BGP – FORGOTTEN BUT USEFUL FEATURES

Piotr Wojciechowski (CCIE #25543)

About me

- Senior Network Engineer MSO at VeriFone Inc.
- Previously Network Solutions Architect at one of top polish IT integrators
- CCIE #25543 (Routing & Switching)
- Blogger – http://ccieplayground.wordpress.com
- Administrator of CCIE.PL board
  - The biggest Cisco community in Europe
  - Over 6800 users
  - 3 admin, 7 moderators
  - 58 polish CCIEs as members, 20 of them actively posting
  - About 150 new topics per month
  - About 1000 posts per month
  - English section available!

Agenda

- BGP Origin AS Validation
- BGP Multipath
- BGP Additional Paths
- BGP Dynamic Neighbors
- BGP Slow Peer

BGP Origin AS Validation

Why We Should Verify Origin AS?

- Prefix hijacking – any AS can advertise any prefix in BGP
  - Human mistake
  - Malicious
- How we can hijack prefix?
  - Advertise someone else's prefix
  - Advertise more specific of someone else's prefix

Why We Should Verify Origin AS?

- Has it happened before?
  - Pakistan Telecom blocks YouTube
    - In February 2008, Pakistan Telecom inadvertently brought down the entire YouTube site worldwide for two hours as it was attempting to restrict local access to the site. When Pakistan Telecom tried to filter access to YouTube, it sent new routing information via BGP to PCCW, an ISP in Hong Kong that propagated the false routing information across the Internet.

Source: NetworkWorld: Six worst Internet routing attacks

Why We Should Verify Origin AS?

- Has it happened before?
  - Northrop Grumman hit by spammers
    - In May 2003, a group of spammers hijacked an unused block of IP address space owned by Northrop Grumman and began sending out massive amounts of unwanted e-mail messages. It took two months for the military contractor to reclaim ownership of its IP addresses and get the rogue routing announcements blocked across the Internet. In the meantime, Northrop Grumman's IP addresses ended up on high-profile spam blacklists

Source: NetworkWorld: Six worst Internet routing attacks

Why We Should Verify Origin AS?

- Has it happened before?
  - Biggest-ever BGP threat unveiled
    - In August 2008, two security researchers demonstrated at DEFCON how an attacker could eavesdrop or change a company's unencrypted data by exploiting BGP. The attacker would reroute all of the company's traffic through their own network and then send it to its destination without the owner's knowledge.

Source: NetworkWorld: Six worst Internet routing attacks

Origin AS Validation

- The network administrator must set up a Resource Public Key Infrastructure (RPKI) server
- The RPKI server handles the actual authentication of public key certificates.

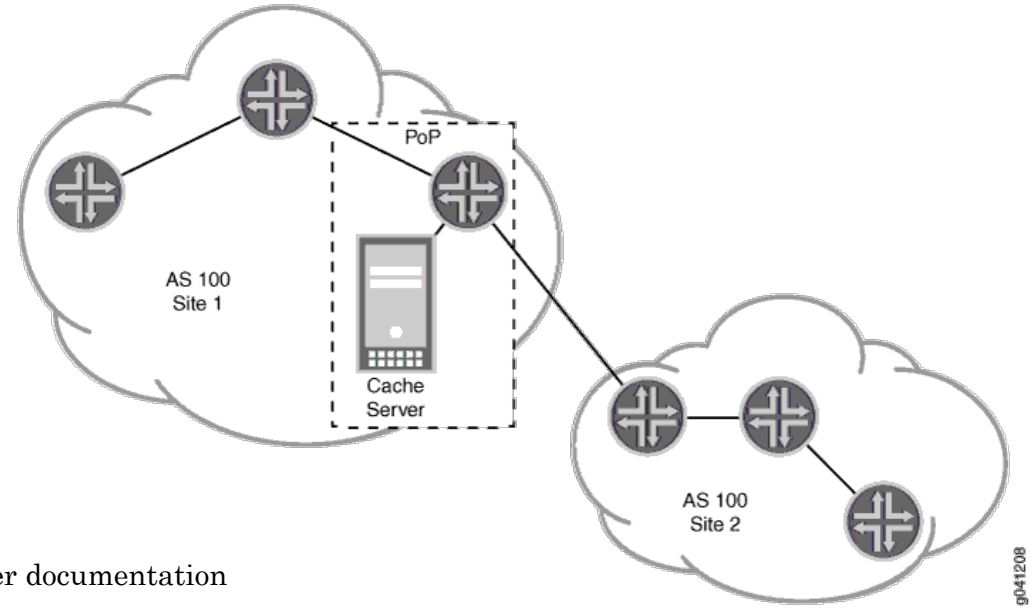

Image source: Juniper documentation

Origin AS Validation

- Router can handle the actual authentication of public key certificates but it's very resource (CPU) intensive and not preferred way of deployment.

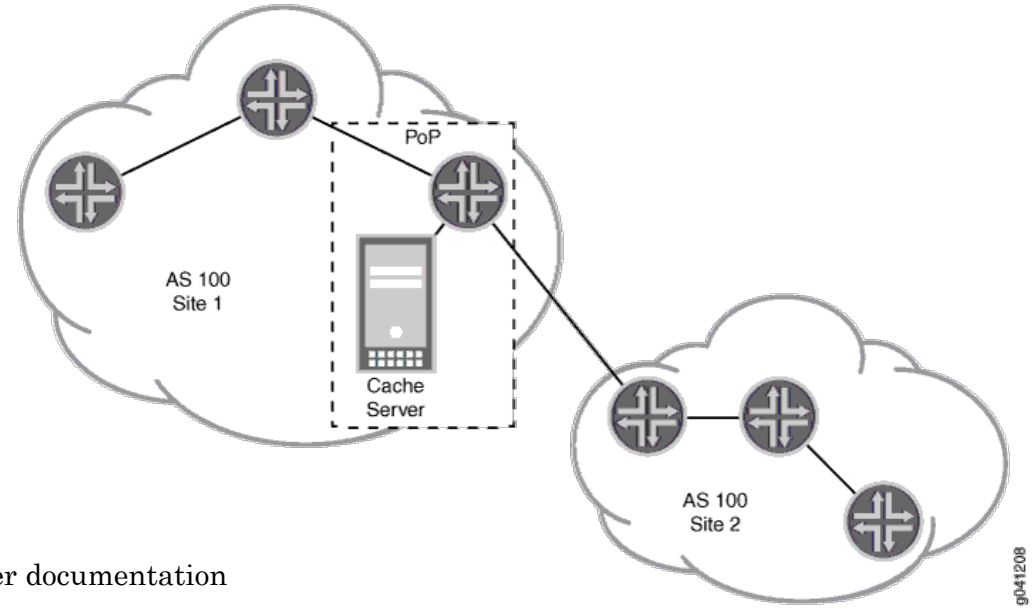

Image source: Juniper documentation

Origin AS Validation

- The server is set up so that certain prefixes or prefix ranges are allowed to originate from certain autonomous systems.
- Router establishes connection to RPKI server and downloads a list of prefixes and permitted origin AS numbers from one or more router/RPKI servers using the RPKI-Router protocol (RTR)

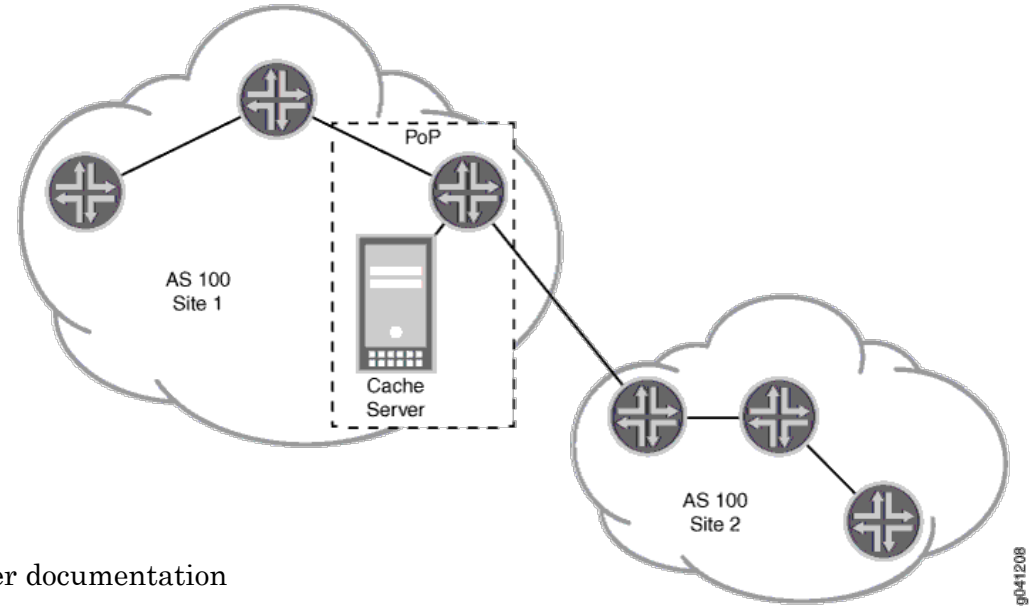

Image source: Juniper documentation

Origin AS Validation

- Components

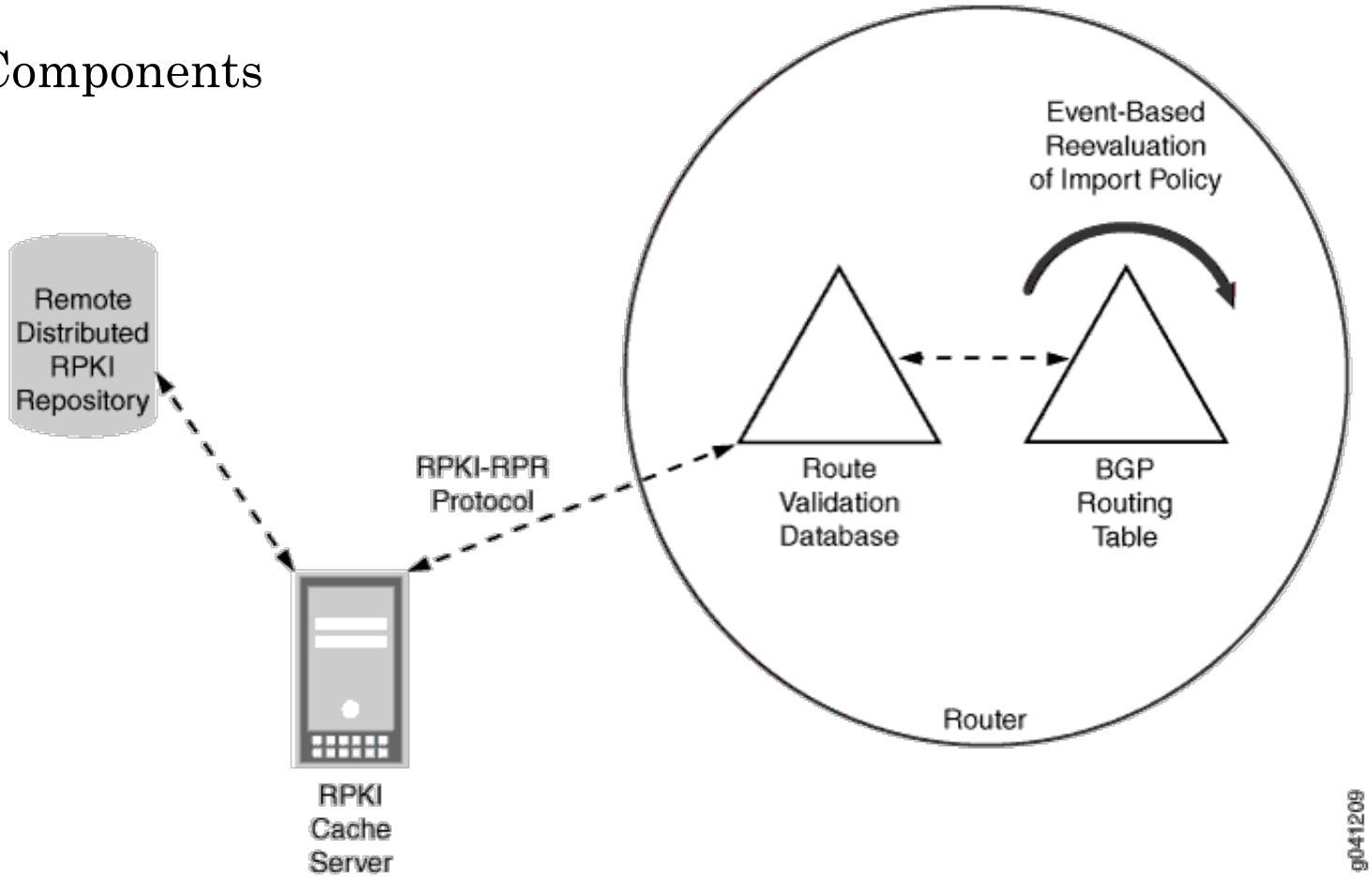

Image source: Juniper documentation

Origin AS Validation

- Route origin authorization (ROA)
  - ROA is a digitally signed object distributed through the RPKI infrastructure
  - Indicates the address prefix holder's explicit authorization that an AS can rightfully originate a prefix
  - Is not directly used in route validation
  - Cache servers exports ROA as RV records to the routers

Origin AS Validation

- Route Validation (RV) record:
  - It's (prefix, maximum length, origin AS) triple.
  - It matches any route whose prefix matches RV prefix, whose prefix length does not exceed maximum length given by RV record, and whose origin AS equals the origin AS given in the RV record.
  - If maximum length not present then AS is only allowed to advertise exact prefix specified in RV

Origin AS Validation

- Example
  - RV – (100.10.1.0/24, /26, AS100)
  - Which of following prefixes are authorized to be announced?

100.10.1.0/24
100.10.1.0/25
100.10.1.128/25
100.10.1.128/26

Answer: All of them!

Origin AS Validation

- Example
  - RV – (100.10.1.0/24, /25, AS100)
  - Which of following prefixes are authorized to be announced?

100.10.1.0/24

100.10.1.0/25

100.10.1.128/25

100.10.1.128/26

Answer: Last one is not valid!

Origin AS Validation

- When prefix is received from an eBGP peer it is examined against import policy and marked:
  - VALID – prefix and AS pair are found in the database
  - INVALID – prefix is found but either corresponding AS or prefix length does not match database
  - UNKNOWN – prefix is not in database
- Prefix marked as INVALID is withdrawn from routing table and not advertised to peers.
- VALID prefix is preferred over UNKNOWN – it affects BGP path selection

Origin AS Validation

- Prefix validation marking done only for eBGP updates
- Validation state is carried across iBGP mesh in an opaque extended community (non-transitive)

Origin AS Validation

- Available on:
  - Cisco
    - IOS 15.2(4)S or later – platforms ME3600, ME3800, 7200-NPE-G2, 7201
    - IOS XE 3.10S or later – platforms ASR1000-RP1, ASR1000-RP2, ASR-1001, ASR1002-X, ISR4451-X, CSR100V
  - Juniper
    - JunOS 12.2 or later (cache server required)

Origin AS Validation

- References
  - PLNOG10: Alex Band (RIPE), **Resource Certification (RPKI)**
  - NANOG49: Pradosh Mohapatra (Cisco Systems), **BGP Prefix Origin Validation**

BGP MULTIPATH

BGP MULTIPATH

- BGP Multipath allows to configure BGP to install multiple paths in the RIB for multipath load sharing.
- Load balancing is performed by CEF in per-packet basis (not recommended) or per-session basis (source and destination pair)
- Can be deployed also in MPLS network as well as with Route Reflectors

BGP MULTIPATH

- Deployment with Route Reflectors:
  - When multiple iBGP paths are installed in a routing table, a route reflector will advertise only one path.
  - If router is behind a route reflector, all routers that are connected to multihomed sites will not be advertised unless a different route distinguisher is configured for each VRF.

BGP MULTIPATH

- The BGP best-path algorithm considers the paths as equal-cost paths if the following attributes are identical:
  - Weight
  - Local preference
  - AS_path
  - Origin code
  - Multi-exit discriminator (MED)
  - IGP cost to the BGP next hop

BGP MULTIPATH

- Available on most platforms, this is not a new feature!

```
Router(config)# router bgp 40000
Router(config-router)# address-family ipv4 vrf RED
Router(config-router-af)# maximum-paths eibgp 6
Router(config-router-af)# end
```

BGP ADDITIONAL PATHS

BGP ADDITIONAL PATHS

- Default BGP behavior
  - BGP routers and Route Reflectors propagate only their best paths over their sessions.
  - Implicit Withdrawal Rule (The advertisement of a prefix replaces the previous announcement of that prefix). This approach can achieve better scaling, but at the cost of path diversity.
  - This leads to path hiding that can prevent efficient use of BGP Multipath, prevent hitless planned maintenance and lead to sub-optimal routing.

BGP ADDITIONAL PATHS

- Multiple information of prefix *p* and *x*
- PE knows only one path (best)

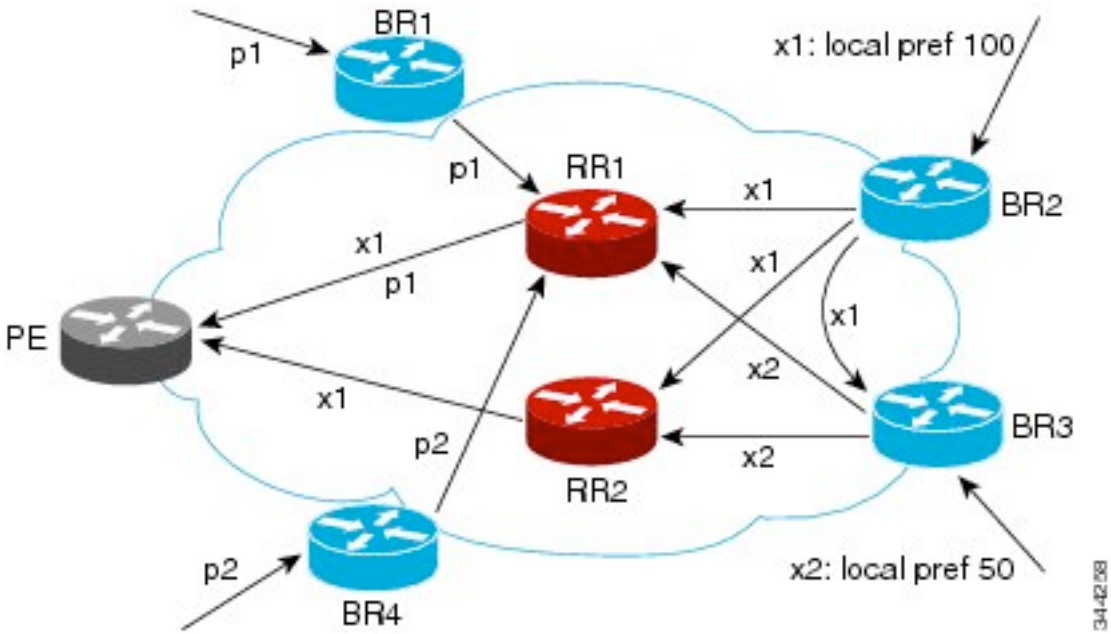

Image source: Cisco documentation

BGP ADDITIONAL PATHS

- BGP Additional Paths feature
  - Path identifier (ID) is added to each path in the NLRI
  - Path ID's are unique to a peering sessions and are generated for each network
  - Path ID us used to prevent a route announcement from implicitly withdrawing the previous one
  - It allows advertisement of more paths in addition to best path.
  - Multiple paths to same prefix can exist in BGP table

BGP ADDITIONAL PATHS

- BGP Additional Paths feature
  - Capability must me negotiated between peer routers
  - Administrator have to configure router to send, receive or both capabilities
  - Neighbors that have negotiated the capability will be grouped together in an update group and in a separate update group from those peers that have not negotiated the capability

BGP ADDITIONAL PATHS

- There are 3 path selection policies that are not mutually exclusive
  - **Best 2** or **Best 3** – best and 2nd best path are advertised, best, 2nd best and 3rd best are advertised
  - **All** – all path with unique next hops are eligible for selection
  - **Group-best** – calculates the group-best for group of prefixes. Refer to documentation for details.

BGP ADDITIONAL PATHS

```
router bgp 1
 address-family ipv4 unicast
  bgp additional-paths select all
  neighbor 192.168.1.2 additional-paths send receive
  neighbor 192.168.1.2 advertise additional-paths all
```

BGP ADDITIONAL PATHS

```
router bgp 1
 neighbor 192.168.101.15 remote-as 1
!
 address-family ipv4 unicast
  bgp additional-paths send receive
  bgp additional-paths select all best 3 group-best
  neighbor 192.168.101.15 activate
  neighbor 192.168.101.15 route-map add_path1 out
  neighbor 192.168.101.15 advertise additional-paths best 2
 exit-address-family
 !
route-map add_path1 permit 10
 match additional-paths advertise-set best 2
 set metric 780
route-map add_path1 permit 20
```

BGP ADDITIONAL PATHS

- Available on:
  - Cisco
    - IOS XR 4.0 – platforms CRS-1, ASR9000
    - IOS XE 3.7S – platforms ASR1000-RP1, ASR1000-RP2, ASR-1001, ASR1002-X, ISR4451-X, CSR100V
    - IOS 15.3(1)T – platforms 3925, 3925E, 3945, 3945E
  - Juniper
    - JunOS 11.3

BGP DYNAMIC NEIGHBORS

BGP DYNAMIC NEIGHBORS

- BGP Dynamic Neighbors feature allows BGP peering to a group of remote neighbors that are defined by a range of IP addresses.
- Each range can be configured as a subnet IP address.
- Another router initiates TCP session
- Dynamic BGP neighbor inherits configuration from peer-group, it does not require any additional configuration
- Available only for IPv4 BGP peering

BGP DYNAMIC NEIGHBORS

- In larger BGP networks implementing BGP dynamic neighbors can reduce the amount and complexity of CLI configuration
- It also saves CPU and memory usage
- Number of allowed dynamic neighbors can be defined in configuration
- Up to 5 optional AS number for listen range can be defined

BGP DYNAMIC NEIGHBORS

- Configuration

```
router bgp 45000
 bgp log-neighbor-changes
 bgp listen limit 200
 bgp listen range 172.21.0.0/16 peer-group group172
 bgp listen range 192.168.0.0/16 peer-group group192
 neighbor group172 peer-group
 neighbor group172 remote-as 45000
 neighbor group192 peer-group
 neighbor group192 remote-as 40000 alternate-as 50000
 neighbor 172.16.1.2 remote-as 45000
 address-family ipv4 unicast
  neighbor group172 activate
  neighbor group192 activate
  neighbor 172.16.1.2 activate
```

BGP DYNAMIC NEIGHBORS

- Available on:
  - Cisco
    - 12.2(33)SXH
    - 15.1(2)T
    - 15.0(1)S
    - 15.1(1)SG
    - Cisco IOS XE Release 3.1S
    - Cisco IOS XE Release 3.3SG

BGP Slow Peer

BGP Slow Peer

- BGP Update Group
  - BGP update generation uses the concept of update groups to optimize performance
  - An update group is a collection of peers with the identical outbound policy
  - Group policy is used to format messages that are then transmitted to the members of the group with every update

BGP Slow Peer

- BGP Update Group
  - Each update group is allocated a quota of formatted messages that it keeps in its cache to maintain fairness in resource utilization
  - Messages are added to the cache when they are formatted by the group, and they are removed when they are transmitted to all the members of the group.

BGP Slow Peer

- BGP Slow Peer
  - A slow peer is a peer that cannot keep up with the rate at which the Cisco IOS software is generating update messages
    - There is packet loss or high traffic on the link to the peer
    - The throughput of the BGP TCP connection is very low
    - The peer has a heavy CPU load and cannot service the TCP connection at the required frequency
  - When a slow peer is present in an update group, the number of formatted updates pending transmission builds up

BGP Slow Peer

- BGP Slow Peer
  - The rest of the members of the group that are faster than the slow peer and have completed transmission of the formatted messages will not have anything new to send, even though there may be newly modified BGP networks waiting to be advertised or withdrawn

BGP Slow Peer

- When peer is not Slow Peer
  - Events that cause large churn in the BGP can cause a brief spike in the rate of update generation.
  - A peer that temporarily falls behind during such events, but quickly recovers after the event, is not considered a slow peer.
  - **Temporary Slowness Does Not Constitute a Slow Peer**
  - In order for a peer for be marked as slow, it must be incapable of keeping up with the average rate of generated updates over a longer period

BGP Slow Peer

- BGP Slow Peer Detection
  - It relies on the timestamp on the update messages in an update group. Update messages are timestamped when they are formatted.
  - When BGP slow peer detection is configured, the timestamp of the oldest message in a peers queue is compared to the current time to determine if the peer is lagging more than the configured slow peer time threshold.
  - Peer can be marked also statically

BGP Slow Peer

- What to do with Slow Peer
  - The slow peer is moved from its normal update group to a slow update group
    - The normal update group will continue to function without being slowed down
  - Peer can be kept in slow update group until manually cleared or router can dynamically move it back to its regular update group as conditions improve

BGP Slow Peer

- Static Slow Peer

```
router bgp 5
address-family ipv4
 neighbor 192.168.12.10 slow-peer split-update-group static
```

BGP Slow Peer

- Dynamic Slow Peer

```
router bgp 13
 no bgp default route-target filter
 no bgp enforce-first-as
 bgp log-neighbor-changes
 neighbor 10.0.101.3 remote-as 13
 address-family ipv4
  neighbor 10.0.101.3 slow-peer split-update-group dynamic
```

BGP Slow Peer

- Available on:
  - Cisco
    - IOS XE 3.1S – platforms ASR1000-RP1, ASR1000-RP2, ASR-1001, ASR1002-X, ISR4451-X, CSR100V
    - IOS 15.0S – platforms 7600-SUP720/MSFC3, 7600-RSP720/MSFC4
  - Juniper
    - JunOS 11.3

Questions?

THANK YOU