CCIE Enterprise Infrastructure Foundation

ciscopress.com

Narbik Kocharians, CCIE® No. 12410

CCIE Enterprise Infrastructure Foundation

Narbik Kocharians

Cisco Press

After you configure this command, the internal path from R11 and the external paths from R2 and R3 are both designated as multipath in the BGP table, as shown above. These paths are once again installed in the routing table, as shown in the output below:

```
R10#show ip route 130.7.1.1

Routing entry for 130.7.1.1/32
  Known via "bgp 400", distance 20, metric 0
  Tag 300, type external
  Last update from 200.2.10.2 00:00:55 ago
  Routing Descriptor Blocks:
  * 200.3.10.3, from 200.3.10.3, 00:00:55 ago
      Route metric is 0, traffic share count is 1
      AS Hops 1
      Route tag 300
      MPLS label: none
    200.2.10.2, from 200.2.10.2, 00:00:55 ago
      Route metric is 0, traffic share count is 1
      AS Hops 1
      Route tag 300
      MPLS label: none
    11.11.11.11, from 11.11.11.11, 00:00:55 ago
      Route metric is 0, traffic share count is 1
      AS Hops 1
      Route tag 300
      MPLS label: none
```

The **maximum-paths eibgp** command includes equal internal and external paths in the calculation, as expressed above. Even though the multipath configuration includes both internal and external paths, there will still be only one best path. In this example, the best path is the path received from R2. All other multipath-capable paths are selected if they contain attributes that are equal to the path R10 received from R2.

Note When entering the **maximum-paths eibgp** command, the following output can be observed:

```
R10(config-router)#maximum-paths eibgp 3

%BGP: This may cause traffic loop if not used properly (command
accepted)
R10(config-router)#
*Jun 28 14:35:08.029: %BGP-4-MULTIPATH_LOOP: This may cause traffic
loop if not used properly (command accepted).
```

Continued

This message serves as a warning when you use the eiBGP multihop feature. In certain situations, such as when an iBGP path may have inconsistent next hops that lead back to the calculating router, loops may be formed when using this form of multipathing.

Note Even though AS_PATH SEQUENCE must match exactly for multipath decision making, the **bgp bestpath as-path multipath-relax** hidden command removes this consideration. When configured, the router will consider all paths with the same AS_PATH length—not necessarily the same SEQUENCE—as potential multipath candidates.

Without this command, the AS_PATH SEQUENCE of the multipath candidate must equal the AS_PATH SEQUENCE of the best path.

Internal Paths Finally, the **maximum-paths ibgp** command can be used to have the router select multipath paths for only internally learned paths. When selecting multipath paths for internal paths, the following additional attributes must be the same between the candidate multipath path and the current best path:

- Both paths must be learned from an internal neighbor.
- The IGP metric to the BGP next hop should be equal to the best path.

To demonstrate this feature, R12's BGP table for the 130.7.1.1 prefix is shown below:

On R12:

```
R12#show ip bgp 130.7.1.1

BGP routing table entry for 130.7.1.1/32, version 112
Paths: (2 available, best #1, table default)
  Advertised to update-groups:
     1
  Refresh Epoch 1
  300, (Received from a RR-client)
    10.10.10.10 (metric 11) from 10.10.10.10 (10.10.10.10)
      Origin IGP, metric 0, localpref 100, valid, internal, best
      rx pathid: 0, tx pathid: 0x0
  Refresh Epoch 1
  300, (Received from a RR-client)
    11.11.11.11 (metric 11) from 11.11.11.11 (11.11.11.11)
      Origin IGP, metric 0, localpref 100, valid, internal
      rx pathid: 0, tx pathid: 0
```

Here, R12 receives two iBGP paths to reach the prefix. It has marked its path from R10 as the best path because of the lower BGP RID. For the path through R11, however, all of the required path attributes for multipath are equal with R12's current best path through R10.

The **maximum-paths ibgp 2** command is configured under BGP router configuration mode on R12 to allow R12 to install the extra path through R11 as multipath in the RIB. The results are shown in the output below. R12 designates both internal paths as multipaths, with the path received from R10 as the best path.

On R12:

```
R12(config)#router bgp 400
R12(config-router)#maximum-paths ibgp 2

R12#show ip bgp 130.7.1.1

BGP routing table entry for 130.7.1.1/32, version 120
Paths: (2 available, best #1, table default)
Multipath: iBGP
  Advertised to update-groups:
     1
  Refresh Epoch 1
  300, (Received from a RR-client)
    10.10.10.10 (metric 11) from 10.10.10.10 (10.10.10.10)
      Origin IGP, metric 0, localpref 100, valid, internal, multipath,
best
      rx pathid: 0, tx pathid: 0x0
  Refresh Epoch 1
  300, (Received from a RR-client)
    11.11.11.11 (metric 11) from 11.11.11.11 (11.11.11.11)
      Origin IGP, metric 0, localpref 100, valid, internal,
multipath(oldest)
      rx pathid: 0, tx pathid: 0
```

As you have seen in this section, the BGP multipath settings allow BGP to act like a normal IGP and install multiple paths into the routing table if certain attributes are equal to its current best path. When configuring BGP multipathing, the following should be considered:

- Potential multipath paths are compared to the current BGP best path.
- It is not possible to configure **maximum-paths** or **maximum-paths ibgp** along with **maximum-paths eibgp**.
- It is possible to configure **maximum-paths** along with **maximum-paths ibgp**.

The first point is just another friendly reminder that the multipath comparison is made by comparing the path attributes of the additional path to the current BGP best path. If multipathing isn't working, start by verifying that the proper set of attributes matches between the two paths.

The second point enforces the requirement that if both internal and external paths to the same prefix are to be considered for multipathing, it is not possible to enable multipathing for only one set of paths. In other words, the administrator must choose whether only internal or external paths to the same prefix are considered for multipathing. This is because the **maximum-paths eibgp** command would include a subset of the other two commands. Including this subset makes it incompatible for consideration.

On the other hand, the second point emphasizes that it is possible to configure the **maximum-paths** and **maximum-paths ibgp** commands together. This use is not detailed in this document.

Step 10: Oldest Route

Note Before starting this section, revert the configuration on all routers to the base initial configuration files provided with the lab.

The next steps of the BGP best-path algorithm introduce a series of tie-breaker conditions that are designed to help deterministically choose a best path from two paths that are virtually identical. At this point, the two paths under consideration have similar attributes and are either both external or both internal paths. Step 10 of the best-path algorithm specifically deals with stability between external peers.

Put simply, at this step, BGP prefers an external path that has already become best over any competing external path with the same attributes. This ensures that BGP does not unnecessarily introduce route flaps into the BGP process run with its external peers by informing it of a path that is only superior to the local BGP router's already chosen best path based on criteria following this step.

The **show ip bgp** output lists paths in reverse order from when they were received. That is, paths at the bottom of the list were received first, and paths at the top of the list were received last.

This step only applies to external paths and is skipped if any of the following are true:

- The router ID is the same for multiple paths, indicating that the path was learned from the same router.
- There is currently no best path, indicating that the current best path was lost or has never been selected.
- The **bgp best-path compare-routerid** command has been enabled. This is explained in step 11 of the best-path algorithm.

Because this section is heavily dependent on a specific order of operations, the following has been applied to the topology in order to achieve consistent results:

- The R17/R10 peering is shut down.
- The R17/R10 peering is brought up.

To demonstrate how this step functions, examine the output of the **show ip bgp 140.15.1.1** output on R17:

On R17:

```
R17(config)#router bgp 200
R17(config-router)#neighbor 200.10.17.10 shutdown

%BGP-5-NBR_RESET: Neighbor 200.10.17.10 reset (Admin. shutdown)

%BGP-5-ADJCHANGE: neighbor 200.10.17.10 Down Admin. shutdown

%BGP_SESSION-5-ADJCHANGE: neighbor 200.10.17.10 IPv4 Unicast topology
base removed from session  Admin. shutdown

R17(config-router)#no neighbor 200.10.17.10 shutdown
```

You should see the following console message:

```
%BGP-5-ADJCHANGE: neighbor 200.10.17.10 Up


R17#show ip bgp 140.15.1.1

BGP routing table entry for 140.15.1.1/32, version 28
Paths: (2 available, best #2, table default)
  Advertised to update-groups:
     1           2
  Refresh Epoch 1
  400
    200.10.17.10 from 200.10.17.10 (10.10.10.10)
      Origin IGP, localpref 100, valid, external
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
  400
    200.9.17.9 from 200.9.17.9 (9.9.9.9)
      Origin IGP, localpref 100, valid, external, best
      rx pathid: 0, tx pathid: 0x0
```

R17 receives two paths for the 140.15.1.1/32 prefix: one from R9 and one from R10. Between the two competing prefixes, the WEIGHT, LOCAL_PREF, AS_PATH, ORIGIN, and MED are all tied. In addition, both paths are external paths. The deciding factor between the two paths is that the path from R9 was received first, as indicated by its being the last path listed in the output. Since R9 is the oldest route, BGP prefers R9.

To prove this point, the peering between R17 and R9 is shut down and brought back up again. R17 removes all prefixes learned from R9 as the peering goes down. At this point, the only path R17 has is from R10, which it now considers best. When the peering is restored to R9, R9 advertises its path back to R17. R17 runs the best-path algorithm for the two paths again. The result of the calculation is that R17 will continue to retain its current best path via R10 as its oldest path.

```
R17(config)#router bgp 200
R17(config-router)#neighbor 200.9.17.9 shut

You should see the following console messages:

%BGP-5-NBR_RESET: Neighbor 200.9.17.9 reset (Admin. shutdown)
%BGP-5-ADJCHANGE: neighbor 200.9.17.9 Down Admin. shutdown
%BGP_SESSION-5-ADJCHANGE: neighbor 200.9.17.9 IPv4 Unicast topology
base removed from session  Admin. shutdown

R17(config-router)#no neighbor 200.9.17.9 shut
You should see the following console messages:
%BGP-5-ADJCHANGE: neighbor 200.9.17.9 Up

R17#show ip bgp 140.15.1.1

BGP routing table entry for 140.15.1.1/32, version 20
Paths: (2 available, best #2, table default)
  Advertised to update-groups:
     1          2
  Refresh Epoch 1
  400
    200.9.17.9 from 200.9.17.9 (9.9.9.9)
      Origin IGP, localpref 100, valid, external
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 2
  400
    200.10.17.10 from 200.10.17.10 (10.10.10.10)
      Origin IGP, localpref 100, valid, external, best
      rx pathid: 0, tx pathid: 0x0
```