

BGP Introduction

BGP is a path vector protocol used to carry routing information between autonomous systems. The term path vector comes from the fact that BGP routing information carries a sequence of AS numbers that identifies the path of AS's that a network prefix has traversed. The path information associated with the prefix is used to enable loop prevention.

BGP uses TCP as its transport protocol (port 179). This ensures that all the transport reliability (such as retransmission) is taken care of by TCP and does not need to be implemented in BGP, thereby simplifying the complexity associated with designing reliability into the protocol itself.

NOTE:

 BGP protocol supports VRF.

BGP Speaker and BGP Peers

Routers that run a BGP routing process are often referred to as BGP speakers. Two BGP speakers that form a TCP connection between one another for the purpose of exchanging routing information, are referred to as neighbors or peers. Peer routers exchange open messages to determine the connection parameters. These messages are used to communicate values such as the BGP speaker's version number.

Gracefully Shutdown

BGP also provides a mechanism to gracefully close a connection with a peer. In other words, in the event of a disagreement between the peers, be it resultant of configuration, incompatibility, operator intervention, or other circumstances, a NOTIFICATION error message is sent, and the peer connection does not get established or is torn down if it's already established. The benefit of this mechanism is that both peers understand that the connection could not be established or maintained and do not waste resources that would otherwise be required to maintain or blindly reattempt to establish the connection. The graceful close mechanism simply ensures that all outstanding messages, primarily NOTIFICATION error messages, are delivered before the TCP session is closed.

BGP Session Establishment

Initially, when a BGP session is established between a set of BGP speakers, all candidate BGP routes are exchanged. After the session has been established and the initial route exchange has occurred, only incremental updates are sent as network information changes.

Routes are advertised between a pair of BGP routers in UPDATE messages. The UPDATE message contains, among other things, a list of <length, prefix> tuples that indicate the list of destinations that can be reached via a BGP speaker. The UPDATE message also contains the path attributes, which include such information as the degree of preference for a particular route and the list of AS's that the route has traversed.

In the event that a route becomes unreachable, a BGP speaker informs its neighbors by withdrawing the invalid route. Withdrawn routes are part of the UPDATE message. These routes are no longer available for use. If information associated with a route has changed or a new path for the same prefix has been selected, a withdrawal is not required; it is enough to just advertise a replacement route.

KEEPALIVE Message

If no routing changes occur, the routers exchange only KEEPALIVE packets.

KEEPALIVE messages are sent periodically between BGP neighbors to ensure that the connection is kept alive. KEEPALIVE packets (19 bytes each) should not cause any strain on the router CPU or link bandwidth because they consume a minimal amount of bandwidth.

BGP Attributes

BGP attributes of the routes can be used in route selection and load balancing. The following describes common BGP route attributes:

- **Origin**

The Origin attribute defines the origin of a route and marks the path of a BGP route. The Origin attribute is classified into three types:

- IGP

A route with IGP as the Origin attribute is of the highest priority. The Origin attribute of the routes imported into a BGP routing table using the **set protocols bgp [vrf <vrf-name>] {ipv4-unicast|ipv6-unicast} network <ipv4/prefixlen> [label-index <label-index>] route-map <ROUTE-MAP-NAME>** command is IGP.

- EGP

A route with EGP as the Origin attribute is of the secondary highest priority. The Origin attribute of the routes obtained through EGP is EGP.

- Incomplete

A route with Incomplete as the Origin attribute is of the lowest priority. The Origin attribute of the routes learned by other means is Incomplete. For example, the Origin attribute of the routes imported by BGP using the `set protocols bgp [vrf <vrf-name>] {ipv4-unicast|ipv6-unicast} redistribute {connected|kernel|ospf|static|table <table-number>} [metric <metric-number>|route-map <route-map>]` command is Incomplete.

- **AS_Path**

The AS_Path attribute records all the ASs that a route passes through from the source to the destination in the vector order. To prevent inter-AS routing loops, a BGP device does not receive the routes of which the AS_Path list contains the local AS number.

When a BGP speaker advertises an imported route:

- If the route is advertised to EBGp peers, the BGP speaker creates an AS_Path list containing the local AS number in an Update message.
- If the route is advertised to IBGP peers, the BGP speaker creates an empty AS_Path list in an Update message.

When a BGP speaker advertises a route learned in the Update message sent by another BGP speaker:

- If the route is advertised to EBGp peers, the BGP speaker adds the local AS number to the leftmost of the AS_Path list. According to the AS_Path list, the BGP speaker that receives the route can learn about the ASs through which the route passes to reach the destination. The number of the AS that is nearest to the local AS is placed on the top of the AS_Path list. The other AS numbers are listed according to the sequence in which the route passes through ASs.
- If the route is advertised to IBGP peers, the BGP speaker does not change the AS_Path attribute of the route.

- **Next_Hop**

The Next_Hop attribute records the next hop that a route passes through. The Next_Hop attribute of BGP is different from that of an IGP because it may not be the neighbor IP address. A BGP speaker processes the Next_Hop attribute based on the following rules:

- When advertising a route to an EBGp peer, a BGP speaker sets the Next_Hop attribute of the route to the address of the local interface through which the BGP peer relationship is established with the peer.
- When advertising a locally originated route to an IBGP peer, the BGP speaker sets the Next_Hop attribute of the route to the address of the local interface through which the BGP peer relationship is established with the peer.
- When advertising a route learned from an EBGp peer to an IBGP peer, the BGP speaker does not change the Next_Hop attribute of the route.

- **Local_Pref**

The Local_Pref attribute indicates the BGP preference of a device and helps determine the optimal route when traffic leaves an AS. When a BGP device obtains multiple routes to the same destination address but with different next hops from different IBGP peers, the BGP device prefers the route with the highest Local_Pref. The Local_Pref attribute is exchanged only between IBGP peers and is not advertised to other ASs. The Local_Pref attribute can be manually configured. If no Local_Pref attribute is configured for a route, the Local_Pref attribute of the route uses the default value 100.

- **MED**

The multi-exit discriminator (MED) attribute helps determine the optimal route when traffic enters an AS. When a BGP device obtains multiple routes to the same destination address but with different next hops from EBGp peers, the BGP device selects the route with the smallest MED value as the optimal route.

The MED attribute is exchanged only between two neighboring ASs. The AS that receives the MED attribute does not advertise it to any other ASs. The MED attribute can be manually configured. If no MED attribute is configured for a route, the MED attribute of the route uses the default value 0.

- **Community**

The Community attribute identifies the BGP routes with the same characteristics, simplifies the applications of routing policies, and facilitates route maintenance and management.

The Community attribute includes self-defined community attributes and well-known community attributes. **Table 1** lists well-known community attributes.

Table 1 Well-known community attributes

Community Attribute	Value	Description
Internet	0 (0x00000000)	A BGP device can advertise the received route with the Internet attribute to all peers.
No_Advertise	4294967042 (0xFFFFFFFF02)	A BGP device does not advertise the received route with the No_Advertise attribute to any peer.
No_Export	4294967041 (0xFFFFFFFF01)	A BGP device does not advertise the received route with the No_Export attribute to devices outside the local AS.
No_Export_Subconfed	4294967043 (0xFFFFFFFF03)	A BGP device does not advertise the received route with the No_Export_Subconfed attribute to devices outside the local AS or to devices outside the local sub-AS.

- **Originator_ID and Cluster_List**

The Originator_ID attribute and Cluster_List attribute help eliminate loops in route reflector scenarios.

