

Configuring BGP on Cisco Routers

Student Guide

Volume 2

Version 4.0

Part Number: 97-3540-01



Americas Headquarters

Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters

Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters

Cisco Systems International BV
Amsterdam,
The Netherlands

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

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED "AS IS" AND AS SUCH MAY INCLUDE TYPOGRAPHICAL, GRAPHICS, OR FORMATTING ERRORS. 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.

Table of Contents

Module 4: Route Selection Using Attributes	1
<u>Lesson 1: Influencing BGP Route Selection with Weights</u>	<u>3</u>
BGP Route Selection Criteria	4
Influencing BGP Route Selection	5
Configuring Per-Nearbor Weights.....	7
Discovery 11: Configure Per-Nearbor Weights	8
Changing Weights with Route Maps	17
BGP Route Selection and Filtering Tools Summary	19
Summary.....	20
<u>Lesson 2: Setting BGP Local Preference</u>	<u>21</u>
Consistent Route Selection Within the AS	22
BGP Local Preference	25
Configuring Default Local Preference	28
Monitoring Local Preference	29
Discovery 12: Configure and Monitor Local Preference.....	30
Configuring Local Preference with Route Maps	37
Discovery 13: Configure Local Preference Using Route Maps.....	38
Summary.....	45
<u>Lesson 3: Using AS Path Prepending</u>	<u>47</u>
Return Path Selection in a Multihomed AS	48
AS Path Prepending	52
AS Path Prepending Design Considerations.....	55
Discovery 14: Configure AS Path Prepending.....	57
BGP Hide Local-Autonomous System	68
Summary.....	70
<u>Lesson 4: Understanding BGP Multi-Exit Discriminators</u>	<u>71</u>
Selecting the Proper Return Path	72
MED Propagation in a BGP Network.....	74
Changing MED.....	75
Discovery 15: Configure MED	77
Troubleshooting the MED.....	85
Advanced MED Configuration	88
Summary.....	91
<u>Lesson 5: Addressing BGP Communities</u>	<u>93</u>
Selecting the Proper Return Path	94
BGP Communities Overview	95
Using Communities.....	98
Configuring BGP Communities.....	100
Discovery 16: Configure Local Preference Using the Communities	106
BGP Named Community Lists	117
BGP Cost Community	121
BGP Link Bandwidth Feature	123
BGP Support for Sequenced Entries in Extended Community Lists	127
Summary.....	130

Lesson 6: Module Summary	131
Lesson 7: Module Self-Check	133
Module Self-Check Answers.....	145
Answer Key.....	145
Module 5: Customer-to-Provider Connectivity with BGP.....	149
Lesson 1: Understanding Customer-to-Provider Connectivity Requirements.....	151
Customer-to-Provider Connectivity Types	152
Customer Redundant Connectivity.....	153
Customer-to-Provider Routing Schemes.....	157
Customer Routing Schemes	158
Customer Addressing Requirements	162
Customer AS Number Allocation	164
Summary.....	167
Lesson 2: Implementing Customer Connectivity Using Static Routing	169
When to Use Static Routing?.....	170
Characteristics of Static Routing.....	171
Designing Static Route Propagation in a Service Provider Network	174
BGP Backup with Static Routes	179
Floating Static Routes with BGP	183
Load Sharing with Static Routes.....	186
Summary.....	190
Lesson 3: Connecting a Customer to a Single Service Provider	191
BGP Configuration on Customer Routers	193
Conditional BGP Advertising in Customer Networks	195
BGP Configuration on Service Provider Routers.....	198
Removing a Private AS Numbers	201
BGP Support for Dual AS Configuration for Network AS Migrations.....	204
Backup Solutions with BGP	208
Load Sharing	210
Load Sharing with BGP Multipath.....	211
Load Sharing with EBGP Multihop	212
Summary.....	214
Lesson 4: Connecting a Multihomed Customer to Multiple Service Providers.....	215
BGP Configuration for Multihomed Customers	217
Multihomed Customer Address Space Selection	219
Multihomed Customer AS Number Selection	220
AS Number Translation	222
Primary and Backup Link Selection	224
BGP Incoming Link Selection	225
Load Sharing with Multiple Providers.....	229
Summary.....	231
Lesson 5: Module Summary	233
References.....	233
Lesson 6: Module Self-Check	235
Module Self-Check Answers.....	245
Answer Key.....	245
Module 6: Scaling Service Provider Networks.....	247
Lesson 1: Scaling IGP and BGP in Service Provider Networks.....	249
Common Service Provider Network.....	250

Route Propagation in Service Provider Networks	253
Scaling Service Provider Routing Protocols	257
Scaling Service Provider Addressing	260
Summary	261
Lesson 2: Introducing and Designing Route Reflectors	263
IBGP Scalability Issues in a Transit AS	264
Route Reflector Split-Horizon Rules	266
Redundant Route Reflectors	270
Route Reflector Clusters	272
Additional Route Reflector Loop-Prevention Mechanisms	274
Network Design with Route Reflectors	275
Potential Network Issues	277
Hierarchical Route Reflectors	278
Summary	280
Lesson 3: Configuring and Monitoring Route Reflectors	281
Route Reflector Backbone Migration	282
Configuring Route Reflectors	283
Discovery 17: Configure Route Reflector	286
Summary	298
Lesson 4: Module Summary	299
Lesson 5: Module Self-Check	301
Module Self-Check Answers	307
Answer Key	307
Module 7: Optimizing BGP Scalability	309
Lesson 1: Improving BGP Convergence	311
BGP Convergence	313
BGP Processes	314
CPU Effects of BGP Processes	315
Improving BGP Convergence	317
PMTU Discovery	320
Increasing Input Queue Depth	322
BGP Prefix Independent Convergence	325
Bidirectional Forwarding Detection for BGP	328
BGP Nonstop Forwarding Awareness	331
BGP Scan Time	336
BGP Advertisement Interval	339
BGP Keepalive and Hold-Down Timers	342
Summary	345
Lesson 2: Limiting the Number of Prefixes Received from a BGP Neighbor	347
BGP Route Limiting	348
Configuring the BGP Route Limiting	350
Discovery 18: Configure BGP Route Limiting	352
Summary	360
Lesson 3: Implementing BGP Peer Groups	361
BGP Peer Groups Overview	362
BGP Peer Groups as a Performance Tool	367
BGP Peer Group Limitations	369
Configuring BGP Peer Groups	370
Discovery 19: Configure BGP Peer Groups	372

BGP Peer Group Configuration Examples	381
BGP Dynamic Update Peer Groups Feature	383
BGP Peer Templates Overview	388
BGP Peer Templates Inheritance	390
BGP Peer Templates Configuration	392
Summary	397
<u>Lesson 4: Using BGP Route Dampening</u>	<u>399</u>
BGP Route Dampening	400
BGP Route Dampening Operation	402
Configuring BGP Route Dampening	404
Discovery 20: Configure BGP Route Dampening	407
Summary	421
<u>Lesson 5: Module Summary</u>	<u>423</u>
References	423
<u>Lesson 6: Module Self-Check</u>	<u>425</u>
<u>Module Self-Check Answers</u>	<u>433</u>
Answer Key	433
<u>Glossary</u>	<u>435</u>

Module 4: Route Selection Using Attributes

Introduction

Routes that are learned via [BGP](#) have properties that are associated with them. They 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. In this module, you will learn the role of BGP attributes, and how their presence influences route selection in BGP.

Understanding how BGP attributes, including weight, local preference, [AS](#) path prepending, [MED](#), and BGP communities, influence route selection is required for the design of robust networks.

Upon completing this module, you will be able to:

- Define how to use BGP weight attribute to influence the route selection
- Define how to use BGP local preference attribute to influence the route selection
- Define how to use BGP AS path prepending to influence the return path that is selected by the neighboring autonomous systems
- Define how to use MED attribute to influence the route selection
- Define how to use BGP community attributes to influence the route selection

Lesson 1: Influencing BGP Route Selection with Weights

Overview

When connections to multiple providers are required, it is important that [BGP](#) selects the optimum route for traffic to use. The optimum, or best, route may not be what you intended during the design, based on the design criteria, administrative policies, or corporate mandate. Fortunately, BGP provides many tools for you to influence route selection. One of these tools is the weight attribute.

In this lesson, you will learn 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. You will also learn how to monitor the BGP table to verify correct weight configuration and properly influence path selection.

Upon completing this lesson, you will be able to:

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

BGP Route Selection Criteria

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, the next criterion, local preference, is checked.

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

© 2015 Cisco and/or its affiliates. All rights reserved.

1

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

Influencing BGP Route Selection

The weight attribute is local to a single router only. [BGP](#) protocol never propagates the weight value, and this value constitutes a routing policy local to the router. You have two different ways how to assign the weight attribute to a route.

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
 - Complex criteria with route maps

© 2015 Cisco and/or its affiliates. All rights reserved.

2

The weight attribute can be assigned to a route using one of the following:

- Assign a default weight value to all of the routes that are received from a specific neighbor. This weight value indicates that the neighbor is preferred over the other neighbors.
- Apply a route map on incoming routes from a neighbor to select some routes and assign weight values to them. Remember that a route map also acts as a filter and will silently drop the routes that the route map does not permit in any statement.

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 can use various selection criteria, such as a network number or AS path to select routes. You can alter some attributes of the selected routes. 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.

Note If you want the routing policy to be exchanged within all the routers in the [AS](#), assign *local preference* attribute to a route. This attribute is carried with the route on all internal BGP sessions. In this situation all other BGP-speaking routers within the 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.

Configuring Per-Neighbor Weights

If you want all routes from the [BGP](#) neighbor to get the specified weight, you have to configure 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.

© 2015 Cisco and/or its affiliates. All rights reserved.3

To assign a weight to a neighbor connection, use the **neighbor weight** router configuration command. To remove a weight assignment, use the **no** form of this command.

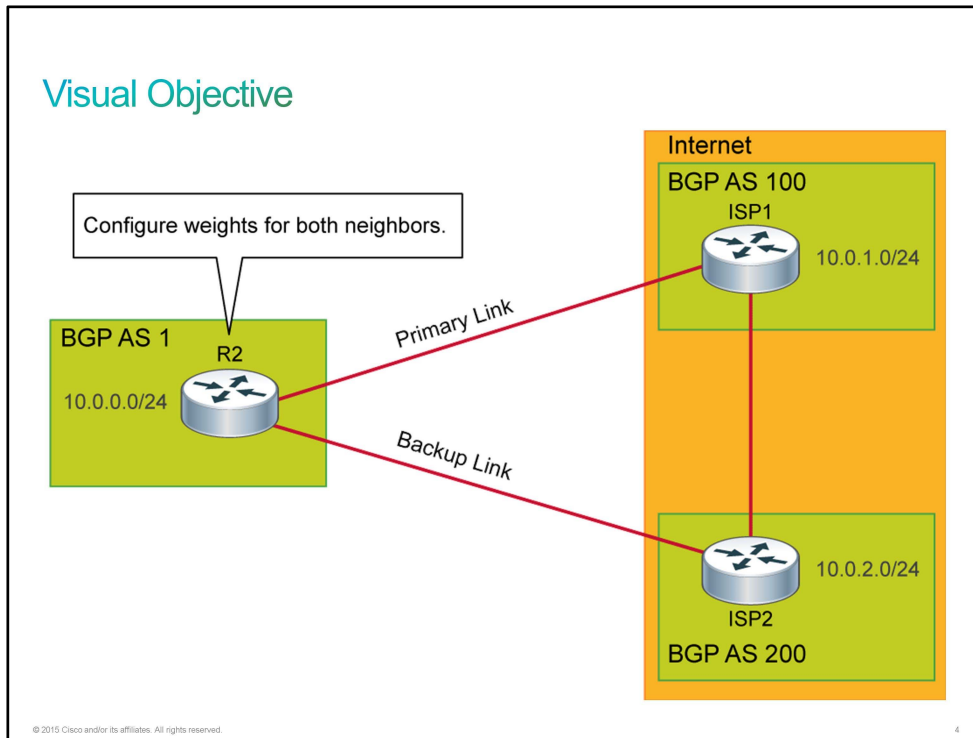
All routes that a router receives from the neighbor after the configuration line is in place get the weight value assigned. To make sure that all routes from the neighbor receive the new weight value, you can restart the BGP session using **clear ip bgp** command, forcing the neighbor to resend all routes.

Note	If no weight value is specified, the default value for routes learned through another BGP peer is 0 and the default value for routes sourced by the local router is 32768.
-------------	--

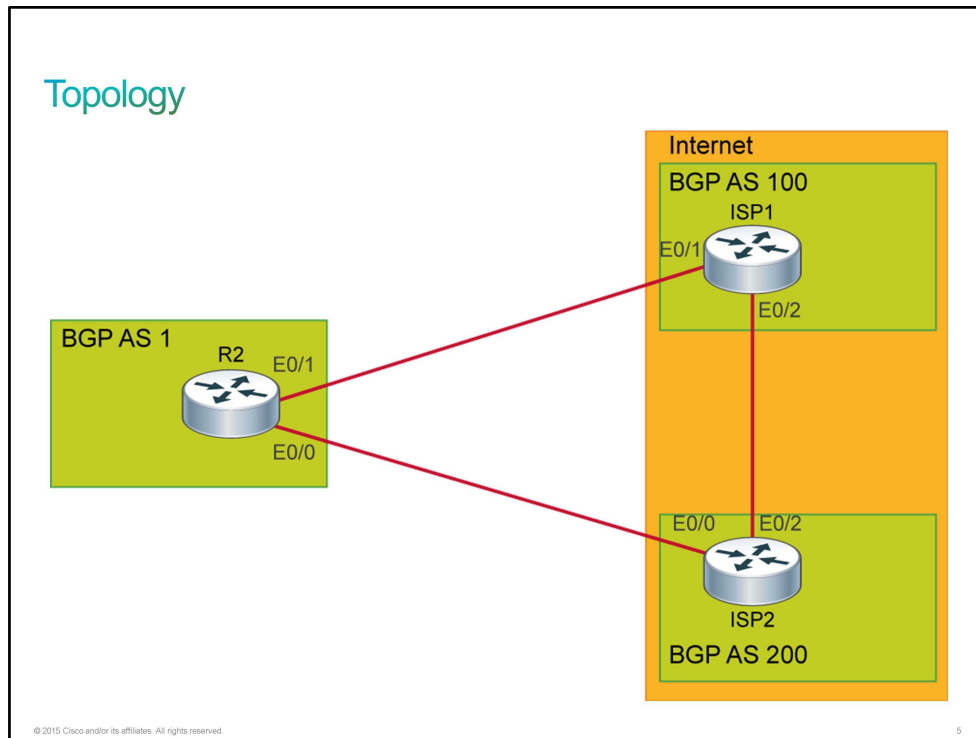
Discovery 11: Configure Per-Neighbor Weights

Overview

Through this discovery, you will learn how to configure per-neighbor weights and how to monitor [BGP](#) route selection. You will configure a multihomed customer's router, R2, to use the link to the primary service provider, ISP1, for all destinations.



Topology



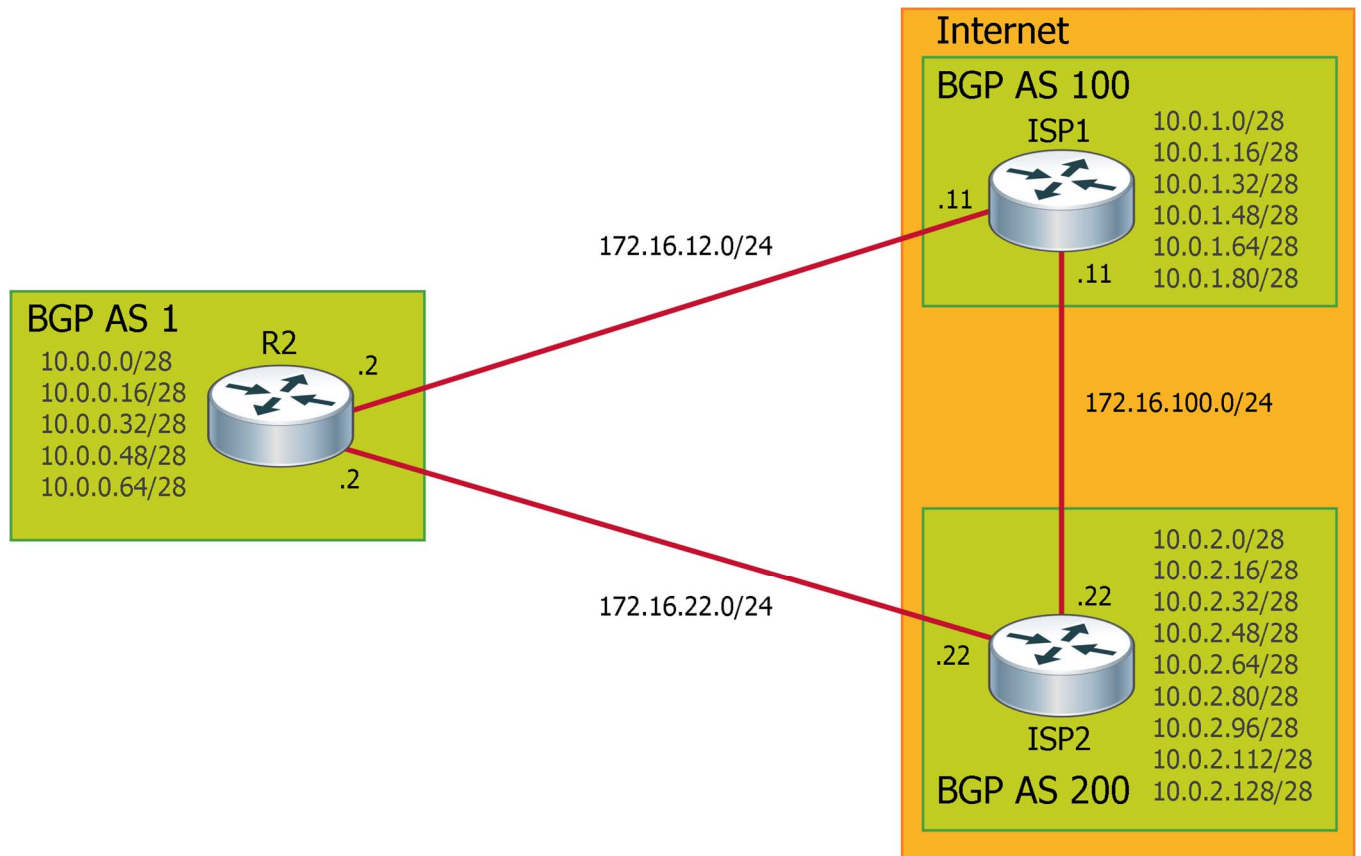
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/1	172.16.12.2/24	Connection to ISP1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/1	172.16.12.11/24	Connection to R2
ISP1	Ethernet 0/2	172.16.100.11/24	Connection to ISP2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Ethernet 0/2	172.16.100.22/24	Connection to ISP1
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in BGP are preconfigured as shown below:



BGP is also preconfigured as EBGP:

- R2 to ISP1
- R2 to ISP2
- ISP1 to ISP2

Configure Per-Neighbor Weights

Discovery Steps

Step 1 On the R2 router, verify the initial state of default weights for all received destinations.

To verify weights, use the **show ip bgp** command.

R2# **show ip bgp**

BGP table version is 65, local router ID is 10.0.0.65

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
x best-external, a additional-path, c RIB-compressed,

Origin codes: i - IGP, e - EGP, ? - incomplete

RPKI validation codes: V valid, I invalid, N Not found

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	0.0.0.0	0		32768	i
*>	10.0.0.16/28	0.0.0.0	0		32768	i
*>	10.0.0.32/28	0.0.0.0	0		32768	i
*>	10.0.0.48/28	0.0.0.0	0		32768	i
*>	10.0.0.64/28	0.0.0.0	0		32768	i
*	10.0.1.0/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
*	10.0.1.16/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
*	10.0.1.32/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
*	10.0.1.48/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
*	10.0.1.64/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.1.80/28	172.16.22.22			0	200 100 i
*>		172.16.12.11	0		0	100 i
*	10.0.2.0/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.16/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.32/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.48/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.64/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.80/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.96/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.112/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i
*	10.0.2.128/28	172.16.12.11			0	100 200 i
*>		172.16.22.22	0		0	200 i

R2 has received routes to the following networks:

- 10.0.0.0/24 networks are injected into the BGP table by this router. By default, locally sourced routes are assigned a weight of 32768.
- 10.0.1.0/24 networks are received from ISP1 and ISP2. All routes have default weight value of 0 assigned.
- 10.0.2.0/24 networks are received from ISP1 and ISP2. All routes have default weight value of 0 assigned.

Step 2 On the R2 router, make sure that route received from the ISP1 router will be preferred over routes received from the ISP2 router.

For routes that are received from ISP1, configure a weight of 150 and for routes that are received from ISP2 configure a weight of 100.

Note that BGP routes with higher weight are preferred.

```
R2(config)# router bgp 1
R2(config-router)# neighbor 172.16.12.11 weight 150
R2(config-router)# neighbor 172.16.22.22 weight 100
```

The weight is configured on R2 on both BGP sessions. A higher weight value is given to the routes that are received from the primary [ISP](#), ISP1, compared to the routes that are received from the backup ISP, ISP2.

Any time that the R2 receives routing information about the same IP network number from both ISPs, the router compares the weights that are assigned to the routes. The routes that are received from the ISP1 will always win this comparison.

Step 3 Re-establish BGP sessions on the R2 router to enforce new weights.

To re-establish BGP session, use the **clear ip bgp** command.

```
R2# clear ip bgp *
```

The * parameter of the command specifies to clear all peers.

Monitoring BGP Route Selection and Weights

Monitoring BGP Route Selection and Weights

```
router> show ip bgp
```

- Displays all BGP routes. Best routes marked with ">"
- With every displayed route, weight is associated.

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

- Displays detailed information about all paths for a single prefix

© 2015 Cisco and/or its affiliates. All rights reserved.6

Without any argument, the **show ip bgp** command displays the entire BGP table. Each route is displayed in one line. 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, [MED](#), local preference, weight, [AS](#) path, and origin code are displayed. 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. The greater-than ("**>**") character indicates the routes that are selected as the best.

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.

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 the router has selected as the best. There are then 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

Step 4 On the R2 router, verify received routes to destination networks and their weights.

To verify received routes, use the **show ip bgp** command.

```

R2# show ip bgp
BGP table version is 20, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	0.0.0.0	0		32768	i
*>	10.0.0.16/28	0.0.0.0	0		32768	i
*>	10.0.0.32/28	0.0.0.0	0		32768	i
*>	10.0.0.48/28	0.0.0.0	0		32768	i
*>	10.0.0.64/28	0.0.0.0	0		32768	i
*	10.0.1.0/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
*	10.0.1.16/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
*	10.0.1.32/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
*	10.0.1.48/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
*	10.0.1.64/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.1.80/28	172.16.22.22			100	200 100 i
*>		172.16.12.11	0		150	100 i
*	10.0.2.0/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.16/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.32/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.48/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.64/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.80/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.96/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.112/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i
*	10.0.2.128/28	172.16.22.22	0		100	200 i
*>		172.16.12.11			150	100 200 i

R2 has received routes to 15 different networks outside of its own AS.

- 10.0.1.0/24 networks are received from ISP1 and ISP2.
- 10.0.2.0/24 networks are received from ISP1 and ISP2.

When the routes were received from the ISP1, the weight value 150 was assigned to each of the routes. When the routes were received from the ISP2, the weight value 100 was assigned to each of the routes. The R2 then makes the route selection—it has two alternative paths for each destination and it selects the path via the ISP1 as the best. It makes this selection based on the weight parameter, regardless of other BGP attributes, such as AS path length.

Consequently, the 10.0.2.0/24 networks that are directly connected to the ISP2 will also be reached via the ISP1. Although they are actually networks in the AS of the ISP2, AS 200.

Step 5 On the R2 router, verify detailed information about the 10.0.2.0/28 network.

To verify detailed information, use the **show ip bgp** command with a network number as an argument.

```
R2# show ip bgp 10.0.2.0/28
BGP routing table entry for 10.0.2.0/28, version 8
Paths: (2 available, best #2, table default)
  Advertised to update-groups:
    2
  Refresh Epoch 2
  200
    172.16.22.22 from 172.16.22.22 (10.0.2.129)
      Origin IGP, metric 0, localpref 100, weight 100, valid, external
  Refresh Epoch 2
  100 200
    172.16.12.11 from 172.16.12.11 (10.0.1.81)
      Origin IGP, localpref 100, weight 150, valid, external, best
```

There are two alternatives to reach network 10.0.2.0/28. Each of them is received from different neighbors, one in AS 100 and another in AS 200. The network 10.0.2.0/28 is created in AS 200.

The route selection mechanism has selected the second route that is listed as the best. It was chosen because the weight value is higher.

Changing Weights with Route Maps

The route map is a powerful tool to select and alter routing information. When you apply a route map to incoming information from a [BGP](#) neighbor, each received update is examined as it passes through the route map. The sequence number of the statements specify the order of execution.

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
```

© 2015 Cisco and/or its affiliates. All rights reserved.7

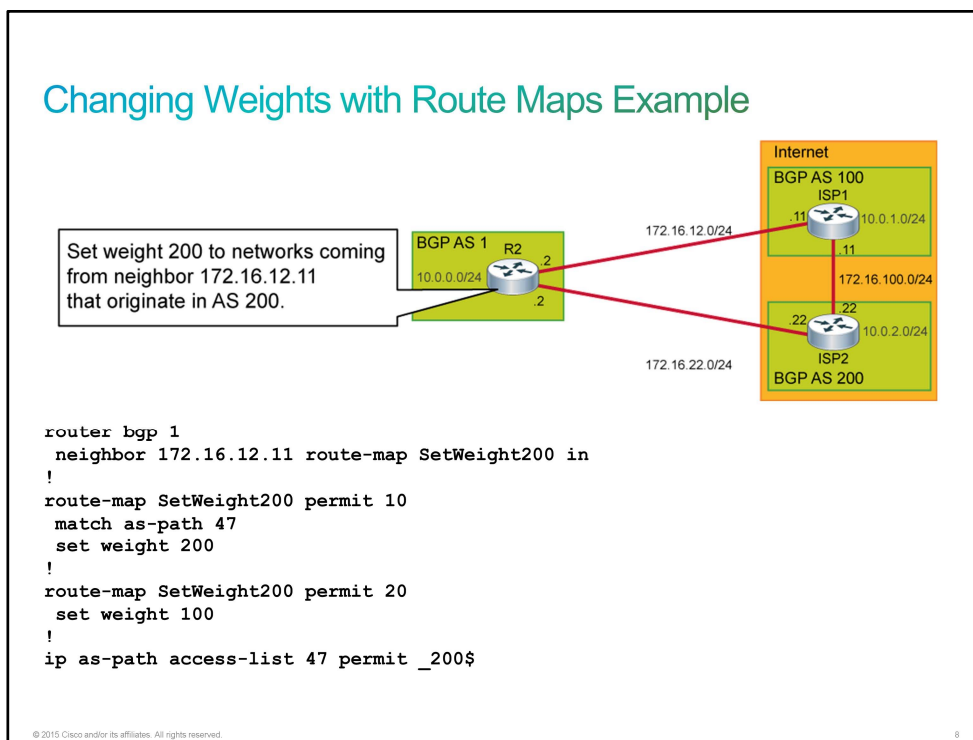
The first statement in the route map for which all the match clauses indicate a match is the one that is used. If the route map says "permit," the set clauses are applied to the route, the route is accepted, and the weight is changed.

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, the route map statement is used and its result, permit, or deny, applied.

If a received route is not matched by any of the route map statements, the end of the route map is reached. The route map logic has an "implicit deny" rule, which means that if no statement selects a route, the route is discarded.

If you do not desire the "implicit deny" rule, add an "explicit permit all" at the end of the route map to 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 this statement does not alter any attributes.

Changing Weights with Route Maps Example



All received routes from ISP1 neighbor have their AS paths checked against the AS path access list 47. AS path access list 47, as referenced by route map statement number 10, permits those routes with an AS path that indicates that they originated in AS 200. Routes that the route map statement number 10 in the SetWeight200 route map permits and selects, will have their weight set to 200. The set clause in the route map indicates the new weight value..

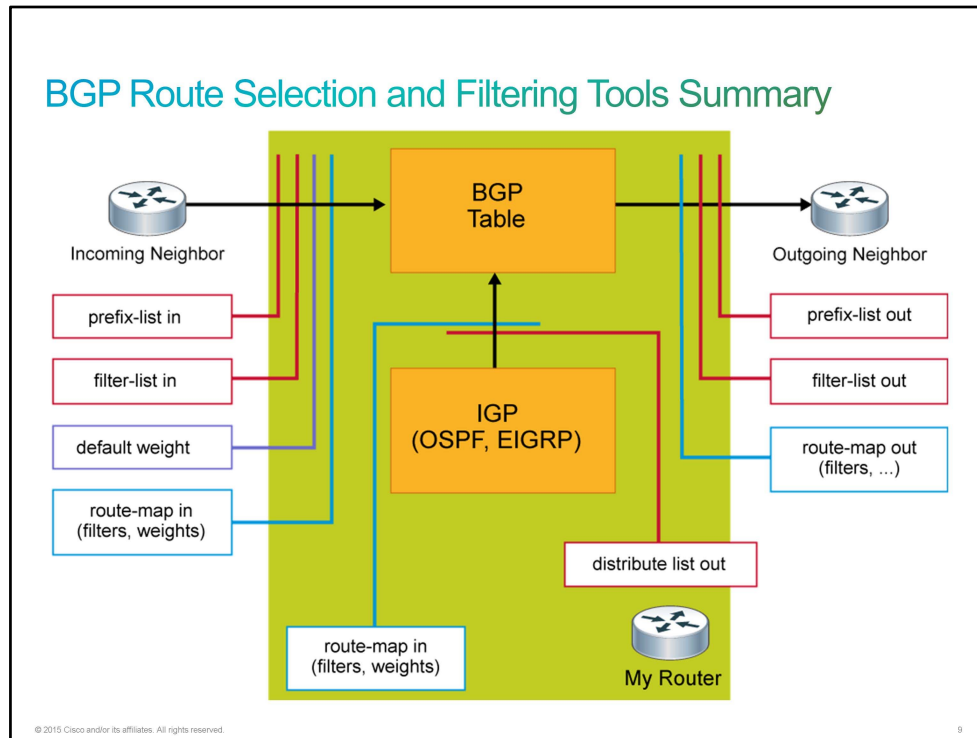
Route map statement number 20 then tests the routes that are not originated in AS 200 (routes that the AS path access list 47 does not permit). This statement does not include a match clause, indicating that all routes are matched. Therefore, the route map statement 20 match all routes that the route map statement does not. The route map has been configured with an “explicit permit all” statement at the end of the route map.

Routes that the route map statement 20 matches, have their weight set to 100. The result is that the router accepts the routes that originated in AS 200 (all 10.0.2.0/24 networks) and assigns the weight 200. All others are accepted and assigned the weight 100. This route map does not discard any route.

Note Specifying weights with filter lists is no longer supported in Cisco IOS Software Release 12.1 and later. 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.

BGP Route Selection and Filtering Tools Summary

Prefix lists and filter lists, both in and out, filter out routes and discard the ones 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.



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

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- BGP uses a number of criteria for best-path route selection—the weight is taken into consideration first.
- The weight attribute can be assigned to a route using one of the following:
 - Default weight value can be assigned to all routes that are received from a specific neighbor.
 - Route maps can be applied to neighbors to set the weight attribute of received routes.
- You can use the **neighbor weight** command to assign a weight value to all routes that are received from a neighbor.
- You can use the **show ip bgp** command to display all BGP routes, the routes that BGP selected as best, and the weight attribute setting for each route.

Lesson 2: Setting BGP Local Preference

Overview

Sometimes you need to influence [BGP](#) route selection. BGP provides many tools to achieve this. One of these tools is also the local preference attribute.

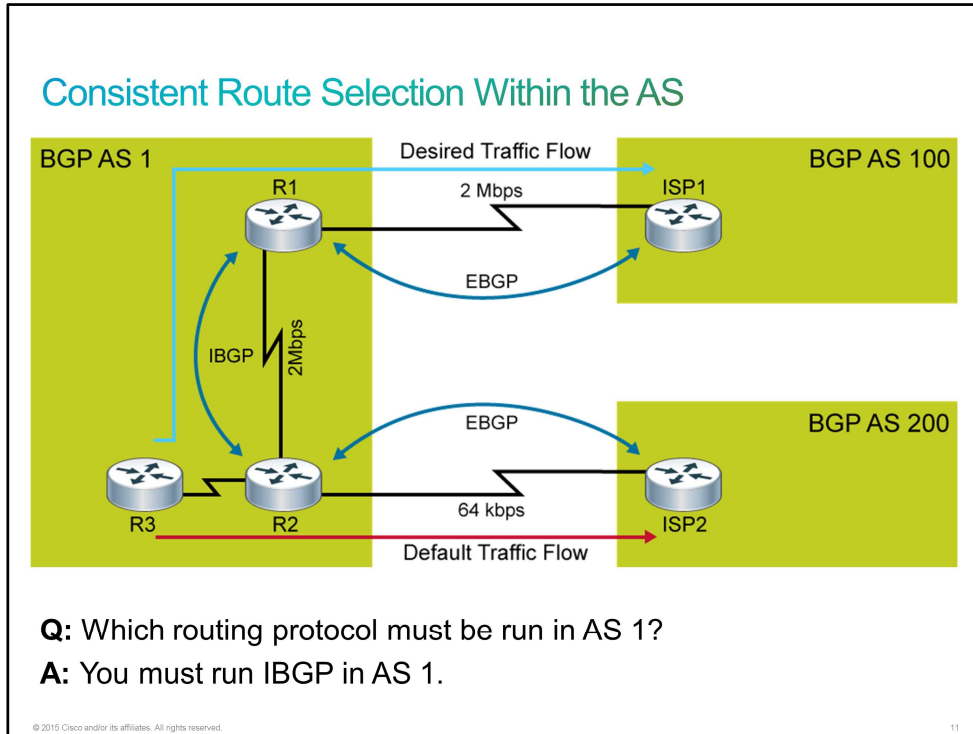
In this lesson, you will learn how to influence BGP route selection by setting the BGP local preference attribute of incoming BGP routes. Local preference is similar to the weight attribute but differs from the BGP weight attribute in that weight is local to a specific router on which it is configured.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

Upon completing this lesson, you will be able to:

- 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 monitor BGP local preference
- Describe how 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
- Describe how to configure BGP local preference using route-maps

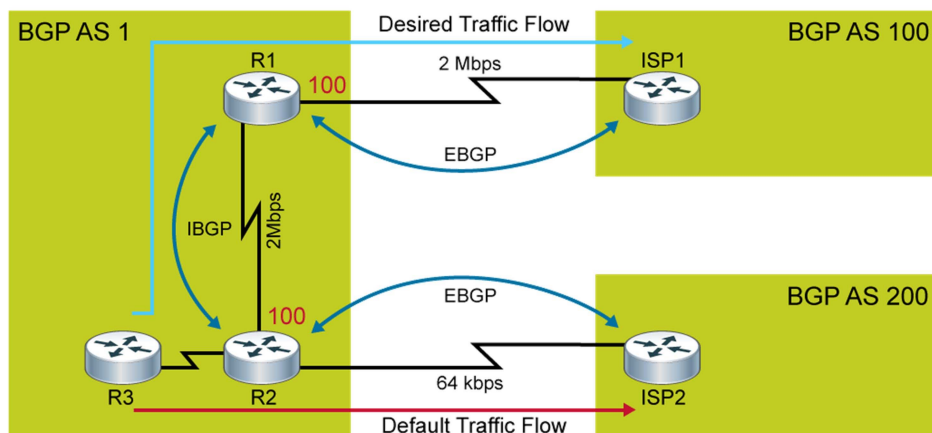
Consistent Route Selection Within the AS

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 the 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.



You have 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. To assure this, you must configure an [IBGP](#) session between the two border routers in AS 1.

Consistent Route Selection Within the AS (Cont.)



Q: How will you influence the route selection on routers in AS 1 so that they select the fastest route?

A: Use weights on EBGP and IBGP sessions.

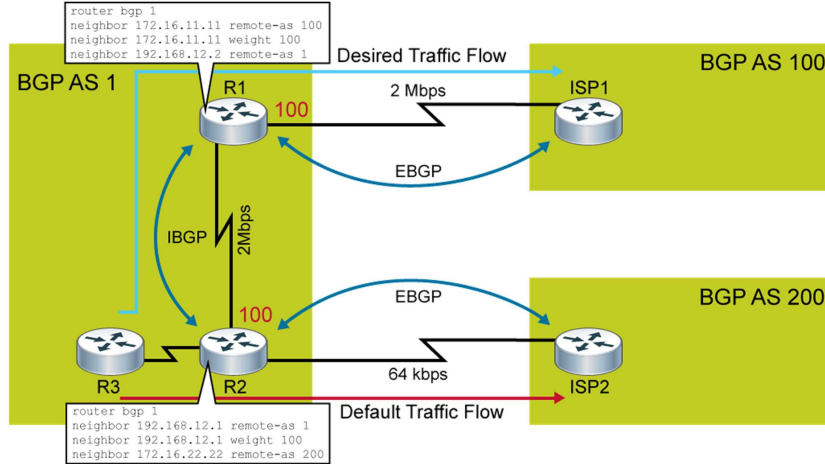
© 2015 Cisco and/or its affiliates. All rights reserved.

12

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, you as a network administrator 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 R1 router in AS 1 sets a weight of 100 for routes that are received over the 2-Mbps link from AS 100 (primary link). R1 prefers these routes 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 200. As a result, all packets will leave the AS through the primary 2-Mbps link.

Consistent Route Selection Within the AS (Cont.)



Run the traffic over the fastest line available.

© 2015 Cisco and/or its affiliates. All rights reserved.

13

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 Cisco IOS Software Release 11.2 and above supports the soft reconfiguration feature..
- Use the **clear ip bgp neighbor address in** command if both neighboring routers support the route refresh. The Cisco IOS Software Release 12.1 and above supports the route refresh feature.

When you are trying to implement this example with weights, it requires two route maps on each router within AS 1. Luckily, BGP has a similar mechanism that you can use for consistent AS-wide route selection: local preference.

BGP Local Preference

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.

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.

© 2015 Cisco and/or its affiliates. All rights reserved.

14

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. 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 [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 that are 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

15

Local preference is the second-strongest [BGP](#) route selection parameter. Remember the route selection rules:

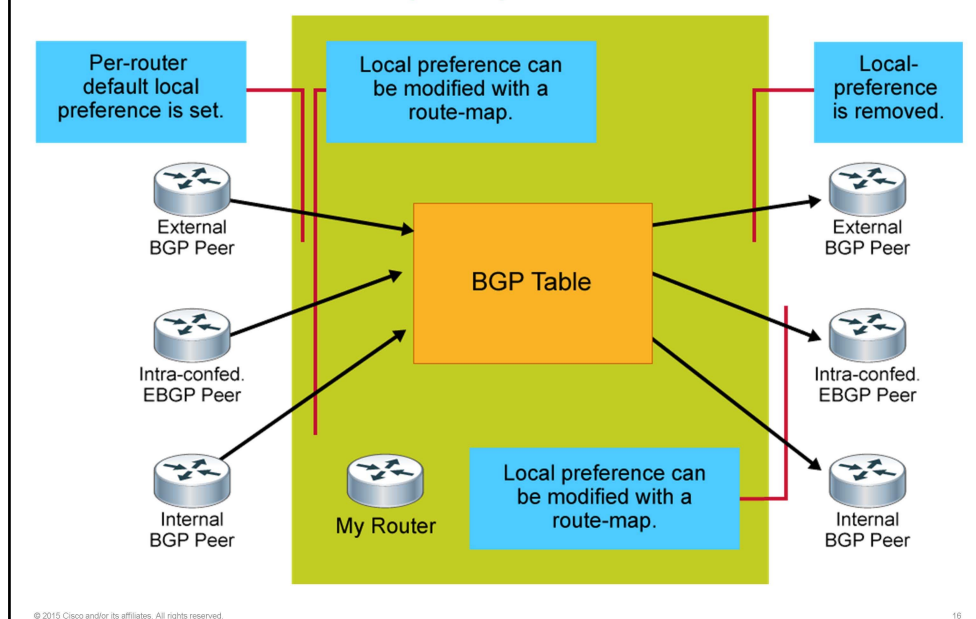
1. Prefer the highest weight (local to router).
2. Prefer the highest local preference (global within AS).
3. Process all remaining BGP route selection rules.

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

You 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. You should avoid changing the local preference attribute on internal sessions to prevent unnecessary complexity and unpredictable behavior.

BGP Local Preference (Cont.)



You 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

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 the routes that are received from external neighbors.

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.

© 2015 Cisco and/or its affiliates. All rights reserved.17

If you set a value lower than the default of 100, the router will prefer internal paths to external (normally a router would prefer external routes).

If you set a value higher than 100, then external paths will be preferred to all internal paths (also the ones with a shorter [AS](#) path).

Monitoring Local Preference

Monitoring Local Preference

- Display nondefault local preference with the **show ip bgp** command.
- Local preference is displayed in the **show ip bgp prefix** command.
- Local preference is also displayed in BGP update debugging (only for inbound updates).

© 2015 Cisco and/or its affiliates. All rights reserved.

18

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.

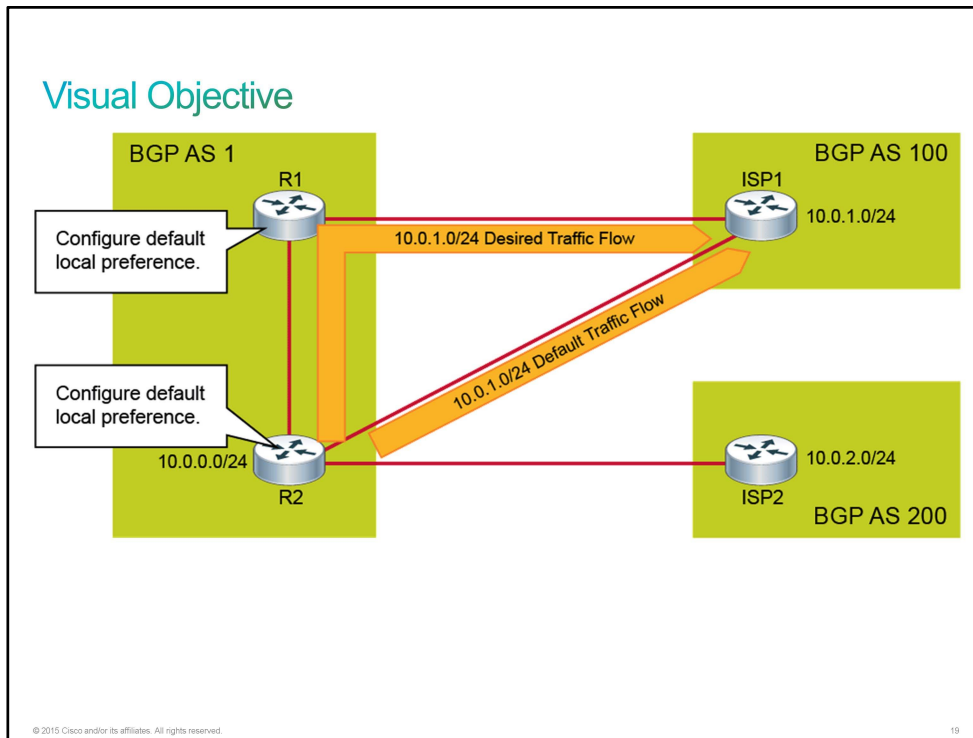
Note	The output of the show ip bgp command will not display the local preference value if the value is the same as the bgp default local-preference value in the local router.
-------------	---

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.

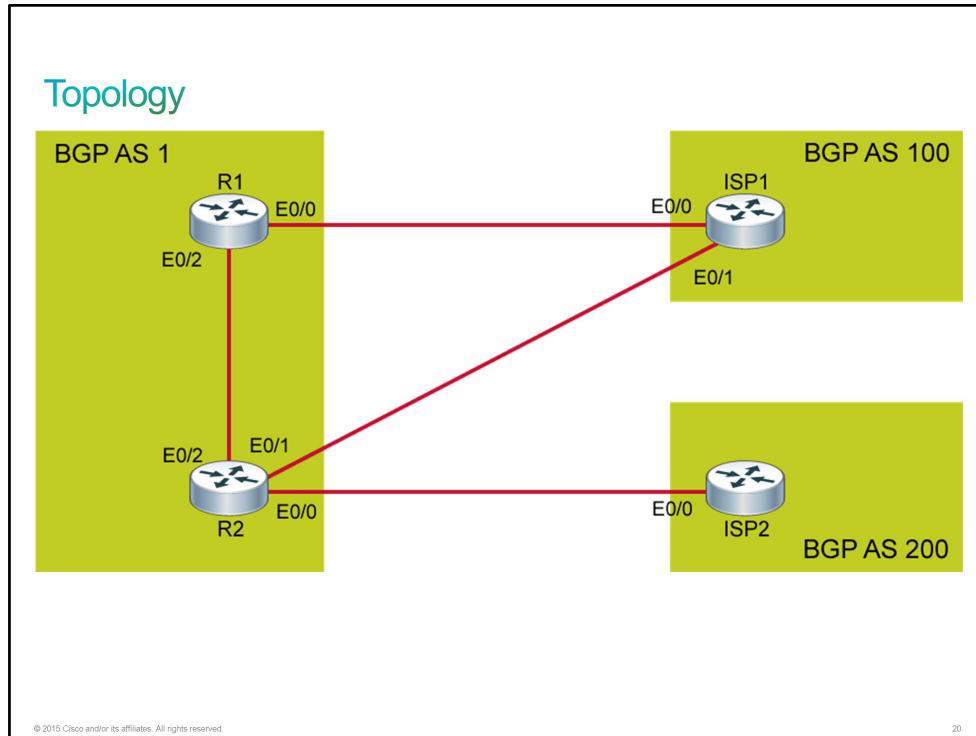
Discovery 12: Configure and Monitor Local Preference

Overview

Through this discovery, you will learn how to configure [BGP](#) default local preference. You will influence R2 router to prefer the path to ISP1 networks via R1, instead of using the directly connected link.



Topology



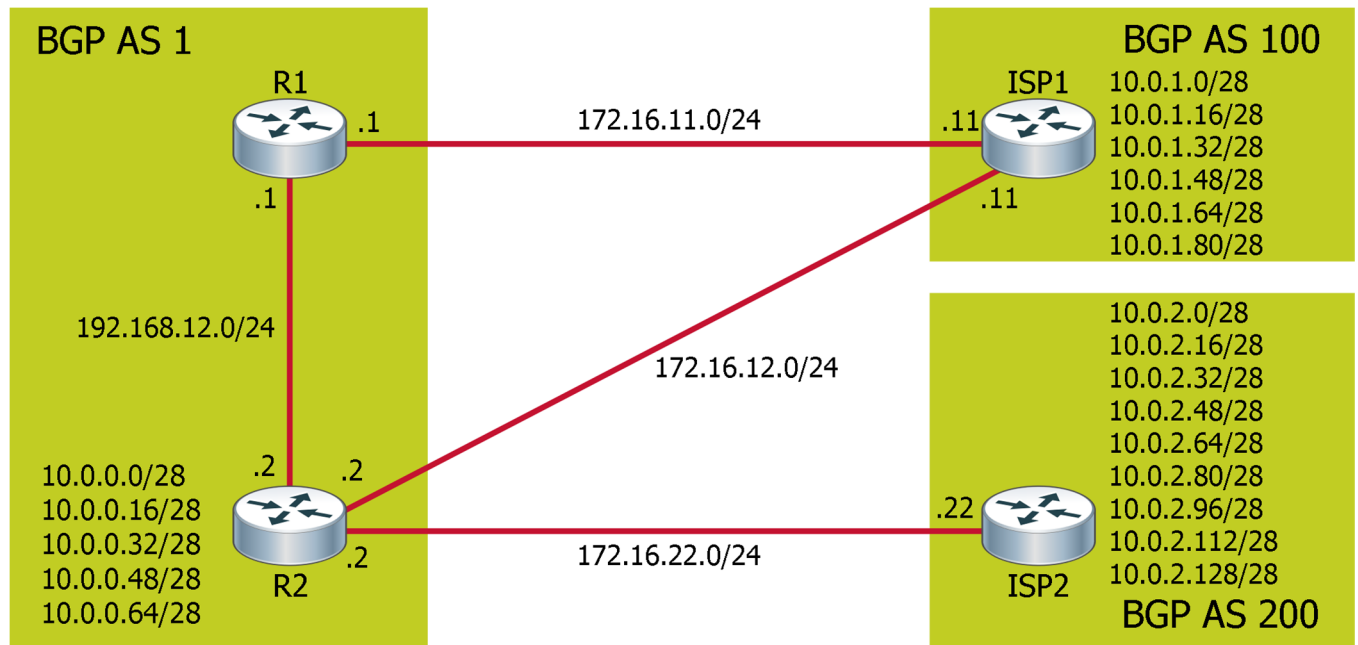
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R1	Ethernet 0/0	172.16.11.1/24	Connection to ISP1
R1	Ethernet 0/2	192.168.12.1/24	Connection to R2
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/1	172.16.12.2/24	Connection to ISP1
R2	Ethernet 0/2	192.168.12.2/24	Connection to R1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/0	172.16.11.11/24	Connection to R1
ISP1	Ethernet 0/1	172.16.12.11/24	Connection to R2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in BGP are preconfigured as shown below:



BGP is also preconfigured:

- [EBGP](#)
 - R1 to ISP1
 - R2 to ISP1
 - R2 to ISP2
- [IBGP](#)
 - R1 to R2

Configure and Monitor Local Preference

Discovery Steps

Step 1 On the R2 router, verify the initial state of default local preferences for all received destinations.

R2# **show ip bgp**

BGP table version is 21, local router ID is 10.0.0.65

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
x best-external, a additional-path, c RIB-compressed,

Origin codes: i - IGP, e - EGP, ? - incomplete

RPKI validation codes: V valid, I invalid, N Not found

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	0.0.0.0	0		32768	i
*>	10.0.0.16/28	0.0.0.0	0		32768	i
*>	10.0.0.32/28	0.0.0.0	0		32768	i
*>	10.0.0.48/28	0.0.0.0	0		32768	i
*>	10.0.0.64/28	0.0.0.0	0		32768	i
* i	10.0.1.0/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.16/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.32/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.48/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.64/28	192.168.12.1	0	100	0	100 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>		172.16.12.11	0		0	100 i
* i	10.0.1.80/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
*>	10.0.2.0/28	172.16.22.22	0		0	200 i
*>	10.0.2.16/28	172.16.22.22	0		0	200 i
*>	10.0.2.32/28	172.16.22.22	0		0	200 i
*>	10.0.2.48/28	172.16.22.22	0		0	200 i
*>	10.0.2.64/28	172.16.22.22	0		0	200 i
*>	10.0.2.80/28	172.16.22.22	0		0	200 i
*>	10.0.2.96/28	172.16.22.22	0		0	200 i
*>	10.0.2.112/28	172.16.22.22	0		0	200 i
*>	10.0.2.128/28	172.16.22.22	0		0	200 i

R2 has received routes to the following networks:

- 10.0.0.0/24 networks (originating locally on R2), for which the applied default local preference of 100 is not displayed.
- 10.0.1.0/24 networks (originating on the ISP1 router) are received from R1 and ISP1:
 - The first path for these networks was received from R1 (with next hop 192.168.12.1), with local preference value of 100. Local preference of internal route is displayed in the output.
 - The second path for these networks was received from ISP1 (with next hop 172.16.12.11). The applied default local preference of 100 is not displayed.
- 10.0.2.0/24 networks (originating on the ISP2 router) are received from ISP2. The applied default local preference of 100 is not displayed.

For 10.0.1.0/24 networks, multiple paths exist and the decision mechanism on R2 prefers the external (EBGP) paths over internal (IBGP).

Step 2 Influence the R2 router to prefer the path to ISP1 networks via R1 router.

On R1, apply local preference of 120 for all routes that are received from external neighbor.

```
R1(config)# router bgp 1
R1(config-router)# bgp default local-preference 120
```

Step 3 On the R2 router, verify configured default local preference for all received destinations.

```
R2# show ip bgp
BGP table version is 20, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	0.0.0.0	0		32768	i
*>	10.0.0.16/28	0.0.0.0	0		32768	i
*>	10.0.0.32/28	0.0.0.0	0		32768	i
*>	10.0.0.48/28	0.0.0.0	0		32768	i
*>	10.0.0.64/28	0.0.0.0	0		32768	i
*>i	10.0.1.0/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
*>i	10.0.1.16/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
*>i	10.0.1.32/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
*>i	10.0.1.48/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
*>i	10.0.1.64/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>i	10.0.1.80/28	192.168.12.1	0	120	0	100 i
*		172.16.12.11	0		0	100 i
*>	10.0.2.0/28	172.16.22.22	0		0	200 i
*>	10.0.2.16/28	172.16.22.22	0		0	200 i
*>	10.0.2.32/28	172.16.22.22	0		0	200 i
*>	10.0.2.48/28	172.16.22.22	0		0	200 i
*>	10.0.2.64/28	172.16.22.22	0		0	200 i
*>	10.0.2.80/28	172.16.22.22	0		0	200 i
*>	10.0.2.96/28	172.16.22.22	0		0	200 i
*>	10.0.2.112/28	172.16.22.22	0		0	200 i
*>	10.0.2.128/28	172.16.22.22	0		0	200 i

R2 has received routes to the following networks:

- 10.0.0.0/24 networks that originate locally on R2, and the applied default local preference of 100 is not displayed.
- 10.0.1.0/24 networks are received from R1 and ISP1:
 - The first path for these networks was received from R1 (next hop 192.168.12.1), with a changed local preference value of 120. Local preference of internal route is displayed in the output. Note that this is the best path for 10.0.1.0/24 networks, marked with > sign.
 - The second path for these networks was received from ISP1 (next hop 172.16.12.11). The applied default local preference of 100 is not displayed.
- 10.0.2.0/24 networks are received from ISP2. The applied default local preference of 100 is not displayed.

For 10.0.1.0/24 networks, multiple paths exist and the decision mechanism on R2 selects a route with higher local preference attribute (120), since weight value is left on default, 0.

Configuring Local Preference with Route Maps

To have more control over setting local preference, you may be forced to use a route map.

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 that the route map entry matches

```
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 that is configured by using a route map with no match condition

© 2015 Cisco and/or its affiliates. All rights reserved.

21

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. You can then apply the route map 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.

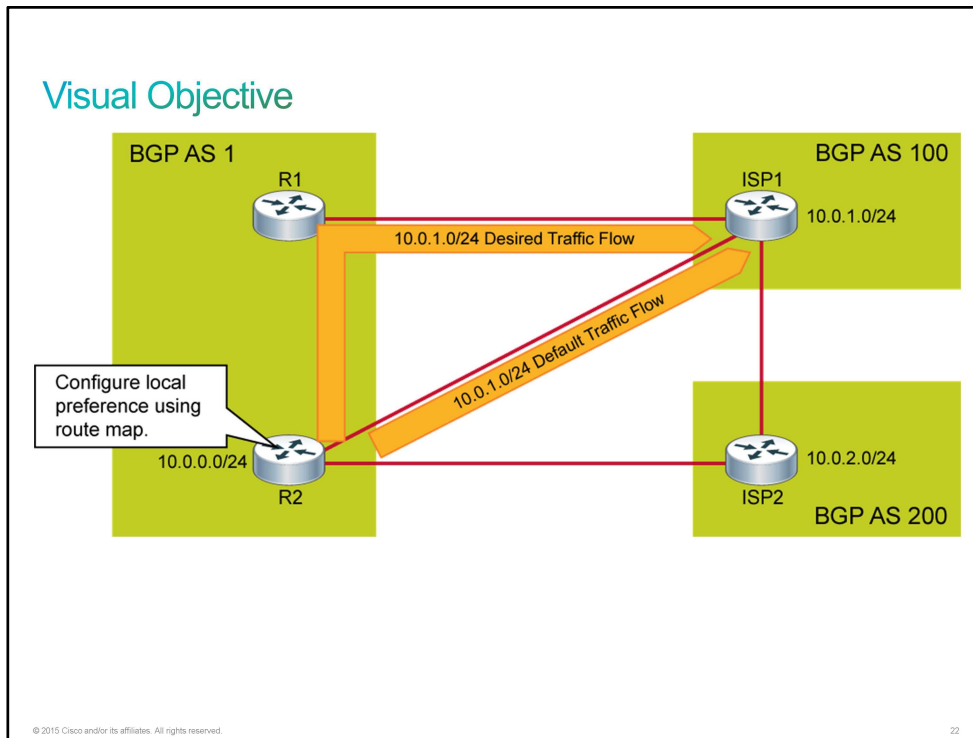
When you use a route map to set local preference, the route map is typically applied to incoming BGP routes that the [EBGP](#) neighbor advertises. The local router uses the local preference attribute in BGP route selection. In addition, the router also propagates the attribute to all [IBGP](#) sessions 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.

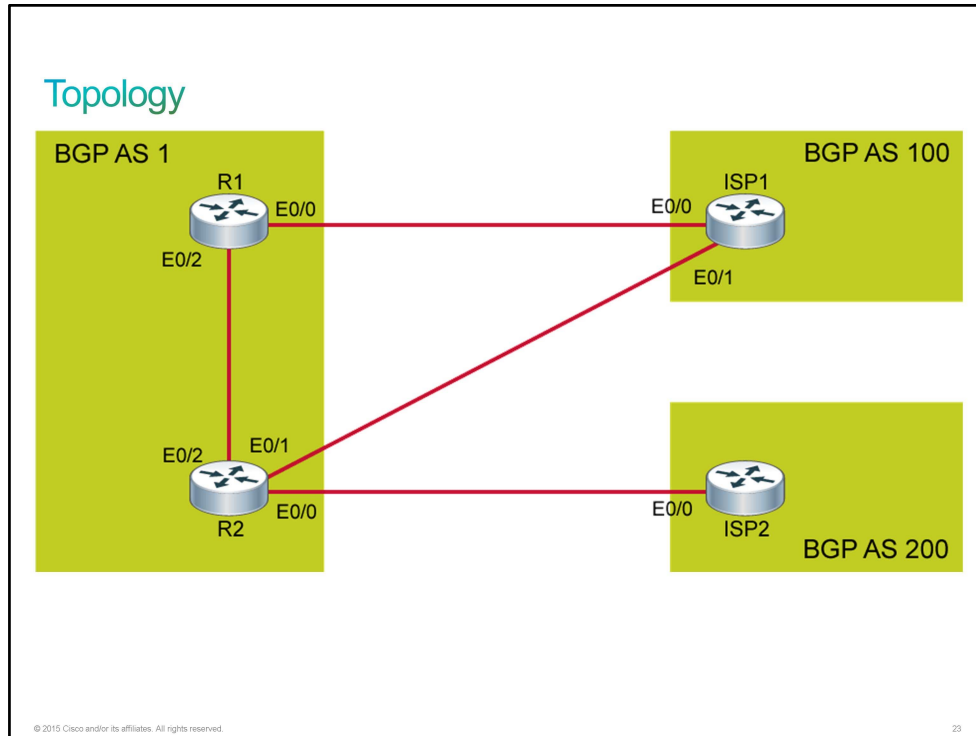
Discovery 13: Configure Local Preference Using Route Maps

Overview

Through this discovery, you will learn how to configure local preference using route maps. You will again configure R2 router to prefer the path to ISP1 networks via R1, instead of using the directly connected link.



Topology



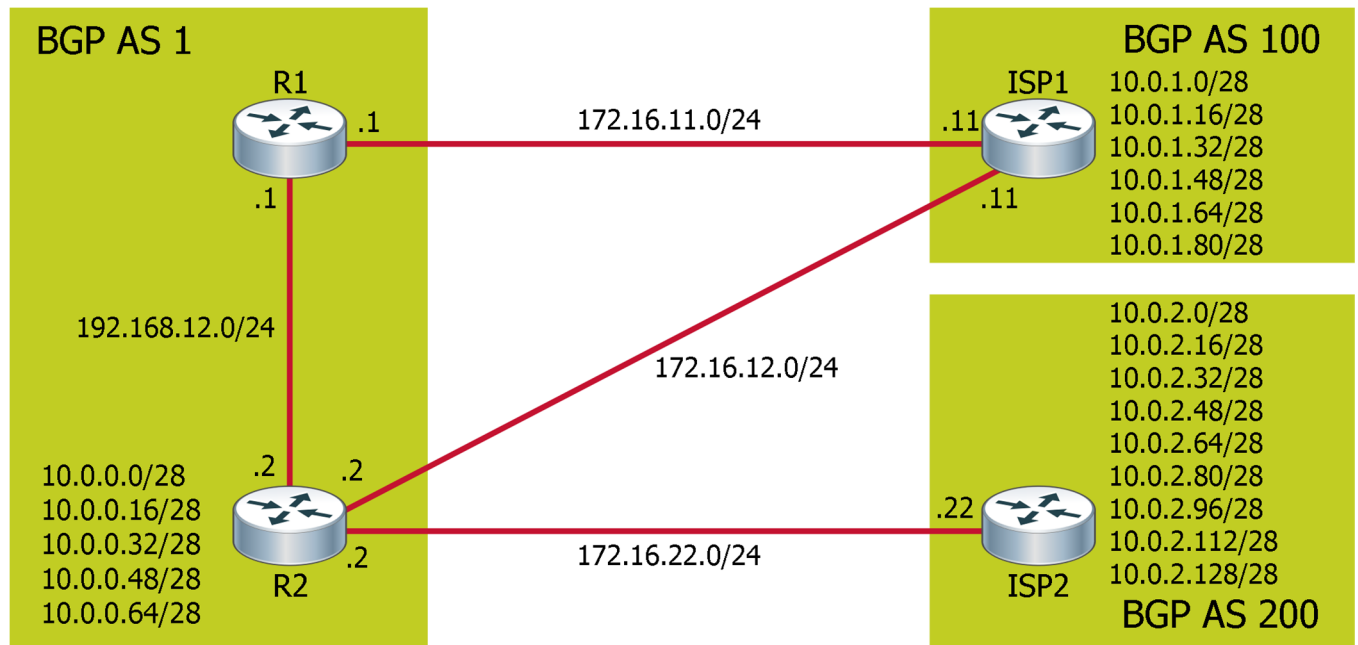
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R1	Ethernet 0/0	172.16.11.1/24	Connection to ISP1
R1	Ethernet 0/2	192.168.12.1/24	Connection to R2
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/1	172.16.12.2/24	Connection to ISP1
R2	Ethernet 0/2	192.168.12.2/24	Connection to R1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/0	172.16.11.11/24	Connection to R1
ISP1	Ethernet 0/1	172.16.12.11/24	Connection to R2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in [BGP](#) are preconfigured as shown below:



BGP is also preconfigured:

- [EBGP](#)
 - R1 to ISP1
 - R2 to ISP1
 - R2 to ISP2
- [IBGP](#)
 - R1 to R2

Configure Local Preference Using Route Maps

Discovery Steps

Step 1 On the R2 router, verify the initial state of default local preferences for all received destinations.

To verify local preferences, use the **show ip bgp** command.

R2# **show ip bgp**

BGP table version is 21, local router ID is 10.0.0.65

Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
x best-external, a additional-path, c RIB-compressed,

Origin codes: i - IGP, e - EGP, ? - incomplete

RPKI validation codes: V valid, I invalid, N Not found

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	0.0.0.0	0		32768	i
*>	10.0.0.16/28	0.0.0.0	0		32768	i
*>	10.0.0.32/28	0.0.0.0	0		32768	i
*>	10.0.0.48/28	0.0.0.0	0		32768	i
*>	10.0.0.64/28	0.0.0.0	0		32768	i
* i	10.0.1.0/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.16/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.32/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.48/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
* i	10.0.1.64/28	192.168.12.1	0	100	0	100 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>		172.16.12.11	0		0	100 i
* i	10.0.1.80/28	192.168.12.1	0	100	0	100 i
*>		172.16.12.11	0		0	100 i
*>	10.0.2.0/28	172.16.22.22	0		0	200 i
*>	10.0.2.16/28	172.16.22.22	0		0	200 i
*>	10.0.2.32/28	172.16.22.22	0		0	200 i
*>	10.0.2.48/28	172.16.22.22	0		0	200 i
*>	10.0.2.64/28	172.16.22.22	0		0	200 i
*>	10.0.2.80/28	172.16.22.22	0		0	200 i
*>	10.0.2.96/28	172.16.22.22	0		0	200 i
*>	10.0.2.112/28	172.16.22.22	0		0	200 i
*>	10.0.2.128/28	172.16.22.22	0		0	200 i

R2 has received routes to the following networks:

- 10.0.0.0/24 networks that originate locally on R2, and the applied default local preference of 100 is not displayed.
- 10.0.1.0/24 networks are received from R1 and ISP1:
 - The first path for these networks was received from R1 (next hop 192.168.12.1), with local preference value of 100. Local preference of internal route is displayed in the output.
 - The second path for these networks was received from ISP1 (next hop 172.16.12.11). The applied default local preference of 100 is not displayed.
- 10.0.2.0/24 networks are received from ISP2. The applied default local preference of 100 is not displayed.

For 10.0.1.0/24 networks, multiple paths exist and the decision mechanism on R2 prefers the external (EBGP) paths over internal (IBGP).

Step 2 On the R2 router, configure route map to set local preference on 10.

Name the route map "LP10."

```
R2(config)# route-map LP10 10
R2(config-route-map)# set local-preference 10
```

Step 3 Influence the R2 router to prefer the path to ISP1 networks via R1 router. On the R2 router, use configured route map to set lower local preference for routes that are received from [AS 100](#).

Apply route map LP10 for routes that are received from AS 100.

```
R2(config)# router bgp 1
R2(config-router)# neighbor 172.16.12.11 route-map LP10 in
```

Local preference of 10 will now be applied to external updates from ISP1. Consequently, R2 will prefer the path to networks in AS 100 via R1, since it has higher local preference value.

Step 4 Clear all BGP sessions on the R2 router.

```
R2# clear ip bgp *
```

Step 5 On the R2 router, verify local preference for all received destinations.

```
R2# show ip bgp
BGP table version is 20, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0/28	0.0.0.0	0		32768	i
*> 10.0.0.16/28	0.0.0.0	0		32768	i
*> 10.0.0.32/28	0.0.0.0	0		32768	i
*> 10.0.0.48/28	0.0.0.0	0		32768	i
*> 10.0.0.64/28	0.0.0.0	0		32768	i
*>i 10.0.1.0/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
*>i 10.0.1.16/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
*>i 10.0.1.32/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
*>i 10.0.1.48/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
*>i 10.0.1.64/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
Network	Next Hop	Metric	LocPrf	Weight	Path
*>i 10.0.1.80/28	192.168.12.1	0	100	0	100 i
*	172.16.12.11	0	10	0	100 i
*> 10.0.2.0/28	172.16.22.22	0		0	200 i
*> 10.0.2.16/28	172.16.22.22	0		0	200 i
*> 10.0.2.32/28	172.16.22.22	0		0	200 i
*> 10.0.2.48/28	172.16.22.22	0		0	200 i
*> 10.0.2.64/28	172.16.22.22	0		0	200 i
*> 10.0.2.80/28	172.16.22.22	0		0	200 i
*> 10.0.2.96/28	172.16.22.22	0		0	200 i
*> 10.0.2.112/28	172.16.22.22	0		0	200 i
*> 10.0.2.128/28	172.16.22.22	0		0	200 i

R2 has received routes to the following networks:

- 10.0.0.0/24 networks that originate locally on R2, and the applied default local preference of 100 is not displayed.
- 10.0.1.0/24 networks are received from R1 and ISP1:
 - The first path for these networks was received from R1 (next hop 192.168.12.1) with a default local preference value of 100. Local preference of internal route is displayed in the output.
 - The second path for these networks was received from ISP1 (next hop 172.16.12.11) with a local preference value of 10, set by the route map. The applied local preference is not left on default and is displayed in the output.
- 10.0.2.0/24 networks are received from ISP2. The applied default local preference of 100 is not displayed.

For 10.0.1.0/24 networks, multiple paths exist and the decision mechanism on R2 selects the route with a higher local preference attribute (100), since weight value is left on default, 0. Consequently the path via R1 will be chosen to reach 10.0.1.0/24 networks.

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

- Local preference is similar to the weight attribute. You can use both to influence BGP path selection, but the 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.
- Configure local preference:
 - Using the **bgp default local-preference** *preference* command.
 - Using route map statements.

© 2015 Cisco and/or its affiliates. All rights reserved.

24

Lesson 3: Using AS Path Prepending

Overview

[BGP](#) provides many tools for administrators to influence route selection. You have already learned how to influence BGP route selection for outgoing traffic by setting weight or local preference.

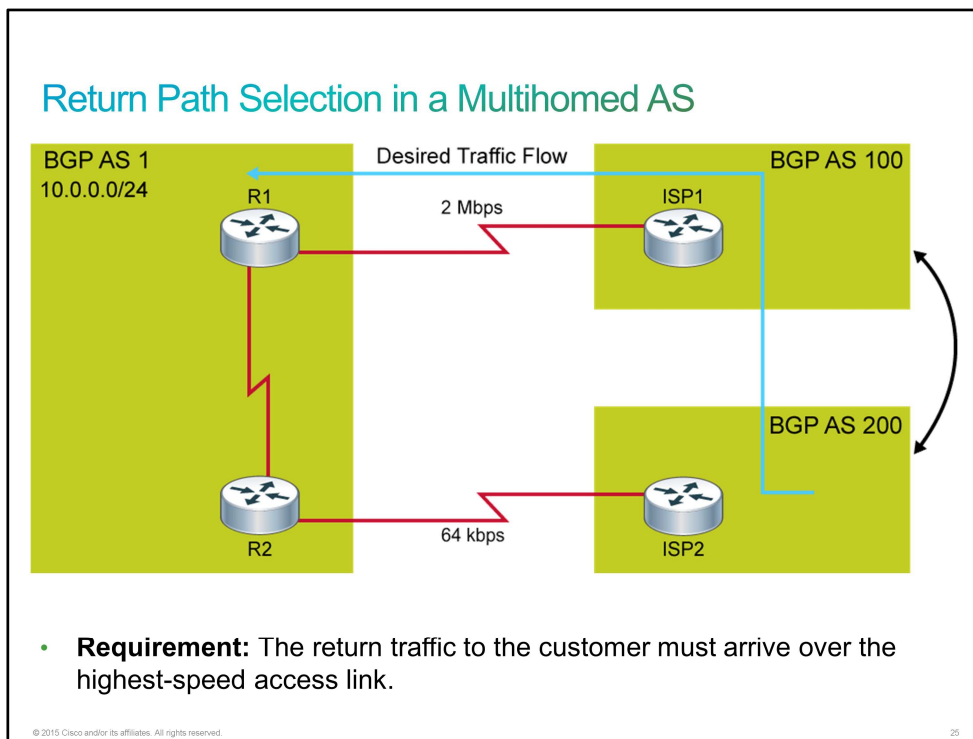
Problems can arise when administrative policies mandate a specific return path to be used for traffic returning to the [AS](#). AS path prepending potentially allows the customer to influence the route selection of its service providers, so the path for returning traffic. In this lesson, you will learn about AS path prepending and the Cisco IOS commands that are required to properly configure and monitor AS path configurations. You will also learn filtering requirements for influencing route selection using AS path prepending.

Upon completing this lesson, you will be able to:

- 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
- Describe how to configure AS path prepending on the customer side and AS path filtering on the service provider side
- Describes how the BGP Hide Local-Autonomous System feature simplifies the task of changing the AS number in a BGP network

Return Path Selection in a Multihomed AS

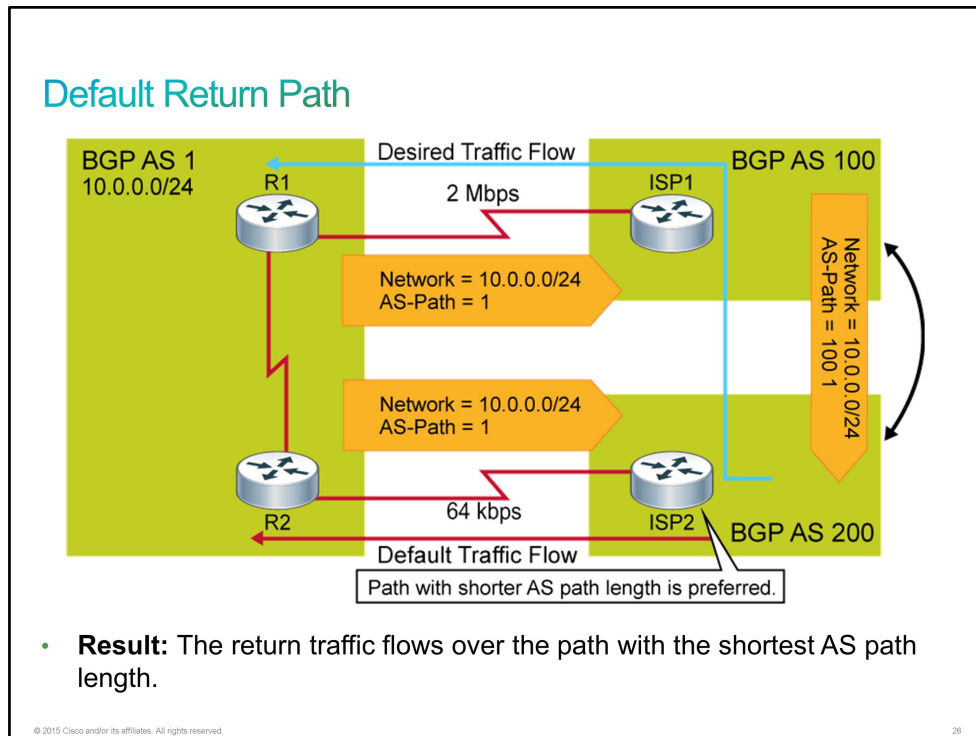
It is fairly easy for an [AS](#) to select the appropriate path for outgoing traffic. It is much more complicated to influence other autonomous systems to select the appropriate path for traffic that is returning to a specific AS.



If you configure the preferred path only for outgoing traffic and not for incoming (return) traffic, you will likely get an asymmetrical traffic flow. It can as well result in 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 200 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 1 requests AS 200 to send packets toward network 10.0.0.0/24 via AS 100. The reason for this request is to improve the network performance and minimize delay. Assuming, of course, that the connectivity between AS 100 and AS 200 is better than the direct 64-kbps link between AS 200 and AS 1.

Default Return Path



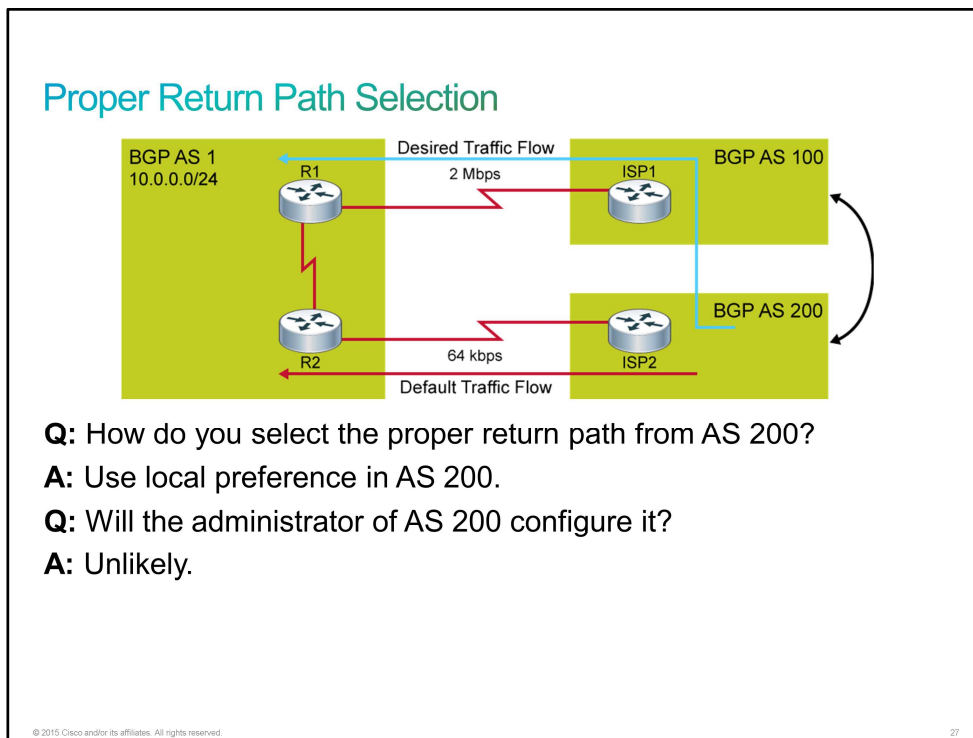
If you do not configure any [BGP](#) path selection tools on the route to influence the traffic flow, AS 200 will use the shortest AS path. This action will result in unwanted behavior because the return traffic to AS 1 will be sent over the low-speed [WAN](#) link.

AS 1 announces network 10.0.0.0/24 over [EBGP](#) sessions to both AS 100 and AS 200. When AS 1 sends EBGp updates, it changes the AS path attribute according to BGP specifications. Both AS 100 and AS 200 receive a BGP update for network 10.0.0.0/24 with the AS path set to 1.

AS 100 selects the route for network 10.0.0.0/24 that it received from AS 1 as its best route. AS 100 uses that route and forwards it to AS 200. According to BGP specifications, AS 100 also changes the AS path attribute. AS 200 receives the route to network 10.0.0.0/24 from AS 100 with an AS path set to 100 1.

AS 200 has now received two alternative routes to network 10.0.0.0/24 (the direct route from AS 1 and the route through AS 100). Remember that you did not configure anything in AS 200 to influence the flow of traffic. So, the router will use the BGP route selection rule of shortest AS path to select the best return path to network 10.0.0.0/24.

Proper Return Path Selection



Remember that the incoming traffic flow (from the perspective of AS 1) will be a result of the route selection for outgoing traffic in AS 200. The traffic that is going out from AS 200 will end up as incoming traffic in AS 1.

AS 200 can configure some changes that cause the route selection process for outgoing traffic to prefer to reach network 10.0.0.0/24 via AS 100. These changes would result in behavior matching the desired administrative policy for AS 1, 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 1 is to reconfigure the router in AS 200, which is receiving EBGp updates directly from AS 1. You can configure the router in AS 200 to assign a local preference value less than the default value (100) to all routes that are received from AS 1. The router in AS 200 is also configured specifically not to set local preference on EBGp routes that are received from AS 100. This configuration results in assignment of the default value of 100 to all routes received from AS 100. This difference in local preference values causes AS 200 routers to select the path via AS 100 as the best one to reach 10.0.0.0/24 networks.

However, all the configuration work to complete this process must be performed in AS 200. The network administrators of AS 200 would be required to modify the router configurations in AS 200 to satisfy the administrative policy requirements of AS 1. All changes must be documented and maintained according to the rules and procedures that have been adopted by AS 200.

If AS 200 is a major [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 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

28

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 you, as the operator of an AS, 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. 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 originate at the router that is performing the route selection. The fourth criterion will apply, however, because the AS paths have different lengths.

If you do not manually manipulate the AS path by some administrative means, the router selects the path going over the fewest number of ASs, regardless of the 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

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. The AS path sent out over the nondesired link becomes longer than the AS path sent out over the preferred path. So, the nondesired link is now less likely to be used as the return path.

AS Path Prepending

AS path prepending is known as manual manipulation of AS path length.

- 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

29

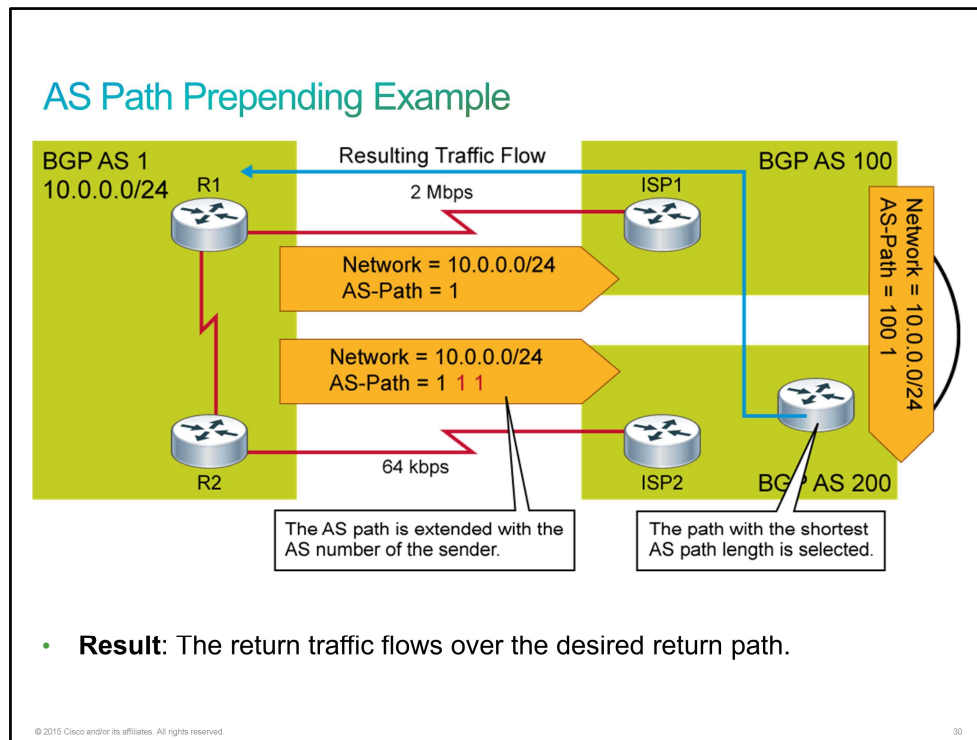
The length of the AS path is extended because extra copies of the AS number of the sender are prepended to (added to the beginning of) the AS path attribute. To avoid clashes with [BGP](#) loop prevention mechanisms, you should prepend no other AS number to the AS path attribute, except that of the sending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

If you prepend another AS number 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.

AS Path Prepending Example

Administrative policy in AS 1 prefers that the low-speed link is used for backup purposes only. As long as the high-speed link between AS 1 and AS 100 is available, all traffic should flow toward AS 1 using the high-speed link.



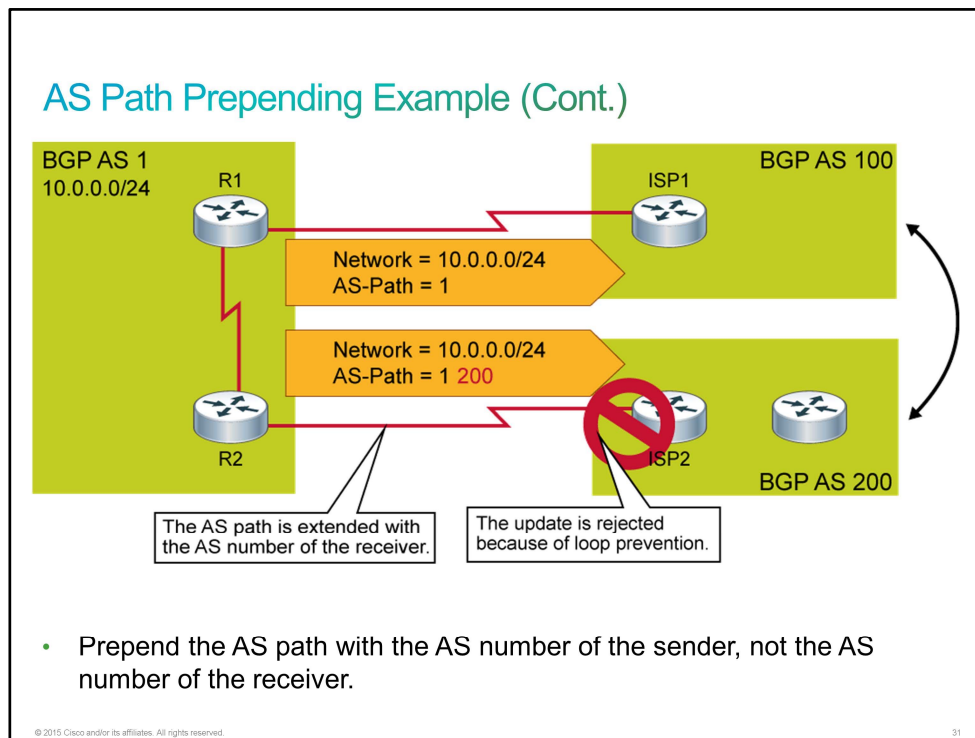
To accomplish this goal, you can configure the router in AS 1, that sends EBGp updates to AS 200, to prepend two copies of AS number "1" to AS path.

AS 200 receives two alternative routes to reach network 10.0.0.0/24:

- The update that it has received directly from AS 1 and has a manipulated AS path with a length of three.
- The update that it has received via AS 100 and was not manually manipulated, and therefore contains an AS path length of two.

When AS 200 starts the selection process for best route to reach network 10.0.0.0/24, it checks the AS path length after the weight and local preference parameters. In this case, neither weight nor local preference has been configured, the length of the AS path will be the deciding factor in the route selection process. Consequently, AS 200 prefers the shortest AS path and thus forwards packets toward network 10.0.0.0/24 via AS 100. The desired administrative policy has been met, and AS 1 will receive incoming traffic over the high-speed link.

If the forwarding path from AS 200 via AS 100 to AS 1 and network 10.0.0.0/24 is later broken, the BGP update to reach network 10.0.0.0/24 is revoked. In case of such a network failure, AS 200 will have only one remaining path to reach network 10.0.0.0/24. The route selection process now has only one choice, the route directly to AS 1 over the low-speed [WAN](#) link. The low-speed link will therefore serve as backup to the high-speed WAN link.



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 1 is prepending AS number 200. The egress router performs AS path prepending when the route is on its way to be transmitted to AS 200. 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 200 receives the BGP update, the AS path contains the value 1 200. The AS number 200 was set by the manual manipulation, and the AS number 1 was prepended automatically by BGP because the update was sent over an EBGP session.

When the edge router in AS 200 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 200. According to the BGP specification, the update will be silently ignored.

Now assume that AS 1 had, for the manual manipulation, used a different AS number, not its own and not AS number 200. Would AS 200 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

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.

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 improved 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.

© 2015 Cisco and/or its affiliates. All rights reserved.32

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, routers forward only routes that are selected as best. So an AS that receives multiple alternatives to a destination will select the route with the shortest AS path and forward only that route.

When 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 then 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 announced AS path to the [EBGP](#) neighbor connected via the backup link, the less likely it is that incoming traffic will be received from that neighbor. You, as a network administrator, can make a clever guess about how many copies of the AS number to prepend. After the prepending is implemented, you have to examine the result. If the expected result is not achieved, you can change the configuration and prepend a few more copies of the AS number.

After AS path prepending has generated the desired results, you 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.

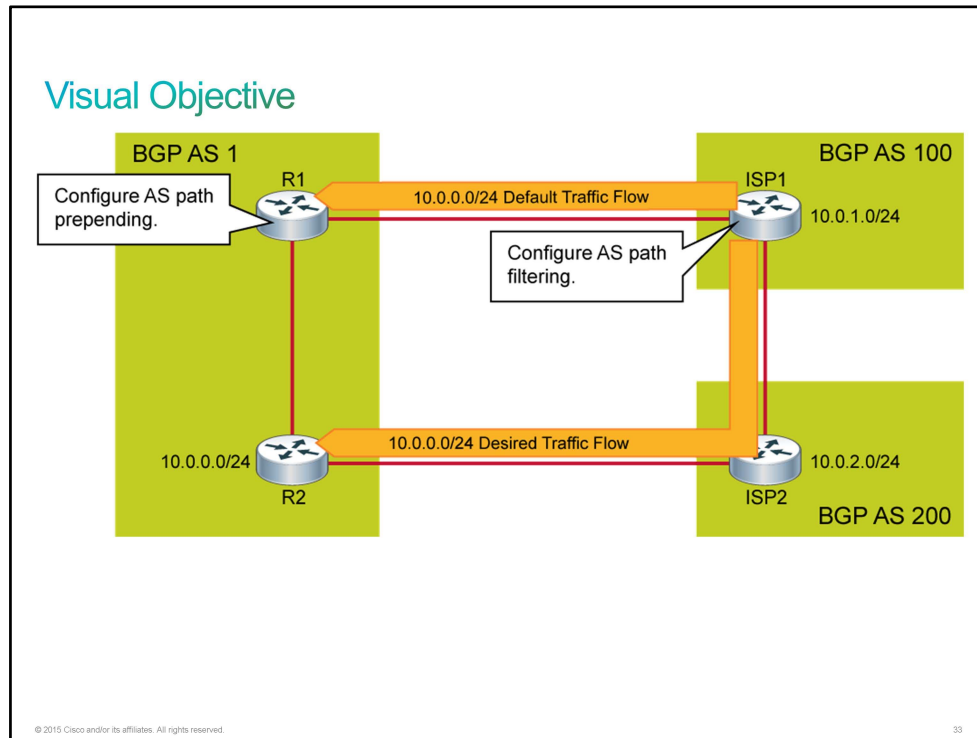
- **Distributing the load of return traffic:** In a multihomed scenario, there is no way to exactly predetermine the volume of traffic that will be received over a particular link. The traffic load on different links will change depending on where the senders are located (which autonomous systems they 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, you 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, you 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, you must prepend more copies of the local AS number. Then the process of resending local routes and monitoring the results starts all over again.

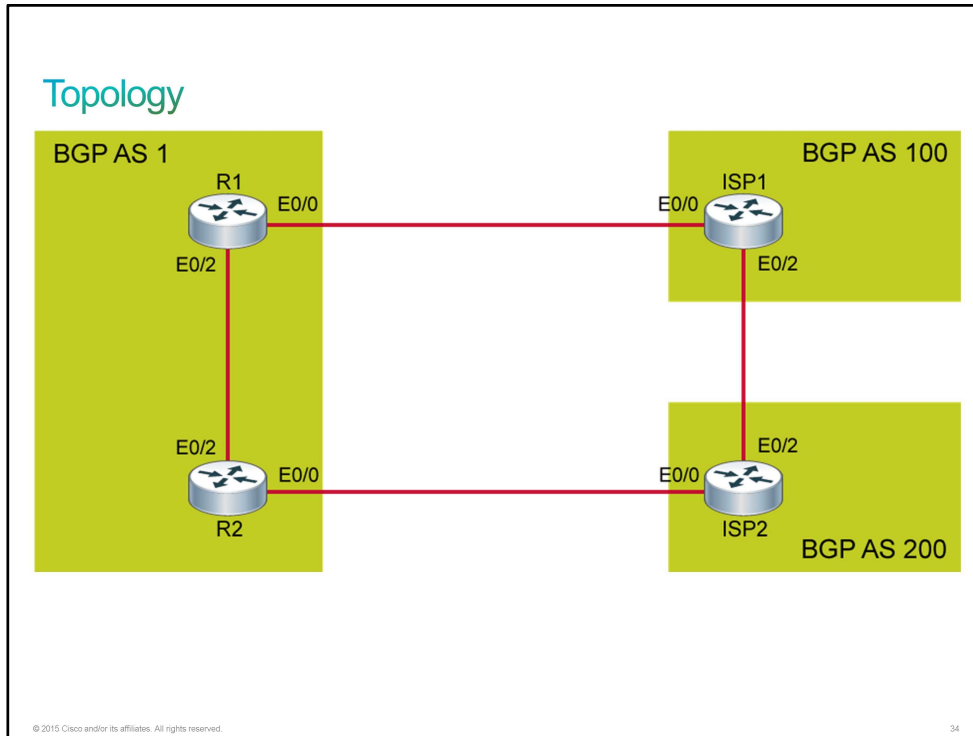
Discovery 14: Configure AS Path Prepending

Overview

In this discovery, you will first configure [AS](#) path prepending on the customer side. Then you will for a moment imagine that you are on a service provider side and you will configure AS path filtering on the service provider side.



Topology



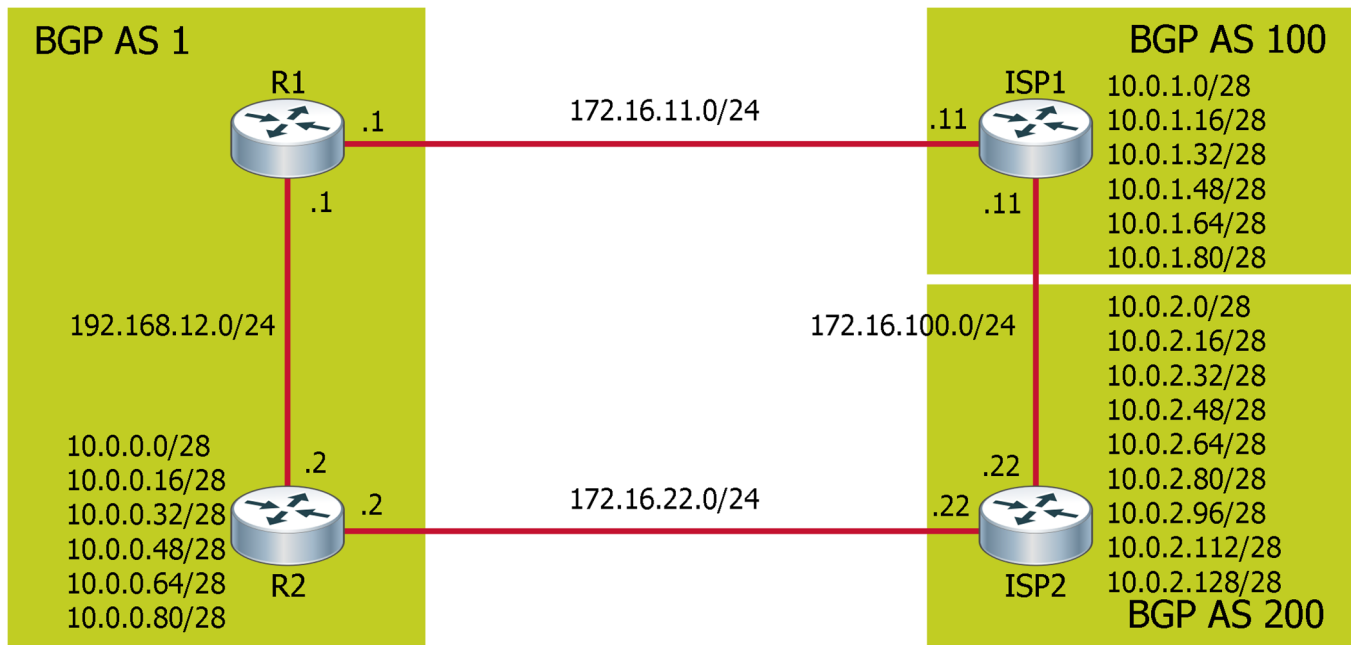
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R1	Ethernet 0/0	172.16.11.1/24	Connection to ISP1
R1	Ethernet 0/2	192.168.12.1/24	Connection to R2
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/2	192.168.12.2/24	Connection to R1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/0	172.16.11.11/24	Connection to R1
ISP1	Ethernet 0/2	172.16.100.11/24	Connection to ISP2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Ethernet 0/2	172.16.100.22/24	Connection to ISP1
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in [BGP](#) are preconfigured as shown below:



BGP is also preconfigured:

- [EBGP](#)
 - R1 to ISP1
 - R2 to ISP2
 - ISP1 to ISP2
- [IBGP](#)
 - R1 to R2

Configure AS Path Prepending

Discovery Steps

Step 1 On the ISP1 router, verify initial best path to reach the 10.0.0.0/24 networks in AS 1.


```
ISP1# show ip bgp
BGP table version is 20, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.0.0/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.16/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.32/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.48/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.64/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*>	10.0.1.0/28	0.0.0.0	0		32768	i
*>	10.0.1.16/28	0.0.0.0	0		32768	i
*>	10.0.1.32/28	0.0.0.0	0		32768	i
*>	10.0.1.48/28	0.0.0.0	0		32768	i
*>	10.0.1.64/28	0.0.0.0	0		32768	i
*>	10.0.1.80/28	0.0.0.0	0		32768	i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.2.0/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.16/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.32/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.48/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.64/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.80/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.96/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.112/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.128/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i

ISP1 received two routes to 10.0.0.0/24 networks:

- The first route is received from ISP2 (next hop 172.16.100.22), with the AS path 200 1.
- The second route is received from R1 (next hop 172.16.11.1), with the AS path 1.

Since weight and local preference attributes are left on default values, ISP1 chooses the path with shortest AS path length as the best (the route that is received from R1).

If you want to influence the path from AS 100 to 10.0.0.0/24 networks in AS 1, you have to configure AS path prepending.

Configuring AS-Path Prepending



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.

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.

```
router(config-router)# neighbor address route-map name out
```

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

© 2015 Cisco and/or its affiliates. All rights reserved.35

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.

Step 2 You will now configure the AS path prepending on the R1 router. First configure route-map on the R1 router.

Name the route map "PREPEND." Select all BGP routes and prepend five copies of "1" to the existing AS path attribute that is attached to each route. This will make sure that all traffic from AS 100 to 10.0.0.0/24 networks in AS1 will flow through ISP2 router in AS 200.

```
R1(config)# route-map PREPEND permit 10  
R1(config-route-map)# set as-path prepend 1 1 1 1 1
```

The lack of match conditions in the route-map indicates that all routes are matched. The AS

path-string "1 1 1 1 1" will be prepended to the existing AS path.

- Step 3** On the R1 router, apply the "PREPEND" route map to all outgoing updates to the ISP1 EBGp neighbor.

```
R1(config)# router bgp 1
R1(config-router)# neighbor 172.16.11.11 route-map PREPEND out
```

AS 100 will now receive a route to network 10.0.0.0/24 from R1 with an AS path containing the AS number 1 six times (1 1 1 1 1 1).

- Step 4** On the R1 router, reestablish BGP sessions to enforce the application of AS path prepending.

To reestablish BGP session, use the **clear ip bgp *** command.

```
R1# clear ip bgp *
```

- Step 5** On the ISP1 router, verify new best path to reach the 10.0.0.0/24 networks in AS 1.

```
ISP1# show ip bgp
BGP table version is 24, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
* > 10.0.0.0/28	172.16.100.22			0	200 1 i
* >	172.16.11.1			0	1 1 1 1 1 1 i
* > 10.0.0.16/28	172.16.100.22			0	200 1 i
* >	172.16.11.1			0	1 1 1 1 1 1 i
* > 10.0.0.32/28	172.16.100.22			0	200 1 i
* >	172.16.11.1			0	1 1 1 1 1 1 i
* > 10.0.0.48/28	172.16.100.22			0	200 1 i
* >	172.16.11.1			0	1 1 1 1 1 1 i
* > 10.0.0.64/28	172.16.100.22			0	200 1 i
* >	172.16.11.1			0	1 1 1 1 1 1 i
* > 10.0.1.0/28	0.0.0.0	0		32768	i
* > 10.0.1.16/28	0.0.0.0	0		32768	i
* > 10.0.1.32/28	0.0.0.0	0		32768	i
* > 10.0.1.48/28	0.0.0.0	0		32768	i
* > 10.0.1.64/28	0.0.0.0	0		32768	i
* > 10.0.1.80/28	0.0.0.0	0		32768	i
Network	Next Hop	Metric	LocPrf	Weight	Path
* 10.0.2.0/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.16/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.32/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.48/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.64/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.80/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.96/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.112/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i
* 10.0.2.128/28	172.16.11.1			0	1 1 1 1 1 1 200 i
* >	172.16.100.22	0		0	200 i

ISP1 received two routes to 10.0.0.0/24 networks:

- The first route is received from ISP2 (next hop 172.16.100.22), with the AS path 200 1.
- The second route is received from R1 (next hop 172.16.11.1), with the AS path 1 1 1 1 1 1.

You can see now, that the received networks have the "1 1 1 1 1" AS path that is prepended to the original AS path. Consequently, the received route from R1 has a longer AS path than the route received from ISP2. Consequently, as weight and local preference value are left on default, ISP1 chooses the path that is received from ISP2 as the best (shortest AS path length).

Note: 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:

- Output of the **debug ip bgp updates** command does not display the prepended paths, because the route map doing the prepending is applied afterward.

- Output of **show route-map** command displays the configuration details of a route map. However, there is no indication of how many routes have been matched by a route-map statement.

A better place for observing AS path prepending is on the router receiving the BGP update that contains the prepended AS path. You can also 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.

AS Path Filtering

Service providers 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 of the rest of the Internet cause the service provider to implement AS path filters on incoming updates.

The network administrator of the service provider can 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 is a better solution. This access list can then be uniformly applied to all incoming routes from all customers.

Note When you configure AS path prepending, you have to agree with the service provider so their AS path filters will be modified to support AS path prepending.

Step 6 Imagine now for a moment that you are a service provider and configure AS path filters to prevent AS path prepending. **NOTE:** A customer normally does not configure the [ISP](#) side. The configuration on the side of ISP is done by their network administrator.

On the ISP1 router, configure AS path filter, a filter list, that will not support AS path prepending but will allow only AS paths with a length of one AS number. Apply this filter list to all incoming updates from R1 EBGP neighbor.

```
ISP1(config)# ip as-path access-list 10 permit ^[0-9]+$
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor 172.16.11.1 filter-list 10 in
```

A filter list allows only AS paths with a length of one AS number—this number can be any number. Consequently, all routes with a prepended AS path are filtered out.

Step 7 On the ISP1 router, reestablish BGP sessions to enforce the application of filter list.

To reestablish BGP session, use the **clear ip bgp *** command.

```
ISP1# clear ip bgp *
```

Step 8 On the ISP1 router, verify received routes to reach the 10.0.0.0/24 networks in AS 1.

```
ISP1# show ip bgp
BGP table version is 20, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	172.16.100.22			0	200 1 i
*>	10.0.0.16/28	172.16.100.22			0	200 1 i
*>	10.0.0.32/28	172.16.100.22			0	200 1 i
*>	10.0.0.48/28	172.16.100.22			0	200 1 i
*>	10.0.0.64/28	172.16.100.22			0	200 1 i
*>	10.0.1.0/28	0.0.0.0	0		32768	i
*>	10.0.1.16/28	0.0.0.0	0		32768	i
*>	10.0.1.32/28	0.0.0.0	0		32768	i
*>	10.0.1.48/28	0.0.0.0	0		32768	i
*>	10.0.1.64/28	0.0.0.0	0		32768	i
*>	10.0.1.80/28	0.0.0.0	0		32768	i
*>	10.0.2.0/28	172.16.100.22	0		0	200 i
*>	10.0.2.16/28	172.16.100.22	0		0	200 i
*>	10.0.2.32/28	172.16.100.22	0		0	200 i
*>	10.0.2.48/28	172.16.100.22	0		0	200 i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.2.64/28	172.16.100.22	0		0	200 i
*>	10.0.2.80/28	172.16.100.22	0		0	200 i
*>	10.0.2.96/28	172.16.100.22	0		0	200 i
*>	10.0.2.112/28	172.16.100.22	0		0	200 i
*>	10.0.2.128/28	172.16.100.22	0		0	200 i

ISP1 now received only one route to 10.0.0.0/24 networks, the one from ISP2 (next hop 172.16.100.22). Since R1 is announcing the networks 10.0.0.0/24 with a prepended AS path, the filter-list for incoming routes to AS 100 filtered those routes out. This filtering results in a situation where the networks 10.0.0.0/24 are not reachable over the link between R1 and ISP1 (AS 1 and AS 100). Therefore, the backup function is not available.

Networks 10.0.0.0/24 are, however, still reachable via the path going through AS 200.

Step 9 If you want to have a backup path available to 10.0.0.0/24 networks, you have to meet an agreement with a customer and allow AS path prepending in your filters.

On the ISP1 router, modify incoming filter to allow one or multiple copies of the same AS number. Configure a filter for all routes that are received directly from AS 1.

```
ISP1(config)# ip as-path access-list 10 permit ^1(_1)*$
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor 172.16.11.1 filter-list 10 in
```

You configured an individual filter for all routes that are received directly from AS 1. The AS path is required to start with 1, then multiple copies of 1 may follow it.

If you had more customers, they would all require an individual filter list, because the customer is explicitly indicated in the regular expression. An alternative would be to implement the AS path filter that allows any AS path containing one or multiple copies of the same AS number. For example using `^(0-9|+)(_1)*$` filter, that matches any AS path beginning with any AS number and continues with no or multiple repetitions of the same AS number.

Step 10 On the ISP1 router, reestablish BGP sessions to enforce the application of new filter list.

To re-establish BGP session, use the **clear ip bgp *** command.

```
ISP1# clear ip bgp *
```

Step 11 On the ISP1 router, verify new received routes to reach the 10.0.0.0/24 networks in AS 1.

```
ISP1# show ip bgp
```

```
BGP table version is 20, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0/28	172.16.100.22			0	200 1 i
*	172.16.11.1			0	1 1 1 1 1 1 i
*> 10.0.0.16/28	172.16.100.22			0	200 1 i
*	172.16.11.1			0	1 1 1 1 1 1 i
*> 10.0.0.32/28	172.16.100.22			0	200 1 i
*	172.16.11.1			0	1 1 1 1 1 1 i
*> 10.0.0.48/28	172.16.100.22			0	200 1 i
*	172.16.11.1			0	1 1 1 1 1 1 i
*> 10.0.0.64/28	172.16.100.22			0	200 1 i
*	172.16.11.1			0	1 1 1 1 1 1 i
*> 10.0.1.0/28	0.0.0.0	0		32768	i
*> 10.0.1.16/28	0.0.0.0	0		32768	i
*> 10.0.1.32/28	0.0.0.0	0		32768	i
*> 10.0.1.48/28	0.0.0.0	0		32768	i
*> 10.0.1.64/28	0.0.0.0	0		32768	i
*> 10.0.1.80/28	0.0.0.0	0		32768	i
Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.2.0/28	172.16.100.22	0		0	200 i
*> 10.0.2.16/28	172.16.100.22	0		0	200 i
*> 10.0.2.32/28	172.16.100.22	0		0	200 i
*> 10.0.2.48/28	172.16.100.22	0		0	200 i
*> 10.0.2.64/28	172.16.100.22	0		0	200 i
*> 10.0.2.80/28	172.16.100.22	0		0	200 i
*> 10.0.2.96/28	172.16.100.22	0		0	200 i
*> 10.0.2.112/28	172.16.100.22	0		0	200 i
*> 10.0.2.128/28	172.16.100.22	0		0	200 i

ISP1 again received two routes to 10.0.0.0/24 networks:

- The first route is received from ISP2 (next hop 172.16.100.22), with the AS path 200 1.
- The second route is received from R1 (next hop 172.16.11.1), with the AS path 1 1 1 1 1 1.

As weight and local preference value are left on default, ISP1 chooses the path that is received from ISP2 as the best based on the shortest AS path length.

BGP Hide Local-Autonomous System

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. However changing of the AS number in a BGP network can be difficult. During the transition, [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. It also ensures that routes can be propagated throughout the AS, while the AS number transition is incomplete.

BGP Hide Local-Autonomous System

- 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

36

To allow customization of the AS number for [EBGP](#) peer groupings, use the **neighbor local-as** command in address family or router configuration mode. To disable this function, use the **no** form of this command.

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

Syntax Description

Parameter	Description
<i>ip-address</i>	IP address of the local BGP-speaking neighbor
<i>peer-group-name</i>	Name of a BGP peer group.
<i>as-number</i>	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. 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 that were discussed in this lesson.

Summary

- If you do not configure the preferred path for incoming (return) traffic, the likely result is an asymmetrical traffic flow as well as suboptimal performance of the return traffic.
- AS path prepending is performed on outgoing EGP updates over the undesired return path. Configure it using a route-map with the **set as-path prepend** command.
- Monitor AS path prepending on the router that is receiving the prepended routes because the prepended path will not be visible on the prepending router.
- If you are a service provider with customers that use AS path prepending, you must create new AS path filters 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. It also ensures that routes can be propagated throughout the AS, while the AS number transition is incomplete.

© 2015 Cisco and/or its affiliates. All rights reserved.

37

Lesson 4: Understanding BGP Multi-Exit Discriminators

Overview

When connections to multiple providers are required, it is important that [BGP](#) selects the optimum route for traffic to use. It is equally important that the selected return path is the optimum return path into the [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 [MED](#) attribute.

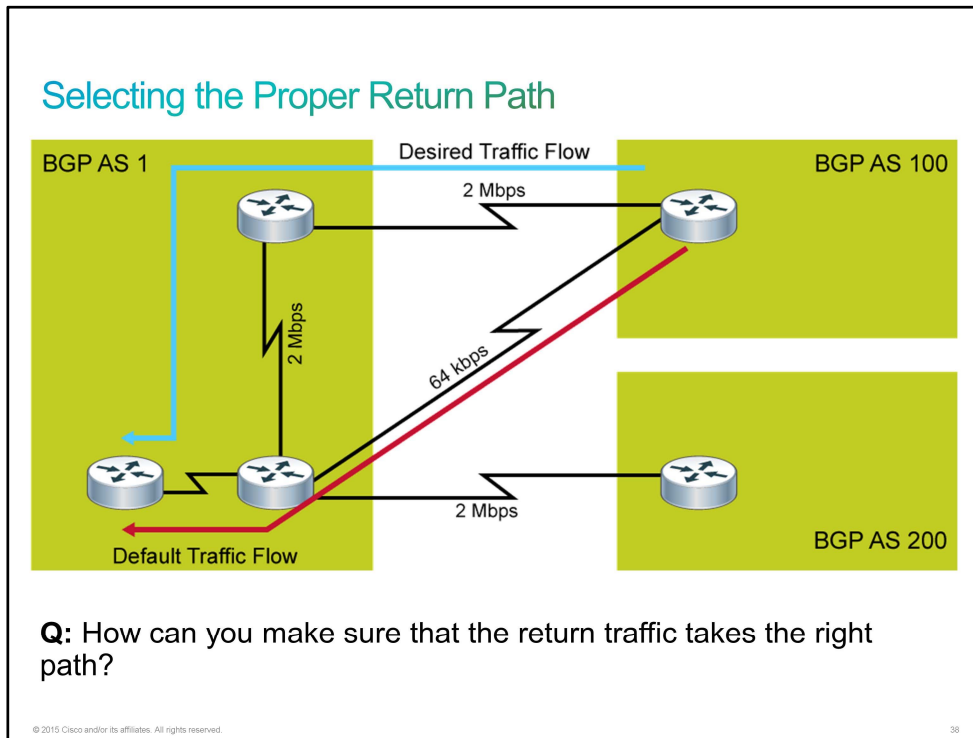
In this lesson, you will learn how to influence BGP route selection by setting the BGP MED attribute of outgoing BGP routes. Two methods that are used to set the MED attribute, the default MED and route maps, are discussed. In addition, you will also learn the basic MED attribute configuration, advanced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

Upon completing this lesson, you will be able to:

- 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 BGP MED
- Describe how to configure and monitor BGP MED
- Identify the Cisco IOS commands that are required to troubleshoot the BGP MED configurations
- Identify the Cisco IOS commands that are required to configure advanced MED features on Cisco routers

Selecting the Proper Return Path

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](#).



More complex half of the problem is how to influence neighboring autonomous systems to choose the correct return path back into the AS.

Multi-Exit Discriminator

The [MED](#) attribute is a hint to external neighbors about the preferred path into an AS when multiple entry points exist.

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.

© 2015 Cisco and/or its affiliates. All rights reserved.

39

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.

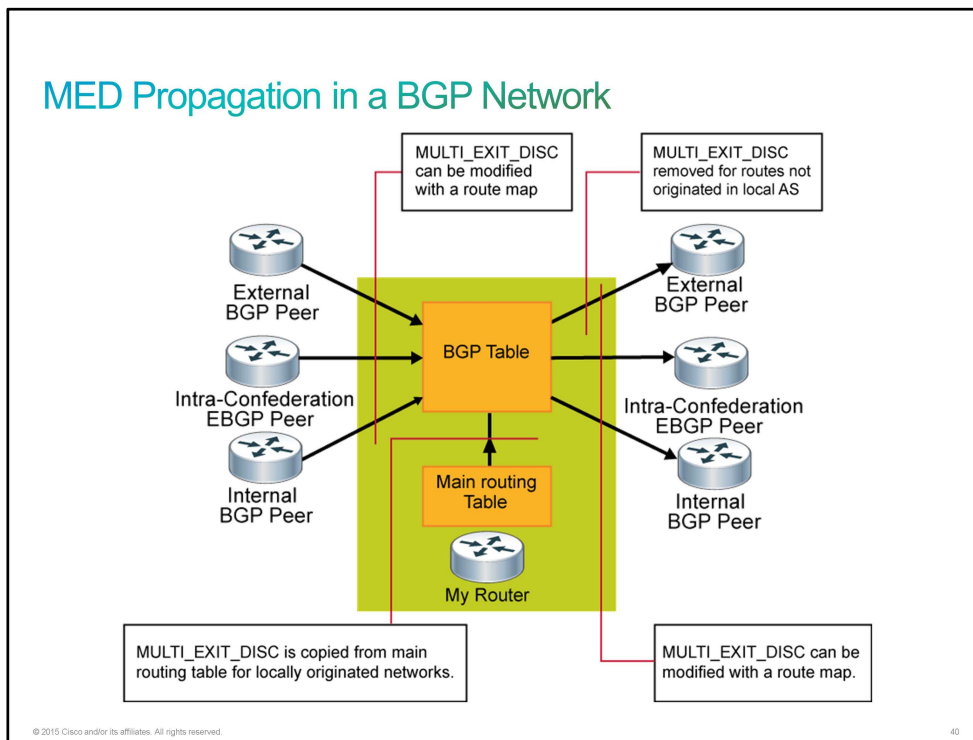
The default value of the MED attribute is 0. A lower value of the MED attribute indicates a more preferred path.

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

The figure shows how the value of the [MED](#) attribute is assigned, depending upon the routing information source.



You must configure a route map 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 MED

The **MED** is not a mandatory attribute, and there no MED attribute 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.

Changing MED

You have two options, either you change the default MED or you change the MED with route map.

- Option 1:

```
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.

© 2015 Cisco and/or its affiliates. All rights reserved.

41

Changing MED (Cont.)

- Option 2:

```
router(config)#  
route-map name permit sequence  
  match condition  
  set metric value
```

- Changes MED for routes that are 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

42

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

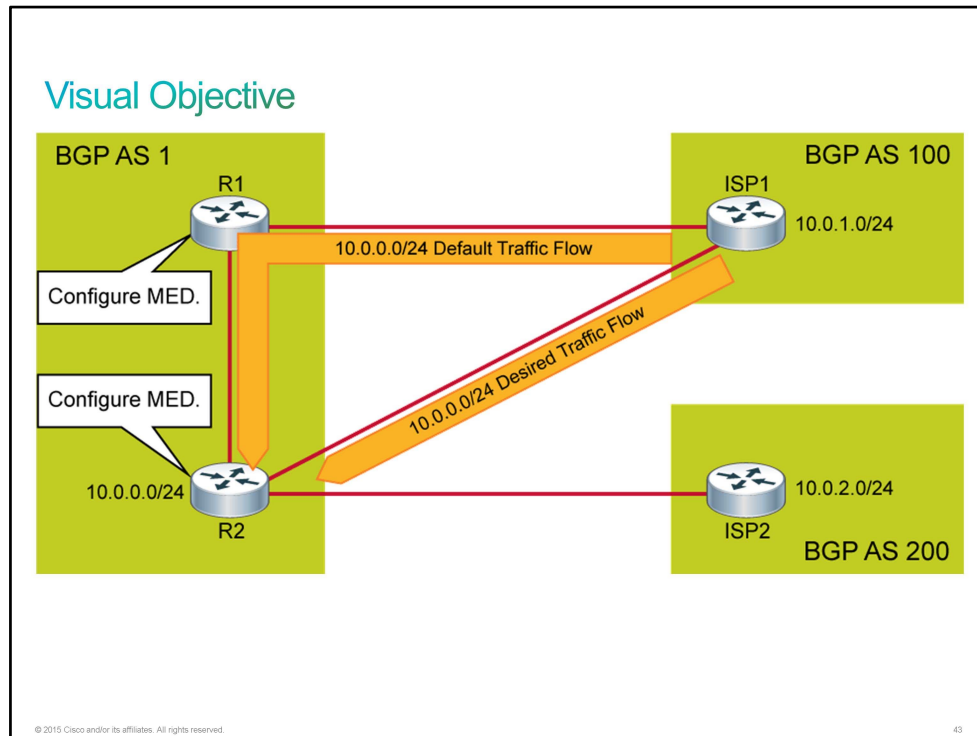
To set the default metric (MED) value for BGP routes, use the **default-metric *number*** command in router configuration mode. *number* is a default metric value that is appropriate for the specified routing protocol. To return to the default state, use the **no** form of this command.

You can also 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.

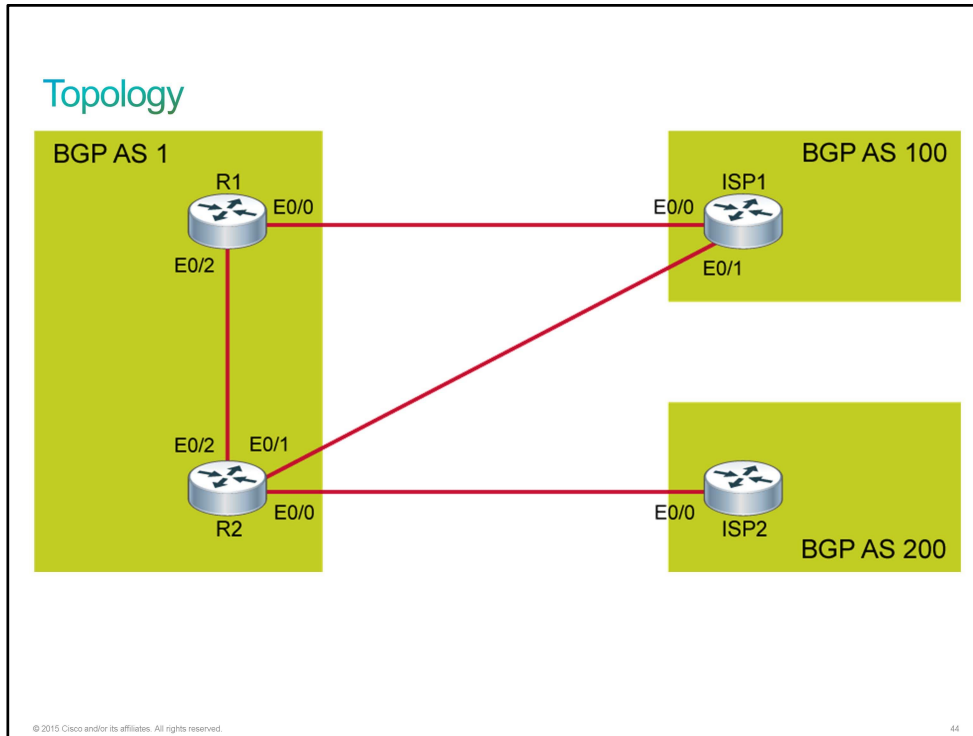
Discovery 15: Configure MED

Overview

Through this discovery, you will learn how to configure [MED](#) using route maps to influence the return path from [AS 100](#) to 10.0.0.0/24 network in AS 1.



Topology



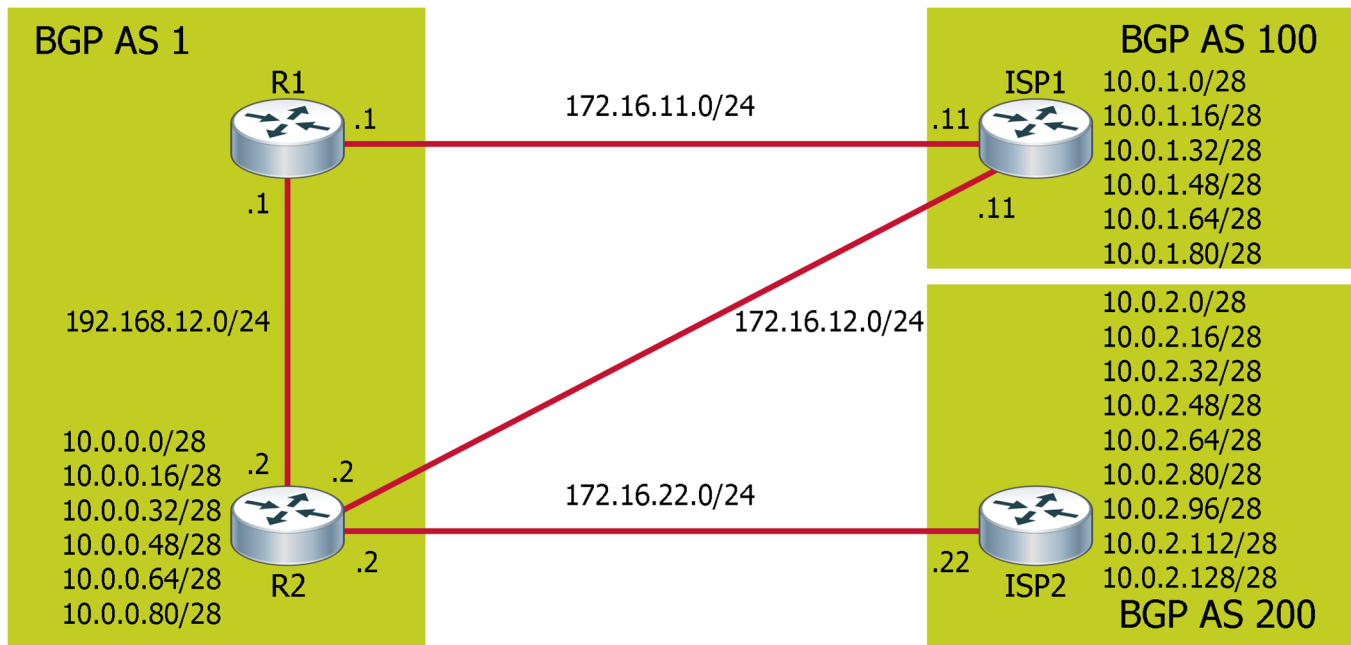
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R1	Ethernet 0/0	172.16.11.1/24	Connection to ISP1
R1	Ethernet 0/2	192.168.12.1/24	Connection to R2
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/1	172.16.12.2/24	Connection to ISP1
R2	Ethernet 0/2	192.168.12.2/24	Connection to R1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/0	172.16.11.11/24	Connection to R1
ISP1	Ethernet 0/1	172.16.12.11/24	Connection to R2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in [BGP](#) are preconfigured as shown below:



BGP is also preconfigured:

- [EBGP](#)
 - R1 to ISP1
 - R2 to ISP1
 - R2 to ISP2
- [IBGP](#)
 - R1 to R2

Configure MED

Discovery Steps

Step 1 On the ISP1 router, re-establish the BGP session.

To re-establish BGP session, use the **clear ip bgp *** command.

```
ISP1# clear ip bgp *
```

Note: Because of the limitation of IOL, you have to reestablish the BGP session on ISP1 to get the same results as described in the next step of the lab.

Step 2 On the ISP1 router, verify the initial best path to reach 10.0.0.0/24 network.

```
ISP1# show ip bgp
BGP table version is 20, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0/28	172.16.11.1				0 1 i
*	172.16.12.2	0			0 1 i
*> 10.0.0.16/28	172.16.11.1				0 1 i
*	172.16.12.2	0			0 1 i
*> 10.0.0.32/28	172.16.11.1				0 1 i
*	172.16.12.2	0			0 1 i
*> 10.0.0.48/28	172.16.11.1				0 1 i
*	172.16.12.2	0			0 1 i
*> 10.0.0.64/28	172.16.11.1				0 1 i
*	172.16.12.2	0			0 1 i
*> 10.0.1.0/28	0.0.0.0	0		32768	i
*> 10.0.1.16/28	0.0.0.0	0		32768	i
*> 10.0.1.32/28	0.0.0.0	0		32768	i
*> 10.0.1.48/28	0.0.0.0	0		32768	i
*> 10.0.1.64/28	0.0.0.0	0		32768	i
*> 10.0.1.80/28	0.0.0.0	0		32768	i
Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.2.0/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.16/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.32/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.48/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.64/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.80/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.96/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.112/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i
*> 10.0.2.128/28	172.16.11.1				0 1 200 i
*	172.16.12.2				0 1 200 i

ISP1 has received 10.0.0.0/24 networks from two different neighbors:

- The first path was received from R1 (next hop 172.16.11.1).
- The second path was received from R2 (next hop 172.16.12.2).

ISP1 chooses the first path (the one received from R1) as the best path.

Step 3 On the ISP1 router, check why was the route that was received from R1 chosen as the best path to 10.0.0.0/24 networks.

```
ISP1# show ip bgp
BGP table version is 20, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0/28	172.16.11.1			0 1	i
*	172.16.12.2	0		0 1	i
*> 10.0.0.16/28	172.16.11.1			0 1	i
*	172.16.12.2	0		0 1	i
*> 10.0.0.32/28	172.16.11.1			0 1	i
*	172.16.12.2	0		0 1	i
*> 10.0.0.48/28	172.16.11.1			0 1	i
*	172.16.12.2	0		0 1	i
*> 10.0.0.64/28	172.16.11.1			0 1	i
*	172.16.12.2	0		0 1	i

<... output omitted ...>

If you follow the route selection criteria, you can determine that:

- The weight is not configured for any neighbor.
- The local preference is not configured for any neighbor.
- Routes do not originate locally.
- Both routes have the same length of AS path.
- Both routes are [EGP](#) routes.

Since both paths are EBGp paths, the oldest (most stable) path is preferred. Next thing that you have to verify is the duration of session between ISP1 and R1, and ISP1 and R2.

```
ISP1# show bgp summary
BGP router identifier 10.0.1.81, local AS number 100
BGP table version is 20, main routing table version 20
19 network entries using 2812 bytes of memory
32 path entries using 2048 bytes of memory
4/3 BGP path/bestpath attribute entries using 544 bytes of memory
2 BGP AS-PATH entries using 48 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
BGP using 5452 total bytes of memory
BGP activity 95/76 prefixes, 160/128 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
172.16.11.1	4	1	25	25	20	0	0	00:17:40	13
172.16.12.2	4	1	28	25	20	0	0	00:17:40	13

The duration of both session is the same. So obviously, the router-ID breaks the tie—the path from router with the lower BGP router-ID is preferred.

```
ISP1# show ip bgp neighbor 172.16.11.1
BGP neighbor is 172.16.11.1, remote AS 1, external link
  BGP version 4, remote router ID 10.0.7.1
  BGP state = Established, up for 00:21:49
  Last read 00:00:42, last write 00:00:54, hold time is 180, keepalive
<... output omitted ...>
```

```
ISP1# show ip bgp neighbor 172.16.12.2
BGP neighbor is 172.16.12.2, remote AS 1, external link
  BGP version 4, remote router ID 10.0.7.17
  BGP state = Established, up for 00:23:40
  Last read 00:00:44, last write 00:00:50, hold time is 180, keepalive
<... output omitted ...>
```

The router-ID of R1 is lower than the router-ID of R2, so the route received from R1 is preferred.

Step 4 On the customer side, influence the returning traffic flow to 10.0.0.0/24 networks, so the ISP1 router will take the path via R2 router instead the default one. Reconfigure MED using route maps.

Configure route map "MED" on both routers, R1 and R2, to prefer the lower link to AS 1.

```
R1(config)# route-map MED
R1(config-route-map)# set metric 500
```

```
R2(config)# route-map MED
R2(config-route-map)# set metric 100
```

Lower MED value is preferred when choosing the best path. If you want R2 to be chosen for the returning traffic, it has to be configured with lower MED.

Step 5 On the R1 and R2 routers, apply route maps for ISP1 neighbor in the outbound direction.

```
R1(config)# router bgp 1
R1(config-router)# neighbor 172.16.11.11 route-map MED out
```

```
R2(config)# router bgp 1
R2(config-router)# neighbor 172.16.12.11 route-map MED out
```

Do not forget that the solution of influencing the returning traffic flow, 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 1.

Step 6 On the ISP1 router, reestablish BGP sessions to enforce the application of MED.

To reestablish BGP session, use the **clear ip bgp** command.

```
ISP1# clear ip bgp *
```

Step 7 On the ISP1 router, verify new best path to reach 10.0.0.0/24 network.

```
ISP1# show ip bgp
BGP table version is 46, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.0.0/28	172.16.11.1	500			0 1 i
*>		172.16.12.2	100			0 1 i
*	10.0.0.16/28	172.16.11.1	500			0 1 i
*>		172.16.12.2	100			0 1 i
*	10.0.0.32/28	172.16.11.1	500			0 1 i
*>		172.16.12.2	100			0 1 i
*	10.0.0.48/28	172.16.11.1	500			0 1 i
*>		172.16.12.2	100			0 1 i
*	10.0.0.64/28	172.16.11.1	500			0 1 i
*>		172.16.12.2	100			0 1 i
*>	10.0.1.0/28	0.0.0.0	0		32768	i
*>	10.0.1.16/28	0.0.0.0	0		32768	i
*>	10.0.1.32/28	0.0.0.0	0		32768	i
*>	10.0.1.48/28	0.0.0.0	0		32768	i
*>	10.0.1.64/28	0.0.0.0	0		32768	i
*>	10.0.1.80/28	0.0.0.0	0		32768	i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.2.0/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.16/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.32/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.48/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.64/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.80/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.96/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.112/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i
*	10.0.2.128/28	172.16.11.1	500			0 1 200 i
*>		172.16.12.2	100			0 1 200 i

ISP1 has received 10.0.0.0/24 networks from two different neighbors:

- The first path was received from R1 (next hop 172.16.11.1), with the MED set to 500.
- The second path was received from R2 (next hop 172.16.12.2), with the MED set to 100.

ISP1 chooses the path that was received from R2 as the best path, since it has lower MED configured.

Note that MED is displayed in the **show ip bgp** printout as the metric field. All BGP-related **show** and **debug** commands display the value of the MED attribute. If the MED is not configured, the metric field is left blank. When looking at detailed information for a specific network (via the **show ip bgp prefix** command), you will see the MED only if the attribute exists.

Troubleshooting the MED

If debugging is necessary to troubleshoot a problem, the [MED](#), among other attributes, is displayed in the outputs.

Troubleshooting the MED

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

```
R1# debug ip bgp updates
BGP updates debugging is on for address family: IPv4 Unicast
R1# clear ip bgp 172.16.11.11 soft out
R1#
*Mar  9 01:58:12.497: BGP(0): (base) 172.16.11.11 send UPDATE (format)
10.0.0.0/28, next 172.16.11.1, metric 500, path Local
*Mar  9 01:58:12.497: BGP(0): (base) 172.16.11.11 send UPDATE (format)
10.0.0.16/28, next 172.16.11.1, metric 500, path Local
*Mar  9 01:58:12.497: BGP(0): (base) 172.16.11.11 send UPDATE (format)
10.0.0.32/28, next 172.16.11.1, metric 500, path Local
```

© 2015 Cisco and/or its affiliates. All rights reserved.

45

The MED attribute was set with an outgoing route map. This MED attribute ("metric 500") is then sent to the neighbor in the update.

Troubleshooting the MED (Cont.)

- MED received from a neighbor is displayed in debugging outputs.

```
ISP1# debug ip bgp updates
BGP updates debugging is on for address family: IPv4 Unicast
ISP1# clear ip bgp 172.16.11.1
ISP1#
*Mar  9 02:31:35.697: BGP(0): 172.16.11.1 rcvd UPDATE w/ attr: nexthop
172.16.11.1, origin i, metric 500, merged path 1, AS_PATH
*Mar  9 02:31:35.697: BGP(0): 172.16.11.1 rcvd 10.0.0.0/28
*Mar  9 02:31:35.697: BGP(0): 172.16.11.1 rcvd 10.0.0.16/28
*Mar  9 02:31:35.698: BGP(0): 172.16.11.1 rcvd 10.0.0.32/28
```

© 2015 Cisco and/or its affiliates. All rights reserved.

46

This debugging example shows the received updates for 10.0.0.0/28, 10.0.0.0/16, and 10.0.0.32/28 networks. The MED attribute ("metric 500") for all three networks is shown in the received update. The MED was set on the neighboring router.

Troubleshooting the MED (Cont.)

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

```
ISP1# show ip bgp neighbor 172.16.11.1 received-routes
BGP table version is 59, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

      Network          Next Hop           Metric LocPrf Weight Path
*    10.0.0.0/28       172.16.11.1           500                0 1 i
*    10.0.0.16/28      172.16.11.1           500                0 1 i
*    10.0.0.32/28      172.16.11.1           500                0 1 i

Total number of prefixes 3
```

© 2015 Cisco and/or its affiliates. All rights reserved.

47

You can change the MED of the received route on the receiving router (on ISP1 for this example). 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.

In the example, the ISP1 router received three networks from 172.16.11.1 neighbor—10.0.0.0/28, 10.0.0.16/28, and 10.0.0.32/28. All three networks have the metric set to 500.

Troubleshooting the MED (Cont.)

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

```
ISP1# show ip bgp 10.0.0.0/28
BGP routing table entry for 10.0.0.0/28, version 56
Paths: (3 available, best #1, table default)
  Advertised to update-groups:
    5
  Refresh Epoch 2
    1
      172.16.11.1 from 172.16.11.1 (10.0.7.1)
        Origin IGP, metric 50, localpref 100, valid, external, best
  Refresh Epoch 2
    1, (received-only)
      172.16.11.1 from 172.16.11.1 (10.0.7.1)
        Origin IGP, metric 500, localpref 100, valid, external
```

© 2015 Cisco and/or its affiliates. All rights reserved.

48

Troubleshooting the MED (Cont.)

- Only the modified route is displayed in the BGP table.

```
ISP1# show ip bgp
BGP table version is 163, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.0.0/28	172.16.11.1	50		0	1 i
*>	10.0.0.16/28	172.16.11.1	50		0	1 i
*>	10.0.0.32/28	172.16.11.1	50		0	1 i

© 2015 Cisco and/or its affiliates. All rights reserved.

49

If you configured the route map to influence the MED of received route, you can see the original updates to the MED attribute by using the **show ip bgp prefix** command. Soft reconfiguration has to be enabled. The original versions are marked with the **received-only** keyword and follow the version that is in the global [BGP](#) table.

In this example, the received update had MED attribute set to 500, but a value of 50 was later applied through an inbound route map.

Advanced MED Configuration

Several rules exist on when and how you should use the [MED](#) attribute.

Advanced MED Configuration

By default, the MED is considered only during selection of routes from the same AS.

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

- With **always-compare-med**, the MED is also considered for routes coming from a different AS.

If the MED is not attached to a BGP route, it is interpreted as value 0, and thus as the best metric.

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

- With this command, a missing MED is interpreted as infinity (worst).

© 2015 Cisco and/or its affiliates. All rights reserved.

50

1. 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.
2. 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.)

By default, the MED is considered only during selection of routes from the same AS, which does not include intra-confederation autonomous systems.

```
router(config-router)# bgp bestpath med confed
```

- Use this command to allow routers to compare paths that are learned from confederation peers.

```
router(config-router)# bgp deterministic-med
```

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

© 2015 Cisco and/or its affiliates. All rights reserved.51

3. 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.
4. 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 [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.

Advanced MED Configuration Example

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 one 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 has been 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. Then it 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.

Summary

This topic summarizes the key points that were 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.
- There is no MED attribute that is attached to a route by default.
 - You can set the default metric value (MED) for BGP routes, using 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.
- You can configure advanced MED parameters to modify the default MED behaviors.
 - Use the **bgp always-compare-med** to force the router to compare the MED even if paths came from different autonomous systems.
 - Use the **bgp bestpath med confed** command when you use the MED within a confederation to influence the route selection process

© 2015 Cisco and/or its affiliates. All rights reserved.

52

Lesson 5: Addressing BGP Communities

Overview

When connections to multiple [ISPs](#) are required, it is important that [BGP](#) select the optimum route for traffic to use. It is equally important that the selected return path is the optimum return path into the [AS](#). BGP provides tools for administrators to influence route selection and BGP community attribute is one such tool.

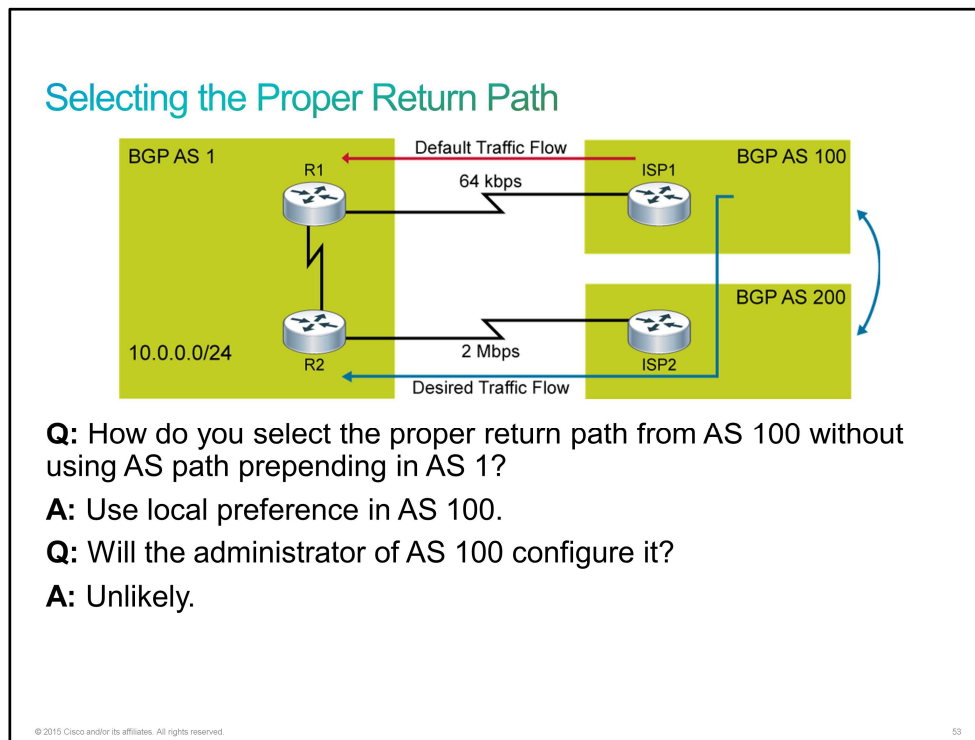
You will learn how to influence BGP route selection by setting the BGP community attribute on outgoing BGP routes to facilitate proper return path selection. You will also learn about the configuration details of BGP communities.

Upon completing this lesson, you will be able to:

- 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 deploy BGP communities
- 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

Selecting the Proper Return Path

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 100 would, therefore, receive networks from AS 1 directly with the MED attribute but would receive networks without a MED attribute from AS 200. In any case, BGP route selection would be based on the length of the AS path. 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 100. The problem with this solution is that service providers normally do not rush to implement every wish that their customers might have.

In this lesson, you will learn a solution to this case study that uses the transitive optional attribute called "BGP community" in conjunction with local preference.

BGP Communities Overview

BGP communities are attributes that are used to group and filter routes. Communities are designed to give you 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.

BGP Communities Overview

- BGP communities are a mean 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

54

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.

© 2015 Cisco and/or its affiliates. All rights reserved.

55

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. This way, the community attribute allows 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 to intra-confederation external neighbors. The **no-export** attribute 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]

© 2015 Cisco and/or its affiliates. All rights reserved.

56

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 that are separated by a colon. The first number (high-order 16 bits) should be the AS number of the AS that defines the community value. 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).

You can also use communities 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

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.

© 2015 Cisco and/or its affiliates. All rights reserved.

57

Designing a [BGP](#) solution around BGP communities usually requires the following steps:

1. Define administrative policy goals that you need to implement.
2. Define the filters and route selection policy that will achieve the required goals.
3. Assign a community value to each goal.
4. Apply communities on incoming updates from neighboring autonomous systems or tell the neighbors to set the communities themselves.
5. Enable community distribution throughout your [AS](#) to allow community propagation.
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.

Using Communities Example

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

Using Communities Example

- Define administrative policy goal:
 - 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 100:17 is used to indicate that the local preference of the route should be lowered to 50.

© 2015 Cisco and/or its affiliates. All rights reserved.

58

This table lists the goals and the community values.

Goal	Community Value
Set local preference of 50.	100:17
Set local preference of 150.	100:18
Prepend the AS path once when sending the network to external neighbors.	100:21
Prepend the AS path twice when sending the network to external neighbors.	100:22
Prepend the AS path three times when sending the network to external neighbors.	100:23

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

Several activities 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

59

To configure BGP communities on customers side, you first have to configure a route map in which you will set the community. Also the community propagation to BGP neighbor has to be configured, otherwise the community values attached to outgoing BGP updates will be striped out.

On the other side, service provider has to define BGP community lists to match routes based on attached BGP communities. Route maps have to be used to filter or modify BGP routing updates based on matching on community lists in incoming or outgoing updates.

1. Configuring Route Tagging with BGP Communities

First step is to set communities, which requires a route map.

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.
- Communities that are specified in the **set** keyword overwrite existing communities unless you specify the **additive** option.

```
router(config-router)# neighbor ip-address route-map map in | out
```

- You have to apply a route map to inbound or outbound BGP updates.

```
router(config-router)# redistribute protocol route-map map
```

- You can apply a route map to redistributed routes

© 2015 Cisco and/or its affiliates. All rights reserved.

60

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 you use the keyword **additive**, 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.

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.

2. Configuring Community Propagation

Second step is to enabling community propagation per neighbor for all internal neighbors. If communities are sent to external neighbors, you must enable community propagation for external 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 many neighbors.

© 2015 Cisco and/or its affiliates. All rights reserved.

61

A command that network administrators commonly forget 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.

3. Defining BGP Community Lists

On a service provider side, community lists have to be created to be used within route maps to match on community values.

Defining BGP Community Lists

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

- This command defines a simple community list.
- You can use the keyword **internet** to match any community.

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

- This command defines an expanded community list.
- Communities that are attached to a route are ordered, converted to string, and matched with regular expression.
- Use "." to match any community value.

© 2015 Cisco and/or its affiliates. All rights reserved.

62

You can use a **standard community 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.

Comparison of Community Lists to Standard IP Access Lists

Similarities	Differences
The router evaluates the lines in the community-list sequentially.	The keyword internet should be used to permit any community value.
If no line matches communities that are attached to a BGP route, the route is implicitly denied.	If more values are listed in a single line, they all have to be in an update to produce a match.

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:
 - "10.0.0.0/24, 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 ".").
5. Use regular expression "." to permit any community.

4. Matching BGP Communities with Route Maps

On a service provider side, route maps where community lists are used to match on community values have to be created. 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).

These route maps are then applied to incoming or outgoing updates.

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

© 2015 Cisco and/or its affiliates. All rights reserved.63

Use route maps to match the networks that carry a subset of communities that the community list permits. Other parameters or attributes can then be set based on community values. A route map with a community list matches a route if at least some communities that are attached to the route match the community list. 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).

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

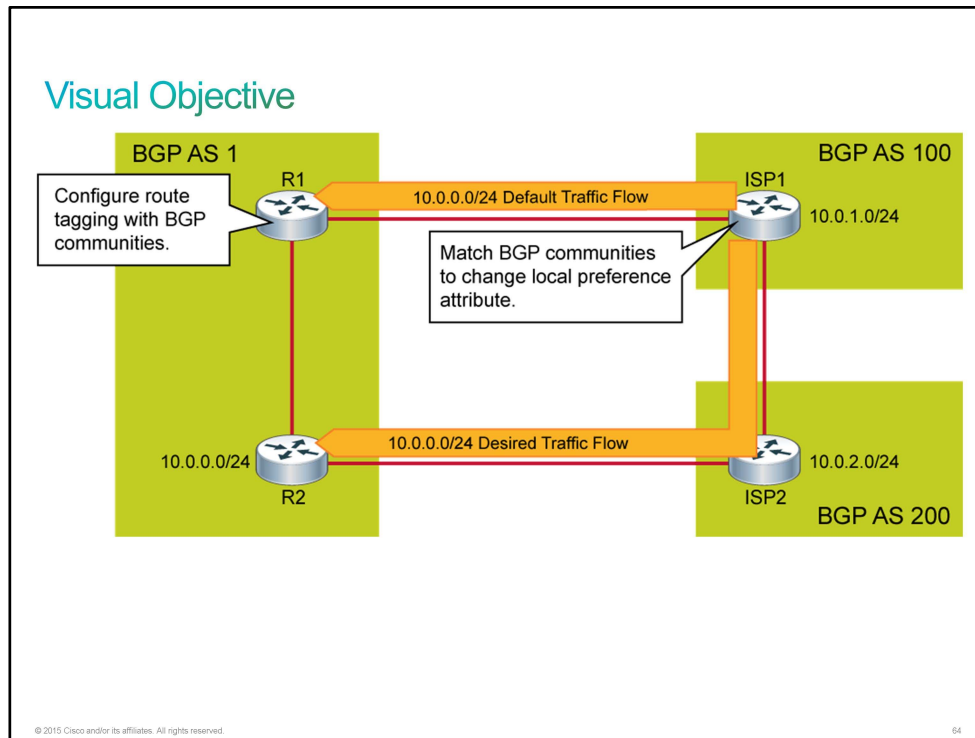
- **no-advertise:** If a router receives an update carrying this community, the router will not forward it to any neighbor.
- **local-as:** This community has a similar meaning to **no-export**, but it keeps a route within the local subautonomous system. The route is not propagated to intra-confederation 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.

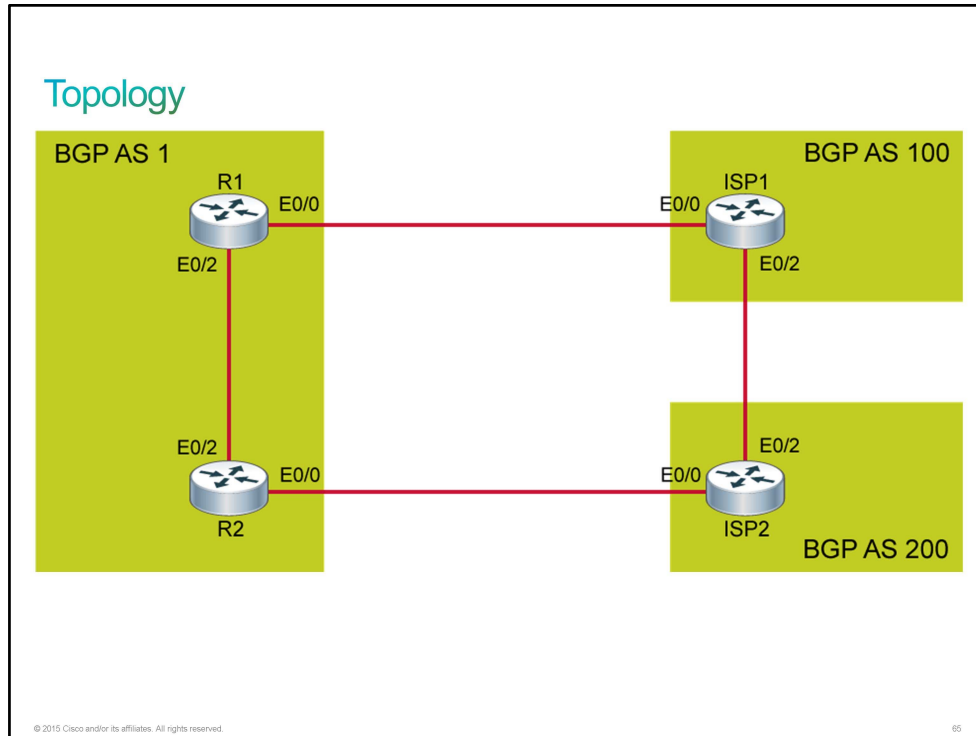
Discovery 16: Configure Local Preference Using the Communities

Overview

In this discovery, you will influence the returning path. You will configure local preference through the communities. You will do the configuration on the customer and on the service provider side.



Topology



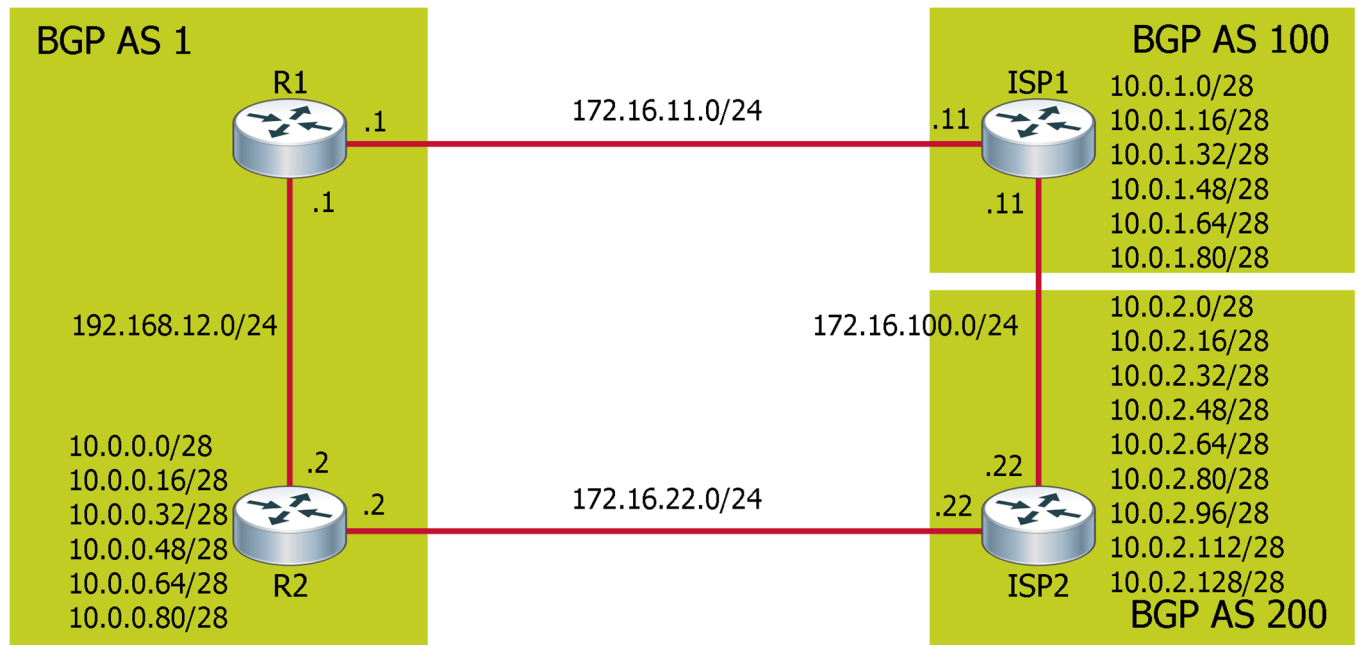
Job Aids

Note If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

Device Information

Device	Interface	IP address	Description
R1	Ethernet 0/0	172.16.11.1/24	Connection to ISP1
R1	Ethernet 0/2	192.168.12.1/24	Connection to R2
R2	Ethernet 0/0	172.16.22.2/24	Connection to ISP2
R2	Ethernet 0/2	192.168.12.2/24	Connection to R1
R2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5	10.0.0.1/28 10.0.0.17/28 10.0.0.33/28 10.0.0.49/28 10.0.0.65/28	Loopbacks simulate LAN networks
ISP1	Ethernet 0/0	172.16.11.11/24	Connection to R1
ISP1	Ethernet 0/2	172.16.100.11/24	Connection to ISP2
ISP1	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6	10.0.1.1/28 10.0.1.17/28 10.0.1.33/28 10.0.1.49/28 10.0.1.65/28 10.0.1.81/28	Loopbacks simulate LAN networks
ISP2	Ethernet 0/0	172.16.22.22/24	Connection to R2
ISP2	Ethernet 0/2	172.16.100.22/24	Connection to ISP1
ISP2	Loopback 1 Loopback 2 Loopback 3 Loopback 4 Loopback 5 Loopback 6 Loopback 7 Loopback 8 Loopback 9	10.0.2.1/28 10.0.2.17/28 10.0.2.33/28 10.0.2.49/28 10.0.2.65/28 10.0.2.81/28 10.0.2.97/28 10.0.2.113/28 10.0.2.129/28	Loopbacks simulate LAN networks

[IP addresses](#) and advertised networks in [BGP](#) are preconfigured as shown below:



BGP is also preconfigured:

- [EBGP](#)
 - R1 to ISP1
 - R2 to ISP2
 - ISP1 to ISP2
- [IBGP](#)
 - R1 to R2

Configure Local Preference Using the Communities

Discovery Steps

Step 1 On the ISP1 router, verify initial best path to reach the 10.0.0.0/24 networks in AS 1.

```
ISP1# show ip bgp
BGP table version is 31, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.0.0/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.16/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.32/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.48/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*	10.0.0.64/28	172.16.100.22			0	200 1 i
*>		172.16.11.1			0	1 i
*>	10.0.1.0/28	0.0.0.0	0		32768	i
*>	10.0.1.16/28	0.0.0.0	0		32768	i
*>	10.0.1.32/28	0.0.0.0	0		32768	i
*>	10.0.1.48/28	0.0.0.0	0		32768	i
	Network	Next Hop	Metric	LocPrf	Weight	Path
*>	10.0.1.64/28	0.0.0.0	0		32768	i
*>	10.0.1.80/28	0.0.0.0	0		32768	i
*	10.0.2.0/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.16/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.32/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.48/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.64/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.80/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.96/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.112/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i
*	10.0.2.128/28	172.16.11.1			0	1 200 i
*>		172.16.100.22	0		0	200 i

ISP1 received two routes to 10.0.0.0/24 networks:

- First route is received from ISP2 (next hop 172.16.100.22)
- Second route is received from R1 (next hop 172.16.11.1),

Since weight and local preference attributes are left on default values, ISP1 chooses the path with shortest [AS](#) path length as the best (the route that is received from R1).

If you want to influence the returning traffic to AS 1 from AS 100 to go via AS 200 (and not directly), you can do it by using community lists in AS 1. In the update from that AS, you can send a community list. This community list can then influence the value of, for example, local preference in AS 100 for all routes that are received from AS 1.

Step 2 Configure community list to influence the returning traffic path to your AS (AS 1). Use community for changing the local preference value on ISP's side. First do the customer's side of the configuration and configure route map on the R1 router to tag routes with BGP communities.

Name the route map "SetCom." Select all BGP routes and set community value of "100:17." You have already agreed with a service provider, that "100:17" community will set local preference of 50 on a service provider's side. Since the default local preference is 100, lower local preference (50) means that the route will not be chosen as the best one.

```
R1(config)# route-map SetCom permit 10
R1(config-route-map)# set community 100:17
```

Because you did not enter any **match** statement, all networks are permitted. If you wanted to set communities on specific routes, you could use a standard access list to match against, with the **match ip address** command in the route map.

Step 3 On the R1 router, apply the "SetCom" route map to all outgoing updates to the ISP1 EBGp neighbor.

```
R1(config)# router bgp 1
R1(config-router)# neighbor 172.16.11.11 route-map SetCom out
```

Step 4 Remember, that you have to enable propagation of community attributes to BGP neighbors. Otherwise the router will strip out any community values attached to outgoing updates.

Enable community propagation on the R1 router.

```
R1(config)# router bgp 1
R1(config-router)# neighbor 172.16.11.11 send-community
```

You enabled community propagation from AS 1 to AS 100.

Step 5 Now move on the service provider's side and define BGP community list that will match communities set in AS 1.

On the ISP1 router, configure community list to match communities that were previously set on the R1 router.

```
ISP1(config)# ip community-list 7 permit 100:17
```

Step 6 The last thing that you have to do is to translate received communities from AS 1 into local preference. On the ISP1 router, configure route map to match BGP communities and to change the local preference value of matched communities.

On the ISP1 router, configure route map "SetLocalPref," that will use the configured community list to find a community 100:17 and will set local-preference of 50. Do not forget to permit all other routes, that do not contain the right community, without changing anything in the update.

```
ISP1(config)# route-map SetLocalPref permit 10
ISP1(config-route-map)# match community 7
ISP1(config-route-map)# set local-preference 50
ISP1(config-route-map)# exit
ISP1(config)# route-map SetLocalPref permit 1000
```

Route map uses a community list, that was configured in previous step, to find community 100:17. If the community list matches one of the community attributes, the **set** command is

executed, the local-preference is changed, and the route is permitted.

If the route does not contain the right community, the route is simply permitted by route map statement 1000 without changing anything in the update.

Step 7 On the ISP1 router, apply the configured route map to all incoming updates from R1 EBGP neighbor.

```
ISP1(config)# router bgp 100  
ISP1(config-router)# neighbor 172.16.11.1 route-map SetLocalPref in
```

The result will be that AS 100 will set local preference of 50 to all routes received from AS 1. So, it will prefer the other path (the path via AS 200) to reach AS 1.

Step 8 On the ISP1 router, re-establish BGP session to enforce the application of new route map.

To reestablish BGP session, use the **clear ip bgp** command.

```
ISP1# clear ip bgp *
```

Step 9 On the ISP1 router, verify new received routes to reach the 10.0.0.0/24 networks in AS 1.

```
ISP1# show ip bgp
BGP table version is 21, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.0.0/28	172.16.100.22			0	200 1 i
*	172.16.11.1		50	0	1 i
*> 10.0.0.16/28	172.16.100.22			0	200 1 i
*	172.16.11.1		50	0	1 i
*> 10.0.0.32/28	172.16.100.22			0	200 1 i
*	172.16.11.1		50	0	1 i
*> 10.0.0.48/28	172.16.100.22			0	200 1 i
*	172.16.11.1		50	0	1 i
*> 10.0.0.64/28	172.16.100.22			0	200 1 i
*	172.16.11.1		50	0	1 i
*> 10.0.1.0/28	0.0.0.0	0		32768	i
*> 10.0.1.16/28	0.0.0.0	0		32768	i
*> 10.0.1.32/28	0.0.0.0	0		32768	i
*> 10.0.1.48/28	0.0.0.0	0		32768	i
Network	Next Hop	Metric	LocPrf	Weight	Path
*> 10.0.1.64/28	0.0.0.0	0		32768	i
*> 10.0.1.80/28	0.0.0.0	0		32768	i
*> 10.0.2.0/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.16/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.32/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.48/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.64/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.80/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.96/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.112/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i
*> 10.0.2.128/28	172.16.100.22	0		0	200 i
*	172.16.11.1		50	0	1 200 i

ISP1 again received two routes to 10.0.0.0/24 networks:

- The first route is received from ISP2 (next hop 172.16.100.22), with default local preference of 100. (Remember that the applied default local preference is not displayed.)
- The second route is received from R1 (next hop 172.16.11.1), with changed local preference of 50.

ISP1 chooses the path that is received from ISP2 as the best based on the highest local-preference value.

Monitoring Communities

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.

Monitoring Communities

- Communities are displayed in **show ip bgp prefix** printout.
- Communities are not displayed in debugging outputs.
- Routes in the BGP table that are tagged with a set of communities or routes that match a community list can be displayed.
- All routes in a BGP table that have at least one attached community are displayed in **show ip bgp community** printout.

© 2015 Cisco and/or its affiliates. All rights reserved.

66

If the keyword **community** is included in **show ip bgp** command, all networks that have at least one community attribute are displayed. However, you can view communities only if you use the **show ip bgp prefix** command.

If you use the **show ip bgp community-list clist** command, all networks that the community list permits are listed. Using the **show ip bgp community as:nn [as:nn ...]** command, all routes in a BGP table having all the specified communities that are attached are displayed. If the keyword **exact** is added at the end, only the networks that match exactly are displayed.

Step 10 On the ISP1 router, verify all the networks, that have at least one community attribute.

```
ISP1# show ip bgp community
BGP table version is 21, local router ID is 10.0.1.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*	10.0.0.0/28	172.16.11.1		50	0 1	i
*	10.0.0.16/28	172.16.11.1		50	0 1	i
*	10.0.0.32/28	172.16.11.1		50	0 1	i
*	10.0.0.48/28	172.16.11.1		50	0 1	i
*	10.0.0.64/28	172.16.11.1		50	0 1	i
*	10.0.2.0/28	172.16.11.1		50	0 1 200	i
*	10.0.2.16/28	172.16.11.1		50	0 1 200	i
*	10.0.2.32/28	172.16.11.1		50	0 1 200	i
*	10.0.2.48/28	172.16.11.1		50	0 1 200	i
*	10.0.2.64/28	172.16.11.1		50	0 1 200	i
*	10.0.2.80/28	172.16.11.1		50	0 1 200	i
*	10.0.2.96/28	172.16.11.1		50	0 1 200	i
*	10.0.2.112/28	172.16.11.1		50	0 1 200	i
*	10.0.2.128/28	172.16.11.1		50	0 1 200	i

You can see that all networks, 10.0.0.0/24 and 10.0.2.0/24, received from AS 1 are listed in the output. They are listed because you configured route tagging with BGP communities on R1 in AS 1, for all updates that are sent to AS 100.

Step 11 On the ISP1 router, verify the community attribute for one of the 10.0.0.0/24 networks, the 10.0.0.0/28 network.

```
ISP1# show ip bgp 10.0.0.0/28
BGP routing table entry for 10.0.0.0/28, version 2
Paths: (2 available, best #1, table default)
  Advertised to update-groups:
    3
  Refresh Epoch 2
  200 1
    172.16.100.22 from 172.16.100.22 (10.0.2.129)
      Origin IGP, localpref 100, valid, external, best
  Refresh Epoch 1
  1
    172.16.11.1 from 172.16.11.1 (172.16.11.1)
      Origin IGP, localpref 50, valid, external
      Community: 6553617
```

You can see that ISP1 received two routes to 10.0.0.0/28 network:

- First route is received from ISP2, with default local preference of 100.
- Second route is received from R1, with the community attribute 6553617 and consequently it has changed local preference of 50.

Note that the community is displayed as a single 32-bit value (6553617). If you want it to be in a form of *as:nn*, you have to configure the support for new-format on the router, using the **ip bgp-community new-format** command.

Step 12 On the ISP1 router, enable new BGP community format.

```
ISP1(config)# ip bgp-community new-format
```

Step 13 On the ISP1 router, again verify the community attribute for 10.0.0.0/28 network.

```
ISP1# show ip bgp 10.0.0.0/28
BGP routing table entry for 10.0.0.0/28, version 2
Paths: (2 available, best #1, table default)
  Advertised to update-groups:
    3
  Refresh Epoch 2
200 1
  172.16.100.22 from 172.16.100.22 (10.0.2.129)
    Origin IGP, localpref 100, valid, external, best
  Refresh Epoch 1
1
  172.16.11.1 from 172.16.11.1 (172.16.11.1)
    Origin IGP, localpref 50, valid, external
    Community: 100:17
```

Now the community is displayed as two colon-separated 16-bit numbers, 100:17.

BGP Named Community Lists

The [BGP](#) Named Community Lists feature introduces a new type of community list that is called the named community list. A named community list can be configured with regular expressions or with numbered community lists.

BGP Named Community Lists

- Allows you 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.
- Increases the number of community lists that a network operator can configure—there is no limitation on the number of named community lists that can be configured.

Commands:

- Use the **ip community-list** command to create a numbered or named community list.
- Use the **match community** command in route-map configuration mode, to match a BGP community.
- Use the **set com-list delete** command in route-map configuration mode to remove communities from the community attribute of an update.

© 2015 Cisco and/or its affiliates. All rights reserved.67

The BGP Named Community Lists feature allows you to assign meaningful names to community lists. All rules of numbered communities apply to named community lists, with one exception. The exception is that there is no limitation on the number of community attributes that can be configured for a named community list. 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 Command

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

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

Syntax Description

Parameter	Description
<i>standard-list-number</i>	Specifies a standard community list number from 1 to 99 that identifies one or more permit or deny groups of communities.
<i>expanded-list-number</i>	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.
<i>community-list-name</i>	The community list name.
permit	Permits access for a matching condition.
deny	Denies access for a matching condition.
<i>community-number</i>	Community number that you configure with the set community command.

Valid value of community number is one of the following:

- A number from 1 to 4294967200. You can specify a single number or multiple numbers and you separate them 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 [EBGP](#) peer. External systems are the systems 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 Command

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

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.

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

Syntax Description

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

set comm-list delete Command

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.

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

```
set comm-list community-list-number / community-list-name delete
```

Syntax Description

Parameter	Description
<i>community-list-number</i>	A standard or expanded community list number.
<i>community-list-name</i>	A standard or expanded community list name.

BGP 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

The cost community is a nontransitive extended community attribute that is passed to [IBGP](#) and confederation peers but not to [EBGP](#) peers.

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.

© 2015 Cisco and/or its affiliates. All rights reserved.

68

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.

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

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

Syntax Description

Parameter	Description
igp	The IGP POI . The configuration of this keyword forces the cost community to be evaluated after the IGP distance to the next hop has been compared.
<i>community-id</i>	The ID for the configured extended community. The range is from 0 to 255.
<i>cost-value</i>	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 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. Then, the BGP best-path selection process will prefer the specific exit path.

BGP Cost Community 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 the routes that the route map permits. This configuration will cause the best-path selection process to prefer this route over other equal-cost paths that this route map sequence does not permit.

```
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 Feature

The [BGP](#) Link Bandwidth feature is used to enable multipath load balancing for external links with unequal bandwidth capacity. This feature supports [IBGP](#), [EBGP](#) multipath load balancing, and EIGRP multipath load balancing in [MPLS 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.

BGP Link Bandwidth Feature

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

© 2015 Cisco and/or its affiliates. All rights reserved.

69

You enable the BGP Link Bandwidth feature under an [IPv4](#) or [VPNv4](#) address-family session by entering the **bgp dmzlink-bw** command.

The BGP Link Bandwidth feature allows BGP to be configured to send traffic over multiple IBGP or EBGP learned paths. The traffic that is sent on those paths 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 Command

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

```
bgp dmzlink-bw
```

Note This command has no keywords or arguments.

neighbor dmzlink-bw Command

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

```
neighbor ip-address dmzlink-bw
```

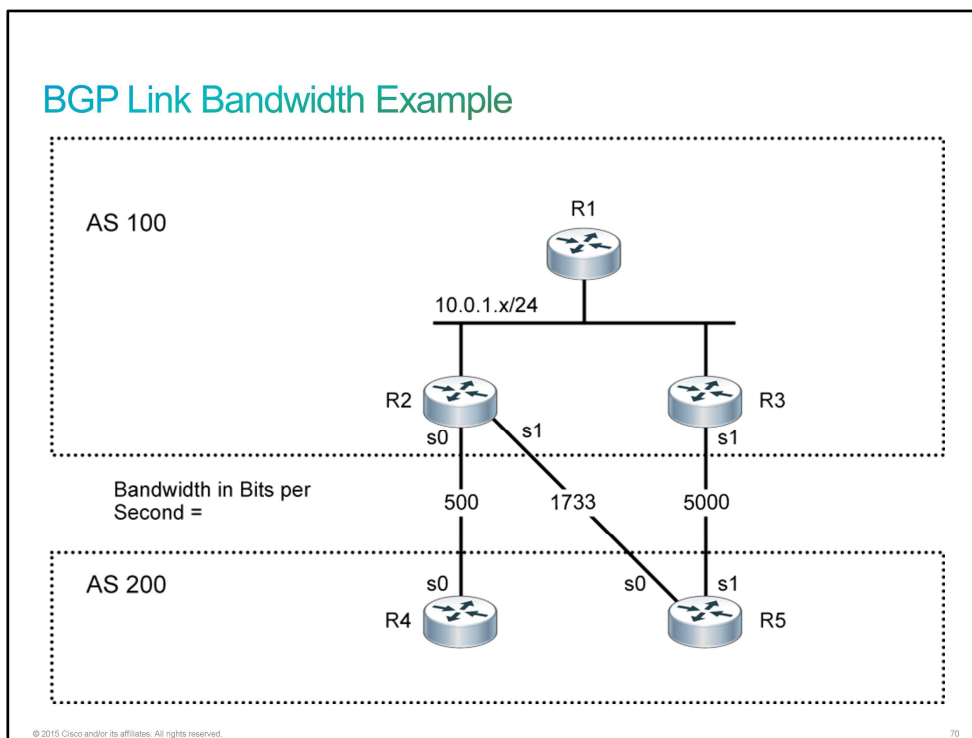
Syntax Description

Parameter	Description
<i>ip-address</i>	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 [DMZ](#) interface that connects two single-hop EBGPs 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 [IGP](#) costs are the same.

BGP Link Bandwidth Example

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



The figure shows two external autonomous systems. Three links connect the two ASs, and each AS carries a different amount of bandwidth (unequal-cost links). Multipath load balancing is enabled, and traffic is balanced proportionally.

First R1 is configured to support IBGP multipath load balancing and to exchange the BGP extended community attribute with IBGP neighbors, R2 on 10.0.1.2 IP address and R3 on 10.0.1.3 IP address.

```
R1(config)# router bgp 100
R1(config-router)# neighbor 10.0.1.2 remote-as 100
R1(config-router)# neighbor 10.0.1.2 update-source Loopback 0
R1(config-router)# neighbor 10.0.1.3 remote-as 100
R1(config-router)# neighbor 10.0.1.3 update-source Loopback 0
R1(config-router)# address-family ipv4
R1(config-router)# bgp dmzlink-bw
R1(config-router-af)# neighbor 10.0.1.2 activate
R1(config-router-af)# neighbor 10.0.1.2 send-community both
R1(config-router-af)# neighbor 10.0.1.3 activate
R1(config-router-af)# neighbor 10.0.1.3 send-community both
R1(config-router-af)# maximum-paths ibgp 6
```

R2 is configured to support multipath load balancing, to distribute R4 (172.16.100.1 IP address) and R5 (172.16.200.2 IP address) link traffic proportionally to the bandwidth of each link, and to advertise the bandwidth of these links to IBGP neighbors as an extended community:

```
R2(config)# router bgp 100
R2(config-router)# neighbor 10.0.1.1 remote-as 100
R2(config-router)# neighbor 10.0.1.1 update-source Loopback 0
R2(config-router)# neighbor 10.0.1.3 remote-as 100
R2(config-router)# neighbor 10.0.1.3 update-source Loopback 0
R2(config-router)# neighbor 172.16.100.1 remote-as 200
R2(config-router)# neighbor 172.16.100.1 ebgp-multihop 1
R2(config-router)# neighbor 172.16.200.2 remote-as 200
R2(config-router)# neighbor 172.16.200.2 ebgp-multihop 1
R2(config-router)# address-family ipv4
R2(config-router-af)# bgp dmzlink-bw
R2(config-router-af)# neighbor 10.0.1.1 activate
R2(config-router-af)# neighbor 10.0.1.1 next-hop-self
R2(config-router-af)# neighbor 10.0.1.1 send-community both
R2(config-router-af)# neighbor 10.0.1.3 activate
R2(config-router-af)# neighbor 10.0.1.3 next-hop-self
R2(config-router-af)# neighbor 10.0.1.3 send-community both
R2(config-router-af)# neighbor 172.16.100.1 activate
R2(config-router-af)# neighbor 172.16.100.1 dmzlink-bw
R2(config-router-af)# neighbor 172.16.200.2 activate
R2(config-router-af)# neighbor 172.16.200.2 dmzlink-bw
R2(config-router-af)# maximum-paths ibgp 6
R2(config-router-af)# maximum-paths 6
```

Also R3 is configured to support multipath load balancing and to advertise the bandwidth of the link with R5 (172.16.300.3 IP address) to IBGP neighbors as an extended community:

```
R3(config)# router bgp 100
R3(config-router)# neighbor 10.0.1.1 remote-as 100
R3(config-router)# neighbor 10.0.1.1 update-source Loopback 0
R3(config-router)# neighbor 10.0.1.2 remote-as 100
R3(config-router)# neighbor 10.0.1.2 update-source Loopback 0
R3(config-router)# neighbor 172.16.300.3 remote-as 200
R3(config-router)# neighbor 172.16.300.3 ebgp-multihop 1
R3(config-router)# address-family ipv4
R3(config-router-af)# bgp dmzlink-bw
R3(config-router-af)# neighbor 10.0.1.1 activate
R3(config-router-af)# neighbor 10.0.1.1 send-community both
R3(config-router-af)# neighbor 10.0.1.1 next-hop-self
R3(config-router-af)# neighbor 10.0.1.2 activate
R3(config-router-af)# neighbor 10.0.1.2 send-community both
R3(config-router-af)# neighbor 10.0.1.2 next-hop-self
R3(config-router-af)# neighbor 172.16.300.3 activate
R3(config-router-af)# neighbor 172.16.300.3 dmzlink-bw
R3(config-router-af)# maximum-paths ibgp 6
R3(config-router-af)# maximum-paths 6
```

BGP Support for Sequenced Entries in Extended Community Lists

The [BGP](#) Sequenced Entries in Extended Community Lists feature allows you automatic sequencing of individual entries in BGP extended community lists. It also provides you the ability to remove or resequence extended community list entries without deleting the entire existing extended community list.

BGP Support for Sequenced Entries in Extended Community Lists

- Allows you automatic sequencing of individual entries in BGP extended community lists.
- Provides you 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

71

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 Command

To create an extended community list and control access to it, use the **ip extcommunity-list** command in global configuration mode. To delete the entire community list, use the **no** form of this command.

```
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]
```

Syntax Description

Parameter	Description
<i>expanded-list-number</i>	An expanded list number from 100 to 500 that identifies one or more permit or deny groups of extended communities.
<i>standard-list-number</i>	A standard list number from 1 to 99 that identifies one or more permit or deny groups of extended communities.
expanded <i>list-name</i>	Creates an expanded named extended community list and enters IP extended community list configuration mode.
standard <i>list-name</i>	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.
<i>regular-expression</i>	(Optional) An input string pattern to match against.
rt	(Optional) Specifies the 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 SOO extended community attribute. The soo keyword can be configured only with standard extended community lists and not expanded community lists.
<i>extcom-value</i>	Specifies the RT or SOO extended community value. The value can be one of the following combinations: <i>autonomous-system-number</i> : <i>network-number</i> or <i>ip-address</i> : <i>network-number</i>
<i>sequence-number</i>	(Optional) The sequence number of a named or numbered extended community list. This value can be a number from 1 to 2147483647.
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.
<i>starting-sequence</i>	(Optional) Specifies the number for the first entry in an extended community list.
<i>sequence-increment</i>	(Optional) Specifies the increment range for each subsequent extended community list entry.

Sequenced and Resequenced Extended Community List Entry 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
```

Resequence the extended community list entries in the named community list that is configured above. 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
```

To display routes that the named extended community list permits, use the **show ip extcommunity-list EXEC** command. The output shows the configuration from the first example after it has been resequenced with user-defined values.

```
Router> show ip extcommunity-list
Standard extended community-list NAMED_LIST
    50 permit RT:64512:10
    150 permit RT:65000:20
    250 permit RT:64535:30
    350 permit SoO:65535:40
```

Summary

This topic summarizes the key points that were 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.
- 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 configure community attribute, you should:
 - Configure a route map with a **set community** command.
 - Configure propagation of BGP communities on the routers on a per-neighbor basis
 - Configure community lists to match against the community attribute as a method of route selection
 - Configure a route map to match networks that carry communities that the community list permits and apply policies to them.

© 2015 Cisco and/or its affiliates. All rights reserved.

72

Summary (Cont.)

- The **BGP Named Community Lists** feature allows you to assign meaningful names to community lists and increases the number of community lists that can be configured.
- The configuration of the **BGP Cost Community** feature allows you to assign cost values to specific routes. This way, you customize the BGP best path selection process for a local AS or confederation.
- The **BGP Link Bandwidth** feature allows you multipath load balancing for external links with unequal bandwidth capacity.
- BGP Support for **Sequenced Entries in Extended Community Lists** allows automatic sequencing of individual entries in BGP extended community lists. It also enables you to remove or resequence extended community list entries without deleting the entire existing extended community list.

© 2015 Cisco and/or its affiliates. All rights reserved.

73

Lesson 6: Module Summary

Overview

This topic summarizes the key points that were discussed in this module.

Module Summary

The following BGP attributes influence the route selection:

- **Weight** (The first criterion in BGP route selection.)
 - You can configure a default weight or use route-maps to configure it.
- **Local preference** (The second-strongest criterion in the route selection process.)
 - You can configure a default local preference or use route-maps to configure it.
- **AS-path**
 - You can perform AS-path prepending on outgoing EBGp updates over the undesired return path.
- **MED** (A "weak" parameter in the route selection process.)
- **Community**
 - Used to tag BGP routes that you can use to manipulate path selection and enforce administrative policies.

Lesson 7: 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.

1. Number the 10 BGP selection criteria in the order in which they are used, from the first to the last, when selecting the BGP path that is submitted to the IP routing table. (Source: "Influencing BGP Route Selection with Weights")

Prefer the oldest route for EBGp paths.	1
Prefer the path with the lowest neighbor BGP router ID.	4
Prefer the route that was originated by the local router.	7
Prefer the shortest AS path.	2
Prefer the highest local preference.	5
Prefer the lowest MED.	8
Prefer the path through the closest IGP neighbor.	3
Prefer the lowest origin code (IGP < EGP < incomplete).	10
Prefer an EBGp path over an IBGP path.	9
Prefer the highest weight.	6

2. 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. It depends on the Cisco IOS release.
3. 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.
4. If you want all routes received from the neighbor to have the new weight value applied, you have to reestablish BGP session. True or false? (Source: "Influencing BGP Route Selection with Weights")
 - A. true
 - B. false
5. 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.
6. 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.
7. 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

8. You can set the weight using a route map applied to a neighbor in the outgoing direction. True or false? (Source: "Influencing BGP Route Selection with Weights")
- A. true
 - B. false
9. 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.
10. 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.
11. What is the default local preference for routes that are received from a BGP neighbor? (Source: "Setting BGP Local Preference")
- A. 0
 - B. 100
 - C. There is no default local preference value.
 - D. It depends on the Cisco IOS release.
12. Which two statements regarding local preference are true? (Choose two.) (Source: "Setting BGP Local Preference")
- A. The higher value for local preference is preferred.
 - B. Local preference is used only between EBGp neighbors
 - C. The lower value for local preference is preferred.
 - D. Local preference is used only between IBGP neighbors
13. 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.

14. Let say we have two routes to the same destination, fist one has a weight of 50 and a local preference of 150, while the other one has a weight of 150 and a local preference of 50. The fist route will be selected as the best one, since it has higher local preference value. True or false? (Source: "Setting BGP Local Preference")
- A. true
 - B. false
15. 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**
16. 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**
17. Locally applied default value can be displayed with **show ip bgp** command. True or false? (Source: "Setting BGP Local Preference")
- A. true
 - B. false
18. 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**
19. Which two statements regarding the AS path are true? (Choose two.) (Source: "Using AS Path Prepending")
- A. The shorter AS path is preferred.
 - B. The longer AS path is preferred.
 - C. The AS path is prepended and exchanged between autonomous systems.
 - D. The AS path is local to an AS.

20. 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.
21. 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.
22. 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. The **set as-path prepend as-number** command in a route-map.
 - D. The interface global command **set as-path prepend as-number**.
23. The configuration below 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. 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")
- ```
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
```
- A. 347 347 347
  - B. 347 347 347 347
  - C. 529 347 347 347
  - D. 529 347 347 347 347
24. Why do you 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.

25. Which regular expression will allow any AS path number, but with a length of one? (Source: "Using AS Path Prepending")
- A. `^[0-9]+$`
  - B. `^[0-9](_[0-9])*$`
  - C. `^_[0-9]$`
  - D. `^[1-8]+$`
26. 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.
27. 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.
28. 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.
29. What is the default MED value? (Source: "Understanding BGP Multi-Exit Discriminators")
- A. 0
  - B. 100
  - C. There is no default MED value.
  - D. It depends on the Cisco IOS release.
30. 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

31. 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.
32. 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.
33. 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.
34. 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.
35. 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.

36. Which two statements regarding the MED are true? (Choose two.) (Source: "Understanding BGP Multi-Exit Discriminators")
- A. The higher value for the MED is preferred.
  - B. The lower value for the MED is preferred.
  - C. The MED is exchanged between autonomous systems.
  - D. The MED is local to an AS.
37. 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.
38. 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.
39. Match the steps with the actions describing how BGP communities can facilitate proper return path selection. (Source: "Addressing BGP Communities")

|                                                                                                                                                                                    |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Define administrative policy goals that you need to implement.                                                                                                                     | Step<br>1 |
| Define the filters and route selection policy that will achieve the required goals.                                                                                                | Step<br>2 |
| Apply communities on incoming updates from neighboring autonomous systems or tell the neighbors to set the communities themselves.                                                 | Step<br>3 |
| Enable community distribution throughout your AS to allow community propagation.                                                                                                   | Step<br>4 |
| 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. | Step<br>5 |
| Assign a community value to each goal.                                                                                                                                             | Step<br>6 |



40. How many community tags can be attached to a single BGP route? (Source: "Addressing BGP Communities")
- A. 1
  - B. 32
  - C. 256
  - D. It depends on the number that is configured with the **ip bgp community** command.
41. 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.
42. 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.
43. 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.
44. 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

45. 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.
46. Match the functions to the Cisco IOS commands that monitor BGP communities. (Source: "Addressing BGP Communities")

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

**show ip bgp community**

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

**show ip bgp community**  
**as:nn [as:nn ...]**

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

**show ip bgp community**  
**as:nn [as:nn ...] exact**

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

**show ip bgp community-list**  
*clist*

47. 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.
48. 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.

49. 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.
50. 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 Answers

---

## Answer Key

1. Prefer the highest weight. — 1  
Prefer the shortest AS path. — 4  
Prefer an EBGp path over an IBGP path. — 7  
Prefer the highest local preference. — 2  
Prefer the lowest origin code (IGP < EGP < incomplete). — 5  
Prefer the path through the closest IGP neighbor. — 8  
Prefer the route that was originated by the local router. — 3  
Prefer the path with the lowest neighbor BGP router ID. — 10  
Prefer the oldest route for EBGp paths. — 9  
Prefer the lowest MED. — 6
2. A
3. D
4. A
5. C
6. B
7. B
8. B
9. B, D
10. B
11. B
12. A, D
13. B, D
14. B

- 15. B
- 16. A
- 17. B
- 18. D
- 19. A, C
- 20. C
- 21. B
- 22. C
- 23. B
- 24. A
- 25. A
- 26. B
- 27. A, B
- 28. D
- 29. A
- 30. A, B, C
- 31. C, D
- 32. A, D
- 33. A, C
- 34. B
- 35. B
- 36. B, D
- 37. B, D
- 38. A
- 39. Define administrative policy goals that you need to implement. — Step 1
  - Define the filters and route selection policy that will achieve the required goals. — Step 2
  - Assign a community value to each goal. — Step 3
  - Apply communities on incoming updates from neighboring autonomous systems or tell the neighbors to set the communities themselves. — Step 4
  - 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. — Step 5
  - Enable community distribution throughout your AS to allow community propagation. — Step 6
- 40. B
- 41. A, B, D
- 42. B, C
- 43. D
- 44. C

45. D

46. Displays all routes in a BGP table that have at least one community attached. — **show ip bgp community**

Displays all routes in a BGP table that have all the specified communities attached. — **show ip bgp community as:nn [as:nn ...]**

Displays all routes in BGP table that have exactly the specified communities attached. — **show ip bgp community as:nn [as:nn ...] exact**

Displays all routes in BGP table that match community-list *clist*. — **show ip bgp community-list clist**

47. A, B

48. A, D, E

49. B, C

50. A, B, E





# Module 5: Customer-to-Provider Connectivity with BGP

---

## Introduction

Customers who are connecting to a service provider must consider various requirements and choose the proper connectivity type. In general, three types of connectivity are available: connecting to a single service provider, connecting to a single service provider using two links, and connecting to two independent service providers. Based on the connectivity type the proper routing methods, [IP addressing](#), and [AS](#) number have to be chosen. After all of these input parameters are known the customer and service provider can implement the desired connectivity type.

Upon completing this module, you will be able to:

- Describe the requirements to connect customer networks to the Internet in a service provider environment
- To implement customer connectivity by using static routing
- Defines how to implement Multihomed customer connectivity using BGP to a single ISP
- Define how to implement Multihomed customer connectivity using BGP 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 [VPNs](#), extranet connectivity with suppliers, and other Internet applications. When planning network connectivity to an [ISP](#), you as a network designer must give careful consideration to the various aspects of the connectivity. These aspects include physical connection types, the redundancy that the chosen connection method provides, [IP addressing](#) requirements, and [AS](#) numbering considerations, if the network design is going to meet both business and technical requirements of the applications that are planned for the network.

In this lesson, you will discuss solutions for connecting customer networks to service providers. You will also discuss customer network redundancy requirements, routing requirements, IP addressing requirements, and AS numbering requirements.

Upon completing this lesson, you will be able to:

- Identify various physical connections that the customers use to connect to a service provider
- Describe the levels of redundancy that is provided by each physical connection type that the customers use to connect to a service provider
- Identify various routing schemes that the customers use to connect to a service provider
- Describe routing schemes that are appropriate for each physical connection type that the customers use to connect to a service provider
- Describe the addressing schemes that the customers use to connect to a service provider
- Describe the AS numbering schemes that the customers use to connect to a service provider

# Customer-to-Provider Connectivity Types

## Customer-to-Provider Connectivity Types

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

- **Single-homed customers:**
  - Single permanent connection to a single service provider.
  - 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.
- **Multihomed customers:**
  - Multiple permanent connections to multiple service providers for maximum redundancy.

© 2015 Cisco and/or its affiliates. All rights reserved.

75

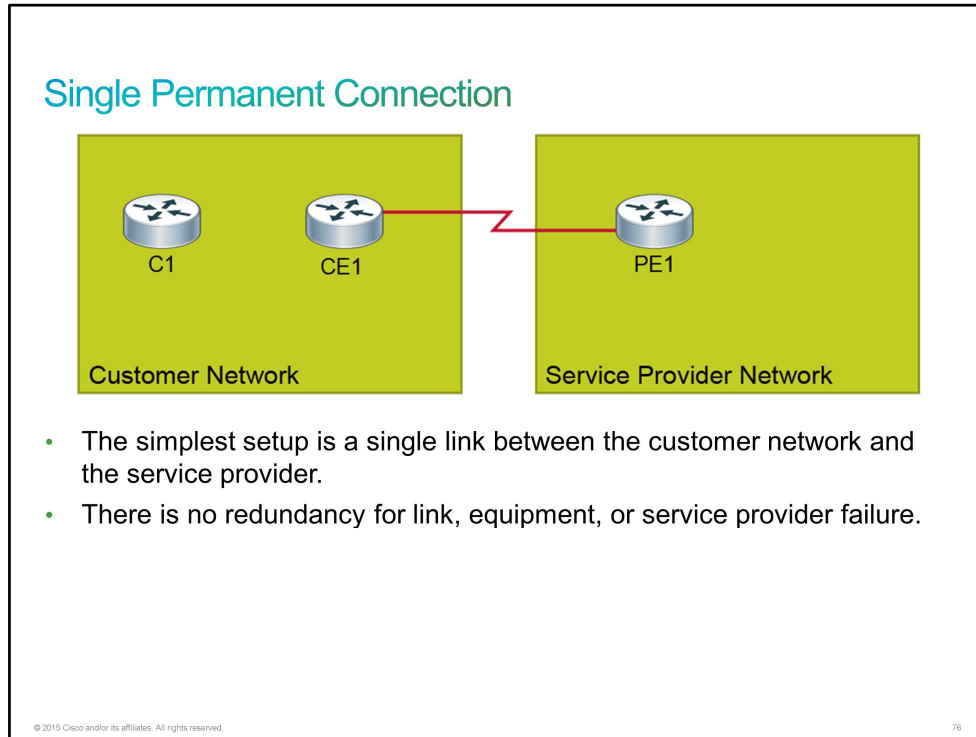
Service provider customers have different requirements, such as redundancy, stability, security, flexibility, for their Internet connectivity. These different requirements result in different solutions:

- A single permanent connection to one [ISP](#). This solution meets the requirements for most of the customers.
- Multiple permanent connections to one ISP, 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. For example, 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 extra bandwidth.
- Multiple permanent connections to more than one ISP. This solution provides the highest level of redundancy, because not only does it cope with link-level failures but also with failures within the network of a service provider.

# Customer Redundant Connectivity

The level of redundancy depends on a physical connection type that the customers use to connect to a service provider.

## Single Permanent Connection

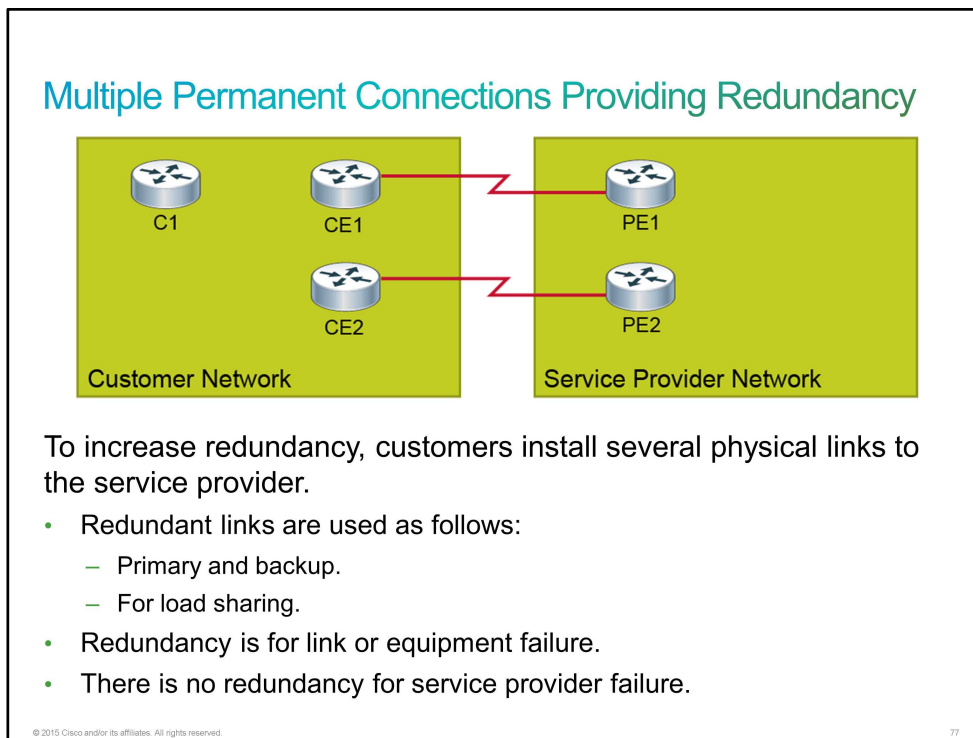


A single permanent connection to one [ISP](#) is the most common setup. This setup is also the simplest to implement.

The customer connects to a single service provider using a single link. The customer has an edge router that is connected to the edge routers of the ISP. This connection is permanent and could be a leased line, a [Frame Relay](#), or [ATM PVC](#), [DSL](#), [FTTx](#), cable, or something equivalent.

You do not have 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



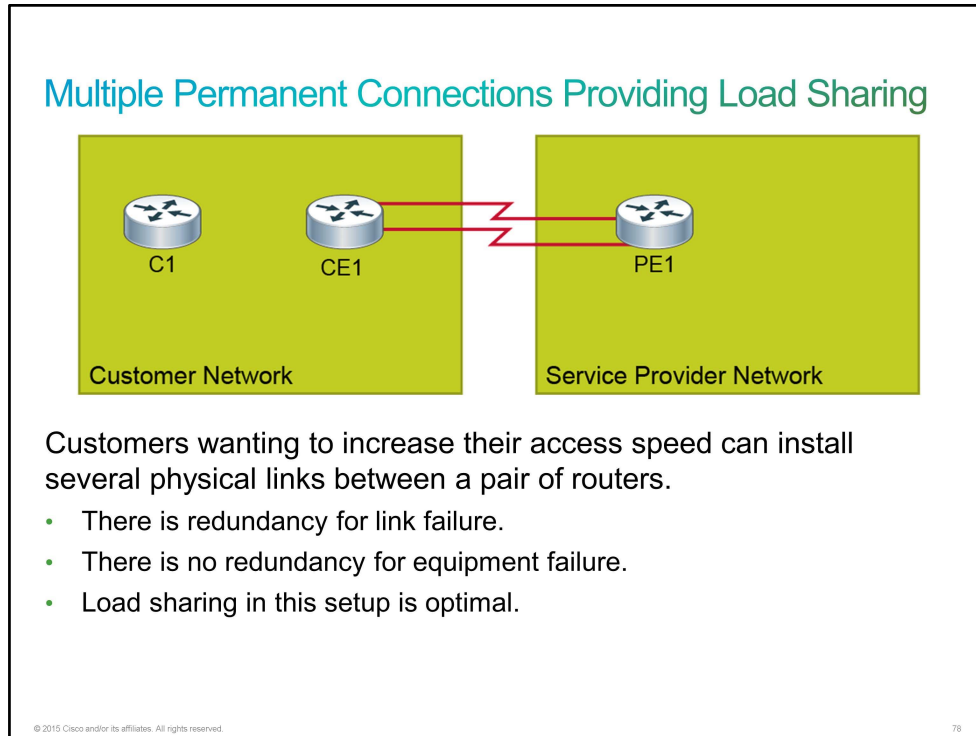
In this setup, you have one CE router that is connected to one PE router, and another CE that is connected to another PE router. If one of these routers fails, only one of the connections breaks down; the other connection is still available.

These two links may be implemented in two scenarios:

- **Primary and backup:** One link is always active, while another one is the backup and is used only when the primary link fails. Backup PVCs in Frame Relay or ATM networks can be very cost-efficient in case that these PVCs carry only a very small volume of traffic and that the primary path is available.
- **Load sharing:** Both links are used at the same time to provide load balancing. 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, redundancy is provided for link, and customer, or ISP device failure. The failures in the ISP network will still affect the customer. This solution can mitigate router and link failures, but it cannot mitigate ISP failures.

## Multiple Permanent Connections Providing Load Sharing

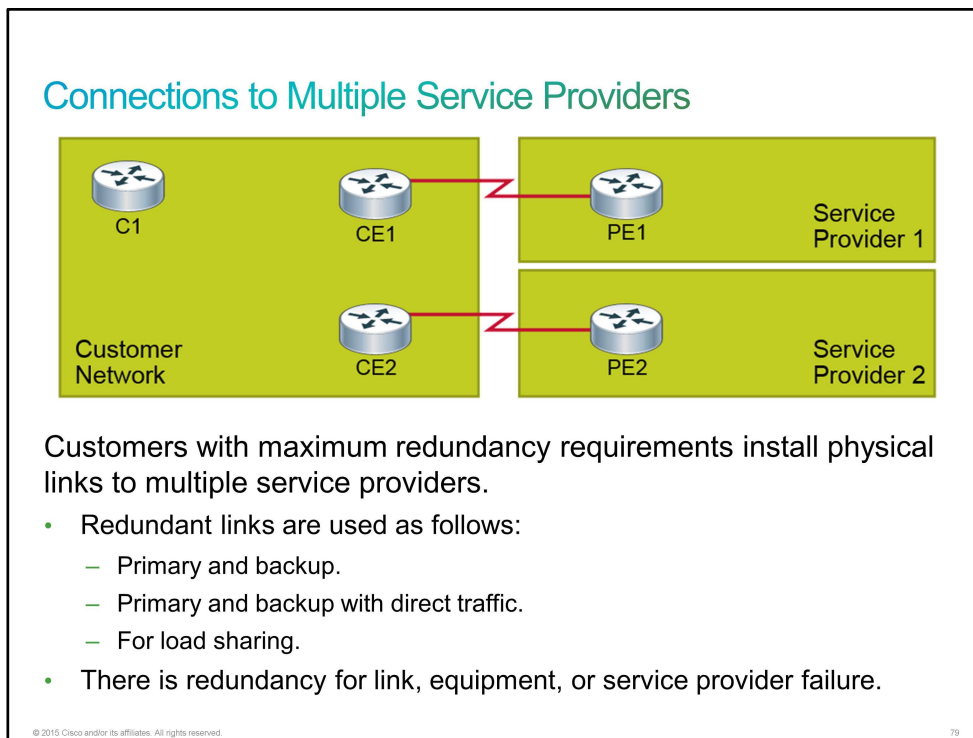


In this setup, you have a single router in the customer network that is connected to a single router in the ISP network. The redundancy is limited to the link level because router failures are not covered. Using two parallel links between two routers allows for an optimal distribution of load over the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

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

As in the previous examples, because the customer is connected to a single ISP, serious ISP network failure will affect the customer regardless of the link backup.

## Connections to Multiple Service Providers



In the example, two **CE** routers have one permanent connection each to a different ISP. Link and device failures are covered in the same way as in the previous examples. However, because the two connections in this example go to two different ISPs, the redundancy also covers problems within one ISP network. If one service provider fails, traffic will still flow through another service provider, and the highest level of redundancy is achieved.

In this setup, three scenarios are possible:

- **Primary and backup:** One link is always active, while another one is the backup and is used only when the primary link fails.
- **Primary and backup with direct traffic to the backup service provider:** One link is active, and another one is the backup. Except for the traffic that should go directly to service provider 2.
- **Load sharing:** Both links are used at the same time to provide load balancing.

In all scenarios, the customer has to cooperate with a service provider to achieve the desired setup. Load sharing between the links 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 router configuration parameters and observing the link traffic load changes can help you achieve an acceptable distribution of traffic between two links.

Avoiding any load on the backup link may require assistance from the ISP to which the backup link is connected.



# Customer-to-Provider Routing Schemes

Different solutions for connecting a customer network to the network of an [ISP](#) require different methods of routing information exchange.

## Customer-to-Provider Routing Schemes

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

© 2015 Cisco and/or its affiliates. All rights reserved.

80

Routing options:

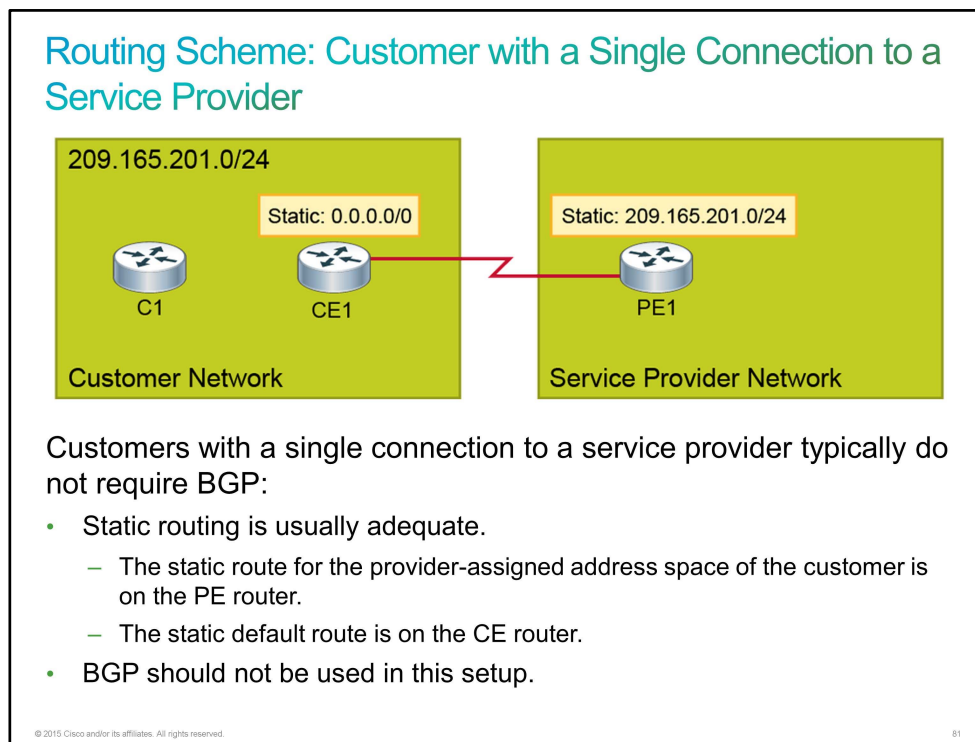
- **Static routing:** Static routing is preferred because of its lower complexity. In a normal case, the customer network must have a default route to the ISP network and the ISP network must have a route to the IP prefixes that the customer has in its network. As always, static routing provides very low redundancy, if any.
- **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 Schemes

Different routing schemes are appropriate for each physical connection type that the customers use to connect to a service provider.

## Routing Scheme: Customer with a Single Connection to a Service Provider

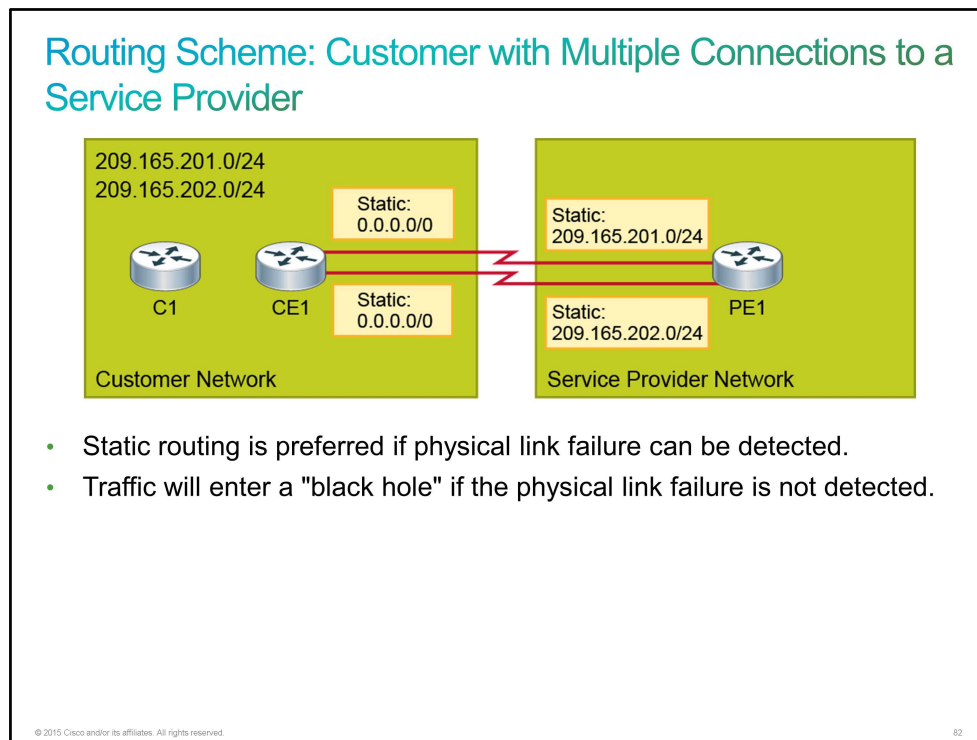
When the customer has a single connection to the service provider, static routing is usually adequate. The physical topology does not provide any redundancy, and it is therefore unnecessary to add the complexity of dynamic routing.



In this case, keep the network simple by avoiding the use of [BGP](#). If there is static routing, a default route is needed on the [CE](#) router that points to the [PE](#) router. A static route for the customer network is needed on the PE router, which points to the CE router.

If you use BGP, the service provider sends a default route to the customer, and the customer advertises its own address space to the service provider.

## Routing Scheme: Customer with Multiple Connections to a Service Provider



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. This configuration enables the link-level procedures to detect a link failure.

With this type of connection, two default routes are configured on the CE router that point to the PE router. Two static routes are configured on PE router, 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 then 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.

## Routing Scheme: Customer with Multiple Connections to a Service Provider (Cont.)



- You can still use static routing if link and service provider equipment failure can be detected.
- BGP between the customer and the service provider is usually used in this setup.
  - The service provider originates the default route (in the primary or secondary scenario) or specific routes (in the load-balancing scenario).
  - The customer originates its address space.

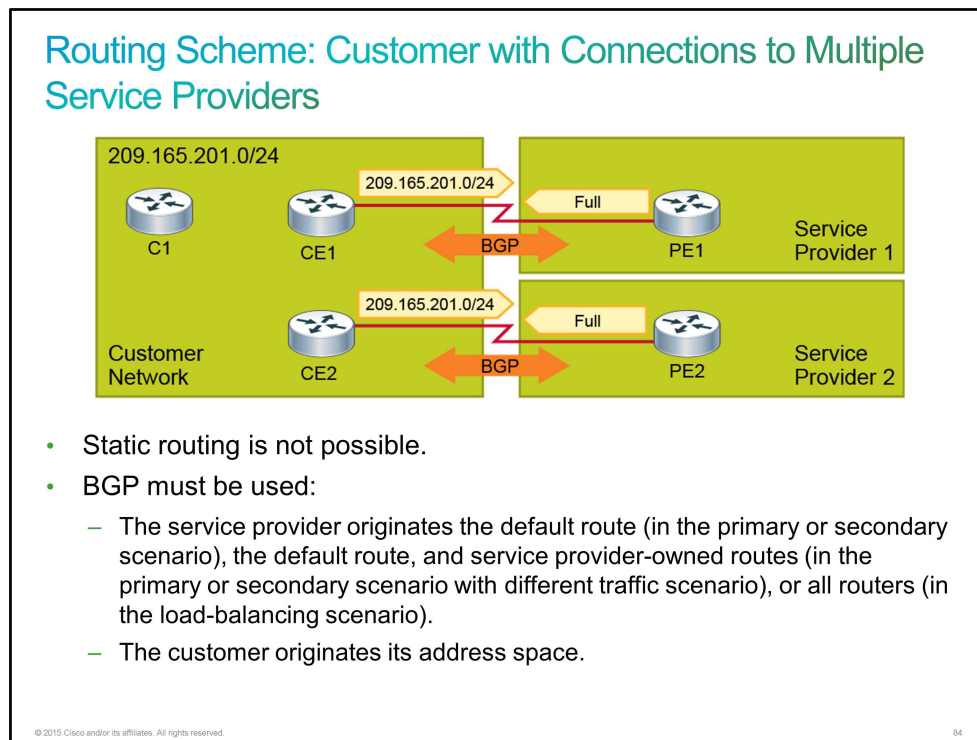
© 2015 Cisco and/or its affiliates. All rights reserved.

83

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. However, you can use static routing only if the link-level procedures can detect link failures. 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 customer's router stops the redistribution of the default route into the customer [IGP](#). The provider's router stops the redistribution of the static route, that points the customer network, into the service provider BGP.

However, customers that require the use of multiple connections and multiple routers very often do not rely on the link-level procedures. A better option to detect link failures is using BGP between CE and PE routers. Because BGP uses handshaking and reliable transfer, it always detects a failed link or failed remote router. BGP can also be used to achieve a primary or secondary scenario or load balancing of traffic over both links.

## Routing Scheme: Customer with Connections to Multiple Service Providers



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 want only 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 ISP network.

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 each ISP can reach at the moment.

CE routers advertise the customer network to the service provider. PE routers advertise only the default route or the default route plus the service provider-owned networks or all routes to the CE routers. The level of routes that the service provider advertises is determined based on the required customer scenario.

# Customer Addressing Requirements

Customers who are connected to a single service provider usually have their address space assigned by the service provider, which is known as [PA addresses](#). 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 [PI addresses](#).

## Customer Addressing Requirements

### Single-homed customers:

- Customer gets the address space from the provider—PA address space.
- Single (link) IP address or multiple (subnet) IP address.
- If service provider changes, readdressing is needed.
- Inside the customer network, private addresses are used, with a combination of NAT.

### Multihomed customers:

- Customer assigns its own address space—PI address space
- Customer has to advertise address space to all service providers.
- If service provider changes, readdressing is not needed.

© 2015 Cisco and/or its affiliates. All rights reserved.

85

**Single-Homed Customers:** Customers that are connected to a single ISP usually get their address space assigned by the ISP. Customers can either request a single [IP address](#), which is used on the access link as well, or they can request an entire IP subnet. A service provider is usually assigned a large address space to delegate to its customers. 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.

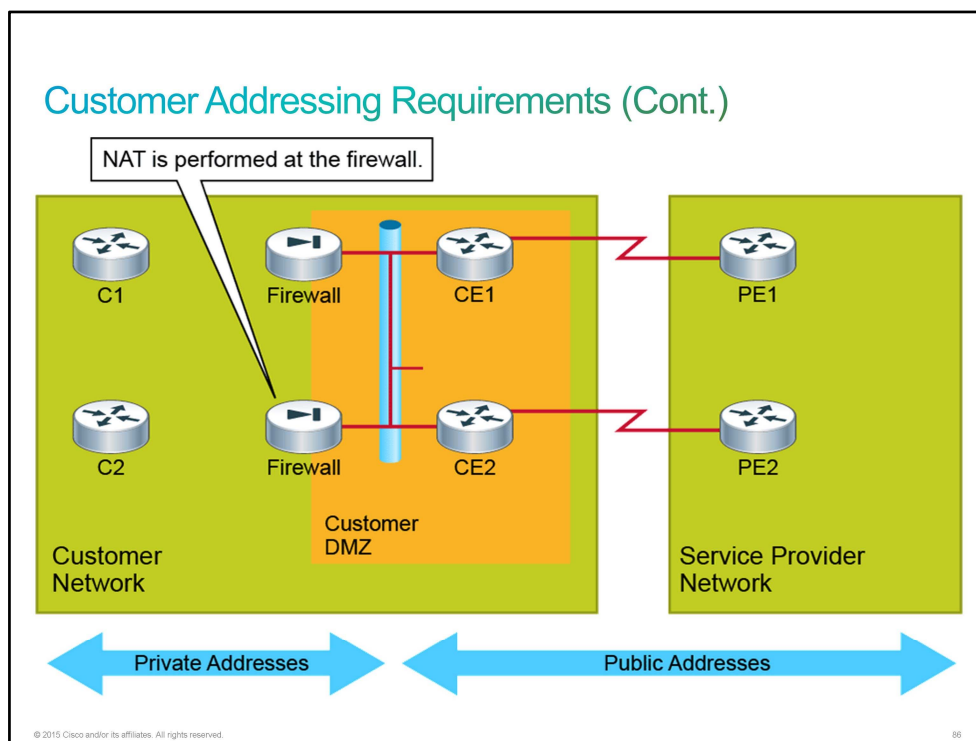
If the customer should decide to change its service provider, the customer must return its PA addresses to the old ISP. The customer then receives a new assignment of PA addresses from the new ISP. Otherwise, the ISPs are no longer able to perform efficient address aggregation. The consequence for the customer is that the customer has to renumber its network when it changes its service provider.

**Multihomed Customers:** Customers that are connected to multiple ISPs usually assign their own address space. None of the service providers advertises the address, and the customer has to advertise it to both service providers. 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 many 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.

If the customer should decide to change its service provider, renumbering of the network is not needed.

**Note** 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.



Some customers decide to use private addresses within their network and do 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 the figure, only customer DMZ has been assigned public addresses. The customer network is connected to the customer DMZ using two alternate firewalls with both firewalls doing NAT.

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, addresses are renumbered only at the NAT point. The rest of the customer network does not need to be renumbered.

# Customer AS Number Allocation

Depending on a customer connectivity to single service provider or to multiple service providers, the assigned [AS](#) number should be private or public.

## Customer AS Number Allocation

**Single-homed customers:**

- If static routing is used, the AS number is not needed.
- If BGP is desired, private AS number (64512 - 65535) can be used.

**Multihomed customers:**

- Public AS number has to be used.

© 2015 Cisco and/or its affiliates. All rights reserved.

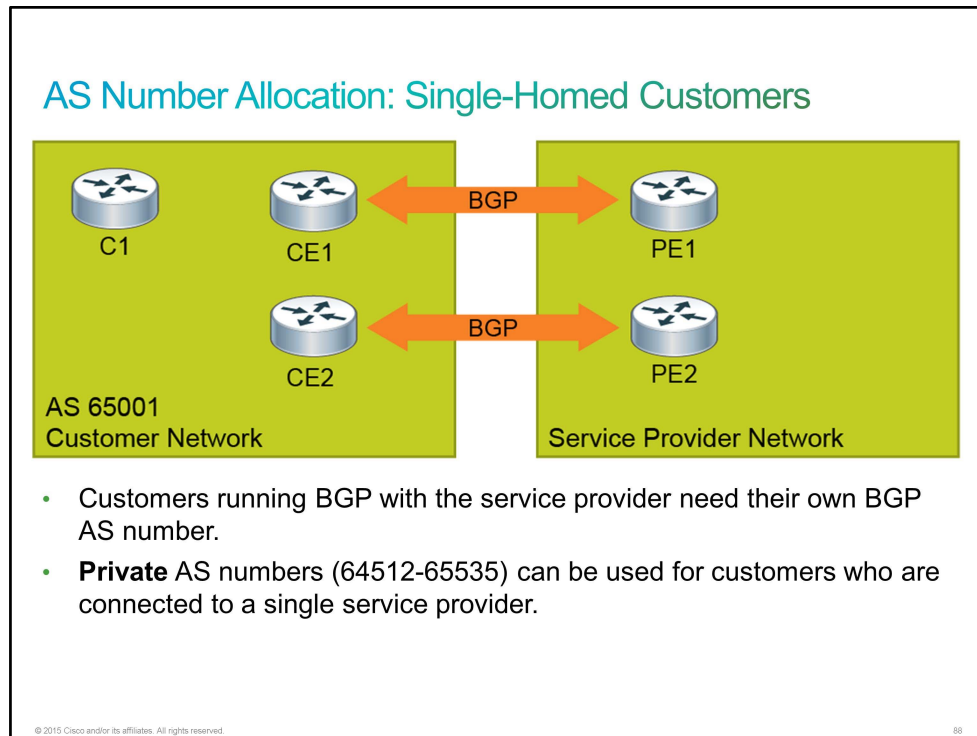
87

Customers who are connected to a single service provider usually do not use [BGP](#), so an AS number is not needed. If a customer needs a BGP to detect link failures, the service provider can assign a private AS number to the customer.

Customers who are connected to multiple service providers use BGP and they should be assigned a public AS number.



## AS Number Allocation: Single-Homed Customers

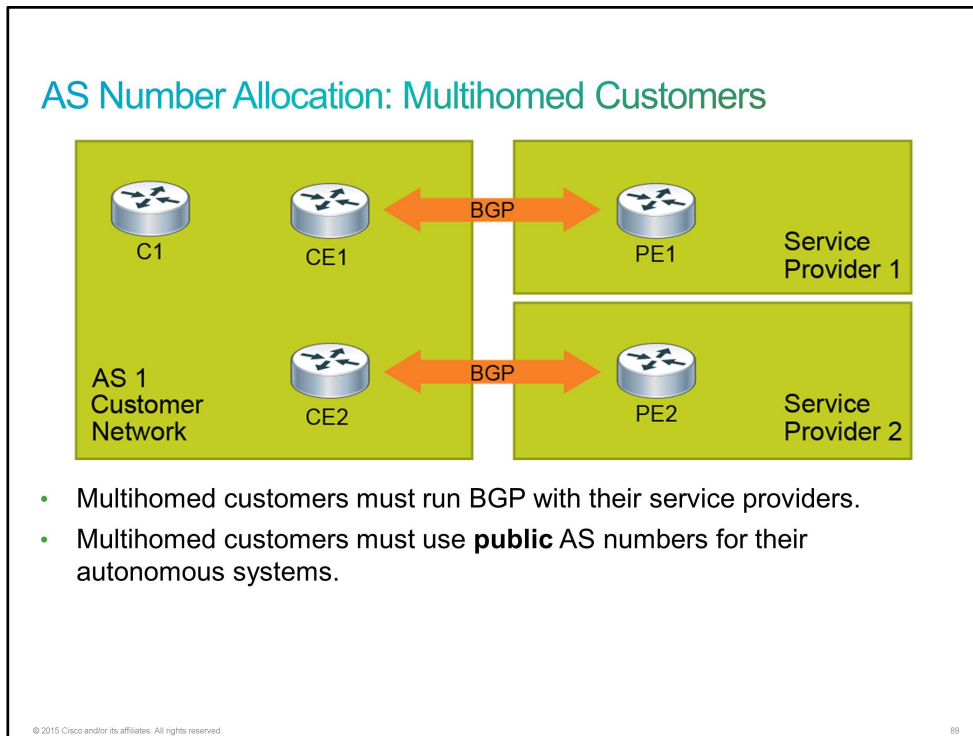


BGP requires the use of AS numbers. When BGP is configured, the AS number is mandatory information. Public AS numbers are a scarce resource, however. Customers should use public AS numbers only when they are required. A customer that uses BGP to exchange routing information with only one [ISP](#) does 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. The ISP treats this 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. 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



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 service provider should propagate it to the rest of the Internet. The customer network is now reachable by the rest of the Internet through both providers. Internet endpoints use the route with the shortest AS path as the best route to the customer network.

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Different customers have different requirements, especially for redundancy for their Internet connections. These connectivity options include:
  - **Single-homed:** One or multiple connections to a single ISP.
  - **Multihomed:** Multiple connections to different ISPs.
- Routing, static or dynamic, that is used between a customer and service provider depends on the selected connectivity type.
  - If dynamic routing is required, BGP is the only option.
- IP addressing assignment depends on the selected connectivity type:
  - **Single-homed:** PA address space.
  - **Multihomed:** PI address space.
- AS numbering depends on the selected connectivity type:
  - **Single-homed:** No AS number, or private AS number
  - **Multihomed:** Public AS number.

© 2015 Cisco and/or its affiliates. All rights reserved.

90



# 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 [ISP](#), static routing is the simplest routing approach to implement between the 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.

Upon completing this lesson, you will be able to:

- 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

# When to Use Static Routing?

Static routing is the best solution to implement when you do not have redundancy in the network topology.

## When to Use Static Routing?

- Static routing should be used only with single-homed customers in the following examples:
  - 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 with multihomed customers.

© 2015 Cisco and/or its affiliates. All rights reserved.

91

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.

The [CE](#) router that is using an [IGP](#) must conditionally announce a static default route. If the link to one of the CE 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 [PE](#) 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." Since the router cannot detect a failure at the link level in these cases, 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 true regardless of which link failure detection mechanisms are in use.

# Characteristics of Static Routing

When static routing is implemented between the customer network and the [ISP](#) network, the [CE](#) 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. Different routers within the customer network must be able to select the best route to the exit point of the network.

## 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-connected routes should be redistributed into service provider BGP.
  - Redistribute static routes into BGP, not core IGP.
- Specific customer routes should not be advertised to other autonomous systems:
  - Mark customer routes with the **no-export** community when redistributed into BGP.
  - Use static route tags for consistent tagging.

© 2015 Cisco and/or its affiliates. All rights reserved.

92

Different IGPs use different methods of announcing a router as a gateway of last resort. [EIGRP](#) uses the concept of a default network, while [OSPF](#) and [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 CE router as a gateway of last resort.

When static routing is used between the customer and the provider, the [PE](#) must propagate a static route that points to the customer network. It must propagate it to all other routers within the ISP network, and also to the rest of the Internet. The route is redistributed 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.

Imagine that a customer uses [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. The service provider should not propagate the routes of this customer to the rest of the Internet. The ISP should carefully redistribute only those connected routes that a single [IP address](#) customer uses. The rest of the Internet should receive only an announcement containing the larger block of addresses, a summary route.

An easy way of achieving this setup is to use communities within the ISP network. You can make individual routes ineligible for advertising outside of the local [AS](#) by setting the predefined BGP **no-export** community. To ensure that the BGP communities are propagated, at least over all [IBGP](#) sessions, the network operators of the ISP network must configure a **send-community** option for all IBGP neighbors. If a router in a provider network receives an update carrying the **no-export** community, the router will not propagate the update to any external BGP neighbors. Meaning the update is not propagated to other service providers.

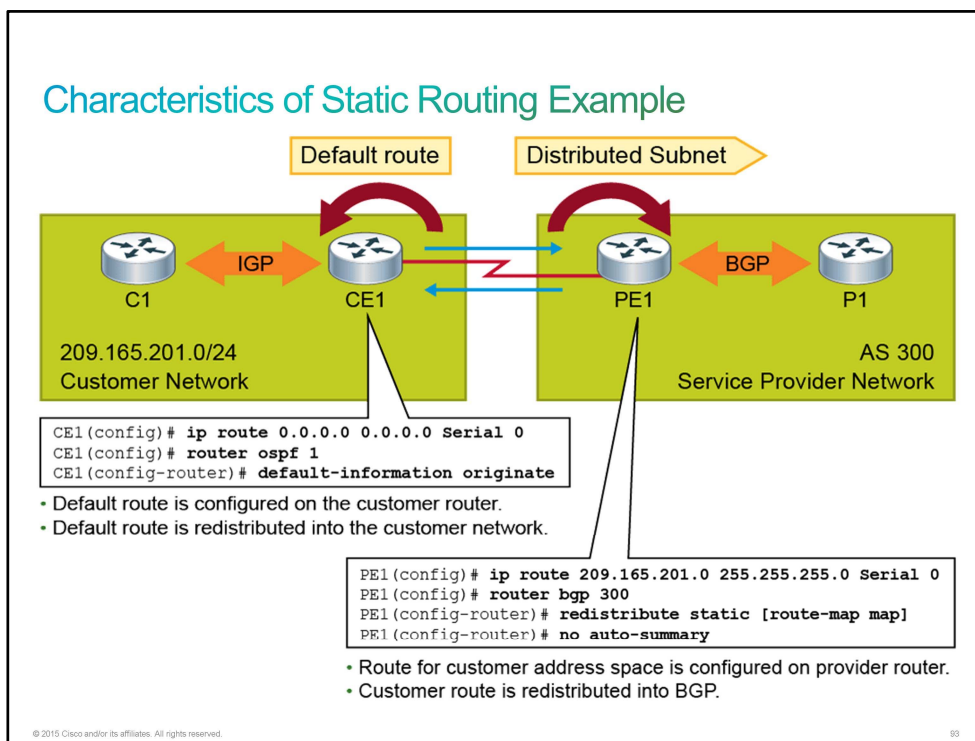
---

**Note** Communities are set by using route maps. A route map can select routes based on various attributes. One of these attributes is the route tag. Through configuration, the router can assign a route tag to specific static routes. This option means that the network operators of the ISP network can invent a scheme of tagging. For example 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.

---

## Characteristics of Static Routing Example

In the figure, the customer network is connected to the Internet by using a single permanent connection to a single service provider.



In this case, a routing protocol does not add any redundancy and would only add complexity.

The configuration of CE and PE router is shown in the example. Access link is used, and directly connected routes are redistributed into BGP on the PE router. The CE router is configured with static default route, which points to the serial 0 interface. If the serial interface goes down, the route becomes invalid. The default route is also redistributed into OSPF, which is used in the customer network. Therefore the router announces a default route into OSPF as long as it has a valid default route itself.



The PE router also has a static route, declaring the customer IP network number 209.165.201.0/24 as reachable over the serial 0 interface. It also becomes invalid if the interface goes into the down state. The PE 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, the BGP announces it. To the rest of the Internet, the customer network appears as reachable within the 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

You can easily extend the principle of using tags when you are configuring static routes. Based on those tags, you can assign different communities to implement a more complex routing policy.

## Designing Static Route Propagation in a Service Provider Network

- Identify all possible combination of services that are offered to a customer, including QoS services.
- Assign a tag to each combination of services.
- Configure a route map that matches the 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

84

To propagate static routes in a service provider network, complete these steps:

1. Identify all different service levels that are offered to customers and then all the different combinations of these service levels.
2. Assign each combination its own tag value and its own community.
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 [PE](#) 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.

---

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.

## Static Route Propagation Example

A scenario with varied service levels is shown. Static route propagation is configured in a BGP environment.

| Static Route Propagation Example                                                                                                                                                                                                                                                                                                                                                                 |     |                          |          |                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|--------------------------|----------|---------------------|
| Sample service offering:                                                                                                                                                                                                                                                                                                                                                                         |     | Advertise Customer Route | QoS Type | Route Tag           |
| <ul style="list-style-type: none"> <li>Addressing: <ul style="list-style-type: none"> <li>PA address blocks are not propagated to upstream ISPs.</li> <li>PI address blocks are propagated to upstream ISP.</li> </ul> </li> <li>Quality of service: <ul style="list-style-type: none"> <li>Normal customers.</li> <li>Gold customers.</li> </ul> </li> <li>Define static route tags.</li> </ul> |     |                          |          |                     |
|                                                                                                                                                                                                                                                                                                                                                                                                  | No  | Normal                   | 1000     | no-export 300:31000 |
|                                                                                                                                                                                                                                                                                                                                                                                                  | Yes | Normal                   | 1001     | 300:31000           |
|                                                                                                                                                                                                                                                                                                                                                                                                  | No  | Gold                     | 2000     | no-export 300:32000 |
|                                                                                                                                                                                                                                                                                                                                                                                                  | Yes | Gold                     | 2001     | 300:32000           |

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 that the ISP announces covers these customers. Other customers use [PI addresses](#) that must be explicitly announced to the Internet by the service provider.

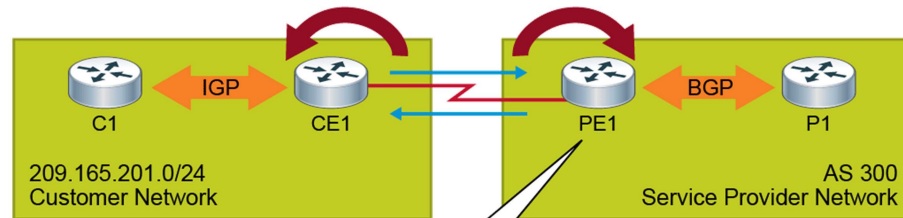
There are two different [QoS](#) services, Normal and Gold, and there are both PA and PI addresses. So the total number of combinations to cover the network policy is four:

- Normal QoS routes that the ISP assigns and should not be explicitly announced.
- Normal QoS routes that are [PI](#) routes and should be explicitly announced.
- Gold QoS routes that the ISP assigns 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 Example (Cont.)

### Configure Route Maps



```
PE1(config)# route-map IntoBGP permit 10
PE1(config-route-map)# match tag 1000
PE1(config-route-map)# set community no-export 300:31000
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 20
PE1(config-route-map)# match tag 1001
PE1(config-route-map)# set community 300:31000
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 30
PE1(config-route-map)# match tag 2000
PE1(config-route-map)# set community no-export 300:32000
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 40
PE1(config-route-map)# match tag 2001
PE1(config-route-map)# set community 300:32000
```

- Every combination of service that are offered to the customer has to be matched individually.
- Do not insert `permit all` at the end. Only routes with proper tags are redistributed into BGP.

© 2015 Cisco and/or its affiliates. All rights reserved.

96

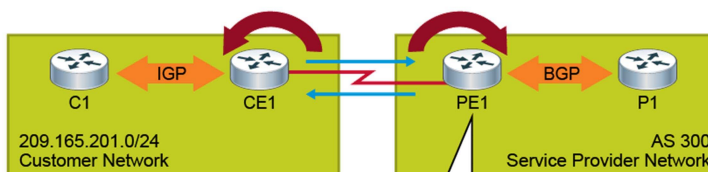
Network operators configure a route map in the PE1 router that has the static routes to the customer network. Redistribution of the configured customer static routes into BGP is also performed at the PE1 router.

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 PE 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 the administrator catch configuration entry errors, thus giving all customers the service combination that they are entitled to.

## Static Route Propagation Example (Cont.)

### Redistribution and Customer Routes



```

PE1(config)# route-map IntoBGP permit 10
PE1(config-route-map)#match tag 1000
PE1(config-route-map)#set community no-export 300:31000
PE1(config-route-map)#exit
PE1(config)# route-map IntoBGP permit 20
PE1(config-route-map)#match tag 1001
PE1(config-route-map)#set community 300:31000
PE1(config-route-map)#exit
PE1(config)# route-map IntoBGP permit 30
PE1(config-route-map)#match tag 2000
PE1(config-route-map)#set community no-export 300:32000
PE1(config-route-map)#exit
PE1(config)# route-map IntoBGP permit 40
PE1(config-route-map)#match tag 2001
PE1(config-route-map)#set community 300:32000
PE1(config-route-map)#exit
PE1(config)# router bgp 300
PE1(config-router)# redistribute static route-map IntoBGP
PE1(config-router)# neighbor 1BGP-neighbor send-community
PE1(config-router)# no synchronization
PE1(config-router)# exit
PE1(config)# ip route 209.165.201.0 255.255.255.0 serial 0 tag 1000

```

© 2015 Cisco and/or its affiliates. All rights reserved.

97

The figure shows how the PE1 router uses the route map named "IntoBGP" when redistributing static routes into BGP. Because the route map assigns community values that other routers within the ISP network will use, network operators must configure all **IBGP** neighbors with the **send-community** qualifier.

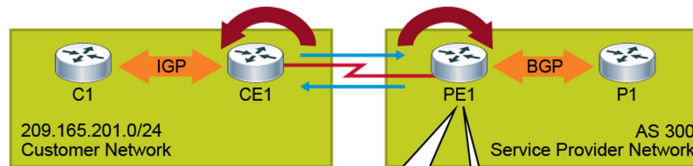
When connecting customers, the network operators identify which service combination to use for this particular customer. The three services that are associated with this particular customer are as follows:

- 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 Example (Cont.)

Static Routes on the PE Router



```
PE1# show ip bgp 209.165.201.00
BGP routing table entry for 209.165.201.0/24, version 2
Paths: (1 available, best #1, table default, not advertised to EBGP peer)
Not advertised to any peer
Refresh Epoch 1
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: 300:31000 no-export
```

© 2015 Cisco and/or its affiliates. All rights reserved.

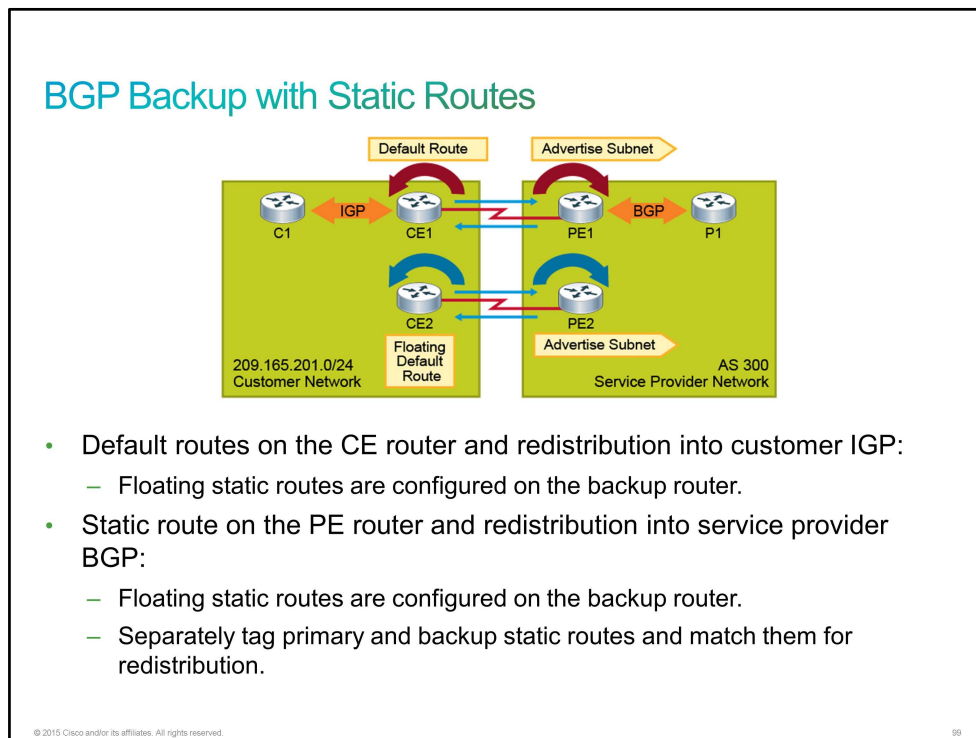
98

The **show ip route** command displays information from the routing table about subnet 209.165.201.0/24. The route is learned by static configuration and is redistributed via BGP. The router has assigned a tag value of 1000 to the customer route by using a statically assigned tag. 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 209.165.201.0/24. The route is local within this AS and this router sourced it. The BGP communities 300:31000 and **no-export** have been assigned by the router to the redistributed customer route. They were assigned by using the provider-defined route map before inserting the customer route into the BGP table.

# BGP Backup with Static Routes

Imagine a case where the customer network has two connections to a single service provider, using two routers on each side. 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 between the customer and provider networks.



As in the previous example, where no backup link is available, the CE1 router has a static default route toward the ISP. The PE1 router 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. This causes 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.

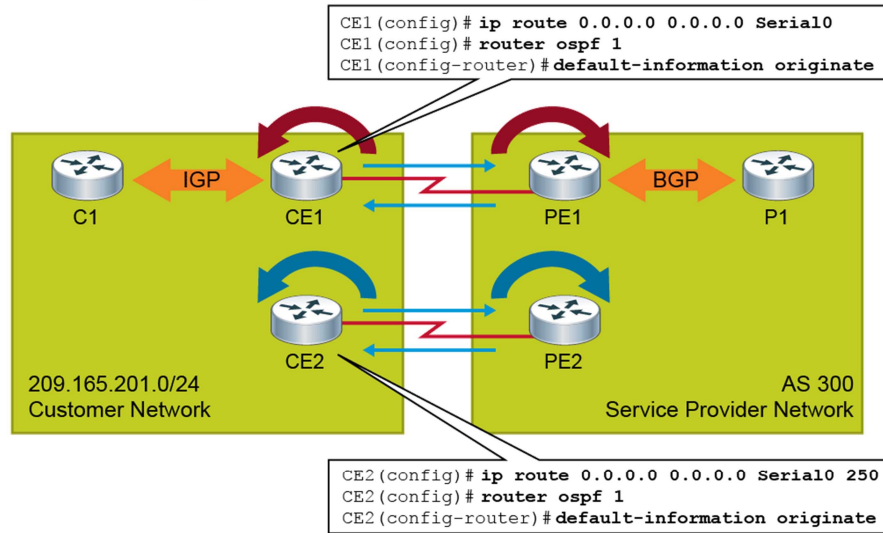
The appearance of the route in the routing table conditions the redistribution of routes into any routing protocol. Thus, if the interface goes down, the router removes the static route from its routing table, and the route is withdrawn from the routing protocol. When the static route reappears, the redistribution process inserts it into the routing protocol again.

The CE2 router also uses a default static route toward the ISP via the backup link. The CE2 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 [AD](#)—higher than the administrative distance of the customer IGP. As long as the primary link works, the IGP provides the CE2 router with the primary default route. Because of the higher AD, the backup static default route is not installed into the CE2 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 CE2 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 then redistribute it into the IGP.

The PE2 router can also use floating static routes, which are redistributed into the ISP [BGP](#) process.

## BGP Backup with Static Routes (Cont.)

### Customer Configuration



In the figure, the customer network and the ISP network are connected using leased lines. Both, CE1 and CE2, 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 CE2 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 CE2 receives the default route by OSPF, the static default route is not used.

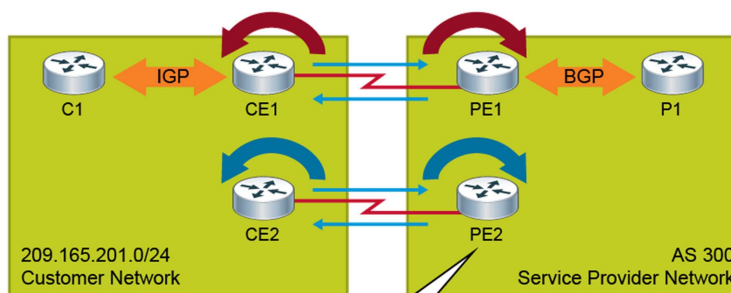
When the primary link goes down, the static default route in the primary router is not valid. The OSPF protocol stops announcing the default route, because the **default-information originate** command makes OSPF contingent on the availability of that static default route in the routing table before announcement.

The CE2 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 CE2 as the gateway of last resort.



## BGP Backup with Static Routes (Cont.)

### Provider Configuration



```
PE2(config)#ip route 209.165.201.0 255.255.255.0 serial 0 tag 1000 250
PE2(config)# route-map IntoBGP permit 10
PE2(config-route-map)#match tag 1000
PE2(config-route-map)#set community no-export 300:31000
PE2(config-route-map)#exit
PE2(config)#router bgp 300
PE2(config-router)# 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 there.

© 2015 Cisco and/or its affiliates. All rights reserved.

101

When floating static routes are configured on the PE 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 209.165.201.0/24. In the PE2 router, the floating static route is assigned the same tag value as the tag value being used in the PE1 router. The route map "IntoBGP" is the same as in the PE1 router and provides the routes to the customer network with the same communities. These communities are based on the same QoS level and the indication whether to explicitly announce route prefix to the rest of the Internet.

The floating static route has the administrative distance set to 250. This value is higher than any routing protocol. When the PE2 router no longer receives any routing protocol information about the customer networks, the router will automatically install the floating static route and then redistribute it into BGP.

Based on BGP route selection rules, the redistributed floating static route will always remain the preferred path if an extra BGP configuration is not performed on the PE router. This preference means that regardless of whether the primary link comes back, the PE2 selects the locally sourced route as the best route.

Therefore, the PE2 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.)

```
PE2# show ip bgp 209.165.201.0
BGP routing table entry for 209.165.201.0/24, version 7
Paths: (2 available, best #1, table default, 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 (10.0.3.6)
 Origin incomplete, metric 0, localpref 100, weight 32768, valid, sourced,
best
 Community: 300:31000 no-export
 Local
 10.0.3.2 (metric 128) from 10.0.3.5 (1.0.0.2)
 Origin incomplete, metric 0, localpref 100, valid, internal
 Originator: 1.0.0.2, Cluster list: 10.0.3.5
 Community: 300:31000 no-export
```

- The BGP table on the service provider backup router contains the floating static route.

© 2015 Cisco and/or its affiliates. All rights reserved.

102

In this example, the **show ip bgp** command is used in the PE2 router of the provider to display the information about the customer network 209.165.201.0/24. The primary link has come back, so the PE2 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 PE2 received by [IBGP](#) from the PE1 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 PE1 router is not a candidate to be installed in the routing table. It 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

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.

## 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 that are redistributed into BGP, to make sure that other routes take precedence.
- Administrative distance cannot be matched with a route map; extra tags need to be defined for static routes.

© 2015 Cisco and/or its affiliates. All rights reserved.

103

Whenever you use floating static routes in combination with redistribution into BGP, you will need to take extra configuration steps. You have 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 that is assigned to the floating static route must be reduced (for example set to 0). Otherwise, the floating static route will always be selected as the best BGP route after the first failure of the primary link occurs.
- The router must also assign local preference values to the floating static route, so that the floating static route has a lower local preference than the primary route (for example 50). This assignment ensures that the primary route is selected as the best BGP route on other routers within Provider network.

These two requirements must be specified on the [PE](#) 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 for the route map to distinguish between normal static routes with normal weight and local preference and the floating static ones with modified values.

The solution is to create extra tag values for this set of static routes. The tag value must not only reflect the [QoS](#) level and whether to announce the route. The tag value must also indicate if it is a primary route or a backup route.

## Floating Static Routes with BGP (Cont.)

| Advertise Customer Route | Backup | QoS Type | Tag  | Community Values    | Local Preference |
|--------------------------|--------|----------|------|---------------------|------------------|
|                          |        | Normal   | 1000 | no-export 300:31000 | 100              |
|                          | Yes    | Normal   | 1010 | no-export 300:31000 | 50               |
| Yes                      |        | Normal   | 1001 | 300:31000           | 100              |
| Yes                      | Yes    | Normal   | 1011 | 300:31000           | 50               |
|                          |        | Gold     | 2000 | no-export 300:32000 | 100              |
|                          | Yes    | Gold     | 2010 | no-export 300:32000 | 50               |
| Yes                      |        | Gold     | 2001 | 300:32000           | 100              |
| Yes                      | Yes    | Gold     | 2011 | 300:32000           | 50               |

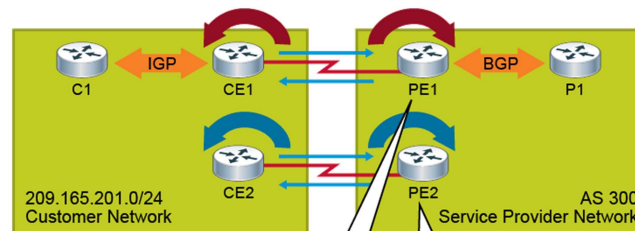
© 2015 Cisco and/or its affiliates. All rights reserved.

104

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 PE 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. It must set the appropriate weight and community and local preference values to these combinations.

## Floating Static Routes with BGP (Cont.)



```

PE1(config)# route-map IntoBGP permit 10
PE1(config-route-map)# match tag 1000
PE1(config-route-map)# set community no-export 300:31000
PE1(config-route-map)# set local-preference 100
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 20
PE1(config-route-map)# match tag 1001
PE1(config-route-map)# set community 300:31000
PE1(config-route-map)# set local-preference 100
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 30
PE1(config-route-map)# match tag 1010
PE1(config-route-map)# set community no-export 300:31000
PE1(config-route-map)# set local-preference 50
PE1(config-route-map)# set weight 0
PE1(config-route-map)# exit
PE1(config)# route-map IntoBGP permit 40
PE1(config-route-map)# match tag 1011
PE1(config-route-map)# set community 300:31000
PE1(config-route-map)# set local-preference 50
PE1(config-route-map)# set weight 0

```

The redistribution route map needs to be updated on all PE routers.

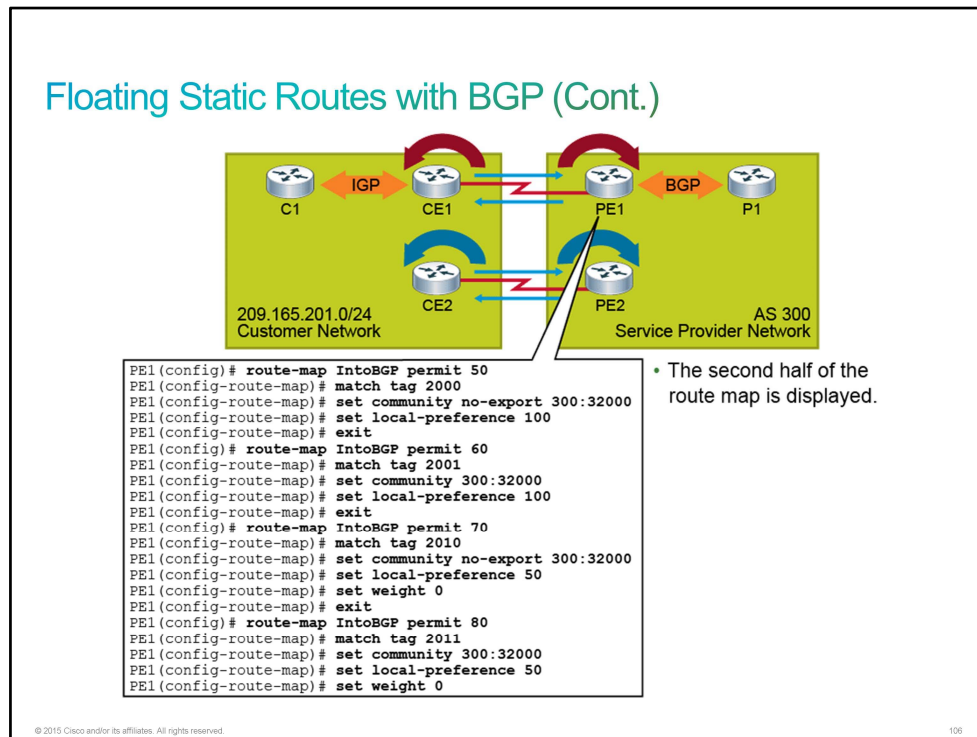
- Only the first half of the route map is displayed.

© 2015 Cisco and/or its affiliates. All rights reserved.

105

The configuration output in the figure displays the first half of the "IntoBGP" route map. The output shows how match clauses identify four of the eight different tags. For each of the tag values, the route map sets the community, the local preference, and, in some cases, the weight.

The displayed half of the route map deals only with the four tags that indicate QoS Normal. So all statements in the configuration display have set the BGP community attribute to 300:31000.



The second part of the route map that is shown deals with the four tags that indicate QoS Gold, which are configured to set the BGP community attribute to 300:32000.

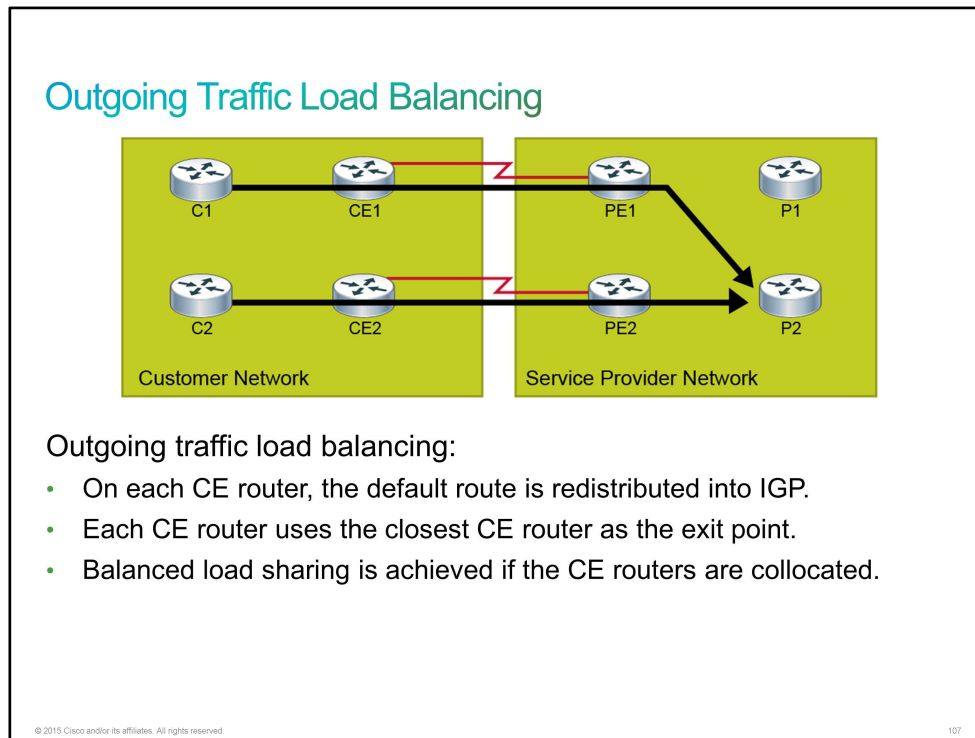
Tag values of 1000, 1010, 2000, and 2010 indicate that the route should not be explicitly propagated. The routes that a provider should not explicitly advertise to the rest of the Internet are assigned the **no-export** community by the route map.

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 PE 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

A customer that has two connections to a service provider can also decide to balance traffic over both links. If both links are active, traffic should be balanced across both links. If there is link failure, one link should care for all traffic.

## Outgoing Traffic Load Balancing

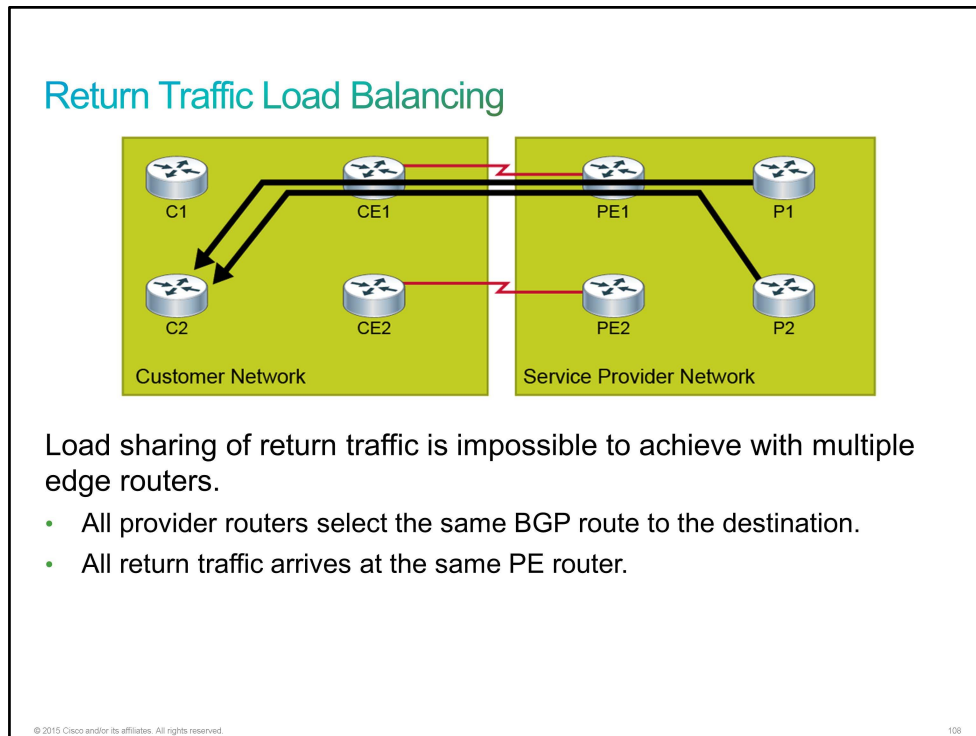


Load sharing of outgoing customer traffic is accomplished by configuring a standard default static route in both CE routers. Each static route is valid as long as the serial link from the CE router to PE router is up. When both static routes are valid, both CE routers announce the default route into 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 of both exit points.

## Return Traffic Load Balancing



The service provider routers know about two routes to the customer network. Those routes are learned through [BGP](#). BGP was designed to select only one route, which is the best route from a collection of routes to destination, allowing no load sharing. The PE routers that receive the same BGP route from two CE routers will always select the closer CE 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 PE routers select the same IBGP route based on router-ID (because the IGP metrics are always equal). So 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.

---

## Return Traffic Load Balancing (Cont.)

You can optimize return traffic load sharing:

- Each PE router advertises only part of the customer address space into the provider backbone.
- Every PE 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.**

© 2015 Cisco and/or its affiliates. All rights reserved.

109

To obtain better control of the return traffic load, the customer address space must be advertised to the PE 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 the 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.

## Load Sharing with Static Routes Example

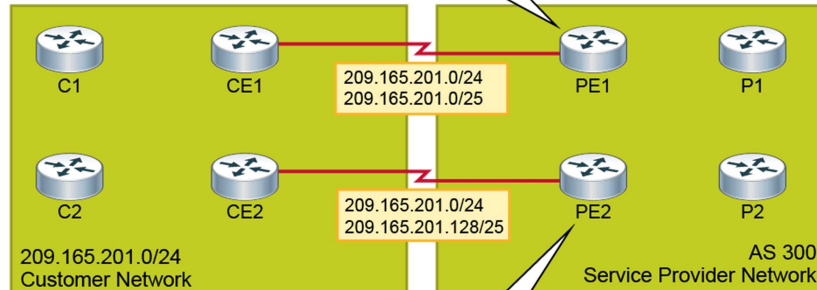
You can achieve load balancing of the incoming traffic by advertising different parts of the customer network by different PE routers. The upper PE router could advertise one-half of the address space, and the lower PE router could advertise the other half. For backup reasons, the PE routers also should advertise the entire address space as a larger route summary.

In the example, the customer address space 209.165.201.0/24 is partitioned into two smaller blocks: 209.165.201.0/25 and 209.165.201.128/25.



## Load Sharing with Static Routes Example

```
PE1(config)# ip route 209.165.201.0 255.255.255.128 Serial0 tag 1000
PE1(config)# ip route 209.165.201.0 255.255.255.0 Serial0 tag 1000
PE1(config)# route-map IntoBGP permit 10
PE1(config-route-map)# match tag 1000
PE1(config)# router bgp 300
PE1(config-router)# redistribute static route-map IntoBGP
```



```
PE2(config)# ip route 209.165.201.128 255.255.255.128 Serial0 tag 1000
PE2(config)# ip route 209.165.201.0 255.255.255.0 Serial0 tag 1000
PE2(config)# route-map IntoBGP permit 10
PE2(config-route-map)# match tag 1000
PE2(config)# router bgp 300
PE2(config-router)# redistribute static route-map IntoBGP
```

© 2015 Cisco and/or its affiliates. All rights reserved.

110

The PE1 router advertises the route to 209.165.201.0/25, and the PE2 router advertises the route to 209.165.201.128/25. Both PE routers also advertise the entire address space 209.165.201.0/24.

As long as both paths are available, the traffic from service provider to the customer uses the most explicit route. Two explicit routes are used to send traffic. Traffic to destinations in the 209.165.201.0/25 range is sent to the upper connections point. Traffic to destination in the 209.165.201.128/25 range is directed to the lower connection point.

---

**Note** 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.

---

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Static routing can be used for:
  - Customers with single connection to service provider.
  - Customers with multiple connections to service provider—primary and backup path, or load balancing.
- 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 ASs.
- When you are using static routes in a backup scenario, floating static routes are used on the backup routers.
- 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.

# Lesson 3: Connecting a 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 are correctly configured. They must ensure proper interaction between the customer and the service provider network. It is also important to understand how to configure routing protocols so that customer backup or load-balancing requirements are met.

In this lesson, you will learn about the use of multiple connections between a customer and a single [ISP](#) for backup and load-sharing purposes.

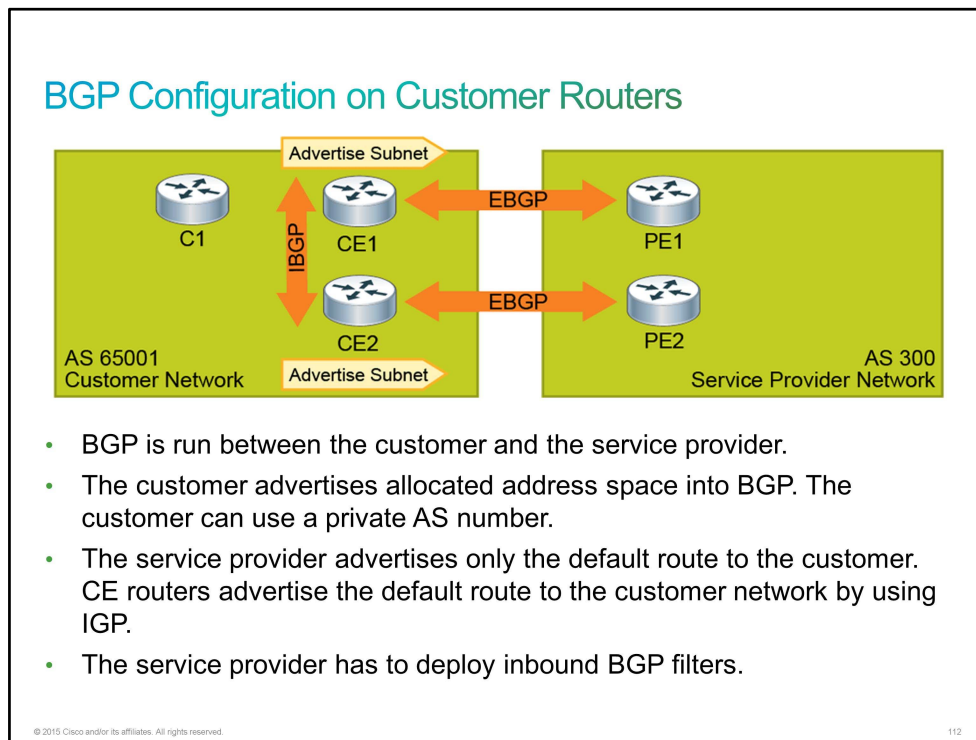
Upon completing this lesson, you will be able to:

- 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

# BGP Configuration on Customer Routers

In the example, the customer network is connected to a service provider network by 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 the 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.

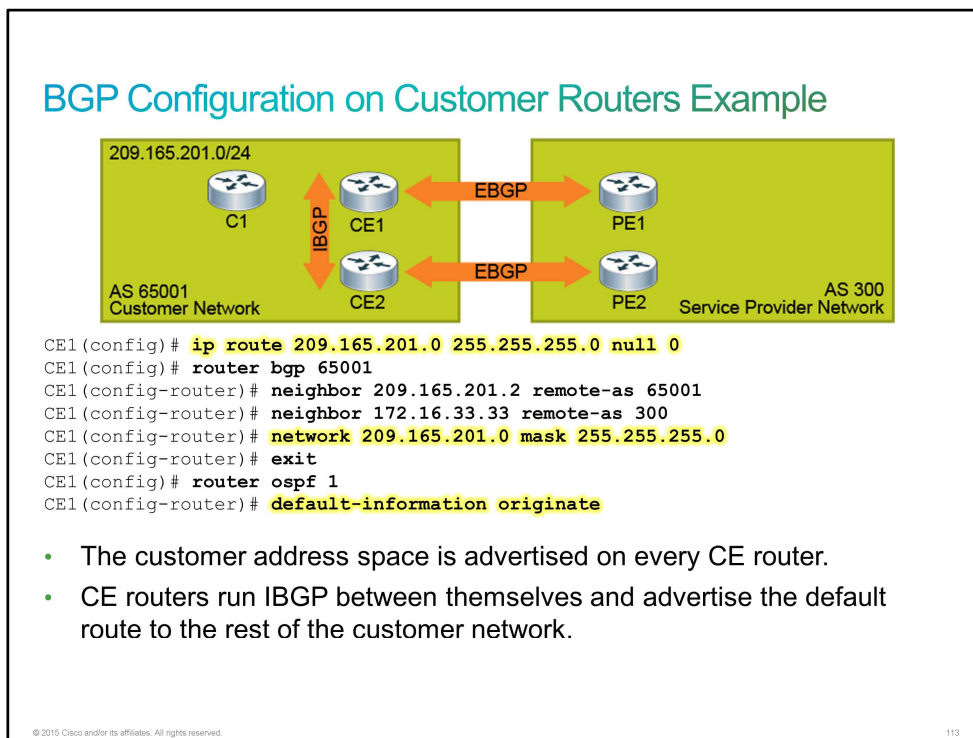
There is a significant difference in this case, as compared to a network scenario where static routes and redistribution are used. Routers within the private AS of the customer now advertise the customer routes via BGP. Thus, the customer is responsible for announcing its own address space. The ISP receives routes from the customer and conditionally propagates them (similar to static routing). If the customer uses [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.

The customer is now creating BGP routes that the ISP receives. So any error that the customer makes can influence routing operation within the provider 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 comes 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.

# BGP Configuration on Multihomed Customer Routers Example

In the figure, the customer has been assigned the private AS number 65001.

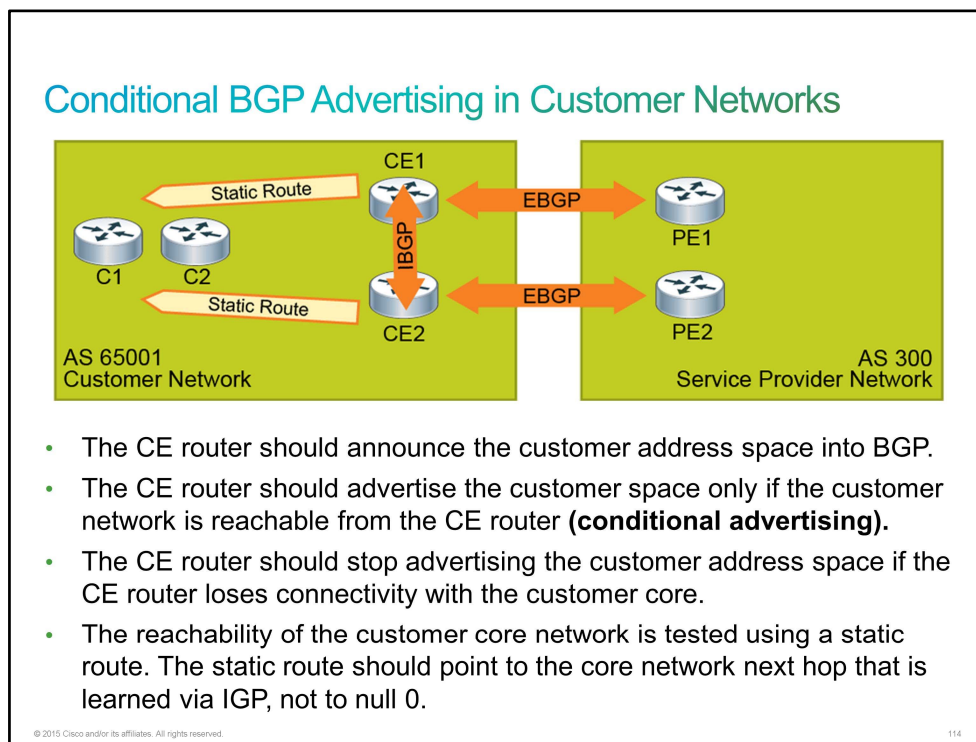


Both [CE](#) routers are configured to run BGP and should advertise all of the customer networks with the **network** command. If only one router advertises the network, a single point of failure has been introduced. The two customer edge routers must also run [IBGP](#) between them to make common decisions regarding BGP routing information.

Each CE router has an [EBGP](#) session with the [PE](#) 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 [IGP](#), in this case [OSPF](#), of the customer.

# Conditional BGP Advertising in Customer Networks

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 [CE](#) routers using the **network** command. The appearance of a corresponding network or subnet in the routing table of the edge router conditions the route advertisement. Because there is a chance that the customer subnetted the address space into smaller subnets, the complete customer address space will not be presented in the CE routing table. The solution is to manually enter the route into the routing table by configuring a static route to null 0. However, the static route is always present in the routing table. And the BGP advertisement is always performed from the CE router, even if the CE router cannot access the rest of the customer network.



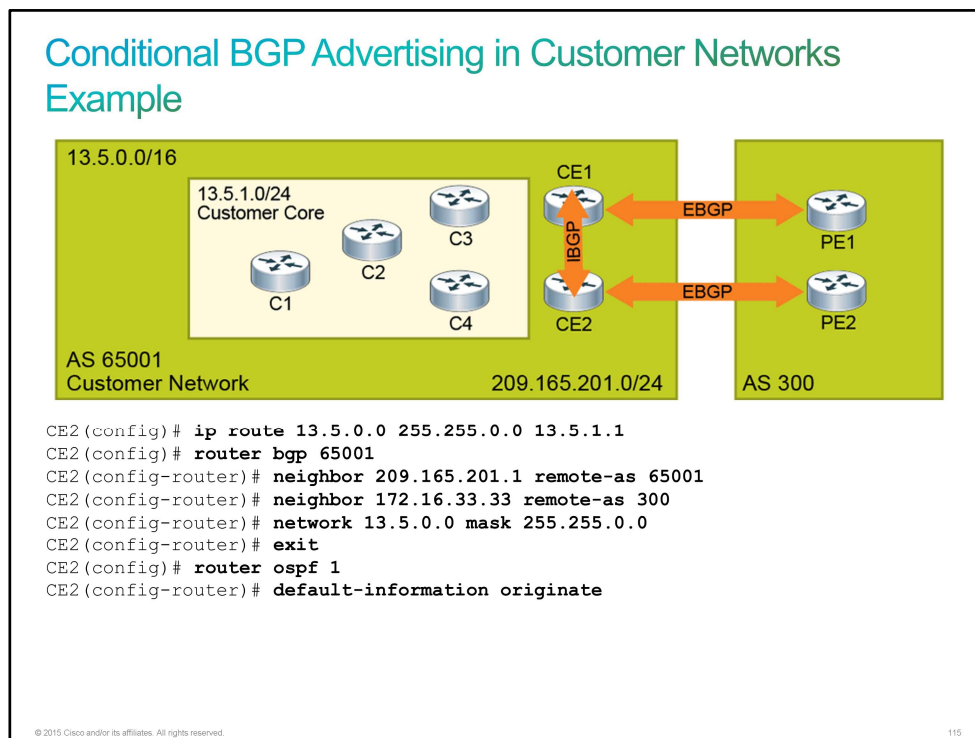
If the CE router loses connectivity to the rest of the customer network but is still connected to the [ISP](#) network, the BGP advertisement must cease. In this case, BGP advertisement can be stopped if BGP advertisements are bound to the reachability status of a specific subnet in the core of the customer 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 CE 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 CE router. Traffic is sent to this router, but because the router has lost connectivity with the rest of the network, the traffic is dropped. Or the traffic is routed to the null 0 interface using the static route.

The solution is to create a static route for the customer-assigned addresses, which points to a next hop that was learned through an IGP. The static route will be in the routing table only if the next hop is reachable. The static route would be valid and the route would be advertised through BGP. If the CE router loses connectivity from the rest of the customer network, the next hop becomes unavailable and the static route is not valid anymore. Because the customer route is not present in the routing table anymore, BGP will cease to advertise the customer route to the ISP. The customer network will then be accessible over the backup connection, if available.

## Conditional BGP Advertising in Multihomed Customer Networks Example

In this example, the customer network also uses the address space 13.5.0.0/16.



The address space is further subnetted at the customer site. One of the subnets (subnet 13.5.1.0/24) is identified as being the central part of the customer core network.

The CE routers participate in the IGP routing of the customer. This participation means that these routers have information about which 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 CE routers is configured to advertise the full address space that the customer uses. When the CE routers advertise this route, the ISP network, and the rest of the Internet, see customer's complete address space as one single route, 13.5.0.0/16.

The appearance of the static route, IP route 13.5.0.0 255.255.0.0 13.5.1.1 conditions the advertisement of the customer address space by BGP. If the static route is valid, then the BGP route 13.5.0.0/16 is advertised. The static route is a recursive route—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 IGP announces the subnet 13.5.1.0/24. 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.



---

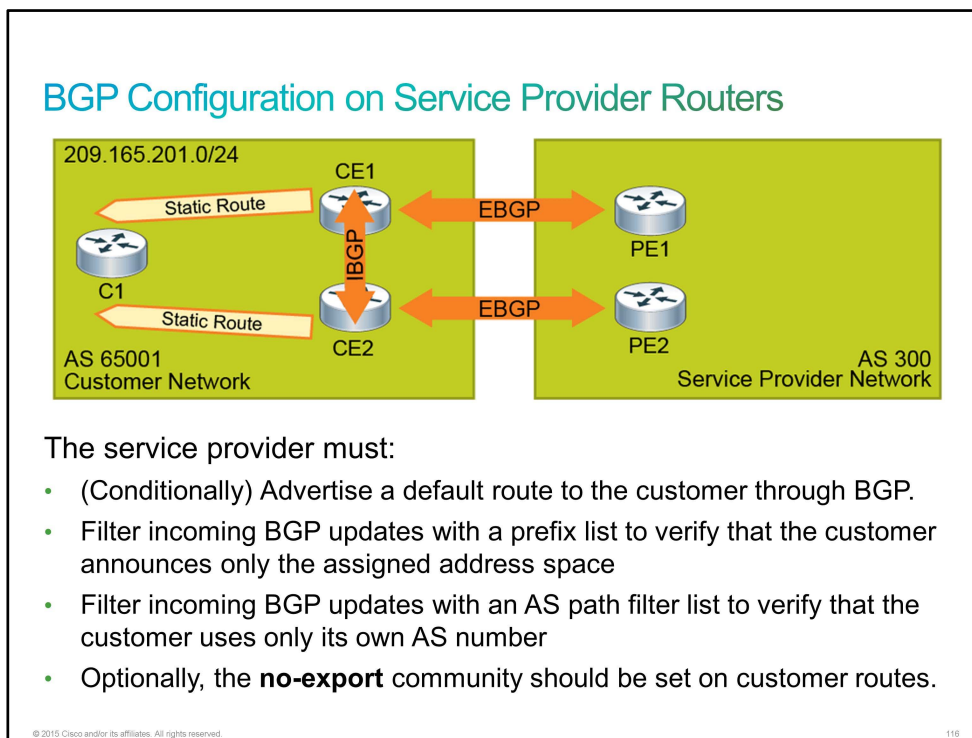
|             |                                                                                                                                                                  |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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 CE routers.

# BGP Configuration on Service Provider Routers

In the [ISP](#) network, the two [PE](#) routers must have [BGP](#) sessions configured with the customer. There is no point in feeding the full Internet routing table to the customer. 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 filter that allows only customer-assigned routes and denies any other route. This filter ensures that private address space or any other illegal networks that the customer erroneously announces, 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 should be allowed. A filter list performs this check.

If the customer address space is [PAaddress](#) space and it represents only a small part of a larger block that the ISP announces, the explicit BGP routes that are received from the customer do not need to be advertised to the rest of the Internet. The ISP can announce the large 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 [PE](#) router can tag the BGP routes that are received from the customer with the **no-export** well-known community. This community prevents the ISP from sending the routes to any other AS.

## BGP Configuration 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.

The default route is not sent through the outbound BGP filters (prefix list, filter list, or route map).

© 2015 Cisco and/or its affiliates. All rights reserved.

117

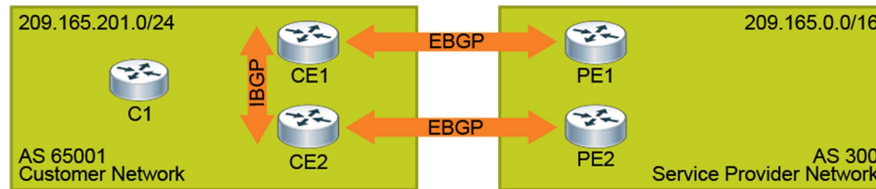
The default route, 0.0.0.0/0, is not advertised in outgoing BGP updates unless it is explicitly configured. The **neighbor default-originate** router configuration command is used to initiate the advertisement of the default route to a neighbor.

BGP does no checking 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.

## BGP Configuration on Service Provider Routers Example

This example shows the configuration of the PE router.

## BGP Configuration on Service Provider Routers Example



```

PE1(config)# router bgp 300
PE1(config-router)# neighbor 172.16.33.3 remote-as 65001
PE1(config-router)# neighbor 172.16.33.3 default-originate
PE1(config-router)# neighbor 172.16.33.3 prefix-list DefaultOnly out
PE1(config-router)# neighbor 172.16.33.3 prefix-list Customer1 in
PE1(config-router)# neighbor 172.16.33.3 filter-list 15 in
PE1(config-router)# neighbor 172.16.33.3 route-map AllCustomerIn in
PE1(config-router)# exit
PE1(config)# ip as-path access-list 15 permit ^65001(_65001)*$
PE1(config)# ip prefix-list DefaultOnly permit 0.0.0.0/0
PE1(config)# ip prefix-list Customer1 permit 209.165.201.0/24 le 32
PE1(config)# ip prefix-list Provider permit 209.165.0.0/16 le 32
PE1(config)# route-map AllCustomersIn permit 10
PE1(config-route-map)# match ip prefix-list Provider
PE1(config-route-map)# set community no-export additive
PE1(config-route-map)# exit
PE1(config)# route-map AllCustomersIn permit 1000

```

© 2015 Cisco and/or its affiliates. All rights reserved.

118

The customer is assigned the private AS number 65001. The BGP session is opened with the customer [IP address](#), 172.16.33.3.

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.

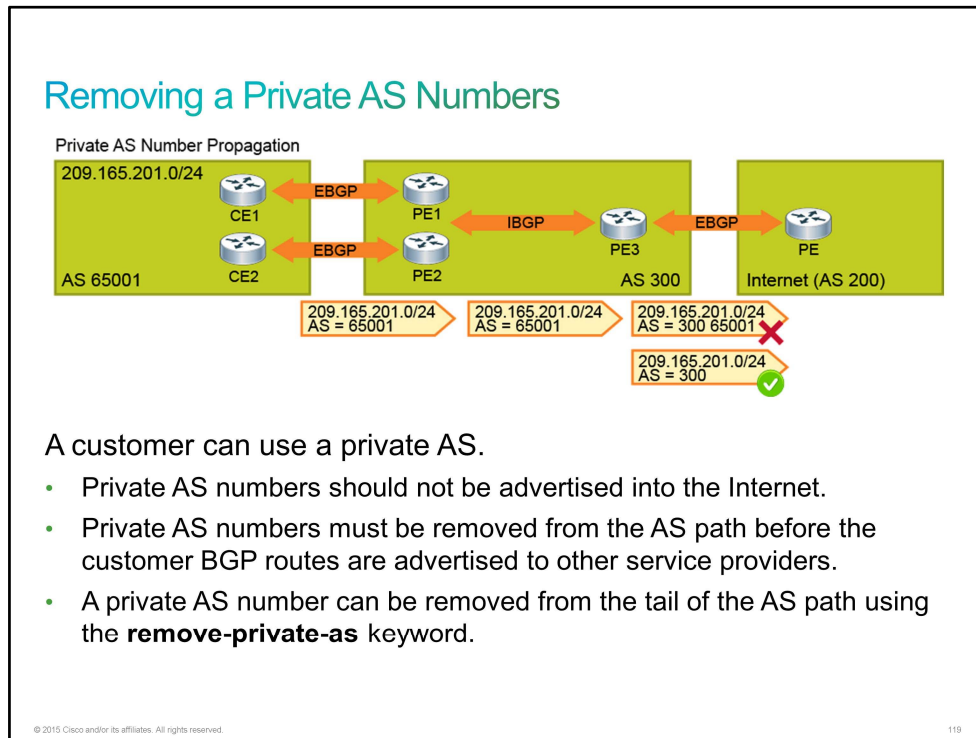
Received routes from the customer must first pass the prefix list Customer1. There is one dedicated prefix list for each individual customer; the prefix list permits only the routes that the customer is allowed to announce. If the prefix list allows the routes, the routes must also pass the filter list named "15." 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 AS path prepending configurations on the customer side. If both prefix lists and the filter list allow the received route, 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. 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 a Private AS Numbers

Routes that the [ISP](#) receives from the customer are propagated to the rest of the Internet only if they are part of the [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 the service provider propagates the customer routes 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. The service provider has to make sure that the private AS number is removed from the BGP route and that the service provider AS number is present in the tail of the AS path.

---

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 occurs on the EBGP session between AS 300 and the Internet.

## Removing a Private AS Numbers (Cont.)

```
router(config-router)# neighbor ip-address remove-private-as
```

- The command modifies AS path processing on outgoing updates that are sent to the 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

120

To remove private AS numbers from the AS path in outbound routing updates, use the **neighbor remove-private-as** router configuration command. To disable this function, use the **no** form of this command.

```
neighbor {ip-address | peer-group-name} remove-private-as
```

### Syntax Description

| Parameter              | Description                                             |
|------------------------|---------------------------------------------------------|
| <i>ip-address</i>      | <a href="#">IP address</a> of the BGP-speaking neighbor |
| <i>peer-group-name</i> | Name of a BGP peer group                                |

Use this command on the service provider egress routers. Before the service provider advertises any of the customer routes of the ISP to the rest of the Internet, the private AS numbers must be removed. The command removes those AS numbers if they are at the tail end of the AS path.

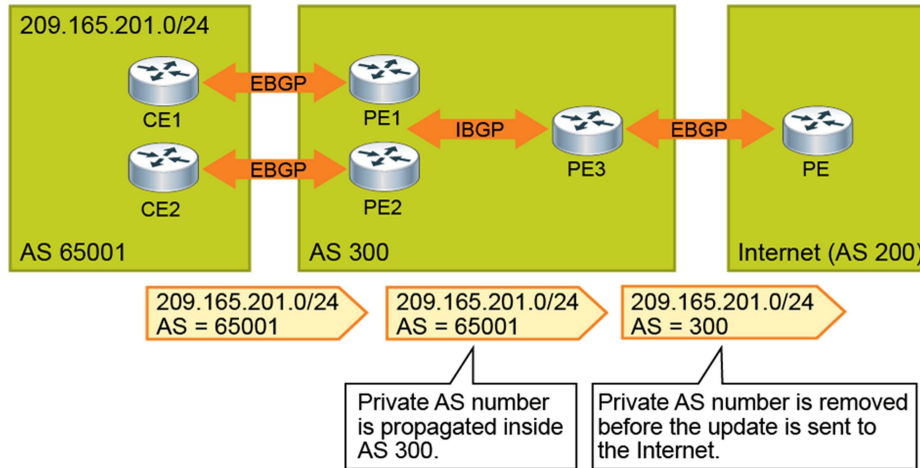
|             |                                                                                                                                             |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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.

## Removing a Private AS Numbers Example

In this example, the service provider AS (300) receives routes from the customer.

## Removing a Private AS Numbers Example



```
PE3(config)# router bgp 300
PE3(config-router)# neighbor 10.2.3.3 remote-as 200
PE3(config-router)# neighbor 10.2.3.3 remove-private-as
```

© 2015 Cisco and/or its affiliates. All rights reserved.

121

The ISP assigns the private AS number 65001 to the customer; therefore, the routes that the provider receives 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 200 in this example).

The PE3 router in AS 300 has been configured to remove private AS numbers on EBGp routes toward AS 200. If private AS numbers appear in the tail end of the AS path (before AS 300 is added), they are removed.

This configuration must be applied to all egress routers in AS 300 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

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. However, the amount of configuration that is required in the customer networks can make this migration a cumbersome task that is difficult to complete without disrupting service.

## BGP Support for Dual AS Configuration for Network AS Migrations

- Allows you to merge a secondary AS under the 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 ASs and then later migrate customers to new configurations during normal service windows without disrupting existing peering arrangements
- If misconfigured, increases the possibility of creating routing loops

© 2015 Cisco and/or its affiliates. All rights reserved.

122

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 AS 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 that are received from EBGp neighbors. Use this command in address-family or router configuration mode. To disable AS path attribute customization, use the no form of this command.

```
neighbor ip-address local-as [as-number[no-prepend [replace-as [dual-as]]]]
```



## Syntax Description

| Parameter         | Description                                                                                                                                                                                                                              |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i> | Specifies the <a href="#">IP address</a> of the EBGP neighbor.                                                                                                                                                                           |
| <i>as-number</i>  | Specifies an AS number to prepend to the AS path attribute.<br>The range of values for this argument is any valid AS number from 1 to 65535.                                                                                             |
| <b>no-prepend</b> | (Optional) Does not prepend the local AS number to any routes received from the EBGP neighbor.                                                                                                                                           |
| <b>replace-as</b> | (Optional) Prepends only the local AS number to the AS path attribute. The AS number from the local BGP routing process is not prepended.                                                                                                |
| <b>dual-as</b>    | (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 that is configured with the <b>ip-address</b> argument ( <b>local-as</b> ) |

---

|             |                                                                                                                                                                                                                                                                                                                                                          |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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. |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

---

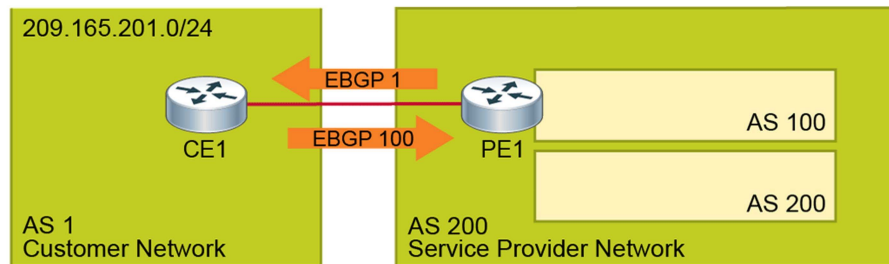
|             |                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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 Dual AS Configuration for Network AS Migrations feature is used to merge two autonomous systems without interrupting peering arrangements with the customer network.

## Dual-AS Configuration Example



- If a service provider migrates to a new AS number, customers have to change the BGP configuration.
- A merged service provider can present itself to the customer using the old AS number, so the customer is not required to change the BGP configuration.

© 2015 Cisco and/or its affiliates. All rights reserved.

123

The **neighbor local-as** command is configured to allow PE1 to maintain peering sessions through AS 100 and AS 200. CE1 is a customer router that runs a BGP routing process in AS 1 and is configured to peer with AS 200.

```
PE1(config)# router bgp 100
PE1(config-router)# no synchronization
PE1(config-router)# neighbor 172.16.33.3 remote-as 1
PE1(config-router)# neighbor 172.16.33.3 local-as 200 no-prepend replace-as dual-as

PE1(config)# router bgp 200
PE1(config-router)# neighbor 172.16.33.3 remote-as 1

CE1(config)# router bgp 1
CE1(config-router)# neighbor 172.16.33.34 remote-as 200
```

After the transition is complete, the configuration on CE1 can be updated to peer with AS 100 during a normal maintenance window or during other scheduled downtime.

```
CE1(config-router)# neighbor 172.16.33.34 remote-as 100
```

## Dual-AS Confederation Configuration Example

The following example can be used in place of the PE1 configuration in the previous example. The only difference between these configurations is that in this example PE1 is configured to be part of a confederation.

```
PE1(config)# router bgp 65100
PE1(config-router)# no synchronization
PE1(config-router)# bgp confederation identifier 100
PE1(config-router)# neighbor 172.16.33.3 remote-as 1
PE1(config-router)# neighbor 172.16.33.3 local-as 200 no-prepend replace-as dual-as
```

## Replace-AS Configuration Example

The following example strips private AS 65100 from outbound routing updates for the 172.16.33.3 neighbor and replaces it with public AS 1.

```
PE1(config)# router bgp 65100
PE1(config-router)# neighbor 172.16.33.3 local-as 1 no-prepend replace-as
```

# Backup Solutions with BGP

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 or backup scenario or in a load-sharing scenario.

## Backup Solutions with BGP

The CE routers control the route selection entirely.

- Outgoing traffic:
  - Local preference is used to select primary or backup link.
- Incoming traffic:
  - MED is used to advertise the primary or backup link to the ISP. This is not reliable, because the ISP can change the local preference.
- No service provider configuration is required.

© 2015 Cisco and/or its affiliates. All rights reserved.

124

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 [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 worse (or high) MED value on the backup link is set. Also a good (or low) value on the primary link is set.

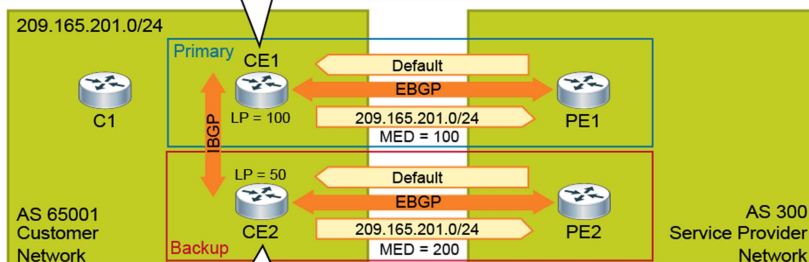
The receiving [EBGP](#) peer checks MED and AS path length only if the weight and local preference attributes have not been configured. In this case, the ISP should not use any of these configuration options. The ISP should rely solely on the attributes that it has received from the customer. However, if the ISP sets, for example, the local preference for customer routes, the ISP will control the path selection for traffic to the customer. This path selection is why you cannot rely on the fact that your traffic distribution scenario will work as intended.

## Primary/Backup Link Selection Example

The customer is connected to the ISP over two permanent connections. The customer uses the upper connection as the primary connection and the lower connection as the backup.

## Primary/Backup Link Selection Example

```
CE1(config)# router bgp 65001
CE1(config-router)# bgp default local-preference 100
CE1(config-router)# neighbor 172.16.33.33 remote-as 300
CE1(config-router)# neighbor 172.16.33.33 route-map LowMED out
CE1(config-router)# exit
CE1(config)# route-map LowMED permit 10
CE1(config-route-map)# set metric 100
```



```
CE2(config)# router bgp 65001
CE2(config-router)# bgp default local-preference 50
CE2(config-router)# neighbor 172.16.33.34 remote-as 300
CE2(config-router)# neighbor 172.16.33.34 route-map HiMED out
CE2(config-router)# exit
CE2(config)# route-map HiMED permit 10
CE2(config-route-map)# set metric 200
```

© 2015 Cisco and/or its affiliates. All rights reserved.

125

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 CE1 router is configured to set local preference to the value 100 on all EBGP routes that are received. The CE2 router sets the local preference attribute to a value of 50. This configuration means that the outgoing traffic toward any destination that the ISP announces is primarily sent over the upper link. It makes the primary router the preferred one.

Incoming traffic to the customer is directed to the primary link by using the MED. On the CE1 router, all routes that are sent to the ISP have their MED attribute set to the value 100 by the route map named "LowMED out." On the CE2 router, all routes that are sent to the ISP have their MED attribute set to the value 200 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. This makes the primary router the preferred one from the service provider point of view.

# Load Sharing

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

You can implement load sharing of return traffic in a number of ways:

- Outgoing traffic:
  - Load sharing is identical to the static routing scenario. The IGP determines it.
- Incoming traffic:
  - 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

126

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 [CE](#) 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. 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. 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 regarding the IGP of the ISP. If an ISP router has an equivalent distance to both connection points, the use of the **maximum-paths** (or [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

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.

## Load Sharing with BGP Multipath

### BGP Multipath:

- By default, BGP selects a single path as the best path and installs it in the IP routing table.
- If configured, a BGP router can select up to six identical BGP routes as the best routes.

```
router(config-router)# maximum-paths number
```

- 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.

© 2015 Cisco and/or its affiliates. All rights reserved.127

A router can use either only [IBGP](#) routes or only [EBGP](#) routes for load balancing but not both at the same time because this could lead to routing loops. BGP multipath can be enabled separately for IBGP and EBGp routes. A 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.

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. To restore the default value, use the **no** form of this command.

```
maximum-paths number
```

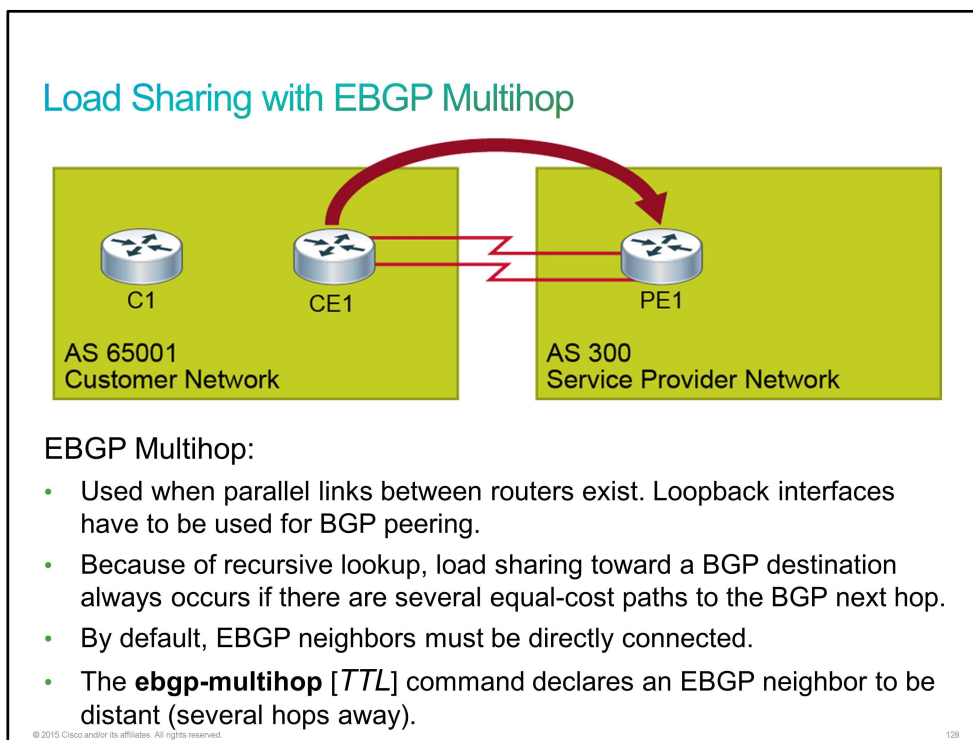
## Syntax Description

| Parameter     | Description                                                                                                        |
|---------------|--------------------------------------------------------------------------------------------------------------------|
| <i>number</i> | Maximum number of parallel routes that an IP routing protocol installs in a routing table, in the range of 1 to 6. |

|             |                                                                                                                                                                                                             |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | Load sharing between alternative BGP routes is achieved only if the EBGp routes are identical according to all BGP route selection rules and <b>maximum-paths</b> is configured with a value larger than 1. |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

# Load Sharing with EBGP Multihop

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. It is also used in the recursive lookup when the router installs the routes 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.

The load balancing is automatic as a result of the recursive route lookup of BGP next hops. The IP routing table will contain a single routing entry to reach the customer prefix but the Cisco Express Forwarding (CEF) table (which contains the results of the recursive lookup) will have two equal-cost paths to reach the customer prefix.

By default, EBGP neighbors must be directly connected. A router 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 [TTL](#) value set to the value 1. This value means that they cannot be routed.

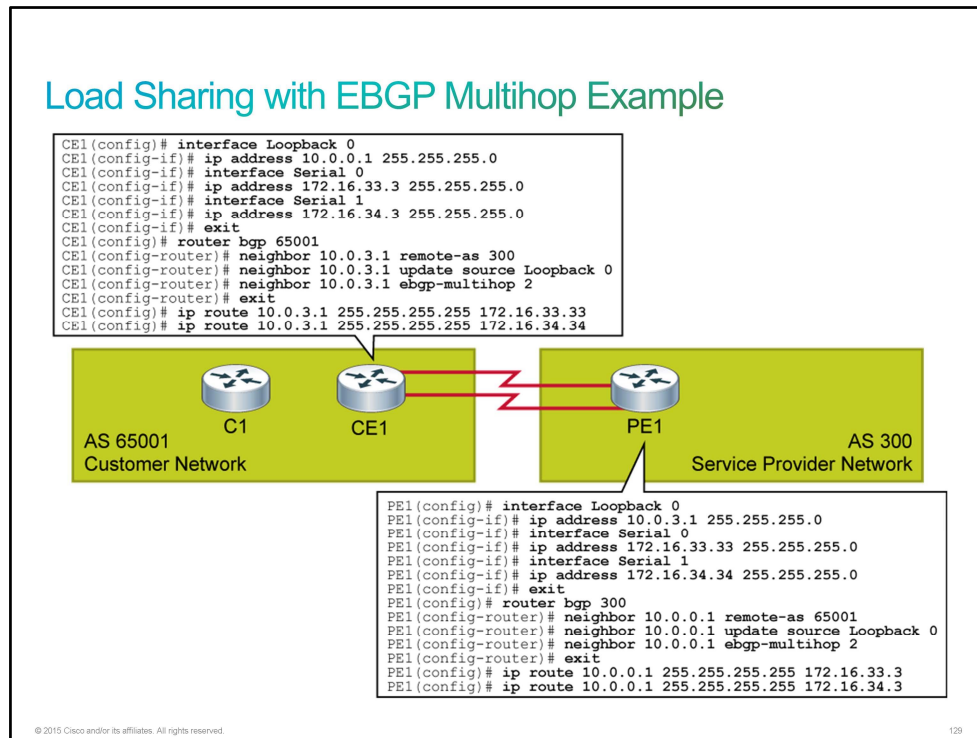
The **ebgp-multihop** neighbor configuration command changes this behavior. Although the neighbor is several hops away, the session goes into the Active state, and packets start to be exchanged. The TTL value of the IP packets is set to a value larger than 1. If no value is specified on the command line, 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**. 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.

## Load Sharing with EBGP Multihop Example

In the figure, the customer network and the [ISP](#) network are connected via single router on each side. Two parallel links are used.



In this case, only one EBGP session is configured between the customer and provider routers. The session should be established from the loopback interface on one router to the loopback interface on 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 10.0.3.1. The [CE](#) router performs a recursive lookup and finds that it can reach 10.0.3.1 via 172.16.33.33 and via 172.16.34.34. 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 that were 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.
- Customer does not necessarily use a public AS number, it can also use private AS numbers that range from 64512 to 65535. BGP must remove private AS numbers.
- The ISP should advertise a default route to the customer through BGP. The provider should also use incoming filters to ensure that the customer advertises only the correct address space and AS number.
- 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.
- By announcing portions of its address space, a customer can use maximum paths and EBGp multihop to provide load sharing over multiple links.

© 2015 Cisco and/or its affiliates. All rights reserved.

130

# 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 [AS](#) number selection are important considerations that affect the network implementation. It is also important to understand how to configure routing protocols so that customer backup or load-sharing requirements are met.

In this lesson, you will discuss multiple connections between a customer and multiple service providers for backup and load-sharing purposes. Including the [BGP](#) characteristics that are used to configure such connectivity. You will also discuss address selection, private AS number translation, and configuration of the network to support either backup links or load sharing.

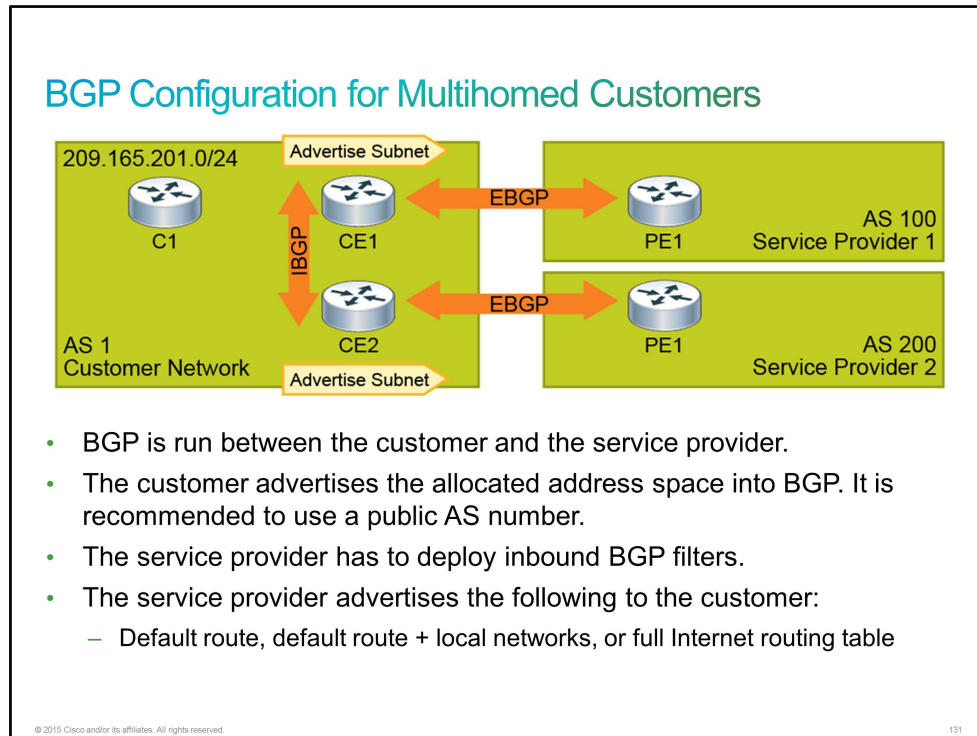
Upon completing this lesson, you will be able to:

- 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

# BGP Configuration for Multihomed Customers

The highest level of redundancy 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.



Because the customer has to use [PI address](#) space, it is the responsibility of the customer to advertise the customer address space. The only routing option that can be used in this scenario is [BGP](#) because service providers run only BGP with customers. 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.
- Edge router failure on either the customer or the service provider side.
- Link failure or router failure within the customer network that causes the [CE](#) 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 [PE](#) router to lose connectivity with the rest of the Internet

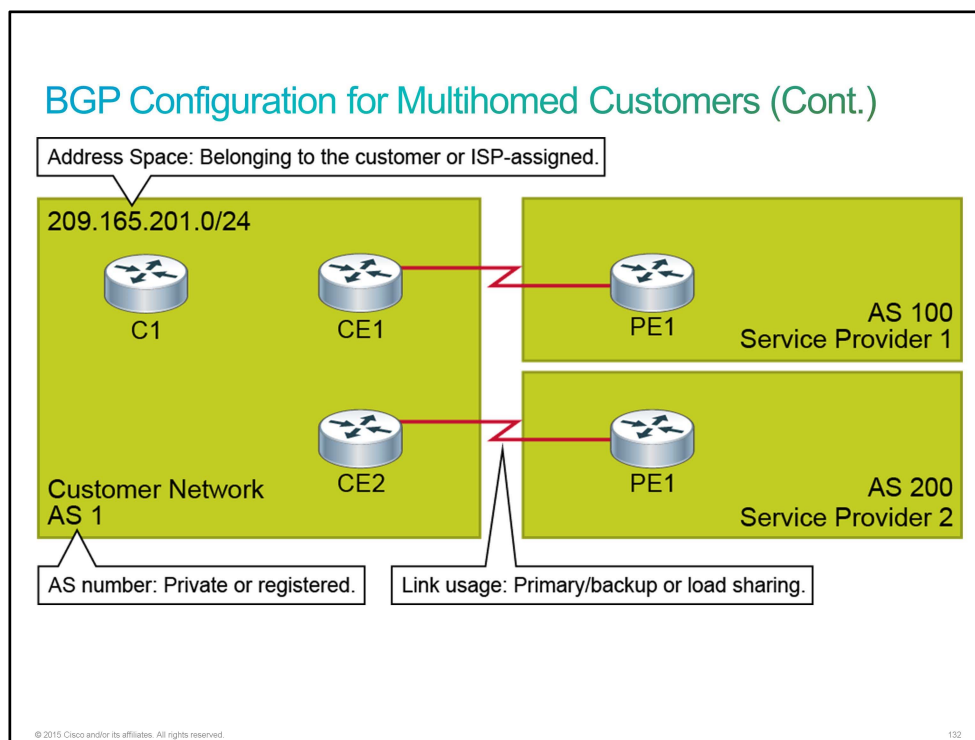
Multihomed customers have multiple permanent links to different ISPs. The links should terminate in different edge routers in the customer network. Otherwise, one of the major advantages, resilience to router failure, is lost.

Multihomed customers should use BGP with both ISPs. The customer should advertise its address space to both providers. The route advertisement should be configured in both CE 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 the one configured when you are connecting a multihomed customer network to a single service 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. These filters will 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 PE 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 the customer requirements, the volume of BGP routes the ISP provides can vary. It can be the default route only, the default route with local routes as well, or the full Internet routing table.



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 [PA](#) or [PI](#)?
- What AS number is the customer using? (Is the customer using a public or a private AS number?)

# Multihomed Customer Address Space Selection

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.

## 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 the reverse path.

© 2015 Cisco and/or its affiliates. All rights reserved.

133

However, if the customer uses a small block of addresses that one of the [ISPs](#) assigned, an alternative design that does not involve [BGP](#) exist. The customer can use two different [PA address](#) spaces and do [NAT](#). With NAT, the router translates traffic going out over one of its connections to one of the PA addresses. If the 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

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 from 64512 to 65535 is sufficient.

## 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: Each service provider assigns a private AS number to the customer and the customer then uses one of them internally; the others have to be translated.
- Causes inconsistent routing information.

© 2015 Cisco and/or its affiliates. All rights reserved.

134

Whenever the customer is assigned a public AS number, there are no conflicts in the BGP setup because the number is guaranteed to be unique within the Internet. Both, the customer and the service provider make route announcements without tampering with the AS path. As a result, the service provider propagates a consistent AS path information 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 make this assignment independently of each other. 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 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.

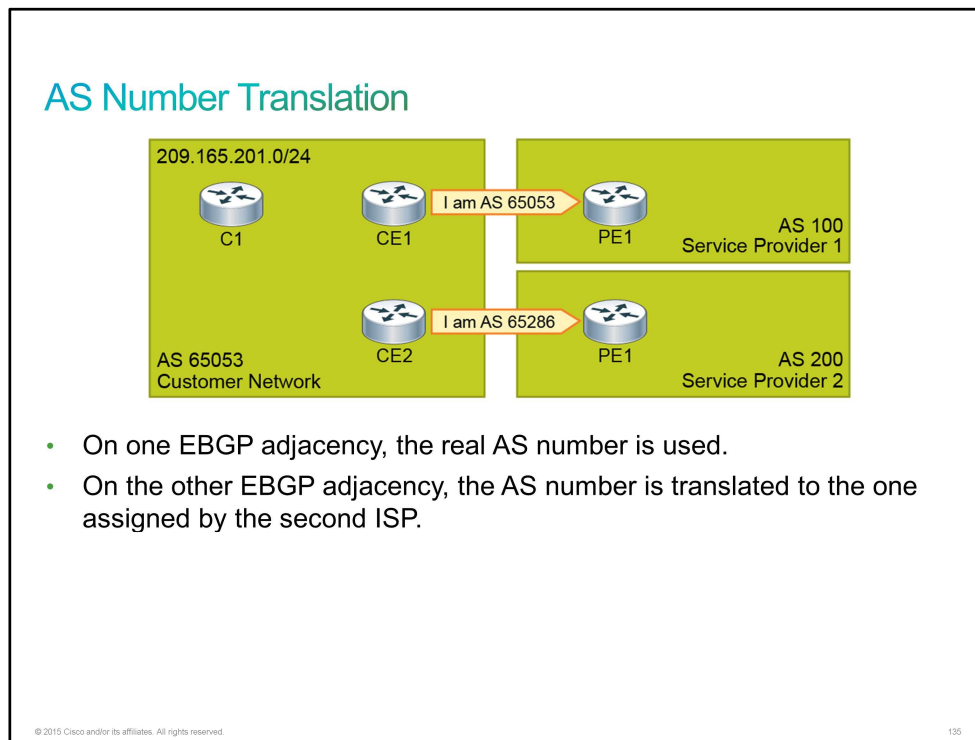
No router should ever propagate private AS numbers to the rest of the Internet. An ISP can keep track of the private AS numbers that it has assigned to its customers and avoid reuse or conflicts within that scope. However, as soon as the scope is widened to include other ISPs, conflicts will happen. 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 first propagated into each of the autonomous systems of its ISPs. In the next step, the private AS number is removed from the routes, as the routes are propagated to the rest of the Internet. Now the customer routes appear to be originating in the ASs 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 ASs.

# AS Number Translation

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 1 has assigned the private AS number 65053 to the customer. Service provider 2 did not agree to use this private AS number when connecting to the customer. Instead, service provider 2 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 [EBGP](#) session to AS 200.

In the example, Service Provider 1 (AS 100) has an EBGP session to the customer where the AS number 65053 is used at the customer end. Service provider 2 (AS 200) 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 [CE](#) router as part of the EBGP session to AS 200.

## AS Number Translation (Cont.)

```
router(config-router)# neighbor ip-address local-as private-as [no-prepend
[replace-as]]
```

- Optionally, the service providers can assign two different private AS numbers to the customer.
- Internally, the customer can use an ISP-assigned AS number or even any other private AS number.
- Externally, the customer appears as one private AS to ISP 1 and as a different private 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

136

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.

When a keyword **no-prepend** is used, router does not prepend the local autonomous system number to any routes received from the EBGP neighbor.

When a keyword **replace-as** is used, router replaces the real autonomous system number with the local autonomous system number in the eBGP updates. The autonomous system number from the local BGP routing process is not prepended.

---

|             |                                                                                                                                                                                                              |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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 and Backup 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 and backup, or a load-sharing configuration.

In the primary and backup scenario, only the default route is required from each service provider. The customer implements routing policies.

## Primary and Backup Link Selection

- **Outgoing link selection:**
  - You can use the same solution as with multihomed customers connected to one service provider.
  - A higher local preference for the default route comes from the primary 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.
- Only the default route is required from both service providers.

© 2015 Cisco and/or its affiliates. All rights reserved.

137

If one link is primary and the other is backup, the customer can use the local preference attribute in the configuration to direct all outgoing traffic over the primary link. The customer selects a primary service provider for outgoing traffic by setting a higher local preference for routes that are received from the preferred [ISP](#). 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 [MED](#) when the customer connects to multiple providers because the updates are sent to two different [ASs](#). 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. The customer can configure AS path prepending of routes that are sent to the backup service provider to lengthen the AS path. The returning traffic originating from the Internet will therefore choose a shorter path through the primary service provider.

# BGP Incoming Link Selection

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.

## BGP Incoming Link Selection

### BGP communities:

- Customer sets the appropriate BGP community attribute on updates that are 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.

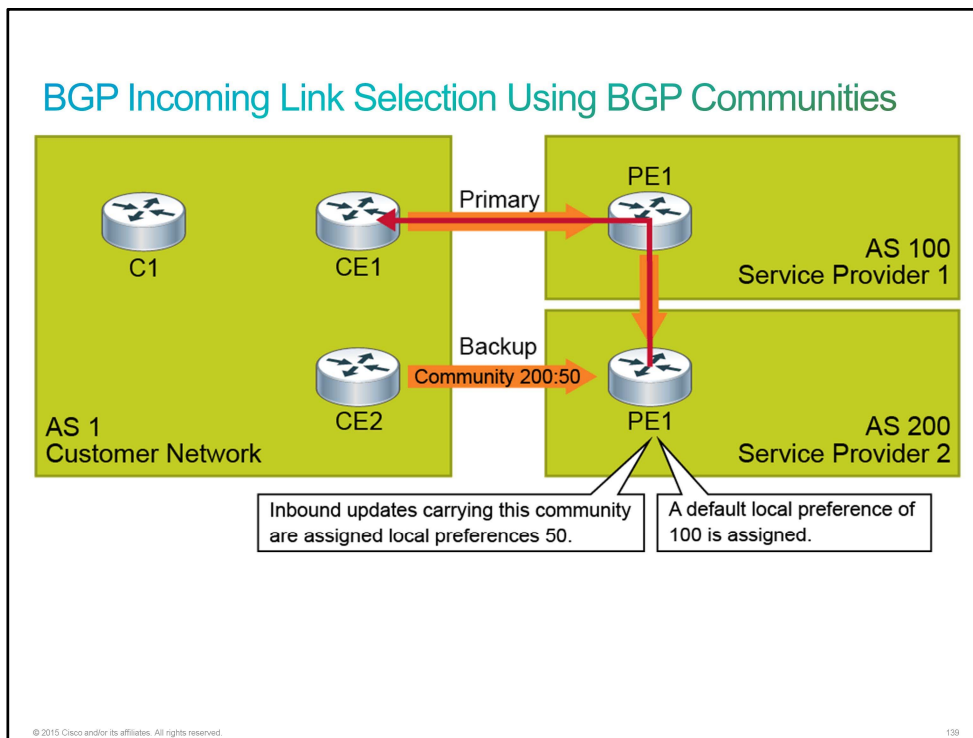
© 2015 Cisco and/or its affiliates. All rights reserved.138

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



In the example here, the backup service provider 2 (AS 200) has defined the meaning of community 200:50. When AS 200 receives routes with this community, the local preference is set to 50.

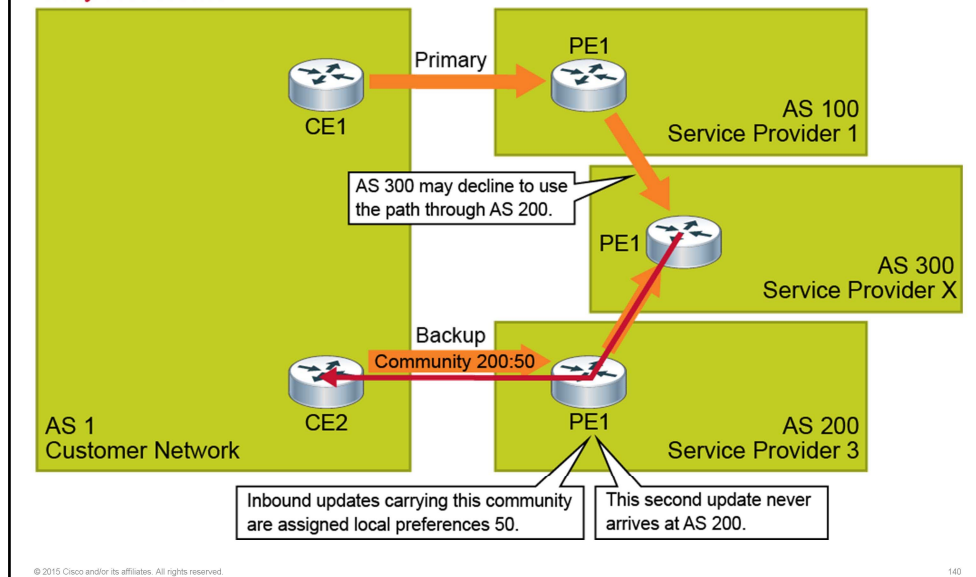
The customer AS 1 is advertising the route over the primary link without any communities. It is received by AS 100 and propagated to AS 200. When AS 200 receives the route via AS 100, there is no community set. AS 200 therefore assigns the default local preference value of 100.

The customer is also advertising the route over the backup link. However, in this case, the route has the community 200:50 set. When AS 200 receives this route, it recognizes the community, and the local preference value is set to 50.

Route selection is now performed in AS 200. The route that has been received via AS 100 is preferred based on the local preference values.

## BGP Incoming Link Selection Using BGP Communities (Cont.)

May Not work?



Even when the use of communities is correctly configured, the desired load distribution may not always be achieved. As this example shows, AS 200 does not always receive the primary route, although nothing is wrong in the network.

The customer AS 1 sends routes with community 200:50 over the backup link to AS 200. AS 200 receives the routes and sets the local preference to 50. If AS 200 over some time selects the directly connected path to AS 1 as the best, it propagates the route to AS 300. As the route is propagated over the [EBGP](#) session between AS 200 and AS 300, the local preference value that is used within AS 200 is lost.

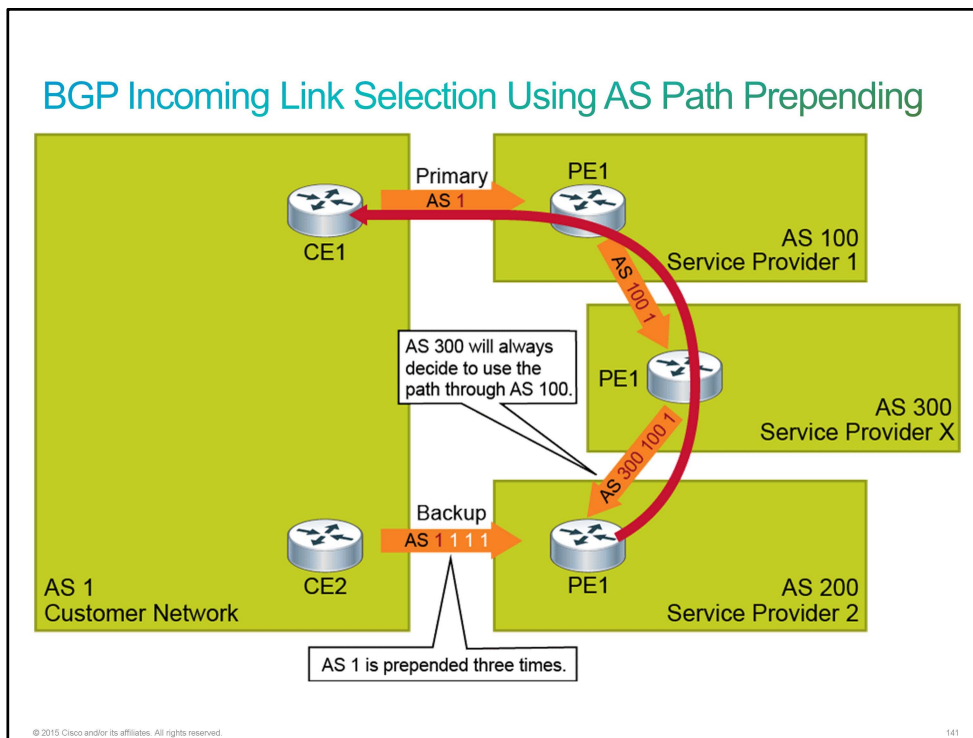
AS 300 does not have any use for the community 200:50 because this community is defined and implemented only within AS 200. Potentially, the community value can also be stripped off during BGP route propagation.

Customer AS 1 also sends the routes over the primary link to AS 100. The routes are propagated to AS 300, which now sees two alternative routes to the destination networks within AS 1. The router in AS 300 does not use weight nor local preference as criteria for reaching AS 1. Both alternatives have equal AS path lengths.

The route selection decision that will be made in AS 300 is hard to predict, but the outcome definitely influences the route selection decision that was made in AS 200. If AS 300 prefers the route to the customer network via AS 200 for any reason, then the second-best alternative via AS 100 and the primary link is never propagated to AS 200.

In this case, AS 200 never sees the primary path and has to stick to the backup link and announce the route to AS 300. 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



In this example, the customer AS 1 is performing AS path prepending on the backup link. Three copies of the customer AS number (1) 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 200 receives routes from AS 1 over the backup link with an AS path length of four. This is the original AS 1 plus three prepended copies that the CE router applied to the AS path attribute.

The customer advertises networks without AS path prepending over the primary link. AS 100 receives routes with an AS path length of one and propagates these routes to AS 300, which then receives them with an AS path length of two.

If, for a short time, AS 300 received the customer routes via AS 200, the AS path length of those routes would have been five. In that case, AS 300 selects the route from AS 100 as the best and propagates it to AS 200.

AS 200 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 300 have an AS path length of three. Because no weight or local preference is configured in this example, AS 200 selects the route via AS 300 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

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 the same way as when a multihomed customer connects to a single service provider.

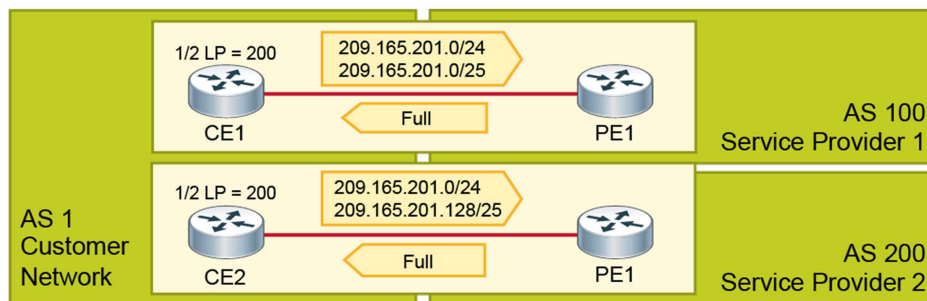
## Load Sharing with Multiple Providers

- **Load sharing for *outgoing* traffic:**
  - You can use the same solution as with multihomed customers that are connected to one service provider.
  - A higher local preference for half of the routes comes from one service provider. A higher local preference for the other half of the routes comes from the other service provider.
- **Load sharing for *incoming* traffic:**
  - The only load-sharing option that you can use in this setup is to separate the address space into two or more smaller address blocks.
  - Some traffic analysis is needed to fine-tune the 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

142

## Load Sharing with Multiple Providers (Cont.)



© 2015 Cisco and/or its affiliates. All rights reserved.

143

The customer can control the load distribution of incoming traffic based on the 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. 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. 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 each part of the customer address space. 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 that were discussed in this lesson.

## Summary

- A customer that is multihomed to multiple BGP service providers must advertise its address space to both ISPs, but has to make sure that it does not become a transit network.
- Customers who are connected to multiple ISPs must use an AS number that all ISPs agree to.
  - 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.
- Primary and backup connectivity:
  - Outgoing route selection is achieved by using local preference.
  - Incoming route selection is achieved by using either BGP communities or AS path prepending.
- The key for load-sharing is to divide the customer address space.

© 2015 Cisco and/or its affiliates. All rights reserved.

144



# Lesson 5: Module Summary

---

## Overview

This topic summarizes the key points that were discussed in this module.

### Module Summary

- Customer requirements influence the planning of connectivity between a customer and an ISP, including:
  - Physical connection type (the redundancy provided)
  - IP addressing
  - AS numbering
- Static or dynamic routing can be used between a customer and a service provider.
  - **Static routing:** Customers with a single or multiple connections to the service provider (link detection and device failure can be detected)
  - **Dynamic routing (BGP):** Multihomed customers.
- When using dual links and BGP, BGP attribute manipulation is required to achieve the required traffic distribution pattern:
  - Primary and backup link
  - Load sharing

© 2015 Cisco and/or its affiliates. All rights reserved.

145

## References

For additional information, refer to these resources:

- Cisco Systems, Inc. *Sample Configuration for BGP with Two Different Service Providers (Multihoming)*.  
[http://www.cisco.com/en/US/tech/tk365/technologies\\_configuration\\_example09186a008009456d.shtml](http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a008009456d.shtml)

- Cisco Systems, Inc. *Removing Private Autonomous System Numbers in BGP*.  
<http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13756-32.html>
- Cisco Systems, Inc. *How BGP Routers Use the Multi-Exit Discriminator for Best Path Selection*.  
<http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13759-37.html>
- Cisco Systems, Inc. *Configuring the BGP Local-AS Feature*.  
<http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13761-39.html>
- Cisco Systems, Inc. *Load Sharing with BGP in Single and Multihomed Environments: Sample Configurations*.  
[http://www.cisco.com/en/US/tech/tk365/technologies\\_configuration\\_example09186a00800945bf.shtml](http://www.cisco.com/en/US/tech/tk365/technologies_configuration_example09186a00800945bf.shtml)

# Lesson 6: 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.

1. If a customer requires extra 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.
  
2. Which two routing options could be used when connecting a customer to a service provider? (Choose two.) (Source: "Understanding Customer-to-Provider Connectivity Requirements")
  - A. static routing
  - B. OSPF
  - C. IS-IS
  - D. BGP
  
3. A customer would like to connect to a service provider. Which two requirements should be considered before deciding on a type of connectivity? (Choose two.) (Source: "Understanding Customer-to-Provider Connectivity Requirements")
  - A. application availability
  - B. redundancy
  - C. cost ownership
  - D. flexibility

4. 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
5. 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
6. 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.
7. 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.
8. 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
9. 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.



10. Floating static routes do not demand any additional configuration to work with BGP. True or false? (Source: "Implementing Customer Connectivity Using Static Routing")
- A. True
  - B. False
11. When static routing is implemented between the customer network and the ISP network, the customer must announce a default route. True or false? (Source: "Implementing Customer Connectivity Using Static Routing")
- A. True
  - B. False
12. 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.
13. 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
14. When you are designing a 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.
15. 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

16. Which three key pieces of information can you derive from the following router command output? (Choose three.) (Source: "Implementing Customer Connectivity Using Static Routing")

**CE2# show ip bgp 209.165.201.0**

BGP routing table entry for 209.165.201.0/24, 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: 1:31000 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: 1:31000 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.
17. 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.
18. What are the 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.

19. What are the three responsibilities of the customer when the customer has multiple connections to a single service provider? (Choose three.) (Source: "Connecting a 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.
20. Given the following router command output, which method has been used to influence return traffic in a primary/backup link implementation for the customer with multiple connections to a service provider? (Source: "Connecting a Customer to a Single Service Provider")

ISP# **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
21. What are three responsibilities of the provider router when supporting a customer with multiple connections? (Choose three.) (Source: "Connecting a 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.
22. What will occur if private AS numbers are advertised to the Internet? (Source: "Connecting a 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.

23. Which two BGP configurations are required to properly implement a backup solution for a customer that is connected to a single provider via multiple connections? (Choose two.) (Source: "Connecting a 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.

24. 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 Customer to a Single Service Provider")

CE1# **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.

25. Which three methods can you use to provide load sharing over network links between a customer and a single provider? (Choose three.) (Source: "Connecting a 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.

26. 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 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.

27. 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 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.
28. A multihomed customer is using AS number 65550 internally. The customer is connected to two different providers. ISP1 (in AS 200) has assigned the customer an AS number of AS 65101. ISP2 (in AS 300) 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 ISP2? (Source: "Connecting a Multihomed Customer to Multiple Service Providers")
- A. 65550 i
  - B. 65201 i
  - C. 65201 65550 i
  - D. 300 65201 i
29. 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.
30. 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.

31. A multihomed customer is using AS number 1024 and is connected to two different providers (ISP1: AS 200 and ISP2: AS 300). The customer has configured the MED to ensure a proper return path so that ISP1 is the primary provider and ISP2 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 ISP2.
32. 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.
33. 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.
34. 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")

ISP# **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

35. You are a customer connected to two service providers with one link to each provider. You would like to achieve primary and backup traffic distribution. Which two BGP attributes do you have to manipulate? (Choose two.) (Source: "Connecting a Multihomed Customer to Multiple Service Providers")
- A. local preference for incoming traffic
  - B. local preference for outgoing traffic
  - C. AS path for incoming traffic
  - D. MED for incoming traffic
36. A multihomed customer can get only one private AS number, regardless the number of service providers. True or false? (Source: "Connecting a Multihomed Customer to Multiple Service Providers")
- A. True
  - B. False





# Module Self-Check Answers

---

## Answer Key

1. D
2. A, D
3. B, D
4. A
5. D
6. B
7. B
8. A, C
9. A, C
10. B
11. A
12. A, B
13. C
14. A, C, D
15. B
16. A, C, D
17. B, D
18. A, B, C
19. A, C, D
20. A
21. A, B, C
22. B
23. A, C

- 24. B
- 25. A, B, C
- 26. B
- 27. A, C
- 28. C
- 29. A, C, D
- 30. A, C, D
- 31. D
- 32. A, C, D
- 33. A
- 34. C, D
- 35. B, C
- 36. B

# Module 6: Scaling Service Provider Networks

---

## Introduction

In standard [BGP](#) implementations, all BGP routers within an [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 as scalability mechanism that allows you 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.

Upon completing this module, you will be able to:

- Describe common routing scalability issues in service provider networks
- Describe the function of route reflectors in a BGP environment
- Configure proper operation of route reflectors to modify IBGP split-horizon rules in an existing IBGP network



# Lesson 1: Scaling IGP and BGP in Service Provider Networks

---

## Overview

Properly scaling [IP addressing](#), [IGPs](#), and [BGP](#) is a common area of concern to all service providers and can be the difference between a successful and a problematic BGP implementation. 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 you support 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 [ISP](#) network and discussion of the propagation of internal and customer routing information. Also scaling considerations for IGPs and BGP, and scaling of IP addressing in service provider networks are described.

Upon completing this lesson, you will be able to:

- 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

# Common Service Provider Network

The common service provider network runs [EBGP](#) or static routing with customers. EBGP is always used as the routing protocol between different service providers.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved. 146

[IBGP](#) is required in the provider network because all EBGP-speaking routers in an [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 [OSPF](#) and [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.)

The diagram illustrates a common service provider network architecture. It features four Autonomous Systems (ASes): AS 100, AS 200, AS 300, and AS 400. AS 100 and AS 200 are connected to AS 300 via Peer Service Providers. AS 300 is connected to AS 400 via a Customer. The network includes several routers (A, B, C, D, E, F, G, H) and various interconnections labeled with protocols like EBGP, IBGP, and IGP. A central cloud represents AS 300, and a bottom cloud represents AS 400. A Customer is shown connected to AS 400. The diagram also includes labels for 'Peer Service Providers', 'IGP Within Service Provider Network', and 'IBGP Between BGP Speakers of Service Provider'.

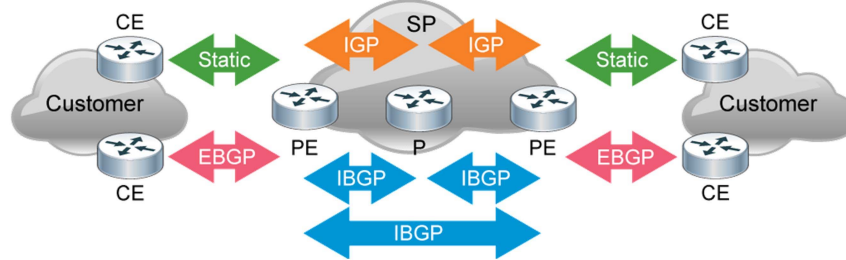
© 2015 Cisco and/or its affiliates. All rights reserved.

147

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.

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.)



- PE routers use EBGP or static routing with CE routers.
- PE and P routers use full-mesh IBGP routing.
- The provider core IGP is a single instance of IS-IS or OSPF and is used only within the service provider core network.
- Optimal routing between PEs is desired.

© 2015 Cisco and/or its affiliates. All rights reserved.

148

CE routers connect via IP to PE routers. In many cases, the PE routers use static routing to customer networks. The PE routers advertise static routes to the rest of the service provider network and to other autonomous systems using BGP. PE and P routers use full-mesh IBGP routing to exchange BGP routing information.

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

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.

## 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 resolution of BGP next hop and internal routes
- Do not redistribute BGP into IGP.
  - IGP performance and convergence time suffer if many are carried.
  - No IGP is capable of carrying full Internet routes.
  - A full Internet routing table has exceeded 600,000 routes.

© 2015 Cisco and/or its affiliates. All rights reserved.

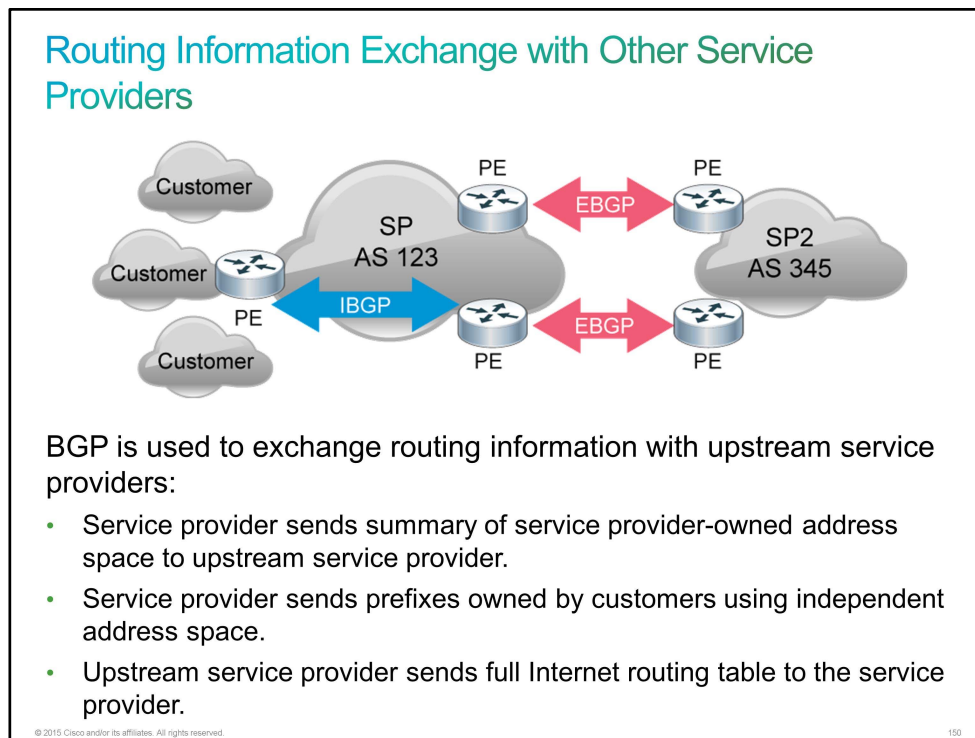
149

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 the IGP carries. 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 600,000 routes. You should therefore never redistribute the routing information that the BGP received into the IGP, because no IGP is capable of carrying several tens of thousands of routes.

# Routing Information Exchange with Other Service Providers

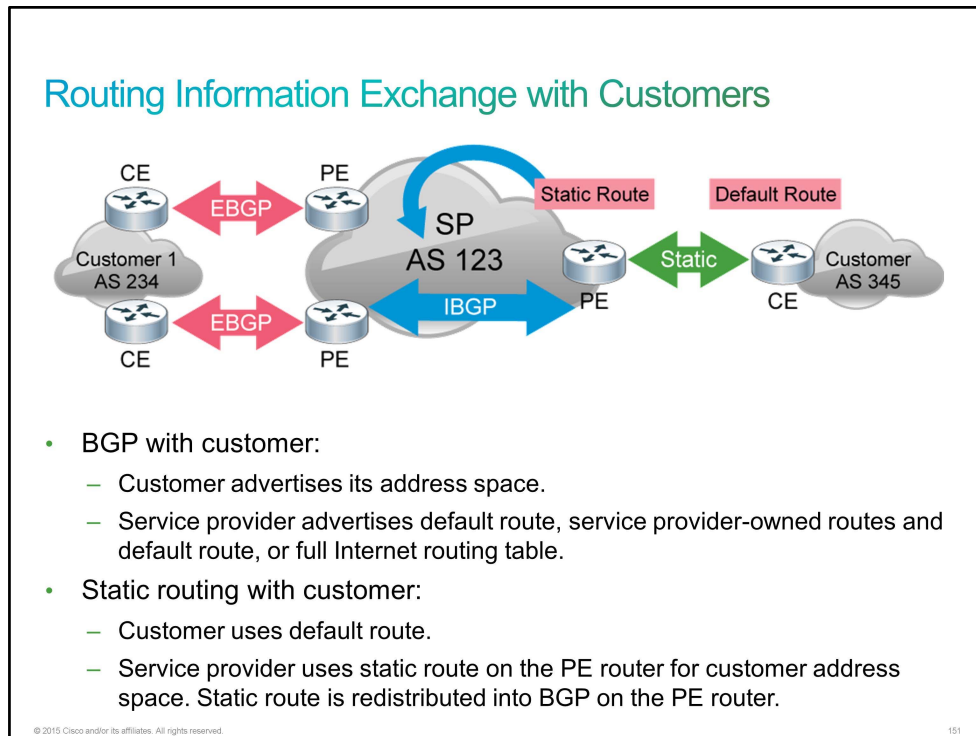


The figure shows how routes are exchanged between service providers.

Provider edge routers use BGP to exchange routing information with other service provider networks for redundancy and scalability reasons.

Routing information from customers is received at PE routers using EBGP and then propagated using IBGP to the rest of the service P-network. At another PE router, the routing information is further propagated to a different service provider using EBGP with other autonomous systems. The service provider sends a summary of service provider-owned space to upstream service providers, along with prefixes owned by the customer using provider-independent address space. The upstream service provider sends the full Internet routing table to the service provider, which allows the service provider to implement a primary/backup scenario or a load-balancing scenario when peering with multiple upstream service providers.

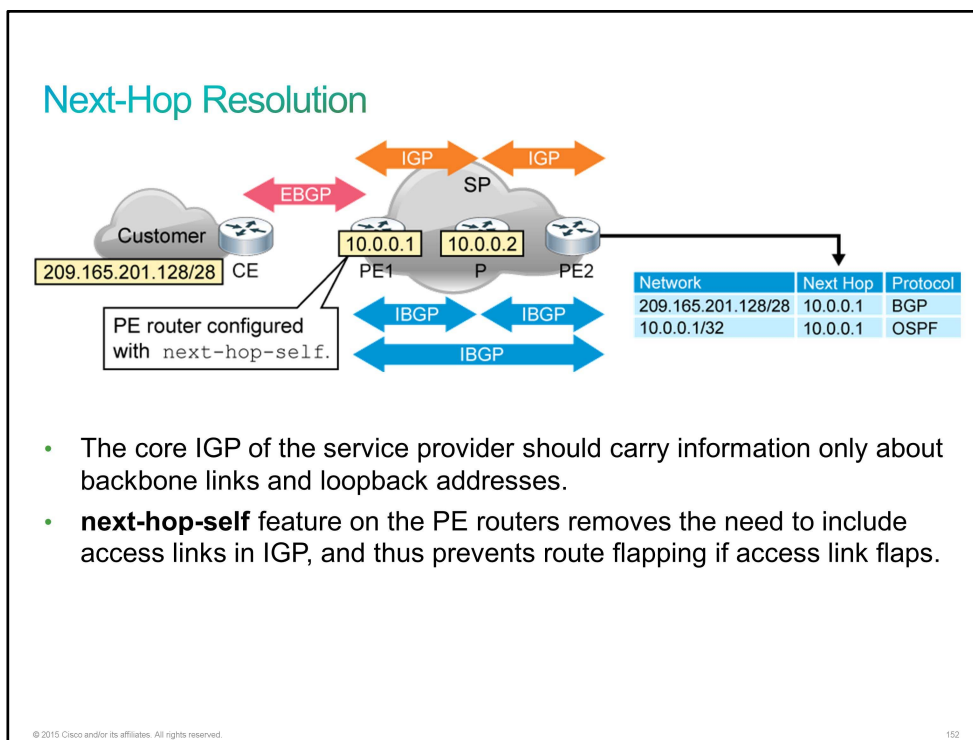
# Routing Information Exchange with Customers



The figure shows how customers exchange routes with the service provider. In this case, the customer advertises its address space to the service provider. The service provider advertises to the customer either the default route, service provider-owned routes and the default route, or the full Internet routing table, based on the customer requirements.

When the customer does not require redundancy, static routing is used. In this case, the customer configures a static default route that points to the PE router of the service provider. The service provider configures a static route for the customer network that points to the [CE](#) router. The PE router redistributes static routes into BGP. The service P-network then uses BGP to propagate the information to the rest of the service P-network using IBGP. The service provider also advertises customer routing information to other autonomous systems using BGP.

# Next-Hop Resolution



The figure shows why IGP is needed in the service provider core network. 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.

As shown in the routing table on the PE2 router, the next hop for the customer network points to the address of the PE1 router. This is usually the [IP address](#) of the loopback interface. Because the PE1 and PE2 routers are not directly connected, an IGP routing protocol is needed that carries information about core links and loopbacks across the service provider network.

# Scaling Service Provider Routing Protocols

[BGP](#) and [IGP](#) have different responsibilities 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

153

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

In scaling an IGP, it is important to limit the number of routes that the IGP carries. Optimally, the IGP carries only loopback interfaces and internal links.

## 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).

© 2015 Cisco and/or its affiliates. All rights reserved.

154

The number of routes that the IGP carried 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 uses optimal routing to the next hop. Also, in an [MPLS](#) environment, an [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

Scaling BGP includes scaling both IBGP and EBGP as well as scaling BGP updates and routing table size.

## Scaling BGP Routing

- 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.
  - Full-mesh IBGP is not needed since there are other technologies and features available
- Updates and table size scaling:
  - Route summarization is the key to scalability.

© 2015 Cisco and/or its affiliates. All rights reserved.

155

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

Internal [IP addressing](#) in the service provider core network can be simplified to reduce the use of public addresses and to simplify configuration.

## Scaling Service Provider Addressing

This list shows how to scale addressing in Service Provider Core Network:

- Private or public IP addresses can be used.
- Private addresses on core links and loopbacks display private IP addresses in a traceroute when run from customers.
  - MPLS with TTL propagation disabled solves the traceroute issue.
- Private addresses on loopbacks and core links call for careful external routing to prevent advertisement of private addresses to customers or upstream service providers.
- Otherwise, use public addresses in service provider core networks.

© 2015 Cisco and/or its affiliates. All rights reserved.

156

For [IPv4](#), public or private IP addresses can be used on core links. However, using private addresses has some drawbacks. Private addresses on the provider internal links will cause trouble for the traceroute application. When the **traceroute** command is executed from the customer, the customer will see private IP addresses in the traceroute printout. Using [MPLS](#) without [TTL](#) propagation in the service [P-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. It can be used to prevent accepting private addresses from any other AS if the other AS, by mistake, announces private addresses.



# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- Service providers most commonly use integrated IS-IS and OSPF as interior gateway protocols and BGP as the exterior gateway protocol.
- BGP is used to carry customer routes while IGP's are used to carry service provider internal prefix reachability information.
- BGP allows ISP clients to acquire information about all or some networks reachable through the ISP.
- ISP can use static routing or BGP to direct traffic going to the customers to the correct links.

© 2015 Cisco and/or its affiliates. All rights reserved.

157

## Summary (Cont.)

- Next-hop-self can be used to avoid redistributing transit segments into IGP on iBGP neighbors.
- When BGP networks grow, various actions must be taken to make them scalable, for iBGP scalability use route reflectors.
- When IP networks grow, several aspects of addressing need to be considered to reduce sizes of routing tables and to avoid consuming too many addresses.

© 2015 Cisco and/or its affiliates. All rights reserved.

158



# Lesson 2: Introducing and Designing Route Reflectors

---

## Overview

Large [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 also describes the various redundancy mechanisms that are used with route reflectors, including route reflector clusters. The lesson introduces network design rules that you 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.

Upon completing this lesson, you will be able to:

- 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
- 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

# IBGP Scalability Issues in a Transit AS

Classic [IBGP](#) split-horizon rules specify that updates that are received on an [EBGP](#) session should be forwarded on all IBGP and EBGP sessions. However, 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 [AS](#) directly through a separate IBGP session to each of them.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved.159

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. This information is then 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.

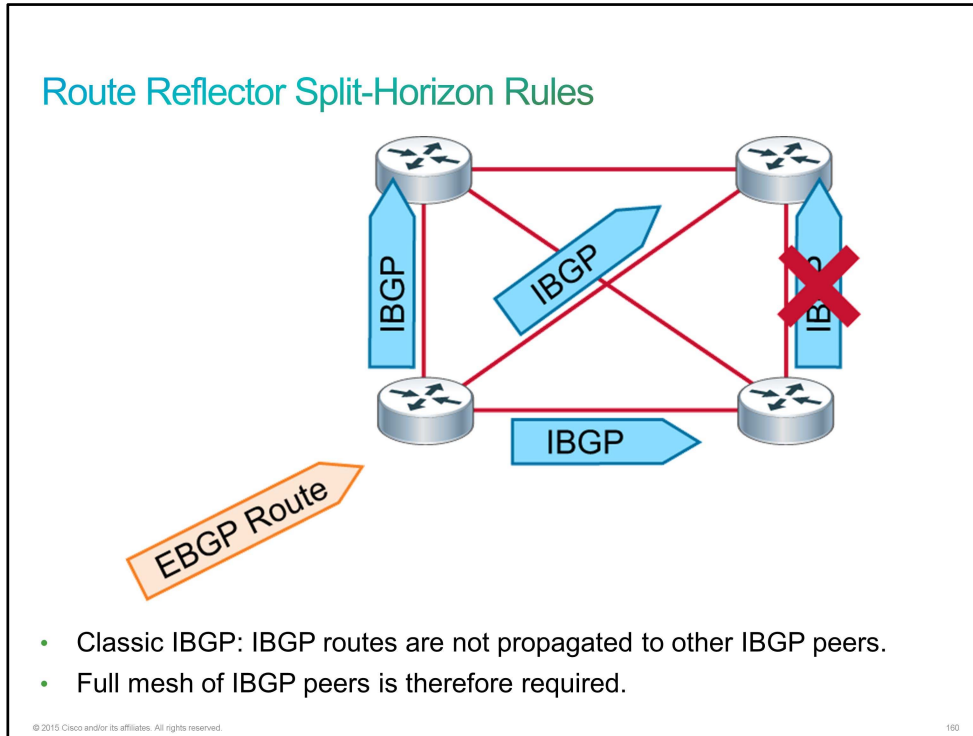
Route reflectors are BGP scalability mechanisms that enable routing information to be redistributed to all routers within an AS while eliminating the need for a fully meshed topology within the AS. This feature reduces the number of TCP sessions that must be maintained, lowering network overhead and [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 (not covered in this course) 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.

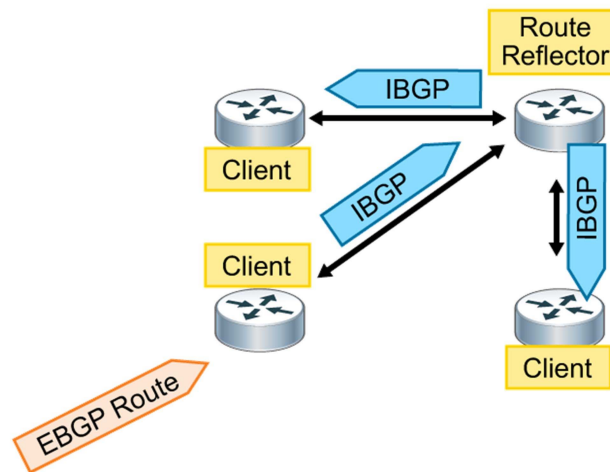
# Route Reflector Split-Horizon Rules

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](#). It uses a dedicated IBGP session for each router. Also, the BGP boundary router forwards routes that a router sources in the same way. To allow every router to update every other router, a full mesh of IBGP sessions is required.



The figure shows how the router receives routing update on EBGP session and forwards this update to all IBGP neighbors. When IBGP neighbors receive this update on IBGP sessions, these updates are not forwarded to any other IBGP sessions by default.

## Route Reflector Split-Horizon Rules (Cont.)



- Route reflector can propagate IBGP routes to other IBGP peers.
- Full mesh of IBGP peers is no longer required.

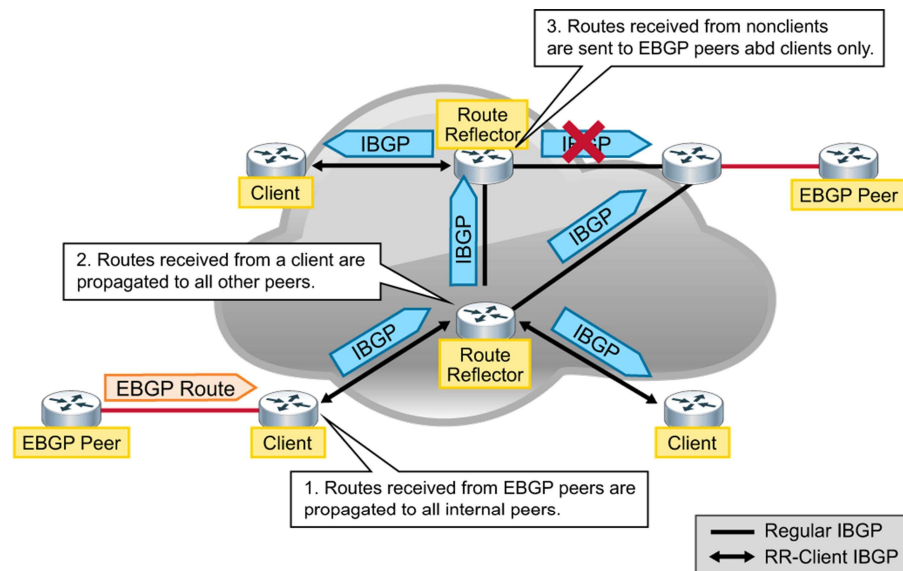
© 2015 Cisco and/or its affiliates. All rights reserved.

161

The IBGP route reflector design relaxes the need for a full mesh. The router that is 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.)



© 2015 Cisco and/or its affiliates. All rights reserved.

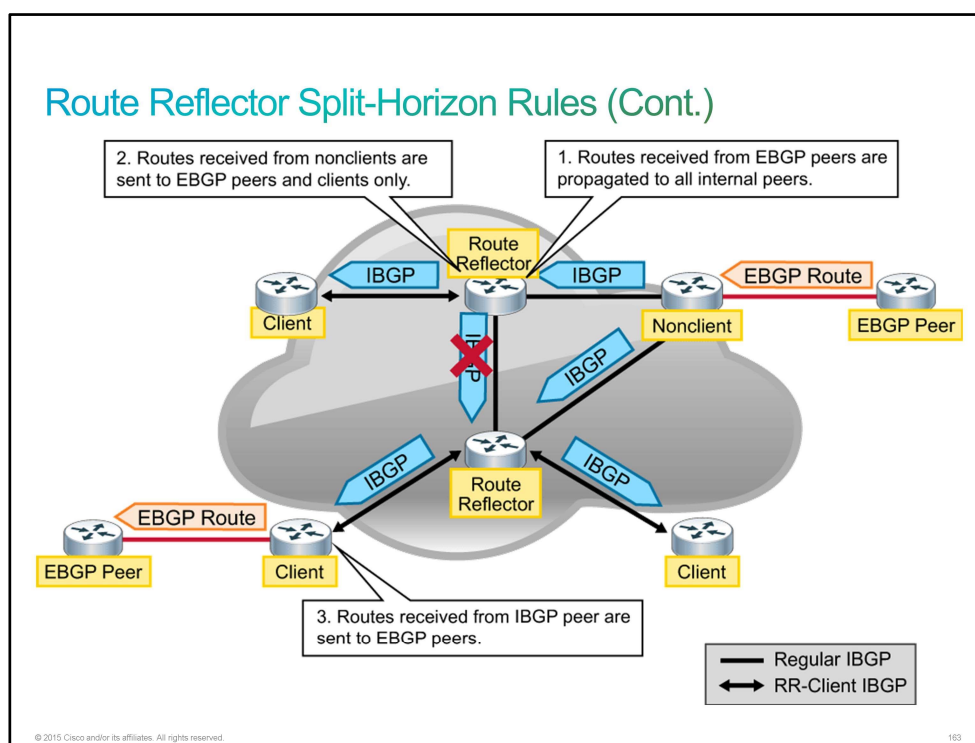
162

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 with the recursive routing, regardless of the way that it has received the BGP route.



Another example of route propagation in a route reflector-enabled network:

- Nonclients conform to the classic IBGP split-horizon rules and forward a received route from EBGP on their IBGP neighbor sessions.
- The upper route reflector that receives a route from a nonclient sends the route to EBGP peers and clients only. It is the reason why the IBGP update is not sent to the bottom route reflector as well.
- The client receives the IBGP update and sends it to the EBGP peers.

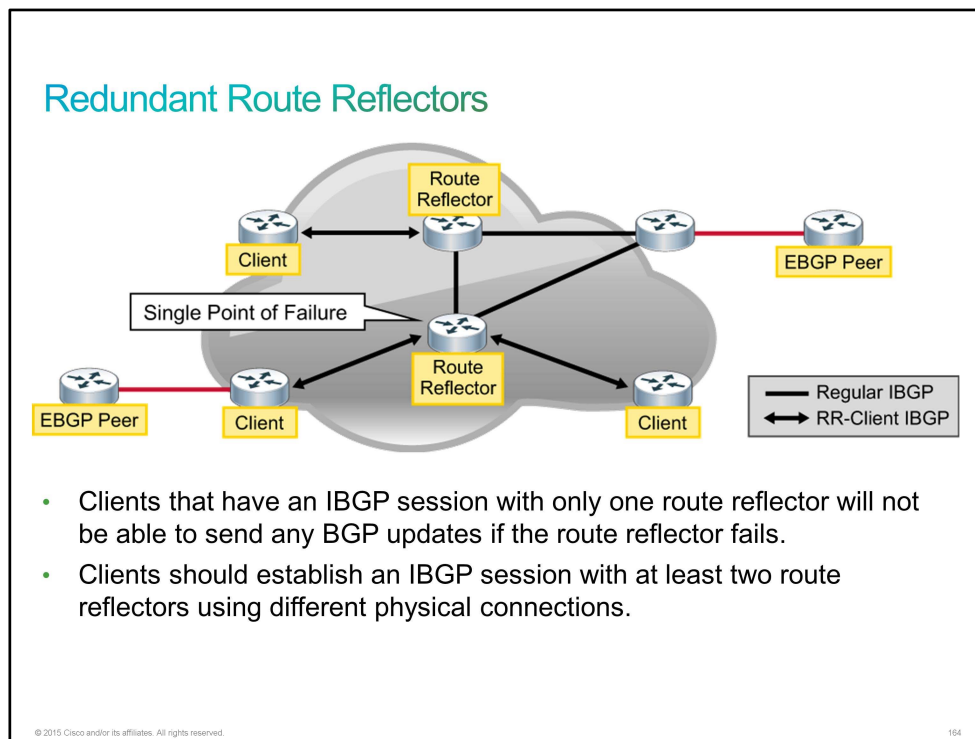
The table presents detailed IBGP split-horizon rules as modified by the introduction of BGP route reflectors. For purposes of definition, a "route reflector" is a BGP speaker that can advertise IBGP learned routes to another IBGP peer and, hence, can reflect routes. IBGP peers of the route reflector 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                |

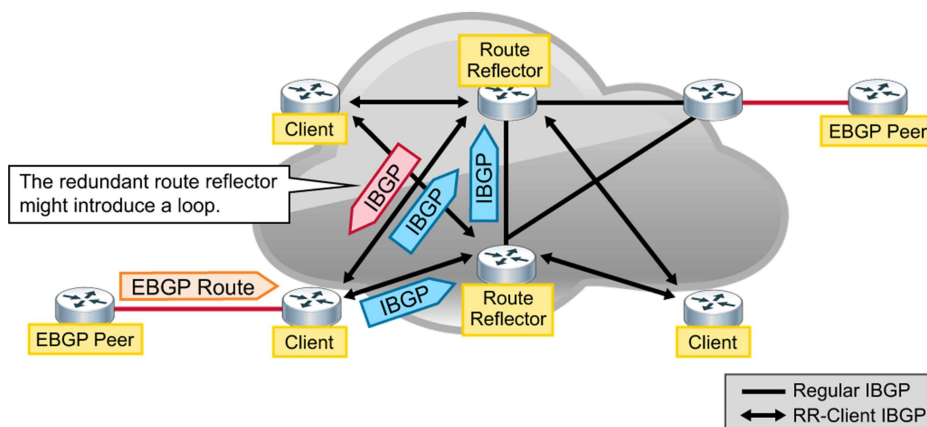
# Redundant Route Reflectors

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.)

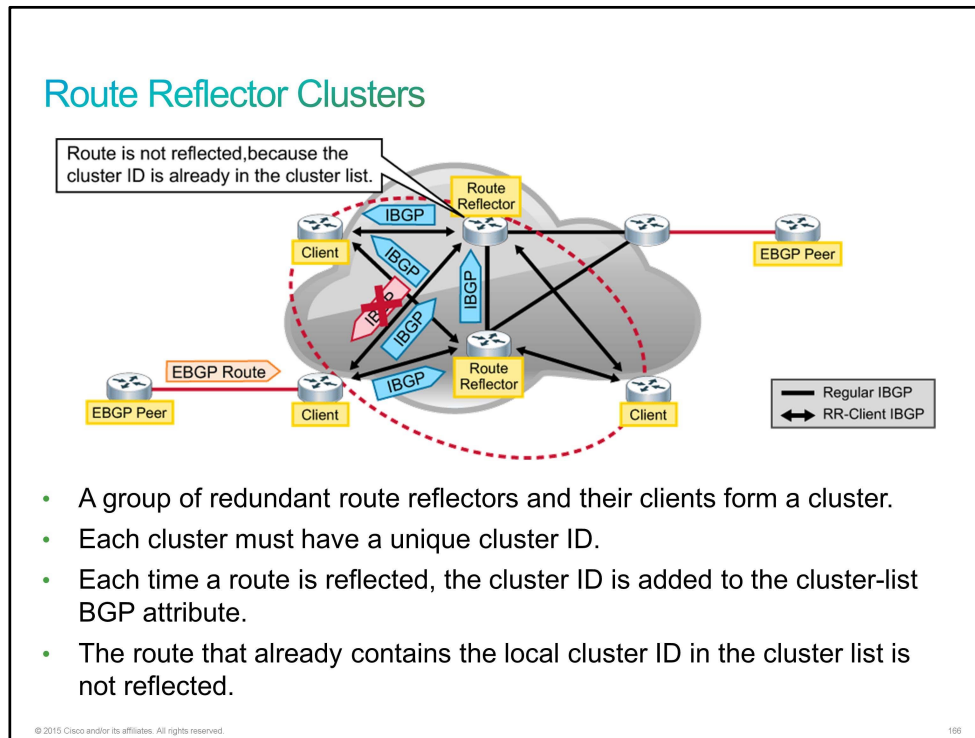


- Redundant reflectors solve the high-availability requirement.
- The concept of clusters is introduced to prevent IBGP routing loops between route reflectors.

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. And this routing loop is impossible to detect. Extra BGP attributes are thus necessary to prevent these routing loops.

# Route Reflector Clusters

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 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 a [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.

In the example, 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

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.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

167

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 [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.

# Network Design with Route Reflectors

The physical topology of the network could serve as a guide to route reflector design.

## 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.

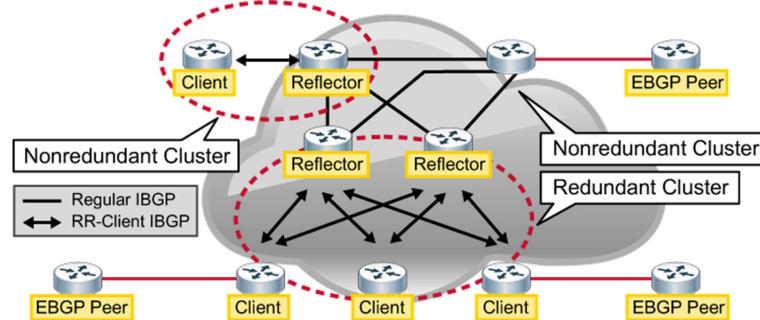
© 2015 Cisco and/or its affiliates. All rights reserved.

168

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 to form a cluster.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

169

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, 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.

In the example, there are two clusters and one standalone BGP routers. In the area that is labeled “redundant cluster,” the three client routers and the two route reflector routers make up the cluster. Each of the three client routers has an IBGP session with the two route reflectors and only with those two route reflectors.

In the nonredundant area, the client router has a single physical connection to a route reflector router. These two routers form a nonredundant cluster. The router that is designated as the route reflector in the cluster is already a single point of failure in this physical design. 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. The client has a single IBGP session to the route reflector.

The standalone router that is 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. This router has a full mesh of IBGP sessions with all route reflectors, each reflector representing its cluster.



# Potential Network Issues

## 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. |

© 2015 Cisco and/or its affiliates. All rights reserved.

170

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 client will forward the BGP update from 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

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.

## 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.

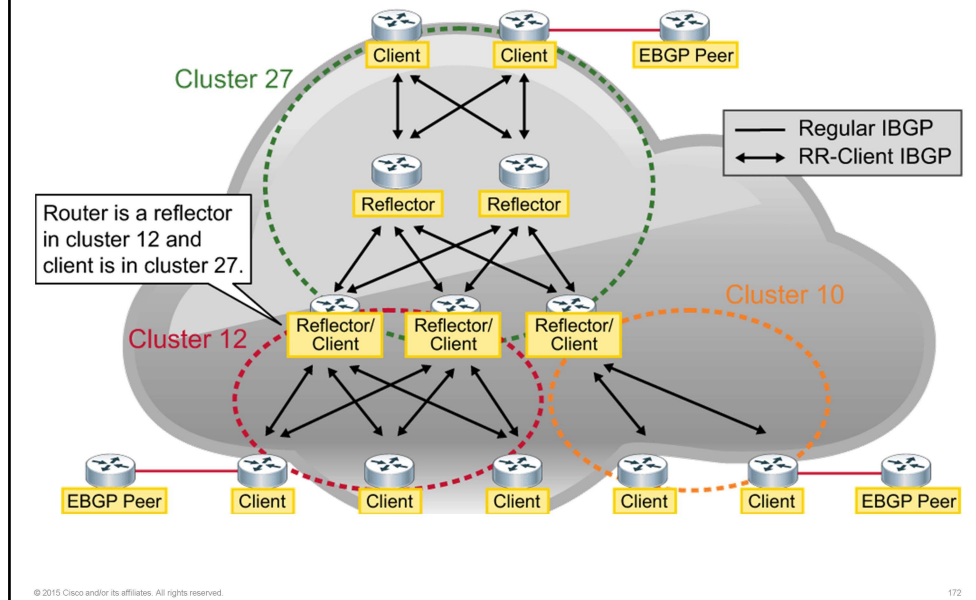
© 2015 Cisco and/or its affiliates. All rights reserved.171

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, so 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 extra level of route reflectors.

## Hierarchical Route Reflectors (Cont.)



The figure shows an example of hierarchical route reflectors.

The first level of hierarchy reduced the original full mesh of 12 routers (all routers in the service provider network) to a full mesh of seven routers (the lower three route reflectors and the upper two route reflectors and two clients). The second level of route reflector clusters was built by creating cluster 27. This second step further reduced the full mesh of seven routers to a full mesh consisting of only two routers (upper route reflectors). Only the two route reflectors in cluster 27 should be connected in a full mesh.

When a client in the lowest level receives an [EBGP](#) update, it will forward the update 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 that were 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. All routes that are received from a route reflector client are sent to all other IBGP neighbors. 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.
- Route reflector clusters were introduced in the BGP route reflector architecture to support redundancy, preventing IBGP routing loops in redundant route reflector designs.

© 2015 Cisco and/or its affiliates. All rights reserved.

173

## Summary (Cont.)

- The originator-ID and cluster-list BGP attributes were introduced to prevent routing loops in route reflector environments.
- 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.
- 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

174

# Lesson 3: Configuring and Monitoring Route Reflectors

---

## Overview

Large [BGP](#) networks cannot properly scale without relying on performance-enhancing tools such as route reflectors. 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 [AS](#) to BGP route reflectors. It also lists the Cisco IOS commands that are required to configure and monitor route reflectors.

Upon completing this lesson, you will be able to:

- 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

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.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

175

The following planning and preparation steps are required before you migrate from a full mesh of [IBGP](#) sessions to a route reflector design:

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.
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 are 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

## Configuring Route Reflectors

Configuration changes that are required to configure BGP 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 that IBGP neighbor is removed on both ends of the IBGP session.

© 2015 Cisco and/or its affiliates. All rights reserved.

176

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.)

Cisco IOS commands that are required to configure BGP route reflectors:

```
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

```
router(config-router)# neighbor ip-address route-reflector-client
```

- On route reflector configures an IBGP neighbor to be a client of this reflector.

© 2015 Cisco and/or its affiliates. All rights reserved.

177

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                                                                                      |
|-------------------|--------------------------------------------------------------------------------------------------|
| <i>cluster-id</i> | Cluster-ID of the router acting as a route reflector.<br>The cluster-ID is a maximum of 4 bytes. |

This command is 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

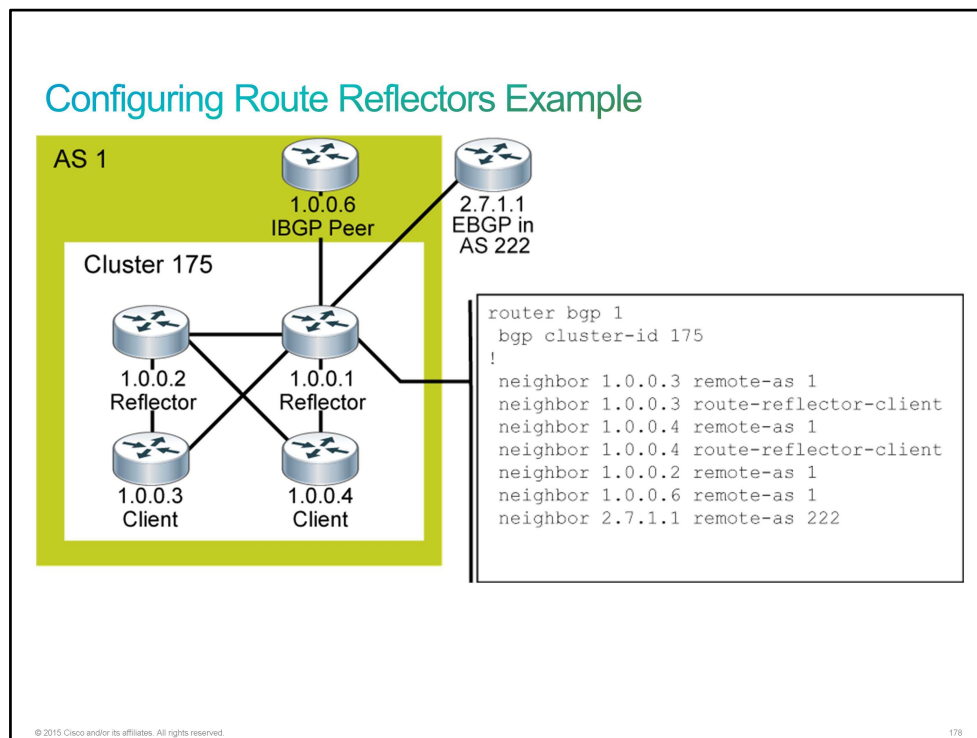
| Parameter         | Description         |
|-------------------|---------------------|
| <i>ip-address</i> | Neighbor IP address |

By default, there is no route reflector in the [AS](#).



# Configuring Route Reflectors Example

In this example, the router 1.0.0.1 has been configured as a route reflector.



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.

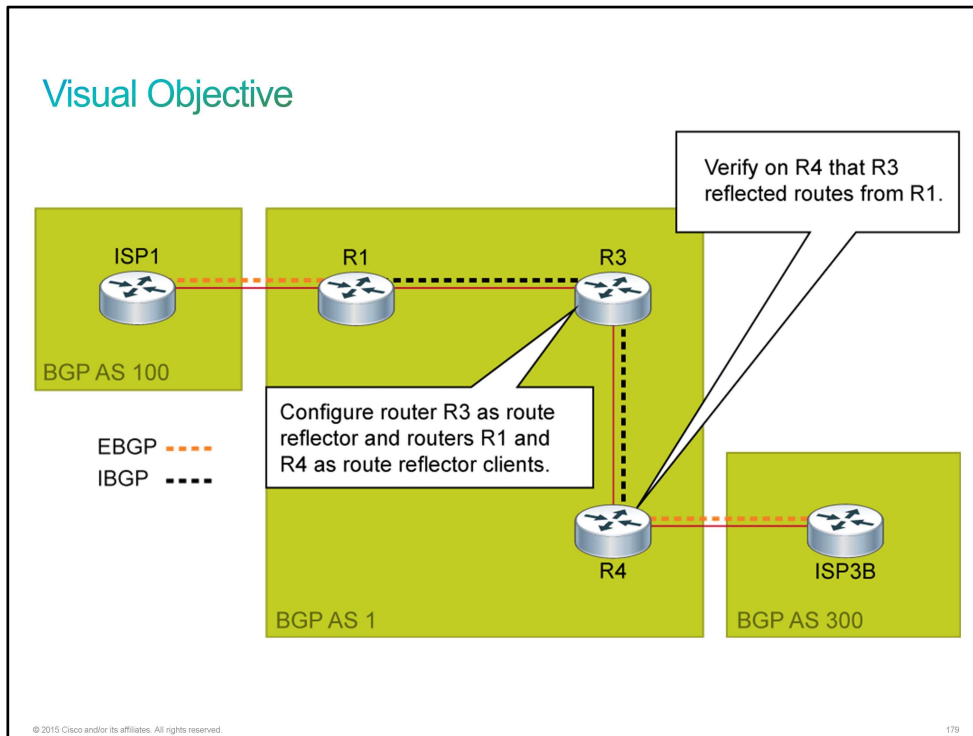
The figure shows a portion of the configuration in router 1.0.0.1. The cluster-ID is assigned to the router under the **router bgp** process definition of the router configuration. After the router has been assigned, the route reflector client configuration is added under the **router bgp** process for the two neighbors that identify the two sessions reaching clients.

Both clients need to have established IBGP session to both route reflectors. IBGP session between route reflectors is also needed. Route reflector 1.0.0.1 also has established IBGP session to nonclient router 1.0.0.6 and [EBGP](#) session to router 2.7.1.1 on autonomous system 222.

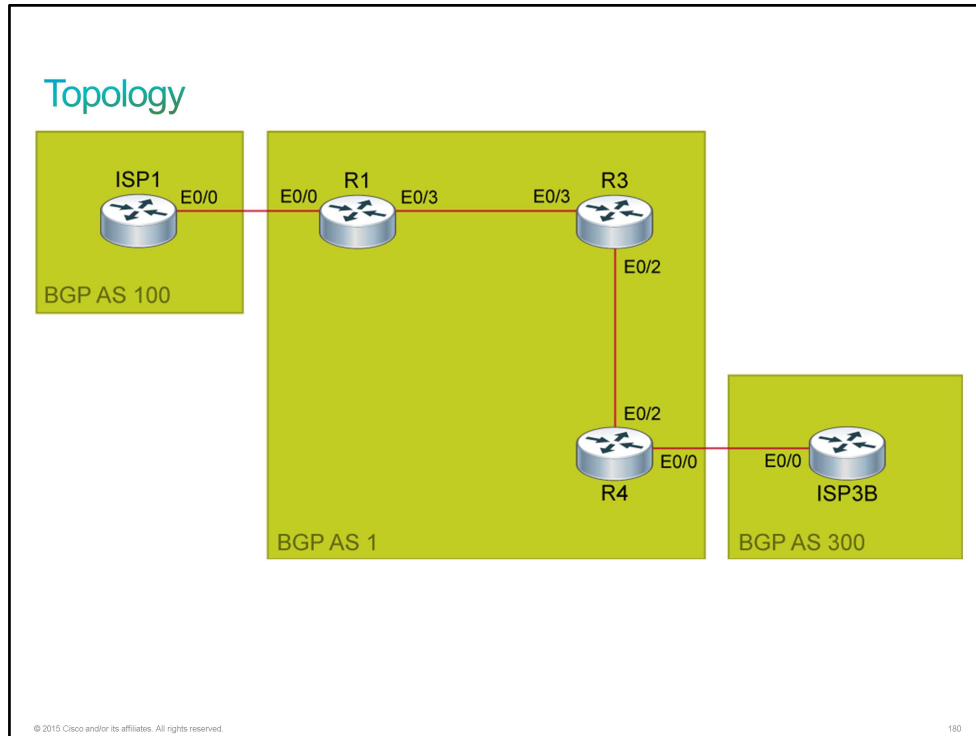
# Discovery 17: Configure Route Reflector

## Overview

Through this discovery, you will learn how to eliminate the need for full mesh [IBGP](#) sessions by introducing the route reflector feature into the autonomous system.



# Topology



## Job Aids

---

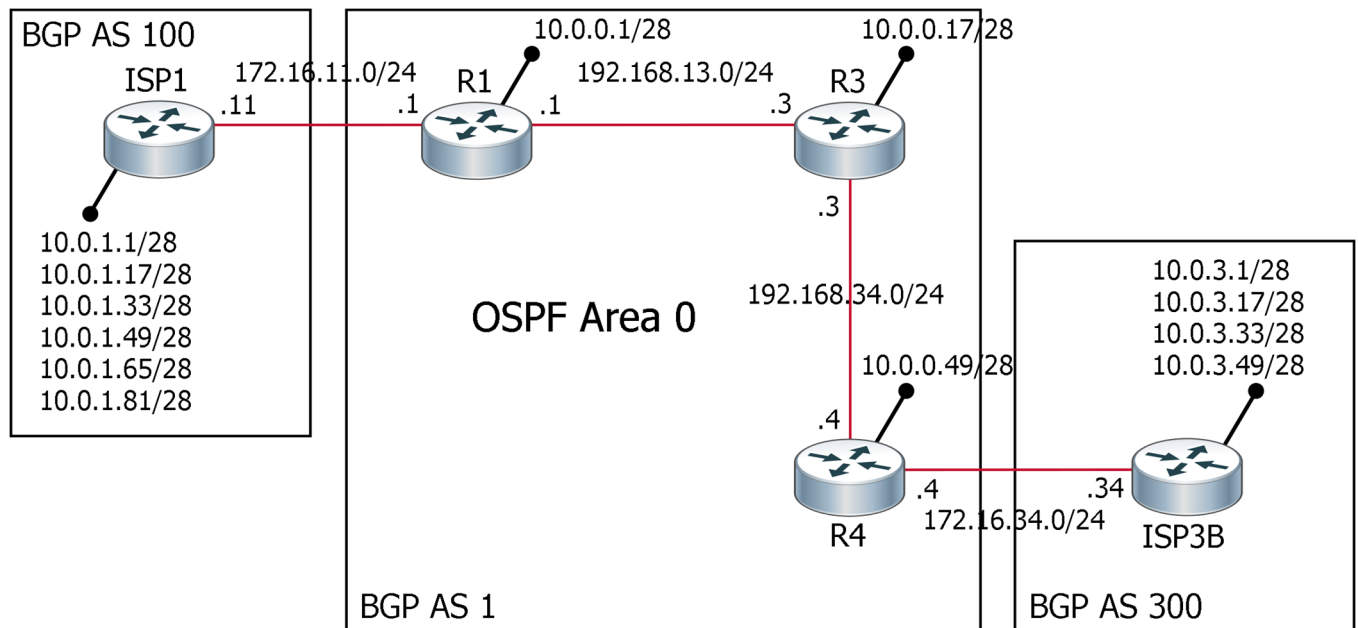
**Note** If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, IOL behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

---

## Device Information

| Device | Interface                                                                        | IP address                                                                                  | Description                     |
|--------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|---------------------------------|
| ISP1   | Ethernet 0/0                                                                     | 172.16.11.11/24                                                                             | Connection to R1                |
| ISP1   | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4<br>Loopback 5<br>Loopback 6 | 10.0.1.1/28<br>10.0.1.17/28<br>10.0.1.33/28<br>10.0.1.49/28<br>10.0.1.65/28<br>10.0.1.81/28 | Loopbacks simulate LAN networks |
| R1     | Ethernet 0/0                                                                     | 172.16.11.1/24                                                                              | Connection to ISP1              |
| R1     | Ethernet 0/3                                                                     | 192.168.13.1/24                                                                             | Connection to R3                |
| R1     | Loopback 1                                                                       | 10.0.0.1/28                                                                                 | Loopback simulates LAN network  |
| R3     | Ethernet 0/2                                                                     | 192.168.34.3/24                                                                             | Connection to R4                |
| R3     | Ethernet 0/3                                                                     | 192.168.13.3/24                                                                             | Connection to R1                |
| R3     | Loopback 1                                                                       | 10.0.0.17/28                                                                                | Loopback simulates LAN network  |
| R4     | Ethernet 0/0                                                                     | 172.16.34.4/24                                                                              | Connection to ISP3B             |
| R4     | Ethernet 0/3                                                                     | 192.168.24.4/24                                                                             | Connection to R2                |
| R4     | Loopback 1                                                                       | 10.0.0.49/28                                                                                | Loopback simulates LAN network  |
| ISP3B  | Ethernet 0/0                                                                     | 172.16.34.34/24                                                                             | Connection to R4                |
| ISP3B  | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4                             | 10.0.3.1/28<br>10.0.3.17/28<br>10.0.3.33/28<br>10.0.3.49/28                                 | Loopbacks simulate LAN networks |

[IP addresses](#) and [IGP](#) are preconfigured as shown in the topology below:



These [BGP](#) sessions have been preconfigured:

- [EBGP](#) session between ISP1 and R1 routers
- EBGP session between ISP3B and R4 routers
- IBGP session between R1 and R3 routers
- IBGP session between R3 and R4 routers

Note that there is no IBGP session between R1 and R4 routers. [OSPF](#) protocol has also been preconfigured as IGP routing protocol, making sure that all internal links, loopbacks, and next hops are accessible in autonomous system AS 1.

## Configuring Route Reflector

### *Discovery Steps*

**Step 1** On the R1 router, verify which BGP sessions are established.

```

R1# show bgp summary
BGP router identifier 10.0.0.1, local AS number 1
BGP table version is 101, main routing table version 101
8 network entries using 1184 bytes of memory
8 path entries using 512 bytes of memory
4/3 BGP path/bestpath attribute entries using 544 bytes of memory
1 BGP rrinfo entries using 24 bytes of memory
2 BGP AS-PATH entries using 48 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
BGP using 2312 total bytes of memory
BGP activity 55/42 prefixes, 86/78 paths, scan interval 60 secs

```

| Neighbor     | V | AS  | MsgRcvd | MsgSent | TblVer | InQ | OutQ | Up/Down  | State/PfxRcd |
|--------------|---|-----|---------|---------|--------|-----|------|----------|--------------|
| 10.0.0.17    | 4 | 1   | 8       | 6       | 101    | 0   | 0    | 00:00:22 | 1            |
| 172.16.11.11 | 4 | 100 | 109     | 118     | 101    | 0   | 0    | 01:33:46 | 6            |

You should only see the BGP sessions established:

- EBGP session to the ISP1 router (neighbor 172.16.11.11)
- IBGP session to the R3 router (neighbor 10.0.0.17)

**Step 2** On the R3 router, verify which BGP sessions are established.

```

R3# show bgp summary
BGP router identifier 10.0.0.17, local AS number 1
BGP table version is 198, main routing table version 198
13 network entries using 1924 bytes of memory
13 path entries using 832 bytes of memory
4/4 BGP path/bestpath attribute entries using 544 bytes of memory
2 BGP AS-PATH entries using 48 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
BGP using 3348 total bytes of memory
BGP activity 37/24 prefixes, 109/96 paths, scan interval 60 secs

```

| Neighbor  | V | AS | MsgRcvd | MsgSent | TblVer | InQ | OutQ | Up/Down  | State/PfxRcd |
|-----------|---|----|---------|---------|--------|-----|------|----------|--------------|
| 10.0.0.1  | 4 | 1  | 7       | 8       | 198    | 0   | 0    | 00:00:51 | 7            |
| 10.0.0.49 | 4 | 1  | 6       | 7       | 198    | 0   | 0    | 00:00:38 | 5            |

You should only see the BGP sessions established:

- IBGP session to the R1 router (neighbor 10.0.0.1)
- IBGP session to the R4 router (neighbor 10.0.0.49)

**Step 3** On the R4 router, verify which BGP sessions are established.

```
R4# show bgp summary
BGP router identifier 10.0.0.49, local AS number 1
BGP table version is 145, main routing table version 145
6 network entries using 888 bytes of memory
6 path entries using 384 bytes of memory
3/3 BGP path/bestpath attribute entries using 408 bytes of memory
1 BGP AS-PATH entries using 24 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
BGP using 1704 total bytes of memory
BGP activity 61/55 prefixes, 109/103 paths, scan interval 60 secs
```

| Neighbor     | V | AS  | MsgRcvd | MsgSent | TblVer | InQ | OutQ | Up/Down  | State/PfxRcd |
|--------------|---|-----|---------|---------|--------|-----|------|----------|--------------|
| 10.0.0.17    | 4 | 1   | 8       | 7       | 145    | 0   | 0    | 00:01:03 | 1            |
| 172.16.34.34 | 4 | 300 | 109     | 122     | 145    | 0   | 0    | 01:34:45 | 4            |

You should only see the BGP sessions established:

- EBGP session to the ISP3B router (neighbor 172.16.34.34)
- IBGP session to the R3 router (neighbor 10.0.0.17)

**Note** that there is no IBGP session between routers R1 and R4.

**Step 4** On the R1 router, verify which routes are advertised to the router R3.

```
R1# show ip bgp neighbors 10.0.0.17 advertised-routes
BGP table version is 53, local router ID is 10.0.0.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network         | Next Hop     | Metric | LocPrf | Weight | Path  |
|-----------------|--------------|--------|--------|--------|-------|
| *> 10.0.0.0/28  | 0.0.0.0      | 0      |        | 32768  | i     |
| *> 10.0.1.0/28  | 172.16.11.11 | 0      |        | 0      | 100 i |
| *> 10.0.1.16/28 | 172.16.11.11 | 0      |        | 0      | 100 i |
| *> 10.0.1.32/28 | 172.16.11.11 | 0      |        | 0      | 100 i |
| *> 10.0.1.48/28 | 172.16.11.11 | 0      |        | 0      | 100 i |
| *> 10.0.1.64/28 | 172.16.11.11 | 0      |        | 0      | 100 i |
| *> 10.0.1.80/28 | 172.16.11.11 | 0      |        | 0      | 100 i |

You should see that R1 router advertises these prefixes:

- Locally originated prefix 10.0.0.0/28
- All prefixes originating on the ISP1 router (10.0.1.0/24 subnets) in AS 100.

**Step 5** On the R3 router, verify the content of the BGP table.

```
R3# show ip bgp
BGP table version is 102, local router ID is 10.0.0.17
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|     | Network      | Next Hop  | Metric | LocPrf | Weight | Path  |
|-----|--------------|-----------|--------|--------|--------|-------|
| *>i | 10.0.0.0/28  | 10.0.0.1  | 0      | 100    | 0      | i     |
| *>  | 10.0.0.16/28 | 0.0.0.0   | 0      |        | 32768  | i     |
| *>i | 10.0.0.48/28 | 10.0.0.49 | 0      | 100    | 0      | i     |
| *>i | 10.0.1.0/28  | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.16/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.32/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.48/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.64/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.80/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i | 10.0.3.0/28  | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i | 10.0.3.16/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i | 10.0.3.32/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i | 10.0.3.48/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |

You should see that prefixes originating on the ISP1 router (10.0.1.0/24 prefixes) were received on the R3 router.

**Step 6** On the R3 router, verify which routes are advertised to the router R4.

```
R3# show ip bgp neighbors 10.0.0.49 advertised-routes
BGP table version is 102, local router ID is 10.0.0.17
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | Next Hop | Metric | LocPrf | Weight | Path |
|----|--------------|----------|--------|--------|--------|------|
| *> | 10.0.0.16/28 | 0.0.0.0  | 0      |        | 32768  | i    |

Total number of prefixes 1

You should see that only locally originated prefix (10.0.0.16/28) is advertised to the R4 router. Split horizon rule prevents router R3 to pass received updates from router R1 to be sent to the R4 router.

**Step 7** On the R4 router, verify the content of the BGP table.



```
R4# show ip bgp
BGP table version is 81, local router ID is 10.0.0.49
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|     | Network      | Next Hop     | Metric | LocPrf | Weight | Path  |
|-----|--------------|--------------|--------|--------|--------|-------|
| *>i | 10.0.0.16/28 | 10.0.0.17    | 0      | 100    | 0      | i     |
| *>  | 10.0.0.48/28 | 0.0.0.0      | 0      |        | 32768  | i     |
| *>  | 10.0.3.0/28  | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.16/28 | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.32/28 | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.48/28 | 172.16.34.34 | 0      |        | 0      | 300 i |

You should see that prefixes originating on the ISP1 router (10.0.1.0/24 prefixes) were not received on the R4 router.

**NOTE:** For the prefixes originating on the ISP1 router to be received on the R4 router, either IBGP session between R1 and R4 needs to be configured or the router R3 should be configured as the route reflector to break the split horizon rule for IBGP sessions.

**Step 8** On the ISP3B router, verify the content of the BGP table.

```
ISP3B# show ip bgp
BGP table version is 75, local router ID is 10.0.3.49
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | Next Hop    | Metric | LocPrf | Weight | Path |
|----|--------------|-------------|--------|--------|--------|------|
| *> | 10.0.0.16/28 | 172.16.34.4 |        |        | 0      | 1 i  |
| *> | 10.0.0.48/28 | 172.16.34.4 | 0      |        | 0      | 1 i  |
| *> | 10.0.3.0/28  | 0.0.0.0     | 0      |        | 32768  | i    |
| *> | 10.0.3.16/28 | 0.0.0.0     | 0      |        | 32768  | i    |
| *> | 10.0.3.32/28 | 0.0.0.0     | 0      |        | 32768  | i    |
| *> | 10.0.3.48/28 | 0.0.0.0     | 0      |        | 32768  | i    |

You should see as a consequence of the IBGP split horizon rule that prefixes originating on the ISP1 router (10.0.1.0/24 prefixes) were not received on the ISP3B router.

**Step 9** Configure the BGP cluster ID of 1 on the R3 router. Configure the R3 router to be route reflector for client router R1 and R4.

```
R3(config)# router bgp 1
R3(config-router)# bgp cluster-id 1
R3(config-router)# neighbor 10.0.0.1 route-reflector-client
R3(config-router)# neighbor 10.0.0.49 route-reflector-client
```

**NOTE:** IBGP session is restored automatically after route reflector client is configured on the route reflector router.

# Monitoring Route Reflectors

## Monitoring Route Reflectors

### Monitoring Route Reflectors

Commands that are used to monitor route reflector operation:

```
router# show ip bgp neighbors
```

- Displays whether a neighbor is a route reflector client.

```
router# show ip bgp network/length
```

- Displays the routes that are received from the client as seen on the reflector.
- Displays reflected routes as seen on the client (with originator ID and cluster ID information)

© 2015 Cisco and/or its affiliates. All rights reserved.181

To display information about the TCP and BGP connections to neighbors, use the **show ip bgp neighbors** EXEC command.

```
show ip bgp neighbors [address]
```

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. The command that is issued on the route reflector router, indicates that the neighbor is a route reflector client.

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

```
show ip bgp [network/length [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. If you issue this command on the route reflector, you could see that a particular entry was received from a route reflector client. If you issue this command on the route reflector client, you could see entries in the BGP table, that were reflected by the route reflector.

**Step 10** On the R3 router (which acts as a route reflector), verify that the R1 router is configured as route reflector client.

```

R3# show ip bgp neighbors 10.0.0.1
BGP neighbor is 10.0.0.1, remote AS 1, internal link
 BGP version 4, remote router ID 10.0.0.1
 BGP state = Established, up for 00:07:39
 Last read 00:00:41, last write 00:00:22, hold time is 180, keepalive interval is 60
seconds
 Neighbor sessions:
 1 active, is not multisession capable (disabled)
 Neighbor capabilities:
 Route refresh: advertised and received(new)
 Four-octets ASN Capability: advertised and received
 Address family IPv4 Unicast: advertised and received
 Enhanced Refresh Capability: advertised and received
 Multisession Capability:
 Stateful switchover support enabled: NO for session 1
 Message statistics:
 InQ depth is 0
 OutQ depth is 0

 Sent Rcvd
Opens: 1 1
Notifications: 0 0
Updates: 7 3
Keepalives: 10 10
Route Refresh: 0 0
Total: 18 14
Default minimum time between advertisement runs is 0 seconds

For address family: IPv4 Unicast
Session: 10.0.0.1
BGP table version 150, neighbor version 150/0
Output queue size : 0
Index 5, Advertise bit 1
Route-Reflector Client
5 update-group member
Slow-peer detection is disabled
Slow-peer split-update-group dynamic is disabled

```

**Step 11** On the R3 router (which acts as a route reflector), verify additional information in the BGP table for prefix 10.0.1.0/28, which is originated on the ISP1 router. Make sure that the route was received from the route reflector client.

```

R3# show ip bgp 10.0.1.0/28
BGP routing table entry for 10.0.1.0/28, version 134
Paths: (1 available, best #1, table default)
 Advertised to update-groups:
 5
 Refresh Epoch 1
 100, (Received from a RR-client)
 10.0.0.1 (metric 11) from 10.0.0.1 (10.0.0.1)
 Origin IGP, metric 0, localpref 100, valid, internal, best

```

**Step 12** On the R4 router (which acts as a route reflector client), verify additional information in the BGP table for prefix 10.0.1.0/28, which is originated on the ISP1 router. Make sure that the route was reflected from the route reflector router R3.

```

R4# show ip bgp 10.0.1.0/28
BGP routing table entry for 10.0.1.0/28, version 99
Paths: (1 available, best #1, table default)
 Advertised to update-groups:
 6
 Refresh Epoch 2
 100
 10.0.0.1 (metric 21) from 10.0.0.17 (10.0.0.17)
 Origin IGP, metric 0, localpref 100, valid, internal, best
 Originator: 10.0.0.1, Cluster list: 0.0.0.1

```

You should see that originator of this reflected route was route reflector client R1 (10.0.0.1) and the BGP cluster ID is 1.

**Step 13** On the R3 router, verify which routes are advertised to the R4 router.

```

R3# show ip bgp neighbors 10.0.0.49 advertised-routes
BGP table version is 150, local router ID is 10.0.0.17
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

```

| Network          | Next Hop  | Metric | LocPrf | Weight | Path  |
|------------------|-----------|--------|--------|--------|-------|
| *>i 10.0.0.0/28  | 10.0.0.1  | 0      | 100    | 0      | i     |
| *> 10.0.0.16/28  | 0.0.0.0   | 0      |        | 32768  | i     |
| *>i 10.0.0.48/28 | 10.0.0.49 | 0      | 100    | 0      | i     |
| *>i 10.0.1.0/28  | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.1.16/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.1.32/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.1.48/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.1.64/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.1.80/28 | 10.0.0.1  | 0      | 100    | 0      | 100 i |
| *>i 10.0.3.0/28  | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i 10.0.3.16/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i 10.0.3.32/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |
| *>i 10.0.3.48/28 | 10.0.0.49 | 0      | 100    | 0      | 300 i |

```

 Network Next Hop Metric LocPrf Weight Path
Total number of prefixes 13

```

You should see that prefixes originated on the ISP1 (10.0.1.0/24 subnets) router are being advertised from the R3 router to the IBGP neighbor R4 because of the router R3 acting as a route reflector.

**Step 14** On the R4 router, verify the content of the BGP table.

R4# **show ip bgp**

BGP table version is 106, local router ID is 10.0.0.49

Status codes: s suppressed, d damped, h history, \* valid, > best, i - internal,  
r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,  
x best-external, a additional-path, c RIB-compressed,

Origin codes: i - IGP, e - EGP, ? - incomplete

RPKI validation codes: V valid, I invalid, N Not found

|     | Network      | Next Hop     | Metric | LocPrf | Weight | Path  |
|-----|--------------|--------------|--------|--------|--------|-------|
| *>i | 10.0.0.0/28  | 10.0.0.1     | 0      | 100    | 0      | i     |
| *>i | 10.0.0.16/28 | 10.0.0.17    | 0      | 100    | 0      | i     |
| *>  | 10.0.0.48/28 | 0.0.0.0      | 0      |        | 32768  | i     |
| *>i | 10.0.1.0/28  | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.16/28 | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.32/28 | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.48/28 | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.64/28 | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>i | 10.0.1.80/28 | 10.0.0.1     | 0      | 100    | 0      | 100 i |
| *>  | 10.0.3.0/28  | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.16/28 | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.32/28 | 172.16.34.34 | 0      |        | 0      | 300 i |
| *>  | 10.0.3.48/28 | 172.16.34.34 | 0      |        | 0      | 300 i |

You should see that prefixes originating on the ISP1 router (10.0.1.0/24 prefixes) were received on the R4 router, because the router R3 is configured as a route reflector.

**Step 15** On the ISP3B router, verify the content of the BGP table.

ISP3B# **show ip bgp**

BGP table version is 84, local router ID is 10.0.3.49

Status codes: s suppressed, d damped, h history, \* valid, > best, i - internal,  
r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,  
x best-external, a additional-path, c RIB-compressed,

Origin codes: i - IGP, e - EGP, ? - incomplete

RPKI validation codes: V valid, I invalid, N Not found

|    | Network      | Next Hop    | Metric | LocPrf | Weight | Path    |
|----|--------------|-------------|--------|--------|--------|---------|
| *> | 10.0.0.0/28  | 172.16.34.4 |        |        | 0      | 1 i     |
| *> | 10.0.0.16/28 | 172.16.34.4 |        |        | 0      | 1 i     |
| *> | 10.0.0.48/28 | 172.16.34.4 | 0      |        | 0      | 1 i     |
| *> | 10.0.1.0/28  | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.1.16/28 | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.1.32/28 | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.1.48/28 | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.1.64/28 | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.1.80/28 | 172.16.34.4 |        |        | 0      | 1 100 i |
| *> | 10.0.3.0/28  | 0.0.0.0     | 0      |        | 32768  | i       |
| *> | 10.0.3.16/28 | 0.0.0.0     | 0      |        | 32768  | i       |
| *> | 10.0.3.32/28 | 0.0.0.0     | 0      |        | 32768  | i       |
| *> | 10.0.3.48/28 | 0.0.0.0     | 0      |        | 32768  | i       |

You should see that prefixes originating on the ISP1 router (10.0.1.0/24 prefixes) were received on the ISP3B router.

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- A proper migration plan is important because the change to BGP confederation involves a major reconfiguration of BGP routing.
- The **neighbor ip-address route-reflector-client** command specifies an IBGP neighbor to be a client of this reflector
- The **show ip bgp neighbors** command has been modified to display whether a BGP neighbor is part of a BGP confederation.

# Lesson 4: Module Summary

---

## Overview

This topic summarizes the key points that were 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.
- There are only two Cisco IOS commands that are used to configure route reflectors: **bgp cluster-id** and neighbor *ip-address* route-reflector-client.





# Lesson 5: 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.

1. 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.
  
2. 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.
  
3. The IGP protocol is not needed in the service provider network, as information about next hop attribute gets advertised via BGP updates. True or false? (Source: "Scaling IGP and BGP in Service Provider Networks")
  - A. true
  - B. false

4. 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.
5. In service provider network It is recommended to to redistribute BGP routes into IGP protocol. True or false? (Source: "Scaling IGP and BGP in Service Provider Networks")
- A. true
  - B. false
6. When using next hop self feature on the PE router in the service provider network, IGP can be relieved of the burden of carrying information about the access link. True or false? (Source: "Scaling IGP and BGP in Service Provider Networks")
- A. true
  - B. false
7. 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
8. 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
9. 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. True or false? (Source: "Scaling IGP and BGP in Service Provider Networks")
- A. true
  - B. false

10. What is the main problem that is solved by implementing BGP route reflectors? (Source: "Introducing and Designing 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
11. How does a route reflector modify the IBGP split-horizon rule? (Source: "Introducing and Designing 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
12. When incoming BGP update is received from EBGp peer on route reflector, the update is forwarded to both IBGP and EBGp peers. True or false? (Source: "Introducing and Designing Route Reflectors")
- A. true
  - B. false
13. Why are redundant route reflectors mandatory in any high-availability network design? (Source: "Introducing and Designing 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.
14. What is the main reason for implementing redundant route reflectors with clusters? (Source: "Introducing and Designing 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
15. How does the originator-ID attribute assist in the elimination of routing loops that are caused by redundant route reflector designs? (Source: "Introducing and Designing 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.

16. What can occur if a client has IBGP neighbor relationships with other routers that are not configured as route reflectors? (Source: "Introducing and Designing 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.
17. Which potential problem can occur if a client does not have an IBGP session with all route reflectors in a cluster? (Source: "Introducing and Designing 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.
18. Which problem are hierarchical route reflectors designed to solve? (Source: "Introducing and Designing 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
19. 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
20. 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
21. 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**

22. The command **neighbor *ip-address* route-reflector-client** command should be used on client. True or false? (Source: "Configuring and Monitoring Route Reflectors")
- A. true
  - B. false



# Module Self-Check Answers

---

## Answer Key

1. B, C, D
2. C
3. B
4. A
5. B
6. A
7. A, B, D
8. A, B, D
9. A
10. B
11. C
12. A
13. D
14. A
15. C
16. D
17. B
18. B
19. A, D
20. A, B, D
21. B
22. B





# Module 7: Optimizing BGP Scalability

---

## Introduction

[BGP](#) is designed for reliability and scalability. As such, it has become the standard protocol that is used to carry the more than 600,000 prefixes in the Internet today. BGP also has a tremendous amount of flexibility with regard to administrative policy controls, route selection, performance tuning, and scalability features.

You will learn about 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.

Upon completing this module, you will be able to:

- 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 on the router in a service provider will increase. [BGP](#) processing affects both router resources and network convergence time. It is important that the network convergence is as fast as possible to ensure accurate routing information between domains. It is also important that router resources are optimized whenever possible. Cisco IOS performance improvements for BGP are designed to aid network administrators in achieving these goals.

In this lesson, you will learn about 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. Improvements that are mentioned are [PMTU](#) discovery, input hold queue, BGP [PIC](#), [BFD](#), BGP [NSF](#) awareness, BGP scan time, advertisement interval, keepalive, and hold-down timers.

Upon completing this lesson, you will be able to:

- Describe convergence in BGP networks
- Describe the BGP router processes and their functions
- Describe the effects of BGP processes on router CPU resources
- Describe the features that can be used to improve BGP convergence
- Describe using PMTU discovery
- Describe increasing the input queue depth
- Describe the BGP prefix independent convergence feature and configuration
- Describe the BGP BFD feature and configuration
- Describe the function of the BGP Nonstop Forwarding Awareness feature

- Explain BGP scan timer configuration
- Explain BGP advertisement interval configuration
- Explain BGP keepalive and hold-down timer configuration

# BGP Convergence

As the number of routes in the Internet routing table grows, service providers and large enterprise customers are experiencing a dramatic increase in the 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.

## 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 300,000 prefixes.
- Network convergence times can range from 10 minutes to more than 1 hour.
- BGP is considered converged when:
  - All routes have been accepted.
  - All routes have been installed in the routing table.
  - The input queue and output queue for all peers is 0.
  - The table version for all peers equals the table version of the BGP table.

© 2015 Cisco and/or its affiliates. All rights reserved.

184

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 input queue and output queue for all peers is 0.
- The table version for all peers equals the table version of the BGP table.

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

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.

[BGP](#) consists of several processes, each of them running at different times, depending on the task that is handles.

| 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 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.                                      | Once per second and when adding, removing, or soft-reconfiguring a BGP peer. |

- BGP scanner and BGP router are responsible for many calculations and can lead to high CPU utilization.

© 2015 Cisco and/or its affiliates. All rights reserved.

185

Several 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.

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.

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 this specific process handles. Because BGP scanner and BGP router are responsible for many calculations, you may notice high CPU utilization during the running of either one of these processes.

# CPU Effects of BGP Processes

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.

© 2015 Cisco and/or its affiliates. All rights reserved.186

On routers that carry a large Internet routing table, you can expect high CPU utilization stemming from the BGP scanner process 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 [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.

The BGP scanner runs through the entire BGP table. So the duration of the high CPU utilization condition that the BGP scanner process causes 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 [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, and is suspended, then ICMP runs and is suspended, then 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.

## CPU Effects of BGP Processes Example

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 on a router during BGP convergence as BGP exchanges prefixes with two [EBGP](#) peers.

- Capture a baseline for normal CPU utilization before starting the test.

```
R1# show process cpu
CPU utilization for five seconds: 0%/0%; one minute: 4%; five minutes: 5%
```

- After the test starts, the CPU reaches 100 percent utilization. The **show process cpu** command shows that the BGP router causes high CPU condition, denoted by 139 (the Cisco IOS process ID for the BGP router) in the following output:

```
R1# 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
```

- 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.

```
R1# show ip bgp summary
Neighbor V AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/PfxRcd
192.168.12.2 4 64512 309453 157389 19981 0 253 22:06:44 111633
172.16.11.11 4 65101 188934 1047 40081 41 0 00:07:51 58430
```

- 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.

```
R1# show process cpu
CPU utilization for five seconds: 3%/0%; one minute: 82%; five minutes: 91%
```

- 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

[BGP](#) convergence can often be an issue in networks requiring quick propagation of routing information.

## Improving BGP Convergence

- Queuing to TCP peer connections
  - BGP automatically queues data aggressively from the BGP output queue
- Deploy BGP peer groups
  - Simplifies BGP configuration and enhances BGP scalability.
- Enable the path MTU feature
  - Improves efficiency by dynamically determining the largest MTU that you can use without creating packets that need to be fragmented.
- Implement BFD:
  - Reduces BGP convergence by fast detection of neighbor failure.
- Implement BGP PIC:
  - Reduces convergence by storing BGP backup/alternate path in RIB and FIB.
- Increase interface input queues
  - Improves convergence by reducing dropped TCP ACKs.
- Enable BGP NSF Awareness feature
  - Forwards packets even during failure and thus, saves resources.

© 2015 Cisco and/or its affiliates. All rights reserved.

187

To reduce BGP convergence time and the high [CPU](#) utilization that a running BGP process causes, the following performance improvement features are available:

- **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. 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. 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. It depends on the number of peer group members, the number of prefixes in the table, and the number of prefixes that are advertised. Consequently they can significantly reduce router resource requirements.
- **Bidirectional Forwarding Detection:** [BFD](#) is a detection protocol designed to provide fast forwarding path failure detection times for all media types, encapsulations, topologies, and routing protocols.
- **BGP Prefix Independent Convergence:** The BGP [PIC](#) feature improves BGP convergence after a network failure. This convergence is applicable to both core and edge failures and can be used in both IP and [MPLS](#) networks. The BGP PIC feature creates and stores a backup/alternate path in the [RIB](#), [FIB](#), and Cisco Express Forwarding (CEF). When a failure is detected, the backup/alternate path can immediately take over, thus enabling fast failover.

- **Path MTU feature:** All TCP sessions are bounded by a limit on the number of bytes that a single packet can transport. This limit, which is known as the [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. 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 [ACKs](#) to the advertising BGP speaker, causing the BGP speaker to receive a flood of TCP ACKs in a short 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 [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 helps reduce the number of dropped TCP ACKs, reducing the amount of work that BGP must do to converge.
- **BGP Nonstop Forwarding Awareness feature:** [NSF](#) awareness allows a 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 router-hold timer sets the maximum time that the NSF-aware router will hold known routes for an NSF-capable neighbor. The deployment of BGP NSF awareness can improve the overall network stability by reducing the amount of resources that are normally required for reestablishing peering with a failed router.

### Improving BGP Convergence (Cont.)

BGP convergence can also be improved to some extent by:

- Lowering scan time interval for the BGP scanner process.
- Lowering advertisement interval between BGP neighbors.
- Lowering keepalive and hold-down timers.

#### Limitation:

- Not recommended in routers dealing with large BGP tables.
- Could lead to CPU or memory exhaustion.
- Lower hold-down timers could lead to undesired session terminations.

You also need to improve BGP convergence in certain scenarios; for example, in networks using the conditional advertisement feature. There are four extra BGP parameters that you 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
- **Keepalive timers:** Controlling the BGP session by sending keepalive messages at specific time interval.
- **Hold-down timers:** Controlling the BGP session by waiting for successive keepalive message from a BGP neighbor.

Network administrators must take care when configuring these 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. Lowering the keepalive and hold-down timers could lead to undesired session terminations when keepalives are not received due to network congestion.

# PMTU Discovery

[PMTU](#) discovery feature can be used to reduce [BGP](#) convergence. PMTU discovery is a method for maximizing the use of available bandwidth in the network between the endpoints of a [TCP](#) connection.

## PMTU Discovery

PMTU discovery is used to automatically determine TCP MSS used for TCP connections from a router.

- Prior to Cisco IOS Release 15.0, the default TCP MSS value for BGP was 536 bytes.
- From Cisco IOS Release 15.0, PMTU is enabled by default.
- Small TCP MSS affects BGP convergence:
  - Higher TCP MSS can improve BGP convergence.

```
router(config)# ip tcp path-mtu-discovery [age-timer {minutes | infinite}]
```

- The 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 is 10 minutes).

© 2015 Cisco and/or its affiliates. All rights reserved.189

PMTU discovery works by setting the [DF](#) option bit in the IP headers of outgoing packets. Then, any device along the path whose [MTU](#) is smaller than the packet will drop it. Then it will send back an [ICMP](#) Fragmentation Needed (Type 3, Code 4) message containing its MTU, allowing the source host to reduce its PMTU appropriately. The process repeats until the MTU is small enough to traverse the entire path without fragmentation.

The default [TCP MSS](#) for BGP traffic is 536 bytes. Enabling PMTU discovery, and thus using a higher MSS for BGP traffic, can significantly improve BGP convergence, since it takes fewer packets to send BGP updates.

To enable the PMTU discovery feature for all new TCP connections from the router, use the **ip tcp path-mtu-discovery** global configuration command. To disable the function, use the **no** form of this command:

```
ip tcp path-mtu discovery [age-timer {minutes | infinite}]
```

## Syntax Description

| Parameter                        | Description                                                                                                                                                   |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>age timer</b> <i>minutes</i>  | (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. |
| <b>age-timer</b> <i>infinite</i> | (Optional) Turns off the age timer.                                                                                                                           |

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."

## Monitoring PMTU Discovery

To verify the TCP MSS that is used between BGP neighbors, use the **show ip bgp neighbors**. You can use **include max data** filter to display relevant output only, as shown in the example.

### Monitoring PMTU Discovery

```
router# show ip bgp neighbor | 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 (before Cisco IOS Release 15.0).

```
router# show ip bgp neighbor | 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.

© 2015 Cisco and/or its affiliates. All rights reserved.190

The first example shows the default size of the MSS, 536 bytes, 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. In the example at the bottom, the output shows that the PMTU feature has been enabled. 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

Another feature can be used to improve [BGP](#) convergence is to increase input queue depth. BGP routers might experience packet drops on an interface due to large number of [TCP ACK](#) segments, which are used to acknowledge receipt of BGP updates.

## Increasing Input Queue Depth

- Input queue on an interface specifies how many packets can be queued before dropping the packets.
- BGP routers with several peers might experience packet drops on an interface due to many TCP ACK segments.

```
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 that are caused by input queue drops of TCP ACKs.

© 2015 Cisco and/or its affiliates. All rights reserved.191

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.

To specify the size of the IP input or output queue on an interface, use the **hold-queue** command in interface configuration mode. To restore the default values for an interface, use the **no** form of this command with the appropriate keyword.

```
hold-queue length {in | out}
```

## Syntax Description

| Parameter     | Description                                                                                                                                                                                                  |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>length</i> | Integer that specifies the maximum number of packets in the queue. The range of allowed values is 0 to 65535.                                                                                                |
| <b>in</b>     | 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.  |
| <b>out</b>    | 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. |

---

|             |                                                                                                                                                                     |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | Several considerations should be taken into account when increasing interface input queues. Increasing the interface input queues will increase memory utilization. |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

---

|             |                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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). |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

---

|             |                                                                                                                                                                                                          |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | The Cisco 12000 Series now uses a default <a href="#">SPD</a> headroom value of 1000. It retains the default input queue size of 75. Use the <b>show spd</b> command to view these special input queues. |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

## Monitoring PMTU Input Queue Depth

Use the **show interface** *{interface-identifier}* command to displays the current input queue levels and the number of incoming packets dropped.

## Monitoring Input Queue Depth

```
router# show interface GigabitEthernet0/0
GigabitEthernet0/0 is up, line protocol is up
 Hardware is AmdP2, address is aabb.cc00.4900 (bia aabb.cc00.4900)
 MTU 1500 bytes, BW 10000 Kbit/sec, DLY 1000 usec,
 reliability 255/255, txload 1/255, rxload 1/255
 Encapsulation ARPA, loopback not set
 Keepalive set (10 sec)
 ARP type: ARPA, ARP Timeout 04:00:00
 Last input never, output never, output hang never
 Last clearing of "show interface" counters never
 Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
```

- Displays interface information, including input queue depth.

© 2015 Cisco and/or its affiliates. All rights reserved.

192

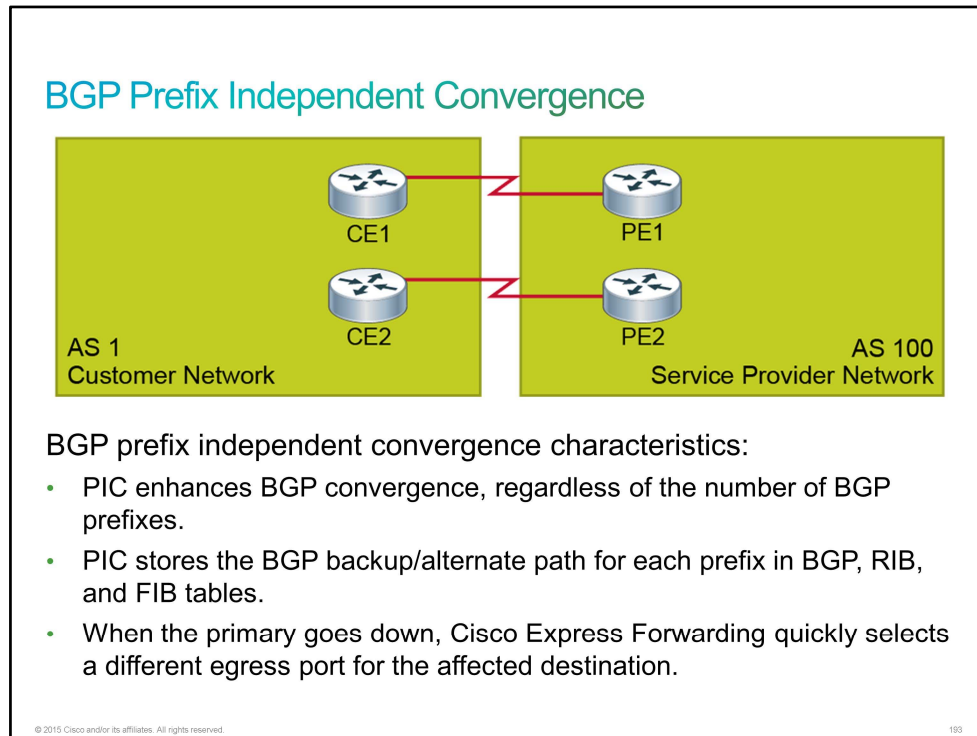
The input queue  $w/x/y/z$  counter displays the current input queue.  $w$  is the current number of packets in the input queue and  $x$  is the queue depth. The drops counter,  $y$ , indicates the number of incoming packets that have been dropped. In the example, there is no packet in the input queue, the depth of queue is left on default, 75, and no packet were 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 Prefix Independent Convergence

[BGP](#) Prefix Independent Convergence feature improves BGP convergence after a network failure. It calculates a second best path, along with the primary best path and stores both paths in the Cisco Express Forwarding (CEF).



Under normal circumstances, BGP can take several seconds to a few minutes to converge after a network change. At a high level, BGP goes through the following process:

1. BGP learns of failures through either [IGP](#) or [BFD](#) events or interface events.
2. BGP withdraws the routes from the [RIB](#), and the RIB withdraws the routes from the [FIB](#) and distributed FIB. This process clears the data path for the affected prefixes.
3. BGP sends withdraw messages to its neighbors.
4. BGP calculates the next best path to the affected prefixes.
5. BGP inserts the next best path for affected prefixes into the RIB, and the RIB installs them in the FIB and distributed FIB.

This process takes a few seconds or a few minutes to complete. The time depends on the latency of the network, the convergence time across the network, and the local load on the devices. The data plane converges only after the control plane converges.

The BGP [PIC](#) functionality is achieved by additional functionality in the BGP, RIB, and CEF.

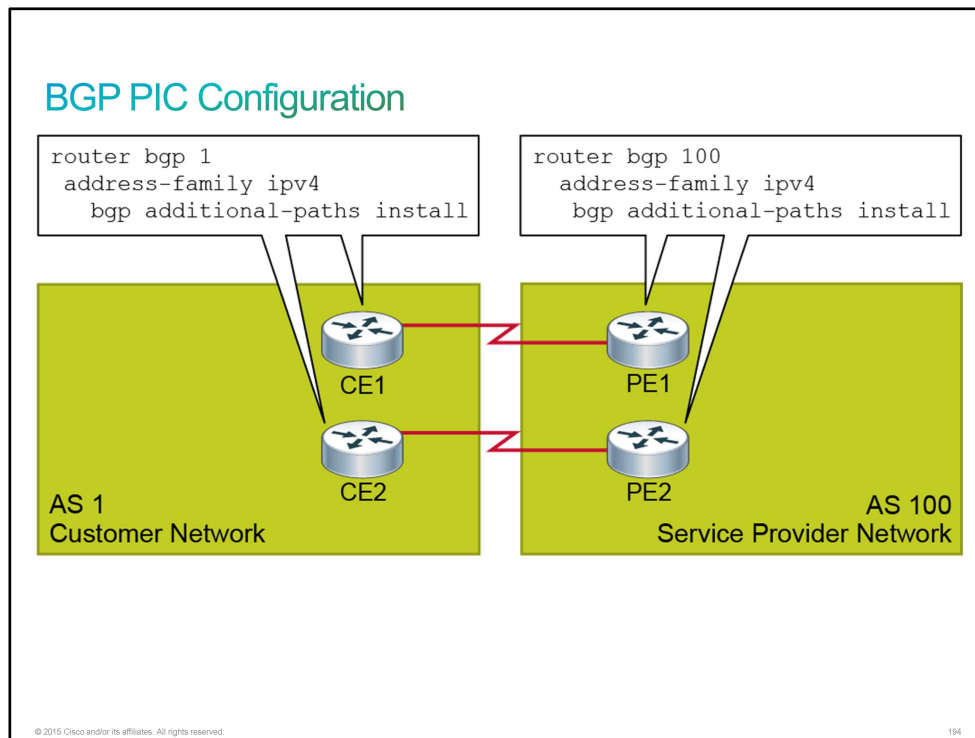
BGP PIC affects prefixes under [IPv4](#) and [VPNv4](#) address families. For those prefixes, BGP calculates a second best path, along with the primary best path. (The second best path is called the backup/alternate path.) BGP installs the best path and the backup/alternate paths for the affected prefixes into the BGP RIB. The backup/alternate path provides a fast reroute mechanism to counter a single network failure.

For BGP PIC, RIB installs an alternate path per route if one is available. With the BGP PIC functionality, a RIB that selects a BGP route containing a backup/alternate path installs the backup/alternate path with the best path.

With BGP PIC, CEF stores an alternate path per prefix. When the primary path goes down, CEF searches for the backup/alternate path in a prefix-independent manner. CEF also listens to BFD events to rapidly detect local failures. Upon detection of a failure, CEF detects the alternate next hop for all prefixes that are affected by the failure.

## BGP PIC Configuration

To enable BGP PIC on Cisco IOS Software routers, first enter BGP configuration mode using the **router bgp** command. Then, enter the appropriate address family configuration mode and use the **bgp additional-paths install** command to enable BGP PIC.

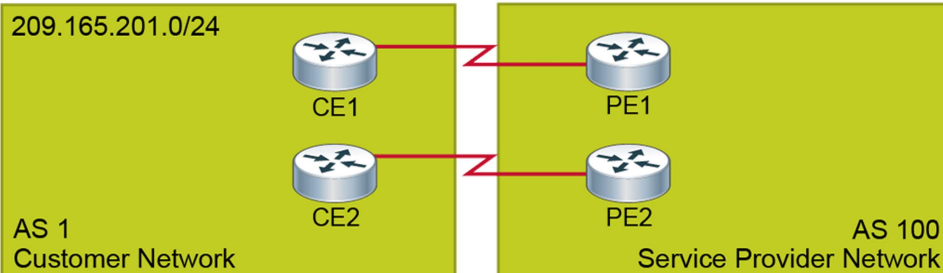


In the example, CE1 and CE2 are configured with the BGP PIC feature. BGP computes PE1 as the best path and PE2 as the backup/alternate path, and installs both the routes into the RIB and CEF plane.

No policies should be set on CE1 and CE2 for the [EBGP](#) peers PE1 and PE2. Both CE routers must point to the EBGp route as the next hop. On CE1, the next hop to reach the Internet is through PE1. On CE2, the best path to reach the Internet is PE2. CE2 advertises itself as the next hop to CE1, and CE1 does the same to CE2. As a result, CE1 has two paths for the specific prefix. It usually selects the directly connected EBGp path over the IBGP path, according to the best path selection rules. Similarly, CE2 has two paths, an EBGp path through PE2 and an IBGP path through CE1-PE1.

If the CE1-PE1 link or PE1 goes down, BGP recomputes the best path, removing the next hop PE1 from RIB and reinstalling CE2 as the next hop into the RIB and CEF. CE1 automatically gets a backup/alternate repair path into CEF, and the traffic loss during forwarding is now in subseconds, achieving fast convergence.

## BGP PIC Configuration (Cont.)



```
PE1# show ip bgp
BGP table version is 65, local router ID is 10.0.3.81
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network             | Next Hop   | Metric | LocPrf | Weight | Path  |
|---------------------|------------|--------|--------|--------|-------|
| *> 209.165.201.0/24 | 1.0.0.101  |        |        |        | 0 1 i |
| *b i                | 10.0.100.2 | 0      | 100    | 0 1    | i     |

© 2015 Cisco and/or its affiliates. All rights reserved.

195

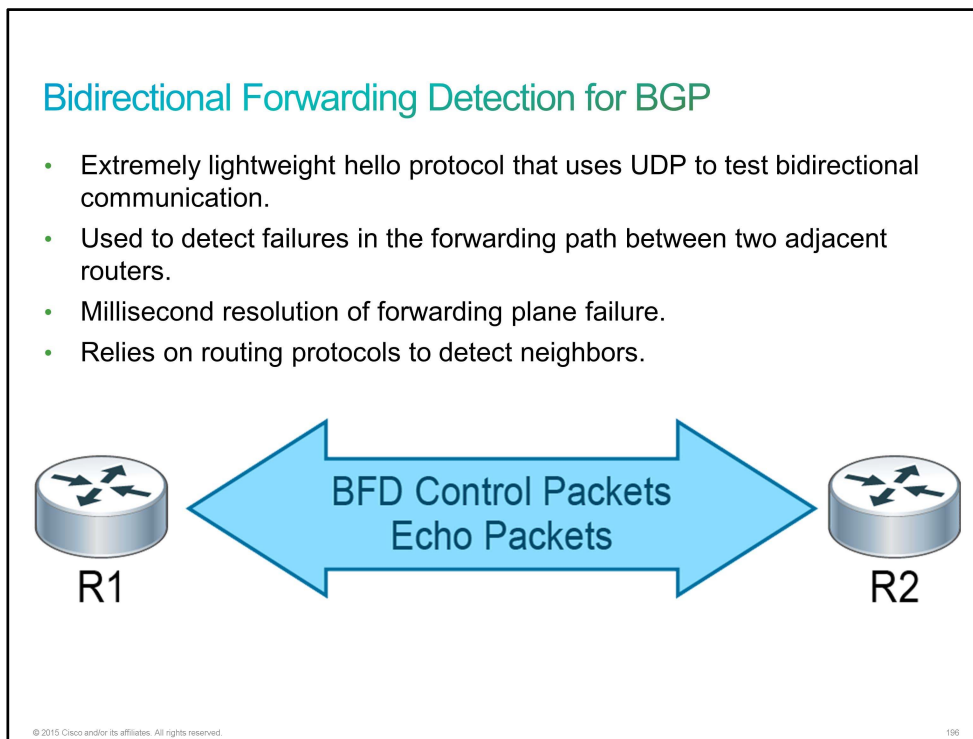
You can verify the backup route by issuing **show ip bgp** command. In the example, you can see that PE1 router has two paths to reach the 209.165.201.0/24 network at the customer side. The internal route, via PE2 router, is marked as the backup route.

Additionally, you can verify that the backup path is really stored in the CEF, by issuing **show ip cef prefix detail** command. The backup route is labeled as "repair." Without BGP PIC feature enabled, only one path is stored in the CEF.

```
PE1# show ip cef 209.165.201.0/24 detail
209.165.201.0/24, epoch 0, flags rib only nlabel, rib defined all labels
recursive via 1.0.0.101
 attached to Serial0
recursive via 10.0.100.2, repair
 attached to Ethernet0/0
```

# Bidirectional Forwarding Detection for BGP

[BFD](#) provides a low-overhead, short-duration method of detecting failures in the forwarding path between two adjacent routers, including the interfaces, data links, and forwarding planes. BFD is a detection protocol that you enable at the interface and routing protocol levels. Cisco supports the BFD asynchronous mode, which depends on the sending of BFD control packets between two systems to activate and maintain BFD neighbor sessions between routers. Therefore, in order for a BFD session to be created, you must configure BFD on both BFD peers. Once BFD has been enabled on the interfaces and at the router level for the appropriate routing protocols, a BFD session is created, BFD timers are negotiated, and the BFD peers will begin to send BFD control packets to each other at the negotiated interval.



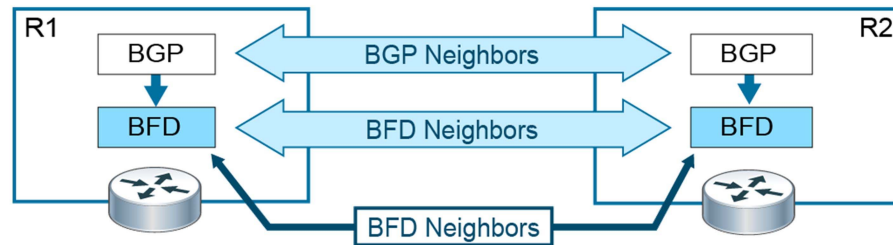
BFD provides fast BFD peer failure detection times independently of all media types, encapsulations, topologies, and routing protocols [BGP](#), [EIGRP](#), [IS-IS](#), and [OSPF](#). By sending rapid failure detection notices to the routing protocols in the local router to initiate the routing table recalculation process, BFD contributes to greatly reduced overall network convergence time.

The BFD protocol has no discovery mechanisms to detect neighbors; it is designed solely as an agent for other applications requiring fast failure detection. Whenever a routing protocol that is configured to use BFD detects a new neighbor, it requests availability tracking from BFD.

BFD can rely on control packets or on echo packets. Echo packets are IP packets addressed to the router itself but sent to the Layer 2 address of the next-hop node. The echo packets thoroughly test the complete bidirectional forwarding path between adjacent routers because they have to be transmitted by the originating router, propagated to the adjacent router, received by its interface logic, switched by its forwarding engine, and sent back to the originator (because the IP packet is addressed to the router itself).

For example, when R1 sends a BFD echo packet, it sets the destination [IP address](#) in the packet to its own interface IP address and the [MAC address](#) in the Layer 2 frame header to the MAC address of the neighbor (R2). When R2 receives the packet, it performs a Layer 3 lookup and sends the packet toward its final destination (back to R1).

## Bidirectional Forwarding Detection for BGP (Cont.)



### BFD operation:

- Routing protocol (BFD client) bootstraps BFD to create a BFD session to a neighbor:
  - BFD client receives link status change notification.
  - Receive and transmit intervals are negotiated and configurable.
- The two systems agree on a method to detect failure.
- In case of failure, BFD notifies the BFD client:
  - The BFD client independently decides on the action.

© 2015 Cisco and/or its affiliates. All rights reserved.

197

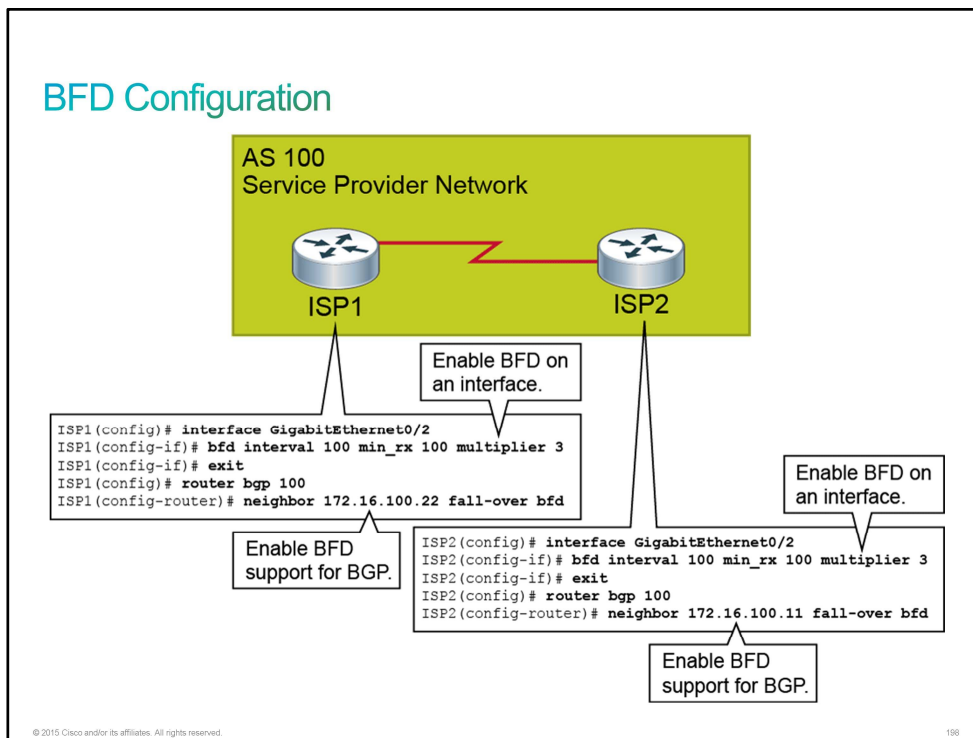
When a routing protocol (BGP, for example) discovers a neighbor, it sends a request to the local BFD process to initiate a BFD neighbor session with the BGP neighbor router. Then the BFD neighbor session with the BGP neighbor router is established. If there is a failure on the link between neighbors, the BFD neighbor session with the BGP neighbor router is torn down. BFD notifies the local BGP process that the BFD neighbor is no longer reachable. The local BGP process tears down the BGP neighbor relationship. If an alternative path is available, the routers will immediately start converging on it.

---

**Note** If multiple routing protocols want to establish BFD sessions with the same remote system for the same routed protocol ([IPv4](#) or [IPv6](#)), all must share a single BFD session.

---

# BFD Configuration



To configure BFD, you first have to enable the BFD, using the **bfd interval** command in the interface configuration mode.

```
bfd interval send-timer min_rx receive-timer multiplier multiplier
```

## Syntax Description

| Parameter                           | Description                                                                                                                                                             |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>interval</b> <i>send-timer</i>   | The rate, in milliseconds, at which BFD control packets will be sent to BFD peers. The valid range for the <i>send-timer</i> is from 50 to 999.                         |
| <b>min_rx</b> <i>receiver-timer</i> | The rate, in milliseconds, at which BFD control packets will be expected to be received from BFD peers. The valid range for the <i>receive-timer</i> is from 50 to 999. |
| <b>multiplier</b> <i>multiplier</i> | The number of BFD packets that can be lost before the BFD peer is declared "down."                                                                                      |

Then you have to enable BFD support for BGP, using the **neighbor fall-over bfd** command.

```
neighbor ip-address fall-over bfd
```

## Syntax Description

| Parameter            | Description                          |
|----------------------|--------------------------------------|
| <b>fall-over bfd</b> | Enables BFD for individual neighbor. |

# BGP Nonstop Forwarding Awareness

Usually, when a networking device restarts, all routing peers of that device detect that the device went down and then came back up. This transition results in what is called a routing flap, which could spread across multiple routing domains. Routing flaps that are caused by routing restarts create routing instabilities, which are detrimental to the overall network performance. Cisco Nonstop Forwarding (NSF), also known as graceful restart, works with the [SSO](#) to minimize the amount of time a network is unavailable to its users following a switchover. The main objective of Cisco NSF is to continue forwarding IP packets following a route processor switchover in platforms with dual [RPs](#), thus reducing network instability.

## Cisco Nonstop Forwarding

### Cisco Nonstop Forwarding

- Cisco NSF is applicable in platforms with dual RPs and works together with SSO.
- Cisco NSF allows:
  - Routing neighbor relationships remain established during SSO.
  - Routes on neighboring routers remain valid.
  - Forwarding of data packets continues while the routing process on the new RP converges.
- Cisco NSF is supported by:
  - Routing protocols (OSPF, IS-IS, EIGRP, BGP)
  - Forwarding operation (Cisco Express Forwarding)
- The device must be Cisco NSF-capable.
- The neighboring device must be Cisco NSF-aware.

© 2015 Cisco and/or its affiliates. All rights reserved.

199

Cisco NSF allows for the forwarding of data packets to continue along known routes while the routing protocol information is being restored following a switchover. With [NSF](#), peer networking devices do not experience routing flaps. Data traffic is forwarded through intelligent line cards or dual forwarding processors, while the standby RP assumes control from the failed active RP during a switchover.

Cisco NSF is supported by four protocols for routing ([BGP](#), [EIGRP](#), [OSPF](#), and [IS-IS](#)), and by Cisco Express Forwarding (CEF) for forwarding. The routing protocols have been enhanced with Cisco NSF capability and awareness, meaning that routers running these protocols can detect a switchover and take the necessary actions to continue forwarding network traffic and to recover route information from the peer devices.

A networking device is said to be Cisco NSF-aware if it is running NSF-compatible software. A device is said to be Cisco NSF-capable if it has been configured to support NSF and can rebuild routing information from NSF-aware or NSF-capable neighbors.

Cisco NSF depends on CEF to continue forwarding packets during switchover while the routing protocols rebuild the [RIB](#) tables. Once the routing protocols have converged, CEF updates the [FIB](#) table and removes old route entries. CEF, in turn, updates the line cards with the new FIB information.



## Cisco Nonstop Forwarding (Cont.)

Cisco NSF overview:

- One RP is active, one is standby.
- Cisco Express Forwarding on the active RP synchronizes the FIB and adjacency table to the standby RP.
- Upon switchover, the new active RP uses the old FIB and adjacency table to forward packets while the routing protocol reconverges.
- Routing protocol must:
  - Establish neighbor relationship without causing a reset of neighbor relationship
  - Learn routing information
- As the routing protocol starts to repopulate the RIB, it updates Cisco Express Forwarding.

© 2015 Cisco and/or its affiliates. All rights reserved.

200

Cisco NSF always operates together with SSO. In specific Cisco networking devices that support dual RPs, SSO establishes one of the RPs as the active processor while the other RP is designated as the standby processor, and then synchronizes the FIB and adjacency table between them. A switchover from the active to the standby processor occurs when the active RP fails, is removed from the networking device, or is manually taken down for maintenance. During the switchover, the new active RP uses the old FIB and adjacency table to forward packets while the routing protocol reconverges. Once the routing protocols have updated the RIB, CEF updates the FIB table and removes old route entries. CEF, in turn, updates the line cards with the new FIB information.

The routing protocols (for example, BGP) run only on the active RP, and they receive routing updates from their neighbor routers. Routing protocols do not run on the standby RP. If Cisco NSF is not configured, a neighboring device will tear down the adjacency during a switchover. The result would be a routing flap that would spread across the network.

Using Cisco NSF, the routing protocols on an NSF-capable device requests that the NSF-aware neighbor devices send routing information to help rebuild the routing tables. The Cisco NSF-aware neighbor will not tear down the neighbor relationship. It will re-establish the neighbor relationship with the Cisco NSF-capable device, and will send routing information to synchronize the routing table on an NSF-capable device.

A Cisco NSF-aware router can peer with Cisco NSF-capable routers and facilitate the resynchronization of routing information with such routers.

## BGP Nonstop Forwarding Awareness

The BGP NSF Awareness feature allows the BGP peers of the failing router to retain the routing information that is advertised by the failing router. It also 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.



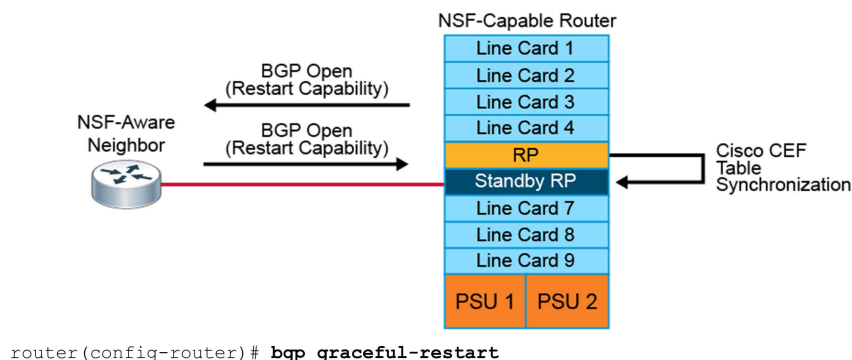
## BGP Nonstop Forwarding Awareness

- 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)

© 2015 Cisco and/or its affiliates. All rights reserved.

201

## BGP Nonstop Forwarding Awareness (Cont.)



© 2015 Cisco and/or its affiliates. All rights reserved.

202

The BGP NSF feature provides an NSF-aware router with the capability to detect a neighbor that is undergoing an SSO operation, maintain the peering session with this neighbor, retain known routes, and continue to forward packets for these routes.

BGP support for Cisco NSF requires that neighbor routers are NSF-aware or NSF-capable. The graceful restart mechanism also enables Cisco NSF awareness in BGP. A router that is NSF-aware functions like a router that is NSF-capable, with one exception: An NSF-aware router cannot perform an SSO operation. However, a router that is NSF-aware can maintain a peering relationship with a NSF-capable neighbor during an NSF SSO operation. It can hold routes for this neighbor during the SSO operation.

---

|             |                                                                                                                                                  |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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 deployment of BGP NSF awareness can minimize 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)

BGP NSF improves the overall network stability by reducing the amount of resources that are normally required for reestablishing peering with a failed router.

When a Cisco NSF-capable router begins a BGP session with a BGP peer, it sends an OPEN message to the peer. Included in the message is a declaration that the Cisco NSF-capable or NSF-aware router has "graceful restart capability." If the BGP peer has received this capability, it is aware that the device sending the message is Cisco NSF-capable. Both the Cisco NSF-capable router and its BGP peers (NSF-aware peers) need to exchange the graceful restart capability in their OPEN messages at the time of session establishment. If both peers do not exchange the graceful restart capability, the session will not be capable of graceful restart.

If the BGP session is lost during the RP switchover, the Cisco NSF-aware BGP peer marks all the routes that are associated with the NSF-capable router as stale. However, it continues to use these routes to make forwarding decisions for a time. This functionality means that no packets are lost while the newly active RP is waiting for convergence of the routing information with the BGP peers.

After an RP switchover occurs, the Cisco NSF-capable router re-establishes the session with the BGP peer. In establishing the new session, it sends a new graceful restart message that identifies the Cisco NSF-capable router as having restarted.

At this point, the routing information is exchanged between the two BGP peers. Once this exchange is complete, the Cisco NSF-capable device uses the routing information to update the RIB and the FIB with the new forwarding information. The Cisco NSF-aware device uses the network information to remove stale routes from its BGP table. Once the stale routes are removed, the BGP protocol is fully converged.

If a BGP peer does not support the graceful restart capability, it will ignore the graceful restart capability in an OPEN message but will establish a BGP session with the Cisco NSF-capable device. This functionality will allow interoperability with non-NSF-aware BGP peers (and without NSF functionality), but the BGP session with non-NSF-aware BGP peers will not be capable of graceful restart.

BGP NSF awareness is not enabled by default. To globally enable it, use **bgp graceful-restart** router configuration command. To disable the BGP graceful restart capability globally for all BGP neighbors, use the **no** form of this command.

```
bgp graceful-restart [restart-time seconds] [stalepath-time seconds]
```

## Syntax Description

| Parameter                            | Description                                                                                                                                                                                                                                                                     |
|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>restart-time</b> <i>seconds</i>   | (Optional) The maximum time period that the local router will wait for a graceful-restart-capable neighbor to return to normal operation after a restart event occurs. The default value for this argument is 120 seconds. The valid range of values is from 1 to 3600 seconds. |
| <b>stalepath-time</b> <i>seconds</i> | (Optional) The maximum time period that the local router will hold stale paths for a restarting peer. All stale paths are deleted after this timer expires. The default value for this argument is 360 seconds. The valid range of values is from 1 to 3600 seconds             |

---

|             |                                                                                                                                     |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | NSF awareness is enabled automatically in supported software images for Interior Gateway Protocols, such as EIGRP, IS-IS, and OSPF. |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------|

---

# BGP Scan Time

[BGP](#) convergence can also be reduced by using certain BGP timers. BGP scan time is one of the timers that can be used to influence BGP convergence speed.

## BGP Scan Time

- Defines how often BGP scanner process scans the BGP table.
- If lowered, can improve convergence
- Needed to confirm that next hops are still available.
- The BGP scanner process is also responsible for advanced features such as conditional advertisement check and performing route dampening.
- Set to 60 seconds by default.

```
router(config-router) # bgp scan-time scanner-interval
```

- Changes the default value of BGP scanner process runs.

© 2015 Cisco and/or its affiliates. All rights reserved.

203

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. The router then propagate the changes 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 BGP should advertise conditional prefixes or perform route dampening.

To configure nondefault time interval for repetitions of the BGP scanner process, use **bgp scan-time** command.

```
bgp scan-time [import] scanner-interval
```

## Syntax Description

| Parameter               | Description                                                                                                                                                                                                   |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>import</b>           | (Optional) Configures import processing of <a href="#">VPNv4</a> unicast routing information from BGP routers into routing tables.                                                                            |
| <i>scanner-interval</i> | 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

You can check the configured BGP scan interval using the **show ip bgp summary** command.

### Monitoring BGP Scan Time

```
R1# show ip bgp summary
BGP router identifier 192.168.12.1, local AS number 1
BGP table version is 87, main routing table version 87
11 network entries using 1628 bytes of memory
16 path entries using 1024 bytes of memory
3/2 BGP path/bestpath attribute entries using 408 bytes of memory
1 BGP AS-PATH entries using 24 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
BGP using 3084 total bytes of memory
BGP activity 44/33 prefixes, 84/68 paths, scan interval 60 secs
<... output omitted ...>
```

- Scan interval is defined per BGP router process and address family.

© 2015 Cisco and/or its affiliates. All rights reserved.204

The configured BGP scan interval will apply to the entire BGP routing protocol process.

## BGP Next-Hop Tracking

BGP monitors the next hop of installed routes to verify next-hop reachability and to select, install, and validate the BGP best path. By default, the BGP scanner is used to poll the [RIB](#) for this information every 60 seconds. During the 60 second time period between scan cycles, [IGP](#) instability or other network failures can cause black holes and routing loops to temporarily form.

The BGP next-hop tracking feature prevents black hole routing during the 60 second period of scan-timer. Scan-timer is already mostly replaced in basic functions by the BGP [NHT](#).

## BGP Next-Hop Tracking

- The replacement for scan-timer.
  - Prevents black-hole routing during the scan-timer period.
- It is event driven.
- Automatically tracks BGP prefixes, and rapidly report next-hop changes to the BGP routing process.
- Enabled by default.

```
router(config-router)# bgp nexthop trigger [delay seconds | enable]
```

- Controls the BGP NGT feature:
  - Enables it, if it was previously disabled.
  - Changes the interval between checks.

© 2015 Cisco and/or its affiliates. All rights reserved.

205

BGP next-hop tracking is an event-driven system that provides faster convergence in on-demand fashion rather than periodic based on scan-timer. It monitors RIB for next-hop related changes for both [EBGP](#) and [IBGP](#) prefixes. It reports changes to the BGP routing process.

This optimization improves overall BGP convergence by reducing the response time to next-hop changes for routes installed in the RIB. When a best path calculation is run in between BGP scanner cycles, only next-hop changes are tracked and processed.

BGP NHT feature is enabled by default. Disabling next hop address tracking may be useful if your network has unstable IGP peers and route dampening is not resolving the stability issues. To control this feature, use the **bgp nexthop trigger** command in the address family configuration mode of the routing process.

```
bgp nexthop trigger { delay seconds | enable }
```

### Syntax Description

| Parameter            | Description                                                                                                                                                                                   |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>delay seconds</b> | Changes the delay interval between checks on updated next-hop routes that are installed in the routing table. Valid values for the delay are from 0 to 100 seconds. The default is 5 seconds. |
| <b>enable</b>        | Enables BGP next-hop tracking.                                                                                                                                                                |

---

|             |                                                                                                                      |
|-------------|----------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | The <b>bgp scan-time</b> command is ignored if your router has BGP next-hop tracking enabled for the address-family. |
|-------------|----------------------------------------------------------------------------------------------------------------------|

---

# BGP Advertisement Interval

[BGP](#) advertisement interval is another option to influence BGP convergence speed.

## BGP Advertisement Interval

- Defines a time which has to elapse between two successive updates about the same destination that are sent to a neighbor.
- If lowered, can improve convergence
- Default values are different for IBGP and EBGP neighbors:
  - 30 seconds for EBGP neighbors
  - 0 seconds for IBGP neighbors

```
router(config-router)# neighbor [ip-address | peer-group-name] advertisement-
interval seconds
```

- Changes the default time interval in the sending of BGP routing updates for a specific neighbor:

© 2015 Cisco and/or its affiliates. All rights reserved.206

Advertisement interval timer controls the rate at which successive advertisements are sent to a BGP neighbor. 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 time equal to the advertisement interval has elapsed. So, 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.

The default values are different for [IBGP](#) and [EBGP](#) peers. For IBGP peers, the interval is set to 0 seconds. For EBGP peers, the interval is set to 30 seconds. For EBGP peers that are configured in a [VRF](#) instance, the interval is also set to 0 seconds.

It is important to note that while the router is waiting for the advertisement interval timer to expire, the router can still receive and process route updates from BGP neighbors. Changing the advertisement interval 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.

You can modify the default advertisement interval for a specific BGP peer using the **neighbor advertisement-interval** router configuration command.

```
neighbor [ip-address | peer-group-name] advertisement-interval seconds
```

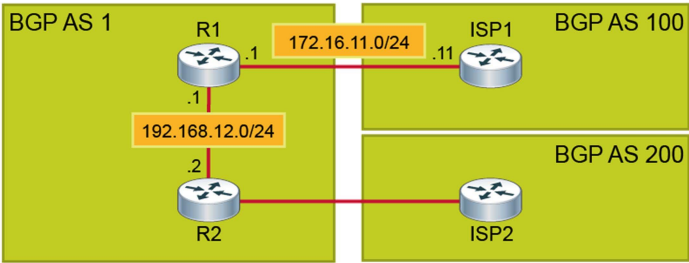
## Syntax Description

| Parameter              | Description                                                                                                                                                                                                  |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>      | Neighbor <a href="#">IP address</a> .                                                                                                                                                                        |
| <i>peer-group-name</i> | Name of a BGP peer group. If a BGP peer group is specified by using the <i>peer-group-name</i> argument, all members of the peer group will inherit the characteristic that is configured with this command. |
| <i>seconds</i>         | Time in seconds. The valid value is from 0 to 600 seconds. The default is 30 seconds for external peers and 5 seconds for internal peers.                                                                    |

**Note** 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.

## Monitoring the BGP Advertisement Interval

### Monitoring the BGP Advertisement Interval



```

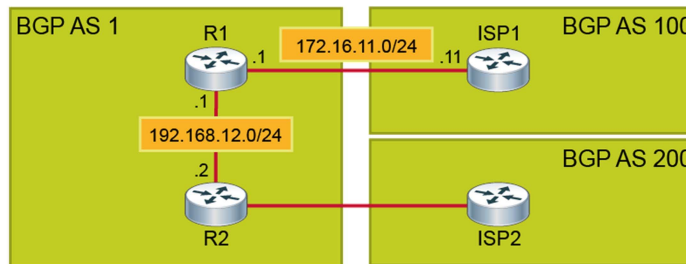
R1# show ip bgp neighbors 192.168.12.2
BGP neighbor is 192.168.12.2, remote AS 1, internal link
<... output omitted ...>
 Default minimum time between advertisement runs is 0 seconds
<... output omitted ...>

```

© 2015 Cisco and/or its affiliates. All rights reserved. 207



## Monitoring the BGP Advertisement Interval (Cont.)



```
R1# show ip bgp neighbors 172.16.11.11
BGP neighbor is 172.16.11.11, remote AS 100, external link
<... output omitted ...>
Default minimum time between advertisement runs is 30 seconds

For address family: IPv4 Unicast
 Session: 172.16.11.11
 BGP table version 102, neighbor version 102/0
<... output omitted ...>
 Minimum time between advertisement runs is 15 seconds
<... output omitted ...>
```

© 2015 Cisco and/or its affiliates. All rights reserved.

208

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 example shows the output for internal neighbor, where the default timer is 0 seconds. The second example shows the changed default timer for EBGP, 30 seconds. After the parameter was changed, the timer switched to 15 seconds.

# BGP Keepalive and Hold-Down Timers

[BGP](#) keepalive and hold-down timers are two BGP timers that influence BGP convergence as well.

## BGP Keepalive and Hold-Down Timers

BGP keepalive timer:

- Defines a time between successive keepalive messages.
- Set to 60 seconds by default.

BGP hold-down timer:

- Defines how long a router will wait from the last received keepalive or update message before declaring the session dead.
- Set to 3x keepalive = 180 seconds by default.

```
router(config-router)# neighbor [ip-address | peer-group-name] timers keepalive
holdtime [min-holdtime]
```

© 2015 Cisco and/or its affiliates. All rights reserved.

209

The keepalive timer defines how often a BGP router will probe a BGP neighbor by sending keepalive messages. The hold down timer defines how long a router will wait from the last received keepalive or update message before declaring the session dead.

Lowering the keepalive and hold-down timers will improve BGP convergence.

By default, the keepalive timer is set to 60 seconds, and the hold-down timer is set to 180 seconds. Even if the IP routing table indicates that the neighbor is no longer reachable, BGP will not trigger the convergence process until the BGP session hold-down timer expires.

To configure a nondefault value of the keepalive and/or hold-down timers, use the **neighbor timers** router configuration command.

```
neighbor [ip-address | peer-group-name] timers keepalive holdtime [min-holdtime]
```

## Syntax Description

| Parameter              | Description                                                                                                                                                                                                                                                            |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>      | Neighbor <a href="#">IP address</a> .                                                                                                                                                                                                                                  |
| <i>peer-group-name</i> | Name of a BGP peer group. If a BGP peer group is specified by using the <i>peer-group-name</i> argument, all members of the peer group will inherit the characteristic that is configured with this command.                                                           |
| <i>keepalive</i>       | Time in seconds with which the keepalive messages are sent to the peer. The valid value is from 0 to 65535 seconds. The default is 60 seconds.                                                                                                                         |
| <i>holdtime</i>        | Time in seconds after not receiving a keepalive message and declaring a peer dead. The valid value is from 0 to 65535 seconds. The default is 180 seconds.                                                                                                             |
| <i>min-holdtime</i>    | (Optional) Interval (in seconds) specifying the minimum acceptable hold-time from a BGP neighbor. The minimum acceptable hold-time must be less than, or equal to, the interval that is specified in the <i>holdtime</i> argument. The valid value is from 0 to 65535. |

---

|             |                                                                                                                                                                                                                           |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | New keepalives will take effect only after new BGP session establishment as they are transferred and negotiated to the lower values from the two peers in the OPEN messages. Therefore you need to reset the BGP session. |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

## Monitoring BGP Keepalive and Hold-Down Timers

### Monitoring BGP Keepalive and Hold-Down Timers

```
R1# show ip bgp neighbors
BGP neighbor is 172.16.11.11, remote AS 100, external link
 BGP version 4, remote router ID 10.0.1.81
 BGP state = Established, up for 06:54:31
 Last read 00:00:02, last write 00:00:22, hold time is 180, keepalive interval
 is 60 seconds
<... output omitted ...>
```

- Verifies keepalive and hold-down timers.

You can examine the currently configured BGP keepalive and hold-down timers for each neighbor with the **show ip bgp neighbors** command.

In the example, both timers are left on default: keepalive timer is 60 seconds and hold-down timer is 180 seconds.

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- BGP is considered converged when there are no outstanding updates to be sent to neighbors.
- Several features are provided in BGP to improve convergence time:
  - **PMTU discovery** benefits BGP as it benefits any transmission over IP.
  - **Input queue depth** can be increased to improve performance.
  - **PIC** enhances BGP convergence regardless of number of BGP prefixes.
  - **BFD** reduces convergence time by providing rapid detection of failed peers.
  - **NSF awareness** reduces the amount of resources by continue forwarding packets during a failure condition.
- You can also lower the BGP timers to influence BGP convergence:
  - Scan timer
  - Advertisement interval
  - Keepalive and hold-down timers

© 2015 Cisco and/or its affiliates. All rights reserved.

211



# Lesson 2: Limiting the Number of Prefixes Received from a BGP Neighbor

---

## Overview

There are currently more than 600,000 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 [ASs](#) that did not originate them.

[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. An advanced BGP configuration tool has been designed to improve BGP scalability and performance by reducing the number of prefixes that a router receives from a BGP neighbor.

Upon completing this lesson, you will be able to:

- 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 route limiting
- Describe how to configure and monitor BGP maximum-prefix function

# BGP Route Limiting

Incoming route policies, which are applied on [BGP](#) routers, indicate the BGP attribute values that a route should have in order to be accepted. Route policies 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 filtering used in route policies is the one that matches the contents of the [AS](#) path attribute.

## BGP Route Limiting

- All 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).
- BGP maximum prefixes limiting is used to establish a hard limit on the number of prefixes that are received from a neighbor.

© 2015 Cisco and/or its affiliates. All rights reserved.

212

An [ISP](#) with a multihomed customer may use route policies 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.

If you configure route policies in a customer router improperly, it may accidentally cause the customer to receive many Internet routes. Even worse, a faulty configuration may cause customer to advertise prefixes 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. The AS would contain 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. As a result, it would direct much of the provider traffic that is intended for the Internet to the customer network.

A route policy that is based on an AS path filter in the ISP router will not prevent this accident. The routes that the customer sends have the anticipated AS path value. A route policy that is based on customer prefixes and that distinctly identifies and permits each of the network numbers that the customer may advertise will prevent the accident. However, such a routing policy is difficult to maintain.



## BGP Route Limiting (Cont.)

BGP router terminates peering when a number of maximum prefixes exceeded.

You can configure a router to:

- Generate a logging message when a specified percentage of the maximum prefixes is reached.
- Reestablish BGP peering after a specified time (from 1 to 65535 minutes).
- Generate a logging message when the maximum prefix limit is exceeded, instead of terminating BGP peering.

© 2015 Cisco and/or its affiliates. All rights reserved.

213

A scalable solution to the need for limiting the number of routes (prefixes) that are received from a BGP neighbor is to use BGP maximum prefixes limiting. This feature enables you to specify how many routes a router can receive from a neighbor.

When the number of prefixes that is received from the peer for a given address family exceeds the maximum limit, a cease notification message is sent to the neighbor. Also the peering with the neighbor is terminated

You can also configure the router to generate a logging message when a configured percentage of the maximum prefixes is reached. You can also configure the router to automatically reestablish a peering session that has been disabled because the maximum prefix limit has been exceeded. Instead of ceasing the peering with a neighbor, you can configure the router to generate and display a warning when a configured percentage of the maximum prefixes has been reached.

# Configuring the BGP Route Limiting

The [BGP](#) route limiting can be configured with the use of maximum-prefix function.

## 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

214

To control how many prefixes 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]
```

## Syntax Description

| Parameter              | Description                                                                                                                                                                                                                            |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>      | IP address of the neighbor.                                                                                                                                                                                                            |
| <i>peer-group-name</i> | Name of a BGP peer group.                                                                                                                                                                                                              |
| <i>maximum</i>         | Maximum number of prefixes that are allowed from this neighbor.                                                                                                                                                                        |
| <i>threshold</i>       | (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.                                                                 |
| <b>warning-only</b>    | (Optional) Allows the router to generate a log message when the maximum is exceeded, instead of terminating the peering.                                                                                                               |
| <b>restart</b>         | (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 65535 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.

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 you to configure the time interval after which a router re-establishes a peering session 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.

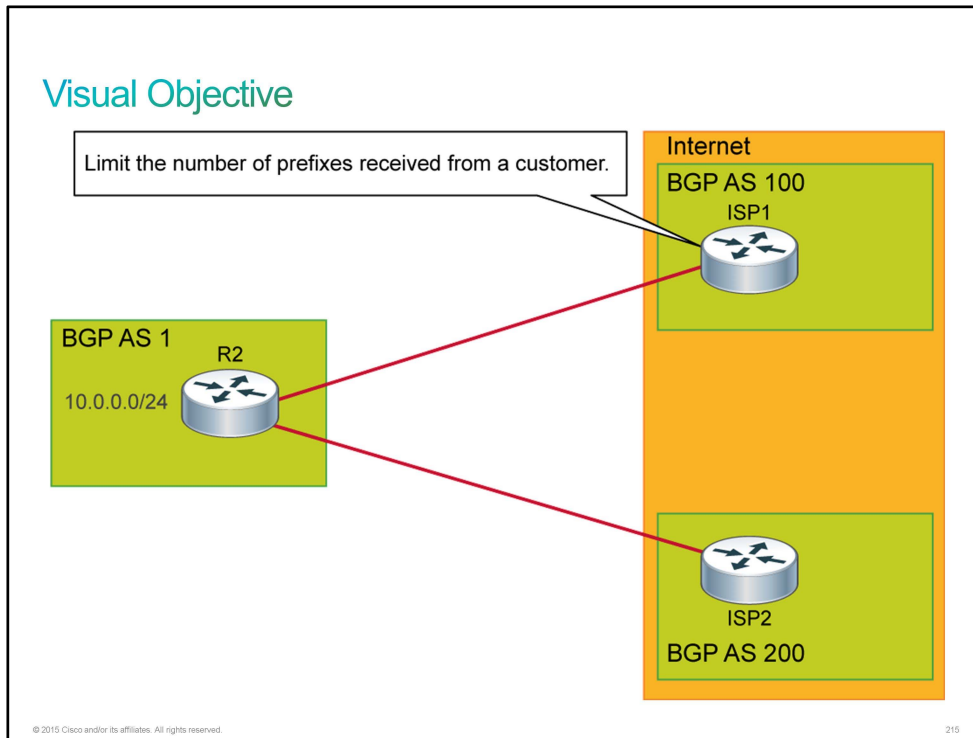
This feature attempts to re-establish a disabled peering session at the time interval that you configure. However, the configuration of the restart timer alone cannot change or correct a peer that is sending an excessive number of prefixes. You will need to reconfigure the maximum-prefix limit or reduce 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 you correct the underlying problem.

|             |                                                                                                                                                                                                                                                                                                                                     |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | You can use the <b>bgp dampening</b> 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 <b>restart</b> command only when you are troubleshooting or tuning a router that is sending an excessive number of prefixes. |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

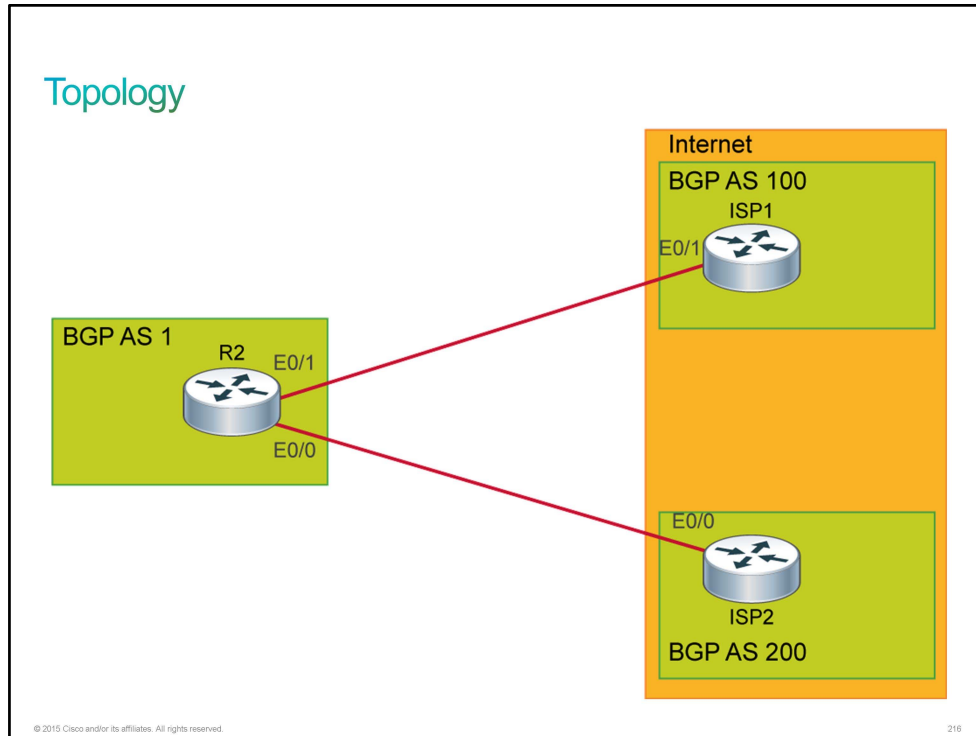
# Discovery 18: Configure BGP Route Limiting

## Overview

Through this discovery, you will learn how to configure [BGP](#) route limiting using the maximum-prefix function. You will configure a service provider's router, ISP1, to limit the maximum number of prefixes that are received from a multihomed customer's router, R2.



# Topology



## Job Aids

---

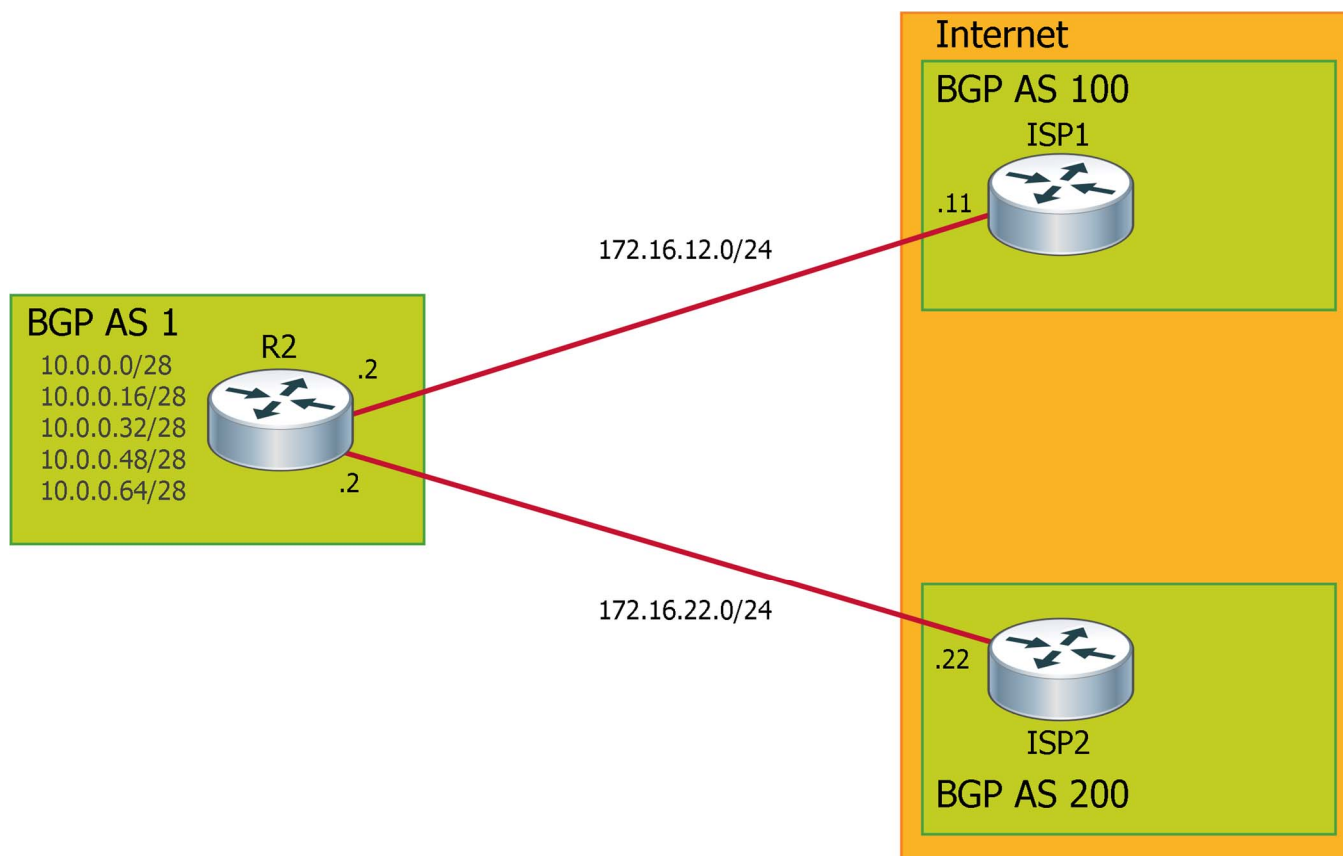
**Note** If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, IOL behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

---

## Device Information

| Device | Interface                                                                                                                  | IP address                                                                                                                                    | Description                     |
|--------|----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| R2     | Ethernet 0/0                                                                                                               | 172.16.22.2/24                                                                                                                                | Connection to ISP2              |
| R2     | Ethernet 0/1                                                                                                               | 172.16.12.2/24                                                                                                                                | Connection to ISP1              |
| R2     | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4<br>Loopback 5<br>Loopback 6<br>Loopback 7<br>Loopback 8<br>Loopback 9 | 10.0.0.1/28<br>10.0.0.17/28<br>10.0.0.33/28<br>10.0.0.49/28<br>10.0.0.65/28<br>10.0.0.81/28<br>10.0.0.97/28<br>10.0.0.113/28<br>10.0.0.129/28 | Loopbacks simulate LAN networks |
| ISP1   | Ethernet 0/1                                                                                                               | 172.16.12.11/24                                                                                                                               | Connection to R2                |
| ISP1   | Ethernet 0/2                                                                                                               | 172.16.100.11/24                                                                                                                              | Connection to ISP2              |
| ISP2   | Ethernet 0/0                                                                                                               | 172.16.22.22/24                                                                                                                               | Connection to R2                |
| ISP2   | Ethernet 0/2                                                                                                               | 172.16.100.22/24                                                                                                                              | Connection to ISP1              |

[IP addresses](#) and advertised networks in BGP are preconfigured as shown below:



BGP is also preconfigured as [EBGP](#):

- R2 to ISP1
- R2 to ISP2

## Configure BGP Route Limiting

### Discovery Steps

**Step 1** On the ISP1 and ISP2 routers, verify initial state of BGP table. Verify that you received routes to four destination networks from R2.

Use the **show ip bgp** command.

```
ISP1# show ip bgp
BGP table version is 5, local router ID is 172.16.12.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network         | Next Hop    | Metric | LocPrf | Weight | Path |
|-----------------|-------------|--------|--------|--------|------|
| *> 10.0.0.0/28  | 172.16.12.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.16/28 | 172.16.12.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.32/28 | 172.16.12.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.48/28 | 172.16.12.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.64/28 | 172.16.12.2 | 0      |        | 0      | 1 i  |

```
ISP2# show ip bgp
BGP table version is 5, local router ID is 172.16.22.22
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network         | Next Hop    | Metric | LocPrf | Weight | Path |
|-----------------|-------------|--------|--------|--------|------|
| *> 10.0.0.0/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.16/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.32/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.48/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.64/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |

Both, ISP1 and ISP2, have received routes to five different networks—10.0.0.0/28, 10.0.0.16/28, 10.0.0.32/28, 10.0.0.48/28, and 10.0.0.64/24.

**Step 2** On the ISP1 router, limit the number of received prefixes from R2.

ISP1 should receive no more than 8 prefixes from R2, and should start generating warning messages when the number of received prefixes exceeds 75% of the maximum specified.

```
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor 172.16.12.2 maximum-prefix 8 75
```

**Step 3** On the R2 router, advertise the 10.0.0.80/28 and 10.0.0.96/28 networks in BGP.

To advertise networks, configure the following commands on R2.

**NOTE:** Before configuring R2, make sure that you also have a console window open for ISP1 router. So you will be able to see the generated warning messages.

```
R2(config)# router bgp 1
R2(config-router)# network 10.0.0.80 mask 255.255.255.240
R2(config-router)# network 10.0.0.96 mask 255.255.255.240
```

**Step 4** On the ISP1 router, observe the warning message generated.

```
ISP1#
*Apr 2 12:07:55.964: %BGP-4-MAXPFX: Number of prefixes received from 172.16.12.2 (afi
0) reaches 7, max 8
```

The warning message was generated, because ISP1 received 7 prefixes from R2. The maximum is set to 8, so the threshold of 75% was exceeded.

**Step 5** On the ISP1 router, verify that you still have all destination, that R2 advertises, in the BGP table.

```
ISP1# show ip bgp
BGP table version is 8, local router ID is 172.16.12.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network         | Next Hop    | Metric | LocPrf | Weight | Path |
|-----------------|-------------|--------|--------|--------|------|
| *> 10.0.0.0/28  | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.16/28 | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.32/28 | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.48/28 | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.64/28 | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.80/28 | 172.16.12.2 | 0      |        | 0 1    | i    |
| *> 10.0.0.96/28 | 172.16.12.2 | 0      |        | 0 1    | i    |

Although the warning was generated, ISP1 still has all 7 networks that are received from R2 in its BGP table.

**Step 6** On the R2 router, also advertise the 10.0.0.112/28 and 10.0.0.128/28 networks in BGP.

To advertise networks, configure the following commands on R2.

**NOTE:** Before configuring R2, make sure that you also have a console window open for ISP1 router. So you will be able to see the generated messages.

```
R2(config)# router bgp 1
R2(config-router)# network 10.0.0.112 mask 255.255.255.240
R2(config-router)# network 10.0.0.128 mask 255.255.255.240
```

**Step 7** On the ISP1 router, observe the generated messages.



```

ISP1#
*Apr 2 12:22:01.859: %BGP-4-MAXPFX: Number of prefixes received from 172.16.12.2 (afi
0) reaches 8, max 8
ISP1#
*Apr 2 12:22:32.575: %BGP-3-MAXPFXEXCEED: Number of prefixes received from 172.16.12.2
(afi 0): 9 exceeds limit 8
*Apr 2 12:22:32.575: %BGP-5-ADJCHANGE: neighbor 172.16.12.2 Down BGP Notification sent
*Apr 2 12:22:32.575: %BGP-3-NOTIFICATION: sent to neighbor 172.16.12.2 3/1 (update
malformed) 0 bytes
*Apr 2 12:22:32.575: %BGP-4-MSGDUMP: unsupported or mal-formatted message received
from 172.16.12.2:
FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF 0037 0200 0000 1B40 0101 0040 0206 0201
0000 0001 4003 04AC 100C 0280 0404 0000 0000 1C0A 0000 80
ISP1#
*Apr 2 12:22:32.575: %BGP_SESSION-5-ADJCHANGE: neighbor 172.16.12.2 IPv4 Unicast
topology base removed from session BGP Notification sent
ISP1#
*Apr 2 12:23:41.141: %BGP-3-NOTIFICATION: sent to neighbor 172.16.12.2 passive 2/8 (no
supported AFI/SAFI) 3 bytes 000101 (timer expired)

```

The beginning of the output first shows that the total number of received prefixes has reached 8, which is over the threshold to generate a warning message. When you started to advertise the second network in BGP, the second message was generated. The total number of received prefixes is now 9, which is above the configured limit (8). As a result, the session with the neighbor, R2, was terminated.

**Step 8** On the ISP1 and ISP2 routers, verify the state of BGP table.

```

ISP1# show ip bgp
ISP1#

```

```

ISP2# show ip bgp
BGP table version is 10, local router ID is 172.16.22.22
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

```

| Network          | Next Hop    | Metric | LocPrf | Weight | Path |
|------------------|-------------|--------|--------|--------|------|
| *> 10.0.0.0/28   | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.16/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.32/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.48/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.64/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.80/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.96/28  | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.112/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |
| *> 10.0.0.128/28 | 172.16.22.2 | 0      |        | 0      | 1 i  |

ISP1 has no routes in the BGP table, since R2 was the only neighbor and the session was terminated due to exceeded maximum number of received prefixes. You did not configure route limiting on ISP2, so you can see routes for all nine destinations in its BGP table.

## Monitoring the BGP Route Limiting

## Monitoring the BGP Route Limiting

```
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.

© 2015 Cisco and/or its affiliates. All rights reserved.217

You can 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.

If the peer exceeded the configured maximum number of prefixes, the output of the **show ip bgp neighbors** command shows the reason for resetting the session. 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, you must issue the **clear ip bgp ip-address** command for the neighbor. Except, if you have specified the restart keyword in the configuration. In this case, the router tries to re-establish the BGP session automatically after the expiration of the configured restart timeout interval.

**Step 9** On the ISP1 router, verify information about the R2 neighbor.

To verify information about the neighbor, use the **show ip bgp neighbors** command with a neighbor IP address as an argument.

```
ISP1# show ip bgp neighbor 172.16.12.2
BGP neighbor is 172.16.12.2, remote AS 1, external link
 BGP version 4, remote router ID 10.0.0.129
 BGP state = Idle
 Neighbor sessions:
 0 active, is not multiseession capable (disabled)
 Stateful switchover support enabled: NO
 Default minimum time between advertisement runs is 30 seconds

For address family: IPv4 Unicast
 BGP table version 17, neighbor version 1/17
 Output queue size : 0
 Index 0, Advertise bit 0
 Address family not supported notification sent
 Slow-peer detection is disabled
 Slow-peer split-update-group dynamic is disabled
 Peer had exceeded the max. no. of prefixes configured.
 Maximum prefixes allowed 8
 Threshold for warning message 75%
 Reduce the no. of prefix and clear ip bgp 172.16.12.2 to restore peering
<... output omitted ...>
```

The state of BGP session with R2 router is Idle. The reason is that the peer had exceeded the maximum number of prefixes configured. You can see that the maximum allowed prefixes for this neighbor are 8 and the threshold for warning messages is set to 75%. These are the values that you configured earlier.

To force the neighbor from the Idle state into the Active state and to re-establish the BGP session, you have to reduce the number of prefixes and clear the BGP session with the neighbor.

# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- An improperly configured filter in a customer router may accidentally cause the customer router to receive many Internet routes.
- Maximum prefix limitation will shut down peering session if too many prefixes are received.
  - Use **neighbor maximum-prefix** command to configure route limiting—a maximum number of prefixes that a BGP router is allowed to receive from a peer.

# Lesson 3: Implementing BGP Peer Groups

---

## Overview

Scaling routers to meet the demands of full Internet routing and associated administrative policies requires protocols like [BGP](#) with embedded scalability mechanisms. In environments where network administrators must configure many 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. These policies can be outbound route maps, distribute lists, filter lists, update source, and so on. You can group neighbors with the same update policies into peer groups to simplify configuration and, more importantly, to make BGP updates more efficient. In this lesson, you will learn about peer groups as a BGP scalability mechanism.

Upon completing this lesson, you will be able to:

- Describe the need for BGP peer groups
- Describe the performance benefits of using BGP peer groups
- Describe the limitations of BGP peer groups
- Identify the Cisco IOS commands that are required to configure BGP peer groups
- Describe the configuration of BGP peer groups
- Explain the BGP peer group configuration examples
- Describe BGP dynamic update groups
- Describe BGP peer templates
- Describe the BGP peer template inheritance
- Identify the Cisco IOS commands that are required to configure BGP peer templates

# BGP Peer Groups Overview

In many cases, you as a network administrator must configure a single router with many neighbors, each neighbor having parameters similar to the others. This situation may cause time-consuming configuration work, because then you have to configure almost the same parameters for all of the neighbors. Imagine you have a service provider network that has an edge router with many customers attached to it, where each customer requires its own [BGP](#) session. You may find that all of the BGP sessions to its customer routers have almost identical configurations.

## BGP Peer Groups Overview

- BGP routers could have many 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 that are specified for the BGP neighbors are identical, with a few exceptions.

**Solution** is to group common parameters in a BGP peer group.

© 2015 Cisco and/or its affiliates. All rights reserved.219

Likewise, [IBGP](#) sessions are almost always identically configured. If a full mesh is deployed within an [AS](#), many 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 lead to a tremendous amount of redundant configuration.

To ease the burden of configuring many neighbors with identical or similar parameters (for example, route maps, filter lists, or prefix lists), the concept of peer groups was introduced. You as a network administrator can configure a template, or 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 you apply 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.

## BGP Peer Groups Overview (Cont.)

- BGP peer group creates a neighbor parameter template.
- Configurable parameters include the following:
  - Community propagation
  - Update-source and next-hop-self
  - EBGp multihop
  - Authentication password
  - Neighbor weight
  - Prefix lists, filter lists, and route maps
- Individual parameters that are specified in a peer group can be overridden on a neighbor-by-neighbor basis.

© 2015 Cisco and/or its affiliates. All rights reserved.

220

On a router, the peer group is created as a template. The template can be, among others, configured to do the following:

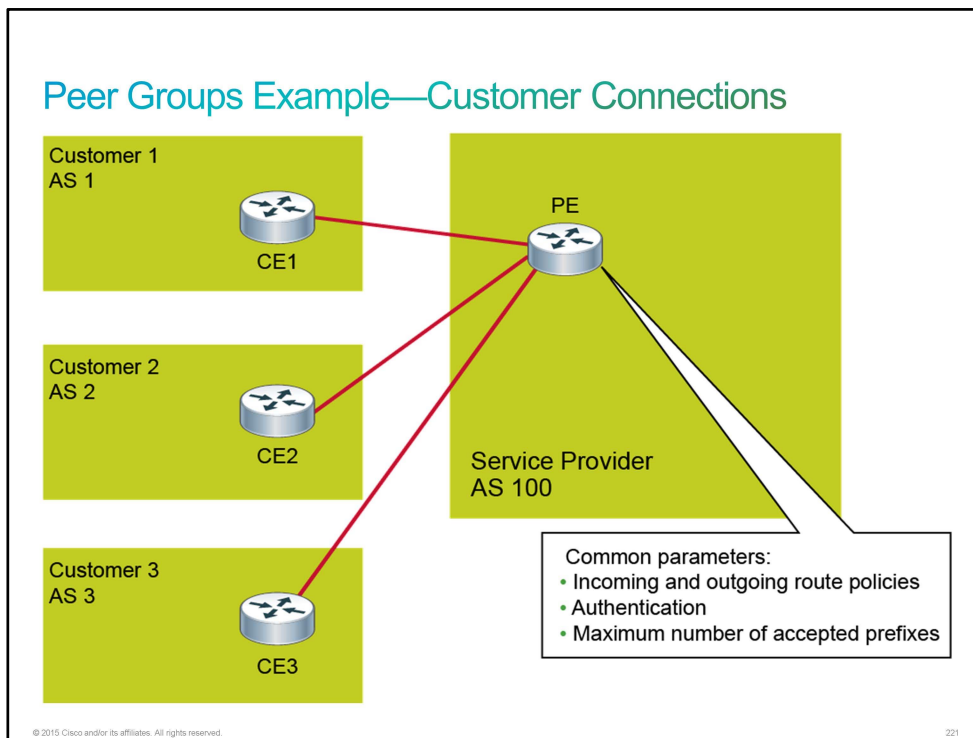
- Propagate, or not propagate, the community attribute.
- Use the [IP address](#) of a specific interface as the source address when opening the [TCP](#) session or use the next-hop-self feature.
- Use, or not use, the [EBGP](#) multihop function.
- Use, or not use, [MD5](#) authentication on the BGP sessions.
- Filter out any incoming or outgoing routes using a prefix list, a filter list, and a route map.
- Assign a particular weight value to the routes that are received.

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 routing policies 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 that are 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.

## Peer Groups Example—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.



The figure shows an example where peer groups are useful. The customer autonomous systems are all assumed to announce local networks only. All customer ASs should receive BGP updates with the same set of Internet routes, and the customer ASs are all assumed to generate only a few 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.

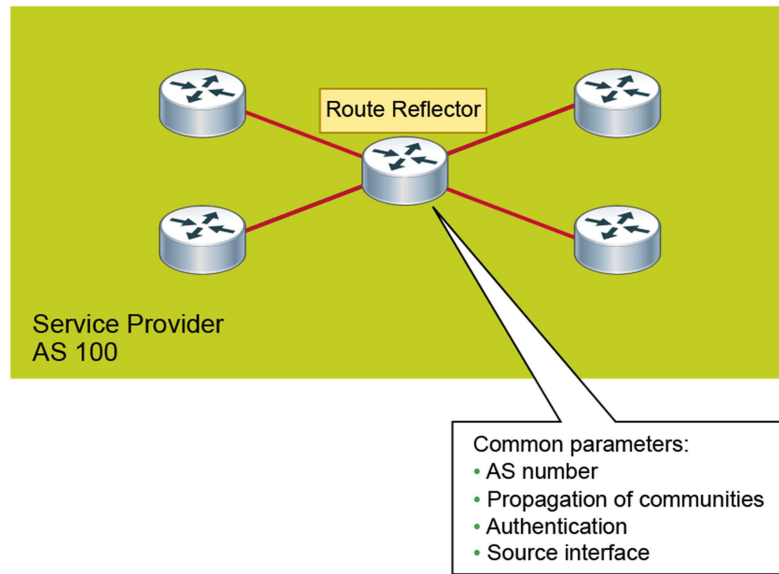
In this scenario, the use of the peer group function is highly desirable. You as a network administrator can configure BGP neighbors in the customer ASs using a single peer group. You configure the peer group template with references to route policies, authentication settings, and the maximum number 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 neighbors.

## Peer Group Example—BGP Route Reflector

This example illustrates the BGP route reflector.



## Peer Group Example—BGP Route Reflector



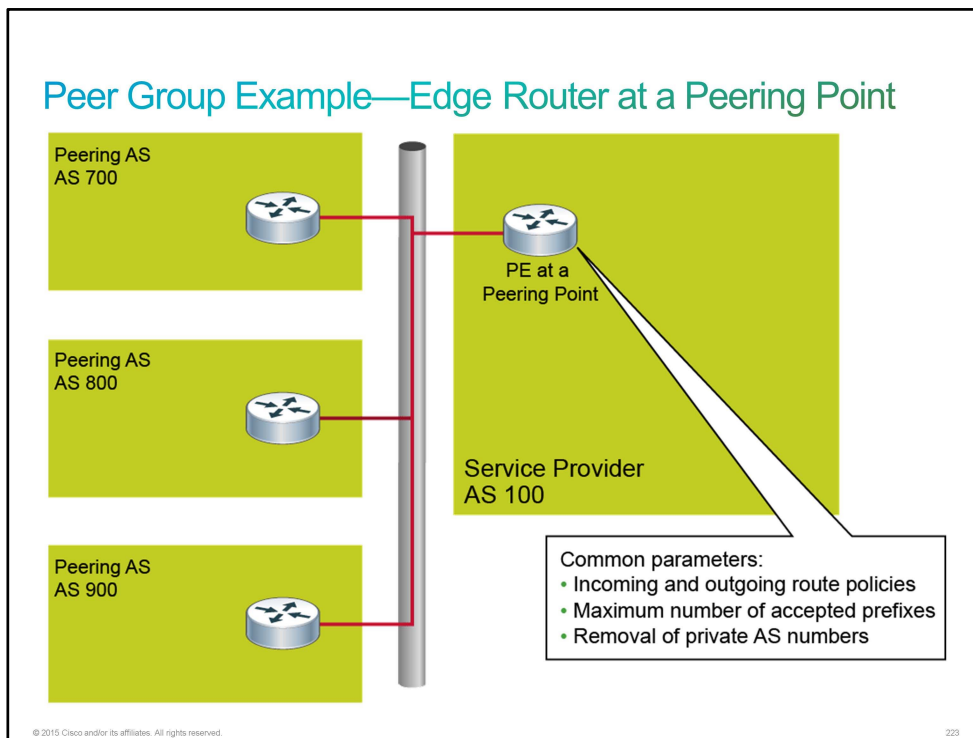
Another example of using peer groups is within the entire local AS, on the route reflectors, where every IBGP session is configured identically. If a router in the AS is supplied with some information, then all the routers should be supplied with the same information. Otherwise, an inconsistent routing 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. You can configure a peer group template with the required parameters. You can configure the neighbor AS number, enable of the **send-community** option, set 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.

## Peer Group Example—Edge Router at a Peering Point

This example illustrates an edge router at a peering point.



Another example of using peer groups is on the edge router that is located in the network where the service provider exchanges routes with other service providers. From the edge router, the service provider AS can peer with many other service providers.

All peering ASs should receive the same set of routes, namely the routes local to the service provider AS and the routes that are received from customer ASs. Also, all routes that are received by the service provider peering router from all peering ASs 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 [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, you can apply these parameters on the peer group template before the actual IP addresses of the neighbors are bound to the peer group.

# BGP Peer Groups as a Performance Tool

As described here, peer groups can be used to combine common [BGP](#) configuration into a peer group. Neighbors are then configured by assigning them to the peer group. By default, router 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 various outgoing filtering mechanisms (filter lists, route maps, and prefix lists). These tasks mean that when a router is configured with a large number of neighbors, the CPU load grows proportionally.

## BGP Peer Groups as a Performance Tool

By default, router builds individual BGP updates for each BGP neighbor.

Common configuration can be combined into a peer group.

- Neighbors are configured by assigning them to the peer group.
- A single BGP update is then built for all members of a BGP peer group.
  - The CPU load does not increase linearly with the increased number of neighbors.
  - Use peer groups wherever possible to reduce the CPU load of the BGP process.

© 2015 Cisco and/or its affiliates. All rights reserved.

224

However, with the use of peer groups, some of the router CPU utilization that is imposed by BGP update generation is significantly reduced. 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.

Router CPU load does increase when there are more neighbors of a router, due of increased TCP workload, but the use of peer groups can significantly reduce the increase. Therefore, you 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.

---

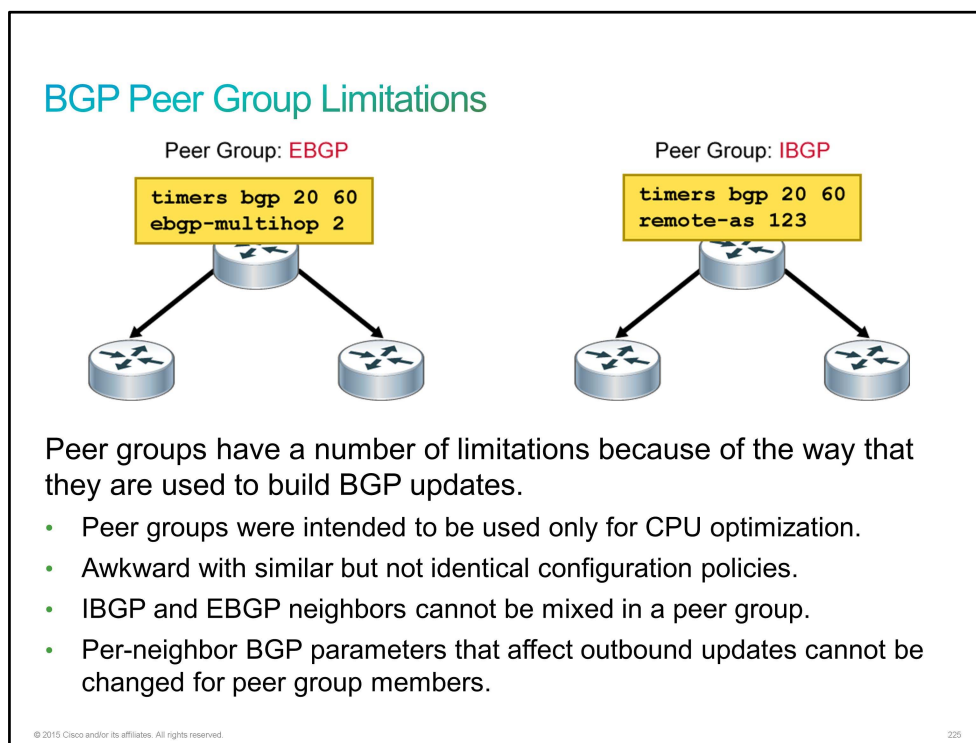
---

**Note**      Routers built BGP updates for each neighbor individually before the introduction of dynamic update groups, which were introduced in Cisco IOS Release 12.0(24)S. Dynamic update groups, which are enabled by default, allow routers to dynamically calculate and optimize updates that are sent to neighbor that share the same outbound policies. Although this new feature effectively made manually defined peer groups obsolete, many administrators still use peer groups as they offer convenience during configuration.

---

# BGP Peer Group Limitations

[BGP](#) peer groups were introduced primarily for [CPU](#) usage optimization. Because the router builds only one update for all members of the same peer group, some restrictions apply to members of the peer group.



BGP parameters and outbound routing policies cannot differ between group members because doing so could cause different updates to be sent to two members of the same peer group. Router creates only one update, which is then replicated to all members. Therefore, peer groups are awkward when used on a router with several neighbors that have similar but not completely identical policies.

Also [EBGP](#) and [IBGP](#) updates are very different. The same peer group cannot be used for EBGP and IBGP neighbors, even though a part of the configuration is the same for both types of neighbors. An example would be if you wanted to change the BGP keepalive and hold-down timers for both IBGP and EBGP sessions. Using peer groups, you must specify the timers in both peer groups (IBGP and EBGP), because they use different configurations due to the different nature of IBGP and EBGP sessions. Also EBGP updates have the [AS](#) path attribute changed. IBGP sessions pass only the local preference attribute. The [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 common BGP parameters must be replicated across all peer groups, and there is no way to use only one peer group for common configuration.

# Configuring BGP Peer Groups

To finish [BGP](#) peer group configuration, you first have to create a peer group, specify parameter for the peer group, and then assign neighbors into the peer group.

## Configuring BGP Peer Groups

```
router(config-router)# neighbor group-name peer-group
```

- Creates a BGP peer group (peer groups names are case-sensitive) .

```
router(config-router)# neighbor group-name any-BGP-parameter
```

- Specifies any BGP parameter for the peer group.

```
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 that are specified for the peer group.

```
router(config-router)# neighbor ip-address any-BGP-parameter
```

- Overrides a BGP parameter that is specified for the peer group with a neighbor-specific parameter.

© 2015 Cisco and/or its affiliates. All rights reserved.226

To configure BGP peer groups on Cisco IOS routers, perform the following steps:

### 1. Create a BGP peer group.

To create a BGP peer group, use the **neighbor peer-group** router configuration command.

```
neighbor peer-group-name peer-group
```

### 2. Specify parameters for the BGP peer group.

After you have created a peer group, 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.

### 3. Create a BGP neighbor.

### 4. Assign this neighbor into the peer group.

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
```

## Syntax Description

| Parameter              | Description                                                                                                 |
|------------------------|-------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>      | <a href="#">IP address</a> of the BGP neighbor that belongs to the peer group that is specified by the tag. |
| <i>peer-group-name</i> | Name of the BGP peer group to which this neighbor belongs.                                                  |

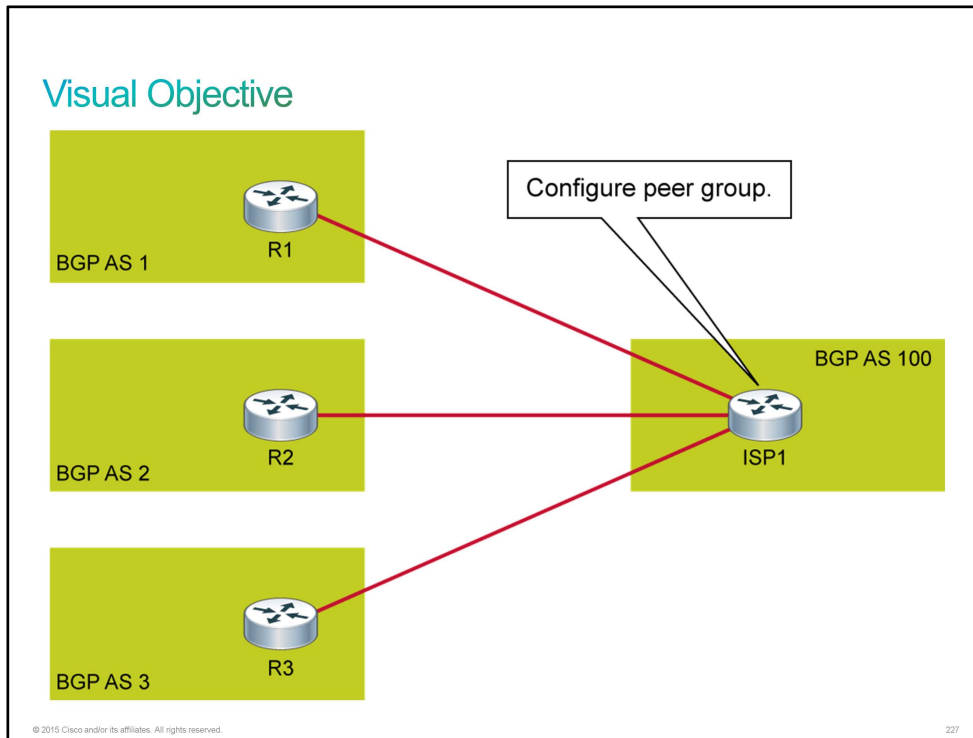
After you have assigned an actual neighbor to be a member of the peer group, all configurations that are made to the peer group template are then applied to all the neighbors that are assigned to that peer group. Through configuration, peer group configurations may be overridden for an individual neighbor, as long as the changes apply only to incoming updates. Remember that outgoing updates are prepared only once and then replicated.

|             |                                                                                                                                                                                                                                                                                                                                                                                |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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 peer group, then that neighbor must be removed from the peer group and configured as an individual neighbor. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

# Discovery 19: Configure BGP Peer Groups

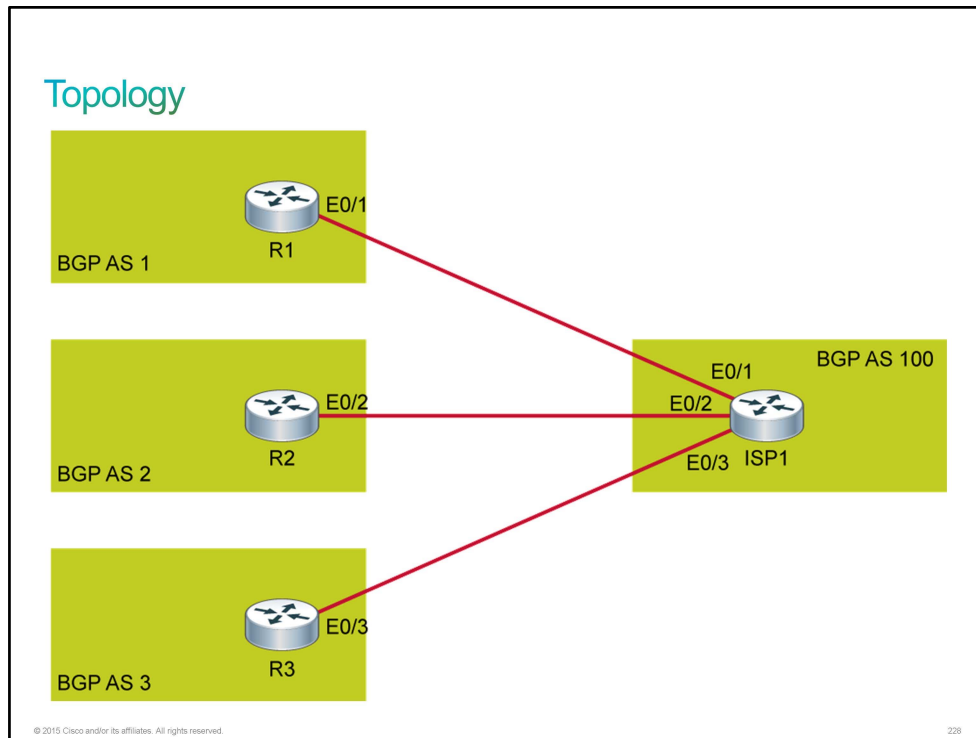
## Overview

Through this discovery, you will learn how to configure [BGP](#) peer group. A peer group will be used for all customer of the service provider because they share identical routing policy. You will put all neighbor configuration in the peer group configuration and the peers will then inherit the peer group configuration.





# Topology



## Job Aids

---

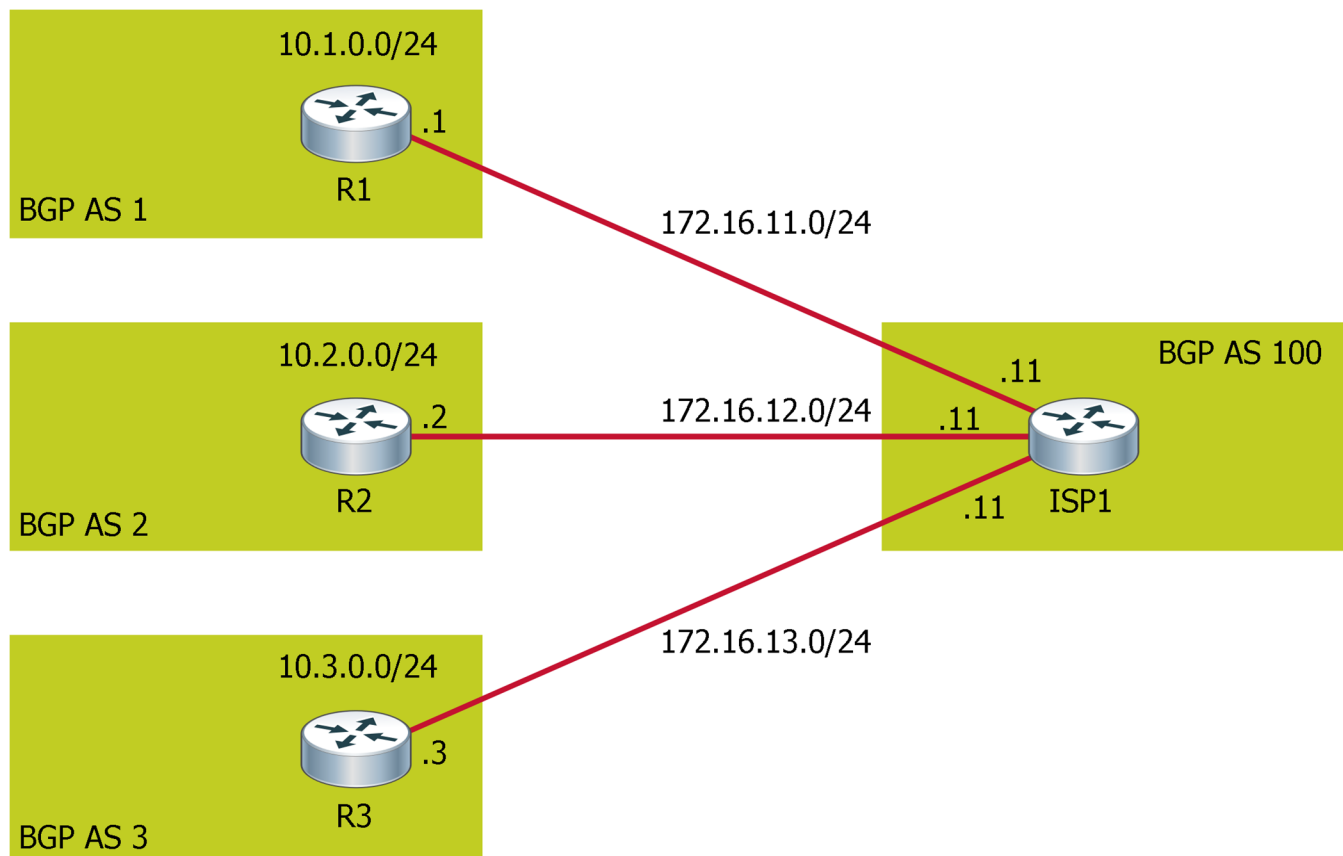
**Note** If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

---

## Device Information

| Device | Interface    | IP address      | Description                                    |
|--------|--------------|-----------------|------------------------------------------------|
| ISP1   | Ethernet 0/1 | 172.16.11.11/24 | Connection to R1                               |
| ISP1   | Ethernet 0/2 | 172.16.12.11/24 | Connection to R2                               |
| ISP1   | Ethernet 0/3 | 172.16.13.11/24 | Connection to R3                               |
| R1     | Ethernet 0/1 | 172.16.11.1/24  | Connection to ISP1                             |
| R1     | Loopback 1   | 10.1.0.1/24     | Loopback simulates <a href="#">LAN</a> network |
| R2     | Ethernet 0/2 | 172.16.12.2/24  | Connection to ISP1                             |
| R2     | Loopback 1   | 10.2.0.1/24     | Loopback simulates LAN network                 |
| R3     | Ethernet 0/3 | 172.16.13.3/24  | Connection to ISP1                             |
| R3     | Loopback 1   | 10.3.0.1/24     | Loopback simulates LAN network                 |

[IP addresses](#) and advertised networks in BGP are preconfigured as shown below:



BGP is also preconfigured as [EBGP](#):

- ISP1 to R1
- ISP1 to R2
- ISP1 to R3

## Configure BGP Peer Groups

### Discovery Steps

**Step 1** On the ISP1 router, verify that you receive networks, that all three customers advertise.

```
ISP1# show ip bgp
BGP table version is 4, local router ID is 172.16.13.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network     | Next Hop    | Metric | LocPrf | Weight | Path |
|----|-------------|-------------|--------|--------|--------|------|
| *> | 10.1.0.0/24 | 172.16.11.1 | 0      | 150    | 0      | 1 i  |
| *> | 10.2.0.0/24 | 172.16.12.2 | 0      | 150    | 0      | 2 i  |
| *> | 10.3.0.0/24 | 172.16.13.3 | 0      | 150    | 0      | 3 i  |

You can see that ISP1 router in AS 100 received three networks, each from different customer:

- 10.1.0.0/24 network originates in the first customer's [AS 1](#).
- 10.2.0.0/24 network originates in the second customer's AS 2.
- 10.3.0.0/24 network originates in the third customer's AS 3.

Note that all three networks have local preference set to 150. Route map is used to set the local preference for all three neighbors.

```
ISP1# show route-map
route-map Filter, permit, sequence 10
Match clauses:
Set clauses:
 local-preference 150
Policy routing matches: 0 packets, 0 bytes
```

**Step 2** On the ISP1 router, verify the BGP configuration.

```
ISP1# show running-config | section bgp
router bgp 100
 bgp log-neighbor-changes
 neighbor 172.16.11.1 remote-as 1
 neighbor 172.16.11.1 prefix-list Customer in
 neighbor 172.16.11.1 route-map Filter in
 neighbor 172.16.11.1 filter-list 10 in
 neighbor 172.16.12.2 remote-as 2
 neighbor 172.16.12.2 prefix-list Customer in
 neighbor 172.16.12.2 route-map Filter in
 neighbor 172.16.12.2 filter-list 10 in
 neighbor 172.16.13.3 remote-as 3
 neighbor 172.16.13.3 prefix-list Customer in
 neighbor 172.16.13.3 route-map Filter in
 neighbor 172.16.13.3 filter-list 10 in
```

All three neighbors are configured with the same parameters. Instead of configuring the same configuration three times, you can put all the configuration into the peer group configuration and assign all the neighbor to this peer group.

**Step 3** On the ISP1 router, create a peer group CUSTOMERS.

```
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor CUSTOMERS peer-group
```

**Step 4** On the ISP1 router, move all per-neighbor BGP configuration into peer group configuration.

Remove the filter list 10, prefix list Customer, and route map Filter from all three neighbors. Apply this three routing policies to the CUSTOMERS peer group

```
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor CUSTOMERS prefix-list Customer in
ISP1(config-router)# neighbor CUSTOMERS route-map Filter in
ISP1(config-router)# neighbor CUSTOMERS filter-list 10 in
ISP1(config-router)# no neighbor 172.16.11.1 prefix-list Customer in
ISP1(config-router)# no neighbor 172.16.11.1 route-map Filter in
ISP1(config-router)# no neighbor 172.16.11.1 filter-list 10 in
ISP1(config-router)# no neighbor 172.16.12.2 prefix-list Customer in
ISP1(config-router)# no neighbor 172.16.12.2 route-map Filter in
ISP1(config-router)# no neighbor 172.16.12.2 filter-list 10 in
ISP1(config-router)# no neighbor 172.16.13.3 prefix-list Customer in
ISP1(config-router)# no neighbor 172.16.13.3 route-map Filter in
ISP1(config-router)# no neighbor 172.16.13.3 filter-list 10 in
```

Applying the routing policies only to the peer group simplifies the BGP configuration and improves BGP scalability. Instead of configuring the same parameters for each neighbor, you now configure them only once and then assign as many neighbor to the peer group as you want.

**Step 5** On the ISP1 router, assign all three neighbors to the CUSTOMERS peer group.

```
ISP1(config)# router bgp 100
ISP1(config-router)# neighbor 172.16.11.1 peer-group CUSTOMERS
ISP1(config-router)# neighbor 172.16.12.2 peer-group CUSTOMERS
ISP1(config-router)# neighbor 172.16.13.3 peer-group CUSTOMERS
```

This configuration enables the neighbors that are assigned to the peer group, to inherit common

configuration from the peer group. You do not need to configure each individual neighbor.

## Monitoring 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

© 2015 Cisco and/or its affiliates. All rights reserved.

229

To display information about BGP peer groups, use the **show ip bgp peer-group** command. If you use **summary** keyword, only a summary of the status of all the members of a peer group will be shown.

To reset the BGP sessions with all the members of a peer group, use the **clear ip bgp** command.

```
clear ip bgp { * | neighbor-address | peer-group-name } [soft [in | out]]
```

## Syntax Description

| Parameter               | Description                                                                                                                                                                           |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>*</b>                | Resets all current BGP sessions.                                                                                                                                                      |
| <i>neighbor-address</i> | Resets only the specified BGP neighbor.                                                                                                                                               |
| <i>peer-group-name</i>  | Resets only the specified BGP peer group.                                                                                                                                             |
| <b>soft</b>             | (Optional) Initiates soft reconfiguration.                                                                                                                                            |
| <b>in   out</b>         | (Optional) Triggers inbound or outbound soft reconfiguration. If you do not specify the <b>in</b> or <b>out</b> option, both inbound and outbound soft reconfiguration are triggered. |

**Step 6** On the ISP1 router, verify that all prefix list, filter list, and route map configuration is now applied to the peer group CUSTOMER.

```
ISP1# show ip bgp peer-group
BGP peer-group is CUSTOMERS
 BGP version 4
 Neighbor sessions:
 0 active, is not multisession capable (disabled)
 Default minimum time between advertisement runs is 30 seconds

For address family: IPv4 Unicast
 BGP neighbor is CUSTOMERS, peer-group external, members:
 172.16.11.1 172.16.12.2 172.16.13.3
 Index 0, Advertise bit 0
 Incoming update prefix filter list is Customer
 Incoming update AS path filter list is 10
 Route map for incoming advertisements is Filter
 Update messages formatted 0, replicated 0
 Number of NLRI's in the update sent: max 0, min 0
```

You can see that all parameters configured for the peer group are listed in the output:

- Prefix list "Customer" for incoming updates.
- Filter list "10" for incoming updates.
- Route map "Filter" for incoming updates.

Note that the peer group should have only [IBGP](#) members or EBGp members. This peer group contains EBGp peers—172.16.11.1, 172.16.12.2, and 172.16.13.3. These are the peers that inherit the configuration from the peer group.

Another option to verify the configuration that is applied to the peer group is to simply view the BGP configuration on the ISP1.

```
ISP1# show running-config | section bgp
router bgp 100
 bgp log-neighbor-changes
 neighbor CUSTOMERS peer-group
 neighbor CUSTOMERS prefix-list Customer in
 neighbor CUSTOMERS route-map Filter in
 neighbor CUSTOMERS filter-list 10 in
 neighbor 172.16.11.1 remote-as 1
 neighbor 172.16.11.1 peer-group CUSTOMERS
 neighbor 172.16.12.2 remote-as 2
 neighbor 172.16.12.2 peer-group CUSTOMERS
 neighbor 172.16.13.3 remote-as 3
 neighbor 172.16.13.3 peer-group CUSTOMERS
```

**Step 7** On the ISP1 router, verify that R1, R2, and R3 peers are part of the CUSTOMERS peer group.

```
ISP1# show ip bgp peer-group CUSTOMERS summary
BGP router identifier 172.16.13.11, local AS number 100
BGP table version is 10, main routing table version 10
3 network entries using 444 bytes of memory
3 path entries using 192 bytes of memory
6/3 BGP path/bestpath attribute entries using 816 bytes of memory
3 BGP AS-PATH entries using 72 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
BGP using 1524 total bytes of memory
BGP activity 24/21 prefixes, 27/24 paths, scan interval 60 secs
```

| Neighbor    | V | AS | MsgRcvd | MsgSent | TblVer | InQ | OutQ | Up/Down  | State/PfxRcd |
|-------------|---|----|---------|---------|--------|-----|------|----------|--------------|
| 172.16.11.1 | 4 | 1  | 18      | 23      | 10     | 0   | 0    | 00:12:10 | 1            |
| 172.16.12.2 | 4 | 2  | 18      | 23      | 10     | 0   | 0    | 00:11:58 | 1            |
| 172.16.13.3 | 4 | 3  | 17      | 21      | 10     | 0   | 0    | 00:11:46 | 1            |

The output lists you all the members of CUSTOMERS peer group

- R1 (172.16.11.1)
- R2 (172.16.12.2)
- R3 (172.16.13.3)

Another option to verify that the neighbor is a member of the peer group is using the show ip bgp neighbor command.

```
ISP1# show ip bgp neighbor 172.16.11.1
BGP neighbor is 172.16.11.1, remote AS 1, external link
Member of peer-group CUSTOMERS for session parameters
 BGP version 4, remote router ID 10.1.0.1
 BGP state = Established, up for 00:12:39
 Last read 00:00:44, last write 00:00:28, hold time is 180, keepalive interval is 60
seconds
<... output omitted ...>
```

```
ISP1# show ip bgp neighbor 172.16.12.2
BGP neighbor is 172.16.12.2, remote AS 2, external link
Member of peer-group CUSTOMERS for session parameters
 BGP version 4, remote router ID 10.2.0.1
 BGP state = Established, up for 00:13:00
 Last read 00:00:13, last write 00:00:12, hold time is 180, keepalive interval is 60
seconds
<... output omitted ...>
```

```
ISP1# show ip bgp neighbor 172.16.13.3
BGP neighbor is 172.16.13.3, remote AS 3, external link
Member of peer-group CUSTOMERS for session parameters
 BGP version 4, remote router ID 10.3.0.1
 BGP state = Established, up for 00:13:08
 Last read 00:00:12, last write 00:00:28, hold time is 180, keepalive interval is 60
seconds
<... output omitted ...>
```

**Step 8** On the ISP1 router, verify that you still receive networks, that all three customers advertise. Verify that the local preference is still set to 150 for all the networks.

```
ISP1# show ip bgp
BGP table version is 10, local router ID is 172.16.13.11
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

| Network        | Next Hop    | Metric | LocPrf | Weight | Path |
|----------------|-------------|--------|--------|--------|------|
| *> 10.1.0.0/24 | 172.16.11.1 | 0      | 150    | 0      | 1 i  |
| *> 10.2.0.0/24 | 172.16.12.2 | 0      | 150    | 0      | 2 i  |
| *> 10.3.0.0/24 | 172.16.13.3 | 0      | 150    | 0      | 3 i  |

You can see that ISP1 router in AS 100 received three networks, each from different customer:

- 10.1.0.0/24 network originates in the first customer's AS 1.
- 10.2.0.0/24 network originates in the second customer's AS 2.
- 10.3.0.0/24 network originates in the third customer's AS 3.

All three networks still have local preference set to 150. Even though the route map that sets the local preference is now included in the peer group configuration, the neighbors inherited the configuration from the peer group.

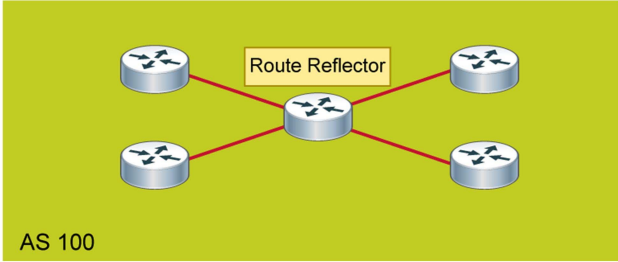


# BGP Peer Group Configuration Examples

## Peer Group Configuration Example—BGP Route Reflector

In this example, a router acting as a [BGP](#) route reflector has four [IBGP](#) neighbors.

### Peer Group Configuration Example—BGP Route Reflector



```
router(config)# router bgp 100
router(config-router)# neighbor IBGP_peers peer-group
router(config-router)# neighbor IBGP_peers remote-as 100
router(config-router)# neighbor IBGP_peers update-source loopback 0
router(config-router)# neighbor IBGP_peers password c73Dx8K
router(config-router)# neighbor IBGP_peers send-community
router(config-router)# neighbor 10.0.1.3 peer-group IBGP_peers
router(config-router)# neighbor 10.0.1.4 peer-group IBGP_peers
router(config-router)# neighbor 10.0.1.6 peer-group IBGP_peers
router(config-router)# neighbor 10.0.1.8 peer-group IBGP_peers
```

© 2015 Cisco and/or its affiliates. All rights reserved. 230

In a large [AS](#), some routers may have many 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, 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 part of the peer group configuration since the AS number is the same for all peer group members.

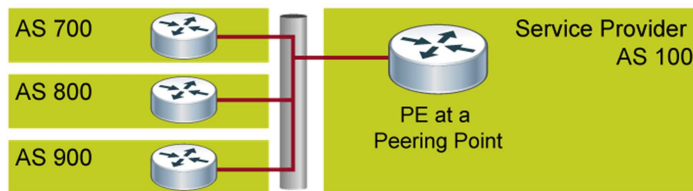
The peer group is a very powerful tool when you are dealing with many IBGP neighbors. You can give all of the neighbors the same configuration to ensure a consistent AS-wide routing policy.

## Peer Group Configuration Example—Edge Router at a Peering Point

In this example, the router in AS 100 is being configured with a peer group named "Peering"

## Peer Group Configuration Example—Edge Router at a Peering Point

```
PE(config)# router bgp 100
PE(config-router)# neighbor Peering peer-group
PE(config-router)# neighbor Peering filter-list 10 in
PE(config-router)# neighbor Peering route-map PeerMap out
PE(config-router)# neighbor Peering maximum-prefix 50
PE(config-router)# neighbor Peering remove-private-as
PE(config-router)# neighbor 209.165.207.7 remote-as 700
PE(config-router)# neighbor 209.165.207.7 peer-group Peering
PE(config-router)# neighbor 209.165.208.8 remote-as 800
PE(config-router)# neighbor 209.165.208.8 peer-group Peering
PE(config-router)# no neighbor 209.165.208.8 maximum-prefix 50
PE(config-router)# no neighbor 209.165.208.8 filter-list 10 in
PE(config-router)# neighbor 209.165.209.9 remote-as 900
PE(config-router)# neighbor 209.165.209.9 peer-group Peering
```



© 2015 Cisco and/or its affiliates. All rights reserved.

231

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 from all AS paths before the routes are sent to the peer AS. These AS numbers are in the range 64512 to 65535 inclusive.

The neighbors in AS 700, AS 800, and AS 900 are then assigned to the peer group, meaning that the router in AS 100 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 800, the individual configuration of **no filter-list 10** will override the peer group configuration. Thus, the filter list will not be used for this neighbor. The limitation on the number of received routes from AS 800 is also removed from the neighbor in AS 800.

---

**Note** 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.

---

# BGP Dynamic Update Peer Groups Feature

In the Cisco IOS 12.0(24)S release Cisco introduced dynamic update peer groups, a feature that enables the routers to automatically group [BGP](#) neighbor into groups and generate update messages accordingly.

## BGP Dynamic Update Peer Groups Feature

- Separates BGP update generation from neighbor configuration
- Introduces a new algorithm that dynamically calculates and optimizes update groups of neighbors that share the same outbound policies.
- Requires no additional configuration.
- Optimal BGP update message generation occurs automatically and independently

© 2015 Cisco and/or its affiliates. All rights reserved.

232

In previous versions of Cisco IOS software, BGP update messages were grouped based on peer group configurations. This method of grouping neighbors for BGP update message generation reduced the amount of system processing resources that are 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 that are 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 groups feature contains an algorithm that dynamically calculates and optimizes update groups of neighbors that share outbound policies and can share the update messages. The BGP dynamic update groups feature separates update group replication from peer group configuration, improving convergence time and flexibility of neighbor configuration. The BGP update groups require no configuration, and a router optimizes BGP update message generation automatically.

When a change to the configuration 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.

---

**Note** A soft reset, which is performed on a per-neighbor basis, does not clear the BGP session and facilitates the application of new policies. There are two methods of performing a soft reset. Dynamic inbound soft reset is used to generate inbound updates from a neighbor. An outbound soft reset is used to send a new set of updates to a neighbor.

---

## BGP Dynamic Update Peer Groups Feature (Cont.)

**clear ip bgp update-group** [*index-group* | *ip-address*]

- Clears BGP update-group member sessions.

**debug ip bgp groups** [*index-group* | *ip-address*]

- Displays information that is related to the processing of BGP update-groups.

**show ip bgp replication** [*index-group* | *ip-address*]

- Displays update replication statistics for BGP update-groups.

**show ip bgp update-group** [*index-group* | *ip-address*] [**summary**]

- Displays information about BGP update-groups.

© 2015 Cisco and/or its affiliates. All rights reserved.

233

If no argument is specified, the **clear ip 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                                                                                                                                                |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>index-group</i> | (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. |
| <i>ip-address</i>  | (Optional) Specifies the <a href="#">IP address</a> of a single peer that will be reset.                                                                   |

## debug ip bgp groups

The output of the **debug ip bgp groups** command displays you 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.

```
debug ip bgp groups [index-group | ip-address]
```

### Syntax Description

| Parameter          | Description                                                                                                                                                                         |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>index-group</i> | (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. |
| <i>ip-address</i>  | (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:

```
ISP1# debug ip bgp groups
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.11.1 Down User reset
5w4d: BGP-DYN(0): Comparing neighbor 172.16.11.1 flags 0x0 cap 0x0 and updgrp 2 fl0
5w4d: BGP-DYN(0): Update-group 2 flags 0x0 cap 0x0 policies same as 172.16.11.1 fl0
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.12.2 Down User reset
5w4d: BGP-DYN(0): Comparing neighbor 172.16.12.2 flags 0x0 cap 0x0 and updgrp 2 fl0
5w4d: BGP-DYN(0): Update-group 2 flags 0x0 cap 0x0 policies same as 172.16.12.2 fl0
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.100.22 Down User reset
5w4d: BGP-DYN(0): Comparing neighbor 172.16.100.22 flags 0x0 cap 0x0 and updgrp 2 fl0
5w4d: BGP-DYN(0): Update-group 2 flags 0x0 cap 0x0 policies same as 172.16.100.22 fl0
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.11.1 Up
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.100.22 Up
5w4d: %BGP-5-ADJCHANGE: neighbor 172.16.12.2 Up
```

---

|             |                                                                                                                                                         |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | The output of this command can be very verbose, so you should not deploy this command in a production network unless you are troubleshooting a problem. |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|

---

## show ip bgp replication

To display update replication statistics for BGP update-groups, use the **show ip bgp replication** command.

```
show ip bgp replication [index-group | ip-address]
```

## Syntax Description

| Parameter          | Description                                                                                                                                                                                      |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>index-group</i> | (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. |
| <i>ip-address</i>  | (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:

```
ISP1# show ip bgp replication
```

| Index | Members | Leader        | MsgFmt | MsgRepl | Csize  | Current Version | Next Version |
|-------|---------|---------------|--------|---------|--------|-----------------|--------------|
| 1     | 2       | 172.16.11.1   | 0      | 0       | 0/1000 | 1/0             |              |
| 2     | 1       | 172.16.100.22 | 0      | 0       | 0/1000 | 1/0             |              |

## show ip bgp update-group

To display information about BGP update-groups, use the **show ip bgp update-group**.

```
show ip bgp update-group [index-group | ip-address] [summary]
```

## Syntax Description

| Parameter          | Description                                                                                                                                                                                                  |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>index-group</i> | (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.             |
| <i>ip-address</i>  | (Optional) Specifies the IP address of a single neighbor for which update-group statistics will be displayed.                                                                                                |
| <b>summary</b>     | (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 <i>index-group</i> or <i>ip-address</i> argument. |

The following sample output from the **show ip bgp update-group** command shows update-group information for all neighbors:

ISP1# **show ip bgp update-group**

BGP version 4 update-group 1, external, Address Family: IPv4 Unicast

BGP Update version : 6/0, messages 0

Topology: global, highest version: 6, tail marker: 6

Format state: Current working (OK, last minimum advertisement interval)

Refresh blocked (not in list, last not in list)

Update messages formatted 1, replicated 2, current 0, refresh 0, limit 1000

Number of NLRI's in the update sent: max 5, min 0

Minimum time between advertisement runs is 30 seconds

Has 1 member:

172.16.100.22

BGP version 4 update-group 2, external, Address Family: IPv4 Unicast

BGP Update version : 6/0, messages 0

Topology: global, highest version: 6, tail marker: 6

Format state: Current working (OK, last minimum advertisement interval)

Refresh blocked (not in list, last not in list)

Route map for outgoing advertisements is CUST1

Update messages formatted 1, replicated 2, current 0, refresh 0, limit 1000

Number of NLRI's in the update sent: max 5, min 0

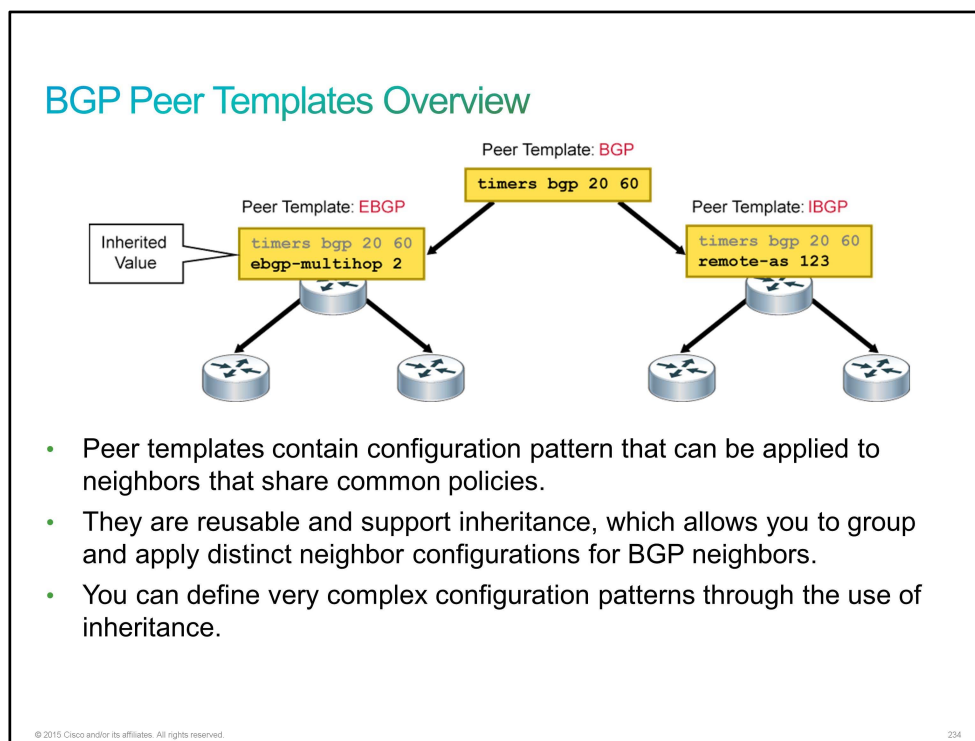
Minimum time between advertisement runs is 30 seconds

Has 2 members:

172.16.12.2      172.16.11.1

# BGP Peer Templates Overview

The [BGP](#) dynamic update 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 peer templates feature was introduced along with the BGP dynamic update 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 you to group and apply distinct neighbor configurations for BGP neighbors that share common policies. Peer templates also allow you to define very complex configuration patterns, since a peer template can inherit a configuration from another peer template.

The example in the figure presents peer template inheritance. One peer template, called BGP, is configured and contains parameters that are common to all BGP neighbors. In the example, such parameters are the BGP keepalive and hold-down timers. Then one peer template is configured for [IBGP](#) and one for [EBGP](#) neighbors. Both templates inherit settings from the common peer template that is called BGP and add other parameters that are common only to IBGP or EBGP settings, respectively. Therefore templates can be created more effectively with fewer configurations when compared to BGP peer groups.



## BGP Peer Templates Overview (Cont.)

Two types of peer templates are available:

- **Peer session template:** Used to group and apply the configuration of general session commands that are common to all address-family configuration modes.
- **Peer policy template:** Used to group and apply the configuration of commands that are applied within specific address-family configuration modes.

© 2015 Cisco and/or its affiliates. All rights reserved.

235

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 configuration modes.
- Peer policy templates are used to group and apply the configuration of commands that are applied within specific address-families 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. BGP peer routers using peer templates also benefit from automatic update group configuration. With the configuration of the BGP peer templates the support of the BGP dynamic update peer groups feature, you no longer need to configure peer groups in BGP. You benefit from improved configuration flexibility and faster convergence.

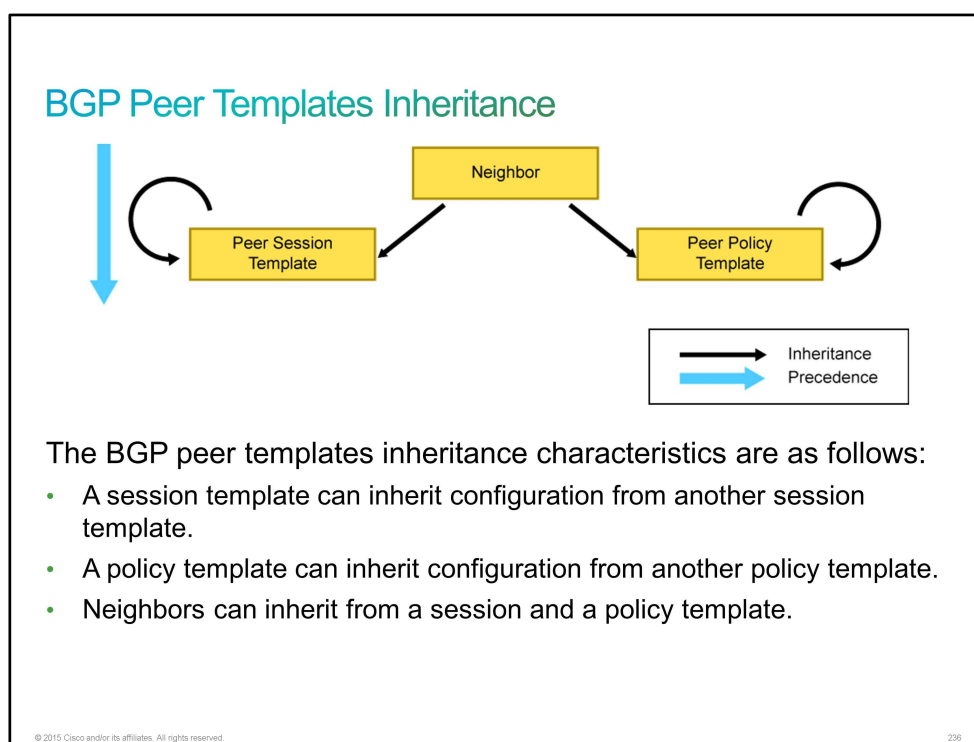
---

|             |                                                                                                                                                                                                                                                                                                                                                                                                            |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | The configuration of BGP peer templates does not conflict with or restrict peer group configuration, and peer groups are still supported in Cisco IOS Releases that support BGP peer templates. However, 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 from peer templates. |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

# BGP Peer Templates Inheritance

The inheritance capability is a key component of peer-template operation. Inheritance in a peer template is similar to the node and tree structures that are 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. Common configuration statements can now 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. It allows you to chain together peer-template configurations to create simple configurations that inherit common configuration statements or complex configurations that apply specific configuration statements along with common inherited configurations.

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.

---

|             |                                                                                                                                      |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | If you attempt to configure more than one inherit statement with a single peer session template, an error message will be displayed. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------|

---

This behavior allows a [BGP](#) neighbor to directly inherit only one session template and indirectly inherit up to seven additional peer session templates. This practice allows 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 configurations are evaluated first and applied starting with the last node in the branch and ending with the directly applied peer session template configuration at the source of the tree. The directly applied peer session template will have priority over inherited peer session template configurations. Any configuration statements that are duplicated in inherited peer session templates will be overwritten by the directly applied peer session template. Meaning that if a general session command is reapplied with a different value, the subsequent value will have priority and overwrite the previous value that was configured in the indirectly inherited template.

Peer policy templates are used to configure BGP policy commands that are configured for neighbors that belong to specific address families. 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

Like peer session templates, a peer policy template supports inheritance. However, there are minor differences. A directly applied peer policy template can directly or indirectly inherit configurations from up to seven peer policy templates. That is, a total of eight peer policy templates can be applied to a neighbor or neighbor group. 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, the difference is that a peer policy template will not collapse like a route map. Every sequence is evaluated, and if a BGP policy command is reapplied with a different value, any previous value from a lower sequence number will be overwritten.

The directly applied peer policy template and the inherit statement with the highest sequence number will always have priority and be applied last. Commands that are reapplied in subsequent peer templates will always overwrite the previous values. This behavior is designed to allow you to apply common policy configurations to large neighbor groups and specific policy configurations only to certain neighbors and neighbor groups, without duplicating individual policy configuration commands.

The configuration of peer policy templates simplifies and improves the flexibility of BGP configuration. A specific policy can be configured once and referenced many times. Because a peer policy supports up to eight levels of inheritance, specific and complex BGP policies can also be created.

# BGP Peer Templates Configuration

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. [BGP](#) policy configuration commands that are configured only for specific address families configuration modes are configured with peer policy templates.

## Peer Session Commands

### Peer Session Commands

```
template peer-session session-template-name
```

- Creates a peer session.

```
inherit peer-session session-template-name
```

- Configures a peer session template to inherit the configuration of another peer session template.

```
neighbor ip-address inherit peer-session session-template-name
```

- Sends a peer session template to a neighbor so that the neighbor can inherit the configuration.

© 2015 Cisco and/or its affiliates. All rights reserved. 237

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
```

### Syntax Description

| Parameter                    | Description                               |
|------------------------------|-------------------------------------------|
| <i>session-template-name</i> | Name or tag for the peer session template |

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
```

## Syntax Description

| Parameter                    | Description                               |
|------------------------------|-------------------------------------------|
| <i>session-template-name</i> | Name or tag for the peer session template |

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.

Use the **neighbor inherit peer-session** command in address family or router configuration mode.

```
neighbor ip-address inherit peer-session session-template-name
```

## Syntax Description

| Parameter                    | Description                                |
|------------------------------|--------------------------------------------|
| <i>ip-address</i>            | <a href="#">IP address</a> of the neighbor |
| <i>session-template-name</i> | Name or tag for the peer session template  |

---

|             |                                                                                                                                                                                                        |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Note</b> | 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. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

---

## Peer Policy Commands

Peer policy templates support only general policy commands. BGP policy configuration commands that are configured only for specific address families configuration modes are configured with peer policy templates.

## Peer Policy Commands

```
inherit peer-policy policy-template-name sequence-number
```

- Configures a peer policy template to inherit the configuration from another peer policy template.

```
neighbor ip-address inherit peer-policy policy-template-name
```

- Sends a peer policy template to a neighbor so that the neighbor can inherit the configuration.

© 2015 Cisco and/or its affiliates. All rights reserved.

238

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
```

### Syntax Description

| Parameter               | Description                                                                                                                                                          |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>peer-policy-name</i> | Name of the peer policy template to be inherited.                                                                                                                    |
| <i>sequence-number</i>  | 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 ip-address inherit peer-policy policy-template-name
```

## Syntax Description

| Parameter                   | Description                              |
|-----------------------------|------------------------------------------|
| <i>ip-address</i>           | IP address of the neighbor               |
| <i>policy-template-name</i> | Name or tag for the peer policy template |

## BGP Peer Templates Verification

### BGP Peer Templates Verification

```
show ip bgp template peer-session [session-template-name]
```

- Displays peer policy template configurations.

```
show ip bgp template peer-policy [policy-template-name]
```

- Displays locally configured peer policy templates.

© 2015 Cisco and/or its affiliates. All rights reserved. 239

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 [session-template-name]
```

## Syntax Description

| Parameter                    | Description                                                   |
|------------------------------|---------------------------------------------------------------|
| <i>session-template-name</i> | (Optional) Name of a locally configured peer session template |

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                                                  |
|-----------------------------|--------------------------------------------------------------|
| <i>policy-template-name</i> | (Optional) Name of a locally configured peer policy template |



# Summary

This topic summarizes the key points that were discussed in this lesson.

## Summary

- BGP peer groups were designed primarily for CPU optimization. They can also be used for configuration optimization.
  - BGP peer groups are used on a router with several neighbors that have similar, but not completely identical policies.
- 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.
  - IBGP and EBGP neighbors cannot be mixed in a peer group.
- To configure BGP peer groups, 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

240

## Summary (Cont.)

- BGP dynamic update groups separate BGP update generation from neighbor configuration.
- BGP peer templates are reusable configuration patterns that support inheritance.
  - Peer session templates—contain configuration that is common to all address families.
  - Peer policy templates—contain configuration that is applied within a specific address family.

© 2015 Cisco and/or its affiliates. All rights reserved.

241



# Lesson 4: Using BGP Route Dampening

---

## Overview

Even when a [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. 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.

Upon completing this lesson, you will be able to:

- 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 and monitor BGP route-dampening

# BGP Route Dampening

[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.

## BGP Route Dampening

- Designed to reduce router processing load that is caused by unstable routes
- Prevents sustained routing oscillations without affecting other well-behaved routes
- Defined in RFC 2439: *BGP Route Flap Dampening*
- A tool that is designed to help minimize the number of BGP updates
- Other update reduction tools:
  - Batching of BGP updates
  - Per-neighbor update timers

© 2015 Cisco and/or its affiliates. All rights reserved.

242

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 helps with 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 extra BGP scalability mechanism, 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

© 2015 Cisco and/or its affiliates. All rights reserved.

243

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.

# BGP Route Dampening Operation

A BGP router with route dampening enabled keeps track of all routes (even the routes that are unreachable) so it can recall the penalties that are assigned to each route.

## BGP Route Dampening Operation

- Each time an EBGp route flaps, it gets 1000 penalty points (IGBP routes are not dampened).
- The penalty that is 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

244

Every time a route flap occurs, the flapping route receives 1000 penalty points. The penalty is gradually decreased by using 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.

## BGP 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

245

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 a BGP neighbor has withdrawn 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

© 2015 Cisco and/or its affiliates. All rights reserved.246

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)

© 2015 Cisco and/or its affiliates. All rights reserved.

247

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 by using 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 that a route map entry matches.
- 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

© 2015 Cisco and/or its affiliates. All rights reserved.

246

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.

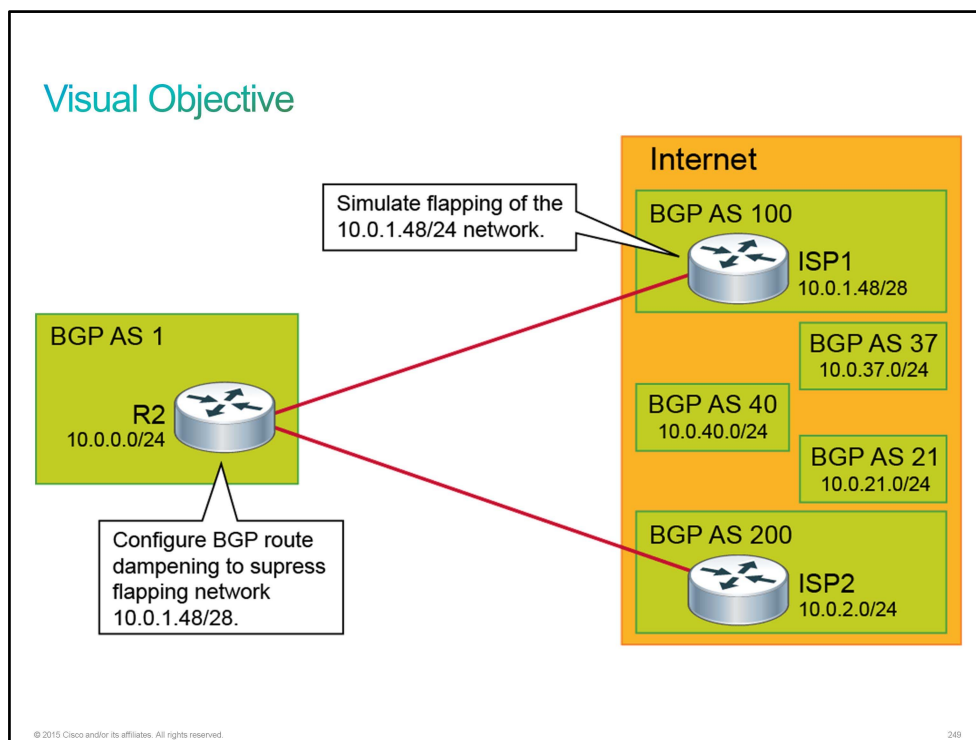
You can use a route-map in combination with a prefix-list to match on prefix length and to set different route-dampening parameters for larger prefixes than for smaller ones. A practical service provider policy is to use a route-map to exclude root [DNS](#) servers from dampening altogether.

You can then attach the route-map to the BGP route-dampening process with the **bgp dampening route-map** command.

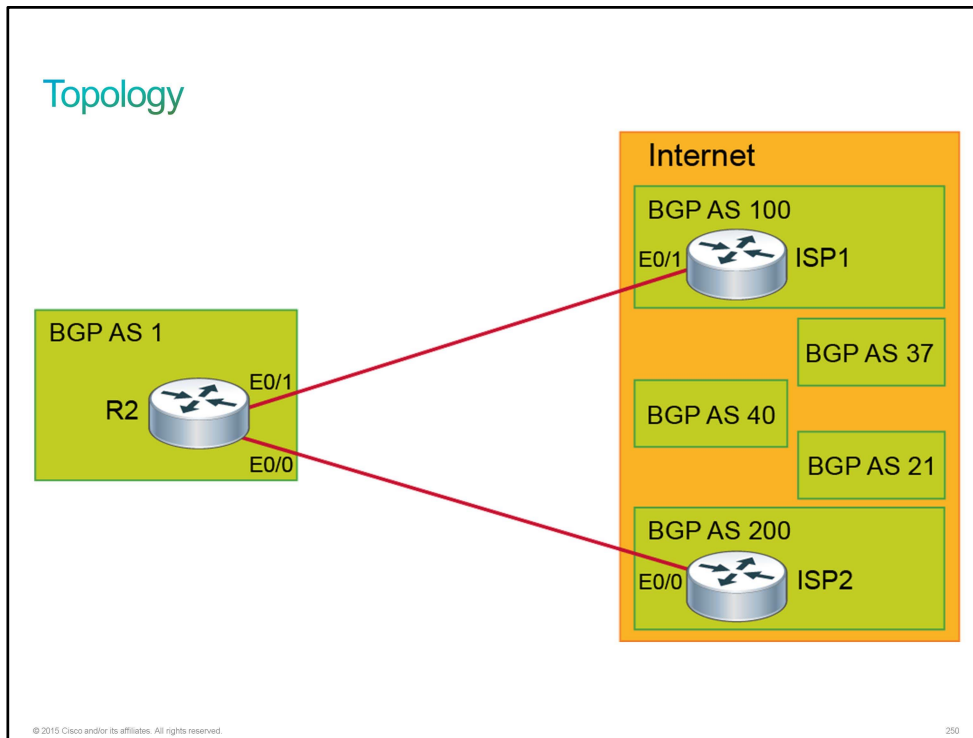
# Discovery 20: Configure BGP Route Dampening

## Overview

Through this discovery, you will learn how to configure [BGP](#) route dampening. You will enable route dampening on the R2 router. You will simulate flapping of one of the advertised networks on the ISP1 router. Consequently, the network flapping will get suppressed on the R2 router, where route dampening is enabled.



# Topology



## Job Aids

---

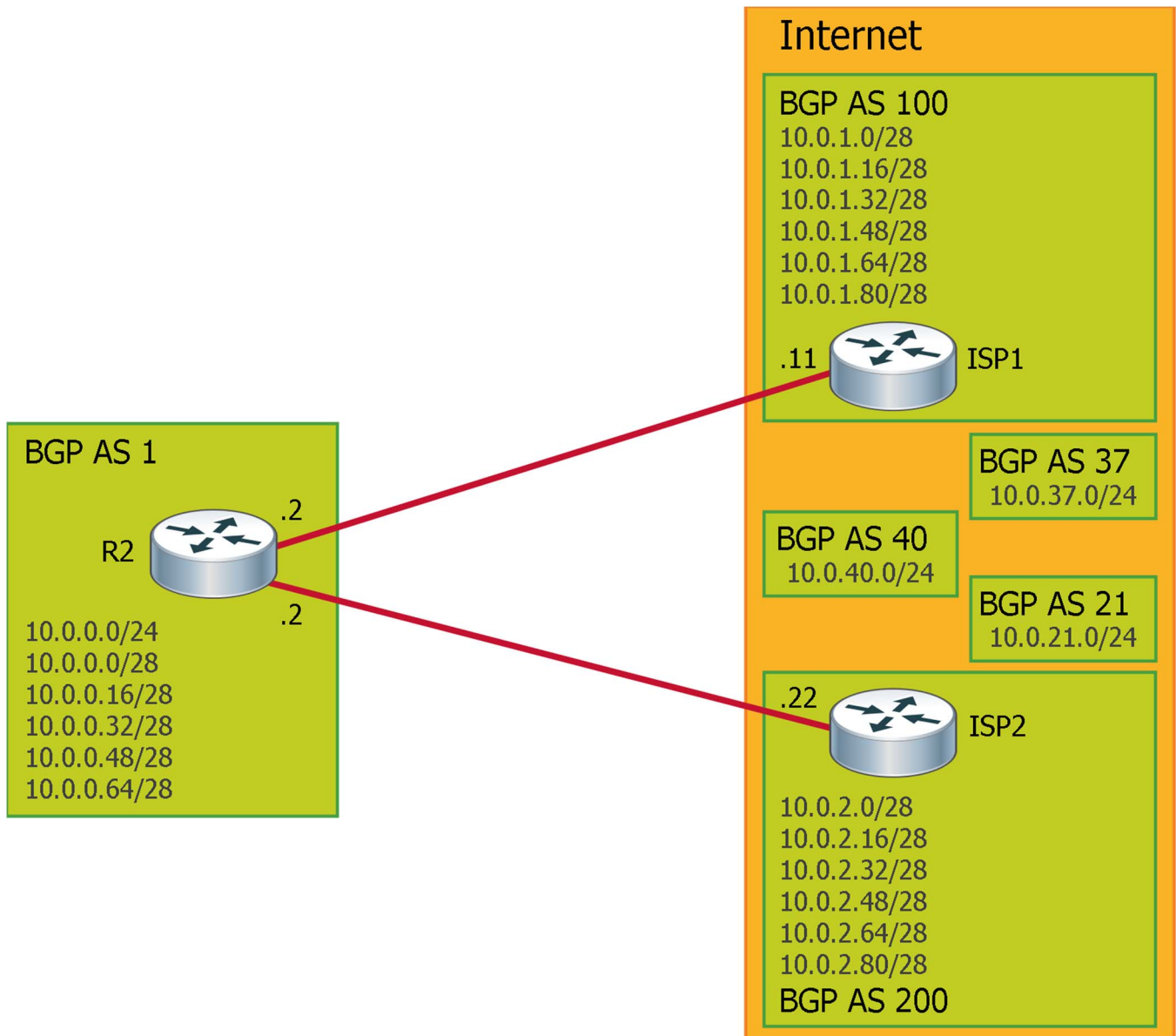
**Note** If you shut down an interface on a real router or switch, the connected device will see it as "down/down." Due to virtualization specifics, [IOL](#) behavior is slightly different. If you shut down an interface on a router or switch, the connected device will see it as "up/up." In IOL, the status of an interface can only be "up/up" or "administratively down/down."

---

## Device Information

| Device | Interface                                                                        | IP address                                                                                  | Description                                                        |
|--------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| R2     | Ethernet 0/0                                                                     | 172.16.22.2/24                                                                              | Connection to ISP2                                                 |
| R2     | Ethernet 0/1                                                                     | 172.16.12.2/24                                                                              | Connection to ISP1                                                 |
| R2     | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4<br>Loopback 5               | 10.0.0.1/28<br>10.0.0.17/28<br>10.0.0.33/28<br>10.0.0.49/28<br>10.0.0.65/28                 | Loopbacks simulate <a href="#">LAN</a> networks                    |
| ISP1   | Ethernet 0/1                                                                     | 172.16.12.11/24                                                                             | Connection to R2                                                   |
| ISP1   | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4<br>Loopback 5<br>Loopback 6 | 10.0.1.1/28<br>10.0.1.17/28<br>10.0.1.33/28<br>10.0.1.49/28<br>10.0.1.65/28<br>10.0.1.81/28 | Loopbacks simulate LAN networks                                    |
| ISP1   | Loopback 21<br>Loopback 37<br>Loopback 40                                        | 10.0.21.1<br>10.0.37.1<br>10.0.40.1                                                         | Loopbacks simulate extra networks in different autonomous systems. |
| ISP2   | Ethernet 0/0                                                                     | 172.16.22.22/24                                                                             | Connection to R2                                                   |
| ISP2   | Loopback 1<br>Loopback 2<br>Loopback 3<br>Loopback 4<br>Loopback 5<br>Loopback 6 | 10.0.2.1/28<br>10.0.2.17/28<br>10.0.2.33/28<br>10.0.2.49/28<br>10.0.2.65/28<br>10.0.2.81/28 | Loopbacks simulate LAN networks                                    |
| ISP2   | Loopback 21<br>Loopback 37<br>Loopback 40                                        | 10.0.21.1<br>10.0.37.1<br>10.0.40.1                                                         | Loopbacks simulate extra networks in different ASs.                |

[IP addresses](#) and advertised networks in BGP are preconfigured as shown below:



BGP is also preconfigured as [EBGP](#) (R2 to ISP1 and R2 to ISP2). The ISP1 router announces these networks:

- 10.0.1.0/28
- 10.0.1.16/28
- 10.0.1.32/28
- 10.0.1.48/28
- 10.0.1.64/28
- 10.0.1.80/28

# Configure BGP Route Dampening

## Discovery Steps

**Step 1** On the R2 router, verify that you receive 10.0.1.48/28 route to the BGP table from the ISP1 neighbor.

```
R2# show ip bgp
BGP table version is 24, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | Next Hop     | Metric | LocPrf | Weight | Path |
|----|--------------|--------------|--------|--------|--------|------|
| *> | 10.0.0.0/24  | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.16/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.32/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.48/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.64/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.1.0/28  | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.16/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.32/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.48/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.64/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.80/28 | 172.16.12.11 | 0      |        | 0 100  | i    |

<... output omitted...>

**Step 2** Enable BGP route dampening on the R2 router with default values for dampening parameters.

```
R2(config)# router bgp1
R2(config-router)# bgp dampening
```

**Step 3** Enable debugging for BGP dampening on the R2 router.

```
R2# debug ip bgp dampening
```

**Step 4** On the ISP1 router, disable and enable interface Loopback 4 (10.0.1.49/28) for at least three times. Allow yourself at least a minute before next command is applied.

```
ISP1(config)# interface Loopback 4
ISP1(config-if)# shutdown
!wait for at least 30 seconds!
ISP1(config-if)# no shutdown
!wait for 30 seconds!
!repeat these steps for at least three times!
```

**NOTE:** You will simulate the flapping of the 10.0.1.48/28 network by enabling and disabling loopback interface few times in short amount of time.

**Step 5** Observe the debugging output. Each time that you disable the interface Loopback 4, the 10.0.1/48 network with existing path gets charged with 1000 points. Once the suppress penalty passes the threshold of 2000 and you enable the interface Loopback 4, network 10.0.1.0/48 gets suppressed.

```
R2#
*Apr 1 08:36:29.423: EvD: charge penalty 1000, new accum. penalty 1000, flap count 1
*Apr 1 08:36:29.423: EvD: unsuppress item left in reuse timer array with penalty 1000
*Apr 1 08:36:29.423: BGP(0): charge penalty for 10.0.1.48/28 path 100 with halflife-
time 15 reuse/suppress 750/2000
*Apr 1 08:36:29.423: BGP(0): flapped 1 times since 00:00:00. New penalty is 1000
R2#
*Apr 1 08:37:12.319: EvD: accum. penalty decayed to 969 after 42 second(s)
R2#
*Apr 1 08:37:30.196: EvD: accum. penalty 957, not suppressed
R2#
*Apr 1 08:38:01.030: EvD: accum. penalty decayed to 935 after 31 second(s)
*Apr 1 08:38:01.030: EvD: charge penalty 1000, new accum. penalty 1935, flap count 2
*Apr 1 08:38:01.030: EvD: unsuppress item left in reuse timer array with penalty 1935
*Apr 1 08:38:01.030: BGP(0): charge penalty for 10.0.1.48/28 path 100 with halflife-
time 15 reuse/suppress 750/2000
*Apr 1 08:38:01.030: BGP(0): flapped 2 times since 00:01:31. New penalty is 1935
R2#
*Apr 1 08:38:12.380: EvD: accum. penalty decayed to 1920 after 11 second(s)
R2#
*Apr 1 08:38:31.858: EvD: accum. penalty 1890, not suppressed
R2#
*Apr 1 08:39:02.687: EvD: accum. penalty decayed to 1846 after 31 second(s)
*Apr 1 08:39:02.687: EvD: charge penalty 1000, new accum. penalty 2846, flap count 3
*Apr 1 08:39:02.687: EvD: unsuppress item left in reuse timer array with penalty 2846
*Apr 1 08:39:02.687: BGP(0): charge penalty for 10.0.1.48/28 path 100 with halflife-
time 15 reuse/suppress 750/2000
*Apr 1 08:39:02.687: BGP(0): flapped 3 times since 00:02:33. New penalty is 2846
R2#
*Apr 1 08:39:12.402: EvD: accum. penalty decayed to 2835 after 9 second(s)
R2#
*Apr 1 08:39:32.484: BGP(0): suppress 10.0.1.48/28 path 100 for 00:28:30 (penalty
2791)
*Apr 1 08:39:32.484: halflife-time 15, reuse/suppress 750/2000
*Apr 1 08:39:32.484: EvD: accum. penalty 2791, now suppressed with a reuse intervals
of 171
```

## Monitoring Route Dampening



## Monitoring Route Dampening

```
router# show ip bgp dampened-paths
```

- Displays the dampened routes

```
router# show ip bgp flap-statistics [{regex 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

© 2015 Cisco and/or its affiliates. All rights reserved.

251

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.

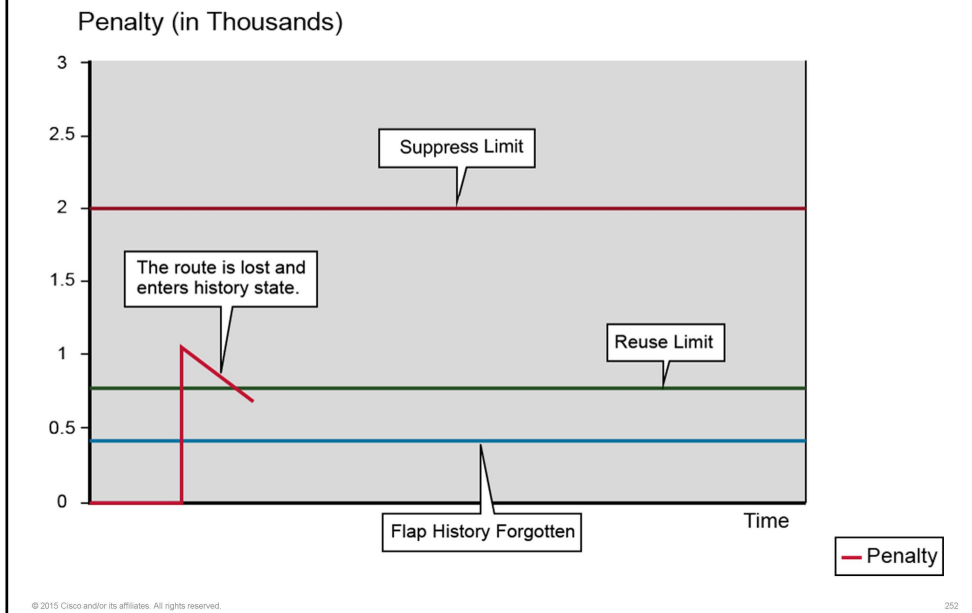
Use the **show ip bgp flap-statistics** command to list all routes that have a penalty that is still above the time-to-forget limit. You can also use this command in combination with regular expressions and filter-lists.

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.

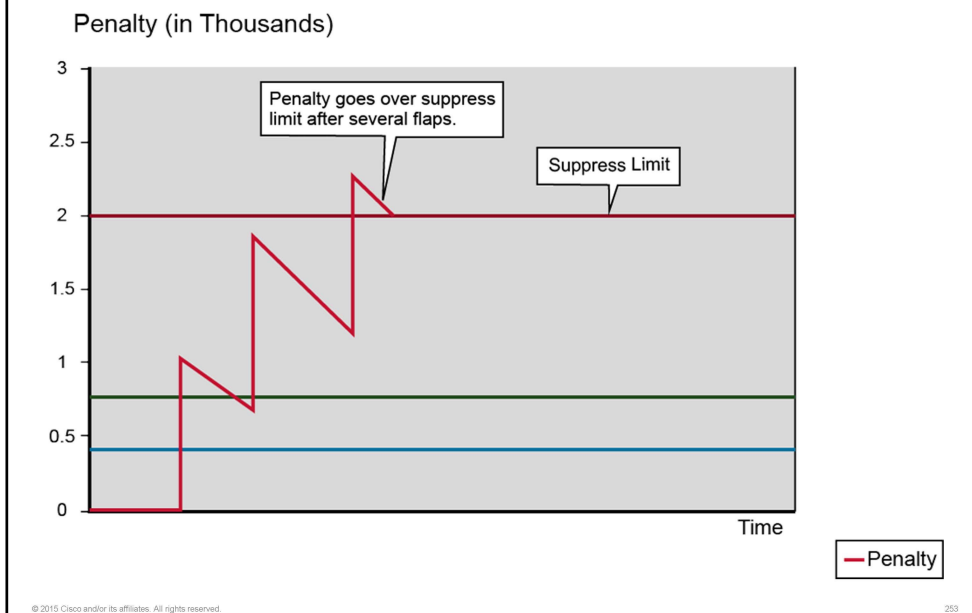
## Monitoring Route Dampening (Cont.)



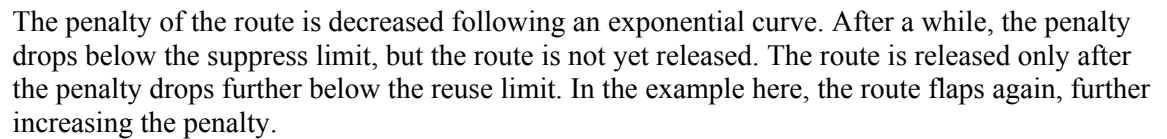
The example shows how, after the first flap (when a route becomes unreachable), the router withdraws the route. However the router keeps the route in its own database to keep track of the penalty. The route enters the history state.

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.

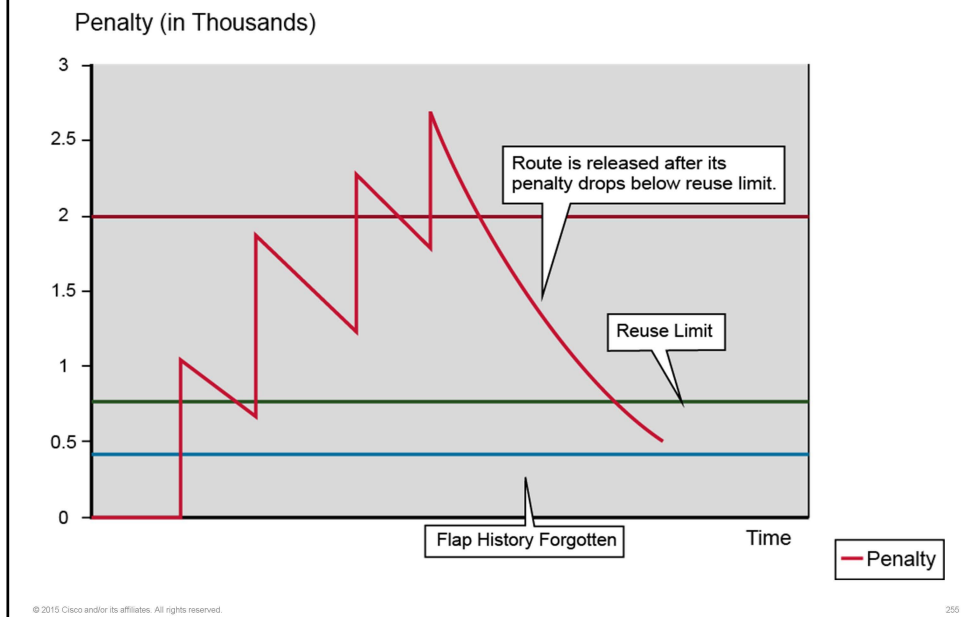
## Monitoring Route Dampening—Example (Cont.)



- 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.)



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 router selects the released route 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 router clears the penalty and the flap history that are associated with the route.

**Step 6** On the R2 router, verify the 10.0.1.0/48 entry in the BGP table.

```
R2# show ip bgp 10.0.1.48/28
BGP routing table entry for 10.0.1.48/28, version 47
Paths: (1 available, no best path)
 Not advertised to any peer
 Refresh Epoch 1
 100, (suppressed due to dampening)
 172.16.12.11 from 172.16.12.11 (10.0.40.1)
 Origin IGP, metric 0, localpref 100, valid, external
 Dampinfo: penalty 2674, flapped 3 times in 00:04:00, reuse in 00:05:39
```

You should see that network 10.0.1.0/48 received via 172.16.12.11 neighbor gets suppressed due to dampening.

**Step 7** On the R2 router, verify the content of the whole BGP table.

```
R2# show ip bgp
BGP table version is 47, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | Next Hop     | Metric | LocPrf | Weight | Path |
|----|--------------|--------------|--------|--------|--------|------|
| *> | 10.0.0.0/24  | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.16/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.32/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.48/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.0.64/28 | 0.0.0.0      | 0      |        | 32768  | i    |
| *> | 10.0.1.0/28  | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.16/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.32/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *d | 10.0.1.48/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.64/28 | 172.16.12.11 | 0      |        | 0 100  | i    |
| *> | 10.0.1.80/28 | 172.16.12.11 | 0      |        | 0 100  | i    |

<... output omitted...>

You should see network 10.0.1.0/48 marked as dampened.

**Step 8** On the R2 router, display all dampened paths.

```
R2# show ip bgp dampening dampened-paths
BGP table version is 47, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | From         | Reuse    | Path  |
|----|--------------|--------------|----------|-------|
| *d | 10.0.1.48/28 | 172.16.12.11 | 00:04:49 | 100 i |

You should see that network 10.0.1.0/48 marked as dampened for a path via 172.16.12.11 neighbor in autonomous system 100.

**Step 9** On the R2 router, display flapping statistics for BGP route dampening.

```
R2# show ip bgp dampening flap-statistics
BGP table version is 47, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

|    | Network      | From         | Flaps | Duration | Reuse    | Path |
|----|--------------|--------------|-------|----------|----------|------|
| *d | 10.0.1.48/28 | 172.16.12.11 | 3     | 00:05:20 | 00:04:19 | 100  |

You should see that network 10.0.1.0/48 flapped three times.

**Step 10** On the R2 router, display BGP dampening parameters.

```
R2# show ip bgp dampening parameters
dampening 15 750 2000 60 (DEFAULT)
Half-life time : 15 mins Decay Time : 2320 secs
Max suppress penalty: 12000 Max suppress time: 60 mins
Suppress penalty : 2000 Reuse penalty : 750
```

You should see the default BGP dampening parameters, for instance for suppress penalty. The default value is 2000.

## Releasing Dampened Routes

### Monitoring Route Reflectors

#### Releasing Dampened Routes

```
router# clear ip bgp ip-address flap-statistics [{regex 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 are also cleared when the BGP session with the neighbor is lost.

© 2015 Cisco and/or its affiliates. All rights reserved.

256

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** *regex*} | {**filter-list** *list-name*} | {*ip-address network-mask*}]

## Syntax Description

| Parameter                           | Description                                                                                                                                                                                                            |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>                   | (Optional) Clears flap statistics for a single entry at this IP address. If this argument is placed before <b>flap-statistics</b> , the router clears flap statistics for all paths from the neighbor at this address. |
| <b>regexp</b> <i>regex</i>          | (Optional) Clears flap statistics for all the paths that match the regular expression.                                                                                                                                 |
| <b>filter-list</b> <i>list-name</i> | (Optional) Clears flap statistics for all the paths that pass the access-list.                                                                                                                                         |
| <i>network-mask</i>                 | (Optional) Network mask that is applied to the <i>ip-address</i> 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                                                                                                                                                                                                            |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>ip-address</i>   | (Optional) Clears flap statistics for a single entry at this IP address. If this argument is placed before <b>flap-statistics</b> , the router clears flap statistics for all paths from the neighbor at this address. |
| <i>network-mask</i> | (Optional) Network mask that is applied to the <i>ip-address</i> argument.                                                                                                                                             |

## Step 11

Clear the BGP dampening information on the R2 router.

```
R2# clear ip bgp dampening
```

**Step 12** On the R2 router, verify the 10.0.1.0/48 entry in the BGP table.

```

R2# show ip bgp 10.0.1.48/28
BGP routing table entry for 10.0.1.48/28, version 48
Paths: (1 available, best #1, table default)
 Advertised to update-groups:
 2
 Refresh Epoch 1
 100
 172.16.12.11 from 172.16.12.11 (10.0.40.1)
 Origin IGP, metric 0, localpref 100, valid, external, best

```

You should see that network 10.0.1.0/48 received via 172.16.12.11 neighbor is not suppressed anymore.

**Step 13** On the R2 router, verify the content of the whole BGP table.

```

R2# show ip bgp
BGP table version is 48, local router ID is 10.0.0.65
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
 r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
 x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

 Network Next Hop Metric LocPrf Weight Path
*> 10.0.0.0/24 0.0.0.0 0 32768 i
*> 10.0.0.16/28 0.0.0.0 0 32768 i
*> 10.0.0.32/28 0.0.0.0 0 32768 i
*> 10.0.0.48/28 0.0.0.0 0 32768 i
*> 10.0.0.64/28 0.0.0.0 0 32768 i
*> 10.0.1.0/28 172.16.12.11 0 0 100 i
*> 10.0.1.16/28 172.16.12.11 0 0 100 i
*> 10.0.1.32/28 172.16.12.11 0 0 100 i
*> 10.0.1.48/28 172.16.12.11 0 0 100 i
*> 10.0.1.64/28 172.16.12.11 0 0 100 i
*> 10.0.1.80/28 172.16.12.11 0 0 100 i
<...output omitted...>

```

You should see that network 10.0.1.0/48 is not marked as dampened anymore.



# Summary

This topic summarizes the key points that were 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

257

## Summary (Cont.)

- You can implement route dampening with the **bgp dampening** command either globally in the BGP process or selectively using a route-map.
- 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

© 2015 Cisco and/or its affiliates. All rights reserved.

258



# Lesson 5: Module Summary

---

## Overview

This topic summarizes the key points that were 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 PMTU discovery, BGP PIC, BFD, as well as by adjusting BGP timers and intervals..
- 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.

© 2015 Cisco and/or its affiliates. All rights reserved.

259

## References

For additional information, refer to these resources:

- Cisco Systems, Inc. *BGP Restart Neighbor Session After Maximum-Prefix Limit Reached*.  
[http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute\\_bgp/configuration/15-mt/irg-15-mt-book/BGP-Sub-Codes-for-BGP-Cease-Notification.pdf](http://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/configuration/15-mt/irg-15-mt-book/BGP-Sub-Codes-for-BGP-Cease-Notification.pdf)

- Cisco Systems, Inc. *BGP Peer Groups*.  
<http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13755-29.html>
- Cisco Systems, Inc. *BGP Case Studies* "BGP Case Studies 4."  
<http://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/26634-bgp-toc.html#case4>

# Lesson 6: 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.

1. What are three characteristics of a converged BGP network? (Choose three.)
  - 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.
  
2. 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.
  
3. 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.

4. 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**

5. Look at the output of the **show interfaces Ethernet0/0** command. How has the interface been modified to improve BGP convergence? (Source: "Improving BGP Convergence")

router# **show interfaces Ethernet0/0**

Ethernet0/0 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

- 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. PMTU discovery has been enabled by setting the interface MSS to 1460 bytes.

6. Refer to the output. Which two parameters would indicate that the BGP network has converged? (Choose two.) (Source: "Improving BGP Convergence")

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            |

- 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.
7. Refer to the output. How frequently the BGP scanner process will run on the router? (Source: "Improving BGP Convergence")

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            |

- 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.

8. 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.
9. Which option can be used to support traffic forwarding during RP switchover? (Source: "Improving BPG Convergence")
- A. Nonstop Forwarding
  - B. Cisco Express Forwarding
  - C. Perfect Forwarding Routing
10. Lowering the keepalive timer between sent keepalive messages to a BGP neighbor will improve BGP convergence. True or false? (Source: "Improving BGP Convergence")
- A. True
  - B. False
11. 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.
12. 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.
13. 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.



14. 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.
15. 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.
16. 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.
17. 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.
18. 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.

19. Which two of the following statements accurately describe the function of the BGP 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.
20. 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
21. 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**
22. 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 effect on other well-behaved routes
23. 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
24. 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.

25. What are two methods of enabling route dampening on a Cisco router? (Choose two.) (Source: "Using BGP Route Dampening", Topic: "Configuring 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
26. 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 Answers

---

## Answer Key

1. A, C, D
2. B, D
3. C
4. B, C
5. D
6. A, C
7. C
8. A, B
9. A
10. A
11. A, B, D
12. A, C
13. A, B
14. A
15. B
16. A, C
17. A, C
18. B, C
19. C, D
20. A, B, D
21. D
22. A, C
23. B, D

24. A, D

25. B, D

26. B, C

# Glossary

## **ACK**

acknowledgment. Notification sent from one network device to another to acknowledge that some event occurred (for example, the receipt of a message).

## **AD**

administrative distance. Rating of the trustworthiness of a routing information source. Administrative distance often is expressed as a numerical value between 0 and 255. The higher the value, the lower the trustworthiness rating.

## **AS**

autonomous system. A collection of networks under a common administration sharing a common routing strategy. Autonomous systems are subdivided by areas. An autonomous system must be assigned a unique 16-bit number by the IANA.

## **ATM**

Asynchronous Transfer Mode. The international standard for cell relay in which multiple service types (such as voice, video, or data) are conveyed in fixed-length (53-byte) cells. Fixed-length cells allow cell processing to occur in hardware, thereby reducing transit delays. ATM is designed to take advantage of high-speed transmission media, such as E3, SONET, and T3.

## **BFD**

Bidirectional Forwarding Detection. A detection protocol that is designed to provide fast forwarding-path failure detection times for media types, encapsulations, topologies, and routing protocols.

## **BGP**

Border Gateway Protocol. Interdomain routing protocol that replaces EGP. BGP exchanges reachability information with other BGP systems. It is defined by RFC 1163.

## **CE**

customer edge. Identifies the network devices, connected to a provider network, that are under the administrative control of the customer.

## **CPU**

central processing unit. The hardware within a computer system or smartphone that carries out the instructions of a computer program by performing the basic arithmetical, logical, and input-output operations of the system.

## **DF**

Do not Fragment

## **DMZ**

demilitarized zone.

## **DNS**

Domain Name System. System used on the Internet for translating names of network nodes into addresses.

## **DPT**

Dynamic Packet Transport

## **DSL**

digital subscriber line. Public network technology that delivers high bandwidth over conventional copper wiring at limited distances. There are four types of DSL: ADSL, HDSL, SDSL, and VDSL. All are provisioned via modem pairs, with one modem located at a central office and the other at the customer site. Because most DSL technologies do not use the whole bandwidth of the twisted pair, there is room remaining for a voice channel.

## **EBGP**

Exterior Border Gateway Protocol

## **EGP**

Exterior Gateway Protocol. It's the Internet protocol for exchanging routing information between autonomous systems. Documented in RFC 904. This is not to be confused with the general term *exterior gateway protocol*. EGP is an obsolete protocol that was replaced by BGP.

## **EIGRP**

Enhanced Interior Gateway Routing Protocol. It's the advanced version of IGRP developed by Cisco. It provides superior convergence properties and operating efficiency, and it combines the advantages of link-state protocols with those of distance vector protocols.

## **FDDI**

Fiber Distributed Data Interface. LAN standard, defined by ANSI X3T9.5, specifying a 100-Mbps token-passing network using fiber-optic cable, with transmission distances of up to 2 km. FDDI uses a dual-ring architecture to provide redundancy.

## **FTTx**

Fiber From The "X", where X stands for different type of deployment (Building, Home, Premises, etc.)

## **FIB**

forwarding information base.

## **Frame Relay**

Industry-standard, packet-switched data link layer protocol that handles multiple virtual circuits between connected devices.

## **Gigabit Ethernet**

Standard for a high-speed Ethernet, approved by the IEEE (Institute of Electrical and Electronics Engineers) 802.3z standards committee in 1996.

## **IBGP**

Internal Border Gateway Protocol.

## **ICMP**

Internet Control Message Protocol. Network layer Internet protocol that reports errors and provides other information that is relevant to IP packet processing. Documented in RFC 792.



## **IGP**

interior gateway protocol. Internet protocol used to exchange routing information within an autonomous system. Examples of common Internet IGPs include IGRP, OSPF, and RIP.

## **IOL**

Special build of Cisco IOS Software for Linux, created specially for virtualized environments.

## **IP address**

A 32-bit address assigned to hosts using TCP/IP. An IP address belongs to one of five classes (A, B, C, D, or E) and is written as 4 octets separated by periods (dotted decimal format). Each address consists of a network number, an optional subnetwork number, and a host number. The network and subnetwork numbers together are used for routing, and the host number is used to address an individual host within the network or subnetwork. A subnet mask is used to extract network and subnetwork information from the IP address. CIDR provides a new way of representing IP addresses and subnet masks. Also called an Internet address.

## **IPv4**

IP version 4. Internet Protocol version 4 is the fourth version in the development of the IP and the first version of the protocol to be widely deployed. Along with IPv6, IPv4 is at the core of standards-based internetworking methods of the Internet. IPv4 is still used to route most traffic across the Internet. IPv4 is a connectionless protocol for use on packet-switched link layer networks (for example, Ethernet). It operates on a best-effort delivery model in that it does not guarantee delivery and does not assure proper sequencing or avoidance of duplicate delivery.

## **IPv6**

IP version 6. Replacement for the current version of IP (version 4). IPv6 includes support for flow ID in the packet header, which can be used to identify flows. Formerly called IPng (next generation).

## **IS-IS**

Intermediate System-to-Intermediate System. OSI link-state hierarchical routing protocol based on DECnet Phase V routing, whereby ISs (routers) exchange routing information based on a single metric to determine network topology.

## **ISP**

Internet service provider. Company that provides Internet access to other companies and individuals.

## **LAN**

local-area network. A high-speed, low-error data network covering a relatively small geographic area (up to a few thousand meters). LANs connect workstations, peripherals, terminals, and other devices in a single building or other geographically limited area. LAN standards specify cabling and signaling at the physical and data link layers of the OSI model. Ethernet, FDDI, and Token Ring are widely used LAN technologies.

## **LSP**

label-switched path.

**MAC address**

a standardized data link layer address that is required for every port or device that connects to a LAN. Other devices in the network use these addresses to locate specific ports in the network and to create and update routing tables and data structures. A MAC address is 6 bytes long and is controlled by the IEEE. It is also known as a hardware address, MAC layer address, and physical address.

**MD5**

Message Digest 5. A one-way hashing algorithm that produces a 128-bit hash. Both MD5 and Secure Hash Algorithm (SHA) are variations on MD4 and are designed to strengthen the security of the MD4 hashing algorithm. Cisco uses hashes for authentication within the IPsec framework. MD5 is also used for message authentication in SNMP v.2. MD5 verifies the integrity of the communication, authenticates the origin, and checks for timeliness.

**MED**

multi-exit discriminator.

**MPLS**

Multiprotocol Label Switching. a switching method that forwards IP traffic using a label. This label instructs the routers and the switches in the network where to forward the packets based on pre-established IP routing information.

**MSS**

maximum segment size. A TCP parameter that specifies the maximum amount of data that a device can receive in a single TCP segment.

**MTU**

maximum transmission unit. The maximum packet size, in bytes, that a particular interface can handle.

**NAT**

Network Address Translation. A mechanism for reducing the need for globally unique IP addresses. NAT allows an organization with addresses that are not globally unique to connect to the Internet by translating these addresses into globally routable address space. Also known as Network Address Translator.

**NHT**

Next hop tracking

**NSF**

Nonstop Forwarding.

**OSPF**

Open Shortest Path First. Link-state, hierarchical IGP routing algorithm proposed as a successor to RIP in the Internet community. OSPF features include least-cost routing, multipath routing, and load balancing. OSPF was derived from an early version of the IS-IS protocol.

**P router**

provider router.

**PA**

provider-assigned.

**PA address**

provider-assigned address.

**PE**

provider edge. Identifies the network devices, under the administrative control of the provider, that connect to CE devices.

**PI**

provider-independent.

**PI address**

provider-independent address.

**PIC**

Prefix-Independent Convergence.

**PMTU**

path maximum transmission unit.

**P-network**

provider network.

**POI**

point of insertion.

**POP**

Post Office Protocol. Protocol that client email applications use to retrieve mail from a mail server.

**PVC**

permanent virtual circuit (or connection). Virtual circuit that is permanently established. PVCs save bandwidth associated with circuit establishment and tear down in situations where certain virtual circuits must exist all the time. In ATM terminology, called a permanent virtual connection.

**QoS**

quality of service. Measure of performance for a transmission system that reflects its transmission quality and service availability.

**RIB**

routing information base.

**RIP**

Routing Information Protocol. A distance-vector routing protocol that uses hop count as a routing metric.

**RP**

route processor.

**RT**

route target.

**SOO**

Site of Origin.

**SPD**

Selective Packet Discard.

**SSO**

Stateful Switchover.

**TCP**

Transmission Control Protocol. Connection-oriented transport layer protocol that provides reliable full-duplex data transmission. TCP is part of the TCP/IP protocol stack.

**TCP/IP**

Transmission Control Protocol/Internet Protocol. Common name for the suite of protocols developed by the U.S. DoD in the 1970s to support the construction of worldwide internetworks. TCP and IP are the two best-known protocols in the suite.

**TTL**

Time to Live. A mechanism that limits the lifespan or lifetime of data in a computer or network.

**VPN**

virtual private network. Enables IP traffic to travel securely over a public TCP/IP network by encrypting all traffic from one network to another. A VPN uses tunneling to encrypt all information at the IP level.

**VPNv4**

virtual private network version 4. Enables IP traffic to travel securely over a public TCP/IP network by encrypting all traffic from one network to another. A VPN uses tunneling to encrypt all information at the IP level.

**VRF**

a VPN routing/forwarding instance. A VRF consists of an IP routing table, a derived forwarding table, a set of interfaces that use the forwarding table, and a set of rules and routing protocols that determine what goes into the forwarding table. In general, a VRF includes the routing information that defines a customer VPN site that is attached to a PE router.

**WAN**

wide-area network. Data communications network that serves users across a broad geographic area and often uses transmission devices provided by common carriers. Frame Relay, SMDS, and X.25 are examples of WANs.