Let's go

The bridge to possible

Border Gateway Protocol(BGP) Fundamentals

Boo Chinnaswamy – Technical Consulting Engineer

BRKENT-1179

cisco Live!

#CiscoLiveAPJC

"Reconciliation"- Dustin Koa Art

Cisco Webex App

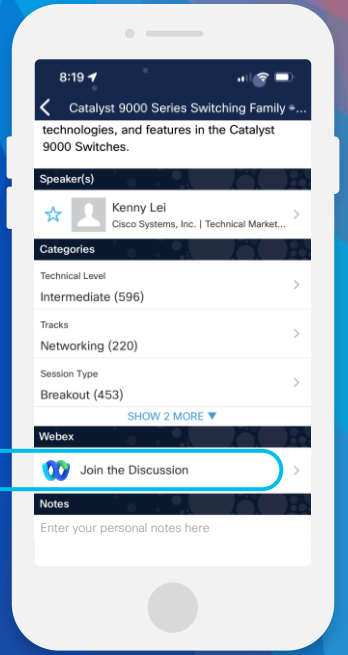
Questions?

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

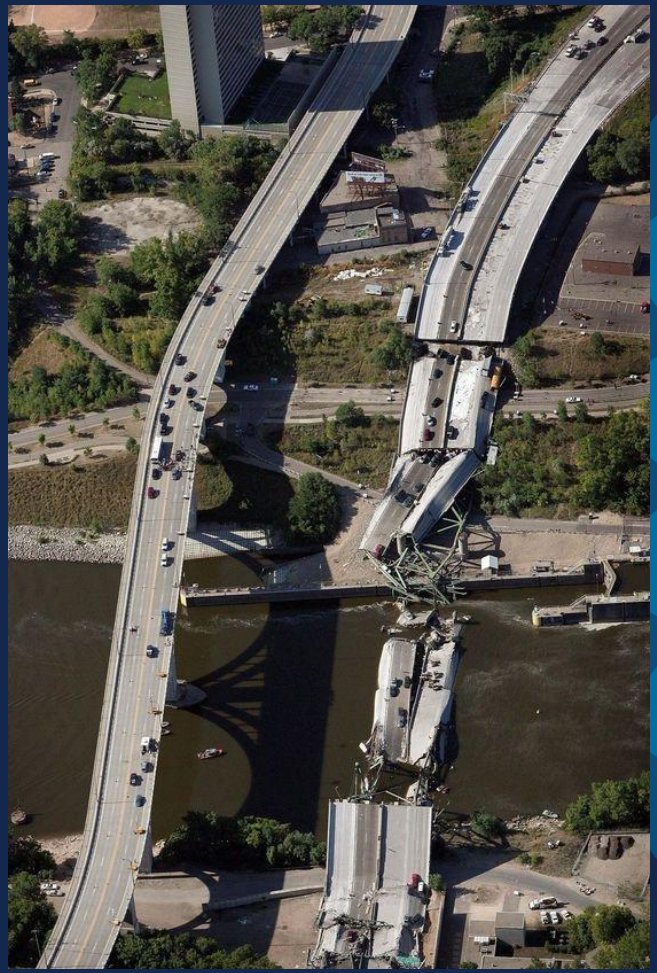
How

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

Webex spaces will be moderated by the speaker until December 22, 2023.



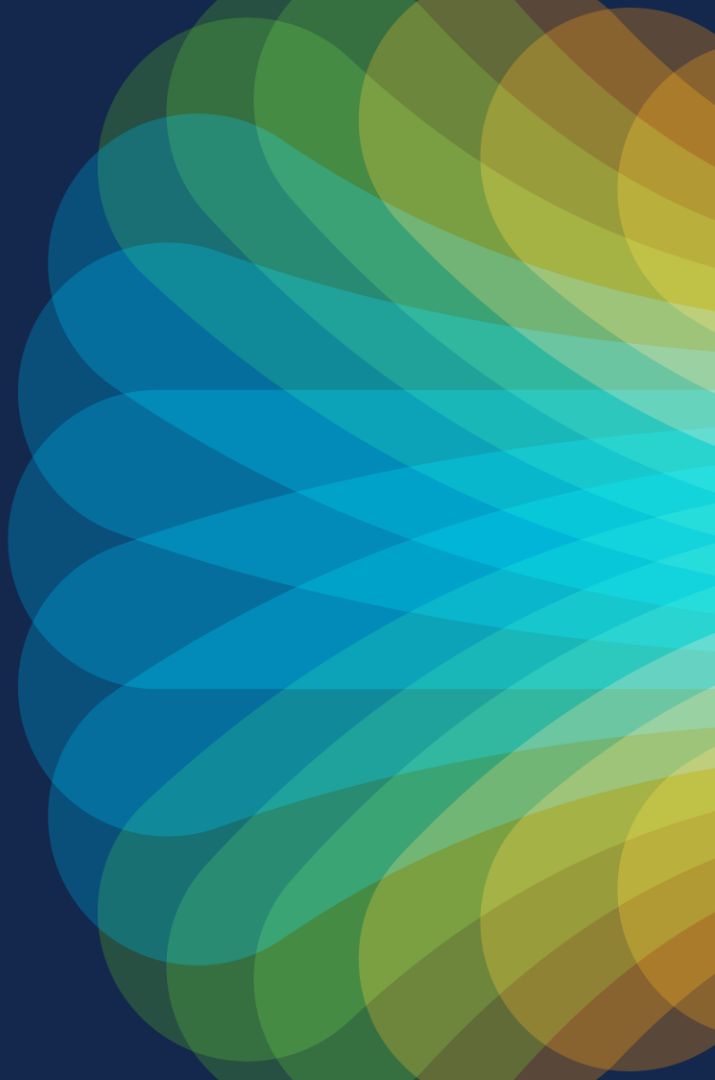
https://ciscolive.ciscoevents.com/ciscolivebot/#BRKENT-1179



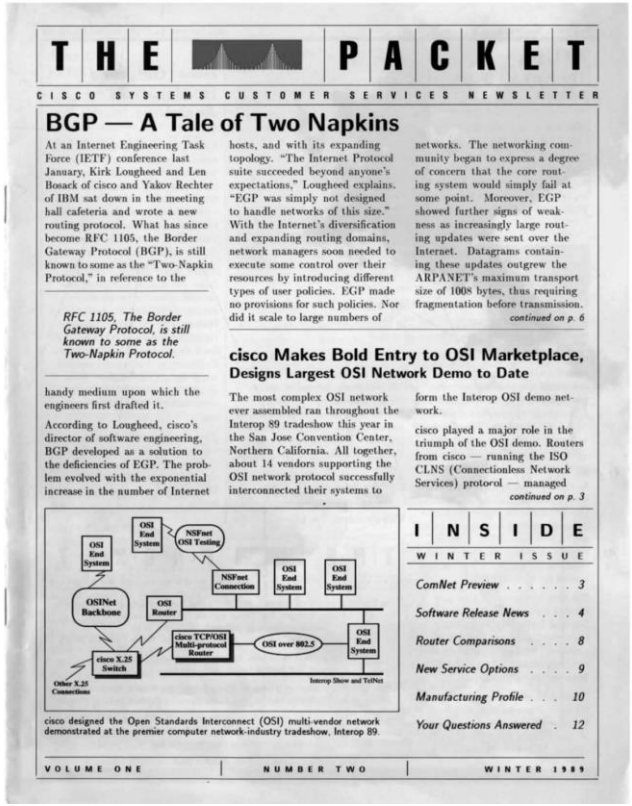
Agenda

- Introduction to BGP
- What and why BGP?
- Messages and states
- Internal vs External BGP
- Attributes
- Best Path Selection Algorithm
- Troubleshooting BGP

What is BGP and Why BGP?



BGP – A Tale of Two Napkins



THE PACKET

CISCO SYSTEMS CUSTOMER SERVICES NEWSLETTER

BGP — A Tale of Two Napkins

At an Internet Engineering Task Force (IETF) conference last January, Kirk Lougheed and Len Bosack of cisco and Yakov Rechter of IBM sat down in the meeting hall cafeteria and wrote a new routing protocol. What has since become RFC 1105, the Border Gateway Protocol (BGP), is still known to some as the "Two-Napkin Protocol," in reference to the handy medium upon which the engineers first drafted it.

RFC 1105, The Border Gateway Protocol, is still known to some as the Two-Napkin Protocol.

According to Lougheed, cisco's director of software engineering, BGP developed as a solution to the deficiencies of EGP. The problem evolved with the exponential increase in the number of Internet hosts, and with its expanding topology. "The Internet Protocol suite succeeded beyond anyone's expectations," Lougheed explains. "EGP was simply not designed to handle networks of this size." With the Internet's diversification and expanding routing domains, network managers soon needed to execute some control over their resources by introducing different types of user policies. EGP made no provisions for such policies. Nor did it scale to large numbers of networks. The networking community began to express a degree of concern that the core routing system would simply fail at some point. Moreover, EGP showed further signs of weakness as increasingly large routing updates were sent over the Internet. Datagrams containing these updates outgrew the ARPANET's maximum transport size of 1008 bytes, thus requiring fragmentation before transmission.

continued on p. 6

cisco Makes Bold Entry to OSI Marketplace, Designs Largest OSI Network Demo to Date

The most complex OSI network ever assembled ran throughout the Interop 89 tradeshow this year in the San Jose Convention Center, Northern California. All together, about 14 vendors supporting the OSI network protocol successfully interconnected their systems to form the Interop OSI demo network.

cisco played a major role in the triumph of the OSI demo. Routers from cisco — running the ISO CLNS (Connectionless Network Services) protocol — managed

continued on p. 3

cisco designed the Open Standards Interconnect (OSI) multi-vendor network demonstrated at the premier computer network-industry tradeshow, Interop 89.

INSIDE

WINTER ISSUE

VOLUME ONE | NUMBER TWO | WINTER 1989

BGP Autonomous System Number (ASN)

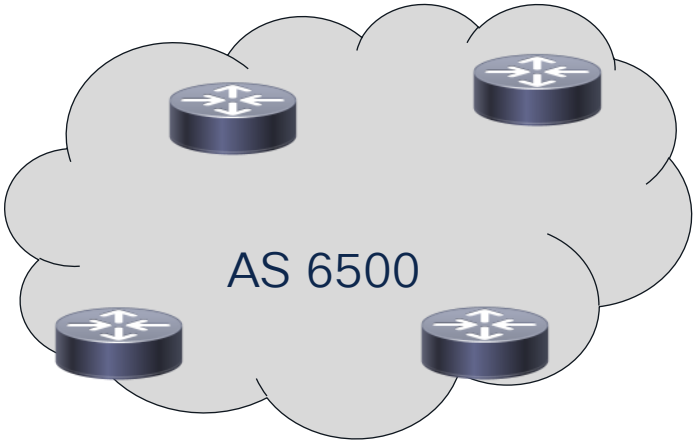
- RFC1105

- Created to address 2-byte ASN depletion
  - 16-bit number
  - 0 to 65535
  - Private range 64512 to 65534

- Interoperable with 2-byte ASNs (range includes 2-byte ASNs)
  - 32-bit number
  - 0 to 4294967295
  - Additional private range 4200000000 to 4294967294

Autonomous System (AS)

A group of one or more IP prefixes (lists of IP addresses accessible on a network) run by one or more network operators that maintain a single, clearly-defined routing policy.



Border Gateway Protocol

- Border Gateway Protocol
  - Reliable and scalable routing protocol designed to operate between autonomous systems
  - Operates on Transmission Control Protocol, listens on port 179
  - No concept of metrics
  - Route selection is based on attributes
  - Assumes routes in AS fully taken care of by an IGP (EIGRP, OSPF, IS-IS)

BGP Stability Considerations

- Events in networks often occur in bursts

- There is always a challenge how to react
  - Reacting fast improves convergence time but may introduce churn
  - Reacting with a delay improves stability but delays convergence

- BGP favors stability
  - It delays sending updates to smoothen out the churn and to collect possibly multiple changes for a single update
  - It only advertises changes (incremental updates)

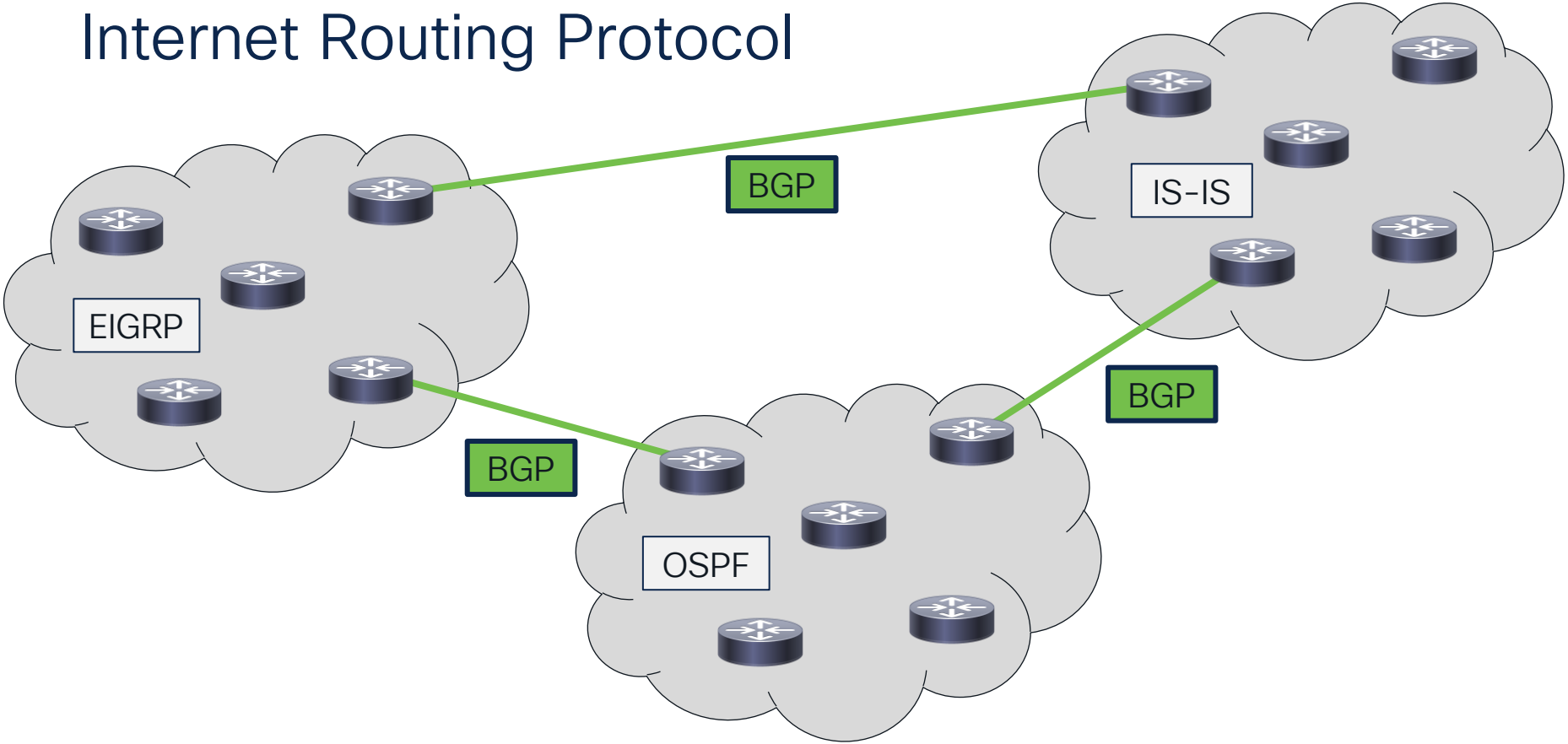
Route Scale & Control Plane Stability

```
R1 # show bgp ipv4 unicast neighbors 10.1.1.2
BGP neighbor is 10.1.1.2,  remote AS 65501, external link
  BGP version 4, remote router ID 10.1.1.2
  BGP state = Established, up for 00:48:19
  Last read 00:00:14, last write 00:00:03, hold time is 180, keepalive interval is 60
seconds

    <Output omitted>

 Default minimum time between advertisement runs is 30 seconds

    <Output omitted>
```

Route Scale & Control Plane Stability

```
route-views> show bgp ipv4 unicast summary | ex never|Active|Idl
BGP router identifier 128.223.51.103, local AS number 6447
BGP table version is 2813468887, main routing table version 2813468887
Path RPKI states: 7708169 valid, 10262796 not found, 14390 invalid
962240 network entries using 238635520 bytes of memory

<Output omitted...>

BGP using 3360540172 total bytes of memory
BGP activity 71481761/70340753 prefixes, 3122764440/3095983216 paths, scan interval 60
secs

Neighbor   V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
4.68.4.46  4         3356 5325851   83510 2813468729   12    0 3w5d       905347
12.0.1.63  4         7018 5117677   19250 2813468729  124    0 1w5d       906235

<Output omitted...>
```

Internet Routing Protocol



BGP Address Families

```
R1(config)# router bgp 6500
R1(config-router)# address-family ?
  ipv4   Address family
  ipv6   Address family
  l2vpn  Address family
  nsap   Address family
  vpnv4  Address family
  vpnv6  Address family
```

```
route-views> show bgp all neighbors 4.68.4.46 | i family
For address family: IPv4 Unicast
    Address family IPv4 Unicast: advertised and received
    Address family IPv4 Multicast: advertised and received
For address family: IPv6 Unicast
For address family: IPv4 Multicast
    Address family IPv4 Unicast: advertised and received
    Address family IPv4 Multicast: advertised and received
For address family: L2VPN E-VPN
For address family: MVPNv4 Unicast
```

Prefix/Length

Multiprotocol BGP

- **IPv4** and **IPv6:** Includes unicast and multicast routes
  - Traditional IPv4 and IPv6 routing

- **VPNv4** and **VPNv6:** Layer-3 VPN information.
  - Supports both the IP versions on Multiprotocol Label Switching) MPLS VPN labels.
  - BGP provides the label switching capability on MPLS.

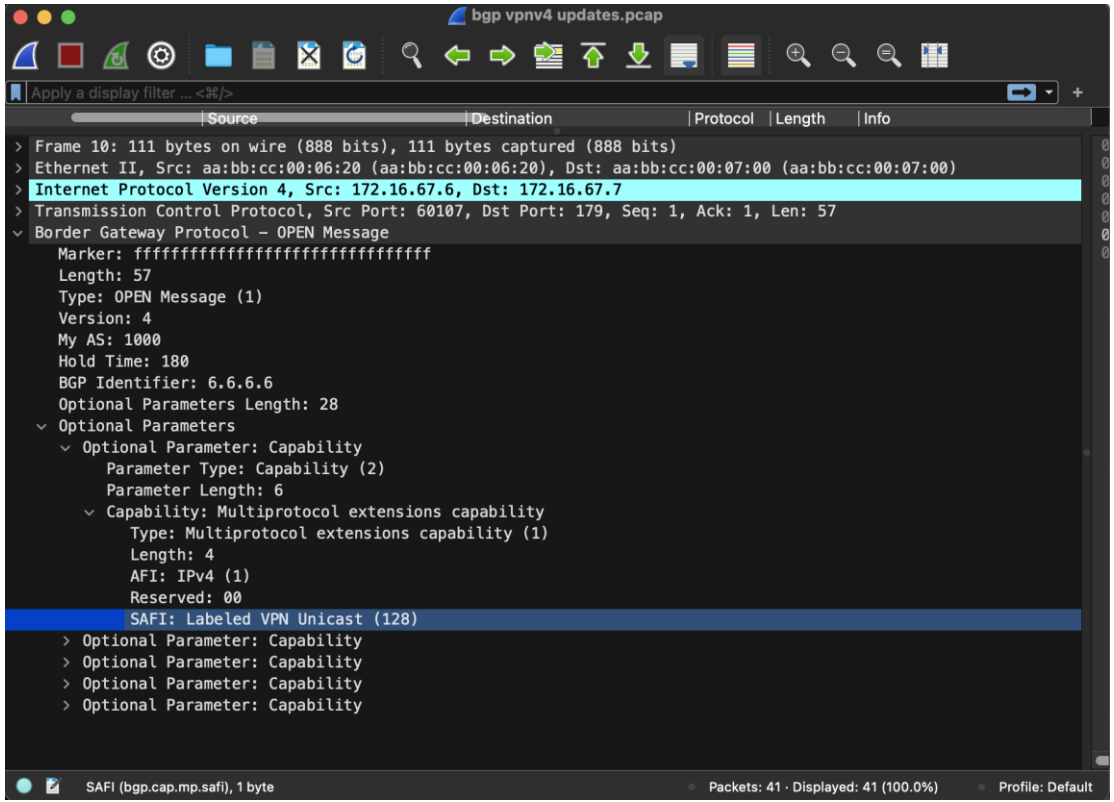
- **L2VPN:** Layer-2 VPN e.g.,VPLS or EVPN-VXLAN
  - Virtual Private LAN Service – Connects multiple sites in a single bridged domain. LAN network can be shared over the internet
  - Virtual Extensible LAN-Ethernet Virtual Private Network – Capability to extend Layer 2 over Layer 3

Types of Address families

- Multiprotocol extension capability are exchanged during neighbour capability exchange process
- AFI and SAFI indicate what address family and subsequent address family routes are transported

Name	Address Family Identifier (AFI)	Subsequent AFI
IPv4 Unicast	1	1
VPNv4	1	128
L2VPN	25	65
EVPN	25	70

VPNv4 AFI and SAFI



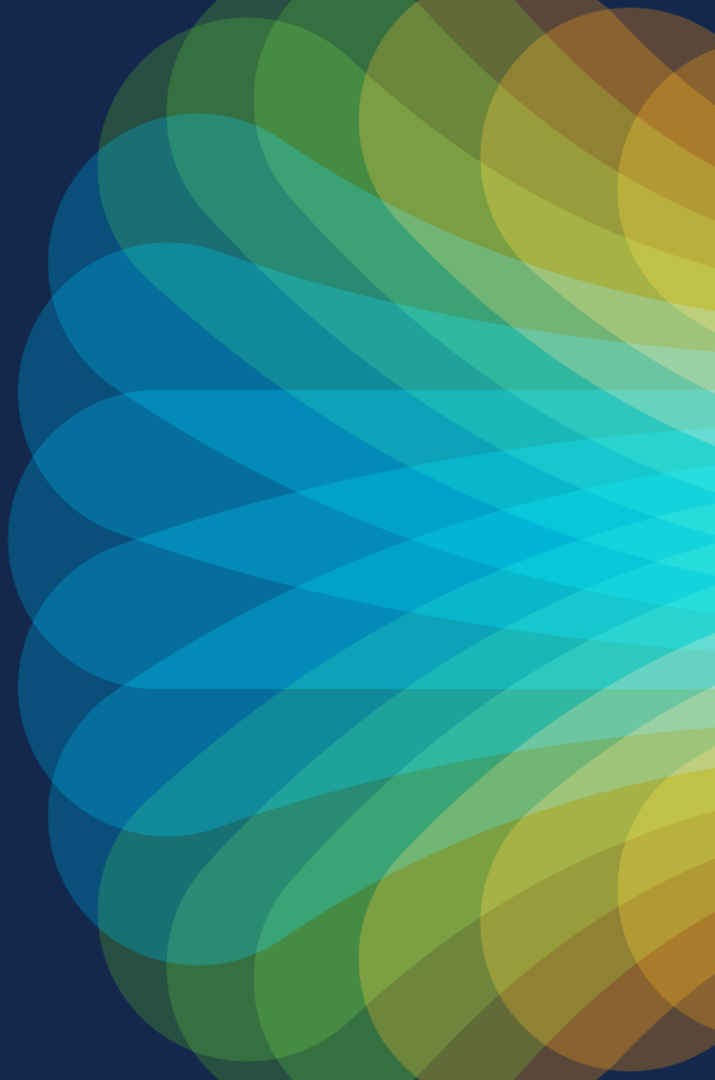
IPv4 Address family

```
R1#show ip bgp neighbors 10.1.1.2
BGP neighbor is 10.1.1.2,   remote AS 6501, external link BGP
  version 4, remote router ID 10.1.1.2
  BGP state = Established, up for 00:02:07
  Last read 00:00:06, last write 00:00:13, hold time is 180, keepalive
   interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received(new)
    Address family IPv4 Unicast: advertised and received Message
  statistics:
    InQ depth is 0
    OutQ depth is 0
```

MVPNv4 Address Family

```
R1#show bgp vpnv4 unicast all summary
BGP router identifier 10.1.1.1, local AS number 6500
BGP table version is 1, main routing table version 1

Neighbor        V    AS    MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.1.1.2        4    6501     5       6        1      0    0  00:01:03        0

R1#show bgp vpnv4 unicast all 10.1.1.2/32
BGP routing table entry for 1:1:10.1.1.2/32, version 14 Paths: (1 available, best #1,
table TAC)
  Advertised to update-groups:

<Output omitted>

      Origin incomplete, metric 2, localpref 100, weight 32768, valid,

<Output omitted>
```

Messages and States



BGP Message Types

- BGP is layer 3 operates on TCP on layer 4
  - Byte stream-oriented
  - Unicast only
  - Connection-oriented and reliable
  - Providing flow and congestion control

- The format of BGP messages partly accommodates TCP specifics
  - Message markers
  - Length indications (lots of!)

BGP Message Types

BGPv4 uses (only) 5 message types

- OPEN
- UPDATE
- NOTIFICATION
- KEEPALIVE
- ROUTE-REFRESH (not part of initial BGPv4 RFC 1654 specification, brought in through RFC 2918 and nearly universally supported)

Message: OPEN

- BGP speakers use OPEN to advertise their configuration and capabilities once their TCP is established

- Session comes up
  - Version advertisement
  - Autonomous System Number advertisement
  - Hold Time advertisement/negotiation
  - BGP Router ID advertisement
  - Optional Capabilities advertisement/negotiation

- If the peer advertises an incompatible configuration, the peering is terminated, and the TCP session closed

Message: OPEN

```
Border Gateway Protocol - OPEN Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 57
 Type: OPEN Message (1)
 Version: 4
 My AS: 64512
 Hold Time: 180
 BGP Identifier: 10.255.255.1
 Optional Parameters Length: 28
▾Optional Parameters
 ▸Optional Parameter: Capability
 ▸Optional Parameter: Capability
 ▸Optional Parameter: Capability
 ▸Optional Parameter: Capability
 ▾Optional Parameter: Capability
   Parameter Type: Capability (2)
   Parameter Length: 6
  ▸Capability: Support for 4-octet AS number capability
```

Message: OPEN – Capability Codes

Capability Codes

Value	Name	RFC
1	Multiprotocol Extensions for BGP-4	2858
2	Route Refresh Capability for BGP-4	2918
3	Outbound Route Filtering Capability	5291
5	Extended Next Hop Encoding	8950
6	BGP Extended Message	8654
7	BGPsec Capability	8205
8	Multiple Labels Capability	8277
9	BGP Role	9234
64	Graceful Restart Capability	4724
65	Support for 4-octet AS number capability	6793
69	ADD-PATH Capability	7911
70	Enhanced Route Refresh Capability	7313

Message: UPDATE

- The UPDATE message is the workhorse of BGP
  - Advertises reachable NLRI's along with their attributes
  - Withdraws unreachable NLRI's

- The format of the UPDATE message targets maximum efficiency
  - The path attributes are included only once, followed by the list of all NLRIs that share them
  - Every NLRI contains only the network prefix (and padding bits to a whole octet if needed)

Message: UPDATE – NEW ROUTES

```
Border Gateway Protocol - UPDATE Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 67
 Type: UPDATE Message (2)
 Withdrawn Routes Length: 0
 Total Path Attribute Length: 28
▾Path attributes
 ▸Path Attribute - ORIGIN: IGP
 ▸Path Attribute - AS_PATH: empty
 ▸Path Attribute - NEXT_HOP: 10.255.255.1
 ▸Path Attribute - MULTI_EXIT_DISC: 1234
 ▸Path Attribute - LOCAL_PREF: 100
▾Network Layer Reachability Information (NLRI)
 ▸192.168.0.0/24
 ▸192.168.1.0/24
 ▸192.168.2.0/24
 ▸192.168.3.0/24
```

Message: UPDATE – WITHDRAWN ROUTES

```
Border Gateway Protocol - UPDATE Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 27
 Type: UPDATE Message (2)
 Withdrawn Routes Length: 4
▾Withdrawn Routes
 ▸192.168.3.0/24
 Total Path Attribute Length: 0
```

Message: UPDATE

Update Error Subcodes

Value	Name	RFC
0	Unspecific	EID 4493
1	Malformed Attribute List	4271
2	Unrecognized Well-known Attribute	4271
3	Missing Well-known Attribute	4271
4	Attribute Flags Error	4271
5	Attribute Length Error	4271
6	Invalid ORIGIN Attribute	4271
8	Invalid NEXT_HOP Attribute	4271
9	Optional Attribute Error	4271
10	Invalid Network Field	4271
11	Malformed AS_PATH Attribute	4271

Message: NOTIFICATION

- The NOTIFICATION message is sent out by a peer who detected an unrecoverable condition and needs to terminate the peering
- After sending out a NOTIFICATION, the sender closes the session
- The NOTIFICATION contents are useful for diagnostics

```
Border Gateway Protocol - NOTIFICATION Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 21
 Type: NOTIFICATION Message (3)
 Major error Code: Cease (6)
 Minor error Code (Cease): Administratively Shutdown (2)
```

Message: NOTIFICATION – Capability Codes

[Error codes]

Value	Name	RFC
1	Message Header Error	4271
2	OPEN Message Error	4271
3	UPDATE Message Error	4271
4	Hold Timer Expired	4271
5	Finite State Machine Error	4271
6	Cease	4271
7	ROUTE-REFRESH Message Error	7313

Message: NOTIFICATION – Capability Codes

Error codes

Value	Name	RFC
1	Message Header Error	4271
2	OPEN Message Error	4271
3	UPDATE Message Error	4271
4	Hold Timer Expired	4271
5	Finite State Machine Error	4271
6	Cease	4271
7	ROUTE-REFRESH Message Error	7313

Select BGP Error Subcodes

https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-3

Message Header Error Subcodes		
Value	Name	RFC
0	Unspecific	EID 4493
1	Connection Not Synchronized	4271
2	Bad Message Length	4271
3	Bad Message Type	4271

OPEN Message Error Subcodes		
Value	Name	RFC
0	Unspecific	EID 4493
1	Unsupported Version Number	4271
2	Bad Peer AS	4271
3	Bad BGP Identifier	4271
4	Unsupported Optional Parameter	4271
6	Unacceptable Hold Time	4271
7	Unsupported Capability	4271
11	Role Mismatch	9234

Select BGP Error Subcodes

https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-3

UPDATE Message Error Subcodes		
Value	Name	RFC
0	Unspecific	EID 4493
1	Malformed Attribute List	4271
2	Unrecognized Well-known Attribute	4271
3	Missing Well-known Attribute	4271
4	Attribute Flags Error	4271
5	Attribute Length Error	4271
6	Invalid ORIGIN Attribute	4271
8	Invalid NEXT_HOP Attribute	4271
9	Optional Attribute Error	4271
10	Invalid Network Field	4271
11	Malformed AS_PATH Attribute	4271

Finite State Machine Error Subcodes		
Value	Name	RFC
0	Unspecified Error	6608
1	Receive Unexpected Message in OpenSent State	6608
2	Receive Unexpected Message in OpenCofirm State	6608
3	Receive Unexpected Message in Established State	6608

Select BGP Error Subcodes

https://www.iana.org/assignments/bgp-parameters/bgp-parameters.xhtml#bgp-parameters-3

Cease Subcodes		
Value	Name	RFC
1	Max Number of Prefixes Reached	4486
2	Administrative Shutdown	4486
3	Peer De-configured	4486
4	Administrative Reset	4486
5	Connection Rejected	4486
6	Other Configuration Change	4486
7	Connection Collision Resolution	4486
8	Out of Resources	4486
9	Hard Reset	8538
10	BFD Down	9384

ROUTE-REFRESH Message Error Subcodes		
Value	Name	RFC
0	Reserved	7313
1	Invalid Message Length	7313

BGP Message : KEEPALIVE

- Instead of relying on TCP keepalives, BGP uses it's KEEPALIVE message to periodically announce a speaker's liveliness

- KEEPALIVE is sent
  - Immediately after receiving an agreeable OPEN message from peer
  - Periodically, with the period being one third of Hold Time by default

```
Border Gateway Protocol - KEEPALIVE Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 19
 Type: KEEPALIVE Message (4)
```

Message: ROUTE-REFRESH

- Not a part of original two napkin invention
  - This is necessary when the inbound route policy changes
  - Vendors worked around this deficiency by storing aside a separate unfiltered copy of all routes from the peer (“Soft Reconfiguration”)

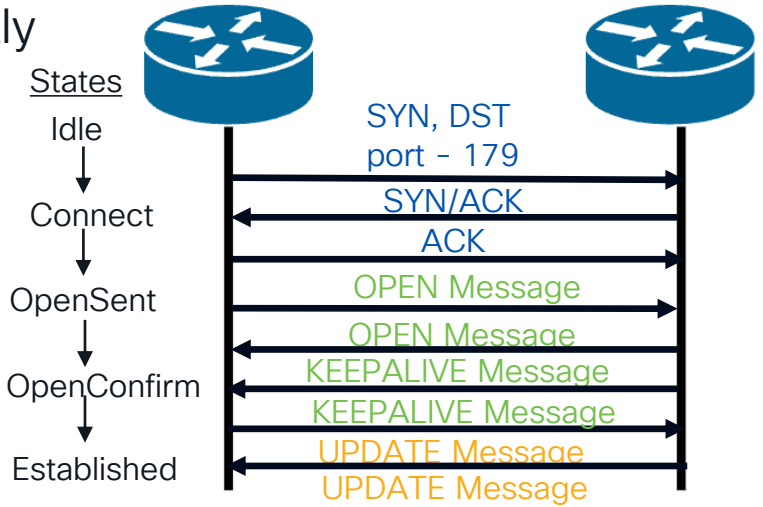
- RFC 2918 brought the ROUTE-REFRESH message allowing to ask a peer to resend all routes of any address family

```
Border Gateway Protocol - ROUTE-REFRESH Message
 Marker: ffffffffffffffffffffffffffffffff
 Length: 23
 Type: ROUTE-REFRESH Message (5)
 Address family identifier (AFI): IPv4 (1)
 Subtype: Normal route refresh request [RFC2918] with/without ORF [RFC5291] (0)
 Subsequent address family identifier (SAFI): Unicast (1)
```

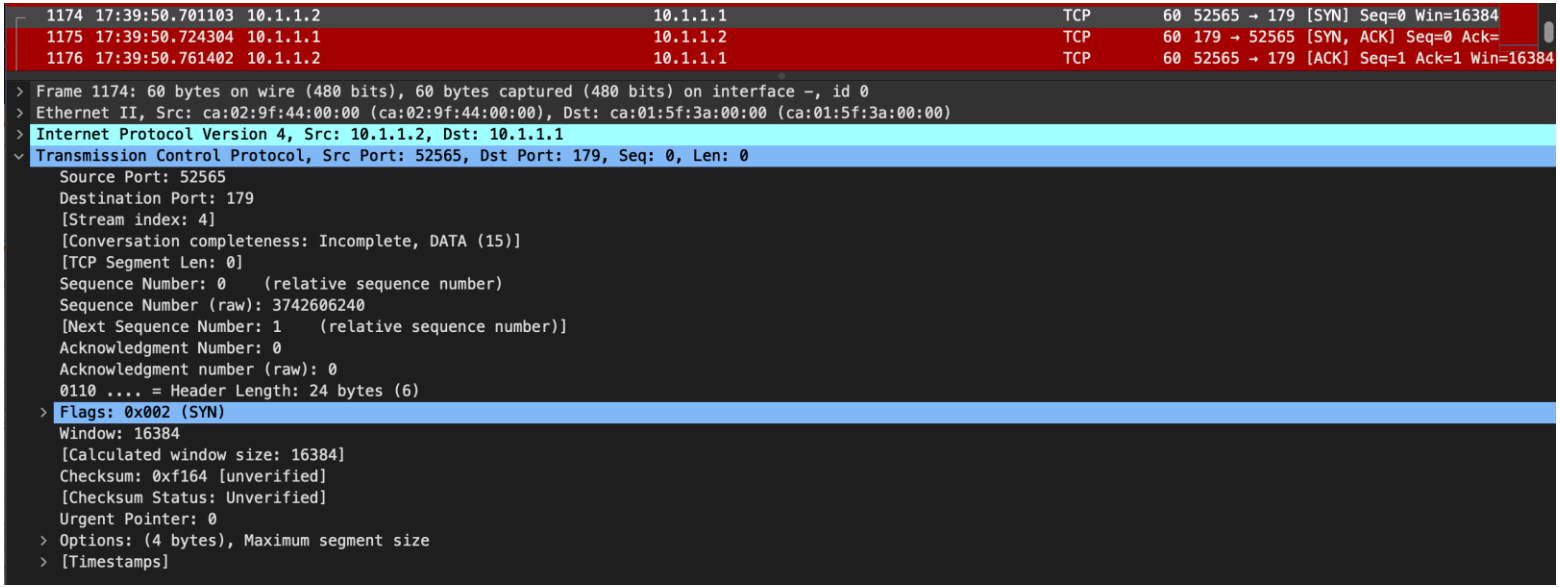
BGP Neighbours

1. Establish neighbourship – Explicitly configured
2. Exchange BGP path attributes
3. Path computation – Best path calculation
4. Inject into routing table

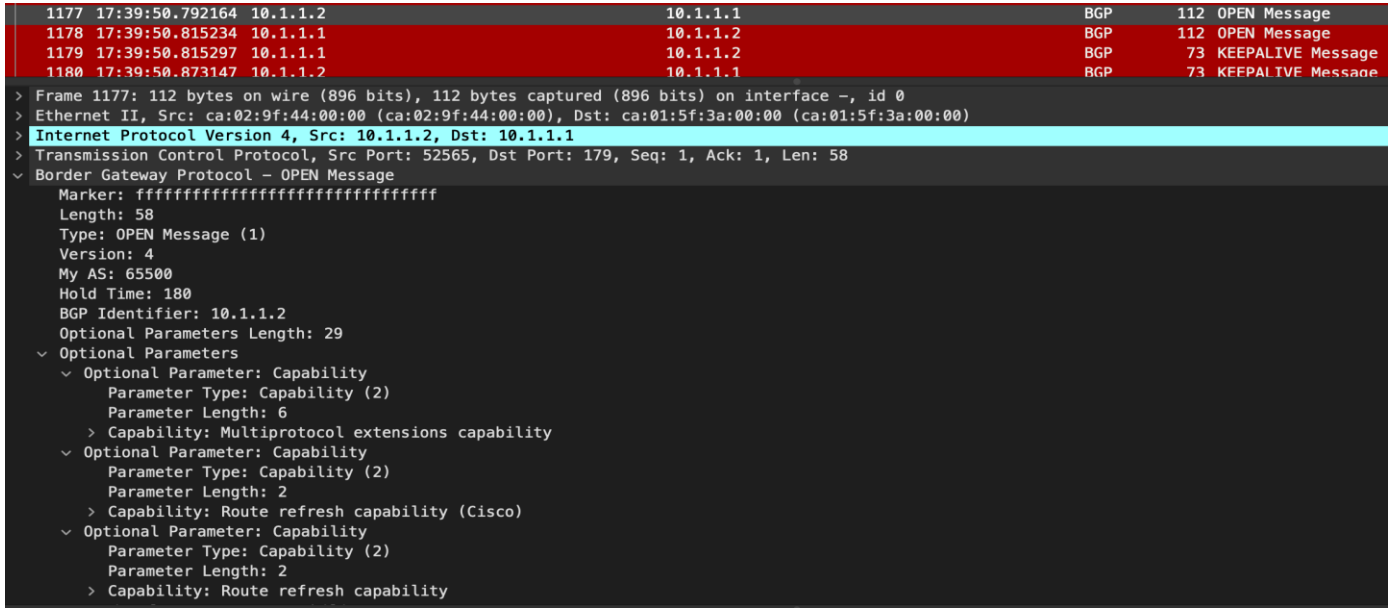
TCP 3-way handshake
BGP Peering Session
Routing Database Synchronization



State: ACTIVE, IDLE & CONNECT



OPEN SENT & OPEN CONFIRMED

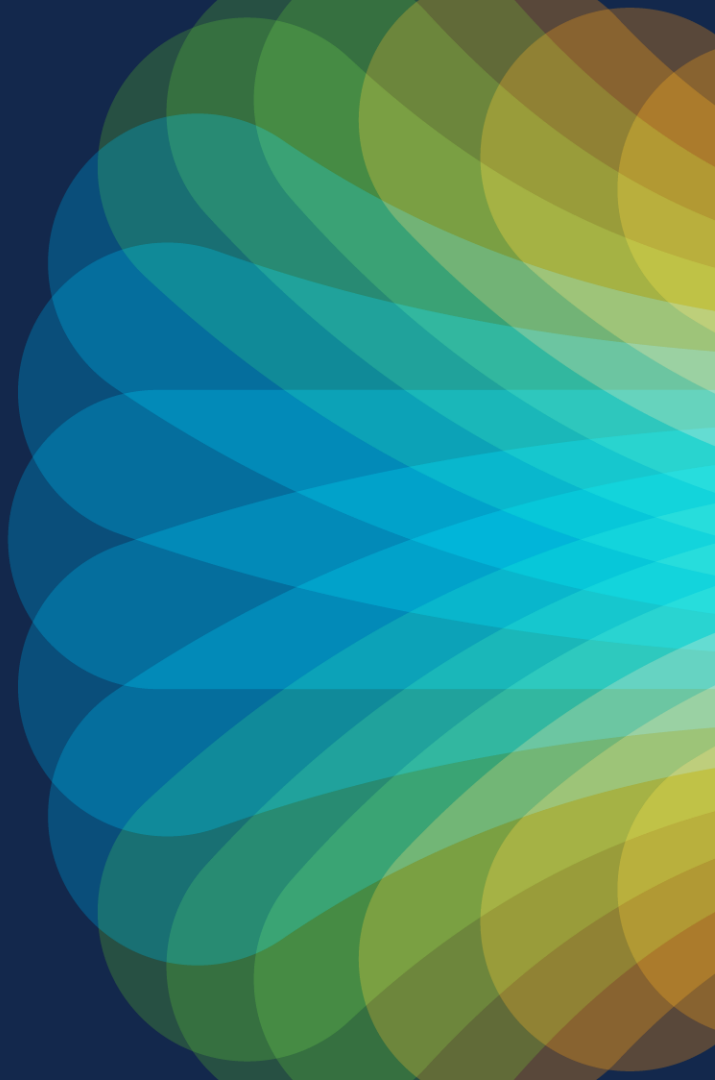


State: ESTABLISHED

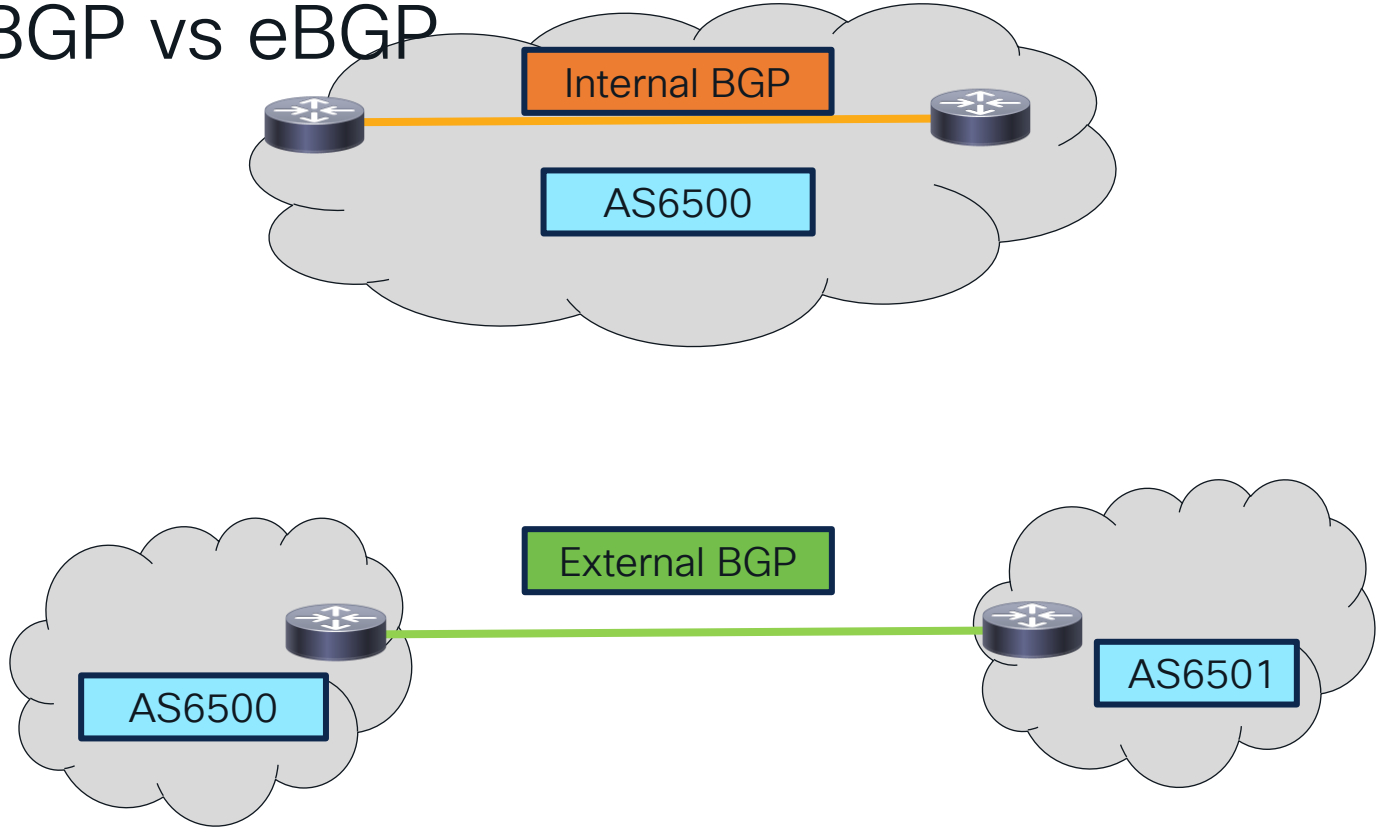
```
R1# show bgp ipv4 unicast summary
BGP router identifier 10.10.1.1, local AS number 6503
BGP table version is 1, main routing table version 1

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.10.1.2       4         6500      15      15        1    0    0 00:10:49        0
10.10.1.3       4         6501       0       0        1    0    0 00:06:29    Idle (Admin)
10.10.1.4       4         6502       0       0        1    0    0 00:00:02    Active
```

```
R1# show bgp ipv4 unicast summary
BGP neighbor is 10.10.1.2,  remote AS 6500
  BGP version 4, remote router ID 10.10.1.2
  BGP state = Established, up for 03:12:40  Last read 03:12:40, last write 05:15:45,
hold time is 180, keepalive interval is 60 seconds

<Output omitted>
```

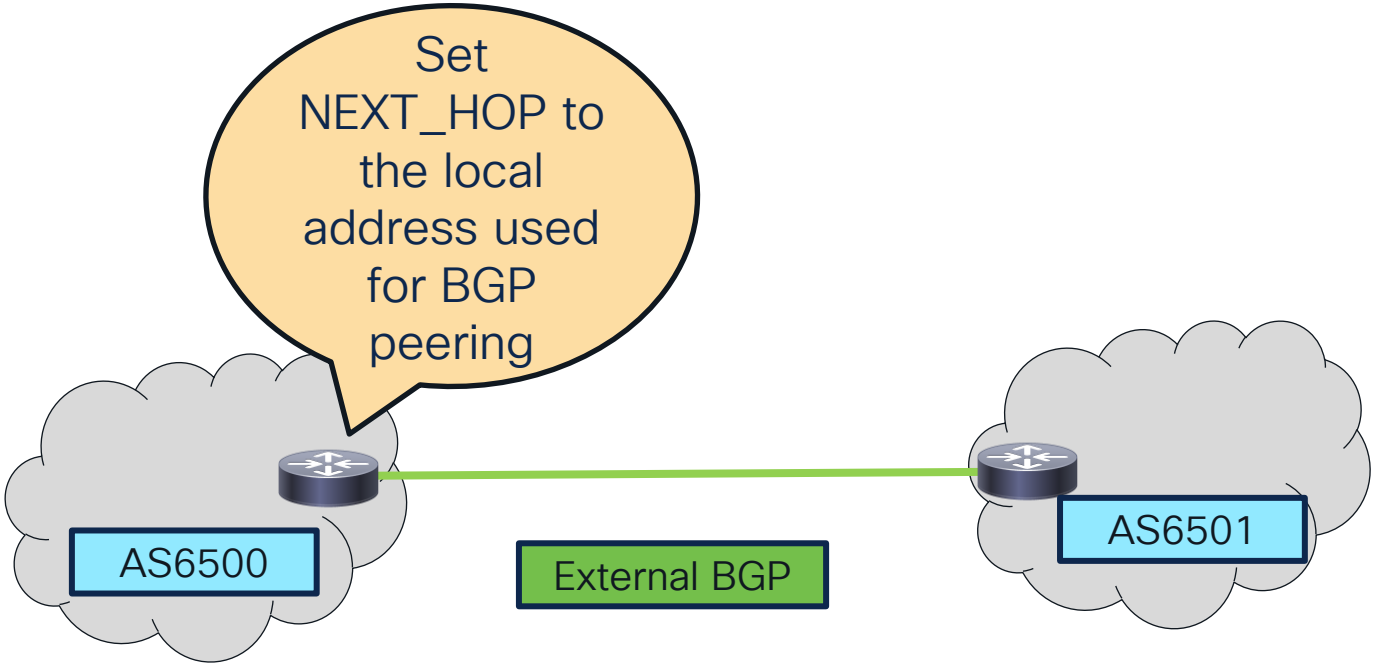
Internal vs External BGP



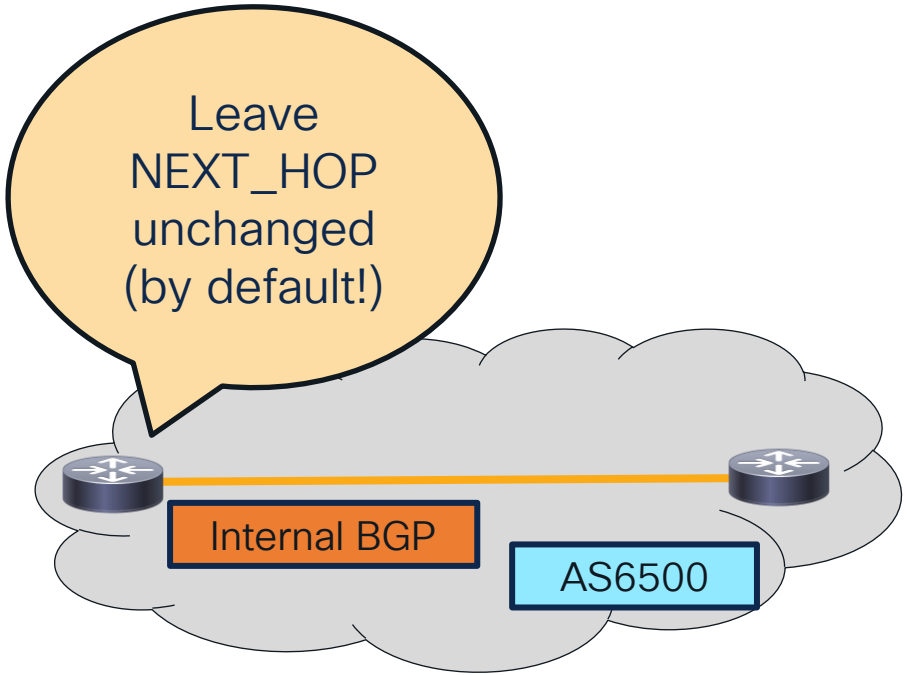
iBGP vs eBGP



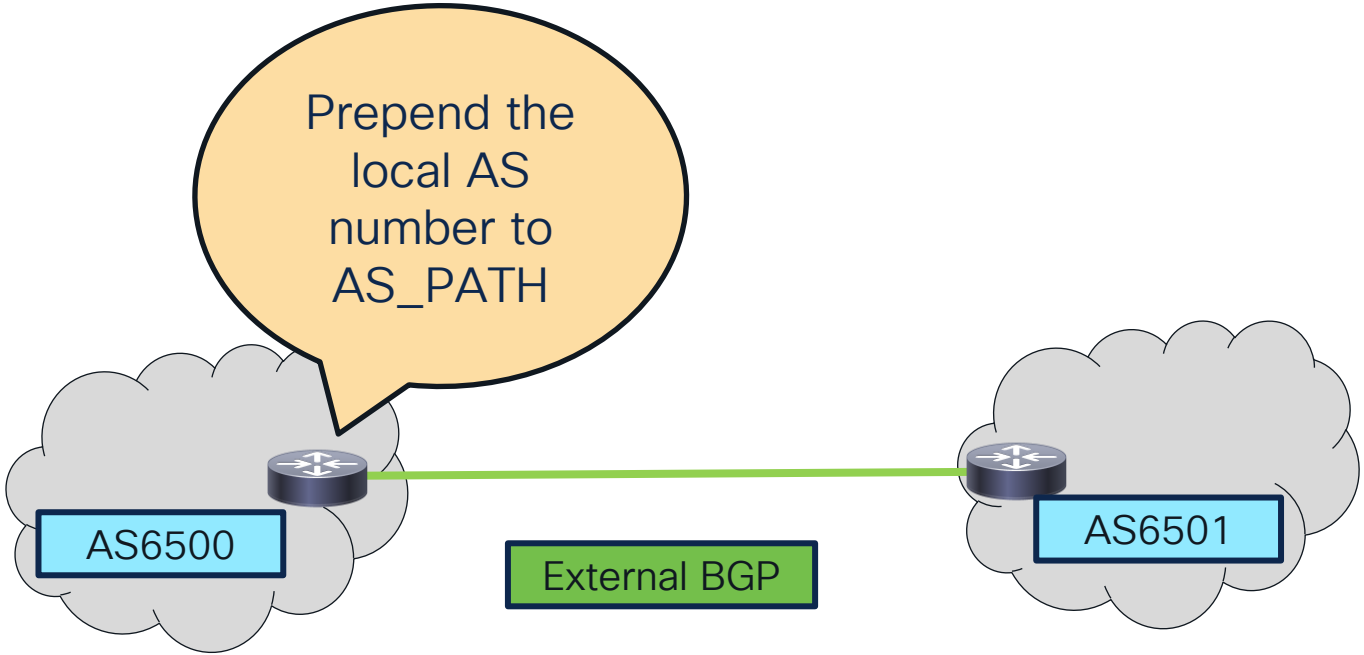
NEXT_HOP in eBGP



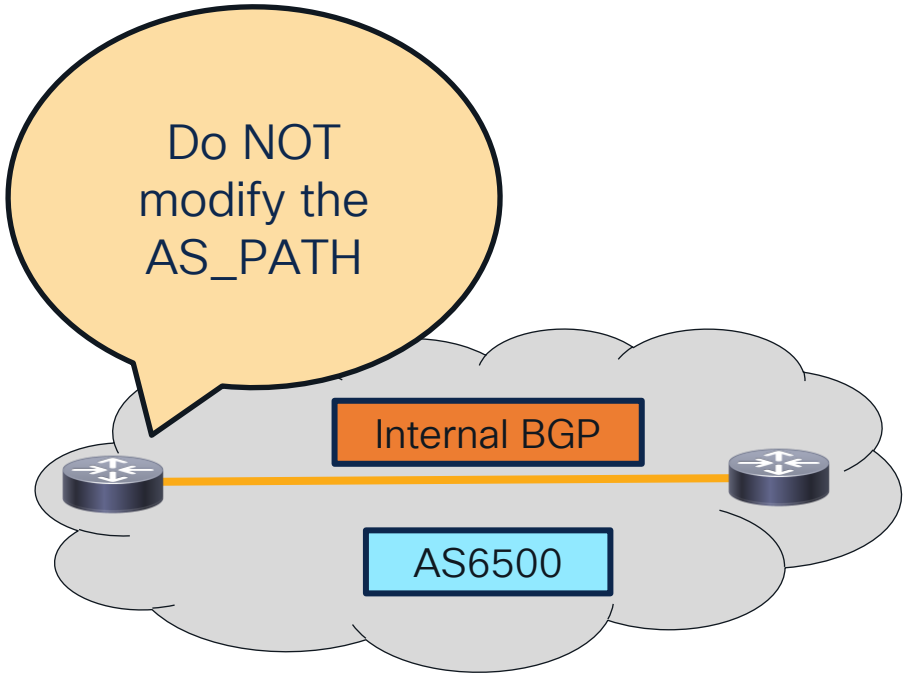
NEXT_HOP in iBGP



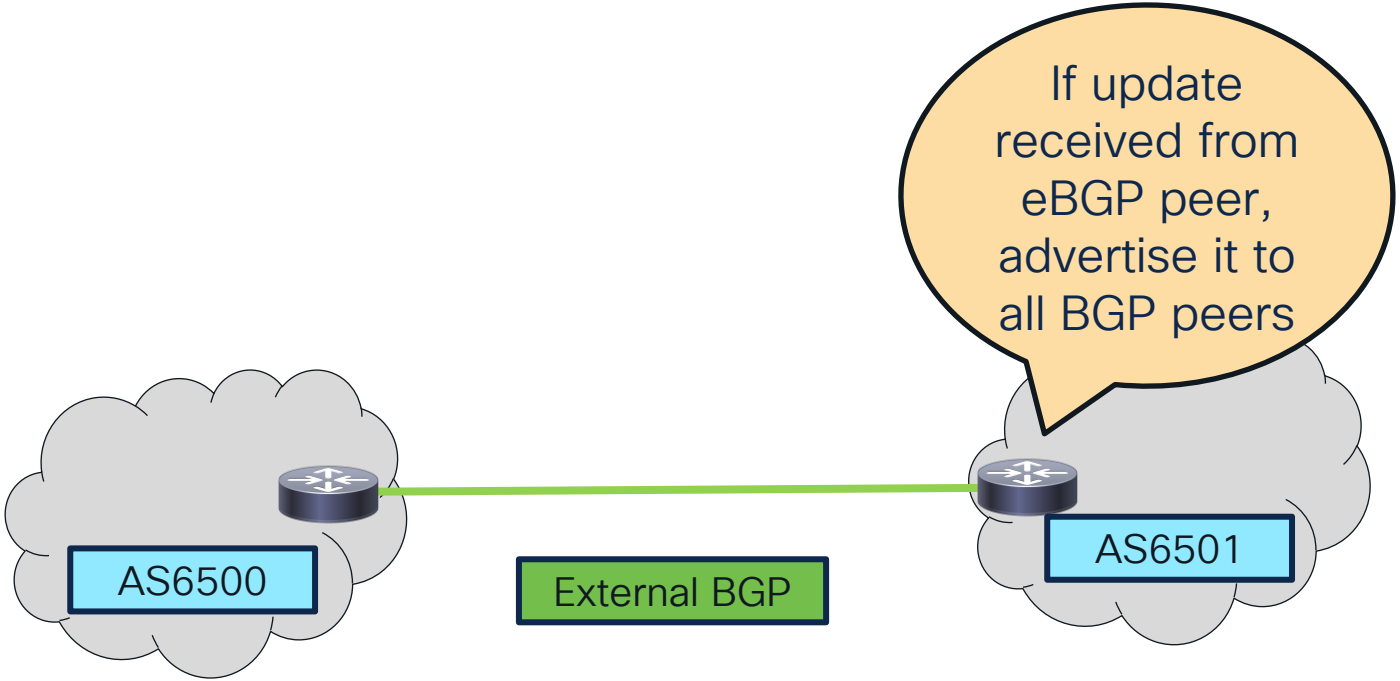
AS_PATH in eBGP



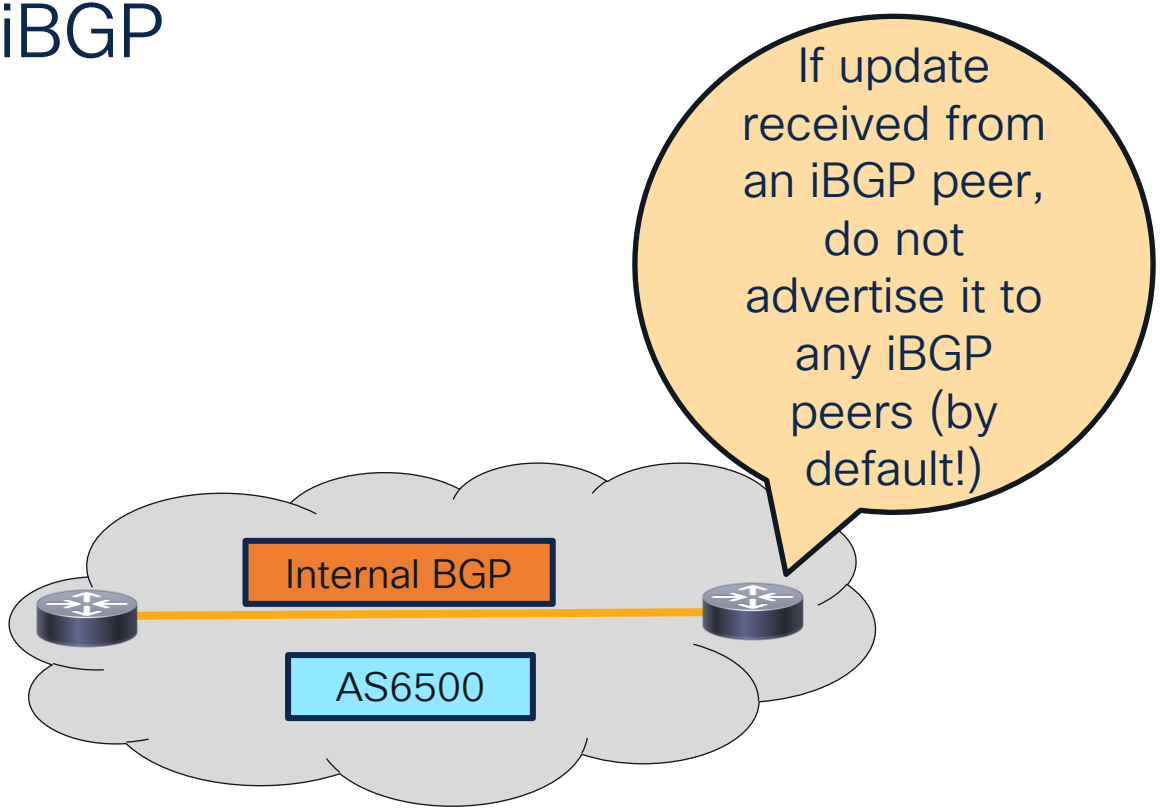
AS_PATH in iBGP



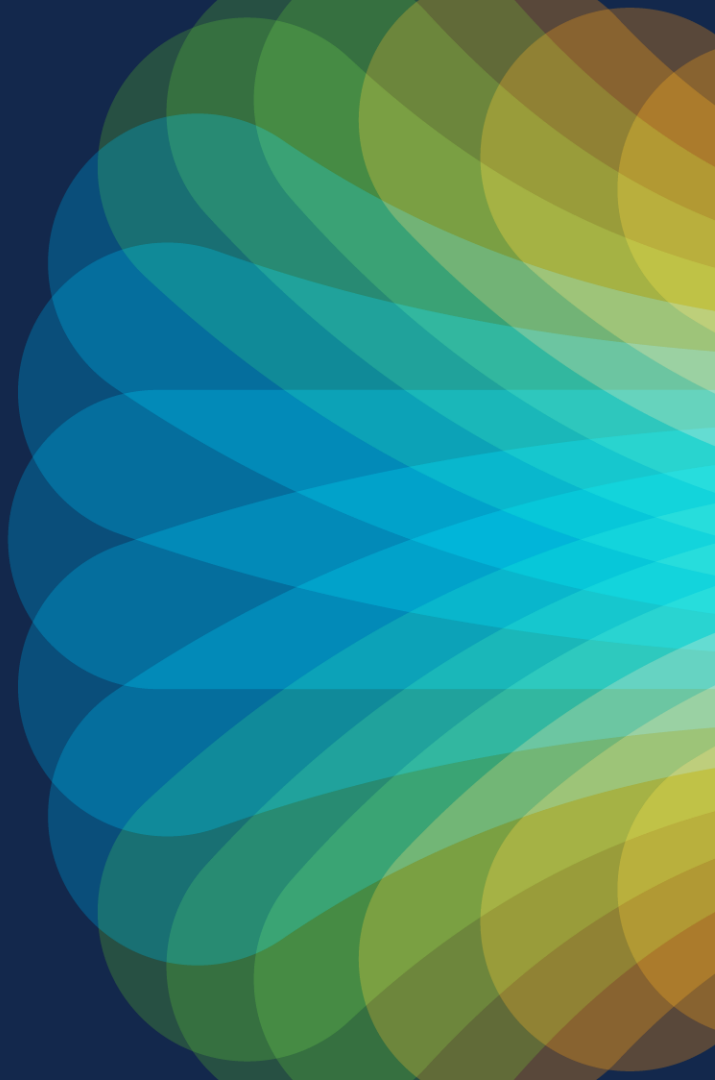
Updates in eBGP



Updates in iBGP



Attributes



BGP Attributes

- An attribute is an additional piece of information accompanying an advertised NLRI

- BGP uses attributes in multiple ways
  - Prevents routing loops
  - Performs best path selection
  - Filter routes

- Basic BGP specification recognizes only a handful of attributes
  - Several new have been added over time for various applications and uses

BGP Attribute Types

- **Well-known**: Every BGP implementation must support it
  - Well-known mandatory: Must always be included with a NLRI
  - Well-known discretionary: May be included with a NLRI as needed

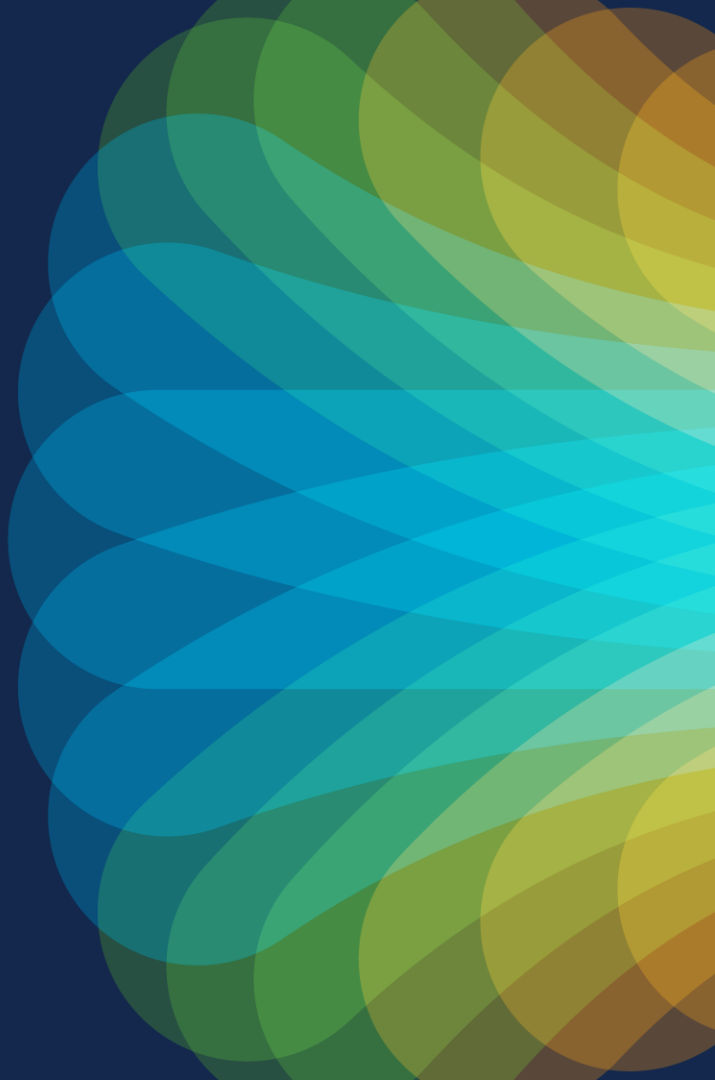
- **Optional**: BGP implementations do not need to support it
  - Optional transitive: When advertising a learned NLRI, keep the attribute with the NLRI even if not recognized
  - Optional non-transitive: When advertising a learned NLRI, remove the attribute from the NLRI if not recognized

Note: All well-known attributes are transitive

BGP Attributes

- Well-known mandatory:
  - AS_PATH
  - NEXT_HOP
  - ORIGIN
- Well-known discretionary
  - LOCAL_PREF
  - ATOMIC_AGGREGATE
- Optional transitive
  - AGGREGATOR
  - COMMUNITIES
  - EXTENDED_COMMUNITIES
- Optional non-transitive
  - MULTI_EXIT_DISC
  - CLUSTER_LIST

Best Path Selection Algorithm



BGP Best Path Selection

BGP Best Path Selection Algorithm as Implemented on Cisco Devices	
1. Highest WEIGHT (Cisco propriety)	7. eBGP-learned route over iBGP-learned one
2. Highest LOCAL_PREF	8. Lowest IGP metric to the next hop
3. Locally originated (injected) route	9. Oldest eBGP-learned route
4. Shortest AS_PATH / AS4_PATH	10. Lowest BGP peer's Router ID
5. Lowest ORIGIN code	11. Shortest CLUSTER_LIST
6. Lowest MULTI_EXIT_DISC	12. Lowest BGP peer's address

BGP Best Path Selection

Reference: Select BGP Best Path Algorithm

1. Next-hop has to be accessible (In the routing table)
2. Route must be synchronised (Better turn synchronisation off)
3. Highest weight (Admin Preference, local to the router)
4. Largest local preference (Admin Preference. Spread within AS)
5. Router originated (Metric= “0 ASes” - Better if we originated it)
6. Shortest AS-PATH (Metric in AS’s)
7. Lowest origin (igp < egp < incomplete)
8. Lowest MED (Metric information from the next AS)

Continued..

BGP Best Path Selection *(continued)*

9. External over internal (Metric better if we are the border router)
10. Closest next-hop (IGP metric - the next-hop must be close)
11. Lowest router-id of Originator (Tie-breaker)
12. Shortest Cluster-list (Tie-breaker)
13. Lowest IP address of neighbour (Tie-breaker)

Path Selection

- Best path is used locally and to advertise to other BGP speakers
- All NLRI, select the first variant as the best one
- When more than one path exists, go through every variant and choose one

Router performing the best path selection

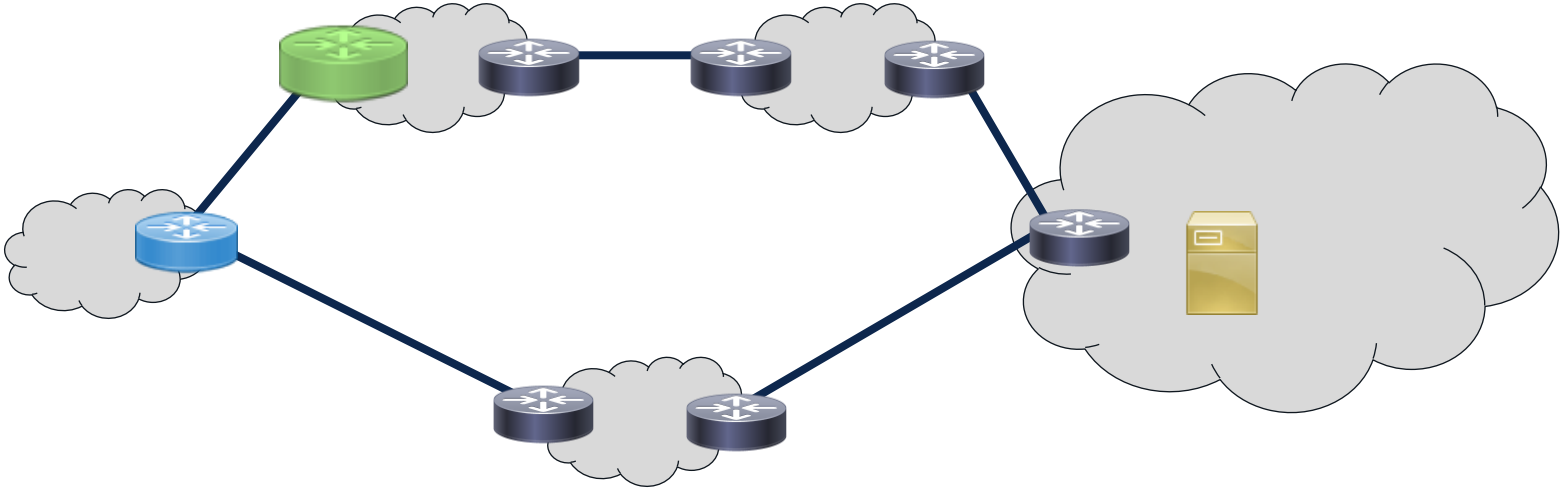
The resulting next hop

Generic router (unspecific)

Destination (NLRI)

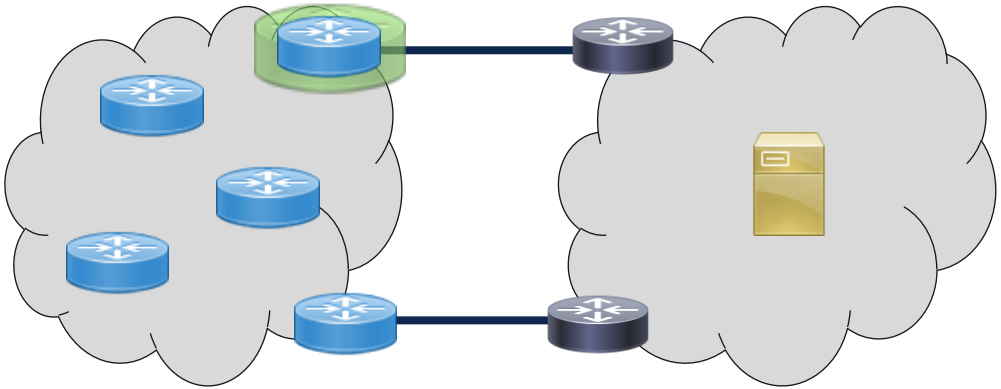
The logic behind BGP best path selection (1)

- Step 1: Prefer the path with the higher WEIGHT
  - Rationale: Always have means to override the path choice locally
  - Note: This is an override rule



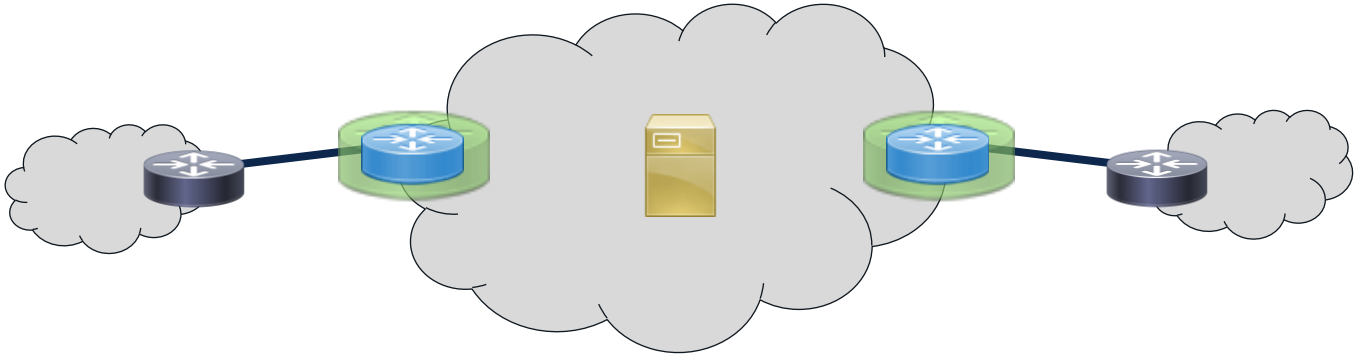
The logic behind BGP best path selection (2)

- Step 2: Prefer the path with the higher LOCAL_PREF
  - Rationale: Have means to override the best path for the entire AS from a single exit point
  - Note: This is an override rule



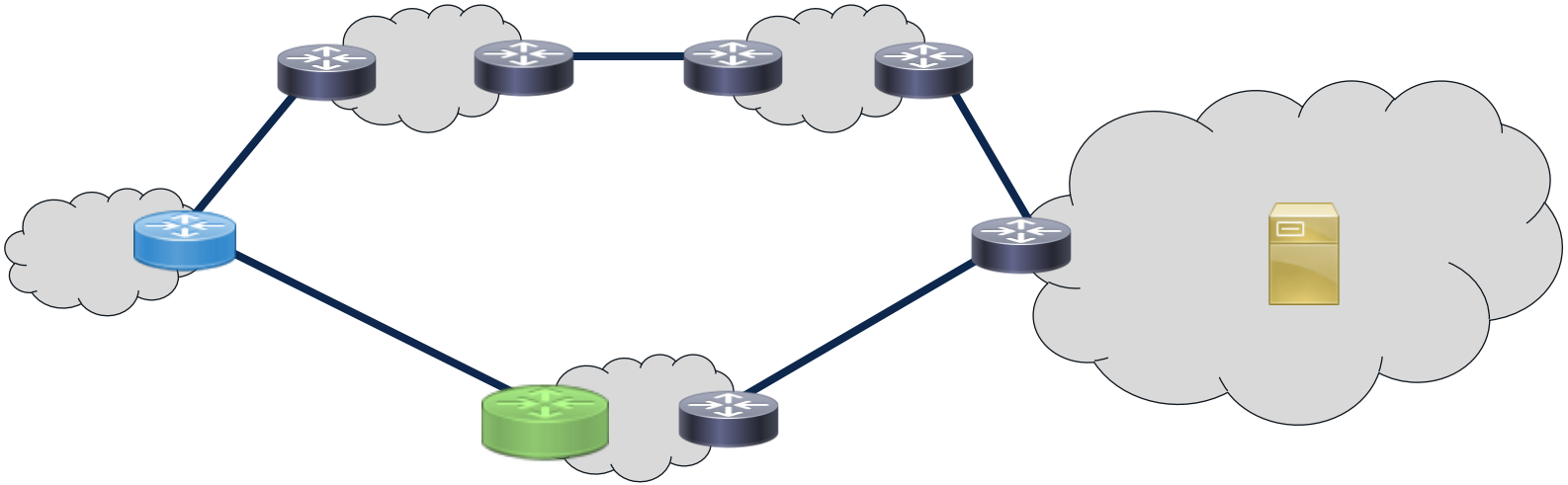
The logic behind BGP best path selection (3)

- Step 3: Prefer the LOCALLY ORGINATED path (network, redistribution, aggregation)
  - Rationale: I get a chance to speak on behalf of my own local AS
  - Note: The best route is not just for me to use but also to advertise to *others* so that *they know*



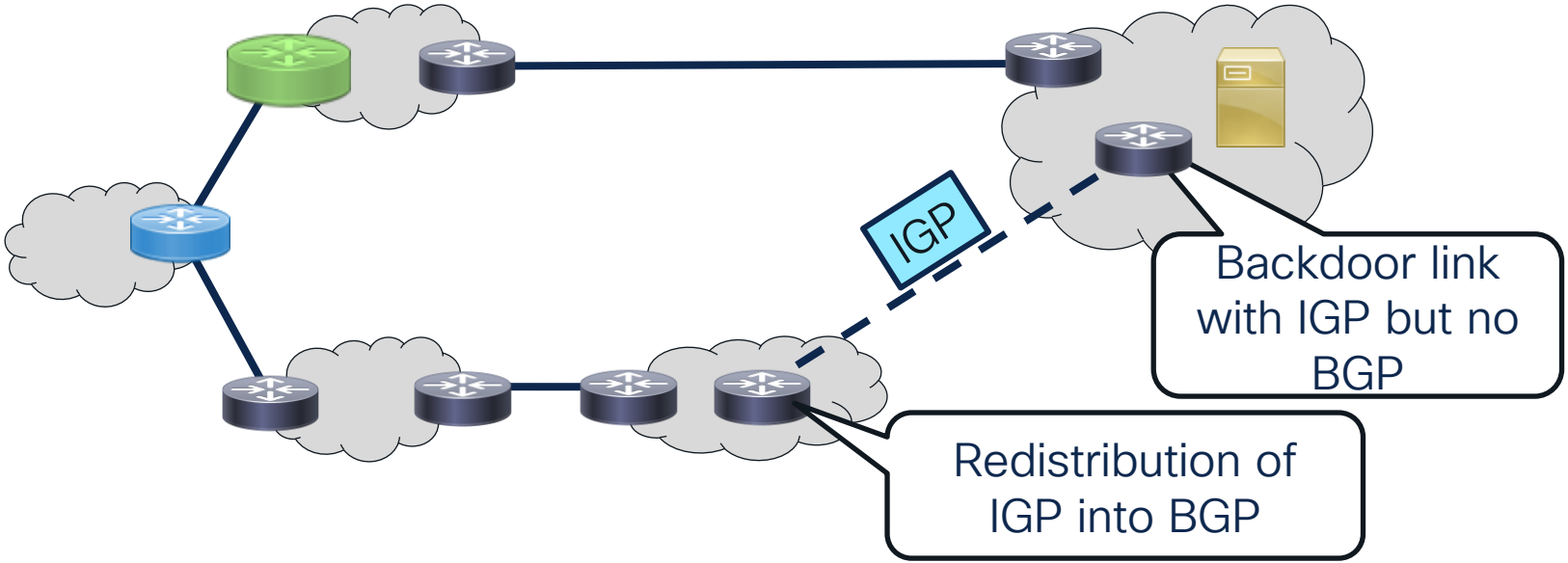
The logic behind BGP best path selection (4)

- Step 4: Prefer the path with the shortest AS_PATH / AS4_PATH
  - Rationale: Traverse the least amount of autonomous systems



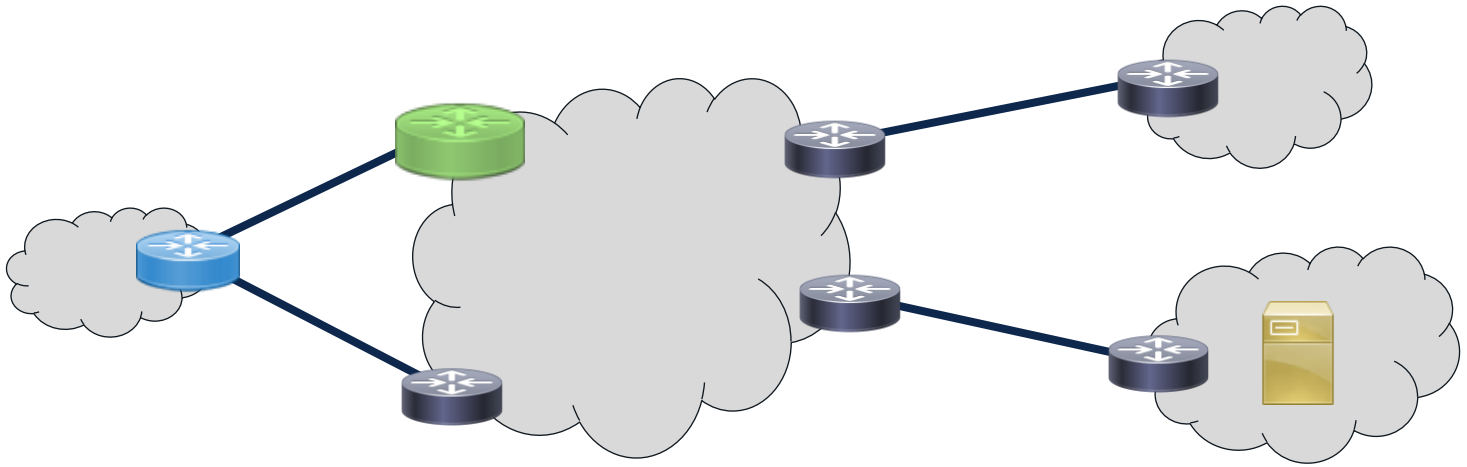
The logic behind BGP best path selection (5)

- Step 5: Prefer the path with the lower ORIGIN code
  - Rationale: Take the *most trustworthy* path
  - Note: IGP is lower than EGP, EGP is lower than Incomplete



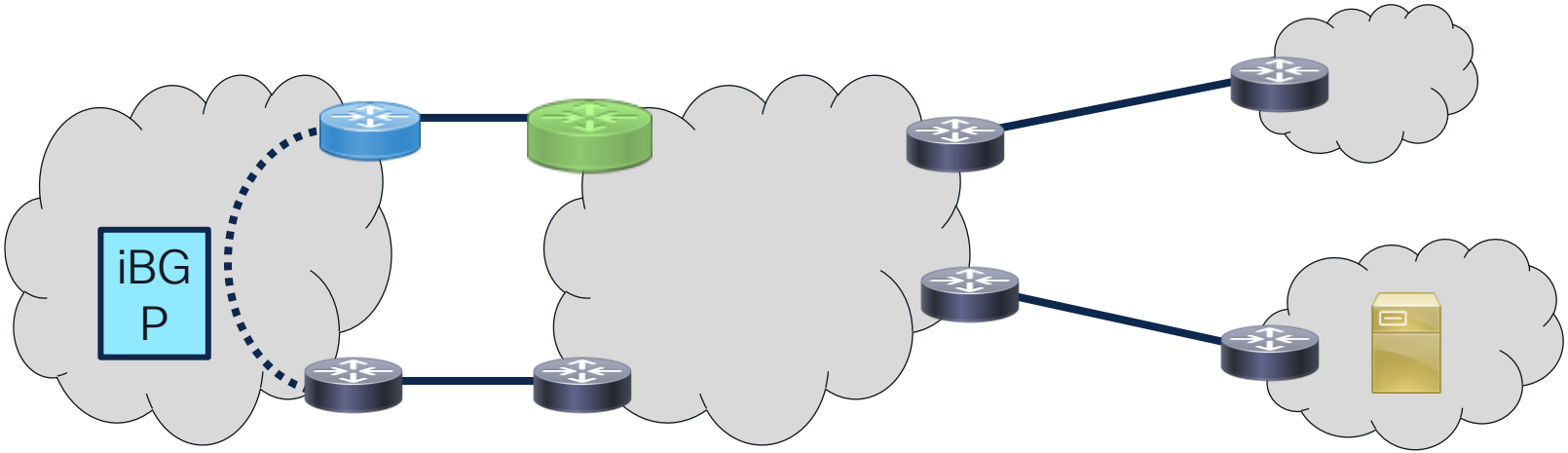
The logic behind BGP best path selection (6)

- Step 6: Prefer the path with the lower MULTI_EXIT_DISC
  - Rationale: Respect the *preferred path hint* indicated by the neighbor AS



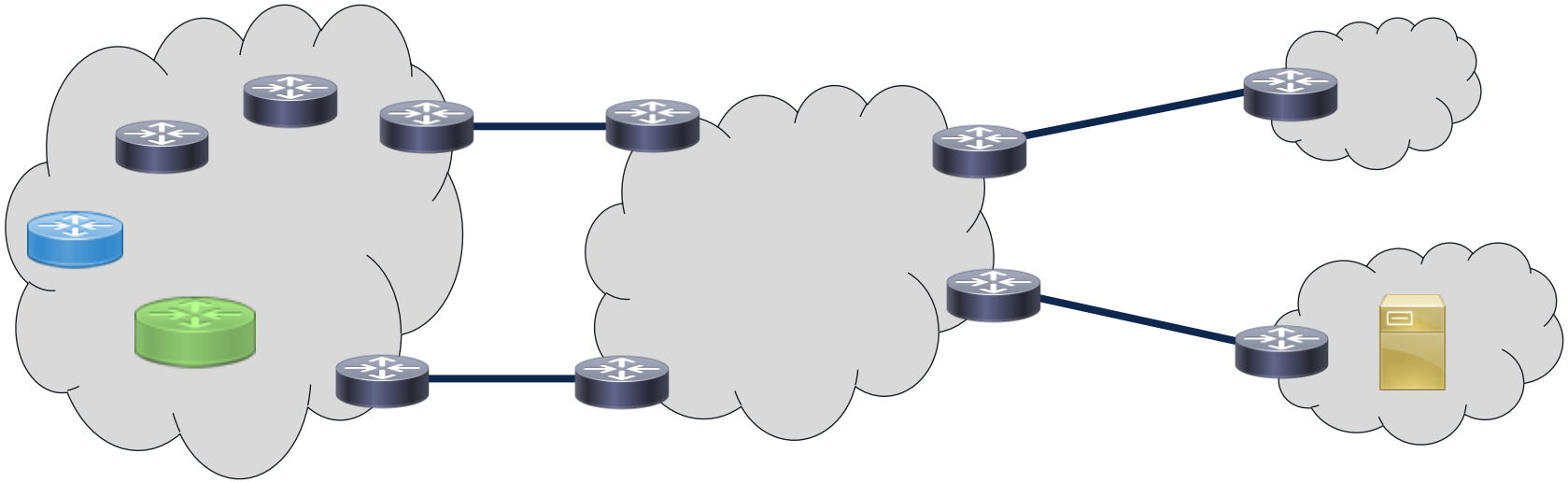
The logic behind BGP best path selection (7)

- Step 7: Prefer eBGP-learned path over iBGP-learned one
  - Rationale: If you need to *leave* the local AS, *leave right away if you can*



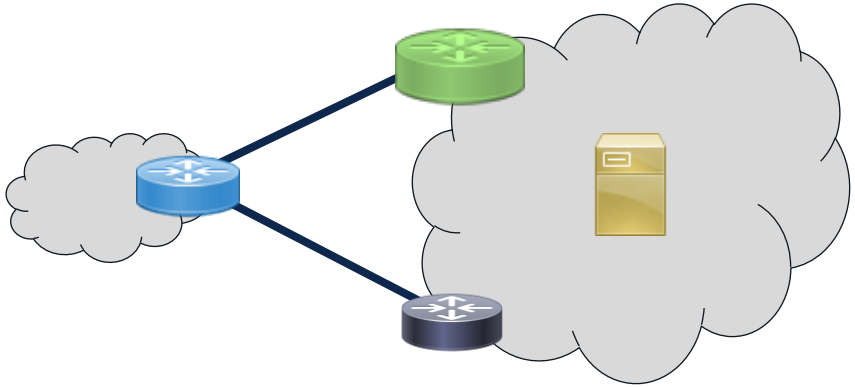
The logic behind BGP best path selection (8)

- Step 8: Prefer the path with the LOWEST IGP METRIC to the next hop
  - Rationale: If you need to *traverse* the local AS but *can't leave* right away, just take the shortest path toward the exit



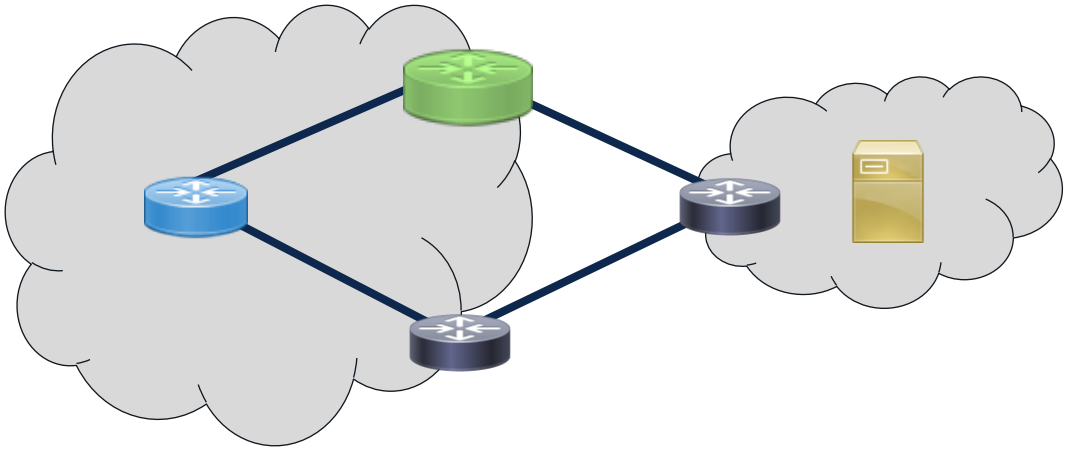
The logic behind BGP best path selection (9)

- Step 9: If both paths are learned via eBGP, prefer the OLDER ONE
  - Rationale: The eBGP paths are, by this point, effectively equal – so don't bother updating anything

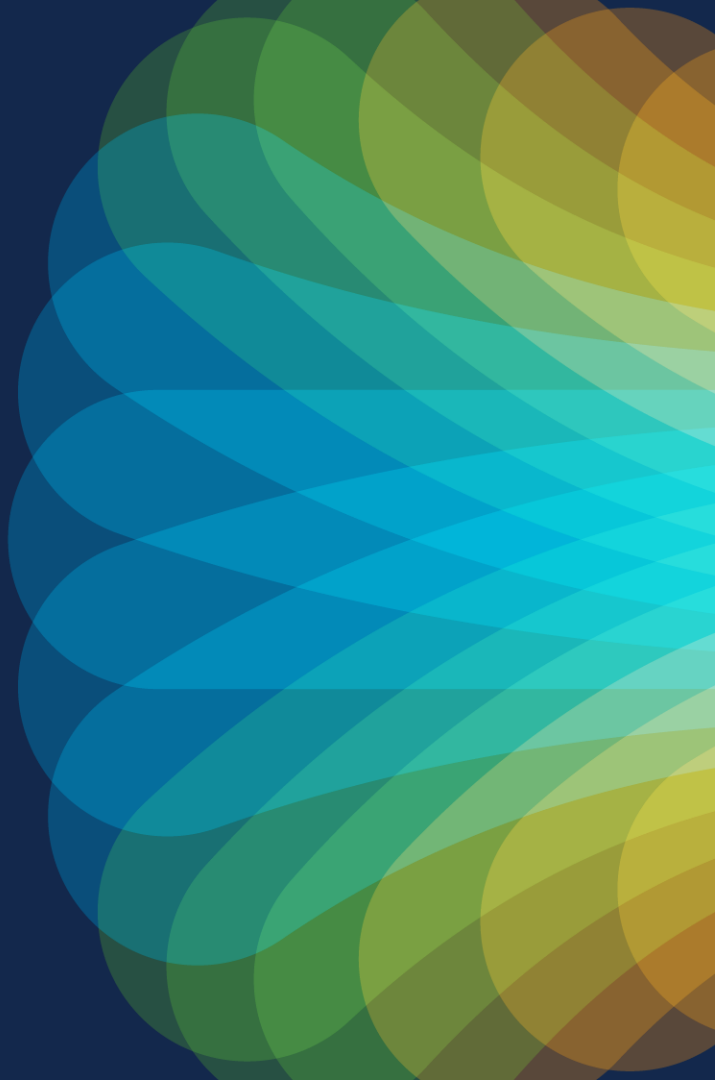


The logic behind BGP best path selection (10-12)

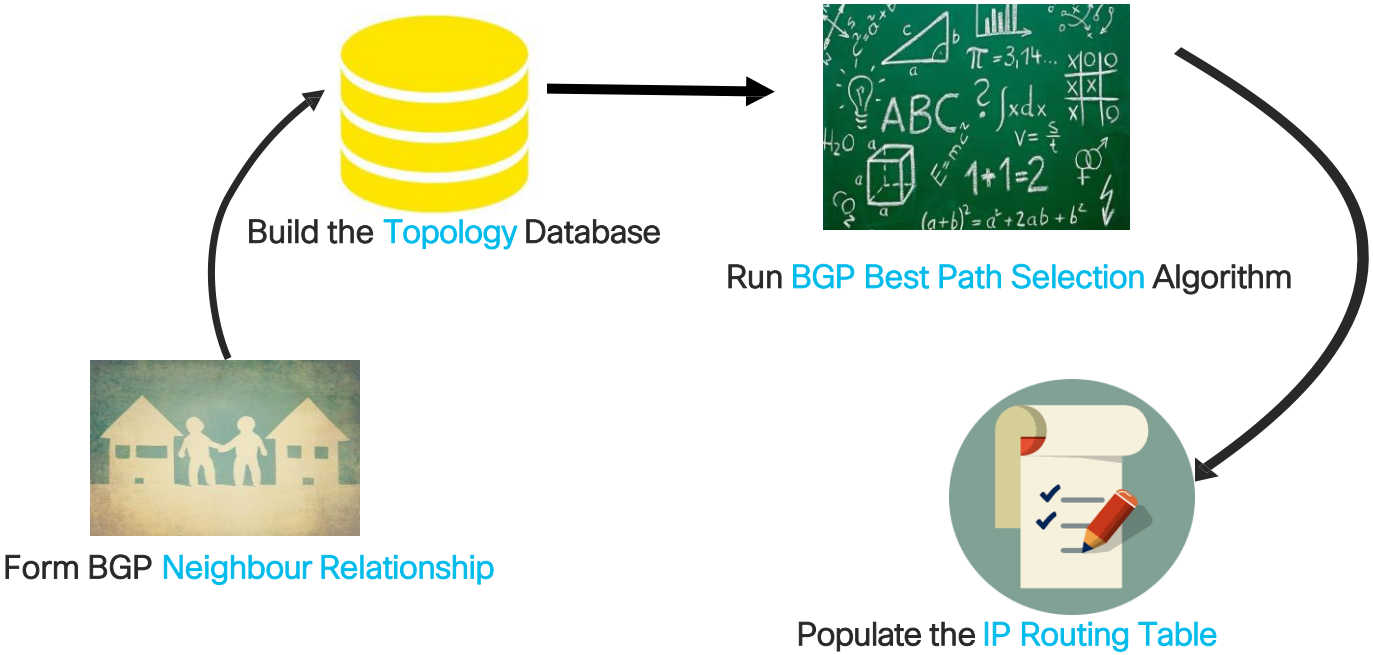
- Steps 10-12: Prefer the path learned from the BGP peer with the LOWER ROUTER ID, then with the SHORTER CLUSTER_LIST, then from the BGP peer with the lower peering IP address
  - Rationale: Technical tiebreakers to arrive at *exactly one best path*



Troubleshooting BGP



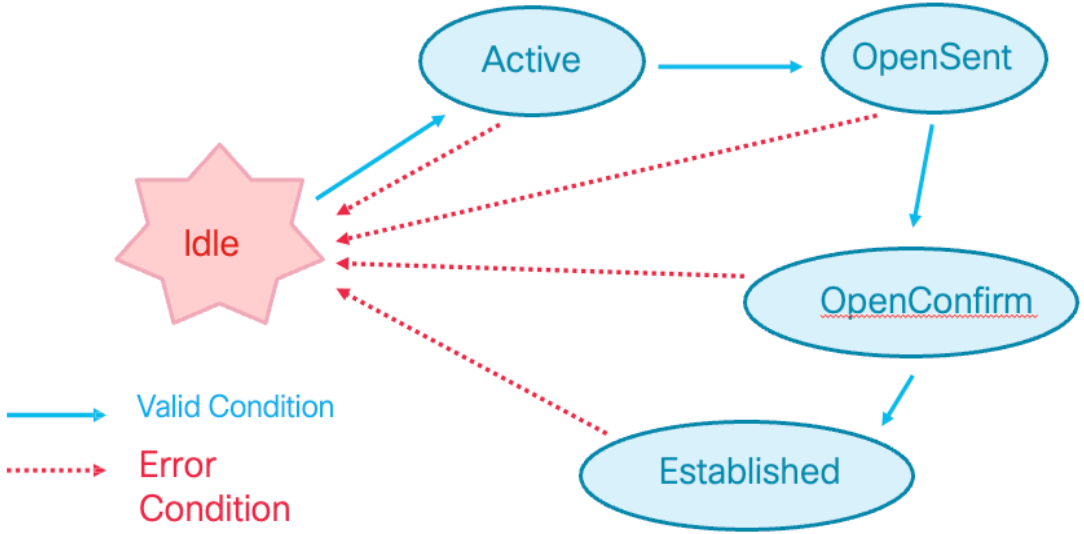
Key Troubleshooting Process



Lifecyle of BGP Peers

- BGP uses TCP port 179
- Peers exchange OPEN messages:
  - Router ID
  - AS #
  - Hold Time
  - Capabilities
- Initial exchange of entire table
- UPDATE/WITHDRAW messages are event based
- Periodic KEEPALIVE messages
- NOTIFICATION messages to signal why BGP closed

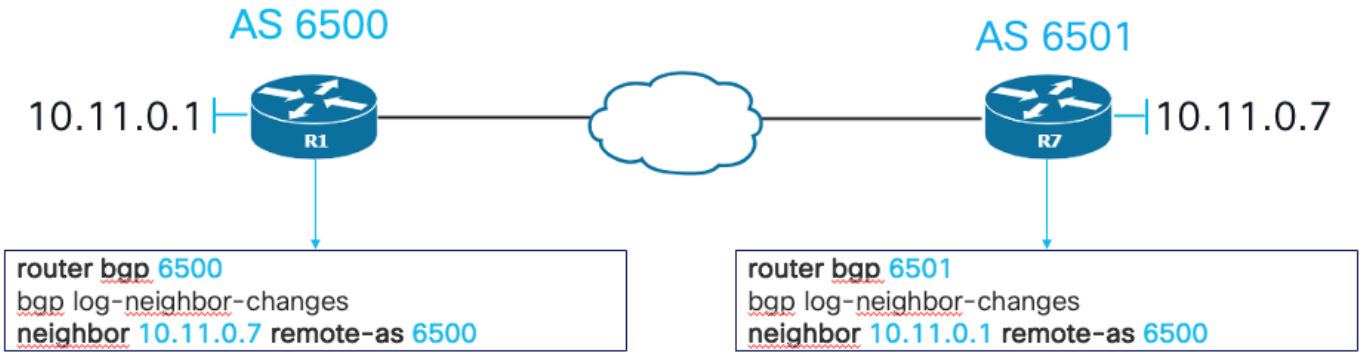
Neighbour Adjacency States



Configuration

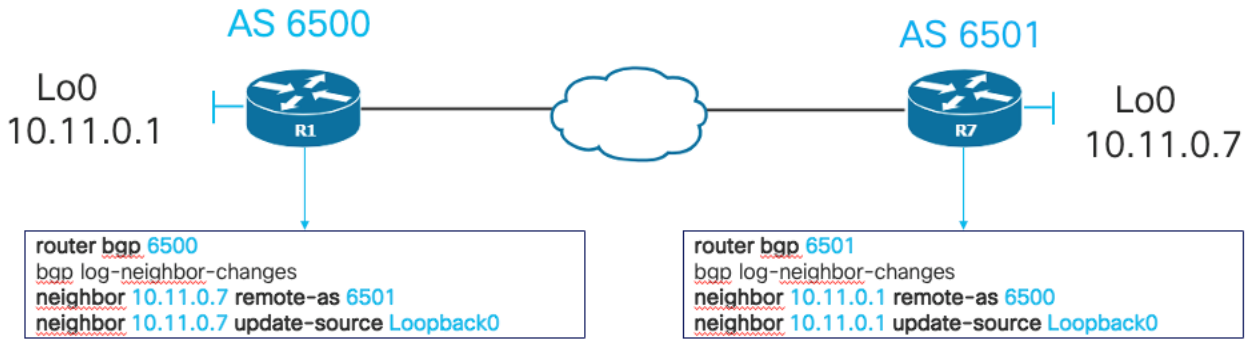
Validate local / remote AS number and peer IP

Will this establish BGP peering?



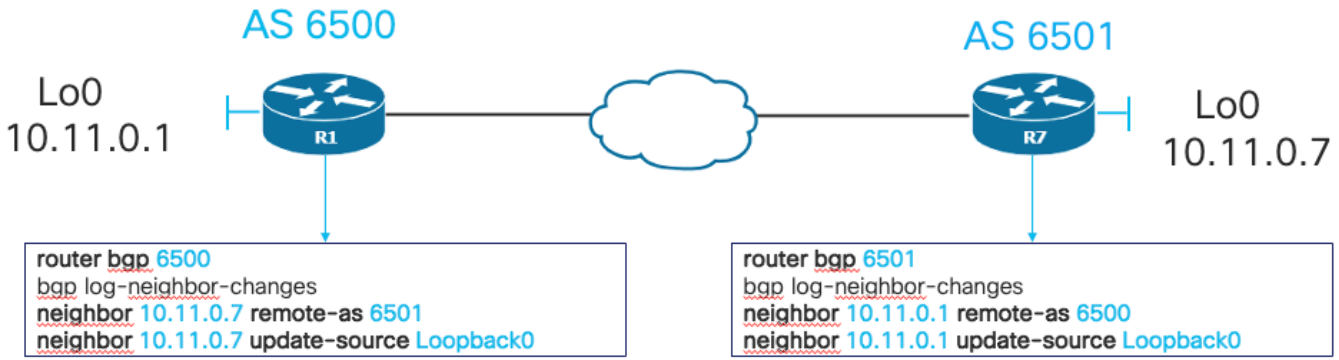
Configuration

For non-directly connected peers specify source interface



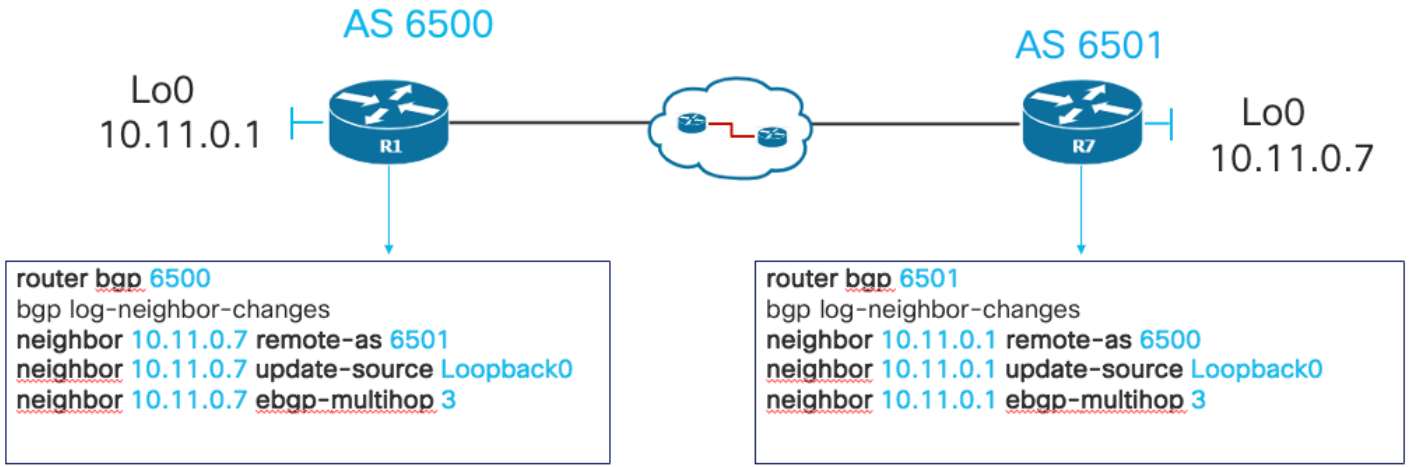
Configuration

For non-directly connected peers specify source interface



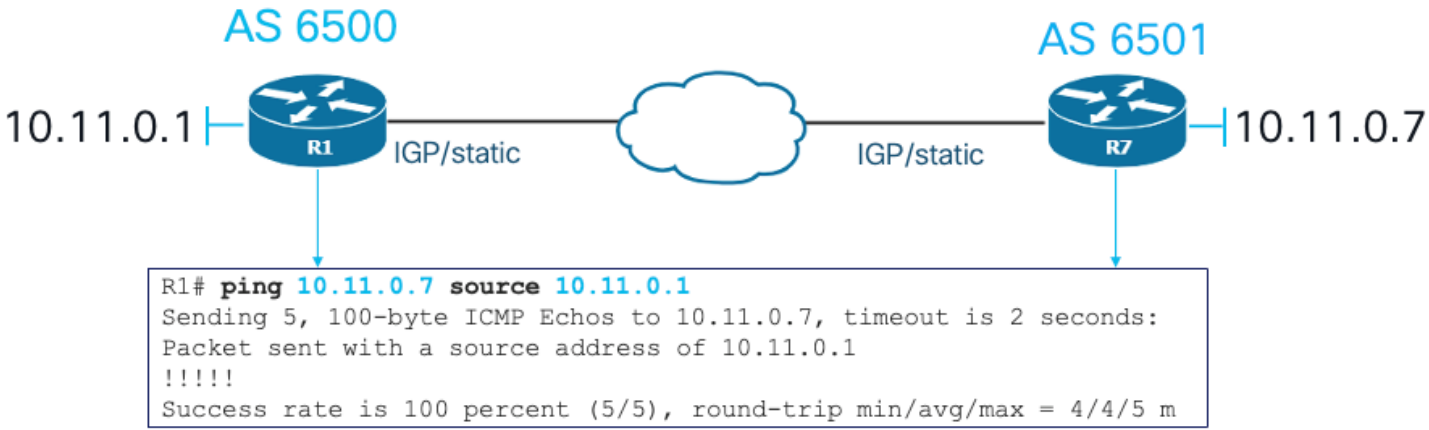
Configuration

eBGP uses TTL=1 by default



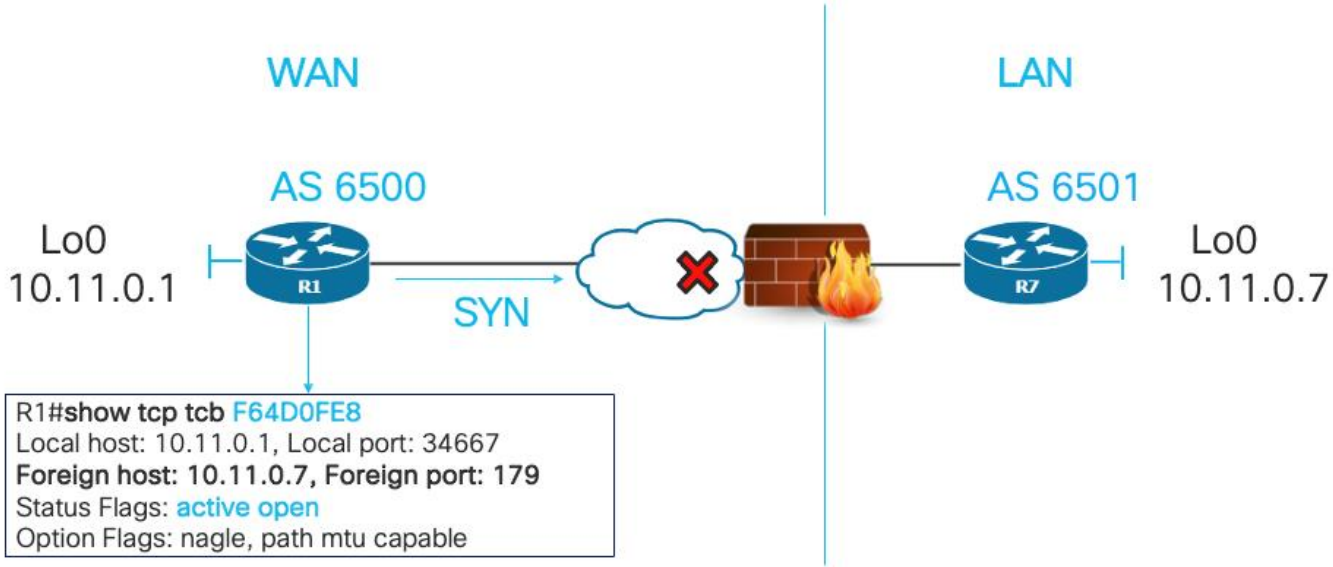
Troubleshoot Neighbourship

Verify layer 3, ICMP reachability



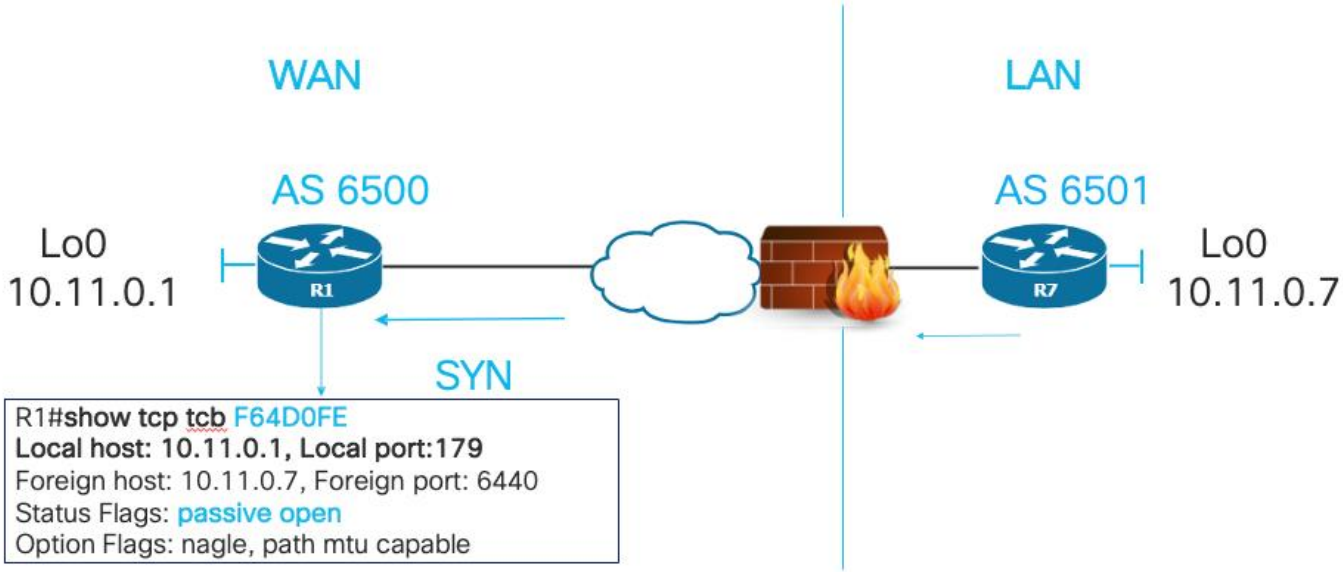
Verify TCP Connection

Transport connection is “active open”



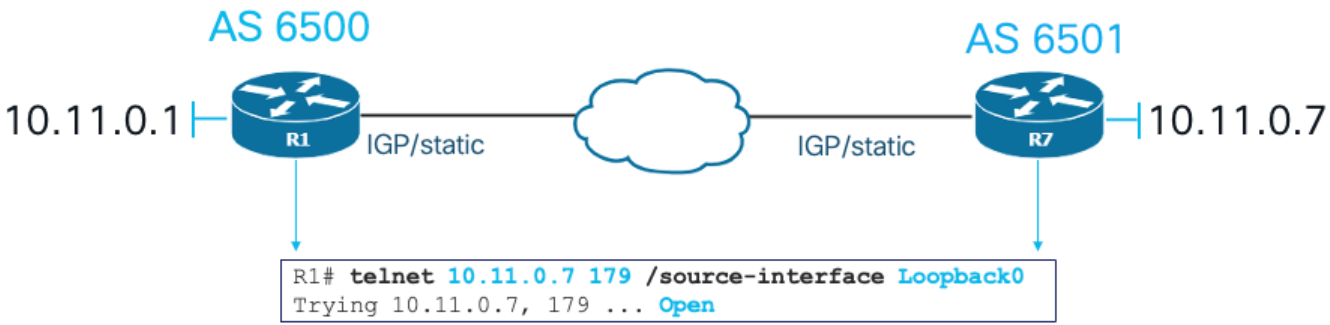
Verify TCP Connection

Change transport connection mode to passive



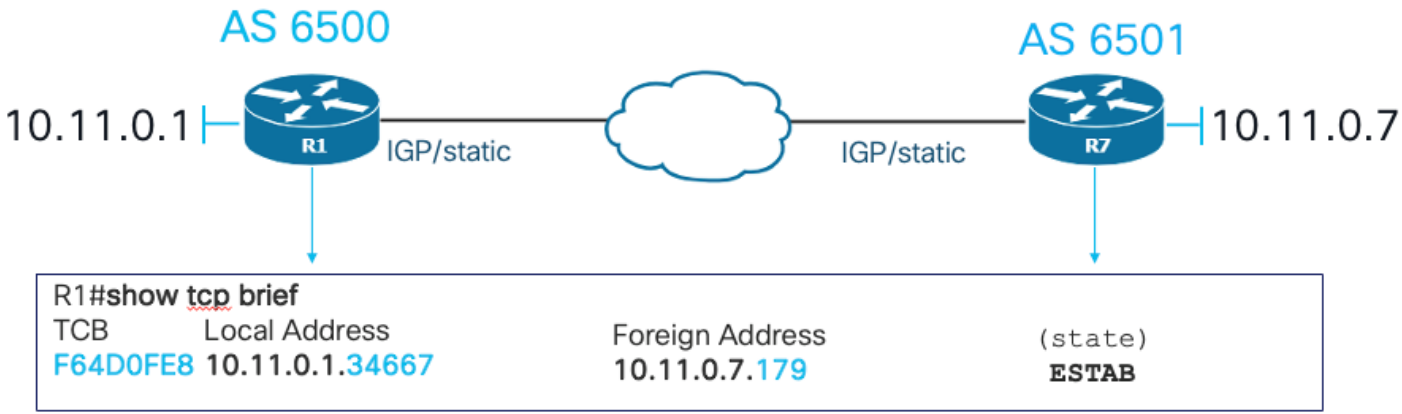
Troubleshoot Neighbourship

Verify layer 4, TCP 179 access between BGP peers



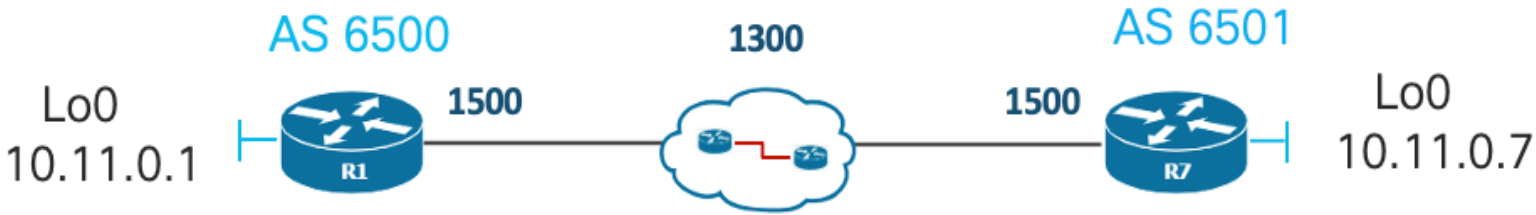
Troubleshoot Neighbourship

Verify TCP's state



Troubleshoot MTU issues

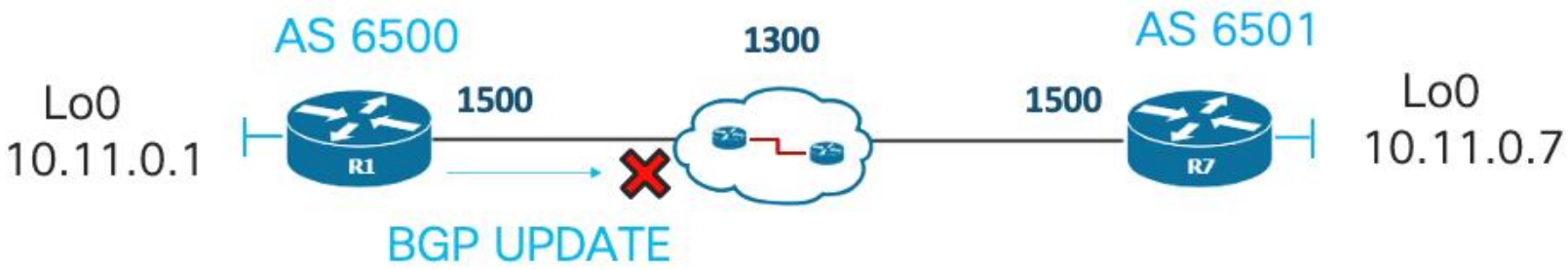
Will this work?



Troubleshoot MTU issues

BGP UPDATE packets are sent with the DF-bit set

BGP relies on Path MTU Discovery (PTMUD) to identify the largest packet size which can be sent without fragmentation



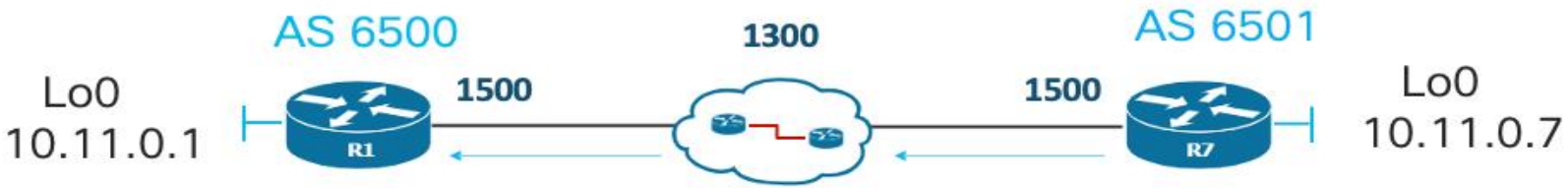
Troubleshoot MTU issues

PTMUD relies on ICMP Message to detect the lowest MTU in the transit path

Troubleshoot MTU issues

Perform df-bit ping to identify highest MTU without fragmentation in path

- ip tcp mss
- ip tcp adjust-mss
- Disable PTMUD to allow BGP to establish with a lower MSS



Concluding Remarks

- BGPv4 is ~30 years old but its core is still the same
  - A credit to its well-thought design

- BGP is a world on its own – where to learn more?
  - Cisco Press textbooks
  - Cisco Communities, Cisco Learning Network
  - IETF RFCs
  - Wireshark
  - Get your hands dirty

cisco Live!

Did you know?

You can have a
one-on-one session with
a technical expert!

Visit Meet the Expert in The HUB
to meet, greet, whiteboard & gain
insights about your unique questions
with the best of the best.

Meet the Expert Opening Hours:

Tuesday	**3:00pm – 7:00pm**
Wednesday	**11:15am – 7:00pm**
Thursday	**9:30am – 4:00pm**
Friday	**10:30am – 1:30pm**

Session Surveys

We would love to know your feedback on this session!

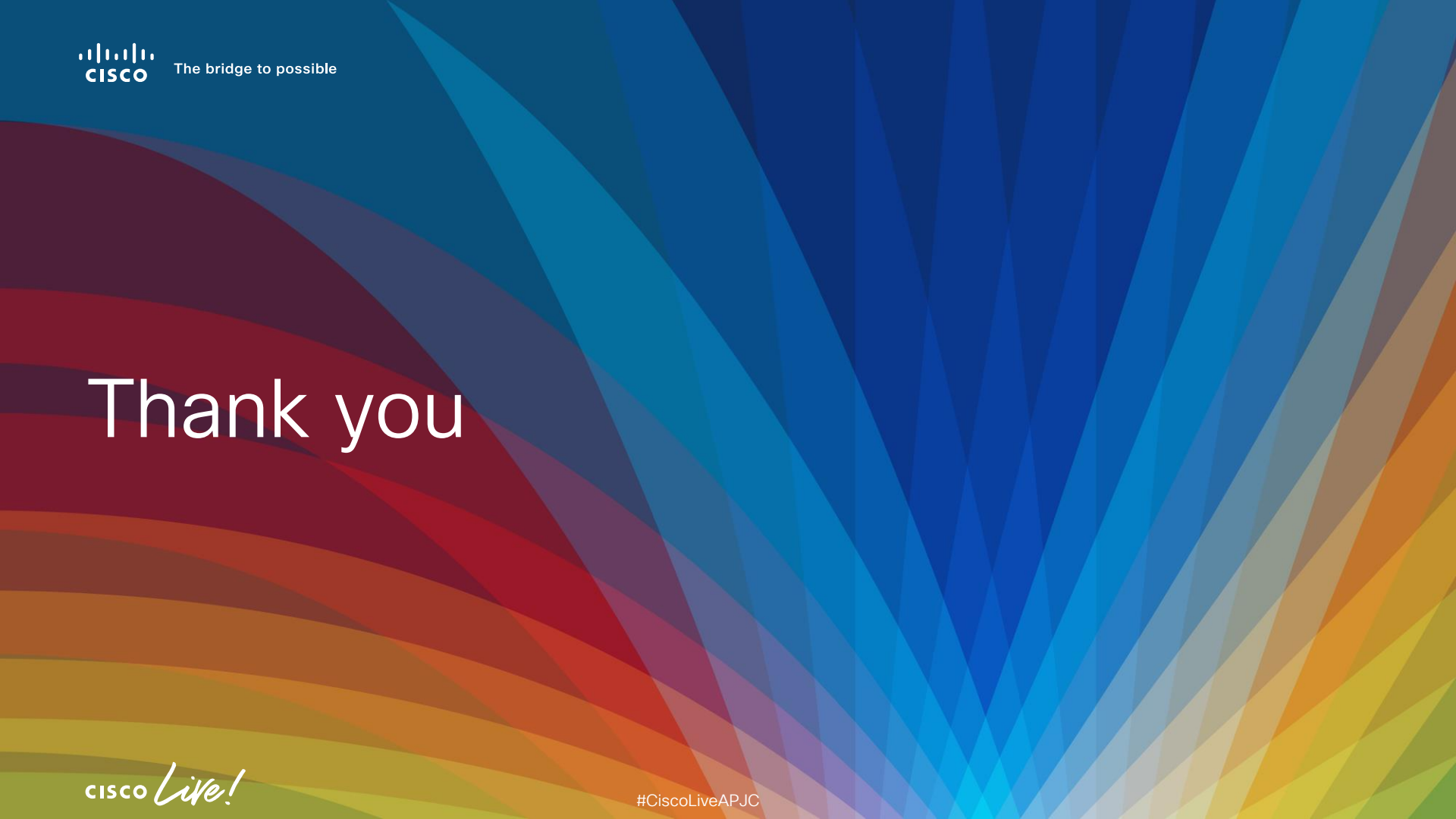
- Complete a minimum of four session surveys and the overall event surveys to claim a Cisco Live T-Shirt

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Expert meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

“Like a fine wine, BGP only gets better with age.”

Anonymous


cisco
The bridge to possible
Thank you
cisco Live!
#CiscoLiveAPJC


cisco Live!
Let's go
#CiscoLiveAPJC