Troubleshoot Border Gateway Protocol Basic Issues

Contents

Introduction

Prerequisites

- Requirements
- Components Used

Background Information

Topology

Scenarios and Problems

- Adjacency Down
  - No Connectivity
  - Configuration Issues
  - TCPSession Issues
- Adjacency Bounces
  - Interface Flap
  - Hold Timer Expired
  - AFI/SAFI Issues
- Path Install and Selection
  - Next Hop
  - RIB Failure
  - Race Condition
- Other Issues
  - BGP Slow Peer
  - Memory Issues
  - High CPU

Related Information

Introduction

This document describes how to troubleshoot the most common issues with the Border Gateway Protocol (BGP) and provides basic solutions and guidelines.

Prerequisites

Requirements

There are no specific prerequisites for this document. Basic BGP protocol knowledge is useful, you can refer to the BGP Configuration Guide for more information.

Components Used

This document is not restricted to specific software and hardware versions, but commands are applicable for Cisco IOS® and Cisco IOS® XE.

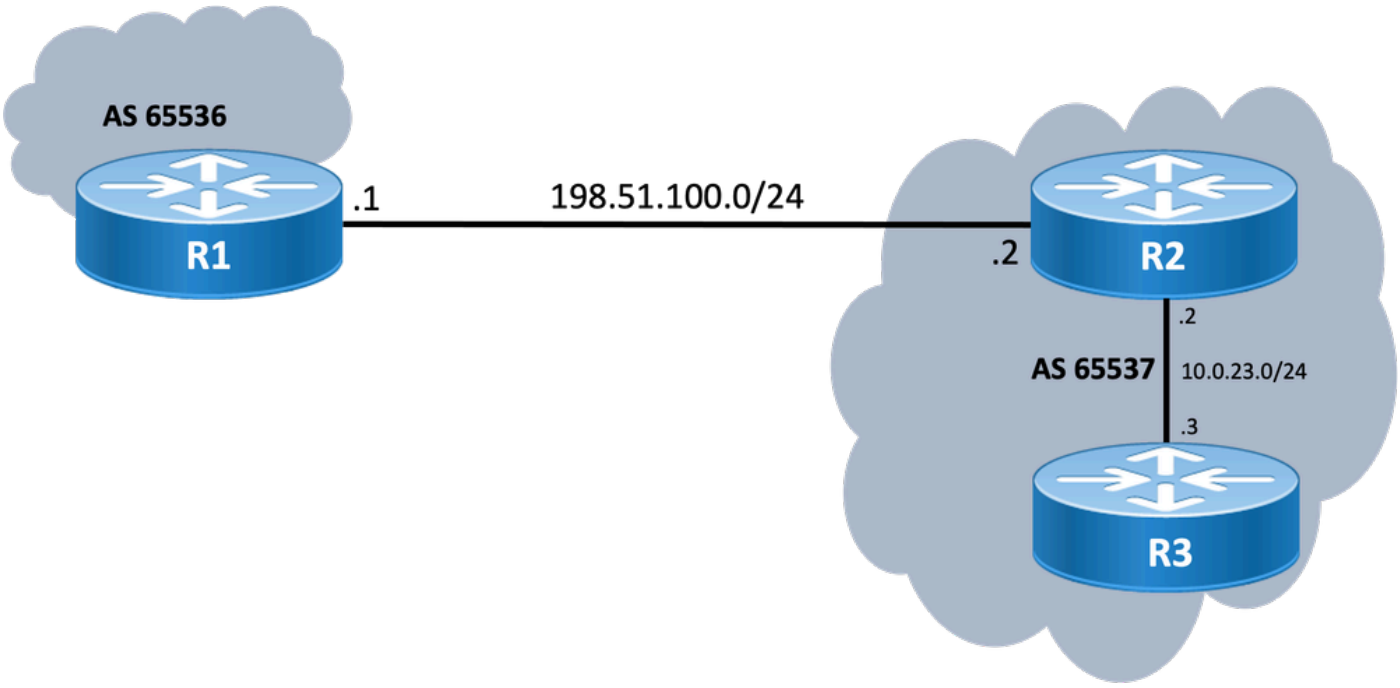
The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Background Information

This document describes a basic guide to troubleshoot the most common issues in Border Gateway Protocol (BGP), provides corrective actions, useful commands/debugs to detect the root cause of problems, and best practices to avoid potential issues. Keep in mind all possible variables and scenarios cannot be considered and a deeper analysis can be required by Cisco TAC.

Topology

Use this topology diagram as reference for the outputs provided in this document.



Scenarios and Problems

Adjacency Down

If a BGP session is offline, run the **show ip bgp all summary** command. This provides the current status of the session:

- If the session is not up state, it can vary between IDLE and ACTIVE (this depends on the Finite State Machine process.)
- If the session is online, you can see the number of prefixes received.

<#root>

```
R2#
```

show ip bgp all summary

```
For address family: IPv4 Unicast
BGP router identifier 198.51.100.2, local AS number 65537
BGP table version is 19, main routing table version 19
18 network entries using 4464 bytes of memory
18 path entries using 2448 bytes of memory
1/1 BGP path/bestpath attribute entries using 296 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 7208 total bytes of memory
```

```
BGP activity 18/0 prefixes, 18/0 paths, scan interval 60 secs
18 networks peaked at 11:21:00 Jun 30 2022 CST (00:01:35.450 ago)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.0.23.3       4        65537       6       5       19    0    0 00:01:34       18

198.51.100.1    4        65536       0       0        1    0    0 never    Idle
```

No Connectivity

The first requirement is the connectivity between both peers, so, the TCP session on port 179 is established. Either they are directly connected or not) and you can use a ping. If peering is established between loopback interfaces, a loopback to loopback ping must be completed. If a ping test is performed without a specific loopback as the source interface, the outgoing physical interface IP address is used as the packet's source IP address instead of the router’s loopback IP address.

If the ping is unsuccessful, consider these reasons:

- No connected route peer or no route at all; run the **show ip route peer_IP_address** command.
- Layer 1 issue: physical interface, SFP (connector), cable or external issue (transport and provider if applicable) must be considered.
- Check any firewall or access lists that can block your connection.

If the ping is successful:

Configuration Issues

- Wrong IP address or AS configured: For wrong IP address, there is no such message displayed, however, ensure the proper configuration are completed. For a wrong AS, you must see a message similar to the **show logging** command.

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.1 passive 2/2 (peer in wrong AS) 2 bytes 1B39
```

Check the BGP configuration on both ends to correct the AS numbers or peer IP address.

- Duplicate router ID:

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.1 passive 2/3 (BGP identifier wrong) 4 bytes 0A0A0A0A
```

Check the BGP identifier on both ends by running the **show ip bgp all summary** command and correct the duplicate issue. This can be achieved manually with the global command **bgp router-id X.X.X.X** under the bgp router configuration. As a best practice, ensure the router ID is set manually to a unique number.

- BGP Source and TTL:

Most iBGP sessions are configured over loopback interfaces, which are reachable via an IGP. This loopback interface must be explicitly defined as the source, You can complete this by running the **neighbor ip-address update-source interface-id** command.

For eBGP peer directly connected interfaces, most are used for peering. There is a check for Cisco IOS/Cisco IOS XE to fulfill this purpose, or it does not try to establish a session. If eBGP is attempted from loopback to loopback on directly connected routers, this check can be disabled for a specific neighbor on both ends by running the **neighbor ip-address disable-connected-check** command.

However, if there are multiple hops between the eBGP peers, a proper hop count is required, ensure the **neighbor ip-address ebgp-multihop [hop-count]**is configured with the correct hop count so each session can be established. If the hop-count is not specified, the default TTL value for iBGP sessions is 255, while the default TTL value for eBGP sessions is 1.

TCP Session Issues

A useful action to test port 179 is a manual telnet from one peer to the other:

<#root>

```
R1#

telnet 198.51.100.2 179

Trying 198.51.100.2, 179 ... Open

[Connection to 198.51.100.2 closed by foreign host]
```

Either open/connection closed, or connection refused by the remote host indicates packets reached remote end. Then, ensure there are no problems with the control plane at the far end. Otherwise, if there is a Destination Unreachable message, check any firewall or access lists that can block TCP port 179, BGP packets, or if there is any packet loss on the path.

If authentication is the problem, messages you can see are:

```
%TCP-6-BADAUTH: Invalid MD5 digest from 198.51.100.1(179) to 198.51.100.2(20062) tableid - 0
%TCP-6-BADAUTH: No MD5 digest from 198.51.100.1(179) to 198.51.100.2(20062) tableid - 0
```

Check authentication methods, password and related configurations, and to further troubleshoot, refer to the MD5 Authentication Between BGP Peers Configuration Example guide.

If the TCP session is not online, use the next commands for isolation:

```
show tcp brief all
show control-plane host open-ports
debug ip tcp transactions
```

Adjacency Bounces

If the session is intermittent, look for the **show log** and you can find some scenarios.

Interface Flap

```
%BGP-5-ADJCHANGE: neighbor 198.51.100.2 Down Interface flap
```

The reason for this failure is due to "Down Interface Flap." Look for any physical issues on the port/SFP, cable, or disconnections.

Hold Timer Expired

```
%BGP-3-NOTIFICATION: sent to neighbor 198.51.100.2 4/0 (hold time expired) 0 bytes
```

This is common and the router did not receive/process a keepalive message or update the message prior to the hold timer expiring. The device sends a notification message and closes the session. The most commons reasons for this issue are:

- **Interface Issues**: Look for input errors, input queue drops, or physical issues on both peer connected interfaces; running the **show interface** command can be used for this purpose.

- Packet Loss in Transit: Sometimes Hello packets can be dropped during transit, to ensure this is a packet capture at interface level:
  - Use the Embedded Packet Capture on Cisco IOS and Cisco IOS XE devices.
  - If packets are seen at the interface level, ensure they reach the control plane, EPC on control plane, or run the **debug bgp [vrf name] ipv4 unicast keepalives** command.

- High CPU: A high CPU condition can cause drops on the control plane, the **show processes cpu [sorted|history]** command can help identify these problems. Based on the platform, you can find the next steps to troubleshoot by referring to the CPU Reference guide.

- CoPP Policy Issues: The troubleshooting methodology varies for each platform and is out of scope for this document.
- MTU Mismatch: If there are MTU discrepancies in the path, and if ICMP messages are blocked in the path from the source to the destination, PMTUD does not function properly and can result in a session flap. Updates are sent with the negotiated MSS value and a DF bit set. If a device in the path or the destination cannot accept the packets with a higher MTU, it sends an ICMP error message back to the BGP speaker. The destination router waits for the BGP keepalive or BGP updates the packet to its hold down timer.

You can check the MSS negotiated by running the **show ip bgp neighbors ip_address** command.

A ping test to a specific neighbor with the df set can show you if the MTU is valid along the path:

<#root>

ping 198.51.100.2 size

max_seg_size

df

If MTU issues are found, an accurate review of the configuration must be completed to ensure the MTU values are consistent throughout the network.

Note: For more information on MTU, refer to the BGP Neighbor Flaps with MTU Troubleshooting guide.

AFI/SAFI Issues

```
%BGP-5-ADJCHANGE: neighbor 198.51.100.2 passive Down AFI/SAFI not supported
%BGP-3-NOTIFICATION: received from neighbor 198.51.100.2 active 2/8 (no supported AFI/SAFI) 3 bytes 000
```

Address-Family Identifier (AFI) is a capability extension added by Multi-Protocol BGP (MP-BGP.) It correlates to a specific network protocol, such as IPv4, IPv6, and the like. Additional granularity through a Subsequent Address-Family Identifier (SAFI), such as unicast and multicast. MBGP achieves this separation with BGP path attributes (PAs) MP_REACH_NLRI and MP_UNREACH_NLRI. These attributes are carried inside BGP update messages and are used to carry network reachability information for different address families.

The message provides you with the numbers of the AFI/SAFI registered by IANA:

- IANA Address Family Numbers
- Subsequent Address Family Identifiers (SAFI) Parameters

- Check the BGP configuration for the address-families intended on both sides to correct any undesired address families.
- Run the **neighbor ip-address dont-capability-negotiate** command on both ends. For more information, refer to the Unsupported Capabilities Cause BGP Peer Malfunction guide.

Path Install and Selection

For additional information on BGP and selecting the best path, refer to BGP Best Path Selection Algorithm.

Next Hop

For a route to be installed into the routing table, the next hop must be reachable, otherwise, even if the prefix is on the Loc-RIB BGP table, it does not move into RIB. As a loop avoidance rule, on Cisco IOS/Cisco IOS XE, iBGP does not change the next hop attribute as it leaves the AS_PATH alone while eBGP rewrites the next hop and prepends its AS_PATH.

You can review the next hop by running the **show ip bgp [prefix]** command as it provides the next hop and inaccessible word. In this example, this is a prefix announced by R1 via eBGP to R2 and learnt by R3 via iBGP connection from R2:

<#root>

```
R3#
```

show ip bgp 192.0.2.1

```
BGP routing table entry for 192.0.2.1/32, version 0

Paths: (1 available, no best path)

  Not advertised to any peer
  Refresh Epoch 1
  65536


198.51.100.1 (inaccessible)

 from 10.0.23.2 (10.2.2.2)
      Origin incomplete, metric 0, localpref 100, valid, internal
      rx pathid: 0, tx pathid: 0
      Updated on Jul 1 2022 13:44:19 CST
```

On the output, the next hop is the outgoing interface of R1, which is not known by R3. To fix this situation, you can advertise the next-hop via IGP, static route, or run the **neighbor ip-address next-hop-self** command on the iBGP peer to modify the next-hop IP (which is directly connected.) On the diagram example, this configuration must be on R2; the neighbor towards R3 (neighbor 10.0.23.3 next-hop-self.)

As a result, the next hop changes (after a **clear ip bgp 10.0.23.2 soft**) to the directly connected interface (reachable) and the prefix is installed:

```
<#root>

R3#

show ip bgp 192.0.2.1

BGP routing table entry for 192.0.2.1/32, version 24

Paths: (1 available, best #1, table default)

  Not advertised to any peer
  Refresh Epoch 1
  65536


10.0.23.2

 from 10.0.23.2 (10.2.2.2)
      Origin incomplete, metric 0, localpref 100, valid, internal, best
      rx pathid: 0, tx pathid: 0x0
      Updated on Jul 1 2022 13:46:53 CST
```

RIB Failure

This happens when a route cannot be installed into the Global RIB, which results in a RIB failure. Common reasons are when the same prefix is already on RIB for another routing protocol with lower administrative distance, but the exact reason for a RIB failure is seen with the command **show ip bgp rib-failure**.

Note: You can identify and correct these issues by referring to the Understand BGP RIB-failure and The Command bgp suppress-inactive guide.

Race Condition

The most common issue seen is when IGP is preferred over eBGP on a mutual redistribution scenario. When an IGP route is redistributed into BGP, it is considered locally generated by BGP and receives a weight of 32,768 by default. All prefixes received from a BGP peer are assigned a local weight of 0 by default. Therefore, if the same prefix must be compared, the prefix with the higher weight is installed in the routing table based on the BGP best path selection process, which is why the IGP route is installed on RIB.

The solution for this problem is to set a higher weight for all routes received from the BGP peer under the router bgp configuration:

<#root>

neighbor

ip-address

weight 40000

Note: For a detailed explanation, refer to the Understand the Importance of BGP Weight Path Attribute in Network Failover Scenarios guide.

Other Issues

BGP Slow Peer

It is a peer that cannot keep up with the rate a sender generates update messages. There are many reasons for

a peer to exhibit this issue; high CPU in one of the peers, excess traffic, traffic loss on a link, bandwidth resource, among others.

Note: To help identify and correct slow peers issues, refer to Use the BGP "Slow Peer" Feature to Resolve Slow Peer Issues.

Memory Issues

BGP uses memory that is assigned to the Cisco IOS process to maintain network prefixes, best paths, polices, and all related configurations to operate properly. The overall processes are seen by running the **show processes memory sorted** command:

<#root>

```
R1#

show processes memory sorted
Processor Pool Total: 2121414332 Used:  255911152 Free: 1865503180

reserve P Pool Total:     102404 Used:         88 Free:     102316
 lsmpi_io Pool Total:    3149400 Used:    3148568 Free:        832

 PID TTY  Allocated      Freed

Holding

    Getbufs    Retbufs Process
   0    0  266231616   81418808  160053760          0          0 *Init*
 662    0   34427640      51720   34751920          0          0 SBC main process
  85    0    9463568          0    8982224          0          0 IOSD ipc task
   0    0   34864888   25213216    8513400    8616279          0 *Dead*
 504    0     696632          0     738576          0          0 QOS_MODULE_MAIN
 518    0     940000       8616

    613760

       0          0

 BGP Router


 228    0     856064     345488     510080          0          0 mDNS
  82    0     547096     118360     417520          0          0 SAMsgThread
   0    0          0          0     395408          0          0 *MallocLite*
```

The processor pool is the memory used; around 2.1 GB in the example. Next, you must look at the Holding

column to identify the sub-process holding most of it. Then, you need to check the BGP sessions you have, how many routes are received, and configuration used.

Common steps to reduce memory holding by BGP:

- BGP filtering: If it is not necessary to receive a full BGP table, use policies to filter routes and install only the prefixes you need.
- Soft reconfiguration: Look for **neighbor ip_address soft-reconfiguration inbound** under BGP configuration; this command allows you to see all prefixes received before any inbound policy (Adj-RIB-in). However, this table needs around half of the current BGP Local RIB table to store this information so you can avoid this configuration unless it is compulsorily required, or your current prefixes are few.

Note: For further information on how to optimize BGP refer to Configure BGP Routers for Optimal Performance and Reduced Memory Consumption.

High CPU

Routers use different processes for BGP to operate. To verify the BGP process is the cause of high CPU utilization, run the **show process cpu sorted** command.

<#root>

```
R3#

show processes cpu sorted


CPU utilization for five seconds: 0%/0%; one minute: 0%; five minutes: 0%
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 PID Runtime(ms)     Invoked      uSecs   5Sec   1Min   5Min TTY Process
 163          36        1463         24  0.07%  0.00%  0.00%   0 ADJ background
  62          28         132        212  0.07%  0.00%  0.00%   0 Exec
   2          39         294        132  0.00%  0.00%  0.00%   0 Load Meter
   1           0           4          0  0.00%  0.00%  0.00%   0 Chunk Manager
   3          27        1429         18  0.00%  0.00%  0.00%   0

BGP Scheduler


   4           0           1          0  0.00%  0.00%  0.00%   0 RO Notify Timers
  63           4          61         65  0.00%  0.00%  0.00%   0

BGP I/O
```

```
  83         924          26       35538  0.00%  0.03%  0.04%   0

BGP Scanner

  96         142       11651          12  0.00%  0.00%  0.00%   0 Tunnel BGP
   7           0           1           0  0.00%  0.00%  0.00%   0 DiscardQ Backgro
```

These are the common processes, causes, and general steps to overcome high CPU utilization due to BGP:

- **BGP Router**: Runs once per second to safeguard faster convergence (this is an important thread.) It reads the bgp update messages, validates the prefixes/networks and attributes, updates the per AFI/SAFI network/prefix table and attribute table, performs best-path calculation, among other tasks. Huge route churn is a very common scenario that leads to this situation.
- **BGP Scanner**: Low-priority process that runs every 60 seconds by default. This process checks the entire BGP table to verify the next-hop reachability and updates the BGP table accordingly. It runs through the Routing Information Base (RIB) for redistribution purposes. Check the platform scale, as more prefixes and routes installed and TCAM used, more resources are needed.

Note: For more information on how to troubleshoot these two processes, refer to the Troubleshoot High CPU Caused by the BGP Scanner or Router Process

---

- **BGP I/O**: Runs when BGP control packets are received and manages the queuing and processing of BGP packets. If there are excessive packets received in the BGP queue for a long period, or if there is a problem with TCP, the router shows symptoms of high CPU due to the BGP I/O process. Usually, the BGP Router is also high in this situation. Look at the message counts to identify peer and capture packets to identify the source of these messages.
- **BGP Open**: Process used on session establishment; not a common high CPU issue unless a session is stuck in an Open State.
- **BGP Event**: Is responsible for next-hop processing. Look for next-hops flaps on prefixes received.

Related Information

- Technical Support & Documentation - Cisco Systems
- BGP Configuration Guide
- MD5 Authentication Between BGP Peers Configuration Example
- Embedded Packet Capture
- BGP Neighbor Flaps with MTU Troubleshooting

- IANA Address Family Numbers

- Subsequent Address Family Identifiers (SAFI) Parameters

- Unsupported Capabilities Cause BGP Peer Malfunction

- BGP Best Path Selection Algorithm
- Understand BGP RIB-failure and The Command bgp suppress-inactive
- Understand the Importance of BGP Weight Path Attribute in Network Failover Scenarios
- Use the BGP "Slow Peer" Feature to Resolve Slow Peer Issues
- Configure BGP Routers for Optimal Performance and Reduced Memory Consumption
- Troubleshoot High CPU Caused by the BGP Scanner or Router Process