BGP Best Path Selection Algorithm

Contents

Introduction

This document describes the function of the Border Gateway Protocol (BGP) best path algorithm. BGP routers typically receive multiple paths to the same destination. The BGP best path algorithm decides which is the best path to install in the IP routing table and to use for traffic forwarding.

Why Routers Ignore Paths

Assume that all paths that a router receives for a particular prefix are arranged in a list. The list is similar to the output of the **show ip bgp longer-prefixes** command. In this case, some paths are not considered as candidates for the best path. Such paths typically do not have the valid flag in the output of the **show ip bgp longer-prefixes** command. Routers ignore paths in these circumstances:

- Paths that are marked as `not synchronized` in the **show ip bgp longer-prefixes** output.If BGP synchronization is enabled, there must be a match for the prefix in the IP routing table in order for an internal BGP (iBGP) path in order to be considered a valid path. BGP synchronization is enabled by default in Cisco IOS® Software. If the route that matches is learned from an Open Shortest Path First (OSPF) neighbor, its OSPF router ID must match the BGP router ID of the iBGP neighbor. Most users prefer to disable synchronization with use of the **no synchronization** BGP subcommand.**Note**:Synchronization is disabled by default in Cisco IOS Software Release 12.2(8)T and later.
- Paths for which the NEXT_HOP is inaccessible.Be sure that there is an Interior Gateway Protocol (IGP) route to the NEXT_HOP that is associated with the path.
- Paths from an external BGP (eBGP) neighbor if the local autonomous system (AS) appears in the AS_PATH. Such paths are denied upon ingress into the router and are not even installed in the BGP Routing Information Base (RIB). The same applies to any path that is denied by a routing policy that is implemented via access, prefix, AS_PATH, or community lists, unless you have configured **neighbor soft-reconfiguration inbound** for the neighbor.
- If you enabled **bgp enforce-first-as** and the UPDATE does not contain the AS of the neighbor as the first AS number in the AS_SEQUENCE.In this case, the router sends a notification and closes the session.
- Paths that are marked as `(received-only)` in the **show ip bgp longer-prefixes** outputThe policy has rejected these paths. However, the router has stored the paths because you have configured **soft-reconfiguration inbound** for the neighbor that sends the path.

How the Best Path Algorithm Works

BGP assigns the first valid path as the current best path. BGP then compares the best path with the next path in the list, until BGP reaches the end of the list of valid paths. This list provides the rules that are used in order to determine the best path:

1. Prefer the path with the highest WEIGHT. **Note**: WEIGHT is a Cisco-specific parameter. It is local to the router on which it is configured.
2. Prefer the path with the highest LOCAL_PREF.**Note**: A path without LOCAL_PREF is considered to have had the value set with the **bgp default local-preference** command, or to have a value of 100 by default.
3. Prefer the path that was locally originated via a **network** or **aggregate** BGP subcommand or through redistribution from an IGP.Local paths that are sourced by the **network** or **redistribute** commands are preferred over local aggregates that are sourced by the **aggregate-address** command.**Note**: Be aware of this item:

   - If AIGP is configured AND the **bgp bestpath aigp ignore** command is not configured, the decision process considers the AIGP metric. See Configure the AIGP Metric Attribute for BGP for further details.
4. Prefer the path with the shortest AS_PATH.**Note**: Be aware of these items:

   - This step is skipped if you have configured the **bgp bestpath as-path ignore** command.

   - An AS_SET counts as 1, no matter how many ASs are in the set.

   - The AS_CONFED_SEQUENCE and AS_CONFED_SET are not included in the AS_PATH length.
5. Prefer the path with the lowest origin type.**Note**: IGP is lower than Exterior Gateway Protocol (EGP), and EGP is lower than INCOMPLETE.
6. Prefer the path with the lowest multi-exit discriminator (MED).**Note**: Be aware of these items:

   - This comparison only occurs if the first (the neighboring) AS is the same in the two paths. Any confederation sub-ASs are ignored.
   In other words, MEDs are compared only if the first AS in the AS_SEQUENCE is the same for multiple paths. Any AS_CONFED_SEQUENCE that precedes is ignored.

   - If **bgp always-compare-med** is enabled, MEDs are compared for all paths.
   You must disable this option over the entire AS. Otherwise, routing loops can occur.

   - If **bgp bestpath med-confed** is enabled, MEDs are compared for all paths that consist only of AS_CONFED_SEQUENCE.
   These paths originated within the local confederation.

   - THE MED of paths that are received from a neighbor with a MED of 4,294,967,295 is changed before insertion into the BGP table. The MED changes to 4,294,967,294.

   - THE MED of paths that are received from a neighbor with a MED of 4,294,967,295 are considered valid and are inserted into BGP table with effect to Codes fixed for Cisco bug ID

CSCef34800.

- Paths received with no MED are assigned a MED of 0, unless you have enabled **bgp bestpath med missing-as-worst**.
If you have enabled **bgp bestpath med missing-as-worst**, the paths are assigned a MED of 4,294,967,294.
If you have enabled **bgp bestpath med missing-as-worst**, the paths are assigned a MED of 4,294,967,295 with effect to Codes fixed for Cisco bug ID CSCef34800.

- The **bgp deterministic-med** command can also influence this step.
Refer to How BGP Routers Use the Multi-Exit Discriminator for Best Path Selection for a demonstration.

7. Prefer eBGP over iBGP paths.If bestpath is selected, go to Step 9 (multipath).**Note**: Paths that contain AS_CONFED_SEQUENCE and AS_CONFED_SET are local to the confederation. Therefore, these paths are treated as internal paths. There is no distinction between Confederation External and Confederation Internal.
8. Prefer the path with the lowest IGP metric to the BGP next hop.Continue, even if bestpath is already selected.
9. Determine if multiple paths require installation in the routing table for BGP Multipath.Continue, if bestpath is not yet selected.
10. When both paths are external, prefer the path that was received first (the oldest one).This step minimizes route-flap because a newer path does not displace an older one, even if the newer path would be the preferred route based on the next decision criteria (Steps 11, 12, and 13).Skip this step if any of these items is true:You have enabled the **bgp best path compare-routerid** command.**Note**: Cisco IOS Software Releases 12.0.11S, 12.0.11SC, 12.0.11S3, 12.1.3, 12.1.3AA, 12.1.3.T, and 12.1.3.E introduced this command.The router ID is the same for multiple paths because the routes were received from the same router.There is no current best path.The current best path can be lost when, for example, the neighbor that offers the path goes down.
11. Prefer the route that comes from the BGP router with the lowest router ID.The router ID is the highest IP address on the router, with preference given to loopback addresses. Also, you can use the **bgp router-id** command in order to manually set the router ID.**Note**: If a path contains route reflector (RR) attributes, the originator ID is substituted for the router ID in the path selection process.
12. If the originator or router ID is the same for multiple paths, prefer the path with the minimum cluster list length.This is only present in BGP RR environments. It allows clients to peer with RRs or clients in other clusters. In this scenario, the client must be aware of the RR-specific BGP attribute.
13. Prefer the path that comes from the lowest neighbor address.This address is the IP address that is used in the BGP **neighbor** configuration. The address corresponds to the remote peer that is used in the TCP connection with the local router.

Example: BGP Best Path Selection

In this example, 9 paths are available for the network 10.30.116.0/23. The **show ip bgp network** command displays the entries in the BGP routing table for the given network.

```
Router R1#show ip bgp vpnv4 rd 1100:1001 10.30.116.0/23
BGP routing table entry for 1100:1001:10.30.116.0/23, version 26765275
Paths: (9 available, best #6, no table)
  Advertised to update-groups:
     1          2          3
  (65001 64955 65003) 65089, (Received from a RR-client)
    172.16.254.226 (metric 20645) from 172.16.224.236 (172.16.224.236)
      Origin IGP, metric 0, localpref 100, valid, confed-internal
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
  (65008 64955 65003) 65089
    172.16.254.226 (metric 20645) from 10.131.123.71 (10.131.123.71)
      Origin IGP, metric 0, localpref 100, valid, confed-external
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
  (65001 64955 65003) 65089
    172.16.254.226 (metric 20645) from 172.16.216.253 (172.16.216.253)
      Origin IGP, metric 0, localpref 100, valid, confed-external
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
  (65001 64955 65003) 65089
    172.16.254.226 (metric 20645) from 172.16.216.252 (172.16.216.252)
      Origin IGP, metric 0, localpref 100, valid, confed-external
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
  (64955 65003) 65089
    172.16.254.226 (metric 20645) from 10.77.255.57 (10.77.255.57)
      Origin IGP, metric 0, localpref 100, valid, confed-external
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
  (64955 65003) 65089
    172.16.254.226 (metric 20645) from 10.57.255.11 (10.57.255.11)
      Origin IGP, metric 0, localpref 100, valid, confed-external, best
      Extended Community: RT:1100:1001
      mpls labels in/out nolabel/362
!--- BGP selects this as the Best Path on comparing
!--- with all the other routes and selected based on lower router ID. (64955 65003) 65089
172.16.254.226 (metric 20645) from 172.16.224.253 (172.16.224.253) Origin IGP, metric 0,
localpref 100, valid, confed-internal Extended Community: RT:1100:1001 mpls labels in/out
nolabel/362 (65003) 65089 172.16.254.226 (metric 20645) from 172.16.254.234 (172.16.254.234)
Origin IGP, metric 0, localpref 100, valid, confed-external Extended Community: RT:1100:1001
mpls labels in/out nolabel/362 65089, (Received from a RR-client) 172.16.228.226 (metric 20645)
from 172.16.228.226 (172.16.228.226) Origin IGP, metric 0, localpref 100, valid, confed-internal
Extended Community: RT:1100:1001 mpls labels in/out nolabel/278
```

BGP selects the best path out of these 9 paths through the consideration of various attributes that are explained in this document. In the output shown here, BGP compares the available paths and selects Path# 6 as the best path based on its lower router-ID.

```
Comparing path 1 with path 2:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
The paths have different neighbor AS's so ignoring MED
Both paths are internal
  (no distinction is made between confed-internal and confed-external)
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 2 is better than path 1 because it has a lower Router-ID.

Comparing path 2 with path 3:
```

```
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are confed-external
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 2 is better than path 3 because it has a lower Router-ID.

Comparing path 2 with path 4:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are confed-external
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 2 is better than path 4 because it has a lower Router-ID.

Comparing path 2 with path 5:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are confed-external
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 5 is better than path 2 because it has a lower Router-ID.

Comparing path 5 with path 6:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are confed-external
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 6 is better than path 5 because it has a lower Router-ID.

Comparing path 6 with path 7:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are internal
  (no distinction is made between confed-internal and confed-external)
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 6 is better than path 7 because it has a lower Router-ID.
```

```
Comparing path 6 with path 8:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
Both paths have the same neighbor AS, 65089, so comparing MED.
Both paths have a MED of 0
Both paths are confed-external
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 6 is better than path 8 because it has a lower Router-ID.

Comparing path 6 with path 9:
Both paths have reachable next hops
Both paths have a WEIGHT of 0
Both paths have a LOCAL_PREF of 100
Both paths are learned
Both paths have AS_PATH length 1
Both paths are of origin IGP
The paths have different neighbor AS's so ignoring MED
Both paths are internal
  (no distinction is made between confed-internal and confed-external)
Both paths have an IGP metric to the NEXT_HOP of 20645
Path 6 is better than path 9 because it has a lower Router-ID.
The best path is #6
```

Customize the Path Selection Process

The extended community attribute, which is called BGP Cost Community, provides a way to customize the best path selection process. An additional step, in which cost communities are compared, is added to the algorithm that the How the Best Path Algorithm Works section describes. This step comes after the required step (point of insertion) in the algorithm. The path with the lowest cost value is preferred.

> **Note**: Be aware of these items:
>
> - This step is skipped if you have issued the **bgp bestpath cost-community ignore** command.
>
> - The cost community set clause is configured with a cost community ID number (0 to 255) and cost number value (0 to 4,294,967,295). The cost number value determines the preference for the path. The path with the lowest cost number value is preferred. Paths that are not specifically configured with the cost number value are assigned a default cost number value of 2,147,483,647. This value is the midpoint between 0 and 4,294,967,295. These paths are then evaluated accordingly by the best path selection process. If two paths are configured with the same cost number value, the path selection process prefers the path with the lowest community ID. If the paths have unequal pre-bestpath cost communities, the path with the lower pre-bestpath cost community is selected as the best path.
>
> - The ABSOLUTE_VALUE is considered the first step in order to determine the degree of preference of a path. For example, when EIGRP is redistributed to BGP VPNv4, the ABSOLUTE_VALUE type is used for cost community. The IGB_Cost is considered after the interior (IGP) distance to the next hop has been compared. This means that cost communities with the IGP_COST point of insertion are considered after step 8 of the algorithm in How the Best Path Algorithm Works.

BGP Multipath

BGP Multipath allows installation into the IP routing table of multiple BGP paths to the same destination. These paths are installed in the table together with the best path for load sharing. BGP Multipath does not affect bestpath selection. For example, a router still designates one of the paths as the best path, per the algorithm, and advertises this best path to its neighbors.

These are the BGP Multipath features:

- eBGP Multipath - **maximum-paths n**
- iBGP Multipath - **maximum-paths ibgp n**
- eiBGP Multipath - **maximum-paths eibgp**

In order to be candidates for multipath, paths to the same destination need to have these characteristics equal to the best-path characteristics:

- Weight
- Local preference
- AS-PATH length
- Origin
- MED
- One of these:Neighboring AS or sub-AS (before the addition of the eiBGP Multipath feature)AS-PATH (after the addition of the eiBGP Multipath feature)

Some BGP Multipath features put additional requirements on multipath candidates.

These are the additional requirements for eBGP multipath:

- The path must be learned from an external or confederation-external neighbor (eBGP).
- The IGP metric to the BGP next hop must be equal to the best-path IGP metric.

These are the additional requirements for iBGP multipath:

- The path must be learned from an internal neighbor (iBGP).
- The IGP metric to the BGP next hop must be equal to the best-path IGP metric, unless the router is configured for unequal-cost iBGP multipath.

BGP inserts up to *n* most recently received paths from multipath candidates in the IP routing table. The maximum value of *n* is currently 6. The default value, when multipath is disabled, is 1.

For unequal-cost load balancing, you can also use BGP Link Bandwidth.

> **Note**: The equivalent next-hop-self is performed on the best path that is selected among eBGP multipaths before it is forwarded to internal peers.

Related Information

- **Troubleshooting BGP**
- **How BGP Routers Use the Multi-Exit Discriminator for Best Path Selection**
- **Configuring BGP**
- **BGP Support Page**

- **Technical Support & Documentation - Cisco Systems**