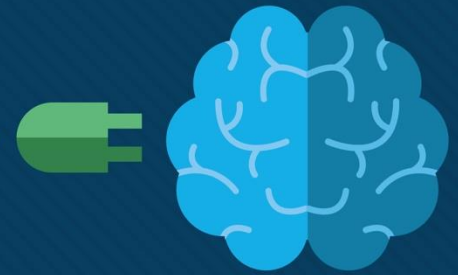

Chapter 13: BGP Path Selection

Instructor Materials

CCNP Enterprise: Advanced Routing

Chapter 13 Content

This chapter covers the following content:

- **Understanding BGP Path Selection -** This section reviews the first step of path selection, which involves selecting the longest prefix length.

- **BGP Best Path -** This section describes the logic used by BGP to identify the best path when multiple routes are installed in the BGP table.

- **BGP Equal-Cost Multipathing -** This section explains how additional paths are presented to the Routing Information Base (RIB) for installation into the routing table.

Understanding BGP Path Selection

- The BGP best-path selection algorithm influences how traffic enters or leaves an *autonomous system* (*AS*).
- Some router configurations modify the BGP attributes to influence inbound traffic, outbound traffic, or inbound and outbound traffic, depending on the network design requirements.
- Many network engineers do not understand the BGP best-path selection, which can often result in suboptimal routing.
- This section explains the logic used by a router that uses BGP when forwarding packets.

BGP Review

- With Border Gateway Protocol (BGP), route advertisements consist of the Network Layer Reachability Information (NLRI) and the path attributes (PAs).
- The NLRI consists of the network prefix and prefix length; BGP attributes such as AS_Path and origin are stored in the PAs.
- A BGP route may contain multiple paths to the same destination network.
- Every path's attributes impact the desirability of the route when a router selects the best path.
- A BGP router advertises only the best path to the neighboring routers.

BGP Recalculates the Best Path

Inside the BGP Loc-RIB table, all the routes and their path attributes are maintained with the best path calculated.

- The best path is then presented to the RIB for installation into the routing table of the router.

- If the best path is no longer available, the router uses the existing paths to quickly identify a new best path.

BGP recalculates the best path for a prefix upon four possible events:

- BGP next-hop reachability change
- Failure of an interface connected to an External BGP (eBGP) peer
- Redistribution change
- Reception of new or removed paths for a route

Router Path Selection

Routers always select the path a packet should take by examining the prefix length.

- The path selected for a packet depends on the prefix length, where the *longest prefix length* is always preferred.
- This logic is used to influence path selection in BGP.
- Path attributes (PAs) could be modified as they are advertised externally to influence the path taken.
- BGP routing policy in the service provider (SP) network could ignore path attributes.

Router Path Selection

To guarantee that paths to a company are selected deterministically outside the organization is to advertise a summary prefix (100.64.0.0/16) out both routers R1 and R2.

- Then advertise a longer matching prefix out the router for one prefix, and then advertise a longer matching prefix out the other router for the second prefix.
- This allows for traffic to enter a network in a deterministic manner while still providing a backup path to the other network in the event that the first router fails.
- Figure 13-1 shows this concept, with R1 advertising the 100.64.1.0/24 prefix, R2 advertising the 100.64.2.0/24 prefix, and both routers advertising the 100.64.0.0/16 summary network prefix.

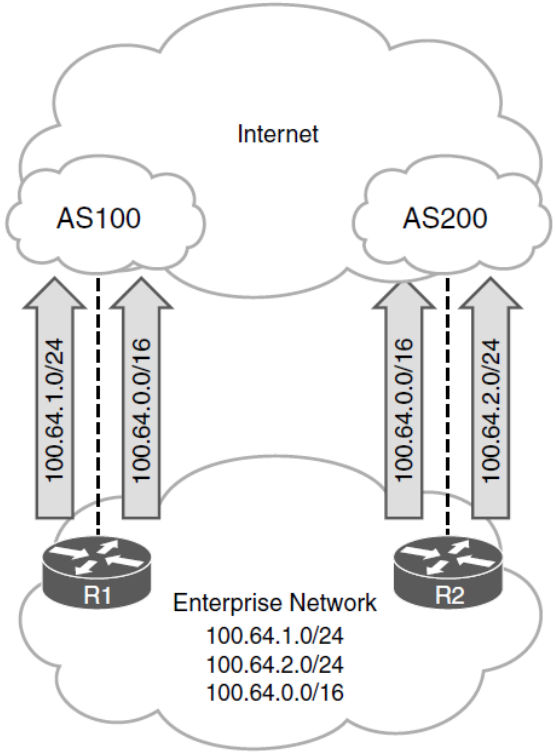

Figure 13-1 *BGP Path Selection Using the Longest Match*

BGP Best Path

- BGP installs the first received path as the best path automatically.
- When additional paths are received for the same network prefix length, the newer paths are compared against the current best path.
- If there is a tie, processing continues until a best path winner is identified.

BGP Best-path Algorithm

The following list provides the attributes that the BGP best-path algorithm uses for the process of selecting the best route. These attributes are processed in the order listed:

1. Prefer the highest *weight*
2. Prefer the highest *local preference*
3. Prefer the route *originated* by the local router
4. Prefer the path with the shorter Accumulated Interior Gateway Protocol (AIGP) metric attribute
5. Prefer the shortest *AS_Path*
6. Prefer the best *origin* code
7. Prefer the lowest multi-exit discriminator (*MED*)
8. Prefer an *external* path over an *internal* path
9. Prefer the path through the *closest IGP neighbor*
10. Prefer the *oldest route* for eBGP paths
11. Prefer the path with the lowest *neighbor BGP RID*
12. Prefer the path with the lowest *neighbor IP address*

Note: All BGP prefixes must pass the route validity check, and the next-hop IP address must be resolvable for the route to be eligible as a best path. Some vendors and publications consider this the first step.

BGP Best-path Algorithm (Cont.)

The BGP routing policy can vary, based on the manipulation of the BGP PAs.

- Because some PAs are transitive and carry from one AS to another AS, those changes could impact downstream routing for other SPs, too.
- Other PAs are non-transitive and influence the routing policy only within the organization.
- Network prefixes are conditionally matched on a variety of factors, such as AS_Path length, specific ASN, and BGP communities.
- Table 13-2 shows which BGP attributes must be supported by all BGP implementations and which BGP attributes are advertised between ASs.

Table 13-2 BGP Path Attribute Classifications

Name	Supported by All BGP Implementations	Advertised Between Autonomous Systems
Well-known mandatory	Yes	Yes
Well-known discretionary	Yes	No
Optional transitive	No	Yes
Optional nontransitive	No	No

Weight

BGP weight is a Cisco-defined attribute and the first step in selecting the BGP best path.

- Weight is a 16-bit value (0 through 65,535) assigned locally on the router; it is not advertised to other routers.
- The path with the higher weight is preferred.
- Weight can be set for specific routes with an inbound route map or for all routes learned from a specific neighbor. Weight is not advertised to peers and only influences outbound traffic from a router or an AS.
- Because it is the first step in the best-path algorithm, it should be used when other attributes should not influence the best path for a specific network prefix.

The command **set weight** *weight* in a route map sets the weight value for a matching prefix. The weight is set for all prefixes received by a neighbor using the BGP address family configuration command **neighbor** *ip-address* **weight** *weight*.

Weight (Cont.)

Figure 13-2 demonstrates the weight attribute and its influence on the BGP best-path algorithm:

- R4, R5, and R6 are in AS 400, with iBGP full mesh peering using loopback interfaces. AS 200 and AS 300 provide transit connectivity to AS 100.
- R4 is an edge router for AS 400 and sets the weight to 222 for the 172.16.0.0/24 prefix received from R2. This ensures that R4 uses R2 for outbound traffic to this prefix.
- R6 is an edge router for AS 400 and sets the weight to 333 for the 172.24.0.0/24 prefix received from R3. This ensures that R6 uses R3 for outbound traffic to this prefix.

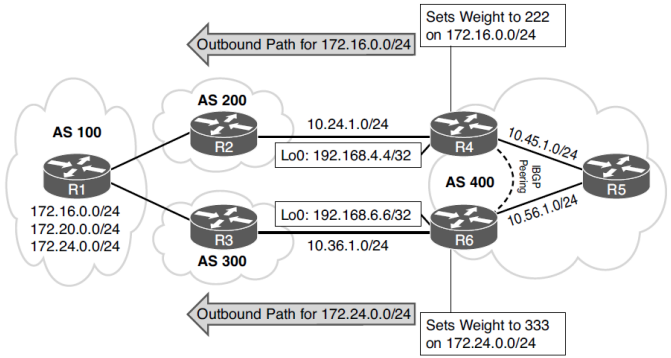

Figure 13-2 *BGP Weight Topology*

BGP Best path

Weight Example

Example 13-1 demonstrates the BGP configuration for manipulating the weight on R4 and R6. R4 uses the default IPv4 address family; R6 does not use the default IPv4 address family but does use BGP peer groups.

Example 13-1 *Weight Manipulation Configuration*

```
R4
ip prefix-list PRE172 permit 172.16.0.0/24
!
route-map AS200 permit 10
 match ip address prefix-list PRE172
 set weight 222
route-map AS200 permit 20
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200 in
 neighbor 192.168.5.5 remote-as 400
 neighbor 192.168.5.5 update-source Loopback0
 neighbor 192.168.5.5 next-hop-self
 neighbor 192.168.6.6 remote-as 400
 neighbor 192.168.6.6 update-source Loopback0
 neighbor 192.168.6.6 next-hop-self
```

```
R6
ip prefix-list PRE172 permit 172.24.0.0/24
!
route-map AS300 permit 10
 match ip address prefix-list PRE172
 set weight 333
route-map AS300 permit 20
!
router bgp 400
 no bgp default ipv4-unicast
 neighbor AS400 peer-group
 neighbor AS400 remote-as 400
 neighbor AS400 update-source Loopback0
 neighbor 10.36.1.3 remote-as 300
 neighbor 192.168.4.4 peer-group AS400
 neighbor 192.168.5.5 peer-group AS400
 !
 address-family ipv4
  neighbor AS400 next-hop-self
  neighbor 10.36.1.3 activate
  neighbor 10.36.1.3 route-map AS300 in
  neighbor 192.168.4.4 activate
  neighbor 192.168.5.5 activate
 exit-address-family
```

Weight Example BGP Table

Example 13-2 shows the BGP table for R4, R5, and R6.

- The weight is only set locally on R4 and R6.
- The weight was not advertised to any of the AS 400 routers and is set to 0 for all other prefixes.
- The > indicates the best path.
- BGP weight is locally significant.
- R4, R5, and R6 use other factors later in the best-path algorithm to select the best path for the prefixes that did not have the weight modified locally.

Example 13-2 *BGP Table After Weight Manipulation*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 * i 172.16.0.0/24    192.168.6.6              0    100      0 300 100 i
 *>                   10.24.1.2                             222 200 100 i
 * i 172.20.0.0/24    192.168.6.6              0    100      0 300 100 i
 *>                   10.24.1.2                             0 200 100 i
 * i 172.24.0.0/24    192.168.6.6              0    100      0 300 100 i
 *>                   10.24.1.2                             0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    100      0 200 100 i
 * i                  192.168.6.6              0    100      0 300 100 i
 *>i 172.20.0.0/24    192.168.4.4              0    100      0 200 100 i
 * i                  192.168.6.6              0    100      0 300 100 i
 *>i 172.24.0.0/24    192.168.4.4              0    100      0 200 100 i
 * i                  192.168.6.6              0    100      0 300 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 * i 172.16.0.0/24    192.168.4.4              0    100      0 200 100 i
 *>                   10.36.1.3                             0 300 100 i
 * i 172.20.0.0/24    192.168.4.4              0    100      0 200 100 i
 *>                   10.36.1.3                             0 300 100 i
 * i 172.24.0.0/24    192.168.4.4              0    100      0 200 100 i
 *>                   10.36.1.3                             333 300 100 i
```

Weight Example BGP Prefix

Example 13-3 shows R4's path information for the 172.16.0.0/24 network prefix.

- Notice that there are multiple paths and that the best path is through R2 because the weight is set to 222.
- The **show bgp ipv4 unicast** *network* command is extremely helpful for viewing and comparing BGP path attributes.

Example 13-3 *Viewing the BGP Prefix for Best-Path Selection*

```
R4# show bgp ipv4 unicast 172.16.0.0/24
BGP routing table entry for 172.16.0.0/24, version 4
Paths: (2 available, best #2, table default)
  Advertised to update-groups:
! Path #1
  Refresh Epoch 4
  300 100
    192.168.6.6 (metric 21) from 192.168.6.6 (192.168.6.6)
      Origin IGP, metric 0, localpref 100, valid, internal
! Path #2
  Refresh Epoch 2
  200 100
    10.24.1.2 from 10.24.1.2 (192.168.2.2)
      Origin IGP, localpref 100, weight 222, valid, external, best
```

Local Preference

Local preference (LOCAL_PREF) is a well-known discretionary path attribute and is included with path advertisements throughout an AS.

- The local preference attribute is a 32-bit value (0 through 4,294,967,295) that indicates the preference for exiting the AS.
- The local preference is not advertised between eBGP peers.
- Set for specific routes using a route map or all routes received from a neighbor.
- A higher value is preferred over a lower value.
- If an edge BGP router does not define the local preference upon receipt of a prefix, the default local preference value of 100 is used.
- You can change the default local preference value from 100 to a different value by using the command **bgp default local-preference** *default-local-preference*.

Local Preference Topology

Figure 13-3 demonstrates modification of the local preference to influence the traffic flow for prefixes 172.24.0.0/24 and 172.16.0.0/24:

- R4, R5, and R6 are in AS 400, with iBGP full mesh peering using loopback interfaces. AS 200 and AS 300 provide transit connectivity to AS 100.
- R4 is an edge router for AS 400 and sets the local preference to 222 for the 172.16.0.0/24 prefix received from R2, making it the preferred path for AS 400.
- R6 is an edge router for AS 400 and sets the local preference to 333 for the 172.24.0.0/24 prefix received from R3, making it the preferred path for AS 400.

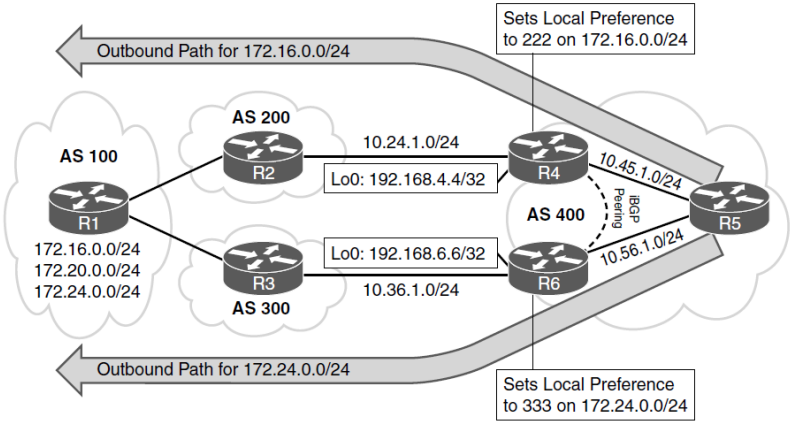

Figure 13-3 *BGP Local Preference Topology*

Local Preference Configuration

Example 13-4 demonstrates the BGP configuration for modifying the local preference on R4 and R6.

Example 13-4 *BGP Local Preference Configuration*

```
R4
ip prefix-list PRE172 permit 172.16.0.0/24
!
route-map AS200 permit 10
 match ip address prefix-list PRE172
 set local-preference 222
route-map AS200 permit 20
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200 in
 neighbor 192.168.5.5 remote-as 400
 neighbor 192.168.5.5 update-source Loopback0
 neighbor 192.168.5.5 next-hop-self
 neighbor 192.168.6.6 remote-as 400
 neighbor 192.168.6.6 update-source Loopback0
 neighbor 192.168.6.6 next-hop-self
```

```
R6
ip prefix-list PRE172 permit 172.24.0.0/24
!
route-map AS300 permit 10
 match ip address prefix-list PRE172
 set local-preference 333
route-map AS300 permit 20
!
router bgp 400
 no bgp default ipv4-unicast
 neighbor AS400 peer-group
 neighbor AS400 remote-as 400
 neighbor AS400 update-source Loopback0
 neighbor 10.36.1.3 remote-as 300
 neighbor 192.168.4.4 peer-group AS400
 neighbor 192.168.5.5 peer-group AS400
 !
 address-family ipv4
  neighbor AS400 next-hop-self
  neighbor 10.36.1.3 activate
  neighbor 10.36.1.3 route-map AS300 in
  neighbor 192.168.4.4 activate
  neighbor 192.168.5.5 activate
```

Local Preference Example

Example 13-5 shows the BGP table for R4, R5, and R6.

In Example 13-5, a network engineer might see that only one path exists on R4 for the 172.16.0.0/24 network prefix, and think R4 deleted the path through AS 300 because it was inferior to the path through AS 200. However, this is not what has happened in this example.

Example 13-5 *R4, R5, and R6 BGP Tables After Local Preference Modification*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24    10.24.1.2                222       0 200 100 i
 * i 172.20.0.0/24    192.168.6.6              0    100       0 300 100 i
 *>                   10.24.1.2                              0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              0    333       0 300 100 i
 *                    10.24.1.2                              0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    222       0 200 100 i
 * i 172.20.0.0/24    192.168.6.6              0    100       0 300 100 i
 *>i                  192.168.4.4              0    100       0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              0    333       0 300 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    222       0 200 100 i
 *                    10.36.1.3                              0 300 100 i
 * i 172.20.0.0/24    192.168.4.4              0    100       0 200 100 i
 *>                   10.36.1.3                              0 300 100 i
 *>  172.24.0.0/24    10.36.1.3                333       0 300 100 i
```

Phase I: Initial BGP Edge Route Processing

Phase I is the phase when routes are initially processed by the BGP edge routers R4 and R6.

This is what happens with R4:

- R4 receives the prefix for 172.16.0.0/24 from R2 and sets the local preference to 222.
- R4 receives the 172.20.0.0/24 and 172.24.0.0/24 prefixes from R2.
- No other paths exist for these prefixes, so all paths are marked as best paths.
- R4 advertises these paths to R5 and R6. (Routes without local preference set are advertised with the local preference 100.)

This is what happens with R6:

- R6 receives the prefix for 172.24.0.0/24 from R3 and sets the local preference to 333.
- R6 receives the 172.16.0.0/24 and 172.20.0.0/24 prefixes from R3.
- No other paths exist for these prefixes, so all paths are marked as best paths.
- R6 advertises these paths to R4 and R5. (Routes without local preference set are advertised with the local preference 100.)

Phase I: Initial BGP Edge Route Processing (Cont.)

Example 13-6 shows the BGP tables on R4 and R6 during this phase.

- Notice the local preference for 172.16.0.0/24 on R4 and the 172.24.0.0/24 network prefix on R6.
- No other entries have values populated for local preference.

Example 13-6 *BGP Table After Phase I Processing*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24     10.24.1.2                222      0 200 100 i
 *>  172.20.0.0/24     10.24.1.2                         0 200 100 i
 *>  172.24.0.0/24     10.24.1.2                         0 200 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24     10.36.1.3                         0 300 100 i
 *>  172.20.0.0/24     10.36.1.3                         0 300 100 i
 *>  172.24.0.0/24     10.36.1.3                333      0 300 100 i
```

Phase II: BGP Edge Evaluation of Multiple Paths

Phase II is the phase when R4 and R6 have received each other's routes and compare each path for a prefix.

- R6 advertises a route withdrawal for the 172.16.0.0/24 network prefix, and R4 advertises a route withdrawal for the 172.24.0.0/24 network prefix.
- R5 receives routes from R4 and R6 at the same time, resulting in both paths being present in the BGP Adj-RIB table.
- Example 13-7 shows the BGP tables for R4, R5, and R6 after Phase II processing.

Example 13-7 *BGP Table After Phase II Processing*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>  172.16.0.0/24    10.24.1.2                222       0 200 100 i
* i                  192.168.6.6              0    100       0 200 100 i
*>  172.20.0.0/24    10.24.1.2                              0 200 100 i
* i                  192.168.6.6              0    100       0 200 100 i
*   172.24.0.0/24    10.24.1.2                              0 200 100 i
*>i                  192.168.6.6              0    333       0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>i 172.16.0.0/24    192.168.4.4              0    222       0 200 100 i
* i                  192.168.4.4              0    100       0 200 100 i
* i 172.20.0.0/24    192.168.6.6              0    100       0 300 100 i
*>i                  192.168.4.4              0    100       0 200 100 i
*>i 172.24.0.0/24    192.168.6.6              0    333       0 300 100 i
* i                  192.168.4.4              0    100       0 200 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*   172.16.0.0/24    10.36.1.3                              0 300 100 i
*>i                  192.168.4.4              0    222       0 200 100 i
*>  172.20.0.0/24    10.36.1.3                              0 300 100 i
* i                  192.168.4.4              0    100       0 200 100 i
*>  172.24.0.0/24    10.36.1.3                333           0 300 100 i
* i                  192.168.4.4              0    100       0 200 100 i
```

Phase III: Final BGP Processing State

Phase III is the last processing phase. In this topology, R4, R5, and R6 process all the route withdrawals.

In this phase:

- R4 and R5 receive R6's withdrawal for the 172.16.0.0/24 network prefix and remove it from the BGP table.
- R5 and R6 receive R4's withdrawal for the 172.24.0.0/24 network prefix and remove it from the BGP table.

Example 13-8 shows the BGP tables for R4, R5, and R6 after Phase III processing.

Example 13-8 *BGP Table After Phase III Processing*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>   172.16.0.0/24    10.24.1.2                222       0 200 100 i
* i  172.20.0.0/24    192.168.6.6              0    100       0 300 100 i
*>                    10.24.1.2                              0 200 100 i
*>i  172.24.0.0/24    192.168.6.6              0    333       0 300 100 i
*                     10.24.1.2                              0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>i  172.16.0.0/24    192.168.4.4              0    222       0 200 100 i
* i  172.20.0.0/24    192.168.6.6              0    100       0 300 100 i
*>i                   192.168.4.4              0    100       0 200 100 i
*>i  172.24.0.0/24    192.168.6.6              0    333       0 300 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>i  172.16.0.0/24    192.168.4.4              0    222       0 200 100 i
*                     10.36.1.3                              0 300 100 i
* i  172.20.0.0/24    192.168.4.4              0    100       0 200 100 i
*>                    10.36.1.3                              0 300 100 i
*>   172.24.0.0/24    10.36.1.3                333           0 300 100 i
```

Locally Originated in the Network or Aggregate Advertisement

The third decision point in the best-path algorithm is to determine whether the route originated locally.

Preference is given in the following order:

1. Routes that were advertised locally
2. Networks that have been aggregated locally
3. Routes received by BGP peers

Accumulated Interior Gateway Protocol (AIGP)

Accumulated Interior Gateway Protocol (AIGP) is an optional nontransitive path attribute that is included with advertisements throughout an AS.

- AIGP provides the ability for BGP to maintain and calculate a conceptual path metric in environments that use multiple ASs with unique IGP routing domains in each AS.
- In Figure 13-4, AS 100, AS 200, and AS 300 are all under the control of the same service provider.

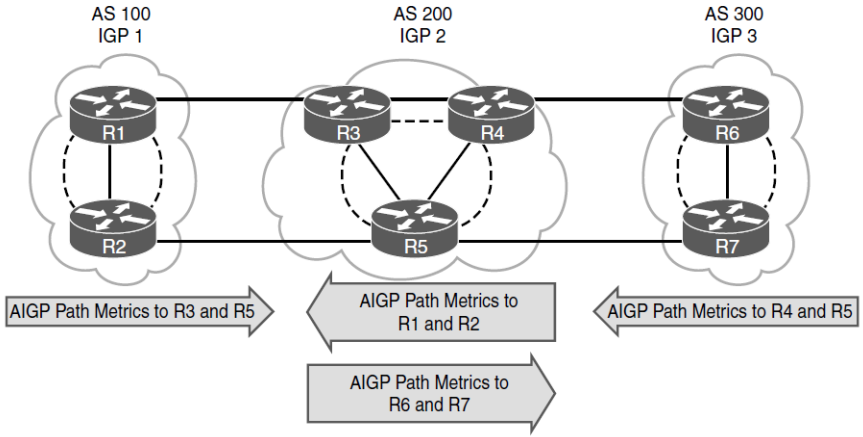

Figure 13-4 *AIGP Path Attribute Exchange Between Autonomous Systems*

BGP Best path

AIGP Metrics

The following guidelines apply to AIGP metrics:

- A path with an AIGP metric is preferred to a path without an AIGP metric.
- If the next-hop address requires a recursive lookup, the AIGP path needs to calculate a derived metric to include the distance to the next-hop address. This ensures that the cost to the BGP edge router is included. The formula is:

  Derived AIGP Metric = (Original AIGP Metric + Next-Hop AIGRP Metric)

  - If multiple AIGP paths exist and one next-hop address contains an AIGP metric and the other does not, the non-AIGP path is not used.
  - The next-hop AIGP metric is recursively added if multiple lookups are performed.
- AIGP paths are compared based on the derived AIGP metric (with recursive next hops) or the actual AIGP metric (nonrecursive next hop). The path with the lower AIGP metric is preferred.
- When a router R2 advertises an AIGP-enabled path that was learned from R1, if the next-hop address changes to an R2 address, R2 increments the AIGP metric to reflect the distance (the IGP path metric) between R1 and R2.

BGP Best path

Shortest AS_Path

The next decision factor for the BGP best-path algorithm is the AS_Path length, which typically correlates to the AS hop count.
A shorter AS_Path is preferred over a longer AS_Path.

Figure 13-5 demonstrates how AS_Path prepending influences outbound traffic pattern.

When working with confederations, AS_CONFED_SEQUENCE (confederation AS_Path) is not counted, and for aggregated addresses with multiple autonomous system numbers (ASNs) under the AS_SET portion of AS_Path, the AS_SET counts for only one AS_Path entry.

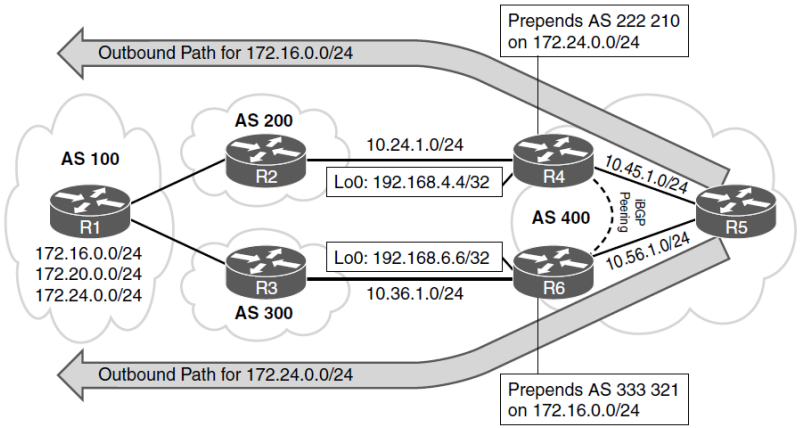

Figure 13-5 *Configuration for Modifying BGP AS_Path*

BGP Best path

Shortest AS_Path Configuration

Example 13-9 shows R4's and R6's configuration for prepending AS_Path on R4 and R6.

Example 13-10 shows the BGP tables for R4, R5, and R6.

Example 13-10 *BGP Tables After AS_Path Prepending*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24    10.24.1.2                              0 200 100 i
 *>  172.20.0.0/24    10.24.1.2                              0 200 100 i
 * i                  192.168.6.6              0    100       0 300 100 i
 *   172.24.0.0/24    10.24.1.2                              0 222 210 200 100 i
 *>i                  192.168.6.6              0    100       0 300 100 i
```

```
R5# show bgp ipv4 unicast | begin Network
 *>i                  192.168.6.6              0    100       0 300 100 i
 *>i 172.24.0.0/24    192.168.6.6              0    100       0 300 100 i
```

```
R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    100       0 200 100 i
 *                    10.36.1.3                              0 333 321 300 100 i
 * i 172.20.0.0/24    192.168.4.4              0    100       0 200 100 i
 *>                   10.36.1.3                              0 300 100 i
 *>  172.24.0.0/24    10.36.1.3                              0 300 100 i
```

Example 13-9 *BGP AS_Path Prepending Configuration*

```
R4
ip prefix-list PRE172 permit 172.24.0.0/24
!
route-map AS200 permit 10
 match ip address prefix-list PRE172
 set as-path prepend 222 210
route-map AS200 permit 20
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200 in
```

```
R6
ip prefix-list PRE172 permit 172.16.0.0/24
!
route-map AS300 permit 10
 match ip address prefix-list PRE172
 set as-path prepend 333 321
route-map AS300 permit 20
!
router bgp 400
 neighbor 10.36.1.3 remote-as 300
 neighbor 10.36.1.3 route-map AS300 in
```

BGP Best path

Origin Type

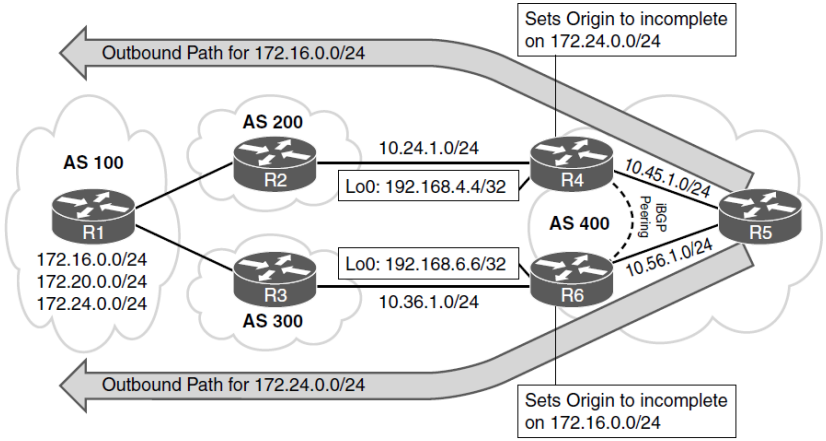

Figure 13-6 *BGP Origin Topology*

The next BGP best-path decision factor is the mandatory BGP attribute named origin.
By default, networks that are advertised on Cisco routers using the network statement are set with the *i* (for IGP) origin, and redistributed networks are assigned the *?* (incomplete) origin attribute.

The origin preference order is as follows:
1. IGP origin (Most)
2. Exterior Gateway Protocol (EGP) origin
3. Incomplete origin (Least)

You can modify a prefix's origin attribute by using the command **set origin** {**igp** | **incomplete**} on a route map. The EGP origin cannot be manually set on IOS XE routers.

Figure 13-6 demonstrates the modification of the origin attribute.

BGP Best path

Origin Type Configuration

Example 13-11 shows R4's and R6's configuration for modifying the BGP origin attribute on R4 and R6.

Example 13-12 shows the BGP tables for R4, R5, and R6.

A path with an incomplete origin is not selected as the best path because the IGP origin is preferred over the incomplete origin. Notice the origin codes (i and ?) on the far right, after the AS_Path information.

Example 13-11 *BGP Origin Manipulation Configuration*

```
R4
ip prefix-list PRE172 permit 172.24.0.0/24
!
route-map AS200 permit 10
 match ip address prefix-list PRE172
 set origin incomplete
route-map AS200 permit 20
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200 in
```

```
R6
ip prefix-list PRE172 permit 172.16.0.0/24
!
route-map AS300 permit 10
 match ip address prefix-list PRE172
 set origin incomplete
route-map AS300 permit 20
!
router bgp 400
 neighbor 10.36.1.3 remote-as 300
 neighbor 10.36.1.3 route-map AS300 in
```

Example 13-12 *BGP Table After Origin Manipulation*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24    10.24.1.2                              0 200 100 i
 * i 172.20.0.0/24    192.168.6.6              0    100      0 300 100 i
 *>                   10.24.1.2                              0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              0    100      0 300 100 i
 *                    10.24.1.2                              0 200 100 ?
```

```
R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    100      0 200 100 i
 * i 172.20.0.0/24    192.168.4.4              0    100      0 200 100 i
 *>i                  192.168.6.6              0    100      0 300 100 i
 *>i 172.24.0.0/24    192.168.6.6              0    100      0 300 100 i
```

```
R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              0    100      0 200 100 i
 *                    10.36.1.3                              0 300 100 ?
 * i 172.20.0.0/24    192.168.4.4              0    100      0 200 100 i
 *>                   10.36.1.3                              0 300 100 i
 *>  172.24.0.0/24    10.36.1.3                              0 300 100 i
```

BGP Best path

Multi-Exit Discriminator

The next BGP best-path decision factor is the non-transitive BGP multi-exit discriminator (MED) attribute.

- The MED uses a 32-bit value (0 to 4,294,967,295) called a *metric*. BGP sets the MED automatically to the IGP path metric during network advertisement or redistribution.
- Figure 13-7 demonstrates the concept in a simple topology.

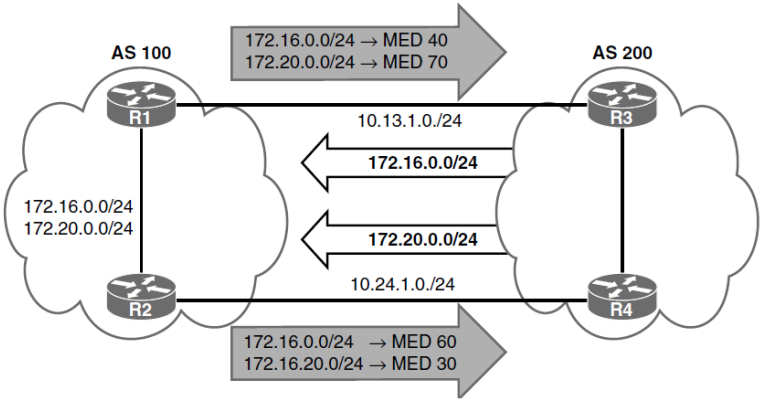

Figure 13-7 *MED Influencing Outbound Traffic*

Multi-Exit Discriminator (Cont.)

You can use an inbound route map to set the MED using the command **set metric** *metric*.

- Figure 13-8 revisits the best-path selection topology but now places R2 and R3 both in AS 200, which is essential for MED to work properly.

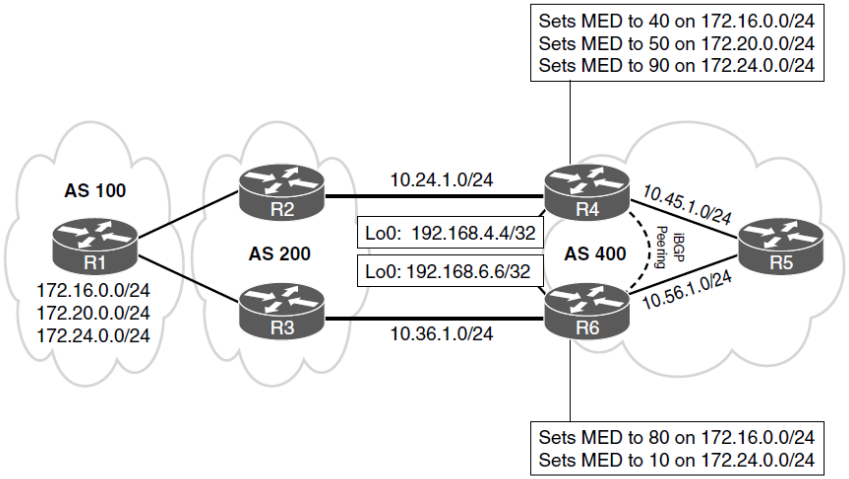

Figure 13-8 *BGP MED Manipulation*

Multi-Exit Discriminator (Cont.)

Example 13-13 shows the configuration for manipulating the MED on R4 and R6, based on the guidelines in Figure 13-8.

Example 13-13 *Configuration to Modify Inbound MED Modification*

R4

```
ip prefix-list PRE172-01 permit 172.16.0.0/24
ip prefix-list PRE172-02 permit 172.20.0.0/24
ip prefix-list PRE172-03 permit 172.24.0.0/24
!
route-map AS200-R2 permit 10
 match ip address prefix-list PRE172-01
 set metric 40
route-map AS200-R2 permit 20
 match ip address prefix-list PRE172-02
 set metric 50
route-map AS200-R2 permit 30
 match ip address prefix-list PRE172-03
 set metric 90
route-map AS200-R2 permit 40
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200-R2 in
```

R6

```
ip prefix-list PRE172-01 permit 172.16.0.0/24
ip prefix-list PRE172-03 permit 172.24.0.0/24
!
route-map AS200-R3 permit 10
 match ip address prefix-list PRE172-01
 set metric 80
route-map AS200-R3 permit 20
 match ip address prefix-list PRE172-03
 set metric 10
route-map AS200-R3 permit 30
!
router bgp 400
 neighbor 10.36.1.3 remote-as 200
 neighbor 10.36.1.3 route-map AS200-R3 in
```

BGP Best path

Multi-Exit Discriminator BGP Tables

Example 13-14 shows the BGP tables for R4, R5, and R6.

All three AS 400 routers send traffic toward the 172.16.0.0/24 network prefix through R4's link to R2 because 40 is lower than 80, and all three AS 400 routers send traffic toward the 172.24.0.0/24 network prefix through R6's link to R3 because 10 is lower than 90.

Example 13-14 *R4, R5, and R6 BGP Tables After MED Modification*

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24    10.24.1.2                40             0 200 100 i
 *>i 172.20.0.0/24    192.168.6.6               0    100       0 200 100 i
 *                    10.24.1.2                50             0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              10    100       0 200 100 i
 *                    10.24.1.2                90             0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              40    100       0 200 100 i
 *>i 172.20.0.0/24    192.168.6.6               0    100       0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              10    100       0 200 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              40    100       0 200 100 i
 *                    10.36.1.3                80             0 200 100 i
 *>  172.20.0.0/24    10.36.1.3                               0 200 100 i
 *>  172.24.0.0/24    10.36.1.3                10             0 200 100 i
```

BGP Best path

Inbound MED Modification

An organization may expect its different SPs to advertise a MED value for every prefix.

- If a MED is missing, the path without a MED is preferred over a path that contains a MED.
- An organization can modify the default behavior so that prefixes without a MED are always selected last.

In Example 13-13, R6's route map is configured to not set the MED on the 172.20.0.0/24 prefix when received by R3. When the MED is not advertised, the value is assumed to be zero (0).

Example 13-13 *Configuration to Modify Inbound MED Modification*

R4

```
ip prefix-list PRE172-01 permit 172.16.0.0/24
ip prefix-list PRE172-02 permit 172.20.0.0/24
ip prefix-list PRE172-03 permit 172.24.0.0/24
!
route-map AS200-R2 permit 10
 match ip address prefix-list PRE172-01
 set metric 40
route-map AS200-R2 permit 20
 match ip address prefix-list PRE172-02
 set metric 50
route-map AS200-R2 permit 30
 match ip address prefix-list PRE172-03
 set metric 90
route-map AS200-R2 permit 40
!
router bgp 400
 neighbor 10.24.1.2 remote-as 200
 neighbor 10.24.1.2 route-map AS200-R2 in
```

R6

```
ip prefix-list PRE172-01 permit 172.16.0.0/24
ip prefix-list PRE172-03 permit 172.24.0.0/24
!
route-map AS200-R3 permit 10
 match ip address prefix-list PRE172-01
 set metric 80
route-map AS200-R3 permit 20
 match ip address prefix-list PRE172-03
 set metric 10
route-map AS200-R3 permit 30
!
router bgp 400
 neighbor 10.36.1.3 remote-as 200
 neighbor 10.36.1.3 route-map AS200-R3 in
```

Missing MED Behavior BGP Tables

An organization may expect its different SPs to advertise a MED value for every prefix.

- If a MED is missing, the path without a MED is preferred over a path that contains a MED.
- An organization can modify the default behavior so that prefixes without a MED are always selected last.
- The command **bgp bestpath med missing-as-worst** is applied to R4, R5, and R6.
- Example 13-15 shows their BGP tables after the change is made. Notice that R6 sets the MED to 4,294,967,295 for the 172.20.0.0/24 route learned from R3.

Example 13-15 *R4, R5, and R6 BGP Tables with* med missing-as-worst

```
R4# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>  172.16.0.0/24    10.24.1.2                40             0 200 100 i
 *>  172.20.0.0/24    10.24.1.2                50             0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              10    100      0 200 100 i
 *                    10.24.1.2                90             0 200 100 i

R5# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
 *>i 172.16.0.0/24    192.168.4.4              40    100      0 200 100 i
 *>i 172.20.0.0/24    192.168.4.4              50    100      0 200 100 i
 *>i 172.24.0.0/24    192.168.6.6              10    100      0 200 100 i

R6# show bgp ipv4 unicast | begin Network
     Network          Next Hop            Metric LocPrf Weight Path
*>i  172.16.0.0/24    192.168.4.4              40    100      0 200 100 i
 *                   10.36.1.3                80             0 200 100 i
 *>i 172.20.0.0/24    192.168.4.4              50    100      0 200 100 i
 *                    10.36.1.3          4294967295           0 200 100 i
 *>  172.24.0.0/24    10.36.1.3                10             0 200 100 i
```

Always Compare MED

The default MED comparison mechanism requires the AS_Path values to be identical because the policies used to set the MED could vary from AS to AS.

- This means that the MED can influence traffic only when multiple links are from the same service provider.
- Typically, organizations use different service providers for redundancy. In these situations, the default BGP rules for MED comparison need to be relaxed to compare MEDs between different service providers.
- The **always-compare-med** feature allows for the comparison of MED regardless of the AS_Path. You enable this feature by using the BGP configuration command **bgp always-compare-med**.

Enable this feature on all BGP routers in the AS, or routing loops can occur.

BGP Deterministic MED

The best-path algorithm compares a route update to the existing best path and processes the paths in the order in which they are stored in the Loc-RIB table.

- The paths are stored in the order in which they are received in the BGP table. If **always-compare-med** is not enabled, the path MED is only compared against the existing best path and not against all the paths in the Loc-RIB table, which can cause variations in the MED best-path comparison process.
- Figure 13-9 demonstrates a topology in which the MED is not compared due to the order of the path advertisement:

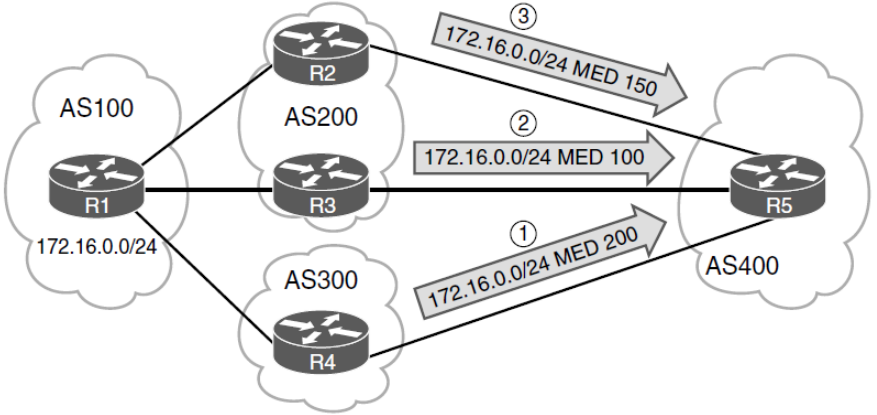

Figure 13-9 *Problems with MED Comparison*

BGP deterministic MED corrects the problem by grouping together paths with identical AS_Path values as part of the best-path identification process.
Each group's MED is compared against the other group's MED.

eBGP over iBGP

The next BGP best-path decision factor is whether the route comes from an iBGP, eBGP, or confederation member AS (Sub-AS) peering.

The best-path selection order is as follows:

1. eBGP peers (most desirable)
2. Confederation member AS peers
3. iBGP peers (least desirable)

Lowest IGP Metric

The next decision step is to use the lowest IGP cost to the BGP next-hop address.

Figure 13-10 illustrates a topology in which R2, R3, R4, and R5 are in AS 400.

- AS 400 peers in a full mesh and establishes BGP sessions using Loopback 0 interfaces. R1 advertises the 172.16.0.0/24 network prefix to R2 and R4.
- R3 prefers the path from R2 compared to the iBGP path from R4 because the metric to reach the next-hop address is lower.
- R5 prefers the path from R4 compared to the iBGP path from R2 because the metric to reach the next-hop address is lower.

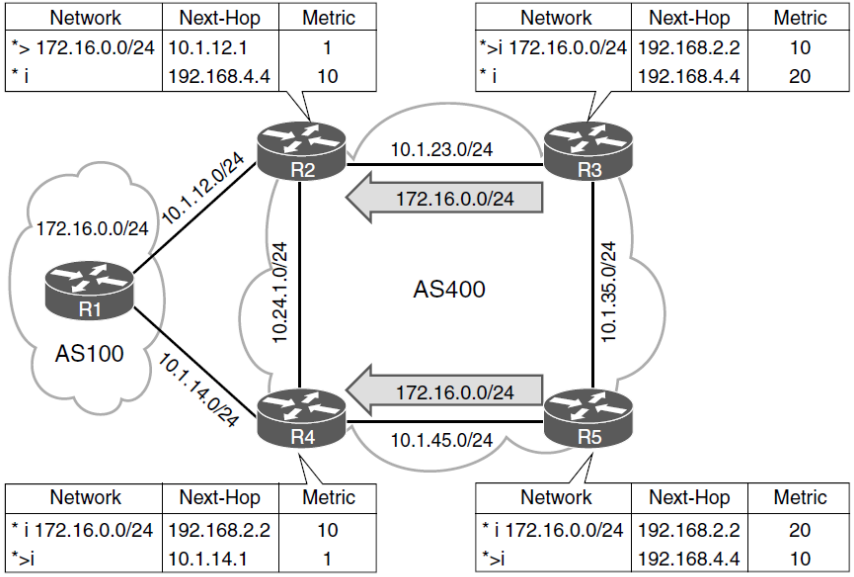

Figure 13-10 *Lowest IGP Metric Topology*

Prefer the Oldest EBGP Path/Router ID

BGP can maintain large routing tables, and unstable sessions result in the BGP best-path calculation executing frequently.

BGP maintains stability in a network by preferring the path from the oldest (established) BGP session. The downfall of this technique is that it does not lead to a deterministic method of identifying the BGP best path from a design perspective.

The next step for the BGP best-path algorithm is to select the best path using the lowest router ID of the advertising EBGP router.

If the route was received by a route reflector, then the originator ID is substituted for the router ID.

Minimum Cluster List Length

The next step in the BGP best-path algorithm is to select the best path using the lowest cluster list length.

- The cluster list is a non-transitive BGP attribute that is appended (not overwritten) by a route reflector with its cluster ID.
- Route reflectors use the **cluster-id** attribute as a loop-prevention mechanism.
- The cluster ID is not advertised between ASs and is locally significant.
- In simplest terms, this step locates the path that has traveled the smallest number of iBGP advertisement hops.

Figure 13-11 demonstrates how the minimum cluster list length is used as part of the BGP best-path calculation.

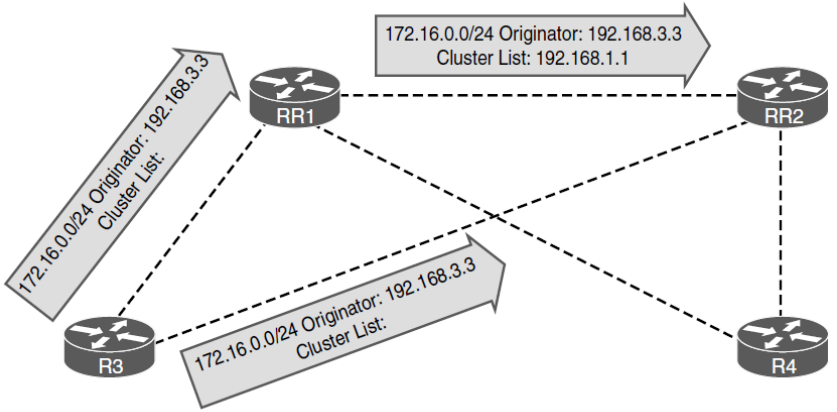

Figure 13-11 *Minimum Cluster List Length*

Lowest Neighbor Address

The last step of the BGP best-path algorithm involves selecting the path that comes from the lowest BGP neighbor address.

- This step is limited to iBGP peerings because eBGP peerings use the oldest received path as the tie breaker.
- Figure 13-12 demonstrates the concept of choosing the router with the lowest neighbor address.
- R1 is advertising the 172.16.0.0/24 network prefix to R2. R1 and R2 have established two BGP sessions using the 10.12.1.0/24 and 10.12.2.0/24 network prefixes. R2 selects the path advertised from 10.12.1.1 as it is the lower IP address.

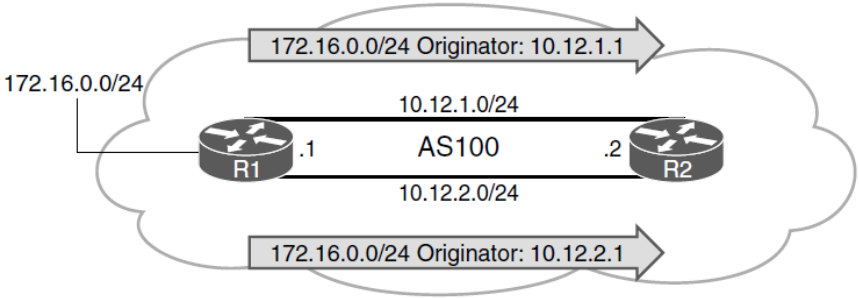

Figure 13-12 *Lowest IP Address*

BGP Equal Cost Multipath

- All the IGP routing protocols explained in this book support equal-cost multipath (ECMP).
- ECMP provides load balancing by installing multiple paths into the RIB for that protocol.
- BGP selects only one best path, but it allows for the installation of multiple routes into the
- RIB.
- BGP multipathing has three different variances in behavior, only the first two of which
- are discussed in this book:
    - eBGP multipath
    - iBGP multipath
    - eBGP and iBGP (eiBGP) multipath

- Enabling BGP multipathing does not alter the best-path algorithm or change the behavior of paths advertisement to other BGP peers. Only the BGP best path is advertised to peers.

BGP Multipath Variances

When you configure BGP multipathing, the additional paths need to match the following best-path BGP path attributes:

- Weight
- Local preference
- AS_Path length
- AS_Path content (although confederations can contain a different AS_CONFED_SEQ path)
- Origin
- MED
- Advertisement method (iBGP or eBGP) (If the prefix is learned from an iBGP advertisement, the IGP cost must match for iBGP and eBGP to be considered equal.)

Enable eBGP multipathing by using the BGP configuration command **maximum-paths** *number-paths*.

- The number of paths indicates the allowed number of eBGP paths to install in the RIB.
- The command **maximum-paths ibgp** *number-paths* sets the number of iBGP routes to install in the RIB. The commands are placed under the appropriate address family.

Prepare for the Exam

Prepare for the Exam

Key Topics for Chapter 13

Description	Description
BGP preference for the longest prefix length	Origin type
Use of summarization to direct traffic flows	Multi-exit discriminator
BGP best path	Missing MED behavior
BGP path attribute classifications	MED comparison
Weight	BGP deterministic MED
Local preference	eBGP over iBGP
Path removal with multiple paths	Lowest IGP metric
Local route origination	BGP equal-cost multipathing
Accumulated Interior Gateway Protocol (AIGP)	Multipathing requirements
Shortest AS_Path	

Key Terms for Chapter 13

Terms
BGP multipathing
Loc-RIB
optional transitive
optional non-transitive
well-known mandatory
well-known discretionary

Command Reference for Chapter 13

Task	Command Syntax
Set the weight in a route map	**set weight** *weight*
Set the weight for all routes learned from this neighbor	**neighbor** *ip-address* **weight** *weight*
Set the local preference in a route map	**set local-preference** *preference*
Set the local preference for all routes learned from this neighbor	**neighbor** *ip-address* **local-preference** *preference*
Enable the advertise of AIGP path attributes	**neighbor** *ip-address* **aigp**
Set the AIGP metric in a route map	**set aigp-metric {igp-metric |** *metric***}**
Set AS_Path prepending in a route map	**set as-path prepend** *as-number*
Set the origin using a route map	**set origin {igp | incomplete}**

Command Reference for Chapter 13

Task	Command Syntax
Set the MED using a route map	**set metric** *metric*
Set the MED to infinity when the MED is not present	**bgp bestpath med missing-as-worst**
Set the MED to the default value when the MED is not present	**default-metric** *metric*
Ensure that MED is always compared, regardless of AS_Path	**bgp always-compare-med**
Group together paths with identical AS_Path values as part of the best-path identification process	**bgp deterministic-med**
Configure eBGP multipathing	**maximum-paths** *number-paths*
Configure iBGP multipathing	**maximum-paths ibgp** *number-paths*


CISCO