
cisco
You make possible

Troubleshooting BGP

Brad Edgeworth, Systems Architect
CCIE# 31574 @BradEdgeworth

BRKRST-3320

cisco Live!
Barcelona | January 27-31, 2020

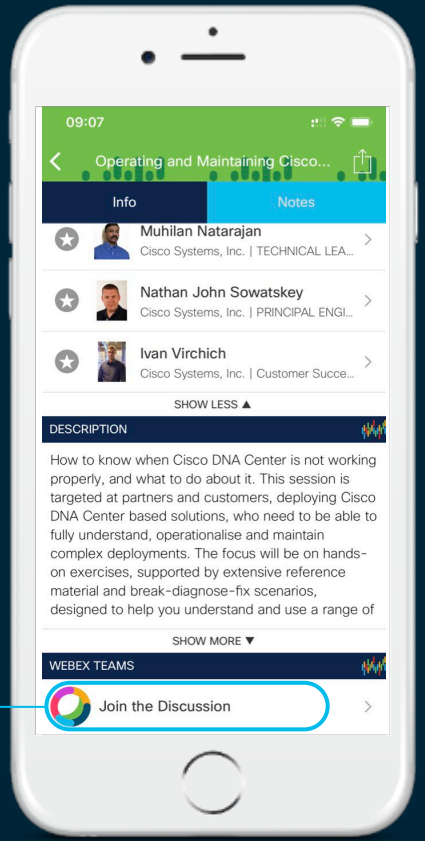
Cisco Webex Teams

Questions?

Use Cisco Webex Teams to chat with the speaker after the session

How

1. Find this session in the Cisco Events Mobile App
2. Click “Join the Discussion”
3. Install Webex Teams or go directly to the team space
4. Enter messages/questions in the team space



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Agenda

- Introduction
- Troubleshooting BGP Peering Issues
- Troubleshooting BGP Routing Issues
  - Missing Routes, Unexpected Routes, Stale Entries
- Troubleshooting BGP Route Churn
- Troubleshooting IPv6 BGP
- Troubleshooting with Cisco NXOS
- Applying Programmability for Troubleshooting
- Conclusion

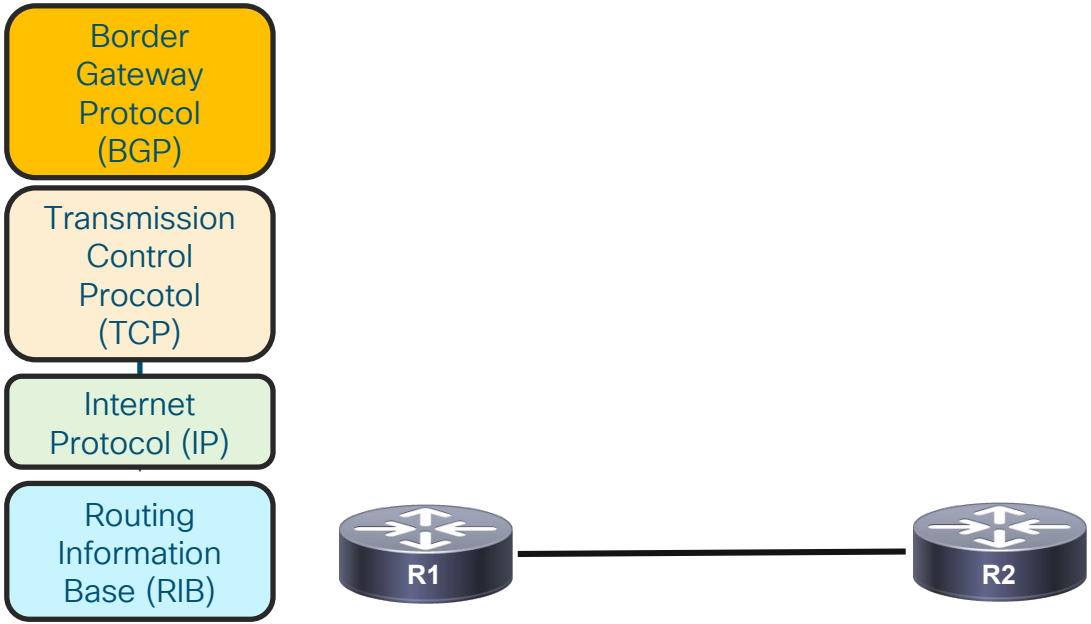
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting BGP Peering Issues

Border Gateway Protocol (BGP)

Is BGP A Routing Protocol?

In some ways we need to think of BGP as a routing application?



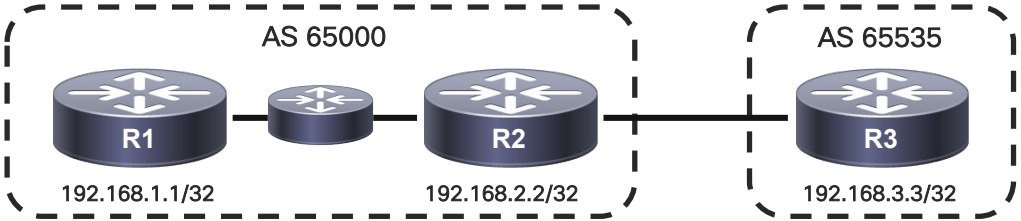
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting BGP Peering Issues

Preliminary Checks

- Verify Configuration
  - ✓ Peering IP Address
  - ✓ AS Number
  - ✓ MD5 Authentication (Optional)
  - ✓ **ebgp-multihop** *hop-count (eBGP only)*

- Verify Reachability
  - ✓ **ping** *remote-ip* **source** *source-ip*
  - If reachability issues found:
    - ✓ Use **traceroute** to verify where the trace is dropping

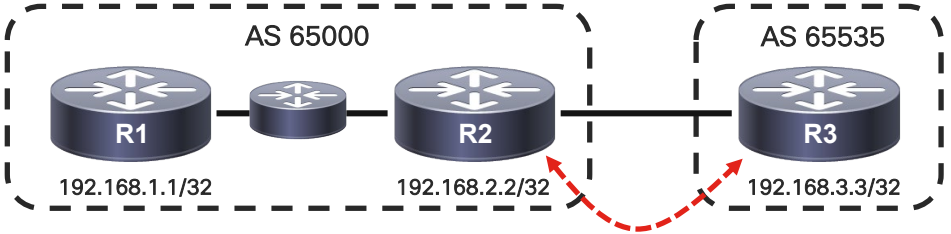



```
router bgp 65000
 bgp router-id 192.168.2.2
 bgp log-neighbor-changes
neighbor 192.168.1.1 remote-as 65000
neighbor 192.168.1.1 update-sour lo0

neighbor 192.168.3.3 remote-as 65535
neighbor 192.168.3.3 password C!sc0
neighbor 192.168.3.3 ebgp-multi 2
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

disable-connected-check



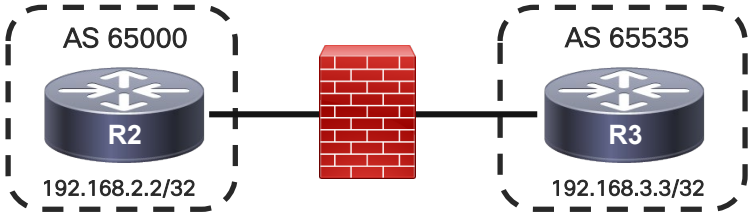
- BGP uses a TTL of 1 for eBGP peers
  - Also verifies if NEXTHOP is directly connected
- For eBGP peers that are more than 1 hop away a larger TTL must be used
  - No longer verifies if NEXTHOP is directly connected
- Use **`neighbor disable-connected-check`**
- Disables the “is the NEXTHOP on a connected subnet” check

```
router bgp 65000
 neighbor 192.168.3.3 remote-as 65535
 neighbor 192.168.3.3 disable-connected-check
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

ACLs & Firewall



- Verify any Firewall / ACLs in the path for TCP port 179

```
R1#telnet 2.2.2.2 179 /source-interface loopback 0
Trying 2.2.2.2 ...
% Destination unreachable; gateway or host down
```

- Ensure BGP Pass-Through configured
  - ASA / PIX offsets TCP sequence number with a random number for every TCP session
    - Causes MD5 authentication to fail
    - ASA strips off TCP option 19

1. Create ACL to permit BGP traffic
2. Create TCP Map to allow TCP option 19
3. Create class-map to match BGP traffic
4. Disable sequence number randomization and Enable TCP option 19 in global policy

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

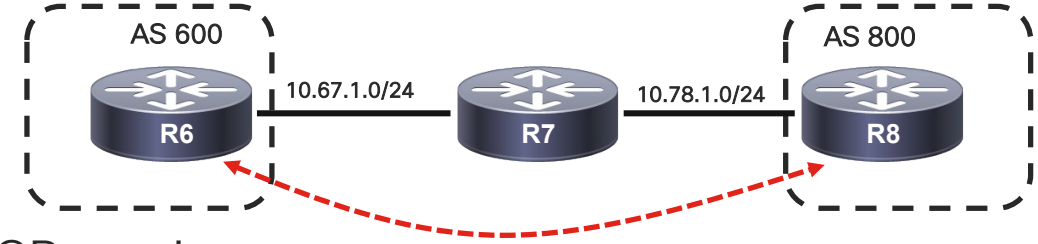
BGP Pass-Through – ASA FW Configuration

```
access-list OUT extended permit tcp host 10.1.12.1 host 10.1.12.2 eq bgp
access-list OUT extended permit tcp host 10.1.12.2 eq bgp host 10.1.12.2
!
access-list BGP-TRAFFIC extended permit tcp host 10.1.110.2 host 10.1.110.10 eq bgp
access-list BGP-TRAFFIC extended permit tcp host 10.1.110.2 eq bgp host 10.1.110.10
!
tcp-map TCP-OPTION-19
tcp-options range 19 19 allow
!
access-group OUT in interface Outside
!
class-map BGP_TRAFFIC
match access-list BGP-TRAFFIC
!
policy-map global_policy
  class BGP_TRAFFIC
    set connection random-sequence-number disable
    set connection advanced-options TCP-OPTION-19
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Troubleshooting Scenario



R6 and R8 cannot establish an eBGP session.

- What do we do?
- Check the configuration for basics.
- View the BGP Summary State
- Can we ping between R6 & R8?
- Can we telnet into each other on port 179?

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Problem with TCP Process (show tcp brief)

PCB	Recv-Q	Send-Q	Local Address	Foreign Address	State
0x48277ea4	**0**	**0**	**:::179**	**:::0**	**LISTEN**
0x48276c50	0	0	0.0.0.0:23	0.0.0.0:0	LISTEN
0x48290da8	0	0	12.26.28.152:23	223.255.254.249:48877	ESTAB
0x4827755c	**0**	**0**	**0.0.0.0:179**	**0.0.0.0:0**	**LISTEN**

- PCB is the internal identifier used by TCP. It can be used as input to other show commands.
- Recv-Q shows how much received data is waiting to be “read” from TCP by application.
- Send-Q shows how much application data is waiting to be “sent” by TCP.
- Local-address and foreign address identify the two end points of the connection.
- State identifies the current state of the connection.

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Most Common TCP States

- LISTEN
  - A listen socket on which incoming connections will be accepted.
- ESTAB
  - An established connection
- CLOSED
  - Socket not fully programmed – most often seen on standby RP by applications that are warm or hot standby.

Connections that are getting established:

- SYNSENT
  - A SYN message was sent to peer.
- SYNRCVD
  - A SYN message was received from peer – socket will move into ESTAB state.

Connections that are getting terminated:

- CLOSEWAIT, CLOSING, LASTACK, TIMEWAIT, FINWAIT1, FINWAIT2

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues
Detailed info about a TCP Socket

```
XR# show tcp detail pcb 0x48277a58

  Connection state is ESTAB, I/O status: 0, socket status: 0
  PCB 0x48277a58, vrfid 0x60000000, Pak Prio: Unspecified, TOS: 16, TTL: 255
  Local host: 12.26.28.152, Local port: 179 (Local App PID: 180393)
  Foreign host: 223.255.254.249, Foreign port: 49017
  Current send queue size in bytes: 0 (max 16384)
  Current receive queue size in bytes: 0 (max 16384)  mis-ordered: 0 bytes
  Current receive queue size in packets: 0 (max 50)
```

- **Pak Prio**: Did the application mark the packet with correct priority? Determines the queuing within the router before it goes out on the wire.
- **TOS**: Type of service, goes out on the wire.
- **TTL**: Important for eBGP for the TTL security check
- **Mis-ordered**: How much of the received data is out-of-order?
- **Receive queue in packets**: how many packets are sitting in receive buffers?

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Malformed Update

- What if a peer sends you a message that causes us to send a NOTIFICATION?
  - Corrupt UPDATE
  - Bad OPEN message, etc.
- View the message that triggered the NOTIFICATION

```
show ip bgp neighbor 1.1.1.1 | begin Last reset

Last reset 5d12h, due to BGP Notification sent, invalid or corrupt AS path
Message received that caused BGP to send a Notification:
FFFFFFFF FFFFFFFF FFFFFFFF FFFFFFFF
005C0200 00004140 01010040 0206065D
1CFC059F 400304D5 8C20F480 04040000
05054005 04000000 55C0081C 329C4844
329C6E28 329C6E29 58F50082 58F5EACE
58F5FA02 58F5FA6E 18D14E70
```

http://bgpaste.convergence.cx/

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Unsupported Capability

```
*Jan  5 18:18:04.667: %BGP-3-NOTIFICATION: sent to neighbor
10.1.12.1 active 2/7 (unsupported/disjoint capability) 0 bytes

R2#

*Jan  5 18:18:04.671: %BGP-4-MSGDUMP: unsupported or mal-formatted
message received from 10.1.12.1:

FFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF 002D 0104 0064 00B4 0101
0101 1002 0601 0400 0100 0102 0280 0002 0202 00
```

- Disable capability negotiation during session establishment process using the below hidden command

`neighbor x.x.x.x dont-capability-negotiate`

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

IPv4 vs IPv6

- BGPv4 carries only 3 pieces of information which are truly IPv4-specific:
  - NLRI in the UPDATE message contains an IPv4 prefix
  - NEXT_HOP path attribute in the UPDATE message contains an IPv4 address
  - BGP Router ID derived from an IPv4 address is in the OPEN message and AGGREGATOR attribute
- Adapting BGP for IPv6 targets primarily the NLRI and NEXT_HOP
  - NLRIs need to carry IPv6 prefixes
  - NEXT_HOP attribute needs to carry an IPv6 address; note that it may carry both global and link-local address
  - Router ID remains a 4-byte address

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

BGP-4 Extension for IPv6 – NH Information & Router-ID

- NEXT_HOP carries a global IPv6 address, may be followed with a link-local
  - Link-local address as a next-hop is only set if the eBGP peer shares the subnet with both routers (advertising and advertised)
- The length of the NEXT_HOP field in the MP_REACH_NLRI attribute is
  - 16 bytes when only global address is present
  - 32 bytes when global and link-local are both present
- Router ID
  - When no IPv4 interface is configured, an explicit 4B BGP RID needs to be set
  - The RID for BGP must not be a “martian” (falling under 0.0.0.0/8, 127.0.0.0/8, 224.0.0.0/4), even though it is not used as a real IPv4 address

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

IPv6 AFI / SAFI Peering (IOS-XR Criteria)

- IPv4 session:
  - Neighbor source interface (update-source) must have a global IPv6 address
- IPv6 session:
  - Neighbor source interface must have at least a link-local IPv6 address
- A BGP session will not come up without the above configurations. These requirements stem from two facts:
  - BGP needs to set a valid IPv6 nexthop when IPv6 updates are being sent over an IPv4 session
  - eBGP needs to send both the global and the LL next hops while sending update to a directly connected IPv6 neighbor, and the neighbor needs both these next hops

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

IPv6 ND and NS

- Ensure **`ipv6 unicast-routing`** is configured globally on IOS / IOS-XE
- Ensure IPv6 ND is completed and no stale IPv6 neighbor (Global) is present
- Ensure correct IPv6 global address is being used for peering

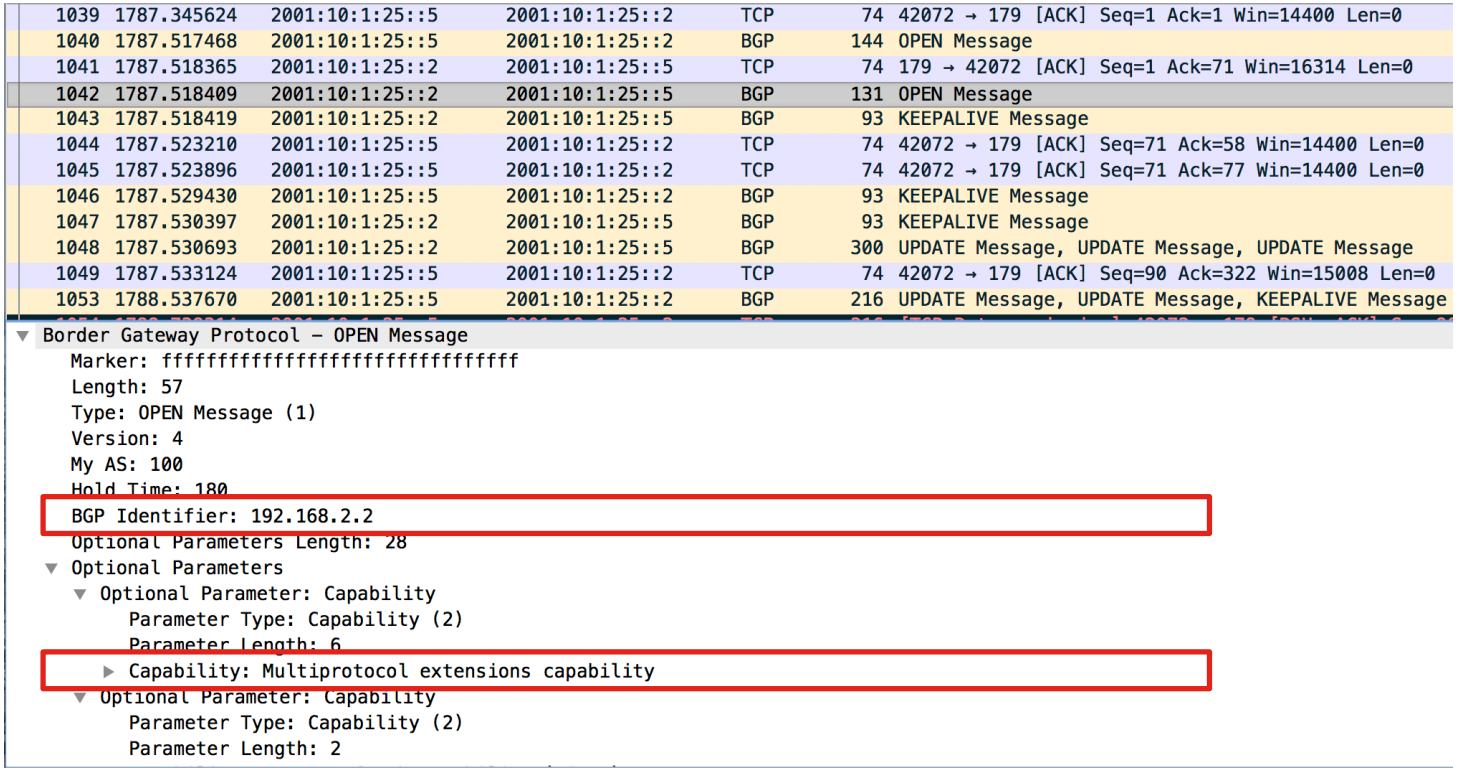
```
R2# show ipv6 neighbors
IPv6 Address                                 Age Link-layer Addr State
Interface
2001:10:1:25::5                                0 0cd1.9c2f.6807  REACH Gi2
FE80::ED1:9CFF:FE2F:6807                      19 0cd1.9c2f.6807  STALE Gi2
```

```
NX5# show ipv6 neighbor
Address         Age       MAC Address     Pref Source      Interface
2001:10:1:25::2 02:24:31  0cd1.9c43.0601  50   icmpv6      Ethernet1/2
fe80::ed1:9cff:fe43:601
                02:18:04  0cd1.9c43.0601  50   icmpv6      Ethernet1/2
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

IPv6 BGP Peering

OPEN Message



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Peering Issues

Stable BGP peers going into Idle State

- BGP Peering has been up for months, but all of a sudden, BGP session goes down and never comes back up
  - IGP goes down as well? Yes
- Debug shows keepalives are getting generated
- Check for the Interface Queue on both sides
  - Interface Queue (both input and output queue) getting wedge can cause this symptom
    - Temporary workarounds – Increase the Queue size, RP Switchover
    - If its input queue wedge, check the **`show buffer input-interface`** *`x/y`* **`packet`** to analyze what packets are stuck in queue. Also checking for incoming traffic rate
    - If its output queue, check for outgoing traffic rate. Check the transmission side

```
R2#show interface gi0/1 | in queue
 Input queue: 0/375/0/0 (size/max/drops/flushes); Total output drops: 0
  Output queue: 1001/1000 (size/max)
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Notifications – Hold Time Expired


```
%BGP-5-ADJCHANGE: neighbor 2.2.2.2 Down BGP Notification sent
%BGP-3-NOTIFICATION: sent to neighbor 2.2.2.2 4/0 (hold time expired)

R1#show ip bgp neighbor 2.2.2.2 | include last reset
Last reset 00:01:02, due to BGP Notification sent, hold time expired
```

- R1 sends hold time expired NOTIFICATION to R2
  - R1 did not receive a KA from R2 for holdtime seconds
- One of two issues
  - R2 is not generating keepalives
  - R2 is generating keepalives but R1 is not receiving them

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Notifications – Hold Time Expired

- Check if R2 is building keepalives (KA)
  - Check for output drops on R2 outgoing interface
  - When did R2 last build a BGP message for R1. (Should be within "keepalive interval" seconds

```
R2#show ip bgp neighbors 1.1.1.1
   Last read 00:00:15, last write 00:00:44, hold time is 180,
     keepalive interval is 60 seconds
```

- R2 is building messages for R1 but possibly R2 is unable to send them
  - Check OutQ and MsgSent Counters – **`show bgp afi safi summary`**
  - OutQ is the number of packets waiting for TCP to Tx to a peer
  - MsgSent is the number of packets TCP has removed from OutQ and transmitted for a peer

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Peering Issues

Notifications – Hold Time Expired

```
R2#show ip bgp sum | begin Neighbor
Neighbor …  MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  State/PfxRcd
1.1.1.1  …       53      284    10167    0   97 00:01:20        0
```

The number of packets transmitted is not increasing ☹

The number of packets generated is increasing

At least one BGP keepalive interval apart

```
R2#show ip bgp sum | begin Neighbor
Neighbor …  MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down  State/PfxRcd
1.1.1.1  …       53      284    10167    0   98 00:02:24        0
```

OutQ is incrementing due to keepalive generation
MsgSent is not incrementing
Something is “stuck” on the OutQ
The keepalives are not leaving R2!!

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Randomly Flapping Peers

Flapping continuously but not at regular intervals…

- What if the BGP peer is flapping continuously, but not at regular intervals.
  - Sometimes it flaps every 2 minutes and sometimes it flaps after 5 minutes

```
R2#show ip bgp sum | begin Neighbor
Neighbor        …    MsgRcvd MsgSent     TblVer  InQ OutQ  Up/Down    State/PfxRcd
10.1.13.3       …        160     284      10167    0    0    00:01:20          10
```

```
R2#show ip bgp sum | begin Neighbor
Neighbor        …    MsgRcvd MsgSent     TblVer  InQ OutQ  Up/Down    State/PfxRcd
10.1.13.3       …        165     296      10167    0    0    00:00:39          10
```

  - Most probable cause could be keepalives are not getting generated in timely manner
  - Or, they are not being forwarded out in a timely manner

Can be identified by using BGP Debugs and/or packet captures

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Randomly Flapping Peers

ASR1k – EPC Capture and Debugs

```
ASR1k(config)#ip access-list extended MYACL
ASR1k(config-acl)#permit tcp host 10.1.13.1 eq bgp host 10.1.13.3
ASR1k(config-acl)#permit tcp host 10.1.13.1 host 10.1.13.3 eq bgp
ASR1k#monitor capture CAP1 buffer circular packets 1000
ASR1k#monitor capture CAP1 buffer size 10
ASR1k#monitor capture CAP1 interface GigabitEthernet0/0/0 in
ASR1k#monitor capture CAP1 access-list MYACL
ASR1k#monitor capture CAP1 start
ASR1k#monitor capture CAP1 stop
ASR1k#monitor capture CAP1 export bootflash:cap1.pcap
```

```
ASR1k#debug ip bgp keepalives
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Randomly Flapping Peers

ASR1k – EPC Capture

```
ASR1k#show monitor capture buffer CAP1 dump
16:25:44.938 JST Aug 21 2015 : IPv4 LES CEF    : Gig0/0 None

F19495B0:                      AABBCC00 0800AABB          *;L...*;
F19495C0: CC000700 08004540 003B1C5D 4000FE06  L.....E@.;.]@.~.
F19495D0: 42020707 07070808 08084A07 00B39372  B.........J..3.r
F19495E0: FFE37CDC E3D35018 3D671161 0000FFFF  .c|\cSP.=g.a....
F19495F0: FFFFFFFF FFFFFFFF FFFFFFFF FD        ............}
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Flapping BGP Peers

Regular Interval Flaps

```
*Jun 22 15:16:23.033: %BGP-3-NOTIFICATION: received from neighbor
192.168.2.2 4/0 (hold time expired) 0 bytes
```

```
*Jun 22 15:16:23.033: %BGP-5-ADJCHANGE: neighbor 192.168.2.2 Down
BGP Notification received
```

```
*Jun 22 15:16:55.621: %BGP-5-ADJCHANGE: neighbor 192.168.2.2 Up
```

```
*Jun 22 15:19:56.409: %BGP-3-NOTIFICATION: received from neighbor
192.168.2.2 4/0 (hold time expired) 0 bytes
```

```
*Jun 22 15:19:56.409: %BGP-5-ADJCHANGE: neighbor 192.168.2.2 Down
BGP Notification received
```

```
*Jun 22 15:20:13.361: %BGP-5-ADJCHANGE: neighbor 192.168.2.2 Up
```

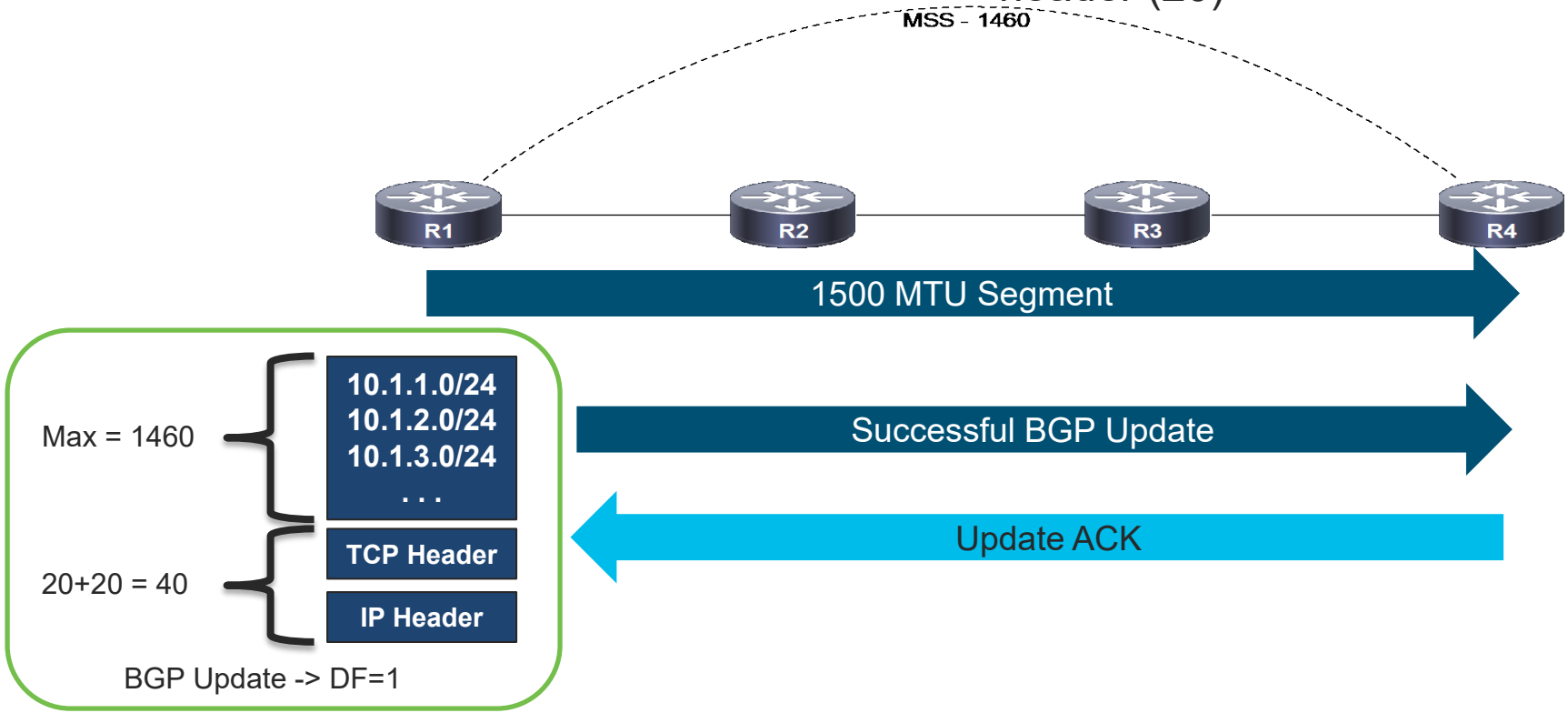

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Flapping BGP Peers

BGP Update Mechanism – Segment 1

MSS Calculation

MSS = MTU – IP Header (20) – TCP header (20)

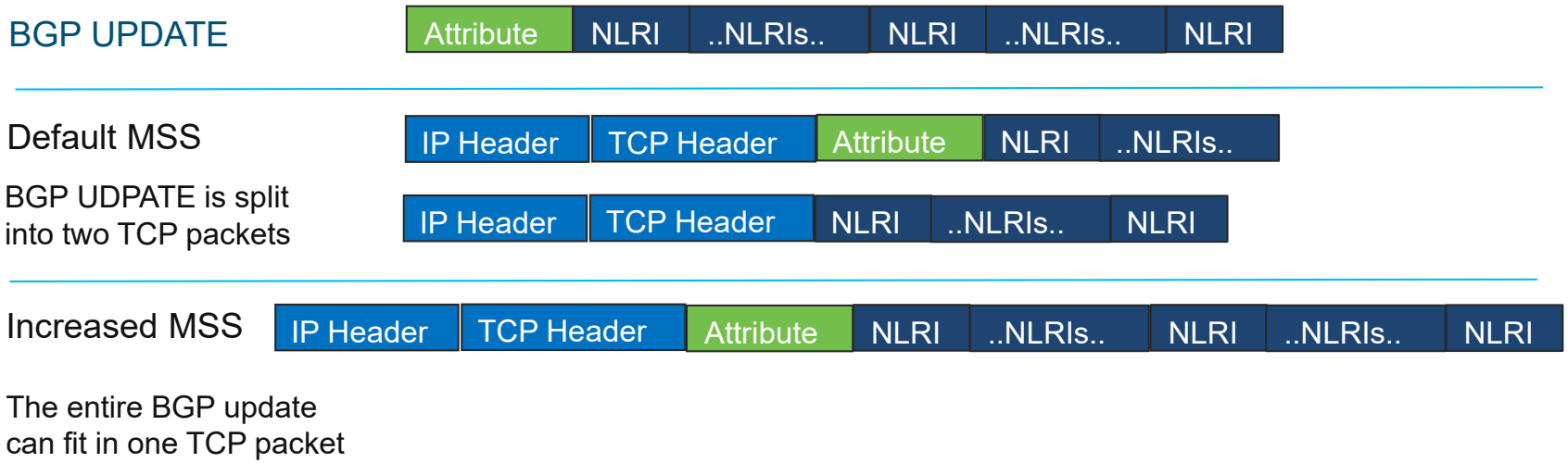


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Update

Role of TCP MSS

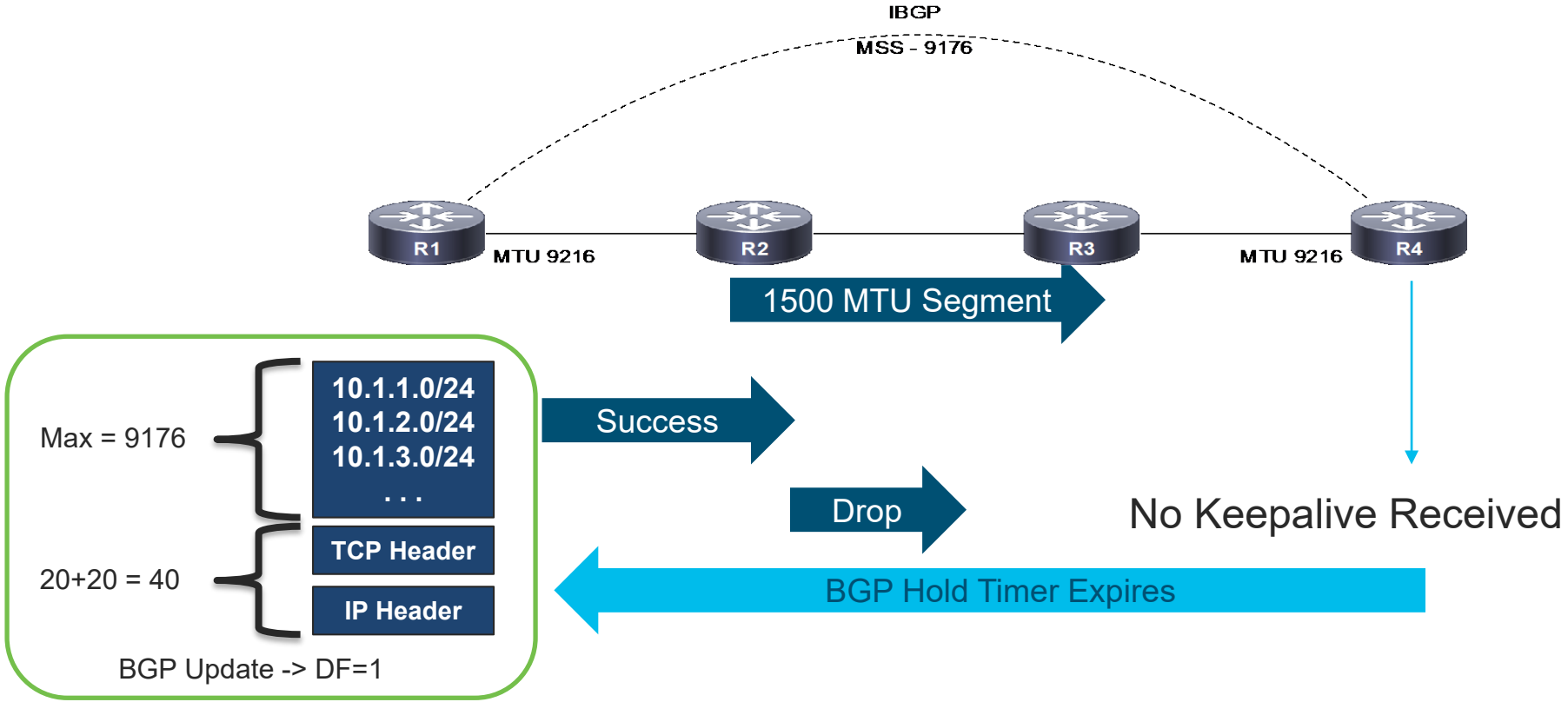
TCP MSS (max segment size) is also a factor in convergence times. The larger the MSS, the fewer TCP packets it takes to transport the BGP updates. Fewer packets means less overhead and faster convergence.



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Flapping BGP Peers

BGP Update Mechanism – Segment 2

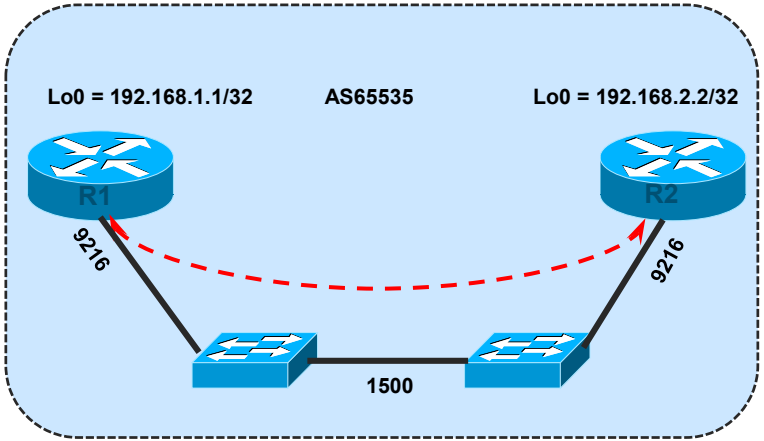


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Flapping BGP Peers

Path MTU Discovery

- R1 sends a packet with packet size of outgoing interface MTU and DF-bit set
- Intermittent device having a lower MTU has three options
  - Fragment and send the packets (if DF-bit is cleared)
  - Drop the packet and send ICMP error message Type 3 Code 4
  - Drop the packet silently ☹
- ICMP error message also has the MTU details in the Next-Hop MTU field
- Upon receiving the message, source can decrease the packet size accordingly



Type 3 – Destination Unreachable
Code 4 – Fragmentation needed and DF-bit set

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Flapping BGP Peers
Notifications – Hold Time Expired

- MSS ping
  - BGP OPENs and Keepalives are small
  - UPDATEs can be much larger
  - Maybe small packets work but larger packets do not?

```
R1# ping 192.168.2.2 source loop0
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 16/21/24 ms

R1# ping 192.168.2.2 source loop0 size 1500 df-bit

Type escape sequence to abort.
Sending 5, 1500-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:
Packet sent with the DF bit set
.....
Success rate is 0 percent (0/5)
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Route Filtering

Missing Routes / Stale Routes

What does it mean?

- Missing Routes
  - The remote peer has not received the route
    - Either the speaker did not advertise the routes, or the remote peer did not receive or process the BGP update
    - Inbound / Outbound route-maps (Filtering)
- Stale Routes
  - A route is present in the local BGP table as learned from remote peer, but the peer no longer has that route present in its own BGP table
    - Either remote speaker did not advertise the withdraw, or the local device did not process the withdraw
    - EOR (End of RIB) received

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Missing Routes

BGP not in read-write mode

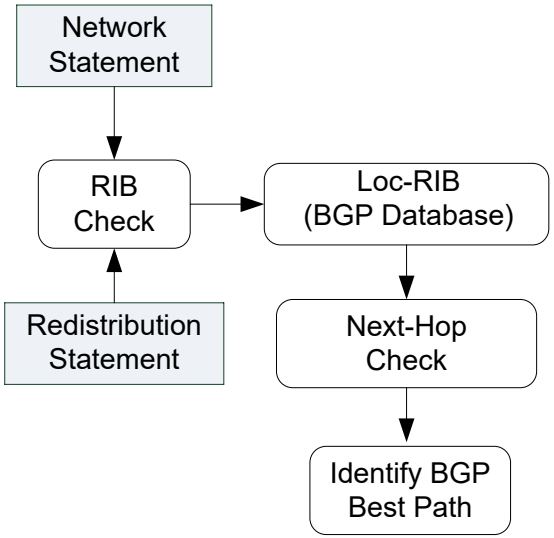
- May not see the routes in BGP table in case BGP remains in read-only mode
  - To have the BGP routes installed, BGP should be in read-write mode
- On XR, use the below commands to verify BGP in read-write mode
  - **`show bgp`**
  - **`show bgp process performance-statistics detail`**
    - At the very bottom of this output, you will see the lines shown below if the device entered the read-write mode

```
First neighbor established:  Jan 23 20:15:45
Entered DO_BESTPATH mode:    Jan 23 20:15:49
Entered DO_IMPORT mode:      Jan 23 20:15:49
Entered DO_RIBUPD mode:      Jan 23 20:15:49
Entered Normal mode:         Jan 23 20:15:49
Latest UPDATE sent:          Jan 23 20:18:39
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route Learning and Propagation Flow

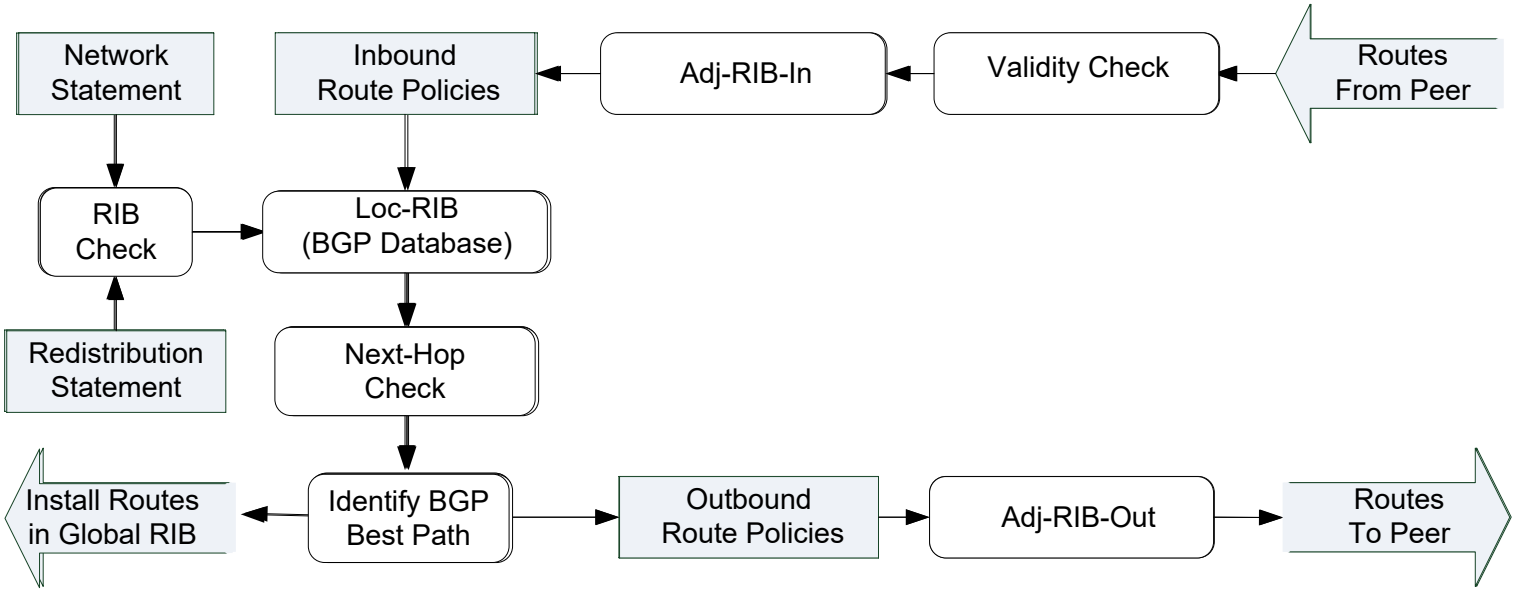


- BGP prefixes are injected by explicit configuration
- Network statement – **`network`** *`prefix`* **`mask`** *`mask`*
  - Prefix/mask needs to match the RIB exactly
  - Does not enable BGP on an interface like IGPs
- Redistribution – **`redistribute`** …
  - Injects prefixes from the specified protocol
  - Does not inject 0.0.0.0/0
- Aggregate route – **`aggregate-address`** *`prefix mask`*
  - Component route must exist in BGP
  - Aggregator attribute is added
- Default route – **`default-information originate`** and **`neighbor`** *`X.X.X.X`* **`default-originate`**

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route Learning and Propagation Flow



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Filtering Techniques

IOS and NX-OS:

- Prefix List (Based on destination networks)
- Filter-list (AS_Path)
- Route-map (Matching on a variety of path attributes in the NLRI

IOS XR:

- Route Policy Language (RPL)
- Programmatic, supports nesting, and multiple operations

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route-Map Behavior

- A route-map consists of one or more sections (blocks), each of them with a unique sequence number and an associated action (permit / deny)
- A route-map processes routes or IP packets in a linear fashion, that is, starting from the section with lowest sequence number
- If the referred policies (for example, prefix lists) within a match statement of a route-map section return either a no-match or a deny-match, the router will inspect the next route-map section
- If there is no applicable match statement in a route-map section, the fate of all routes or packets inspected at this stage is determined by the action of the route-map section being inspected

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route-Map Behavior

- What is the outcome of the above redistribution ?

```
ip prefix-list OSPF2BGP seq 15 permit 10.0.0.0/7 ge 8
!
route-map OSPF2BGP permit 10
 match ip prefix-list FILTERv4
!
router bgp 100
 address-family ipv4 unicast
   redistribute ospf 1 route-map OSPF2BGP
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route-Map Problem

- What is the outcome of the above redistribution ?

```
ip prefix-list FILTERv4 seq 10 permit 10.0.0.0/7 ge 8
ipv6 prefix-list FILTERv6 seq 5 permit 2001:2::/48 ge 48

route-map OSPF2BGP permit 10
 match ip prefix-list FILTERv4
route-map OSPF2BGP permit 20
 match ipv6 prefix-list FILTERv6
!
router bgp 100
 address-family ipv4 unicast
   redistribute ospf 1 route-map OSPF2BGP
 address-family ipv6 unicast
   redistribute ospfv3 1 route-map OSPF2BGP
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

Route-Map Problem

- What is the outcome of the above redistribution ?

```
ip prefix-list FILTERv4 seq 10 permit 10.0.0.0/7 ge 8
ipv6 prefix-list FILTERv6 seq 5 permit 2001:2::/48 ge 48

route-map OSPF2BGP permit 10
 match ip prefix-list FILTERv4
route-map OSPF2BGP permit 20
 match ipv6 prefix-list FILTERv6
!
router bgp 100
 address-family ipv4 unicast
   redistribute ospf 1 route-map OSPF2BGP
 address-family ipv6 unicast
   redistribute ospfv3 1 route-map OSPF2BGP
```

Not a Conditional IPv6 Match

Not a Conditional IPv4 Match

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering
IOS XR RPL

There is an implicit drop at the end of RPL processing.
A route must be given a '**ticket**' to ensure that it has been inspected by the RPL

- Pass – prefix allowed if not later dropped
  - **`pass`** grants a ticket to defeat default drop
  - Execution continues after pass
- Set – value changed, prefix allowed if not later dropped
  - Any **`set`** at any level grants a ticket
  - Execution continues after **`set`**
  - Values can be set more than once
- Drop – prefix is discarded
  - Explicit drop stops policy execution
  - Implicit drop (if policy runs to end without getting a ticket)
- Done – accepts prefix and stops processing

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

IOS XR RPL

RPL to Pass Everything

```
route-policy PASS-ALL
  pass
end-policy
```

RPL to Drop Everything

```
route-policy DROP-ALL
  drop
end-policy
```

RPL for Filtering by Prefixes

```
if destination in (10.0.0.0/8 ge 8, 172.16.0.0/12 ge 12, 192.168.0.0/16 ge 16) then
   pass
 else
   drop
 endif
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Advertisement and Filtering

IOS XR RPL

Bad RPL Logic

```
route-policy METRIC-MODIFICATION
  if destination in (10.0.0.0/8 ge 8) then
    set med 100
  endif
  set med 200
end-policy
```

Overwrites Setting

Good RPL Logic
Option #1

```
route-policy METRIC-MODIFICATION
  if destination in (10.0.0.0/8 ge 8) then
    set med 100
  else
    set med 200
  endif
end-policy
```

Option #2

```
route-policy METRIC-MODIFICATION
  if destination in (10.0.0.0/8 ge 8) then
    set med 100
    done
  endif
  set med 200
end-policy
```

Stops all processing on matched prefixes

Missing Routes
RPL in IOS XR

- IOS and NX-OS by default install routes in the BGP table for prefixes learned from eBGP peers
- IOS XR requires a mandatory RPL policy to have them installed in BGP table.

```
RP/0/0/CPU0: 16:28:06.171 : bgp[1047]: %ROUTING-BGP-6-NBR_NOPOLICY : No inbound IPv4
Unicast policy is configured for eBGP neighbor 10.0.0.1. No IPv4 Unicast prefixes
will be accepted from the neighbor until inbound policy is configured.

RP/0/0/CPU0:16:28:06.171 : bgp[1047]: %ROUTING-BGP-6-NBR_NOPOLICY : No outbound IPv4
Unicast policy is configured for eBGP neighbor 10.0.0.1. No IPv4 Unicast prefixes
will be sent to the neighbor until outbound policy is configured.
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Missing Routes
RPL in IOS XR

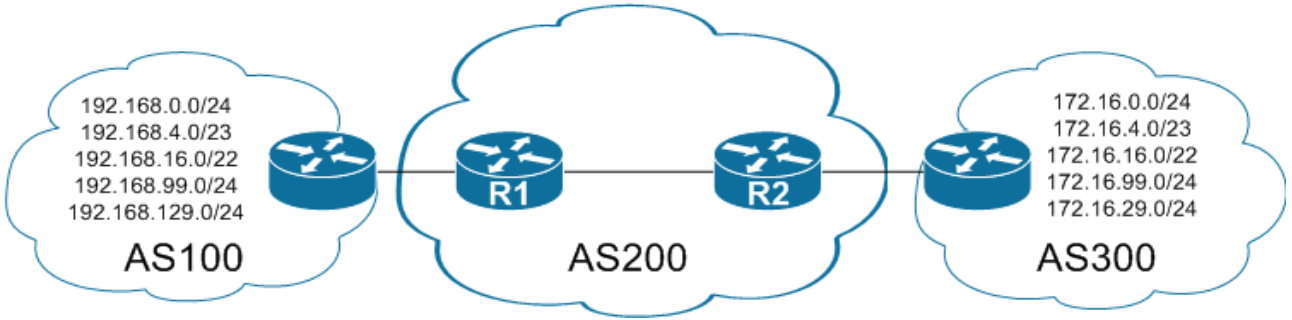
- Th

```
route-policy INBOUND-ROUTES
 if destination in A1-PREFIX-SET then
    pass
 else
    drop
 endif
end-policy

route-policy PASS-ALL
 pass
end-policy

 router bgp 100
 !
  neighbor 10.0.0.1
   remote-as 200
   address-family ipv4 unicast
    route-policy INBOUND-ROUTES in
    route-policy PASS-ALL out
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

Topology




```
R2# show bgp ipv4 unicast
Network          Next Hop       Metric LocPrf Weight Path
*> 172.16.0.0/24    192.168.200.3       0             0 300 80 90 21003 2100 i
*> 172.16.4.0/23    192.168.200.3       0             0 300 1080 1090 1100 1110 i
*> 172.16.16.0/22   192.168.200.3       0             0 300 11234 21234 31234 i
*> 172.16.99.0/24   192.168.200.3       0             0 300 40 i
*> 172.16.129.0/24  192.168.200.3       0             0 300 10010 300 30010 30050 i
*>i192.168.0.0      10.12.1.1           0    100      0 100 80 90 21003 2100 i
*>i192.168.4.0/23   10.12.1.1           0    100      0 100 1080 1090 1100 1110 i
*>i192.168.16.0/22  10.12.1.1           0    100      0 100 11234 21234 31234 i
*>i192.168.99.0     10.12.1.1           0    100      0 100 40 i
*>i192.168.129.0    10.12.1.1           0    100      0 100 10010 300 30010 30050 i
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

Regex Query Modifiers

Modifier	Description
_ (Underscore)	Matches a space
^ (Caret)	Indicates the start of the string
$ (Dollar Sign)	Indicates the end of the string
[] (Brackets)	Matches a single character of the set
- (Hyphen)	Indicates a range of numbers or characters in brackets
[^] (Caret in Brackets)	Excludes the characters listed in brackets; must be the first character
() (Parentheses)	Used for nesting of search patterns
| (Pipe)	Provides 'or' functionality to the query
. (Period)	Matches a single character, including a space
* (Asterisk)	Matches zero or more preceding characters or patterns
+ (Plus Sign)	Matches one or more preceding characters or patterns
? (Question Mark)	Matches at most one preceding characters or patterns

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

Regex

```
R2# show bgp ipv4 unicast regexp _300_
! Output omitted for brevity
     Network          Next Hop      Metric LocPrf Weight Path
*> 172.16.0.0/24    192.168.200.3         0             0 300 80 90 21003 455 i
*> 172.16.4.0/23    192.168.200.3         0             0 300 878 1190 1100 1010 i
*> 172.16.16.0/22   192.168.200.3         0             0 300 779 21234 45 i
*> 172.16.99.0/24   192.168.200.3         0             0 300 145 40 i
*> 172.16.129.0/24  192.168.200.3         0             0 300 10010 300 1010 40 50 i
*>i192.168.129.0    10.12.1.1             0    100      0 100 10010 300 1010 40 50 i
```

```
R2# show bgp ipv4 unicast regexp ^300_
! Output omitted for brevity
     Network          Next Hop      Metric LocPrf Weight Path
*> 172.16.0.0/24    192.168.200.3         0             0 300 80 90 21003 455 i
*> 172.16.4.0/23    192.168.200.3         0             0 300 878 1190 1100 1010 i
*> 172.16.16.0/22   192.168.200.3         0             0 300 779 21234 45 i
*> 172.16.99.0/24   192.168.200.3         0             0 300 145 40 i
*> 172.16.129.0/24  192.168.200.3         0             0 300 10010 300 1010 40 50 i
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

Regex

```
R2# show bgp ipv4 unicast regexp [4-8]0_
! Output omitted for brevity
     Network          Next Hop    Metric LocPrf Weight Path
*> 172.16.0.0/24    192.168.200.3      0             0 300 80 90 21003 455 i
*> 172.16.99.0/24   192.168.200.3      0             0 300 145 40 i
*> 172.16.129.0/24  192.168.200.3      0             0 300 10010 300 1010 40 50 i
*>i192.168.0.0      10.12.1.1          0    100      0 100 80 90 21003 455 i
*>i192.168.99.0     10.12.1.1          0    100      0 100 145 40 i
*>i192.168.129.0    10.12.1.1          0    100      0 100 10010 300 1010 40 50 i
```

```
R2# show bgp ipv4 unicast regexp ^[13]00_[^3-8]
! Output omitted for brevity
     Network          Next Hop    Metric LocPrf Weight Path
*> 172.16.99.0/24   192.168.200.3      0             0 300 145 40 i
*> 172.16.129.0/24  192.168.200.3      0             0 300 10010 300 1010 40 50 i
*>i192.168.99.0     10.12.1.1          0    100      0 100 145 40 i
*>i192.168.129.0    10.12.1.1          0    100      0 100 10010 300 1010 40 50 i
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

Prefix-List Blocking Prefixes

```
RTR# debug bgp ipv4 unicast updates in
BGP updates debugging is on (inbound) for address family: IPv4 Unicast

RTR# clear bgp ipv4 unicast 10.1.45.4 soft in
! Output omitted for brevity
* 18:59:42.515: BGP(0): process 10.1.12.0/24, next hop 10.1.45.4, metric 0 from 10.1.45.4
* 18:59:42.515: BGP(0): Prefix 10.1.12.0/24 rejected by inbound filter-list.
* 18:59:42.515: BGP(0): update denied
```

```
NXOS5# debug bgp updates
NXOS5# clear bgp ipv4 unicast 10.1.45.4 soft in
! Output omitted for brevity
19:02:54 bgp: 300 [8449] UPD: [IPv4 Unicast] 10.1.45.4 Inbound as-path-list 1, action permit
 19:02:54 bgp: 300 [8449] UPD: [IPv4 Unicast] 10.1.45.4 Inbound as-path-list 1, action deny
 19:02:54 bgp: 300 [8449]  UPD: [IPv4 Unicast] Dropping prefix 10.1.12.0/24 from peer 10.1.45.4,
due to attribute policy rejected
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting Filtering

IOS XR BGP RPL Debugging

```
route-policy R4-IN
  if destination in (10.0.0.0/8 le 32) then
    pass
  endif
  if destination in (172.16.0.0/12 le 32) then
    set med 20
  endif
end-policy
```

```
RP/0/0/CPU0:XR# debug bgp policy-execution events
RP/0/0/CPU0:XR# clear bgp ipv4 unicast 10.1.45.4 soft
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]: --Running policy 'R4-IN':---
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Attach pt='neighbor-in-dflt'
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Attach pt inst='default-IPv4-Uni-10.1.45.4'
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]: Input route attributes:
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   as-path: 200 100 600
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   as-path-length: 3
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   as-path-unique-length: 3
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   community: No Community Information
. . .
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   path-type: ebgp
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   aigp-metric: 0
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   validation-state: not-found
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]: Policy execution trace:
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Condition: destination in (10.0.0.0/8 ...)
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Condition evaluated to FALSE
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Condition: destination in (172.16.0.0/12 ...)
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   Condition evaluated to FALSE
RP/0/0/CPU0: 06:19:10.000 : bgp[1053]:   End policy: result=DROP
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

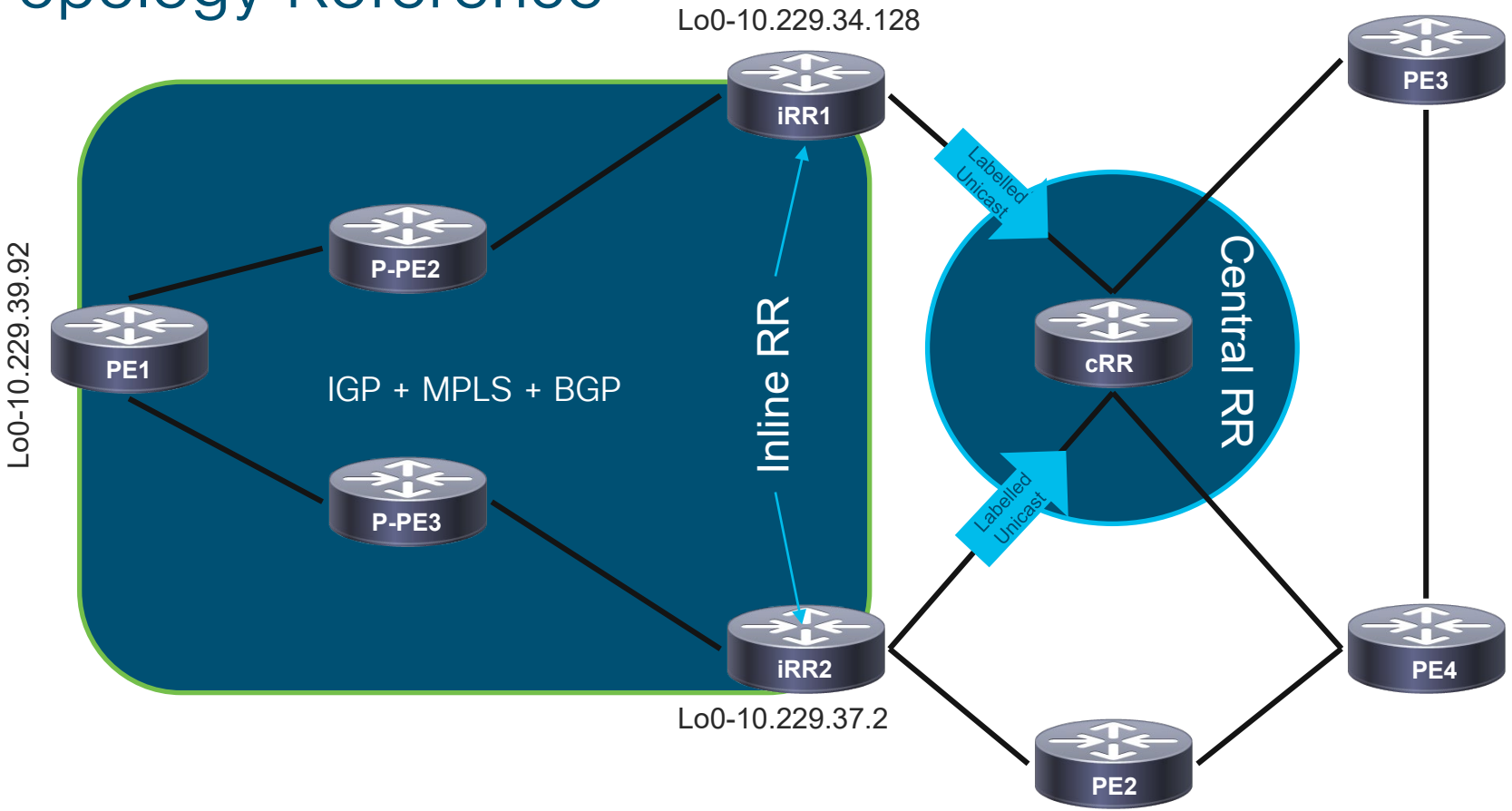
Stale Routes

Symptoms and Possible Causes

- Symptoms
  - Stale Entry to BGP Peer
  - Traffic Black-Hole
  - Outage
- Possible Causes
  - BGP Slow Peer
  - Sender did not send the update
  - Receiver did not process the update

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Topology Reference



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stale Routes

Example – Route on BGP Speaker

```
RP/0/RSP0/CPU0:iRR2# show bgp ipv4 labeled-unicast 10.229.37.92
BGP routing table entry for 10.229.37.92/32
   Local Label: 25528
Last Modified: Jan 13 10:20:52.424 for 11:45:15
Paths: (1 available, best #1)
    Path #1: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.1 0.2 0.3 0.7
  Local
    10.229.34.128 (metric 5) from 192.168.53.9 (10.229.37.92)
      Received Label 26596
      Origin IGP, metric 0, localpref 100, valid, internal, best, group-best
      Received Path ID 1, Local Path ID 0, version 301642
      Community: 65080:109
      Originator: 10.229.37.92, Cluster list: 0.0.254.56, 10.229.34.128
```

iRR1 IP

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stale Routes

Example – Stale Entry on Receiving Router

```
Central-RR# show bgp ipv4 unicast 10.229.37.92
BGP routing table entry for 10.229.37.92/32, version 290518
BGP Bestpath: deterministic-med
Paths: (3 available, best #2, table default)
  Refresh Epoch 1
  Local, (Received from a RR-client)
    10.229.34.128 (metric 116) from 10.229.34.128 (10.229.34.128)
      Origin IGP, metric 0, localpref 100, valid, internal, best2
      Community: 65080:109
      Originator: 10.229.37.92, Cluster list: 10.229.34.128
      mpls labels in/out nolabel/26596
      rx pathid: 0x1A, tx pathid: 0x1
  Local, (Received from a RR-client)
    10.229.37.2 (metric 113) from 10.229.37.2 (10.229.37.2)
      Origin IGP, metric 0, localpref 100, valid, internal, best
      Community: 65080:109
      Originator: 10.229.37.92, Cluster list: 10.229.37.2
      mpls labels in/out nolabel/27183
      rx pathid: 0x7, tx pathid: 0x0
```

iRR1 IP

iRR2 IP

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stale Routes

How to Troubleshoot?

- On IOS, it is difficult to get to the root cause after the problem has occurred
  - Enable conditional debugs and wait for the issue to happen again
  - Reproduce the problem in lab environment (hard but not impossible)
- On IOS XR, use **`show bgp trace`** and BGP debugs to understand if the advertisement has been sent/received
  - Debug
- On NX-OS, use **`show bgp internal event-history { events | errors }`** to figure out if the prefix has been received / advertised

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Stale Routes or Missing Routes / Advertisements

Conditional Debugs

```
IOS-1# show access-list 99
Standard IP access list 99
     permit 10.1.1.0 0.0.0.255

IOS-1# debug ip bgp 2.2.2.2 update 99
```

```
IOS-XR:
route-policy DEBUG_BGP
 if destination in BGP_PREFIX then
  pass
 else
  drop
 endif
end-policy
 prefix-set BGP_PREFIX
  100.1.1.0/24
 end-set
debug bgp update ipv4 unicast [in | out] route-policy DEBUG_BGP
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

BGP Route Churn and Troubleshooting with BGP Table Version

cisco Live!

Route Churn

Symptom – High CPU?

```
Router# show process cpu
  CPU utilization for five seconds: 100%/0%; one minute: 99%; five minutes: 81%
....
139     6795740   1020252       6660 88.34% 91.63% 74.01%   0 BGP Router
```

- Define “High”
  - Know what normal CPU utilization is for the router in question
  - Is the CPU spiking due to “BGP Scanner” or is it constant?
- Look at the scenario
  - Is BGP going through “Initial Convergence”?
- If not then route churn is the usual culprit
  - Illegal recursive lookup or some other factor causes bestpath changes for the entire table

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

High CPU due to BGP Router

- How to identify route churn?
  - Do **`sh ip bgp summary`** and note the table version
  - Wait 60 seconds
  - Do **`sh ip bgp summary`** and compare the table version from 60 seconds ago
- You have 150K routes and see the table version increase by 300
  - This is probably normal route churn
  - Know how many bestpath changes you normally see per minute
- You have 150K routes and see the table version fluctuating by 20K – 50K
  - This is bad and is the likely cause of your high CPU

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

```
Router# show ip bgp all sum | in table
BGP table version is 936574954, main routing table version 936574954
BGP table version is 429591477, main routing table version 429591477
Router#


Router# show ip bgp all sum | in table
BGP table version is 936576768, main routing table version 936575068
BGP table version is 429591526, main routing table version 429591526
Router#

Router# show ip route | in 00:00:0
B       187.164.0.0 [200/0] via 218.185.80.140, 00:00:00
B       187.52.0.0 [200/0] via 218.185.80.140, 00:00:00
B       187.24.0.0 [200/0] via 218.185.80.140, 00:00:00
B       187.68.0.0 [200/0] via 218.185.80.140, 00:00:00
B       186.136.0.0 [200/0] via 218.185.80.140, 00:00:00
. . . .
```

Over 1800 prefixes flapped

< 4 seconds later

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

Table Version Changes?

- What causes massive table version changes?
- Flapping peers
  - Hold-timer expiring?
  - Corrupt UPDATE?
- Route churn
  - Do not try to troubleshoot the entire BGP table at once
  - Identify one prefix that is churning and troubleshoot that one prefix
  - Will likely fix the problem with the rest of the BGP table churn

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

Flapping Routes in BGP

- Figuring out flapping routes from routing table is easy (even in VRF)
  - **`show ip route vrf * | in 00:00:0|VRF`**
- How about identifying flapping routes on the VPNv4 Route Reflector?
  - **`show bgp vpnv4 unicast all summary | in table`**
  - Use the table version as the marker in the below command to see the routes which flapped after the last command that was executed
  - **`show bgp vpnv4 unicast all version [`** *`version-num`* **`| recent`** *`version-num`* **`]`**
  - Use the next-hop of the prefixes from the above command, to see why the prefixes are flapping

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

Flapping Routes in BGP

```
R1# show bgp ipv4 unicast version recent 6
BGP table version is 12, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
     Network          Next Hop            Metric LocPrf Weight Path
r>i 192.168.2.2/32   192.168.2.2              0    100      0 i
r>i 192.168.3.3/32   192.168.3.3              0    100      0 i
*mi 192.168.200.200/32
                     192.168.3.3              0    100      0 200 i
*>i                  192.168.2.2              0    100      0 200 i
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

Flapping Routes in BGP on IOS XR

- IOS XR has a more interesting command for table version updates
  - **show bgp** *afi safi* **version** *start-version end-version*

```
RP/0/0/CPU0:XR1# show bgp ipv4 unicast version 5 7
VRF: default
------------
Status codes: s suppressed, d damped, h history, * valid, > best
              i - internal, r RIB-failure, S stale, N Nexthop-discard
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network            Next Hop            Metric LocPrf    Version Path
*>i192.168.2.2/32     192.168.2.2              0    100          6
i*>i192.168.3.3/32     192.168.3.3             0    100          7
i*>i192.168.200.200/32 192.168.2.2             0    100          5 200 i
i                      192.168.3.3             0    100          5 200 i

Processed 3 prefixes, 4 paths
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Route Churn

Which AFI?

- If there are too many updates coming onto the router, one way to identify it would be

```
show ip traffic | section TCP
```

- Symptom – TCP traffic increasing rapidly, but table version for IPv4 and VPNv4 AFI is only increasing by 200 or 300 or a smaller value
- Check for different AFIs enabled on the router and checking for the table version changes in those AFIs
  - Especially IPv6 or VPNv6 as those can have more impact with fewer prefixes flapping

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Embedded Event Manager (EEM)

- Serves as a powerful tool for high CPU troubleshooting
- Triggered based on event and thresholds
- Multiple actions can be set based on events

```
event manager applet HIGHCPU
event snmp oid "1.3.6.1.4.1.9.9.109.1.1.1.1.3.1" get-type exact entry-op gt entry-val "90"
exit-op lt exit-val "70" poll-interval 5 maxrun 200
action 1.0 syslog msg "START of TAC-EEM: High CPU"
action 1.1 cli command "show clock"
action 1.3 cli command "show ip bgp all summary | append disk0:proc_CPU"
action 2.0 cli command "sh clock | append disk0:proc_CPU"
action 2.1 cli command "show process cpu sorted | append disk0:proc_CPU‟
action 2.2 cli command "show proc cpu history | append disk0:proc_CPU"
action 2.3 cli command " show ip bgp all summary | append disk0:proc_CPU"
action 3.1 cli command "show log | append disk0:proc_CPU"
action 4.0 syslog msg "END of TAC-EEM: High CPU"
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Dissecting NH for IPv6

- Understanding NH for IPv6
- Troubleshooting 6PE and 6VPE

Dissecting NH for IPv6

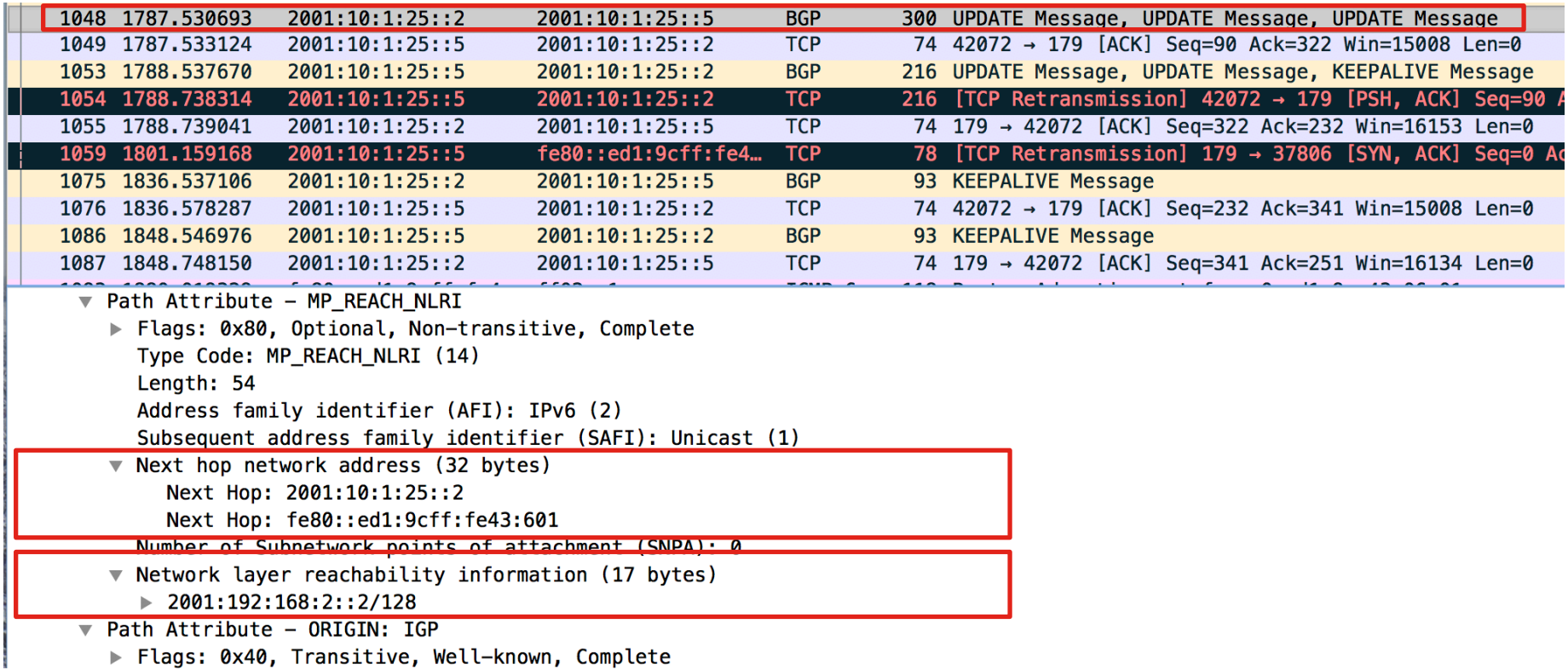
IPv6 Next-Hop

- Assume that BGP neighbors are peered using their global IPv6 addresses
- Given this, for IPv6 NLRIs, the NH will always contain a global IPv6 address, and may contain a link-local IPv6 address
- IPv6 iBGP
  - NH only contains a global IPv6 address
- IPv6 eBGP in IPv6 or VPNv6 address family
  - Between directly connected IPv6 addresses
    - NH contains both a global and a link-local address
  - Between indirect IPv6 addresses
    - NH contains only a global IPv6 address

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Dissecting NH for IPv6

IPv6 BGP Update



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Dissecting NH for IPv6

IPv4 Peering for IPv6 AFI

- When using IPv4 peering, NH is automatically set to IPv4-mapped IPv6 address ::ffff:X.X.X.X (on IOS and IOS-XE) (NX-OS won't install the prefix)
  - This cannot be used as an IPv6 NH since it is an inaccessible address
  - RR cannot advertise this prefix
    - Exception: Can be accessible if the session is for 6PE or 6VPE
- Alternate Solution
  - Have the advertising BGP speaker set an outbound route-map to set the NH to a global IPv6 address
  - Have the receiving BGP speaker set an inbound route-map to set the NH to a global IPv6 address

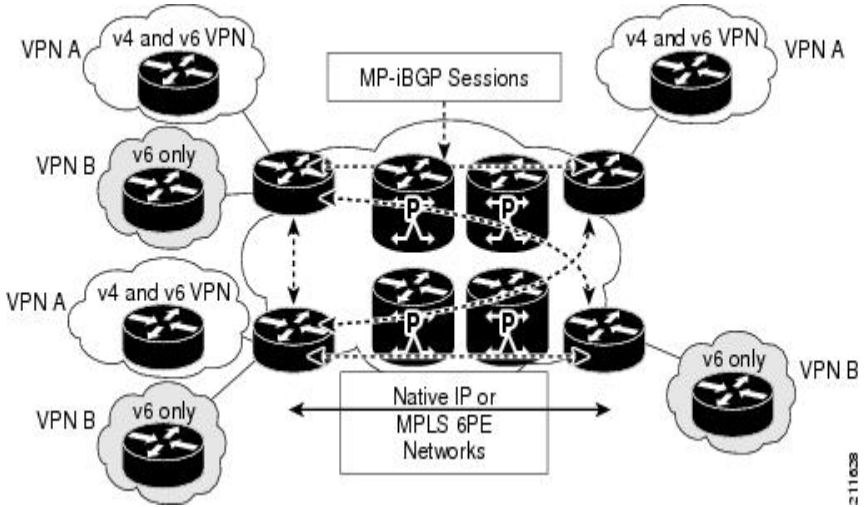
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Overview (RFC4798)

High Level 6PE/6VPE

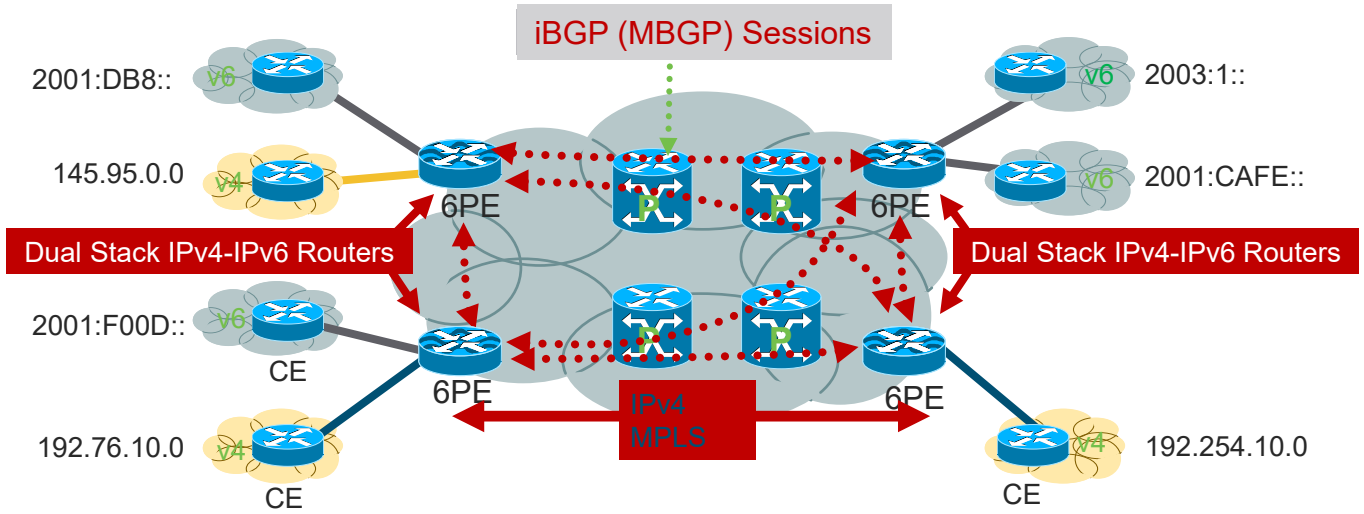
MPLS as a IPv6 tunneling mechanism

- MPLS (VPN) network built on IPv4 can be upgraded to provide tunneling mechanism (6-4-6) using Multi-Protocol BGP
  - 6PE – For global routing
  - 6VPE – For VRF based routing



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

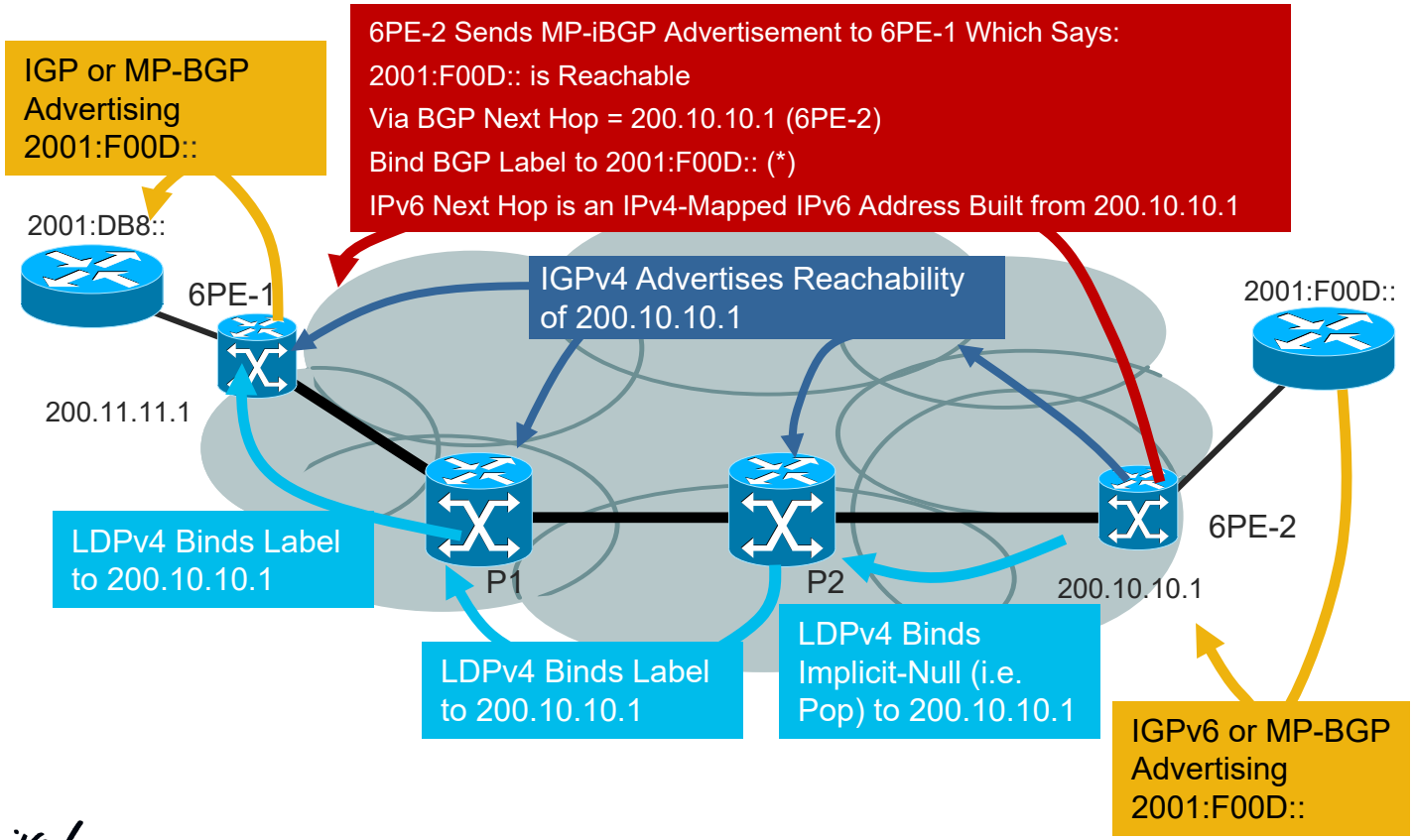
IPv6 Provider Edge Router (6PE) over MPLS



- IPv6 global connectivity over an IPv4-MPLS core
- Transitioning mechanism for providing unicast IPv6 connectivity
- PEs are updated to support dual stack/6PE
- IPv6 reachability exchanged among 6PEs via iBGP (MBGP)
- IPv6 packets transported from 6PE to 6PE inside MPLS using labels

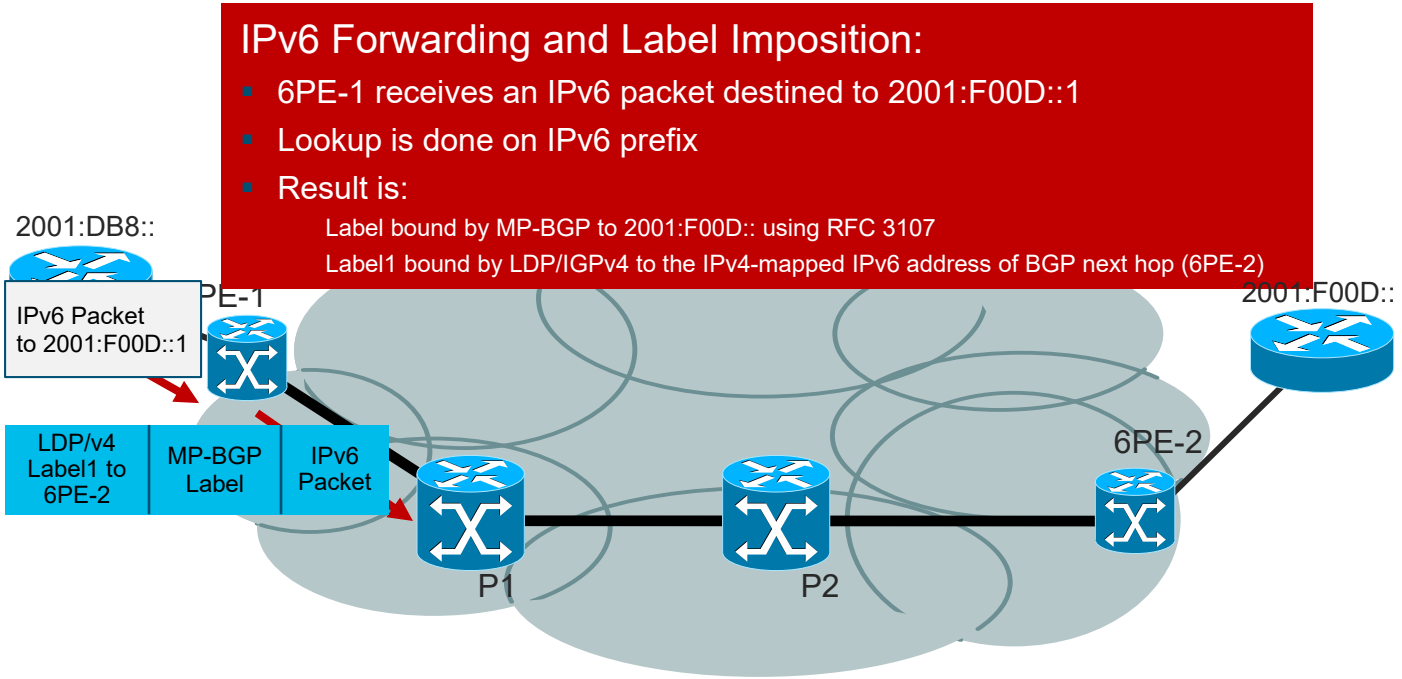
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Routing/Label Distribution



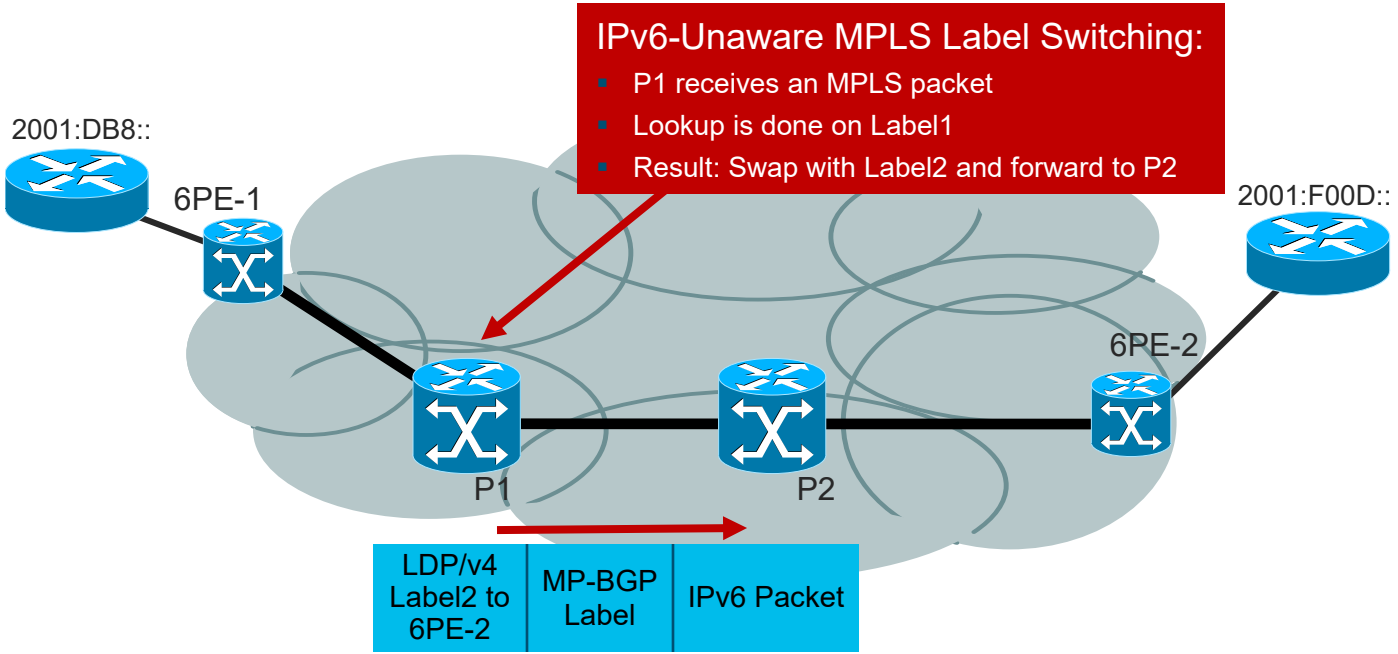
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Forwarding (6PE-1)



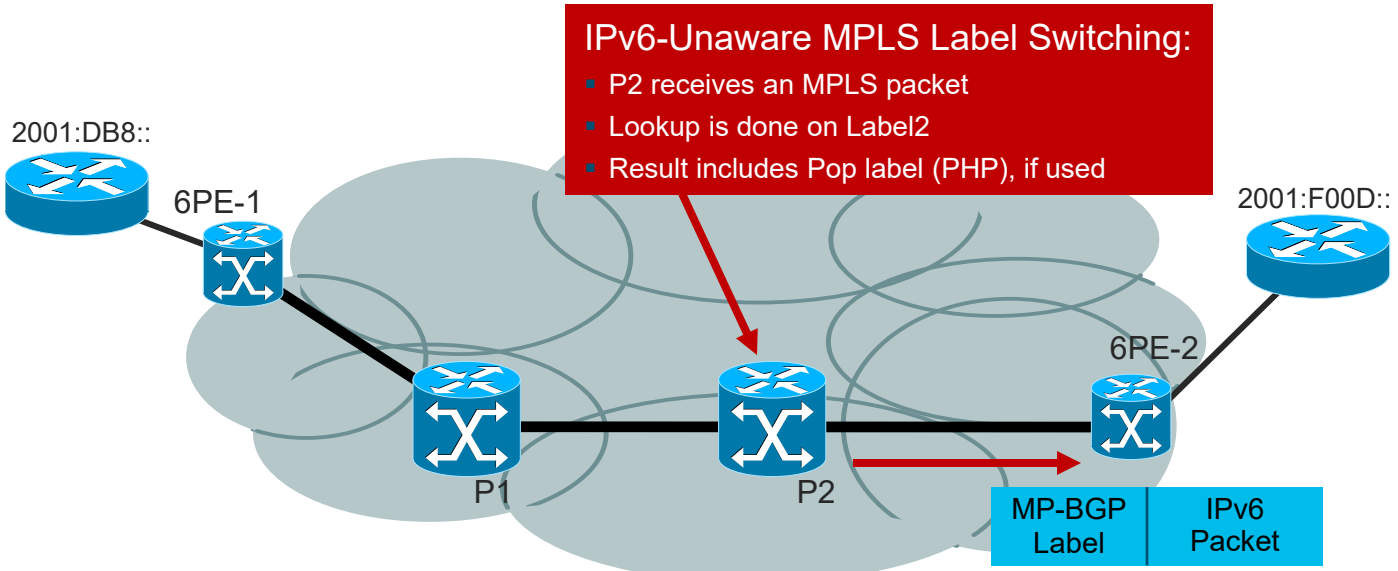
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Forwarding (P1)



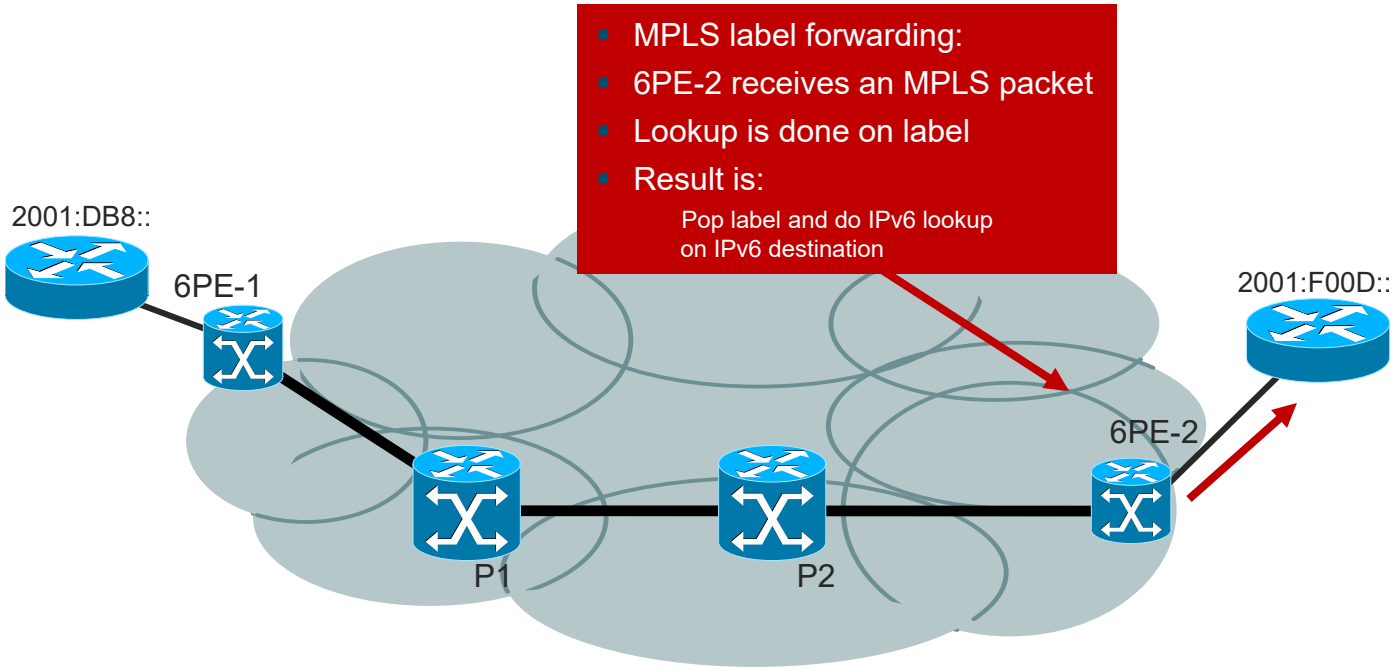
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Forwarding (P2)



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

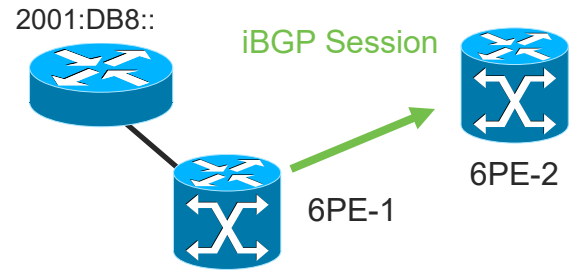
6PE Forwarding (6PE-2)



© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE-1 Configuration

```
ipv6 cef
!
mpls label protocol ldp
!
router bgp 100
 no synchronization
 no bgp default ipv4 unicast
 neighbor 2001:DB8:1::1 remote-as 65014
 neighbor 200.10.10.1 remote-as 100
 neighbor 200.10.10.1 update-source Loopback0
 !
 address-family ipv6
 neighbor 200.10.10.1 activate
 neighbor 200.10.10.1 send-label
 neighbor 2001:DB8:1::1 activate
 redistribute connected
 no synchronization
 exit-address-family
```



2001:DB8:1::1 Is the Local CE

200.10.10.1 Is the Remote 6PE

Send Labels Along with IPv6 Prefixes by Means of MP-BGP Note: Will Cause Session to Flap

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

6PE Show Output

```
6PE-1# show ip route 200.10.10.1
Routing entry for 200.10.10.1/32
 Known via "isis", distance 115, metric 20, type level-2
[snip]
  * 10.12.0.1, from 200.10.10.1, via FastEthernet1/0
    Route metric is 20, traffic share count is 1
```

```
6PE-1# show ipv6 route
B  2001:F00D::/64 [200/0]
   via ::FFFF:200.10.10.1, IPv6-mpls
```

```
6PE-1# show ipv6 cef internal
.. OUTPUT TRUNCATED ..
2001:F00D::/64,
  nexthop ::FFFF:200.10.10.1
 fast tag rewrite with F0/1, 10.12.0.1, tags imposed {17 28}
```

Other Useful Output:

```
show bgp ipv6 neighbors
show bgp ipv6 unicast
show mpls forwarding
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Troubleshooting with NX-OS

Verifying BGP Configuration Parameters

- Sometimes we require verifying some configuration parameters for BGP
- To verify the config / process wide parameters, use the command **`show bgp process`**
- Includes the following:
  - BGP Router ID
  - Confederation ID or Cluster ID
  - Process and Memory state
  - Counts of configured peers and established peers
  - AFI information
    - Redistribution (if any)
    - Route-map

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Show bgp process

```
N7K1# show bgp process
BGP Process Information
BGP Process ID : 5128
BGP Protocol Started, reason: : configuration
BGP Protocol Tag : 1
BGP Protocol State : Running
BGP Memory State : OK
BGP asformat : asplain

BGP attributes information
Number of attribute entries : 15
HWM of attribute entries : 49
Bytes used by entries : 1380
Entries pending delete : 0
HWM of entries pending delete : 0
BGP paths per attribute HWM : 11
BGP AS path entries : 0
Bytes used by AS path entries : 0

Information regarding configured VRFs:
BGP Information for VRF default VRF Id : 1
VRF state : UP
Router-ID : 192.168.1.1
Configured Router-ID : 192.168.1.1
Confed-ID : 0 Cluster-ID : 0.0.0.0
No. of configured peers : 10
No. of pending config peers : 0
No. of established peers : 0
VRF RD : Not configured
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Show bgp process

```
N7K1# show bgp process
[ ... continued ... ]

Information for address family IPv4 Unicast in VRF default
Table Id                   : 1
Table state       : UP
Peers Active-peers Routes Paths Networks Aggregates
5          0        19     20     10        0

Redistribution
   static, route-map static-bgp
   direct, route-map rm-permit-all
   eigrp, route-map rm-permit-all


Default-Information originate enabled

 Nexthop trigger-delay
       critical 3000 ms
       non-critical 10000 ms
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

BGP Event-History

- NX-OS event-history capability is alternate, and preferred, to running debugs
- Event-History Buffer Sizes:
  - Large
  - Medium
  - Small
- Event-History maintained for:
  - Events
  - Errors
  - Detail
  - Msgs
  - CLI

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Processing an Incoming Update – **`show bgp event-history detail`**

- Manually enable detail event-history using the command **`event-history detail size [large | medium | small]`** in BGP process configuration

```
05:28:12.515623: (default) UPD: Received UPDATE message from 10.1.23.2
05:28:12.515616: (default) BRIB: [IPv4 Unicast] (192.168.1.1/32 (10.1.23.2)): returning from
bgp_brib_add, new_path: 0, change: 0, undelete: 0, history: 0, force: 0, (pflags=0x28), reeval=0
05:28:12.515608: (default) BRIB: [IPv4 Unicast] 192.168.1.1/32 from 10.1.23.2 was already in BRIB
with same attributes
05:28:12.515600: (default) BRIB: [IPv4 Unicast] (192.168.1.1/32 (10.1.23.2)): bgp_brib_add:
handling nexthop
05:28:12.515593: (default) BRIB: [IPv4 Unicast] Path to 192.168.1.1/32 via 192.168.2.2 already
exists, dflags=0x8001a
05:28:12.515580: (default) BRIB: [IPv4 Unicast] Installing prefix 192.168.1.1/32 (10.1.23.2) via
10.1.23.2  into BRIB with extcomm
05:28:12.515557: (default) UPD: [IPv4 Unicast] Received prefix 192.168.1.1/32 from peer
10.1.23.2, origin 0, next hop 10.1.23.2, localpref 0, med
005:28:12.515524: (default) UPD: 10.1.23.2 Received attr code 2, length 10, AS-Path: <200 100 >
05:28:12.515503: (default) UPD: Attr code 3, length 4, Next-hop: 10.1.23.2
05:28:12.515454: (default) UPD: Attr code 1, length 1, Origin: IGP
05:28:12.515446: (default) UPD: 10.1.23.2 parsed UPDATE message from peer, len 52 , withdraw len
0, attr len 24, nlri len 5
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Update Generation – **show bgp event-history detail**

```
05:28:11.478903: (default) UPD: [IPv4 Unicast] 10.1.23.2 Created UPD msg (len 52) with prefix
192.168.1.1/32 ( Installed in HW) path-id 1 for peer
05:28:11.478886: (default) UPD: 10.1.23.2 Sending attr code 3, length 4, Next-hop: 10.1.23.3
05:28:11.478880: (default) UPD: 10.1.23.2 Sending attr code 2, length 10, AS-Path: <300 100 >
05:28:11.478870: (default) UPD: 10.1.23.2 Sending attr code 1, length 1, Origin: IGP
05:28:11.478856: (default) UPD: [IPv4 Unicast] consider sending 192.168.1.1/32 to peer 10.1.23.2,
path-id 1, best-ext is off
...
05:28:11.478717: (default) EVT: [IPv4 Unicast] soft refresh out completed for 1 peers
05:28:11.478690: (default) EVT: [IPv4 Unicast] Adding peer 10.1.23.2 for update gen
05:28:11.478686: (default) BRIB: [IPv4 Unicast] Group setting SRM for dest 192.168.3.3/32
05:28:11.478682: (default) BRIB: [IPv4 Unicast] Group setting SRM for dest 192.168.2.2/32
05:28:11.478678: (default) BRIB: [IPv4 Unicast] Group setting SRM for dest 192.168.1.1/32
05:28:11.478666: (default) EVT: [IPv4 Unicast] 1 peer(s) being soft refreshed out
05:28:11.478661: (default) EVT: [IPv4 Unicast] 10.1.23.2 [peer index 2]
05:28:11.478638: (default) EVT: [IPv4 Unicast] Doing soft out BGP table walk for peers
05:28:10.478332: (default) EVT: [IPv4 Unicast] Scheduling peer 10.1.23.2 for soft refresh out
05:28:10.478321: (default) EVT: Received ROUTEREFRESH message from 10.1.23.2
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Conditional Debugging and URIB

```
debug logfile bgp
debug bgp events updates rib brib import
debug-filter bgp vrf vpn1
debug-filter bgp address-family ipv4 unicast
debug-filter bgp neighbor 10.1.202.2
debug-filter bgp prefix 192.168.2.2/32
```

- Troubleshooting URIB

```
show routing internal event-history { add | delete | modify | summ }
show routing internal event-history recursive
show routing internal event-history ufdm
show routing internal event-history ufdm-summary
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Route Policy Manager

- Route-map functionality is provided by a standalone process in NX-OS called Route Policy Manager (RPM)
- RPM handles route-maps, AS-path ACLs, community lists, and prefix lists
- The route-maps are configured the same way they are configured in Cisco IOS, but are managed by RPM
  - If there are any issues seen with route-maps not functioning, RPM process status and traces should be inspected

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Route Policy Manager

```
NX-1# show system internal sysmgr service name rpm
Service "rpm" ("rpm", 203):
        UUID = 0x131, PID = 5265, SAP = 348
        State: SRV_STATE_HANDSHAKED (entered at time Mon Jan
30 03:07:59 2017).
        Restart count: 1
        Time of last restart: Mon Aug 22 03:07:57 2016.
        The service never crashed since the last reboot.
        Tag = N/A
        Plugin ID: 1
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Route Policy Manager

```
template peer-policy PP-Test1
    send-community
    route-map RM-Test1 out
!
neighbor 192.168.2.2 remote-as 65000
      inherit peer-session ps-ebgp-peer-
to-mpls-core
        address-family ipv4 unicast
        inherit peer-policy PP-Test1 5
        send-community
        prefix-list pl-nab-core-devl-
routes in
        no prefix-list pl-cloud-routes out
        route-map RM-Test2 out
        soft-reconfiguration inbound
. . . .
```

```
NX-1# sh route-map RM-Test1
route-map RM-Test1, permit, sequence 10
  Match clauses:
    ip address prefix-lists: sy3-routes
  Continue: sequence 20
  Set clauses:
    community 65135:999
route-map RM-Test1, permit, sequence 999
  Match clauses:
  Set clauses:
!
NX-1# sh route-map RM-Test2
route-map RM-Test1, permit, sequence 10
  Match clauses:
    ip address prefix-lists: pl-cloud-
routes
  Set clauses:
route-map RM-Test1, permit, sequence 20
  Match clauses:
    as-path (as-path filter): as-me1-o365-
ext-routes
  Set clauses:
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

```
NX-2# show system internal rpm route-map
Policy name: RM-Test1                Type: route-map
Version: 6                           State: Ready
Ref. count: 1                        PBR refcount: 0
Stmt count: 5                        Last stmt seq: 999
Set nhop cmd count: 0                Set vrf cmd count: 0
Set intf cmd count: 0                Flags: 0x00000003
PPF nodeid: 0x00000000               Config refcount: 0
PBR Stats: No
Clients:
    bgp-65136 (Route filtering/redistribution)    ACN version: 0
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

```
# show system internal rpm event-history rsw

Routing software interaction logs of RPM
1) Event:E_DEBUG, length:88, at 96760 usecs after Sun Apr 23 22:19:12 2017
    [120] [3959]: Bind ack sent - client bgp-65136 uuid 0x0000011b for policy RM-
Test2  <<<<<  Outbound route-map bound to BGP client
2) Event:E_DEBUG, length:83, at 96717 usecs after Sun Apr 23 22:19:12 2017
    [120] [3959]: Bind request - client bgp-65136 uuid 0x0000011b policy RM-Test2
3) Event:E_DEBUG, length:88, at 782159 usecs after Sun Apr 23 21:51:06 2017
    [120] [3959]: Bind ack sent - client bgp-65136 uuid 0x0000011b for policy RM-
Test2
<snip>
    [120] [3959]: UnBind request succesfull - client bgp-65136 policy RM-
Test1 <<<<< Unbind for route-map referenced in peer-policy
6) Event:E_DEBUG, length:99, at 781950 usecs after Sun Apr 23 21:51:06 2017
    [120] [3959]: UnBind request - client bgp-65136 uuid 0x0000011b policy RM-Test1
7) Event:E_DEBUG, length:102, at 344591 usecs after Sun Apr 23 21:47:39 2017
    [120] [3959]: Bind ack sent - client bgp-65136 uuid 0x0000011b for policy RM-
Test1 <<<<< Route-map referenced in peer-policy
8) Event:E_DEBUG, length:97, at 344557 usecs after Sun Apr 23 21:47:39 2017
    [120] [3959]: Bind request - client bgp-65136 uuid 0x0000011b policy RM-Test1
```


© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Troubleshooting with NX-OS

Route Policy Manager

- Use RPM Event-history when troubleshooting any misbehavior of route policy / redistribution / missing routes / routes not learnt
- In case of issues, collect **`show tech rpm`**
- Use the below commands to troubleshoot RPM issues
  - **`show system internal rpm event-history events`** (For RPM Events)
  - **`show system internal rpm event-history errors`** (For errors with RPM)
  - **`show system internal rpm event-history rsw`** (RPM Interaction with RPM software)
  - **`show system internal rpm event-history msgs`** (RPM Message logs)
  - **`show system internal rpm event-history trace`** (RPM Traces)

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Complete your online session survey

- Please complete your session survey after each session. Your feedback is very important.
- Complete a minimum of 4 session surveys and the Overall Conference survey (starting on Thursday) to receive your Cisco Live t-shirt.
- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Content Catalog on ciscolive.com/emea.

Cisco Live sessions will be available for viewing on demand after the event at ciscolive.com.

cisco Live!

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Continue your education

Demos in the Cisco Showcase

Walk-In Labs

Meet the Engineer 1:1 meetings

Related sessions

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public

Thank you

You make **possible**