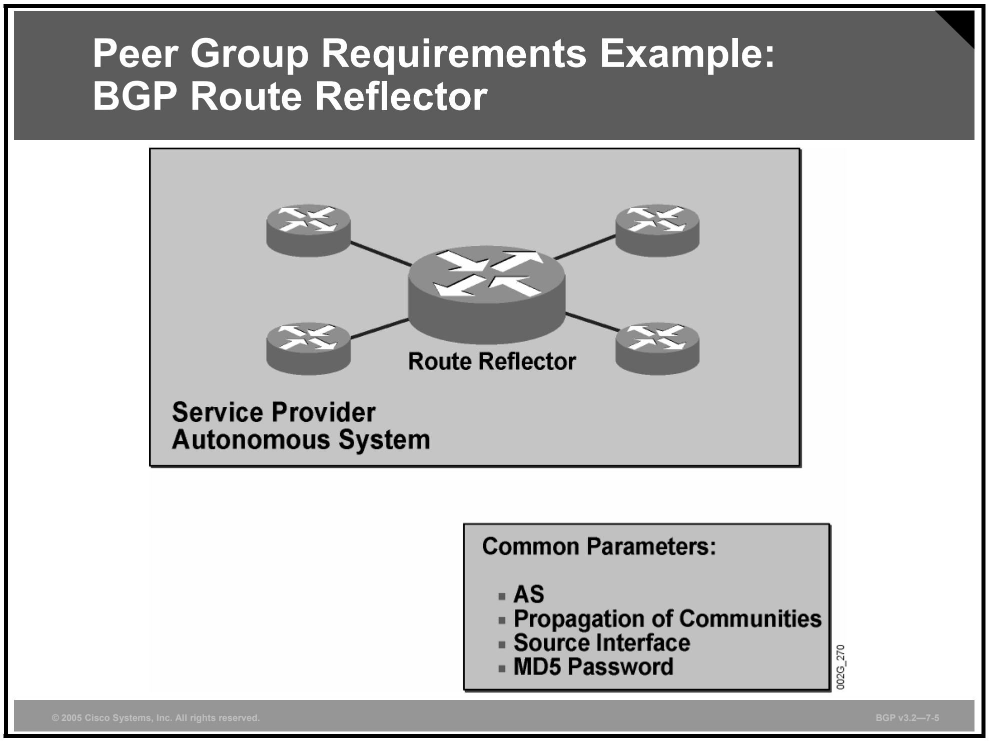
The customer autonomous systems are all assumed to announce local networks only. All customer autonomous systems should receive BGP updates with the same set of Internet routes, and the customer autonomous systems are all assumed to generate only a small number of prefixes.

This situation makes the neighbor configuration almost identical for each of the customers, with only a few changes that are specific to each neighbor.

In this scenario, the use of the peer group function is highly desirable. The network administrator can configure BGP neighbors in the customer autonomous systems using a single peer group. The administrator configures the peer group template with references to route-maps, filter-lists, prefix-lists, and the maximum number of received prefixes. Then the IP addresses of the customer routers are bound to the peer group, and the peer group configuration is applied to all of them.

Example: Peer Group Requirements–BGP Route Reflector

This example illustrates the BGP route reflector.



Peer Group Requirements Example: BGP Route Reflector

Common Parameters:

- AS
- Propagation of Communities
- Source Interface
- MD5 Password

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—7-5

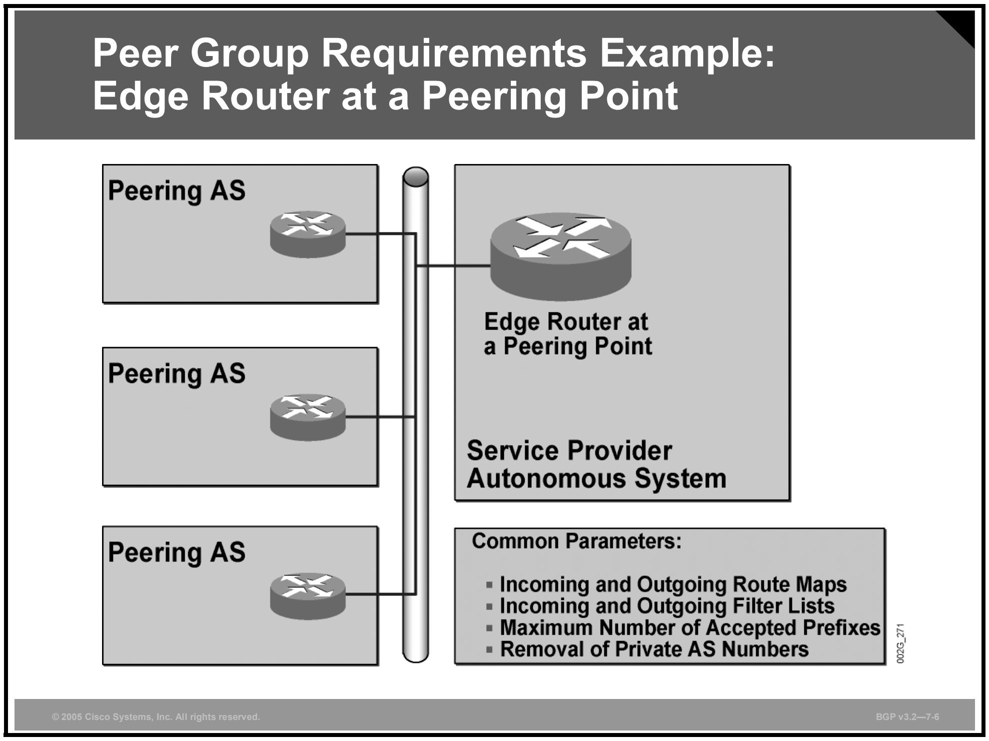
To apply a consistent routing policy within the entire local AS, the network administrator needs to configure every IBGP session identically. If a router in the AS is supplied with some information, then all the routers should be supplied with the same information. Otherwise, an inconsistent routing policy within the AS might cause inconsistent routing or application of BGP policies.

The peer group function is a good tool to make sure that all IBGP peers receive the same configuration information. The network administrator configures a peer group template with the required parameters, such as the neighbor AS number, enabling of the **send-community** option, setting of the update source to a loopback interface, and router authentication mechanisms. Then all the internal neighbor IP addresses are bound to the peer group, and the peer group configuration is applied to all of them. This approach ensures a consistent routing policy within the AS.

In a service provider network, the routers that are assigned as route reflectors are the routers with the largest number of IBGP sessions. These are the routers where the peer group function is most useful.

Example: Peer Group Requirements—Edge Router at a Peering Point

This example illustrates an edge router at a peering point.



Peer Group Requirements Example:
Edge Router at a Peering Point

Peering AS
Peering AS
Peering AS
Edge Router at a Peering Point
Service Provider Autonomous System
002G_271

Common Parameters:

- Incoming and Outgoing Route Maps
- Incoming and Outgoing Filter Lists
- Maximum Number of Accepted Prefixes
- Removal of Private AS Numbers

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-6

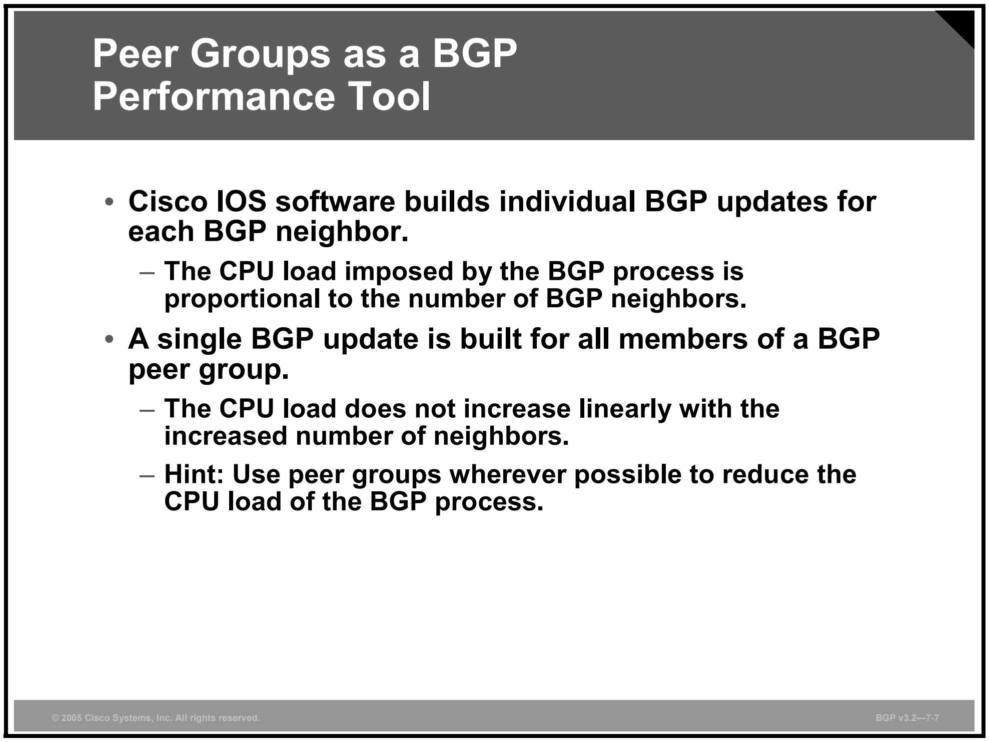
The edge router that is located in the network where the service provider exchanges routes with other service providers is also a suitable place to use peer groups. From the edge router, the service provider AS can peer with a large number of other service providers.

All peering autonomous systems should receive the same set of routes, namely the routes local to the service provider AS and the routes that are received from customer autonomous systems. Also, all routes that are received by the service provider peering router from all peering autonomous systems are processed almost identically. The characteristic of the exchange network is the same regardless of which neighbor the routes are received from. If the peering point is an FDDI, ATM, Gigabit Ethernet, or Dynamic Packet Transport (DPT) network, the preference of using the network for packet exchange may be different. However, for each single peering point, all neighbors are reachable over the same network, and the preference is quite likely to be the same.

Additionally, a number of other parameters could be the same, such as removing private AS numbers and limiting the number of routes received. In these cases, the network administrator can apply these parameters on the peer group template before the actual IP addresses of the neighbors are bound to the peer group.

Peer Groups as a BGP Performance Tool

This topic describes the performance benefits of using BGP peer groups.



Peer Groups as a BGP Performance Tool

- **Cisco IOS software builds individual BGP updates for each BGP neighbor.**
  - **The CPU load imposed by the BGP process is proportional to the number of BGP neighbors.**
- **A single BGP update is built for all members of a BGP peer group.**
  - **The CPU load does not increase linearly with the increased number of neighbors.**
  - **Hint: Use peer groups wherever possible to reduce the CPU load of the BGP process.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-7

By default, Cisco IOS software builds BGP updates for each neighbor individually. Building BGP updates involves a number of router-CPU-consuming tasks, including scanning the BGP table and applying a variety of outgoing filtering mechanisms (filter-lists, route-maps, and prefix-lists). These tasks imply that when a router is configured with a large number of neighbors, the CPU load grows proportionally.

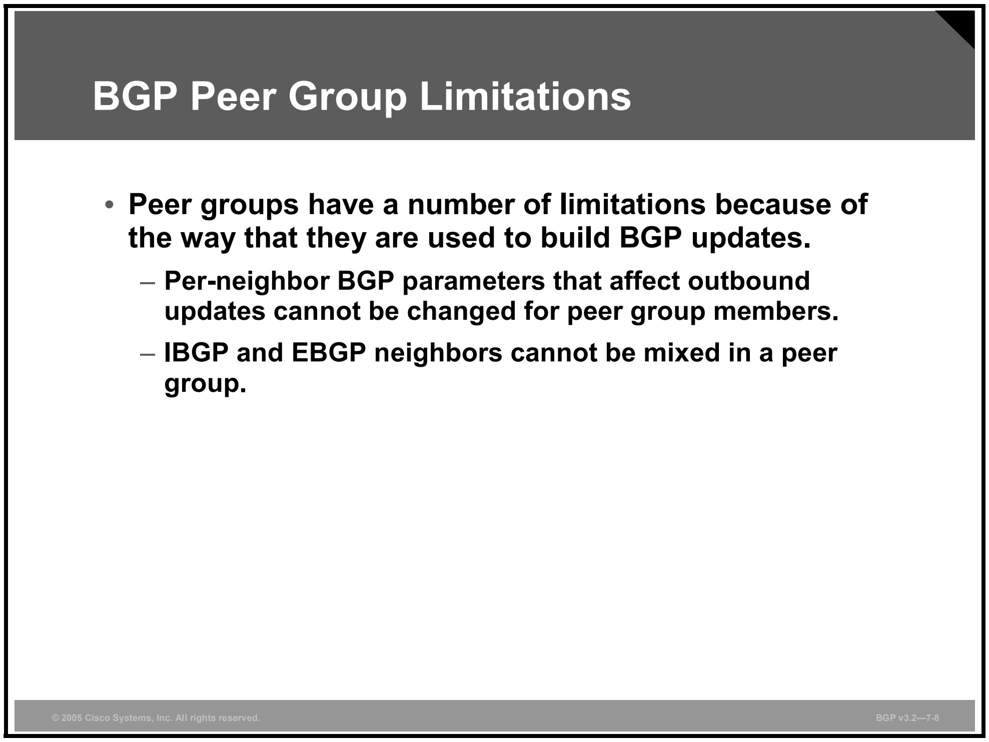
However, with the use of peer groups, some of the router CPU utilization that is imposed by BGP update generation is significantly reduced, because the use of peer groups allows the router to run the BGP update (including all outgoing filter processing) only once for the entire peer group. The router, after it has finished building the BGP update, sends it to each member of the peer group. The actual TCP transmission still has to be done on a per-neighbor basis because of the connection-oriented characteristics of BGP sessions.

So, router CPU load does increase when there are more neighbors of a router, because of increased TCP workload, but the use of peer groups can significantly reduce the increase. Therefore, network administrators should use peer groups whenever possible to reduce the CPU load.

Note	BGP peer groups are the fundamental BGP scalability tool and should be used in all environments where a router has a large number of BGP neighbors.

BGP Peer Group Limitations

This topic describes the limitations of BGP peer groups.



BGP Peer Group Limitations

- **Peer groups have a number of limitations because of the way that they are used to build BGP updates.**
  - **Per-neighbor BGP parameters that affect outbound updates cannot be changed for peer group members.**
  - **IBGP and EBGP neighbors cannot be mixed in a peer group.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-8

Because the router builds only one update for all members of the same peer group, some restrictions apply to members of the peer group:

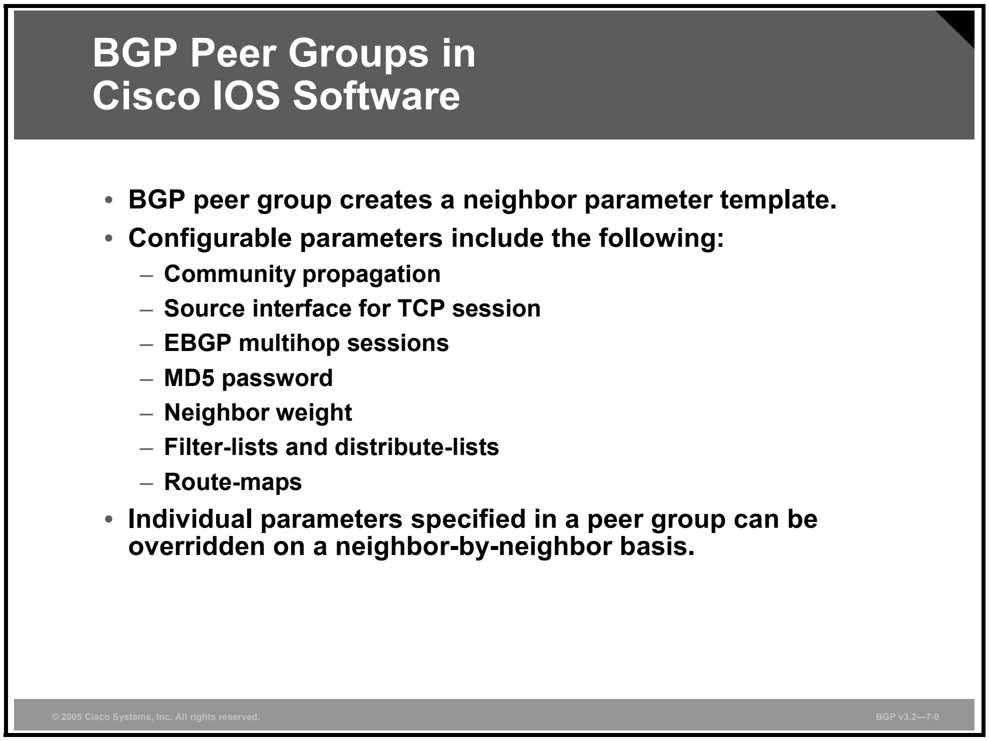
- There cannot be different outbound filters, route-maps, or other means that could possibly cause different updates to be sent to two members of the same peer group. Cisco IOS software creates only one update, which is subsequently replicated to all members. Any violation of this rule could cause unexpected results.
- External Border Gateway Protocol (EBGP) and IBGP updates are very different. EBGP updates have the AS-path attribute changed. IBGP sessions pass only the local preference attribute. The multi-exit discriminator (MED) attribute that is received from a remote AS is passed on to IBGP sessions but is removed before it is sent on an EBGP session. Therefore, you cannot assign IBGP neighbors and EBGP neighbors to the same peer group because they cannot receive a replication of the same update.

Prior to Cisco IOS Software Release 11.1(18)CC, a route reflector client could not be a member of a peer group. When the peer group leader was a route reflector client and an update was received from it, the route reflector split-horizon rules prevented the update from being sent back to the sender. Therefore, no update was generated for any members of the peer group. When using peer groups in combination with route reflectors, make sure that all routers in the AS are running Cisco IOS releases later than 11.1(18)CC, where this restriction is lifted.

Because the router sends the same update to each of the peer group members, the next-hop BGP attribute is replicated. The receivers of the information must be able to use the same next-hop IP address, requiring the receivers to be in the same IP subnet. If two receivers are on different subnets, only one of them will receive a valid next-hop attribute. The other routers will receive a next-hop IP address that is inaccessible. This restriction was also removed in Cisco IOS Release 11.1(18)CC, making BGP peer groups an ideal scalability tool.

BGP Peer Groups in Cisco IOS Software

This topic describes the characteristics of BGP peer groups when they are implemented in Cisco IOS software.



BGP Peer Groups in Cisco IOS Software

- **BGP peer group creates a neighbor parameter template.**
- **Configurable parameters include the following:**
  - **Community propagation**
  - **Source interface for TCP session**
  - **EBGP multihop sessions**
  - **MD5 password**
  - **Neighbor weight**
  - **Filter-lists and distribute-lists**
  - **Route-maps**
- **Individual parameters specified in a peer group can be overridden on a neighbor-by-neighbor basis.**

BGP v3.2—7-9

On a Cisco IOS router, the peer group is created as a template. The template is configured to do the following:

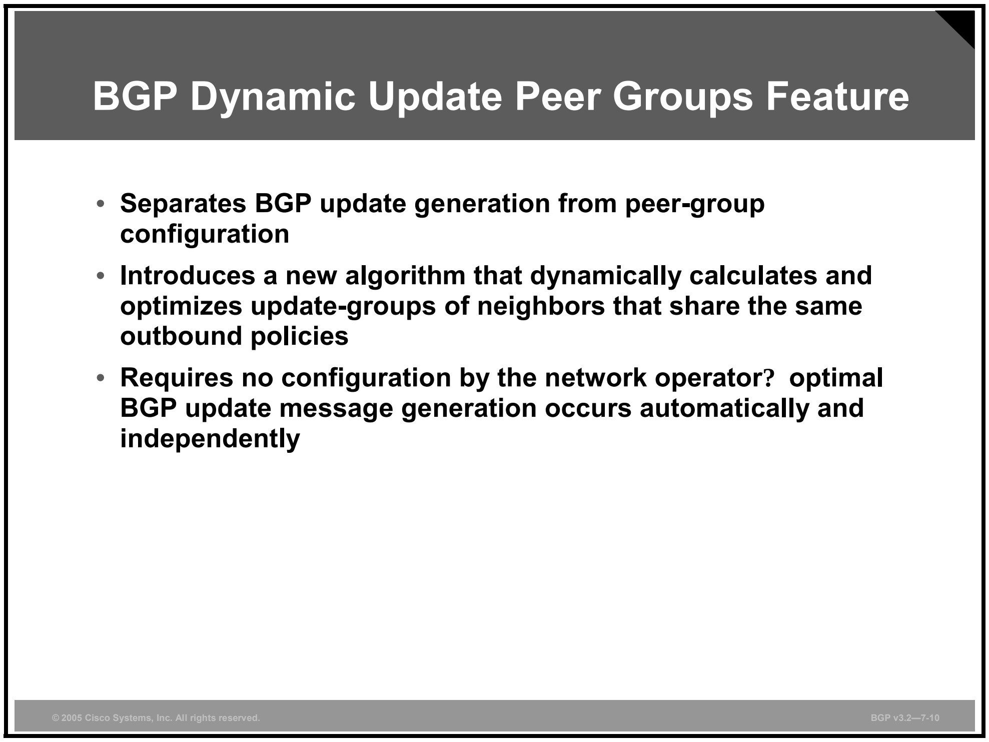
- Propagate, or not propagate, the community attribute
- Use the IP address of a specific interface as the source address when opening the TCP session
- Use, or not use, the EBGP multihop function
- Use, or not use, Message Digest 5 (MD5) authentication on the BGP sessions
- Assign a particular weight value to the routes that are received
- Filter out any incoming or outgoing routes based on the content of the AS-path attribute that are associated with the route or the network number of the route
- Pass the incoming or outgoing routes through a particular route-map

When actual neighboring routers are assigned to the peer group on a router, all of the attributes that are configured for the peer group are applied to all peer group members. Cisco IOS software optimizes the outgoing routes by running through the outgoing filters and route-maps only once and then replicating the results to each of the peer group members. In reality, Cisco IOS software assigns a peer group leader, for which the software generates an update, and this update is replicated by the leader to all other members of the peer group.

Some parameters configured on the peer group can be overridden by neighbor configurations, but only if the individual configurations apply on incoming updates. Outgoing updates are always prepared for the peer group leader and then replicated to the other members of the peer group.

BGP Dynamic Update Peer-Groups Feature

This topic describes the function of the BGP Dynamic Update Peer-Groups feature.



BGP Dynamic Update Peer Groups Feature

- **Separates BGP update generation from peer-group configuration**
- **Introduces a new algorithm that dynamically calculates and optimizes update-groups of neighbors that share the same outbound policies**
- **Requires no configuration by the network operator? optimal BGP update message generation occurs automatically and independently**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-10

In previous versions of Cisco IOS software, BGP update messages were grouped together based on peer-group configurations. This method of grouping neighbors together for BGP update message generation reduced the amount of system processing resources needed to process the routing table. This method, however, had the following limitations:

- All neighbors that shared the same peer-group configuration also had to share the same outbound routing policies.
- All neighbors had to belong to the same peer group and address family. Neighbors configured in different peer groups cannot belong to different address families.

These limitations existed to balance optimal update generation and replication against peer-group configuration. These limitations also caused the network operator to configure smaller peer groups, which reduced the efficiency of update message generation.

The BGP Dynamic Update Peer-Groups feature separates BGP update generation from peer-group configuration, The BGP Dynamic Update Peer-Groups feature introduces a new algorithm that dynamically calculates and optimizes update-groups of neighbors that share the same outbound policies and can share the same update messages.

This feature does not require any configuration by the network operator. Optimal BGP update-message generation occurs automatically and independently. BGP neighbor configuration is no longer restricted by outbound routing policies, and update-groups can belong to different address families. When a change to outbound policy occurs, the router automatically recalculates update-group memberships and applies the changes by triggering an outbound soft reset after a 3-minute timer expires. This behavior is designed to provide the network operator with time to change the configuration if a mistake is made.

If no argument is specified, the **clear up bgp update-group** command recalculates all update-groups. Specific index numbers for update groups and information about update-group membership is displayed in the output of the **show ip bgp update-group** and **debug ip bgp groups** commands.

clear ip bgp update-group

To clear BGP update-group member sessions, use the **clear ip bgp update-group** command in privileged EXEC mode.

- **clear ip bgp update-group** [*index-group* | *ip-address*]

Syntax Description

Parameter	Description
index-group	(Optional) Specifies that the update-group with corresponding index number will be reset. The range of update-group index numbers is from 1 to 4294967295.
ip-address	(Optional) Specifies the IP address of a single peer that will be reset.

The output of the **debug ip bgp groups** command displays information about update-group calculations and the addition and removal of update-group members. Information about peer groups and peer-policy and peer-session templates is also displayed in the output of this command as neighbor configurations change.

Note The output of this command can be very verbose, so this command should not be deployed in a production network unless you are troubleshooting a problem.

debug ip bgp groups

To display information related to the processing of BGP update-groups, use the **debug ip bgp update** privileged EXEC mode.

- **debug ip bgp groups** [*index-group* | *ip-address*]

To disable the display of BGP update information, use the **no** form of this command.

- **no debug ip bgp groups**

Syntax Description

Parameter	Description
index-group	(Optional) Specifies that update-group debugging information for the corresponding index number will be displayed. The range of update-group index numbers is from 1 to 4294967295.
ip-address	(Optional) Specifies that update-group debugging information for a single peer will be displayed.

The following example output from the **debug ip bgp groups** command shows the recalculation of update-groups after the **clear ip bgp groups** command was issued:

```
Router# debug ip bgp groups

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.5 Down User reset

5w4d: BGP-DYN(0): Comparing neighbor 10.4.9.5 flags 0x0 cap 0x0 and updgrp 2 fl0

5w4d: BGP-DYN(0): Update-group 2 flags 0x0 cap 0x0 policies same as 10.4.9.5 fl0

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.8 Down User reset

5w4d: BGP-DYN(0): Comparing neighbor 10.4.9.8 flags 0x0 cap 0x0 and updgrp 2 fl0

5w4d: BGP-DYN(0): Update-group 2 flags 0x0 cap 0x0 policies same as 10.4.9.8 fl0

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.21 Down User reset

5w4d: BGP-DYN(0): Comparing neighbor 10.4.9.21 flags 0x0 cap 0x0 and updgrp 1 f0

5w4d: BGP-DYN(0): Update-group 1 flags 0x0 cap 0x0 policies same as 10.4.9.21 f0

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.5 Up

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.21 Up

5w4d: %BGP-5-ADJCHANGE: neighbor 10.4.9.8 Up
```

show ip bgp replication

To display update replication statistics for BGP update-groups, use the **show ip bgp replication** command in EXEC mode.

- **show ip bgp replication** [*index-group* | *ip-address*]

Syntax Description

Parameter	Description
index-group	(Optional) Specifies that update replication statistics for the update-group with corresponding index number will be displayed. The range of update-group index numbers is from 1 to 4294967295.
ip-address	(Optional) Specifies the IP address of a single neighbor for which update-group statistics will be displayed.

The following sample output from the **show ip bgp replication** command shows update-group replication information for all for neighbors:

```
Router# show ip bgp replication

BGP Total Messages Formatted/Enqueued : 0/0


     Index     Type  Members          Leader   MsgFmt  MsgRepl  Csize  Qsize

         1 internal        1       10.4.9.21        0        0      0      0
```

2 internal 2 10.4.9.5 0 0 0 0

show ip bgp update-group

To display information about BGP update-groups, use the **show ip bgp update-group** command in EXEC mode.

- **show ip bgp update-group** [*index-group* | *ip-address*] [**summary**]

Syntax Description

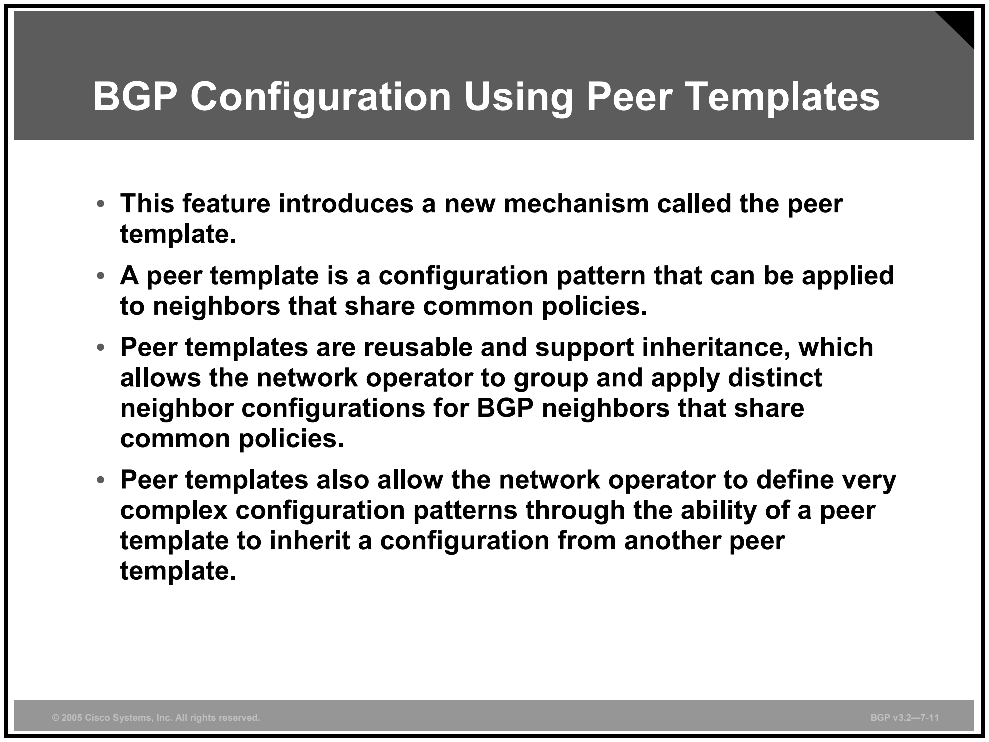
Parameter	Description
index-group	(Optional) Displays the update-group with corresponding index number. The range of update-group index numbers is from 1 to 4294967295.
ip-address	(Optional) Displays the IP address of a single neighbor.
summary	(Optional) Displays a summary of update-group member information. The output can be filtered to show information for a single index-group or peer with the *index-group* or *ip-address* argument.

The following sample output from the **show ip bgp update-group** command shows update-group information for all neighbors:

```
Router# show ip bgp update-group

BGP version 4 update-group 1, internal, Address Family: IPv4 Unicast

  BGP Update version : 0, messages 0/0

  Route map for outgoing advertisements is COST1

  Update messages formatted 0, replicated 0

  Number of NLRIs in the update sent: max 0, min 0

  Minimum time between advertisement runs is 5 seconds

  Has 1 member:

  10.4.9.21


BGP version 4 update-group 2, internal, Address Family: IPv4 Unicast

  BGP Update version : 0, messages 0/0

  Update messages formatted 0, replicated 0

  Number of NLRIs in the update sent: max 0, min 0
  Minimum time between advertisement runs is 5 seconds
  Has 2 members:
10.4.9.5 10.4.9.8
```

BGP Configuration Using Peer Templates

This topic describes the function of the BGP Configuration Using Peer Templates feature.



BGP Configuration Using Peer Templates

- This feature introduces a new mechanism called the peer template.
- A peer template is a configuration pattern that can be applied to neighbors that share common policies.
- Peer templates are reusable and support inheritance, which allows the network operator to group and apply distinct neighbor configurations for BGP neighbors that share common policies.
- Peer templates also allow the network operator to define very complex configuration patterns through the ability of a peer template to inherit a configuration from another peer template.

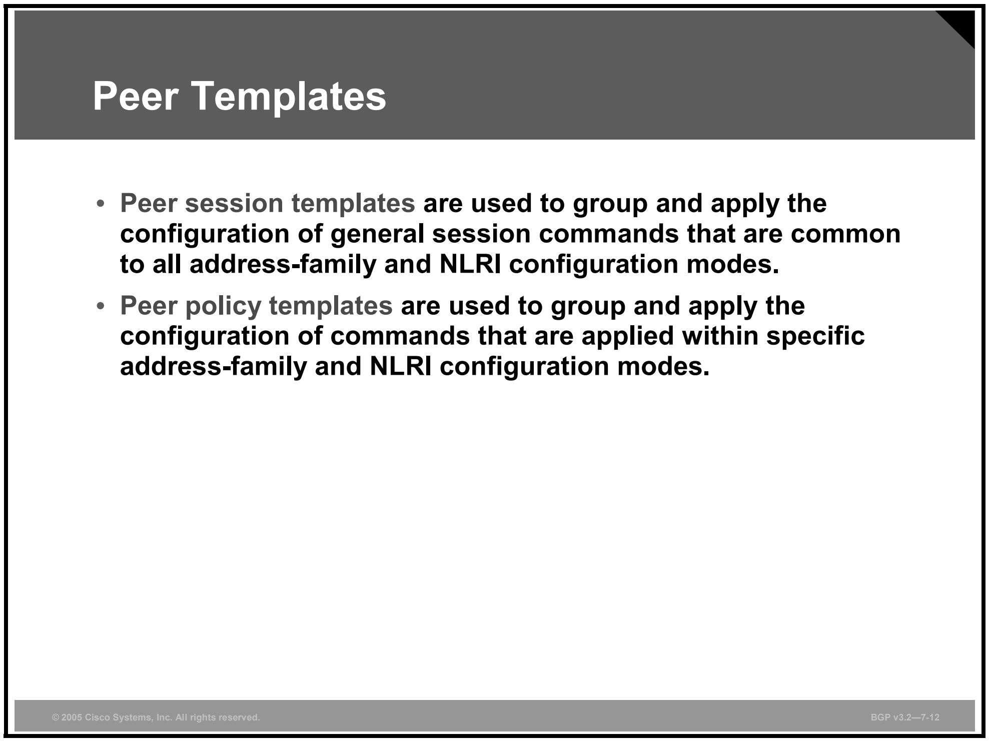
BGP v3.2—7-11

The BGP Dynamic Update Peer-Groups feature separates peer-group configuration from update group generation. BGP neighbor configuration is no longer restricted by outbound routing policies, and update-groups can belong to different address families. Even though BGP update-message generation has been separated from peer-group configuration, peer group configuration still has the following limitations:

- A neighbor can belong only to one peer group.
- Neighbors that belong to different address families cannot belong to the same peer group.
- Different sets of policies cannot be grouped and applied to a neighbor.

To address the limitations of peer groups, the BGP Configuration Using Peer Templates feature was introduced along with the BGP Dynamic Update Peer-Groups feature.

A peer template is a configuration pattern that can be applied to neighbors that share common policies. Peer templates are reusable and support inheritance, which allows the network operator to group and apply distinct neighbor configurations for BGP neighbors that share common policies. Peer templates also allow the network operator to define very complex configuration patterns through the ability of a peer template to inherit a configuration from another peer template.



Peer Templates

- **Peer session templates are used to group and apply the configuration of general session commands that are common to all address-family and NLRI configuration modes.**
- **Peer policy templates are used to group and apply the configuration of commands that are applied within specific address-family and NLRI configuration modes.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-12

There are two types of peer templates:

- Peer session templates are used to group and apply the configuration of general session commands that are common to all address-families and Network Layer Reachability Information (NLRI) configuration modes.
- Peer policy templates are used to group and apply the configuration of commands that are applied within specific address-families and NLRI configuration modes.

Peer templates improve the flexibility and enhance the capability of neighbor configuration. Peer templates also provide an alternative to peer group configuration and overcome some limitations of peer groups. With the configuration of the BGP Configuration Using Peer Templates feature and the support of the BGP Dynamic Update Peer-Groups feature, the network operator no longer needs to configure peer groups in BGP and can benefit from improved configuration flexibility and faster convergence.

The inheritance capability is a key component of peer-template operation. Inheritance in a peer template is similar to the node and tree structures commonly found in general computing—for example, file and directory trees. A peer template can directly or indirectly inherit a configuration from another peer template. The directly inherited peer template represents the tree in the structure, and the indirectly inherited peer template represents a node in the tree. Because each node also supports inheritance, branches can be created that apply the configurations of all indirectly inherited peer templates within a chain that traces back to the directly inherited peer template or the source of the tree. This structure eliminates the need to repeat configuration statements that are commonly reapplied to groups of neighbors, because common configuration statements can be applied once and then indirectly inherited by peer templates that are applied to neighbor groups with common configurations.

Inheritance expands the scalability and flexibility of neighbor configuration by allowing you to chain together peer-template configurations to create simple configurations that inherit common configuration statements or complex configurations that apply very specific configuration statements along with common inherited configurations.

Peer Session Commands

General session commands that are common for neighbors that are configured in different address families can be configured within the same peer session template. Peer session templates are created and configured in peer session configuration mode. Only general session commands can be configured in a peer session template.

General session commands can be configured once in a peer session template and then applied to many neighbors through the direct application of a peer session template or through indirect inheritance from a peer session template. The configuration of peer session templates simplifies the configuration of general session commands that are commonly applied to all neighbors within an AS.

Peer session templates support direct and indirect inheritance. A peer can be configured with only one peer session template at a time, and that peer session template can contain only one indirectly inherited peer session template. However, each inherited session template can also contain one indirectly inherited peer session template. So, only one directly applied peer session template and up to seven additional indirectly inherited peer session templates can be applied, allowing you to apply a maximum of eight peer session configurations to a neighbor: the configuration from the directly inherited peer session template and the configurations from up to seven indirectly inherited peer session templates. Inherited peer session templates are evaluated first, and the directly applied template is evaluated and applied last. So, if a general session command is reapplied with a different value, the subsequent value has priority and overwrites the previous value that was configured in the indirectly inherited template.

Peer session templates support only general session commands. BGP policy configuration commands that are configured only for specific address families or NLRI configuration modes are configured with peer policy templates.

template peer-session

To create a peer session template and enter session template configuration mode, use the **template peer-session** command in router configuration mode.

- **template peer-session** *session-template-name*

To remove a peer session template, use the **no** form of this command.

- **no template peer-session** *session-template-name*

Syntax Description

Parameter	Description
template-name	Name or tag for the peer session template

The **inherit peer-session** command configures a peer session template to inherit the configuration of another peer session template.

inherit peer-session

To configure a peer session template to inherit the configuration of another peer session template, use the **inherit peer-session** command in session-template configuration mode.

- **inherit peer-session** *session-template-name*

To remove an inherit statement from a peer session template, use the **no** form of this command.

- **no inherit peer-session** *session-template-name*

Syntax Description

Parameter	Description
session-template-name	Name of the peer session template to be inherited

The **neighbor inherit peer-session** command is used to send locally configured session templates to the specified neighbor. If the session template is configured to inherit configurations from other session templates, the specified neighbor will also indirectly inherit these configurations from the other session templates. A neighbor can directly inherit only one peer session template and indirectly inherit up to seven peer session templates.

Note A BGP neighbor cannot be configured to work with both peer groups and peer templates. A BGP neighbor can be configured to belong only to a peer group or to inherit policies only from peer templates.

neighbor inherit peer-session

To send a peer session template to a neighbor so that the neighbor can inherit the configuration, use the **neighbor inherit peer-session** command in address family or router configuration mode.

- **neighbor** *ip-address* **inherit peer-session** *session-template-name*

To stop sending the peer session template, use the **no** form of this command.

- **no neighbor** ***ip-address*** **inherit peer-session** *session-template-name*

Syntax Description

Parameter	Description
ip-address	IP address of the neighbor
session-template name	Name or tag for the peer session template

The **show ip bgp template peer-session** command is used to display locally configured peer session templates. The output can be filtered to display a single peer session template with the peer-session-name argument. This command also supports all standard output modifiers.

show ip bgp template peer-session

To display peer policy template configurations, use the **show ip bgp template peer-session** command in EXEC mode.

- **show ip bgp template peer-session** [*session-template-name*]

Syntax Description

Parameter	Description
session-template-name	(Optional) Name of a locally configured peer session template

Peer Policy Commands

Peer policy templates are used to configure BGP policy commands that are configured for neighbors that belong to specific address families and NLRI configuration modes. Like peer session templates, peer policy templates are configured once and then applied to many neighbors through the direct application of a peer policy template or through inheritance from peer policy templates. The configuration of peer policy templates simplifies the configuration of BGP policy commands that are applied to all neighbors within an AS.

Peer policy templates support direct and indirect inheritance from up to eight peer policy templates. Inherited peer policy templates are configured with sequence numbers, like route-maps. An inherited peer policy template, like a route-map, is evaluated starting with the inherit statement with the lowest sequence number and ending with the highest sequence number. However, there is a difference: A peer policy template will not fall through like a route-map. Every sequence is evaluated, and if a BGP policy command is reapplied with different value, it overwrites any previous value from a lower sequence number.

Peer policy templates support only general policy commands. BGP policy configuration commands that are configured only for specific address families or NLRI configuration modes are configured with peer policy templates.

inherit peer-policy

To configure a peer policy template to inherit the configuration from another peer policy template, use the **inherit peer-policy** command in policy-template configuration mode.

- **inherit peer-policy** *policy-template-name sequence-number*

To remove an inherit statement from a peer policy template, use the **no** form of this command.

- **no inherit peer-policy** *policy-template-name sequence-number*

Syntax Description

Parameter	Description
peer-policy-name	Name of the peer policy template to be inherited.
sequence-number	Sequence number that sets the order in which the peer policy template is evaluated. Like a route-map sequence number, the lowest sequence number is evaluated first.

The **neighbor inherit peer-policy** command is used to send locally configured policy templates to the specified neighbor. If the policy template is configured to inherit configurations from other peer policy templates, the specified neighbor will also indirectly inherit these configurations from the other peer policy templates. A directly applied peer policy template can directly or indirectly inherit configurations from up to seven peer policy templates. So, a total of eight peer policy templates can be applied to a neighbor or neighbor group.

neighbor inherit peer-policy

To send a peer policy template to a neighbor so that the neighbor can inherit the configuration, use the **neighbor inherit peer-policy** command in address-family or router configuration mode.

- **neighbor** *ip-address* **inherit peer-policy** *policy-template-name*

To stop sending the peer policy template, use the **no** form of this command.

- **no neighbor** *ip-address* **inherit peer-policy** *policy-template-name*

Syntax Description

Parameter	Description
ip-address	IP address of the neighbor
policy-template-name	Name or tag for the peer policy template

show ip bgp template peer-policy

To display locally configured peer policy templates, use the **show ip bgp template peer-policy** command in EXEC mode.

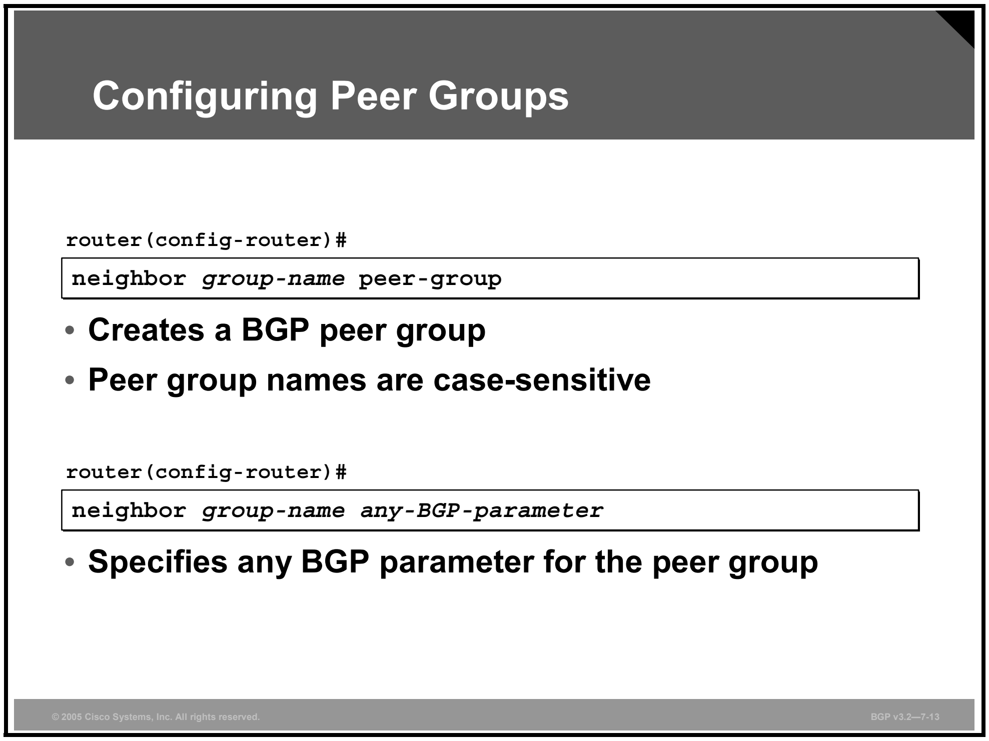
- **show ip bgp template peer-policy** [*policy-template-name*]

Syntax Description

Parameter	Description
policy-template-name	(Optional) Name of a locally configured peer policy template

Configuring Peer Groups

This topic identifies the commands that are used to configure BGP peer groups on Cisco IOS routers.



Configuring Peer Groups

```
router(config-router)#
neighbor group-name peer-group
```

- **Creates a BGP peer group**
- **Peer group names are case-sensitive**

```
router(config-router)#
neighbor group-name any-BGP-parameter
```

- **Specifies any BGP parameter for the peer group**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-13

To configure BGP peer groups on Cisco IOS routers, perform the following steps:

Step 1 Create a BGP peer group.

Step 2 Specify parameters for the BGP peer group.

Step 3 Create a BGP neighbor.

Step 4 Assign a neighbor into the peer group.

neighbor peer-group (Creating)

To create a BGP peer group, use the **neighbor peer-group** router configuration command.

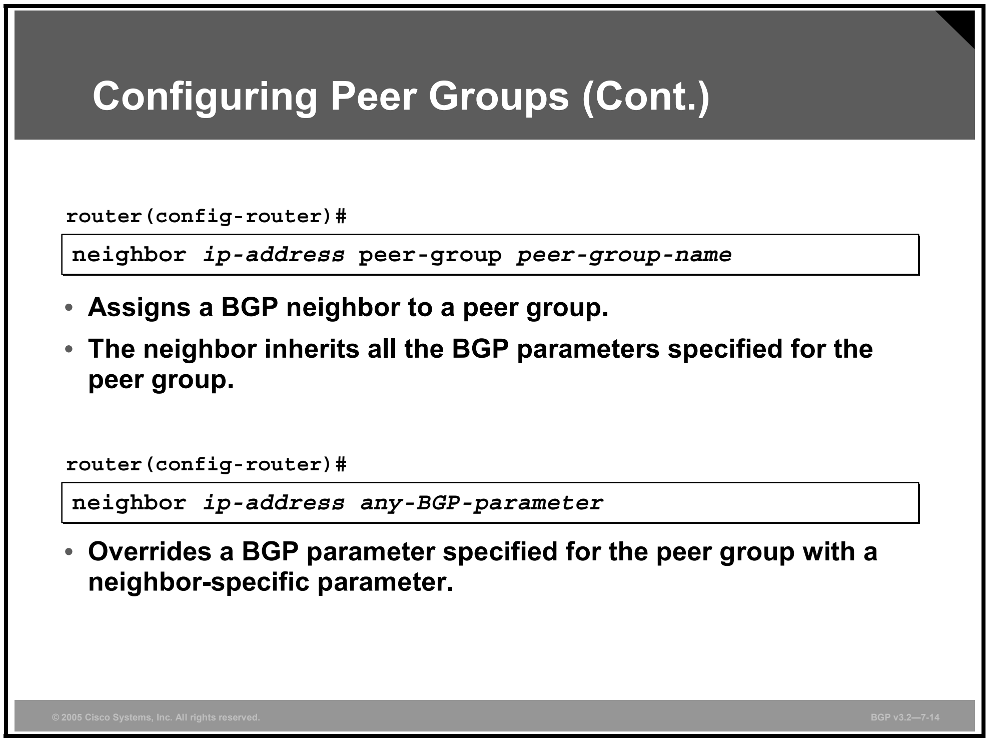
- **neighbor** *peer-group-name* **peer-group**

To remove the peer group and all of its members, use the **no** form of this command.

- **no neighbor** *peer-group-name* **peer-group**

After you have created a peer group using the **neighbor peer-group** command, you can configure it with the **neighbor** commands. By default, members of the peer group inherit all the configuration options of the peer group. You can also configure members to override the options that do not affect outbound updates.

Peer group members will always inherit the following configuration options: **remote-as** (if configured), **version**, **update-source**, **out-route-map**, **out-filter-list**, **out-dist-list**, **minimum-advertisement-interval**, and **next-hop-self**. All peer group members will inherit changes that are made to the peer group.



Configuring Peer Groups (Cont.)

```
router(config-router)#
neighbor ip-address peer-group peer-group-name
```

- Assigns a BGP neighbor to a peer group.
- The neighbor inherits all the BGP parameters specified for the peer group.

```
router(config-router)#
neighbor ip-address any-BGP-parameter
```

- Overrides a BGP parameter specified for the peer group with a neighbor-specific parameter.

neighbor peer-group (Assigning Members)

To configure a BGP neighbor to be a member of a peer group, use the **neighbor peer-group** router configuration command.

- **neighbor** *ip-address* **peer-group** *peer-group-name*

To remove the neighbor from the peer group, use the **no** form of this command.

- **no neighbor** *ip-address* **peer-group** *peer-group-name*

Syntax Description

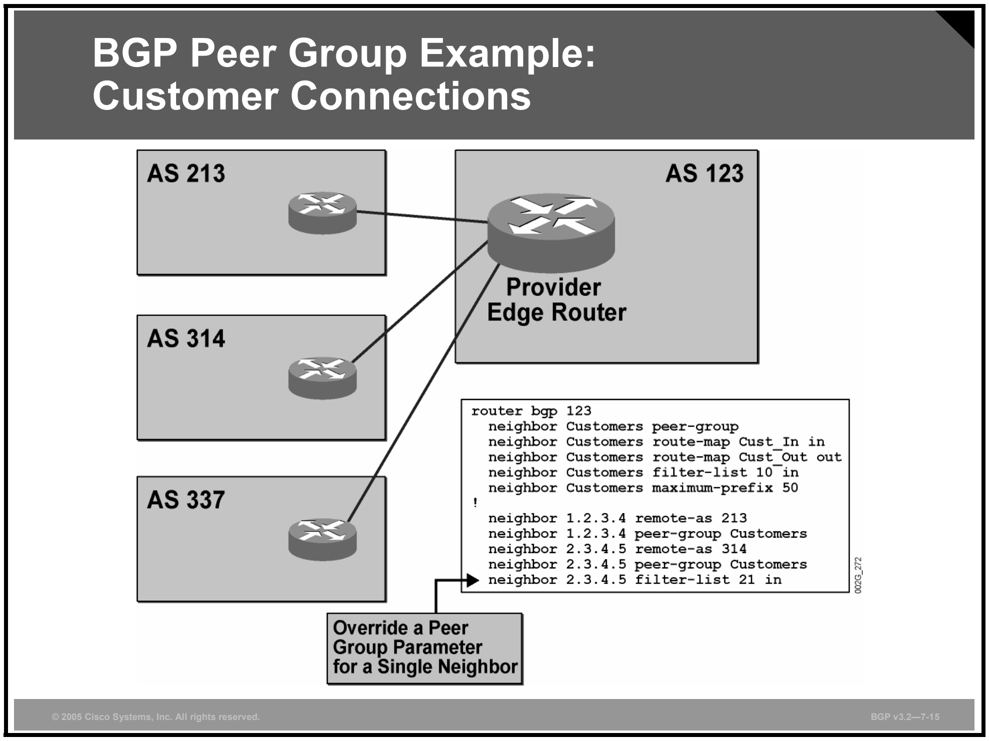
Parameter	Description
ip-address	IP address of the BGP neighbor that belongs to the peer group that is specified by the tag
peer-group-name	Name of the BGP peer group to which this neighbor belongs

After you have assigned an actual neighbor to be a member of the peer group, all configurations made to the peer group template are then applied to all the neighbors that are assigned to that peer group.

Through configuration, peer group configurations may be overridden for an individual neighbor, provided that the changes apply only to incoming updates. Remember that outgoing updates are prepared only once and then replicated.

Example: BGP Peer Group–Customer Connections

In this example, the router in AS 123 is being configured with a peer group named "Customers."



BGP Peer Group Example: Customer Connections

```
router bgp 123
  neighbor Customers peer-group
  neighbor Customers route-map Cust_In in
  neighbor Customers route-map Cust_Out out
  neighbor Customers filter-list 10 in
  neighbor Customers maximum-prefix 50
!
  neighbor 1.2.3.4 remote-as 213
  neighbor 1.2.3.4 peer-group Customers
  neighbor 2.3.4.5 remote-as 314
  neighbor 2.3.4.5 peer-group Customers
  neighbor 2.3.4.5 filter-list 21 in
```

Override a Peer Group Parameter for a Single Neighbor

This peer group is used for all customers of the service provider because they share an almost identical routing policy. The peer group is first created as a template, which is configured with an incoming route-map named "Cust_In" and an outgoing route-map named "Cust_Out," as well as an incoming AS-path filter-list number of 10. The peer group is also configured with a maximum limit of 50 received prefixes.

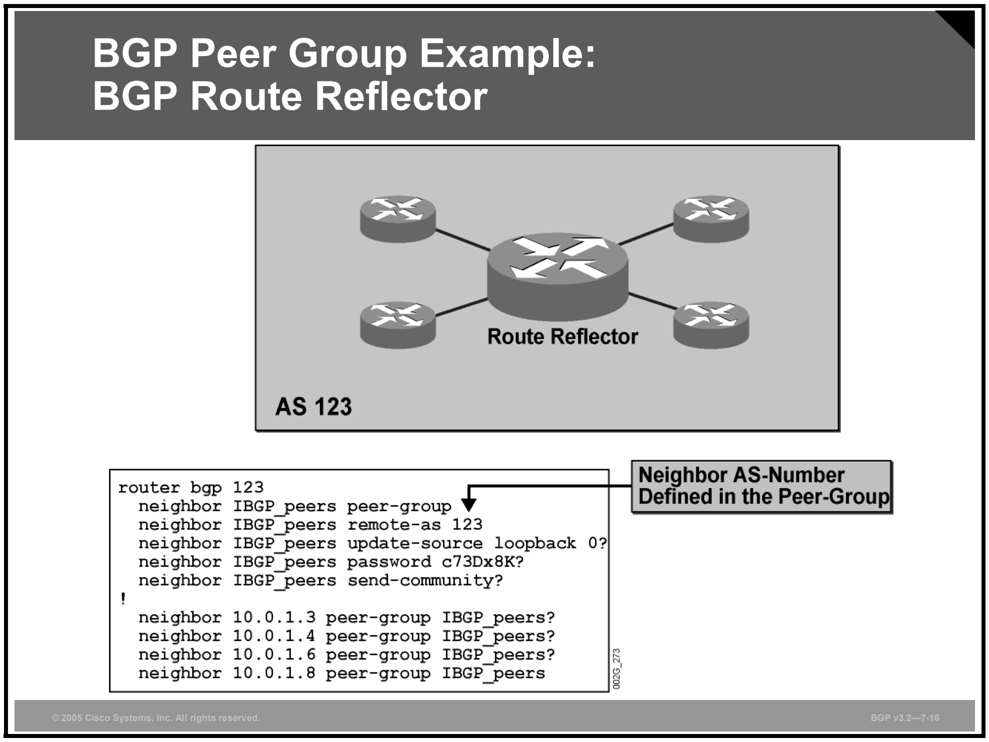
Then neighbors in AS 213 and AS 314 are assigned to the peer group. These additions mean that the router in AS 123 will attempt to open BGP sessions with those routers. If the BGP sessions succeed, the route-maps Cust_In and Cust_Out will be used with both neighbors on incoming and outgoing routes, respectively. The maximum number of received prefixes that can configured in the Customer peer group will also be applied to both neighbors.

Filter-list 10 will be used to filter out any incoming routes from peer group members unless otherwise specified. However, in the case of the neighbor in AS 314, the individual configuration of filter-list 21 will override the peer group configuration, and the AS-path access-list number 21 will be used instead.

The peer group is a very powerful tool when network administrators are dealing with a large number of neighbors with almost identical configurations. However, if any of the customers require routing information that differs from that of other members of the Customer peer group, then that neighbor must be removed from the peer group and configured as an individual neighbor. This would be the reason for not including the AS 337 neighbor in the Customer peer group.

Example: BGP Peer Group—BGP Route Reflector

In this example, a router acting as a BGP route reflector has four IBGP neighbors.



BGP Peer Group Example:
BGP Route Reflector

```
router bgp 123
  neighbor IBGP_peers peer-group
  neighbor IBGP_peers remote-as 123
  neighbor IBGP_peers update-source loopback 0?
  neighbor IBGP_peers password c73Dx8K?
  neighbor IBGP_peers send-community?
!
  neighbor 10.0.1.3 peer-group IBGP_peers?
  neighbor 10.0.1.4 peer-group IBGP_peers?
  neighbor 10.0.1.6 peer-group IBGP_peers?
  neighbor 10.0.1.8 peer-group IBGP_peers
```

Neighbor AS-Number Defined in the Peer-Group

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-16

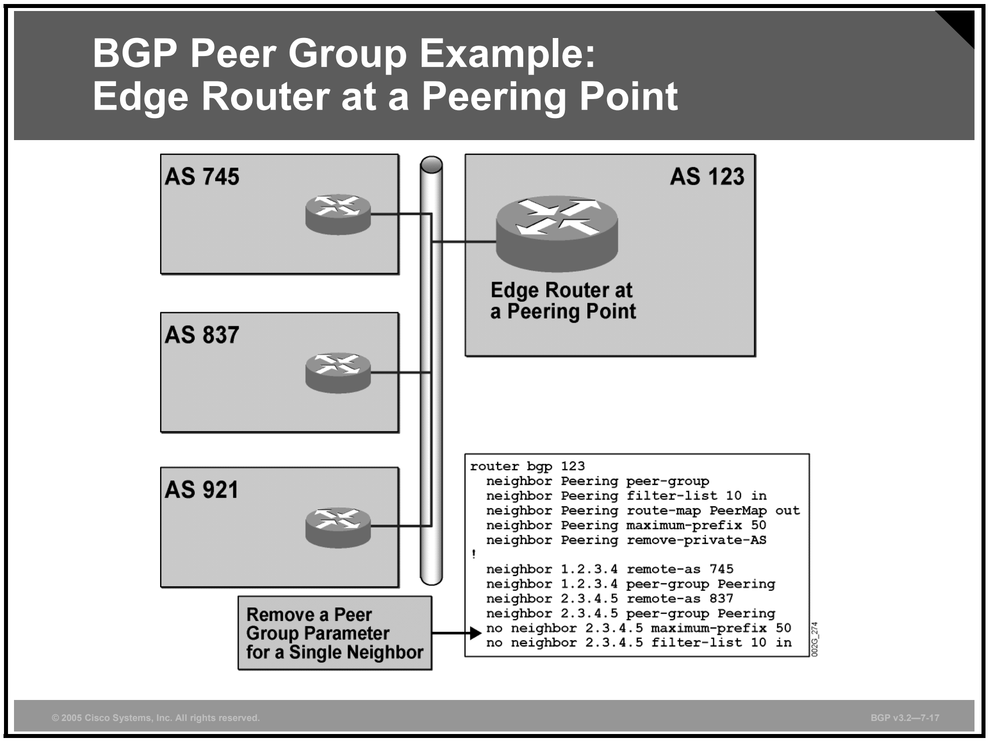
In a large AS, some routers may have a large number of IBGP sessions. A peer group that is named “IBGP_peers” is created to handle all of the IBGP sessions. The peer group is created and configured with the remote AS, update-source, MD5 authentication, and community-passing parameters. When the actual neighbors are configured as members of the peer group, all these configuration parameters will apply to all of the neighbors.

In the case of IBGP, the remote AS can also be configured as a part of the peer group configuration because the AS number is the same for each of the peer group members.

The peer group is a very powerful tool when you are dealing with a large number of IBGP neighbors because you can give all of the neighbors the same configuration to ensure a consistent AS-wide routing policy.

Example: BGP Peer Group—Edge Router at a Peering Point

In this example, the router in AS 123 is being configured with a peer group named “Peering.”



BGP Peer Group Example:
Edge Router at a Peering Point

AS 745
AS 837
AS 921
AS 123
Edge Router at a Peering Point
Remove a Peer Group Parameter for a Single Neighbor
002G_274

```
router bgp 123
  neighbor Peering peer-group
  neighbor Peering filter-list 10 in
  neighbor Peering route-map PeerMap out
  neighbor Peering maximum-prefix 50
  neighbor Peering remove-private-AS
!
  neighbor 1.2.3.4 remote-as 745
  neighbor 1.2.3.4 peer-group Peering
  neighbor 2.3.4.5 remote-as 837
  neighbor 2.3.4.5 peer-group Peering
  no neighbor 2.3.4.5 maximum-prefix 50
  no neighbor 2.3.4.5 filter-list 10 in
```

This peer group is used for all peer providers because they share an almost identical routing policy. The peer group is first created as a template, which is configured with an incoming AS-path filter-list (list 10) and an outgoing route-map named “PeerMap.” The maximum number of received prefixes is also set in the peer group to 50. The peer group has also been configured to remove private AS numbers (AS numbers in the range 64512 to 65535 inclusive) from all AS paths before the routes are sent to the peer AS.

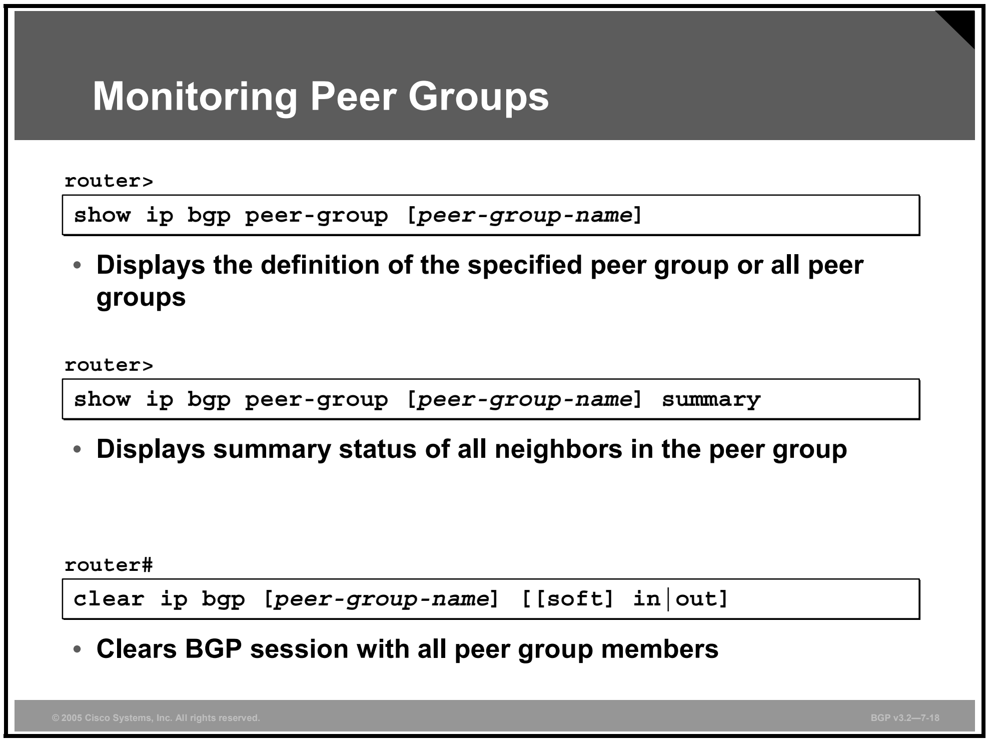
The neighbors in AS 745 and in AS 837 are then assigned to the peer group, meaning that the router in AS 123 will attempt to open BGP sessions with those routers. If the BGP sessions are successfully established, filter-list 10 and the route-map PeerMap, as configured in the peer group, will be applied to incoming and outgoing routes from both neighbors, respectively.

As defined in the router configuration, filter-list 10 filters out any incoming routes from peer group members unless otherwise specified. However, in the case of the neighbor in AS 837, the individual configuration of **no filter-list 10** will override the peer group configuration, and thus, the filter-list will not be used for this neighbor. The limitation on the number of received routes from AS 837 is also removed from the neighbor in AS 837.

The peer group is a very powerful tool when you are dealing with a large number of neighbors with almost identical configurations. However, if any of the customers that are assigned to the peer group require routing information that is different from other members of the peer group, then that neighbor must be removed from the peer group and configured individually.

Monitoring Peer Groups

This topic lists the Cisco IOS commands that are required to monitor BGP peer groups.



Monitoring Peer Groups

```
router>
show ip bgp peer-group [peer-group-name]
```

- **Displays the definition of the specified peer group or all peer groups**

```
router>
show ip bgp peer-group [peer-group-name] summary
```

- **Displays summary status of all neighbors in the peer group**

```
router#
clear ip bgp [peer-group-name] [[soft] in|out]
```

- **Clears BGP session with all peer group members**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-18

show ip bgp peer-group

To display information about BGP peer groups, use the **show ip bgp peer-group** EXEC command.

- **show ip bgp peer-group** [*peer-group-name*] **summary**

Syntax Description

Parameter	Description
`peer-group-name`	(Optional) Displays information about that specific peer group
`summary`	(Optional) Displays a summary of the status of all the members of a peer group

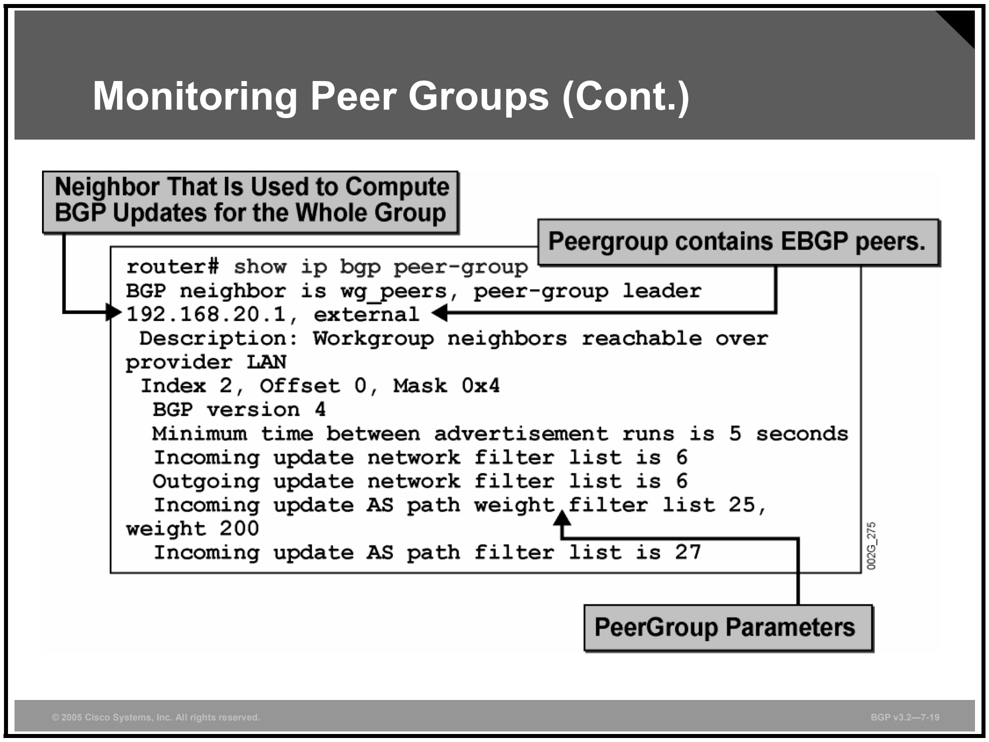
clear ip bgp

To reset the BGP sessions with all the members of a peer group, use the **clear ip bgp** EXEC command.

- **clear ip bgp** {***** | *neighbor-address* | *peer-group-name*} [**soft** [**in** | **out**]]

Syntax Description

Parameter	Description
`*`	Resets all current BGP sessions.
`neighbor-address`	Resets only the identified BGP neighbor.
`peer-group-name`	(Optional) Displays information about that specific peer group.
`soft`	(Optional) Initiates soft reconfiguration.
`in` | **`out`**	(Optional) Triggers inbound or outbound soft reconfiguration. If you do not specify the **in** or **out** option, both inbound and outbound soft reconfiguration are triggered.



Monitoring Peer Groups (Cont.)

Neighbor That Is Used to Compute BGP Updates for the Whole Group

Peergroup contains EBGP peers.

```
router# show ip bgp peer-group
BGP neighbor is wg_peers, peer-group leader
192.168.20.1, external
 Description: Workgroup neighbors reachable over
provider LAN
 Index 2, Offset 0, Mask 0x4
  BGP version 4
  Minimum time between advertisement runs is 5 seconds
  Incoming update network filter list is 6
  Outgoing update network filter list is 6
  Incoming update AS path weight filter list 25,
weight 200
  Incoming update AS path filter list is 27
```

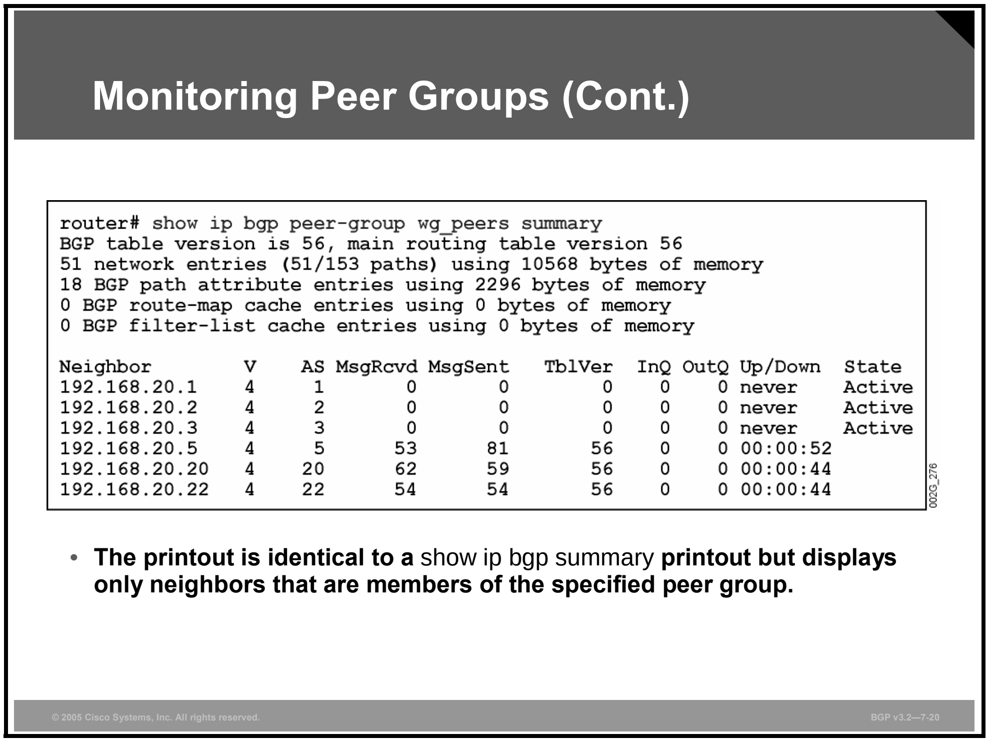
PeerGroup Parameters

In this example, the **show ip bgp peer-group** command displays information about the peer group named “wg_peers.” One of the peer group members (192.168.20.1) has been selected as the peer group leader, meaning that all outgoing BGP updates are processed as if they were being sent to this neighbor. Thereafter, the update is replicated on all the BGP sessions to the other members in the wg_peers peer group.

A peer group should have only IBGP members or EBGP members. In the example shown, the members are EBGP neighbors.

All parameters that are configured for the peer group are listed in the **show ip bgp peer-group** command output. In the example, the peer group wg_peers has been configured in the following manner:

```
neighbor wg_peers description Workgroup neighbors reachable
over provider LAN
neighbor wg_peers distribute-list 6 in
neighbor wg_peers distribute-list 6 out
neighbor wg_peers filter-list 25 weight 200
neighbor wg_peers filter-list 27 in
```



Monitoring Peer Groups (Cont.)

```
router# show ip bgp peer-group wg_peers summary
BGP table version is 56, main routing table version 56
51 network entries (51/153 paths) using 10568 bytes of memory
18 BGP path attribute entries using 2296 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State
192.168.20.1    4     1       0       0        0    0    0 never    Active
192.168.20.2    4     2       0       0        0    0    0 never    Active
192.168.20.3    4     3       0       0        0    0    0 never    Active
192.168.20.5    4     5      53      81       56    0    0 00:00:52
192.168.20.20   4    20      62      59       56    0    0 00:00:44
192.168.20.22   4    22      54      54       56    0    0 00:00:44
```

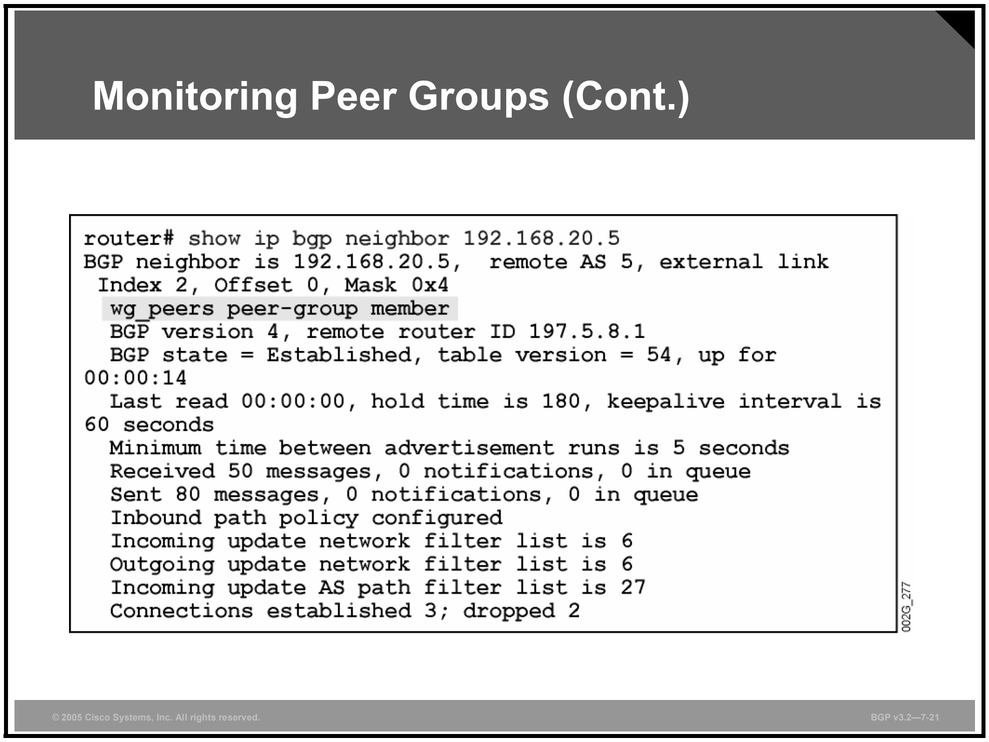
002G_276

- **The printout is identical to a** show ip bgp summary **printout but displays only neighbors that are members of the specified peer group.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-20

The **show ip bgp peer-group** *peer-group-name* **summary** command is used in this example to display only the summary status information about the neighbors who are members of the peer group wg_peers.

Note	The **show ip bgp summary** command is described in the module “BGP Overview.”



Monitoring Peer Groups (Cont.)

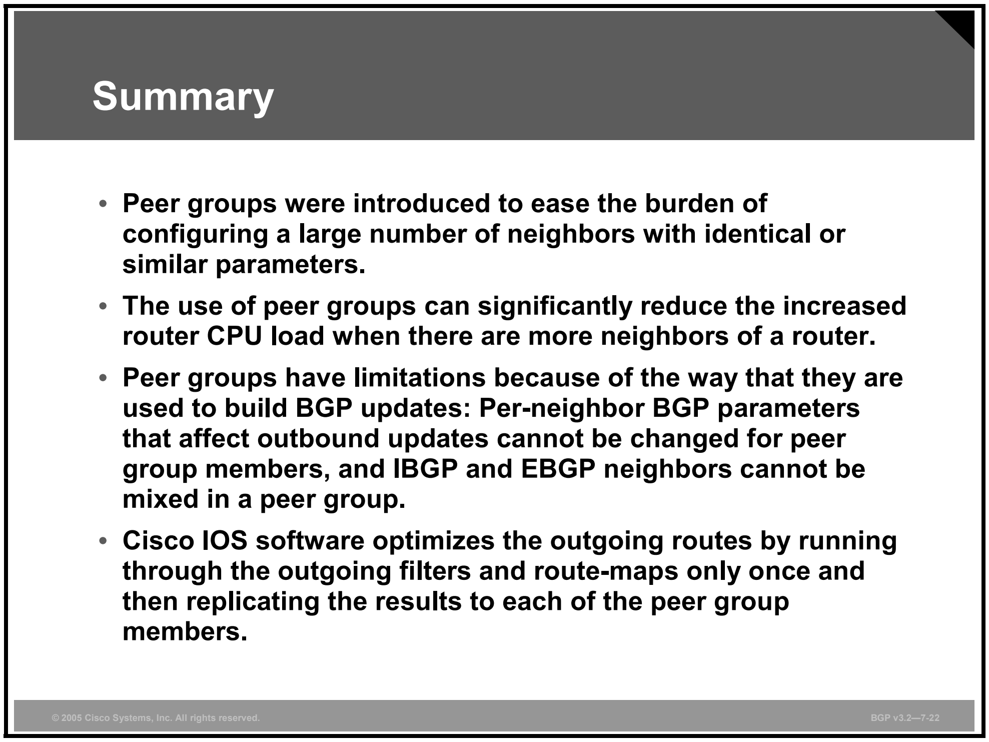
```
router# show ip bgp neighbor 192.168.20.5
BGP neighbor is 192.168.20.5,  remote AS 5, external link
 Index 2, Offset 0, Mask 0x4
  wg_peers peer-group member
  BGP version 4, remote router ID 197.5.8.1
  BGP state = Established, table version = 54, up for
00:00:14
  Last read 00:00:00, hold time is 180, keepalive interval is
60 seconds
  Minimum time between advertisement runs is 5 seconds
  Received 50 messages, 0 notifications, 0 in queue
  Sent 80 messages, 0 notifications, 0 in queue
  Inbound path policy configured
  Incoming update network filter list is 6
  Outgoing update network filter list is 6
  Incoming update AS path filter list is 27
  Connections established 3; dropped 2
```

The **show ip bgp neighbor** command displays additional information about BGP neighbors that are members of a peer group.

In the example here, the membership in the peer group wg_peers is indicated, in addition to other information as displayed by the command.

Summary

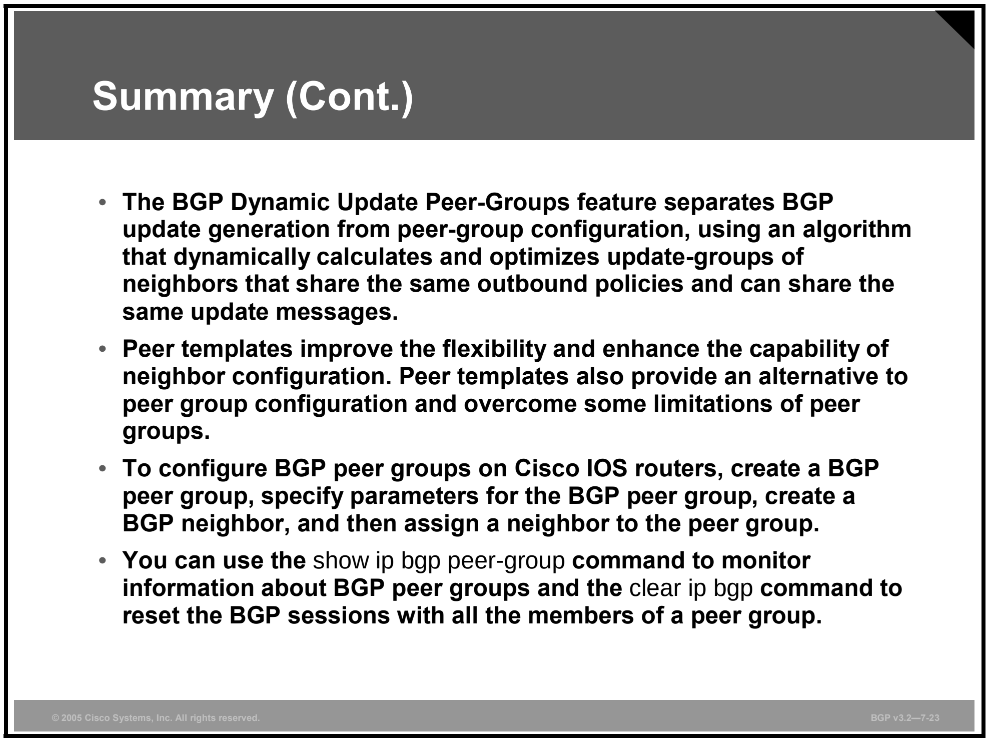
This topic summarizes the key points discussed in this lesson.



Summary

- Peer groups were introduced to ease the burden of configuring a large number of neighbors with identical or similar parameters.
- The use of peer groups can significantly reduce the increased router CPU load when there are more neighbors of a router.
- Peer groups have limitations because of the way that they are used to build BGP updates: Per-neighbor BGP parameters that affect outbound updates cannot be changed for peer group members, and IBGP and EBGP neighbors cannot be mixed in a peer group.
- Cisco IOS software optimizes the outgoing routes by running through the outgoing filters and route-maps only once and then replicating the results to each of the peer group members.

BGP v3.2—7-22



Summary (Cont.)

- The BGP Dynamic Update Peer-Groups feature separates BGP update generation from peer-group configuration, using an algorithm that dynamically calculates and optimizes update-groups of neighbors that share the same outbound policies and can share the same update messages.
- Peer templates improve the flexibility and enhance the capability of neighbor configuration. Peer templates also provide an alternative to peer group configuration and overcome some limitations of peer groups.
- To configure BGP peer groups on Cisco IOS routers, create a BGP peer group, specify parameters for the BGP peer group, create a BGP neighbor, and then assign a neighbor to the peer group.
- You can use the show ip bgp peer-group command to monitor information about BGP peer groups and the clear ip bgp command to reset the BGP sessions with all the members of a peer group.

BGP v3.2—7-23

Lesson 4

Using BGP Route Dampening

Overview

Even when a Border Gateway Protocol (BGP) implementation is correctly configured and highly robust, the performance of the routing process on any given router is limited. Limiting the propagation of unstable routes, specifically when they are not beneficial to the network, becomes an important issue because it reduces the processing requirements of the router that is forced to process routing table state changes.

Route dampening is a BGP feature that has been designed to reduce BGP processing requirements by minimizing the propagation of unstable routes to BGP peers. Autonomous system (AS) border routers, in any BGP implementation, cannot rely upon external peers to sufficiently shield the AS from routing table instability. Route dampening allows route instability to be contained at an AS border router that borders the instability.

This lesson describes the purpose and operation of the route-dampening feature on Cisco IOS routers. Also discussed in this lesson are the Cisco IOS commands that are required to enable route dampening, modify default dampening parameters, and release a route that has been suppressed because of dampening. The Cisco IOS commands that are used to monitor route dampening are also discussed.

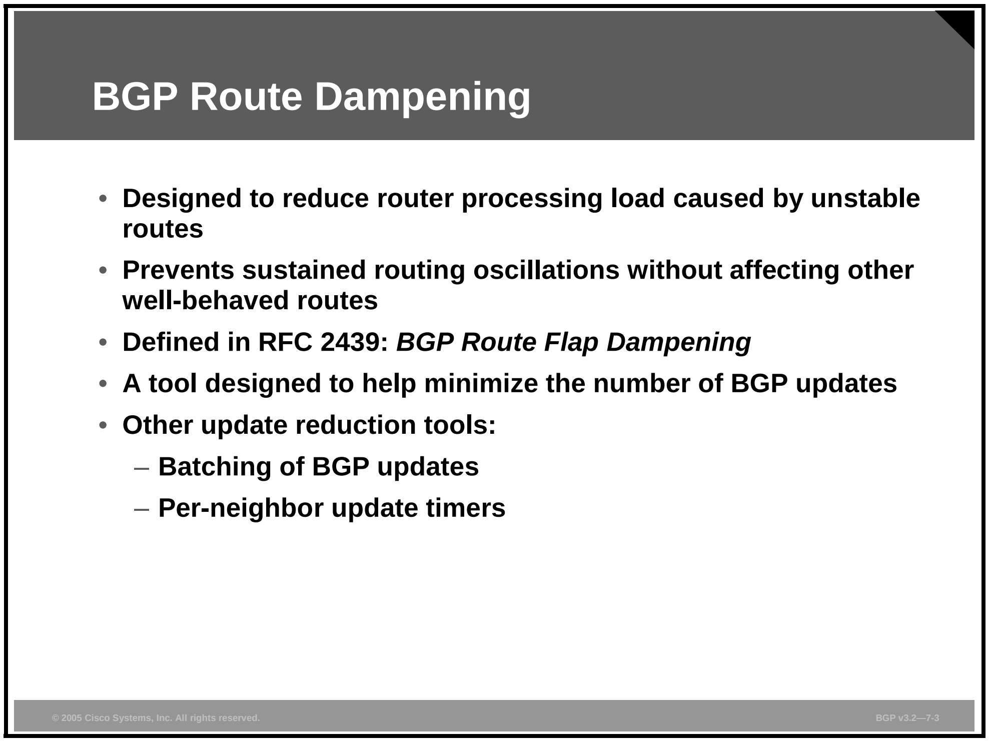
Objectives

Upon completing this lesson, you will be able to use route dampening to minimize the impact of unstable routes. This ability includes being able to meet these objectives:

- Describe the purpose of BGP route dampening
- Describe the operation of BGP route dampening
- Identify the Cisco IOS commands that are required to configure BGP route dampening
- Identify the Cisco IOS commands that are required to release dampened routes
- Identify the Cisco IOS commands that are required to monitor BGP route-dampening

BGP Route Dampening

This topic describes the purpose of BGP route dampening.



BGP Route Dampening

- Designed to reduce router processing load caused by unstable routes
- Prevents sustained routing oscillations without affecting other well-behaved routes
- Defined in RFC 2439: *BGP Route Flap Dampening*
- A tool designed to help minimize the number of BGP updates
- Other update reduction tools:
  - Batching of BGP updates
  - Per-neighbor update timers

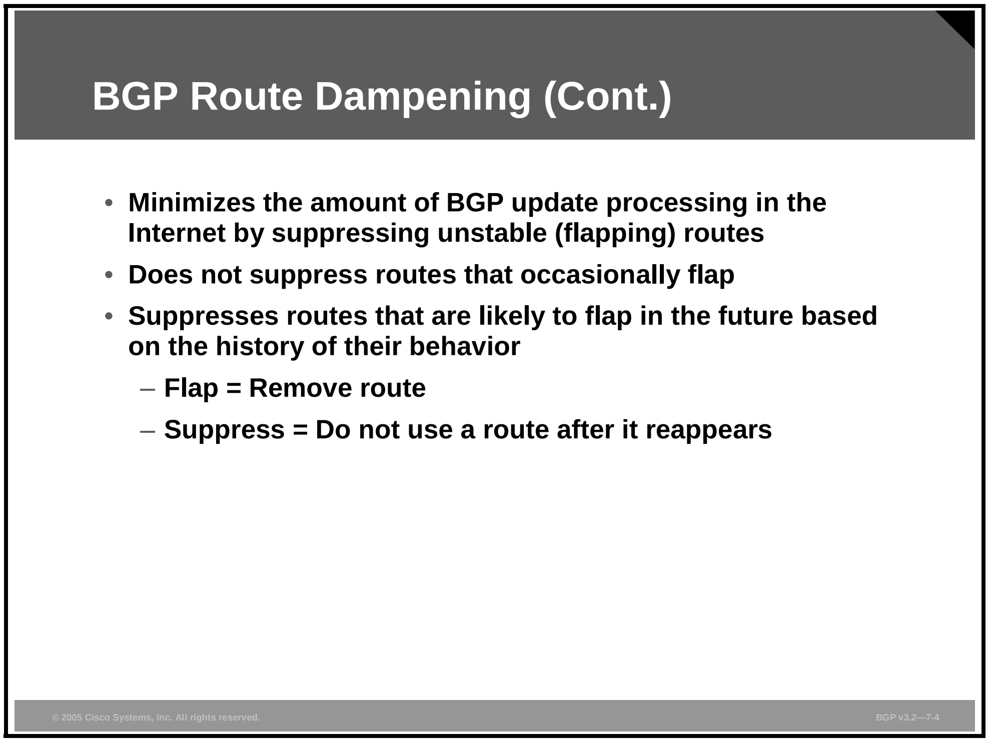
© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-3

BGP is the only routing protocol that is designed for large internetworks with the specific intention of carrying a large number of prefixes. There are several mechanisms that are built into BGP that ensure maximum router stability.

For example, a BGP router does not forward BGP routing updates immediately after receiving them. Every time a BGP router sends an update, it starts a 5-second timer for internal neighbors and a 30-second timer for external neighbors. No new updates can be sent until that timer expires. The result is that if a router contains a link that is flapping repeatedly (available, then unavailable, then available, then unavailable, and so on) at a rate of once per second, external routers see the flap at a much slower rate. Routers that are external to the source of the flap are not forced to recalculate the best path every second but, at most, every 30 seconds.

Reducing the rate at which neighboring routers process flapping routes assists in reducing the requirements to process the BGP update. However, routers that process routing updates for unstable routes are still wasting resources in determining the best route to the destination. Because the unstable route is oscillating between up and down, each route update that a router receives causes it to process the unstable route all over again. A better approach is to remove the update about the route until it can be guaranteed that the destination is more stable.

With this goal in mind, an additional BGP scalability mechanism called route flap dampening was created to reduce route update processing requirements by suppressing unstable routes.



BGP Route Dampening (Cont.)

- Minimizes the amount of BGP update processing in the Internet by suppressing unstable (flapping) routes
- Does not suppress routes that occasionally flap
- Suppresses routes that are likely to flap in the future based on the history of their behavior
  - Flap = Remove route
  - Suppress = Do not use a route after it reappears

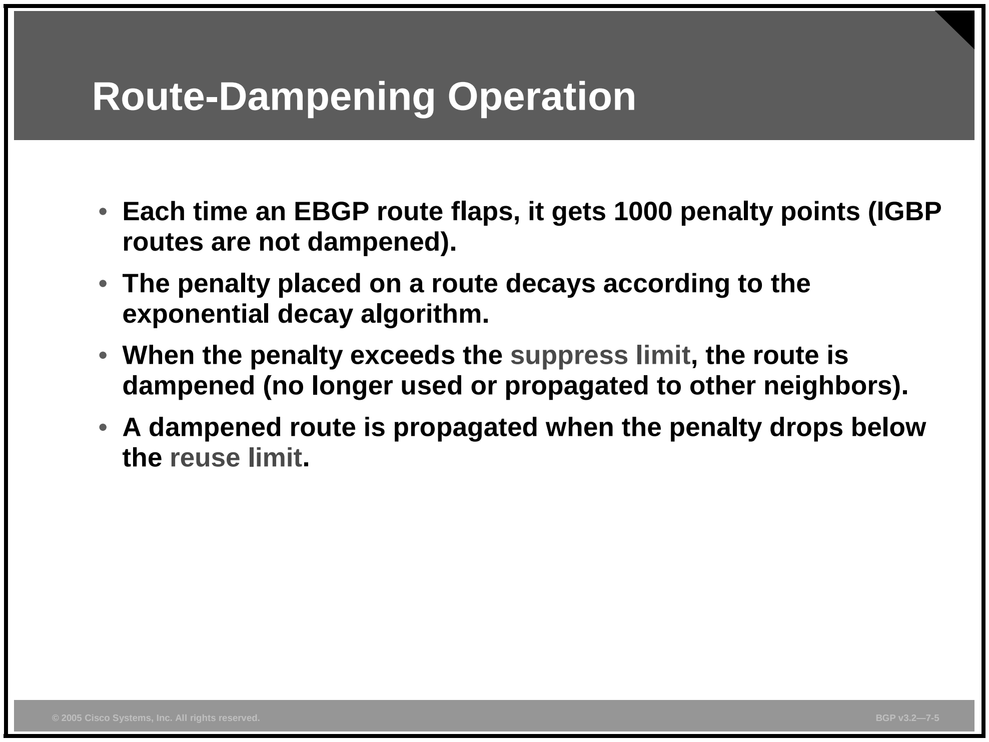
BGP v3.2—7-4

Most service providers hold routing information for the entire Internet. Therefore, a flapping link somewhere in the Internet can cause all routers in the Internet to keep processing changes because of one single link. If, however, one of the autonomous systems in the Internet implements route dampening, the flapping network is suppressed. The route is not propagated further to other autonomous systems until the configured rules of route dampening allow it.

A “flap” refers to a route that is repeatedly available, then unavailable, then available, then unavailable, and so on. If a route flaps once or twice, it is typically not considered a flap from an administrative perspective. If the flapping happens more often, however, there is probably something wrong with the destination and the route should be suppressed. The BGP router stores a suppressed route in the BGP table but does not consider it in the BGP path-selection process and does not therefore propagate it to other BGP neighbors or use it for data forwarding.

Route-Dampening Operation

This topic describes the operation of BGP route dampening.



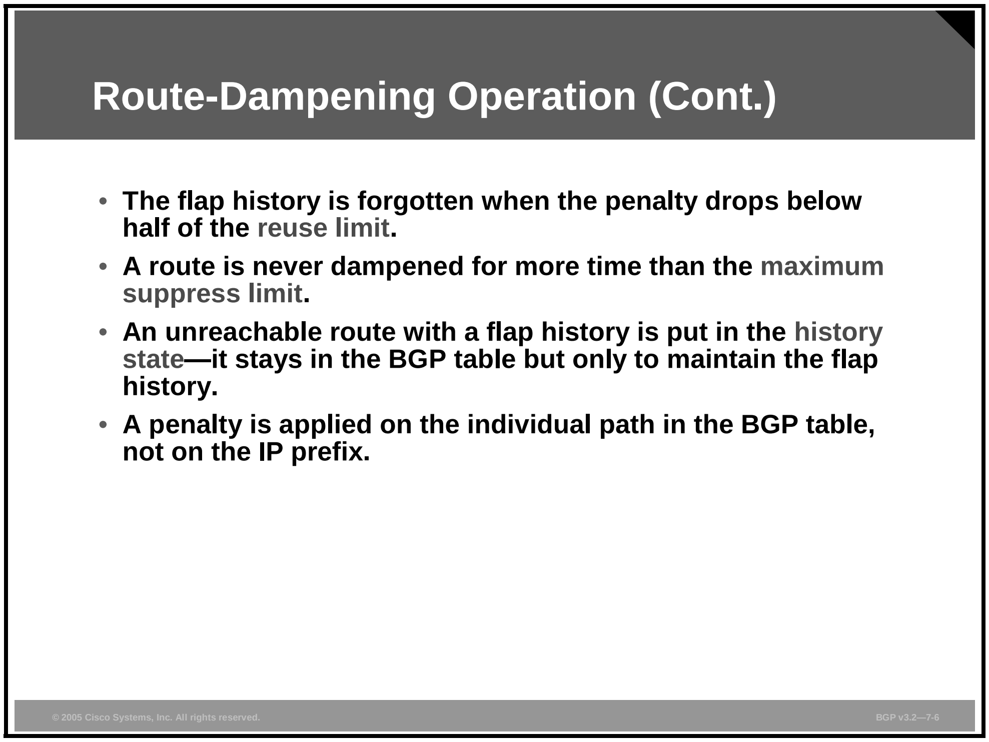
Route-Dampening Operation

- **Each time an EBGP route flaps, it gets 1000 penalty points (IGBP routes are not dampened).**
- **The penalty placed on a route decays according to the exponential decay algorithm.**
- **When the penalty exceeds the suppress limit, the route is dampened (no longer used or propagated to other neighbors).**
- **A dampened route is propagated when the penalty drops below the reuse limit.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-5

A BGP router with route dampening enabled keeps track of all routes (even those that are unreachable) so that it can recall the penalties that are assigned to each route. Every time a route flap occurs, the flapping route receives 1000 penalty points. The penalty is gradually decreased through the use of a decaying algorithm. If a route flaps several times, it will be penalized (gain enough penalty points) and subsequently reach and exceed the suppress limit.

Any route that reaches the suppress limit is no longer forwarded to other neighbors until the assigned penalty is once again below the reuse limit. An exponential decay algorithm reduces penalty points that are applied to a flapping route. After the number of penalty points that are assigned to a route falls below the reuse limit, the BGP router once again advertises the route.



Route-Dampening Operation (Cont.)

- The flap history is forgotten when the penalty drops below half of the reuse limit.
- A route is never dampened for more time than the maximum suppress limit.
- An unreachable route with a flap history is put in the history state—it stays in the BGP table but only to maintain the flap history.
- A penalty is applied on the individual path in the BGP table, not on the IP prefix.

BGP v3.2—7-6

A router stops tracking penalty points when they are below half of the reuse limit.

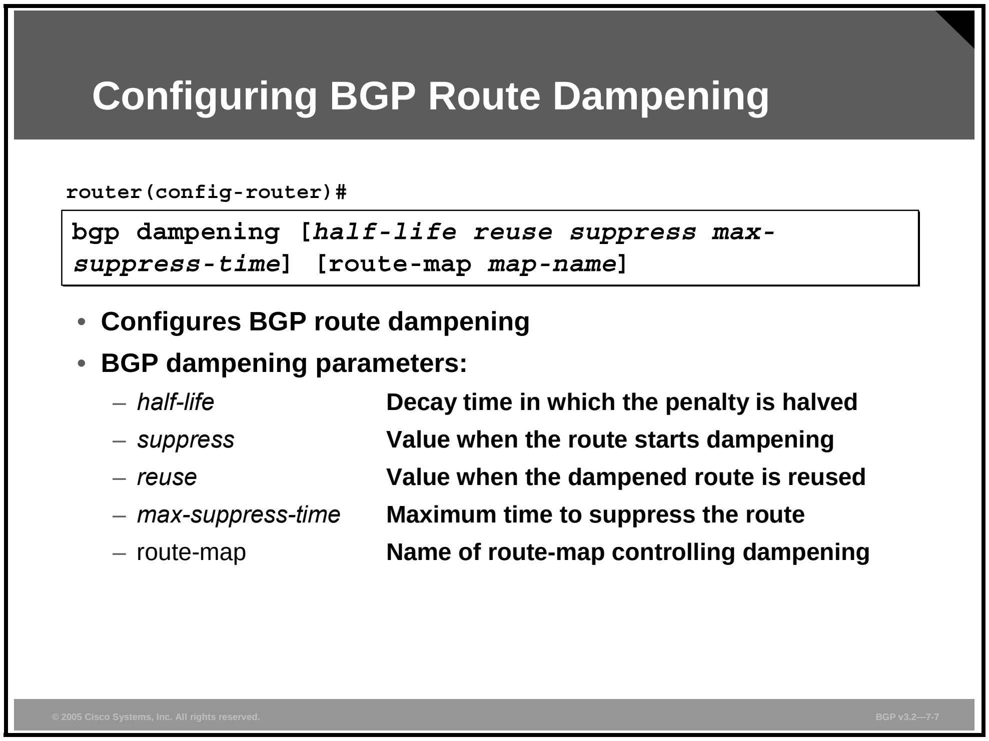
The maximum suppress limit defines the maximum duration that a route can be suppressed after it has been suppressed.

After route dampening is enabled, the router never removes a route from the BGP table. A route that has been withdrawn by a BGP neighbor can still be seen in the BGP table and is marked with an “h” (history state).

A penalty is always applied to a path and not a prefix. If one of the paths is flapping, it does not mean that the destination is flapping.

Configuring BGP Route Dampening

This topic identifies the Cisco IOS commands that are required to configure BGP route dampening.



Configuring BGP Route Dampening

```
router(config-router)#
bgp dampening [half-life reuse suppress max-
suppress-time] [route-map map-name]
```

- **Configures BGP route dampening**
- **BGP dampening parameters:**

– *half-life*	**Decay time in which the penalty is halved**
– *suppress*	**Value when the route starts dampening**
– *reuse*	**Value when the dampened route is reused**
– *max-suppress-time*	**Maximum time to suppress the route**
– route-map	**Name of route-map controlling dampening**

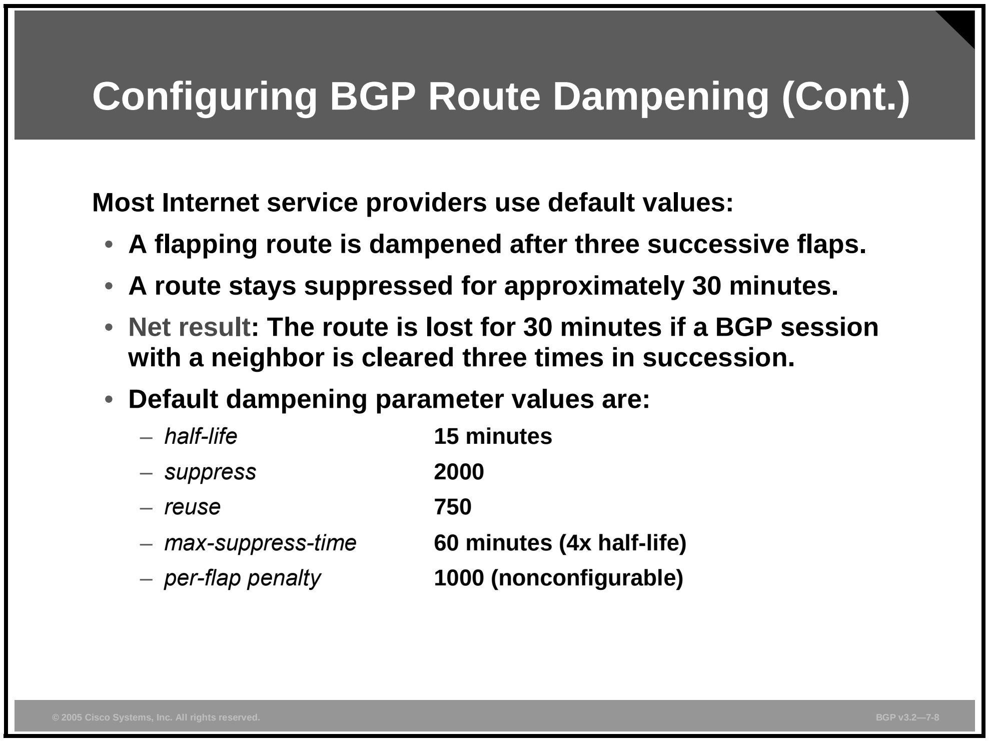
BGP v3.2—7-7

To enable route dampening, use the **bgp dampening** command. Optionally, you can change the default settings of the route-dampening parameters.

Route flap dampening requires the following parameters:

- *half-life*: The half-life is the time that is needed for the penalty to halve (default is 15 minutes).
- *suppress*: When a route has more penalty points than the suppress limit, the route is suppressed (default is 2000).
- *reuse*: After the flapping has stopped and the penalty for a route has fallen below the reuse limit, the route is unsuppressed (default is 750).
- *max-suppress-time*: No route can be suppressed longer than the max-suppress-time minutes (default is 1 hour; maximum is 255 minutes).

You can specify the four route flap-dampening parameters directly with the **bgp dampening** command. Alternatively, you can create a route-map that specifies different dampening parameters for different sets of routes, and then you can apply the route-map with the **bgp dampening route-map** command.



Configuring BGP Route Dampening (Cont.)

Most Internet service providers use default values:

- **A flapping route is dampened after three successive flaps.**
- **A route stays suppressed for approximately 30 minutes.**
- Net result**: The route is lost for 30 minutes if a BGP session with a neighbor is cleared three times in succession.**
- **Default dampening parameter values are:**

– *half-life*	**15 minutes**
– *suppress*	**2000**
– *reuse*	**750**
– *max-suppress-time*	**60 minutes (4x half-life)**
– *per-flap penalty*	**1000 (nonconfigurable)**

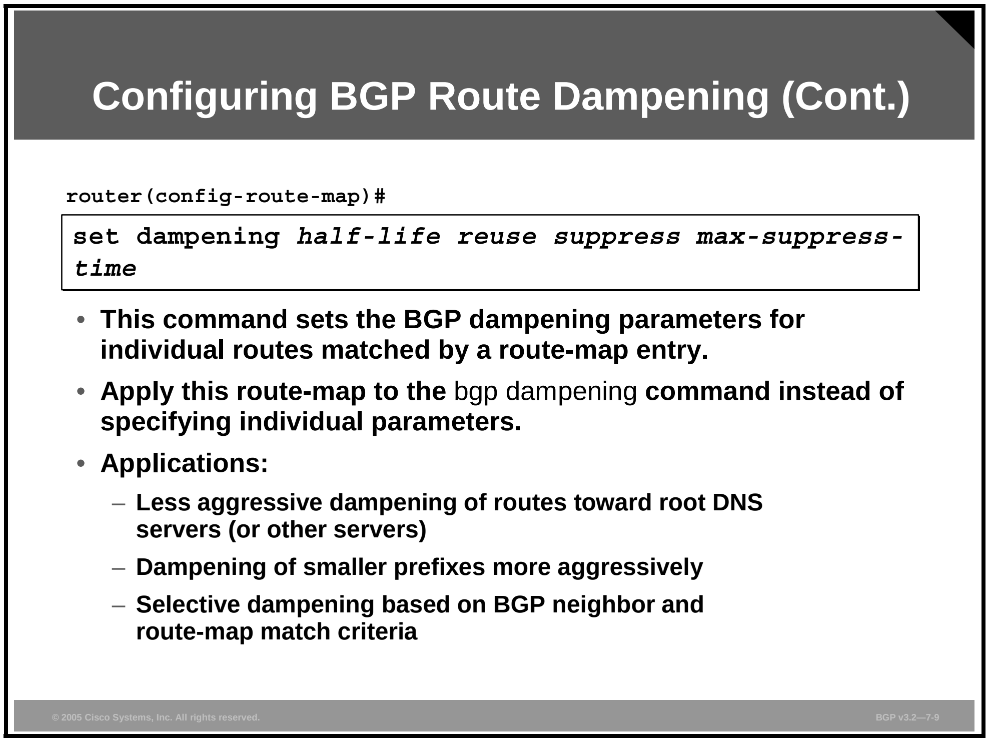
This sample calculation shows how long a route that flaps three times is suppressed with the default values of the Cisco IOS route-dampening parameters. Each time that a route flaps, it accumulates another 1000 points. After the third flap, the route has almost 3000 points. Remember that the penalty is gradually decreased through the use of a decaying algorithm, causing a reduction in the number of points that the route accumulates. It takes 15 minutes for the penalty to drop below 1500 (provided there are no further flaps) and another 15 minutes to drop below the reuse limit of 750.

Many service providers change the default parameters to allow a maximum suppress time of several hours.

Note Using the **clear ip bgp *** command is regarded as a flap to neighboring autonomous systems. Using this command several times may cause neighboring autonomous systems to suppress prefixes for some time even if there is nothing wrong with the route.

Note Using the **clear ip bgp *** [**soft**] [**in** | **out**] command is not regarded as a flap, and this command should be used instead of **clear ip bgp *** for clearing the neighbor relationships.

You can change all default values if you specify them in the optional parameters of the **bgp dampening** command. The per-flap penalty is the only value that is not configurable.



Configuring BGP Route Dampening (Cont.)

```
router(config-route-map)#
```

```
set dampening half-life reuse suppress max-suppress-time
```

- **This command sets the BGP dampening parameters for individual routes matched by a route-map entry.**
- **Apply this route-map to the** bgp dampening **command instead of specifying individual parameters.**
- **Applications:**
  - **Less aggressive dampening of routes toward root DNS servers (or other servers)**
  - **Dampening of smaller prefixes more aggressively**
  - **Selective dampening based on BGP neighbor and route-map match criteria**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-9

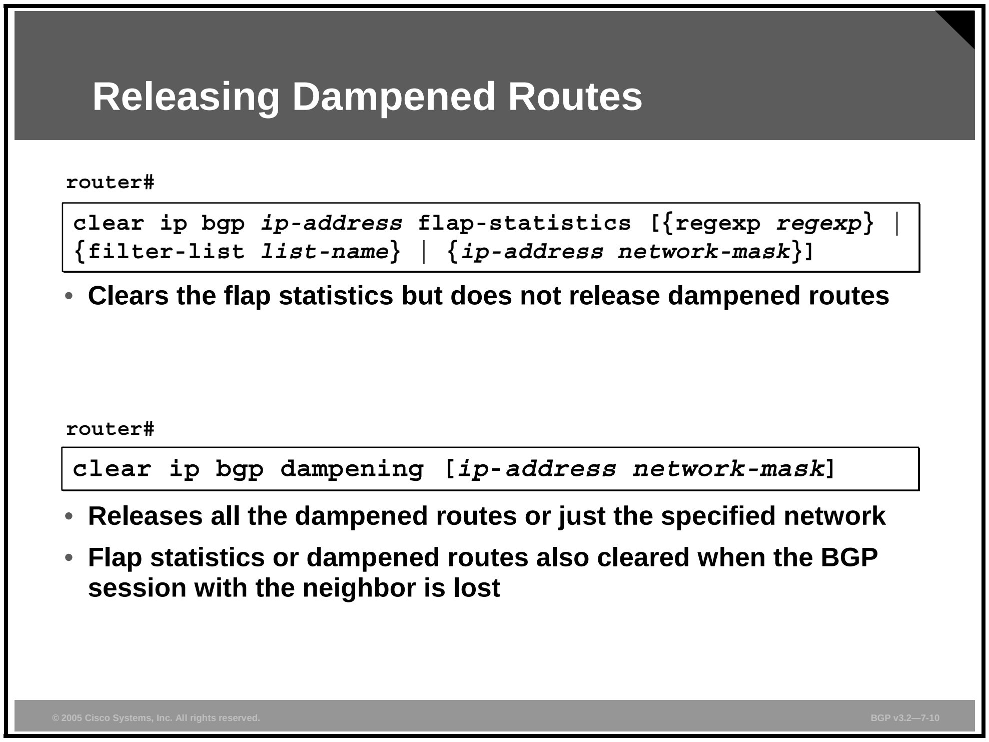
Many service providers prefer to implement selective dampening. Larger prefixes are usually less likely to flap and should not be penalized as aggressively as the smaller prefixes that populate most of the BGP table.

You can use a route-map in combination with a prefix-list to match on prefix length and to set different route-dampening parameters for larger prefixes than for smaller ones. A practical service provider policy is to use a route-map to exclude root Domain Name System (DNS) servers from dampening altogether.

You can then attach the route-map to the BGP route-dampening process with the **bgp dampening route-map** command.

Releasing Dampened Routes

This topic identifies the Cisco IOS commands that are used to release dampened routes.



Releasing Dampened Routes

```
router#
clear ip bgp ip-address flap-statistics [{regexp regexp} |
{filter-list list-name} | {ip-address network-mask}]
```

- **Clears the flap statistics but does not release dampened routes**

```
router#
clear ip bgp dampening [ip-address network-mask]
```

- **Releases all the dampened routes or just the specified network**
- **Flap statistics or dampened routes also cleared when the BGP session with the neighbor is lost**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-10

There are two timers that are calculated for every route when it flaps:

- **Time to forget:** The time that it takes before all flap history is deleted. Using the **clear ip bgp flap-statistics** command deletes all penalty information, but it does not release the dampened paths.
- **Time to reuse:** The time that it takes before a route can be considered for best-path processing. Using the **clear ip bgp dampening** command resets this timer for all networks or for specified networks so that they are no longer suppressed. The flap statistics, however, are still kept, and the next flap will cause the previously dampened paths to be suppressed again.

clear ip bgp flap-statistics

To clear BGP flap statistics, use the **clear ip bgp flap-statistics** privileged EXEC command.

- **clear ip bgp** *ip-address* **flap-statistics** [{**regexp** *regexp*} | {**filter-list** *list-name*} | {*ip-address network-mask*}]

Syntax Description

Parameter	Description
ip-address	(Optional) Clears flap statistics for a single entry at this IP address. If this argument is placed before **flap-statistics**, the router clears flap statistics for all paths from the neighbor at this address.
regexp *regexp*	(Optional) Clears flap statistics for all the paths that match the regular expression.
filter-list *list-name*	(Optional) Clears flap statistics for all the paths that pass the access-list.
network-mask	(Optional) Network mask that is applied to the *ip-address* argument.

clear ip bgp dampening

To clear BGP route-dampening information and unsuppress the suppressed routes, use the **clear ip bgp dampening** privileged EXEC command.

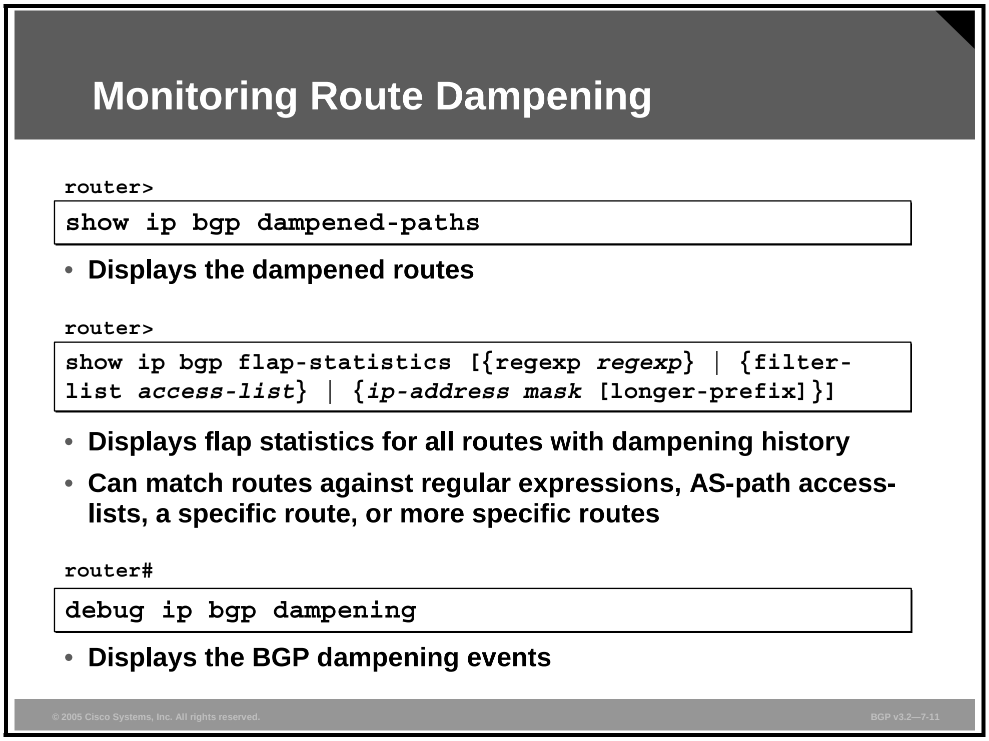
- **clear ip bgp dampening** [*ip-address network-mask*]

Syntax Description

Parameter	Description
ip-address	(Optional) Clears flap statistics for a single entry at this IP address. If this argument is placed before **flap-statistics**, the router clears flap statistics for all paths from the neighbor at this address.
network-mask	(Optional) Network mask that is applied to the *ip-address* argument.

Monitoring Route Dampening

This topic identifies the Cisco IOS commands that are required to monitor BGP route dampening.



Monitoring Route Dampening

router>

```
show ip bgp dampened-paths
```

- Displays the dampened routes

router>

```
show ip bgp flap-statistics [{regexp regexp} | {filter-
list access-list} | {ip-address mask [longer-prefix]}]
```

- Displays flap statistics for all routes with dampening history
- Can match routes against regular expressions, AS-path access-lists, a specific route, or more specific routes

router#

```
debug ip bgp dampening
```

- Displays the BGP dampening events

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-11

The penalty that is placed on a network is decayed until the reuse limit is reached, upon which the route is once again advertised. Every time that a route flap occurs, the penalty is recalculated. In the router, the penalty is encoded as the time that it takes for the penalty to decay below the reuse limit. At half of the reuse limit, the dampening information for the route to the network is removed.

Use the **show ip bgp dampened-paths** command to list all routes that are currently suppressed because of dampening.

Use the **show ip bgp flap-statistics** command to list all routes that have a penalty that is still above the time-to-forget limit. You can also use this command in combination with regular expressions and filter-lists.

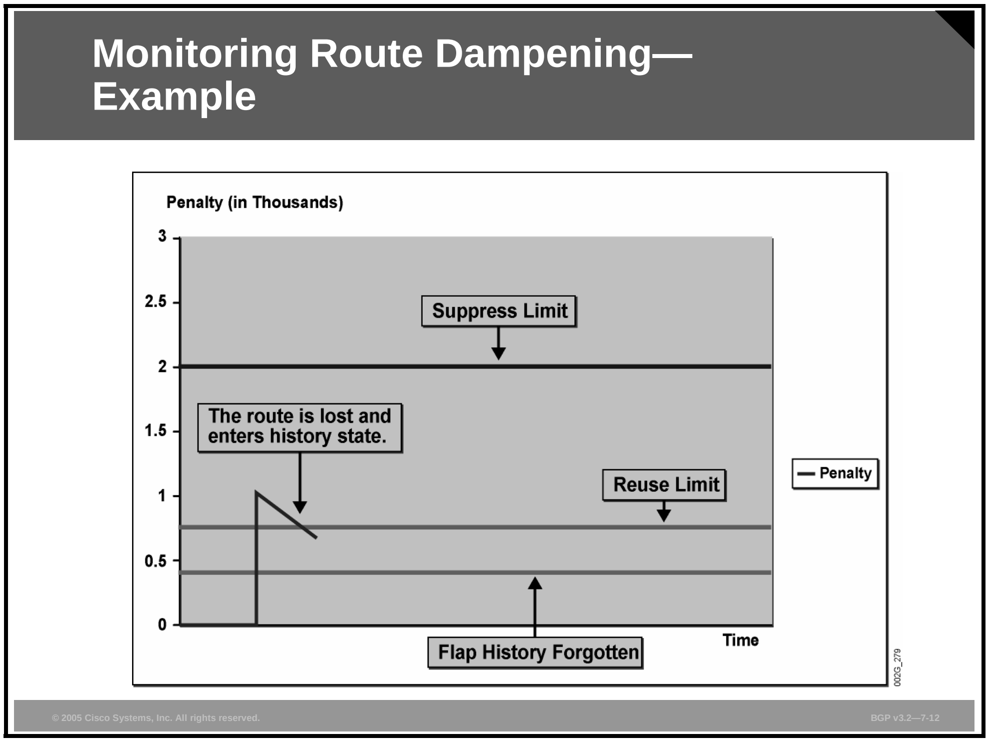
The **show ip bgp flap-statistics prefix** command displays detailed dampening information about a specific network.

The **show ip bgp flap-statistics prefix mask longer-prefix** command displays dampening information about a specific network and its subnets.

Use the **debug ip bgp dampening** command to display BGP dampening events as they occur in real time.

Example: Monitoring Route Dampening

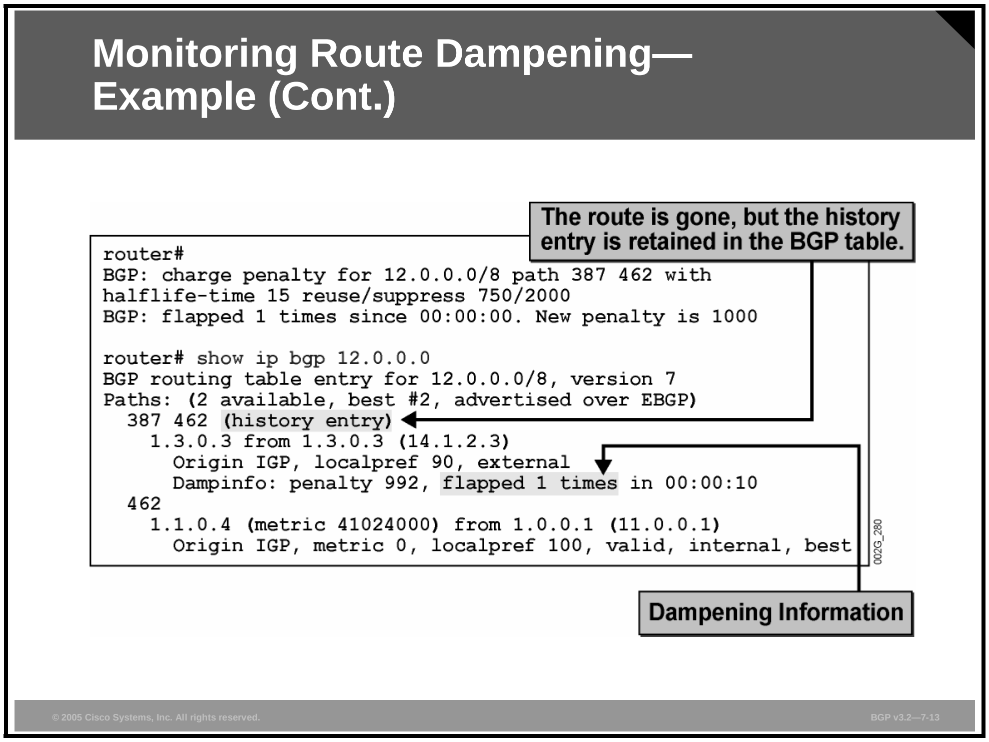
The example shows how, after the first flap (when a route becomes unreachable), the router withdraws the route but keeps it in its own database to keep track of the penalty. The route enters the history state.



Monitoring Route Dampening—Example

© 2005 Cisco Systems, Inc. All rights reserved.

BGP v3.2—7-12



Monitoring Route Dampening—Example (Cont.)

```
router#
BGP: charge penalty for 12.0.0.0/8 path 387 462 with
halflife-time 15 reuse/suppress 750/2000
BGP: flapped 1 times since 00:00:00. New penalty is 1000

router# show ip bgp 12.0.0.0
BGP routing table entry for 12.0.0.0/8, version 7
Paths: (2 available, best #2, advertised over EBGP)
  387 462 (history entry)
    1.3.0.3 from 1.3.0.3 (14.1.2.3)
      Origin IGP, localpref 90, external
      Dampinfo: penalty 992, flapped 1 times in 00:00:10
  462
    1.1.0.4 (metric 41024000) from 1.0.0.1 (11.0.0.1)
      Origin IGP, metric 0, localpref 100, valid, internal, best
```

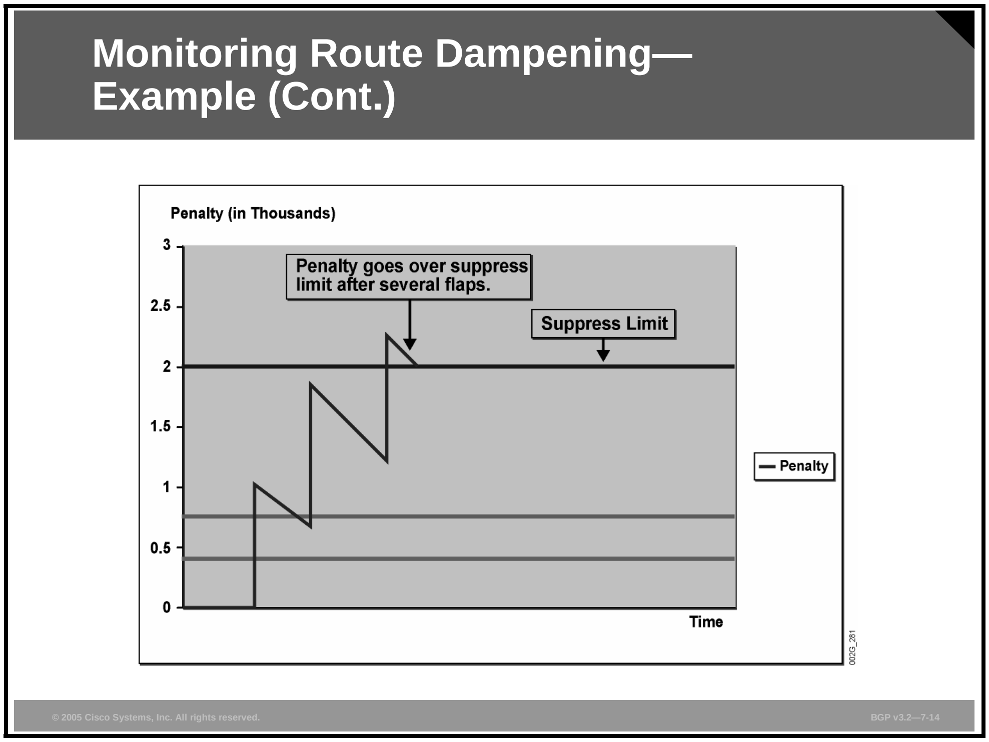
The route is gone, but the history entry is retained in the BGP table.

Dampening Information

Using **show ip bgp** displays suppressed prefixes with the state “h.”

Using **show ip bgp prefix** displays suppressed prefixes that are marked with “history entry.”

If a prefix is in the history state, it means that it is currently unreachable but that the information has been kept in the BGP table to keep track of the penalty.



Monitoring Route Dampening—Example (Cont.)

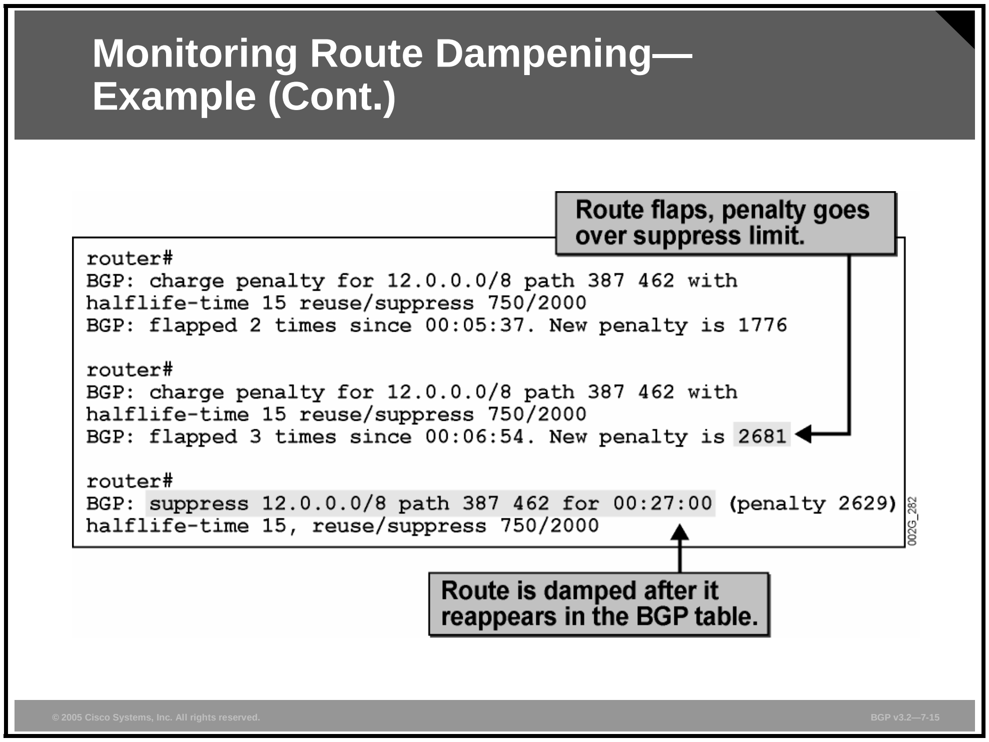
Penalty goes over suppress limit after several flaps.

Suppress Limit

BGP v3.2—7-14

The example here shows how, after the third flap, the penalty of the route exceeds the suppress limit, and the route could be suppressed. When the route exceeds the suppress limit, one of two things could happen:

- The router will put the route in the history state if the route is currently unreachable.
- The router will suppress the route if the route is currently reachable.



Monitoring Route Dampening—Example (Cont.)

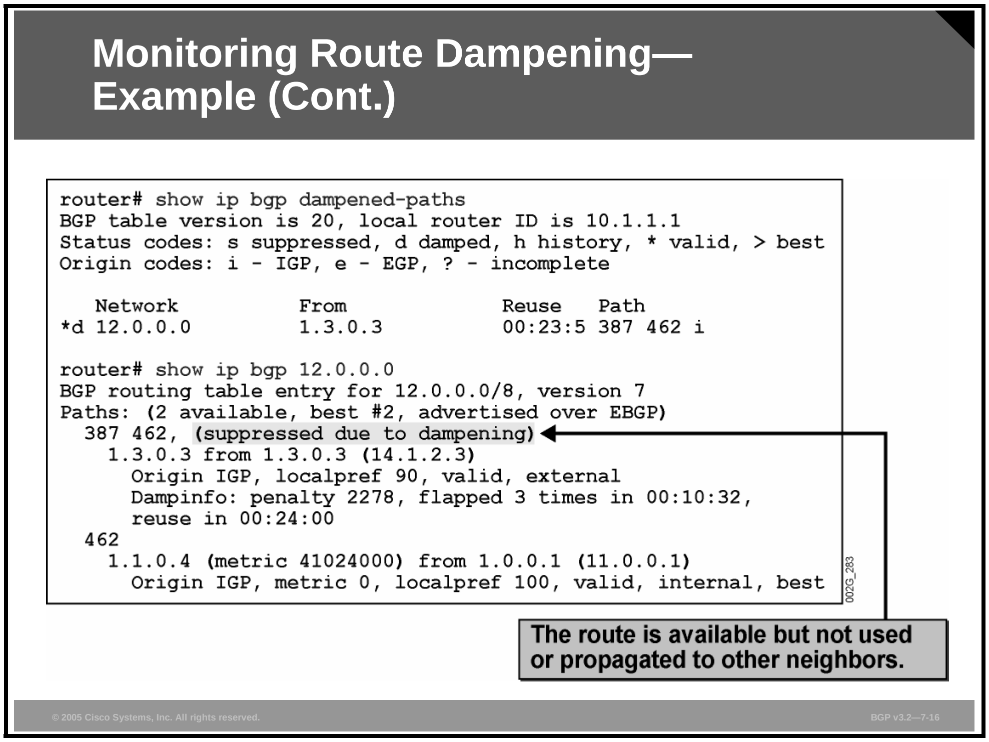
Route flaps, penalty goes over suppress limit.

```
router#
BGP: charge penalty for 12.0.0.0/8 path 387 462 with
halflife-time 15 reuse/suppress 750/2000
BGP: flapped 2 times since 00:05:37. New penalty is 1776

router#
BGP: charge penalty for 12.0.0.0/8 path 387 462 with
halflife-time 15 reuse/suppress 750/2000
BGP: flapped 3 times since 00:06:54. New penalty is 2681

router#
BGP: suppress 12.0.0.0/8 path 387 462 for 00:27:00 (penalty 2629)
halflife-time 15, reuse/suppress 750/2000
```

Route is damped after it reappears in the BGP table.

The figure displays the debugging messages that accompany the three route flaps that were illustrated in the previous figure. After the 12.0.0.0/8 prefix flaps for the third time, the router assigns a new penalty of 2681 to the route. Because the new penalty exceeds the suppress limit of 2000, the 12.0.0.0/8 prefix is suppressed by the router for 27 minutes (if it remains stable).

Using the **show ip bgp** command displays all suppressed prefixes that have their state marked as dampened (“d”).



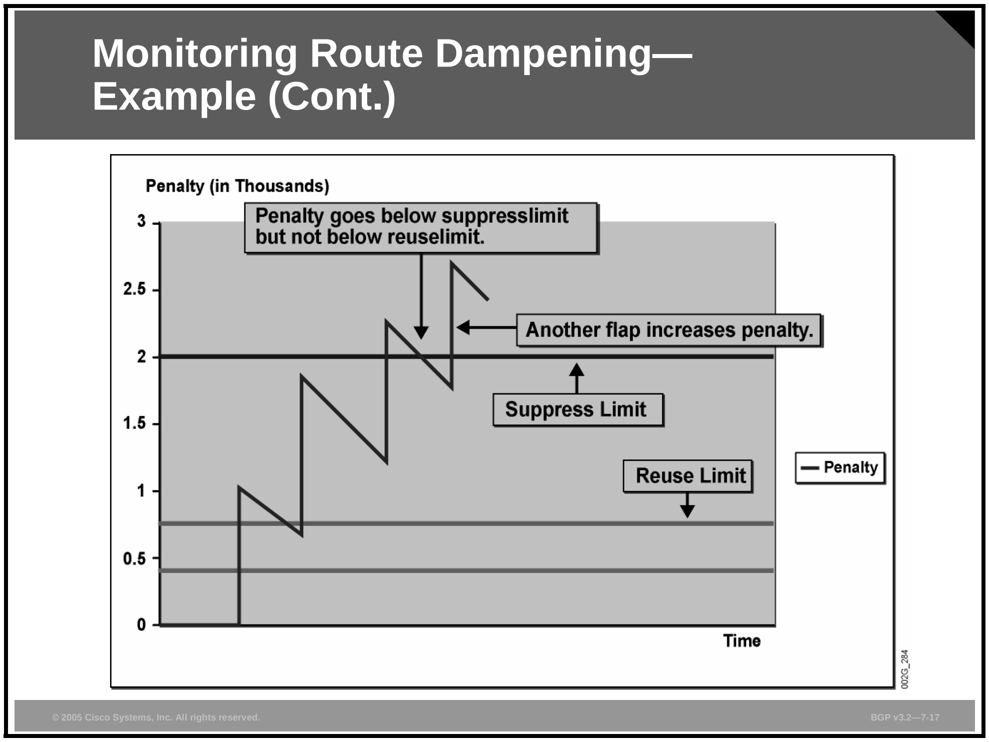
Monitoring Route Dampening—Example (Cont.)

```
router# show ip bgp dampened-paths
BGP table version is 20, local router ID is 10.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          From             Reuse   Path
*d 12.0.0.0         1.3.0.3          00:23:5 387 462 i

router# show ip bgp 12.0.0.0
BGP routing table entry for 12.0.0.0/8, version 7
Paths: (2 available, best #2, advertised over EBGP)
  387 462, (suppressed due to dampening)
    1.3.0.3 from 1.3.0.3 (14.1.2.3)
      Origin IGP, localpref 90, valid, external
      Dampinfo: penalty 2278, flapped 3 times in 00:10:32,
      reuse in 00:24:00
  462
    1.1.0.4 (metric 41024000) from 1.0.0.1 (11.0.0.1)
      Origin IGP, metric 0, localpref 100, valid, internal, best
```

The route is available but not used or propagated to other neighbors.

The **show ip bgp dampened-paths** command is used to list all networks that are currently suppressed because of dampening.

The **show ip bgp prefix** command displays detailed information among other paths about suppressed prefixes. These prefixes are marked with the words “suppressed due to dampening.”



Monitoring Route Dampening—Example (Cont.)

Penalty goes below suppresslimit but not below reuselimit.

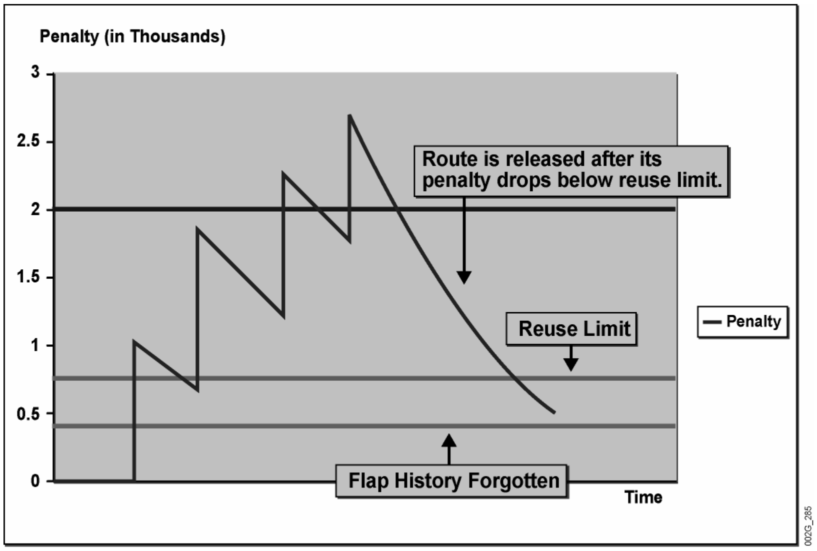
Another flap increases penalty.

Suppress Limit

Reuse Limit

The penalty of the route is decreased following an exponential curve. After a while, the penalty drops below the suppress limit, but the route is not yet released—the route is released only after the penalty drops further below the reuse limit. In the example here, the route flaps again, further increasing the penalty.

Monitoring Route Dampening—Example (Cont.)



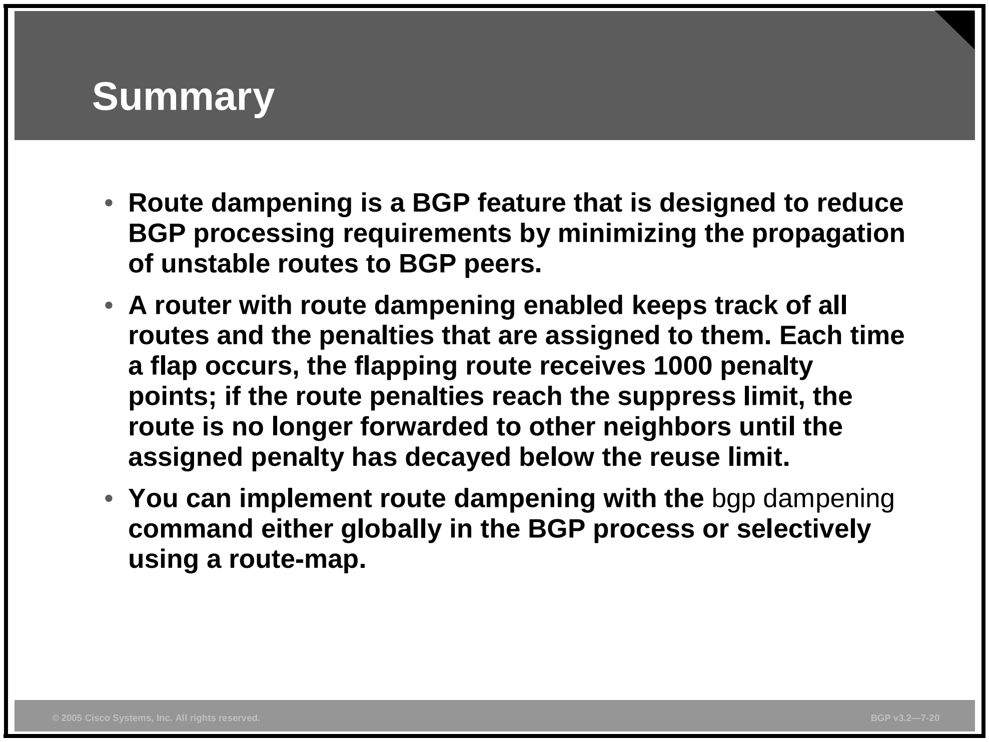
BGP v3.2—7-18

In the example here, the route has stabilized, and the penalty gradually drops below the reuse limit. At that point, the BGP router releases the route, and it can now be selected as the best BGP path. If the released route is selected by the router as the best BGP path, it is also propagated to BGP neighbors and used for data forwarding.

When the penalty that is associated with a route drops below half of the reuse limit, the penalty and the flap history that are associated with the route are cleared by the router.

Summary

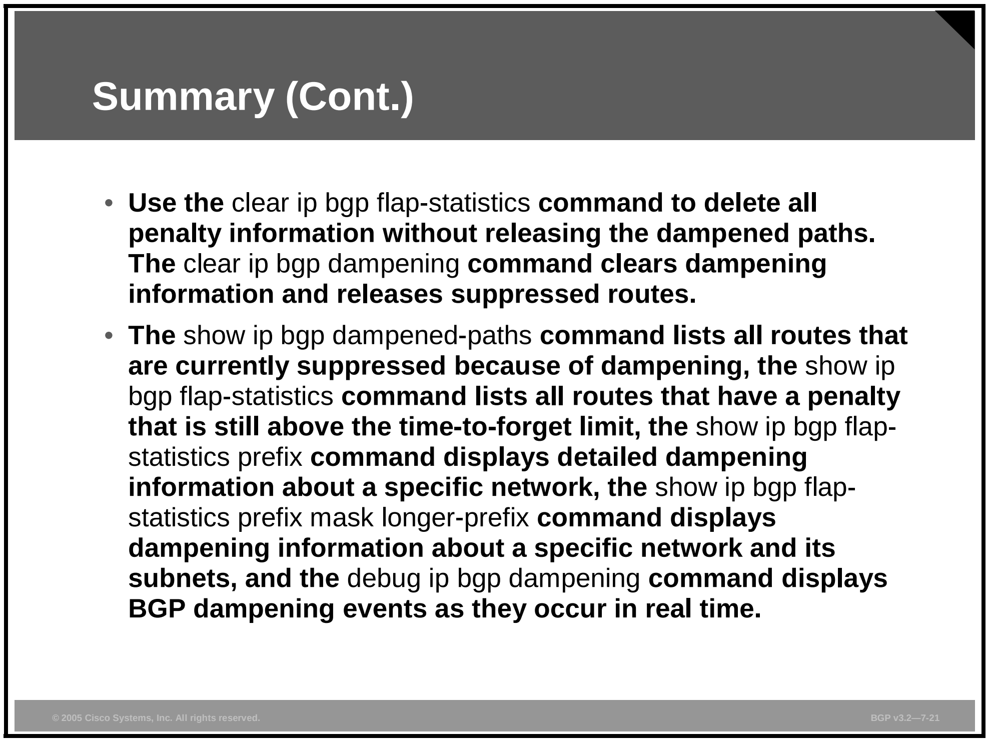
This topic summarizes the key points discussed in this lesson.



Summary

- **Route dampening is a BGP feature that is designed to reduce BGP processing requirements by minimizing the propagation of unstable routes to BGP peers.**
- **A router with route dampening enabled keeps track of all routes and the penalties that are assigned to them. Each time a flap occurs, the flapping route receives 1000 penalty points; if the route penalties reach the suppress limit, the route is no longer forwarded to other neighbors until the assigned penalty has decayed below the reuse limit.**
- **You can implement route dampening with the** bgp dampening **command either globally in the BGP process or selectively using a route-map.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-20



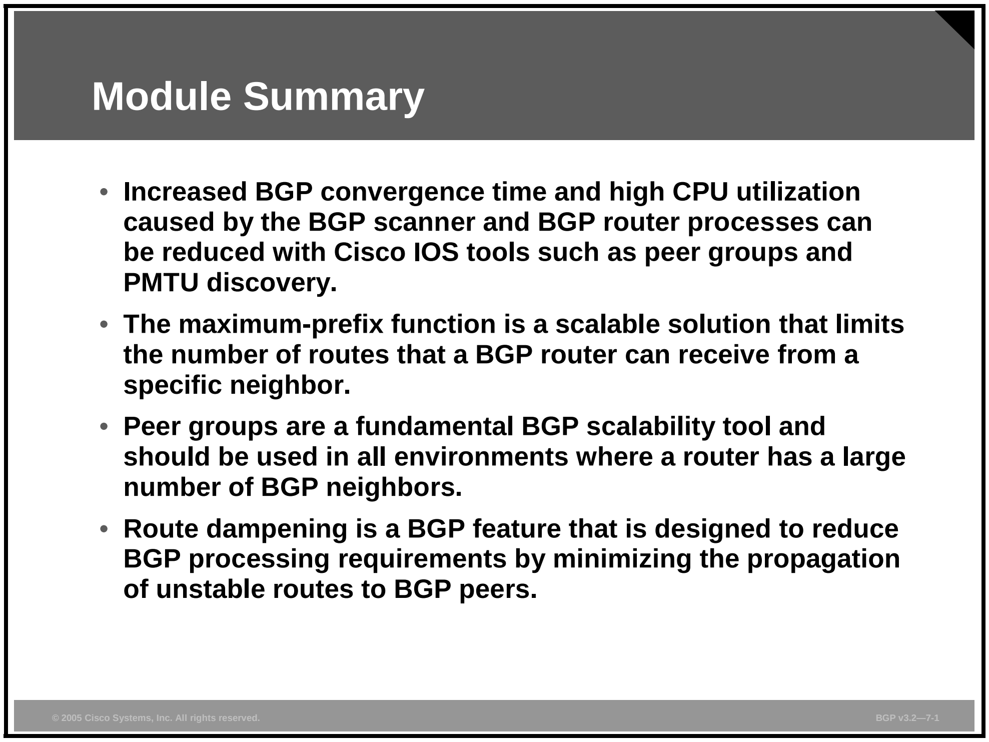
Summary (Cont.)

- **Use the** clear ip bgp flap-statistics **command to delete all penalty information without releasing the dampened paths. The** clear ip bgp dampening **command clears dampening information and releases suppressed routes.**
- **The** show ip bgp dampened-paths **command lists all routes that are currently suppressed because of dampening, the** show ip bgp flap-statistics **command lists all routes that have a penalty that is still above the time-to-forget limit, the** show ip bgp flap-statistics prefix **command displays detailed dampening information about a specific network, the** show ip bgp flap-statistics prefix mask longer-prefix **command displays dampening information about a specific network and its subnets, and the** debug ip bgp dampening **command displays BGP dampening events as they occur in real time.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-21

Module Summary

This topic summarizes the key points discussed in this module.



Module Summary

- **Increased BGP convergence time and high CPU utilization caused by the BGP scanner and BGP router processes can be reduced with Cisco IOS tools such as peer groups and PMTU discovery.**
- **The maximum-prefix function is a scalable solution that limits the number of routes that a BGP router can receive from a specific neighbor.**
- **Peer groups are a fundamental BGP scalability tool and should be used in all environments where a router has a large number of BGP neighbors.**
- **Route dampening is a BGP feature that is designed to reduce BGP processing requirements by minimizing the propagation of unstable routes to BGP peers.**

© 2005 Cisco Systems, Inc. All rights reserved. BGP v3.2—7-1

This module introduced advanced BGP configuration tools designed to improve BGP scalability and performance. The first lesson described various Cisco IOS performance improvements that have been designed to reduce BGP convergence time. The second lesson explained how to configure BGP to limit the number of prefixes received from a neighbor. The third lesson described how to use BGP peer groups to share common configuration parameters between multiple BGP peers. The final lesson described how to utilize route dampening to minimize the impact of unstable routes.

References

For additional information, refer to these resources:

- Cisco Systems, Inc. *Cisco IOS IP and IP Routing Command Reference, Release 12.1*. http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/ip_r/index.htm.
- Griffin, T. and Wilfong, G.T. *An Analysis of BGP Convergence Properties*. SIGCOMM 1999.
- Pei, D., Xiaoliang, Z., Wang, L., Massey, D. Mankin, A., Wu, S.F., and Zhang, L. *Improving BGP Convergence Through Consistency Assertions*. 2002.
- Cisco Systems, Inc. *BGP Restart Session After Maximum-Prefix Limit*. http://www.cisco.com/en/US/partner/products/ps6566/products_feature_guide09186a00801545d5.html .
- Cisco Systems, Inc. *BGP Peer Groups*. http://www.cisco.com/warp/public/459/29.html.
- Cisco Systems, Inc. *Configuring BGP*. http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/12cgcr/np1_c/1cprt1/1cbgp.htm#xtocid45.
- Cisco Systems, Inc. *BGP Case Studies*. "BGP Case Studies 4." http://www.cisco.com/warp/public/459/bgp-toc.html#flapdampen.
- Cisco Systems, Inc. *Configuring BGP*. http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/ip_c/ipcprt2/1cdbgp.htm#xtocid61.

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) What are three characteristics of a converged BGP network? (Choose three.) (Source: Improving BGP Convergence)

A) The input queue and output queue for all peers is 0.
B) All routes in the BGP table have been installed in the routing table.
C) The table version for all peers equals the table version of the BGP table.
D) All routes have been accepted.

Q2) Which two of the following modifications result in improved BGP convergence? (Choose two.) (Source: Improving BGP Convergence)

A) increasing the default value of BGP hold time
B) lowering the default value of BGP scan time
C) increasing the default value of the neighbor advertisement intervals
D) lowering the default value of the neighbor advertisement intervals

Q3) What is the main task of the BGP scanner process? (Source: Improving BGP Convergence)

A) sends routing updates to BGP neighbors
B) walks the BGP table for routes to enter into the IP routing table
C) confirms the reachability of BGP next hops
D) scans the router configuration to establish and maintain BGP neighbors

Q4) One of your BGP core routers is experiencing periodic slow responses to ping packets that are being directed to it from the network management console. The router has just been configured to receive full Internet routes, and you suspect that the BGP router process is causing CPU utilization issues in the core router. Which two router commands should you use to confirm your suspicion? (Choose two.) (Source: Improving BGP Convergence)

A) **show ip route**
B) **show ip bgp summary**
C) **show process cpu**
D) **show memory**

Q5) The output of a **show interfaces fastethernet 0/0** command follows:

```
Fast Ethernet0 is up, line protocol is up
  Hardware is DEC21140, address is 0000.0c0c.1111 (bia 0002.eaa3.5a60)
  Internet address is 112.64.101.17 255.255.255.240
  MTU 1460 bytes, BW 100000 Kbit, DLY 100 usec, rely 255/255, load 200/255
  Encapsulation ARPA, loopback not set, keepalive not set, hdx, 100BaseTX
  ARP type: ARPA, ARP Timeout 4:00:00
  Last input never, output 0:00:16, output hang 0:28:01
  Last clearing of "show interface" counters 0:20:05
  Output queue 25/40, 0 drops; input queue 50/500, 1470 drops
  5 minute input rate 21666400 bits/sec, 1855 packets/sec
  5 minute output rate 72221 bits/sec, 618 packets/sec
```

How has the interface been modified to improve BGP convergence? (Source: Improving BGP Convergence)

A) The output queue has been decreased to expedite packet forwarding out the Fast Ethernet interface.
B) The drop threshold of the input queue has been set to begin randomly discarding packets after the queue reaches 50 packets deep.
C) PMTU discovery has been enabled by setting the interface MSS to 1460 bytes.
D) The size of the input queue has been increased to support up to 500 incoming packets.

Q6) Refer to the following Cisco IOS router output:

```
router# show ip bgp summary
BGP router identifier 172.16.0.4, local AS number 1
BGP table version is 16, main routing table version 16
20 network entries and 20 paths using 2826 bytes of memory
8 BGP path attribute entries using 480 bytes of memory
7 BGP AS-PATH entries using 168 bytes of memory
3 BGP community entries using 72 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
3 BGP filter-list cache entries using 36 bytes of memory
BGP activity 20/0 prefixes, 24/4 paths, scan interval 120 secs

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
172.16.0.1      4     1      30      30       16    0    0 00:23:13        5
172.16.0.2      4     1      33      30       16    0    0 00:23:15        5
172.16.0.3      4     1      27      30       16    0    0 00:23:14        5
192.168.21.99   4    99      31      35       16    0    0 00:23:04        5
```

Which two parameters would indicate that the BGP network has converged? (Choose two.) (Source: Improving BGP Convergence)

A) The TblVer for all neighbors is 16.
B) V is set to 4 for all neighbors.
C) The InQ and OutQ for all neighbors is 0.
D) All neighbors are in the Established state and have the same PfxRcd value.

Q7) Using the command output from item 6, identify how frequently the BGP scanner process will run on the router. (Source: Improving BGP Convergence)

A) By default, the process will run every 60 seconds.
B) The process has run 16 times and will run again when the next BGP update arrives.
C) The process will run on this router every 120 seconds.
D) It cannot be determined from this output.

Q8) What are two potential issues that are caused by modifying the default scan time and advertisement interval on a BGP router? (Choose two.) (Source: Improving BGP Convergence)

A) Router CPU resources can be exhausted.
B) Router memory resources can be depleted.
C) Routing loops are more likely.
D) BGP could converge faster than the IGP and cause network black holes.

Q9) Which two of the following statements about the EIGRP Nonstop Forwarding Awareness feature are accurate? (Choose two.) (Source: Improving BGP Convergence)

A) NSF-aware routers are sometimes incompatible with non-NSF-aware or -capable neighbors in an EIGRP network.
B) EIGRP NSF awareness must be enabled by an administrator.
C) The deployment of EIGRP NSF awareness can minimize the effects of well-known failure conditions.
D) NSF awareness allows an NSF-aware router to assist NSF-capable and NSF-aware neighbors to continue forwarding packets during a switchover operation or during a well-known failure condition.

Q10) What are three reasons to limit the number of BGP prefixes that are received from a neighbor? (Choose three.) (Source: Limiting the Number of Prefixes Received from a BGP Neighbor)

A) to prevent denial-of-service attacks
B) to protect against incorrect router configuration on the neighbor side
C) to prevent redundant routing information from being loaded into the BGP table
D) to avoid overloading router memory and CPU resources

Q11) In which two situations would a directly connected BGP neighbor stay in the Idle state? (Choose two.) (Source: Limiting the Number of Prefixes Received from a BGP Neighbor)

A) The neighbor has exceeded the maximum number of allowed prefixes.
B) The maximum-prefix threshold has been reached.
C) The **restart** option has not been specified with the **maximum-prefix** command.
D) The neighbor is more than one hop away.

Q12) Which two of the following characteristics accurately describe the **show ip bgp neighbors** command? (Choose two.) (Source: Limiting the Number of Prefixes Received from a BGP Neighbor)

A) for neighbors with the maximum-prefix function configured, displays the maximum number of prefixes and the warning threshold
B) for neighbors exceeding the maximum number of prefixes, displays the reason that the BGP session is idle
C) for neighbors with unstable routes, displays the feasible successor for those routes
D) for neighbors in confederations, displays the route reflector status of those neighbors

Q13) What is the need for BGP peer groups? (Source: Implementing BGP Peer Groups)

A) can be used to configure the same set of parameters for a number of BGP neighbors in a common template
B) can be used to allow anonymous BGP neighbors
C) allow EBGP peers to be configured with the same AS number and parameters
D) can be used to hide the identity of BGP peers from external neighbors

Q14) Which of the following statements about the benefit of BGP peer groups is accurate? (Source: Implementing BGP Peer Groups)

A) With BGP peer groups, all of the router CPU utilization that is imposed by BGP update generation is significantly reduced.
B) With BGP peer groups, some of the router CPU utilization that is imposed by BGP update generation is significantly reduced.
C) Network administrators should use peer groups to make smaller networks more productive.
D) With BGP peer groups, neighbor relationships are automatically created.

Q15) What are two limitations of BGP peer groups on Cisco routers? (Choose two.) (Source: Implementing BGP Peer Groups)

A) EBGP and IBGP neighbors cannot be members of the same peer group.
B) All routers in the peer group must belong to the same AS.
C) Peer group members cannot contain different outbound filtering mechanisms.
D) Peer group members must have the same inbound filtering mechanisms.

Q16) Which two of the following characteristics accurately describe BGP peer groups? (Choose two.) (Source: Implementing BGP Peer Groups)

A) A BGP peer group creates a neighbor parameter template.
B) When actual neighboring routers are assigned to the peer group on a router, all of the attributes that are configured for the peer group are applied to selected peer group members.
C) One of the configurable parameters includes community propagation.
D) Individual parameters specified in a peer group cannot be overridden on a neighbor-by-neighbor basis.

Q17) Which two of the following characteristics accurately describe the function of the BGP Dynamic Update Peer-Groups feature? (Choose two.) (Source: Implementing BGP Peer Groups)

A) does not provide the operator with time to change the configuration if a mistake is made
B) separates BGP update generation from peer-group configuration
C) does not require any configuration by the network operator
D) requires minimal configuration by the network operator

Q18) Which two of the following statements accurately describe the function of the BGP Configuration Using Peer Templates feature? (Choose two.) (Source: Implementing BGP Peer Groups)

A) Network operators must still configure some peer groups in BGP, even if using the BGP Configuration Using Peer Templates feature.
B) Peer templates overcome all limitations of peer groups.
C) Peer templates improve the flexibility and enhance the capability of neighbor configuration.
D) You can chain together peer template configurations to create simple or complex configurations.

Q19) What are three steps that are required to properly configure BGP peer groups on Cisco routers? (Choose three.) (Source: Implementing BGP Peer Groups)

A) specify parameters for the BGP peer group
B) create a BGP peer group
C) enable the peer group by clearing the BGP session
D) assign a neighbor into the peer group

Q20) Which command do you use to display the summary status of all neighbors in a peer group? (Source: Implementing BGP Peer Groups)

A) **show ip bgp**
B) **show peer-group summary**
C) **show ip bgp neighbor**
D) **show ip bgp peer-group summary**

Q21) Which two descriptions of the purpose of BGP route dampening are accurate? (Choose two.) (Source: Using BGP Route Dampening)

A) a tool designed to help minimize the number of BGP updates
B) suppresses routes that occasionally flap
C) designed to reduce router processing load caused by unstable routes
D) prevents sustained routing oscillation, with some affect on other well-behaved routes

Q22) Which two mechanisms are built into BGP to make it more scalable by reducing the route-processing requirements of BGP routers? (Choose two.) (Source: Using BGP Route Dampening)

A) split horizon
B) route dampening
C) synchronization
D) per-neighbor update timers

Q23) What are two things that happen to an EBGP route that has become unreachable when BGP route dampening is used? (Choose two.) (Source: Using BGP Route Dampening)

A) It is removed from the IP routing table.
B) It is removed from the BGP table.
C) It remains in the IP routing table as long as its penalty remains greater than 50 percent of the reuse limit.
D) It is kept in the BGP table and marked as a history entry

Q24) What are two methods of enabling route dampening on a Cisco router? (Choose two.) (Source: Using BGP Route Dampening)

A) globally, by enabling route dampening in global router configuration mode
B) globally, by enabling route dampening under the BGP routing process
C) on specific routes by enabling route dampening on a specific interface
D) by using a route-map in the BGP process to apply route dampening to specific routes

Q25) Which two things could happen to a BGP route that is penalized above the reuse limit but has an assigned penalty that is under the suppress limit? (Choose two.) (Source: Using BGP Route Dampening)

A) The route is suppressed from BGP updates if it is reachable.
B) The route is marked as a history entry in the BGP table.
C) The route is withdrawn from the IP routing table.
D) The route continues to be advertised.

Module Self-Check Answer Key

Q1)	A, C, D
Q2)	B, D
Q3)	C
Q4)	B, C
Q5)	D
Q6)	A, C
Q7)	C
Q8)	A, B
Q9)	C,D
Q10)	A, B, D
Q11)	A, C
Q12)	A, B
Q13)	A
Q14)	B
Q15)	A, C
Q16)	A, C
Q17)	B, C
Q18)	C, D
Q19)	A, B, D
Q20)	D
Q21)	A, C
Q22)	B, D
Q23)	A, D
Q24)	B, D
Q25)	B, C