AOS-CX 10.6 Update
November, 2020

aruba
a Hewlett Packard Enterprise company

BGP Confederation

Aruba Switching TME

Confederation

Agenda

1. Overview
2. Use Cases
3. Details / Caveats
4. Configuration
5. Best Practices
6. Troubleshooting
7. Demo

aruba
a Hewlett Packard
Enterprise company

Overview

Definitions

Acronyms

MP-BGP	**M**ulti-**P**rotocol **B**order **G**ateway **P**rotocol
AF	**A**ddress **F**amily (Ex: IPv4, IPv6 or EVPN address families used in MP-BGP)
AS	**A**utonomous **S**ystem : a BGP administrative domain
iBGP	**i**nternal **BGP**, refers to peering between nodes inside the same AS.
eBGP	**e**xternal **BGP**, refers to peering between nodes from different AS.
BGP RR	**R**oute-**R**eflector concept applies to iBGP only. RR propagates routes to RR-Clients
RR-Clients	**R**oute-**R**eflector Clients
AS Confederation	A collection of autonomous systems represented and advertised as a single AS number to BGP speakers that are not members of the local BGP confederation. “AS Confederation” and “Confederation” are used interchangeably in the industry.
AS Confederation Identifier	An externally visible autonomous system number that identifies a BGP confederation as a whole.
Member-AS	An autonomous system that is contained in a given AS confederation.
Member-AS Number	An autonomous system number identifier visible only within a BGP confederation, and used to represent a Member-AS within that confederation. “Member-AS” and “Member-AS Number” are used interchangeably in the industry

Overview

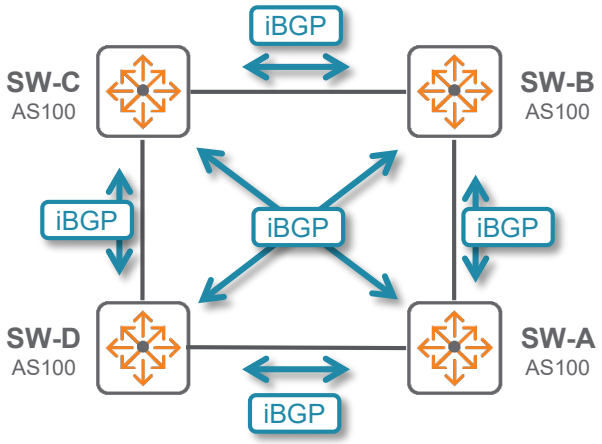
Reminder: iBGP Split Horizon Rule

- The BGP Spit Horizon Rule states that a BGP router that receives a BGP route via an iBGP peering shall not advertise that route to another iBGP Peer.
- Because of the split horizon rule, the information learnt cannot be advertised to other peers in the iBGP network. To overcome this constraint, 3 solutions are possible:
  1. implementation of a **full mesh routing topology.**
     For **n** BGP speakers within an AS, **n*(n-1)/2** unique iBGP sessions are required. Full mesh deployment **does not scale** when there are a large number of iBGP speakers within the Autonomous System.
     This method should be reserved for a very small number of iBGP nodes (preferably less or equal to 4).
  2. **Route Reflector**
     An iBGP RR advertises routes to other iBGP speakers, called route-reflector clients. No other iBGP peering is configured on the RR-client beside the one to the RRs.
  3. **Confederation**
     New in 10.6.

Reminder

iBGP Full mesh routing topology

- n*(n-1)/2 iBGP peering configuration is required. (here 6)



BGP Confederation

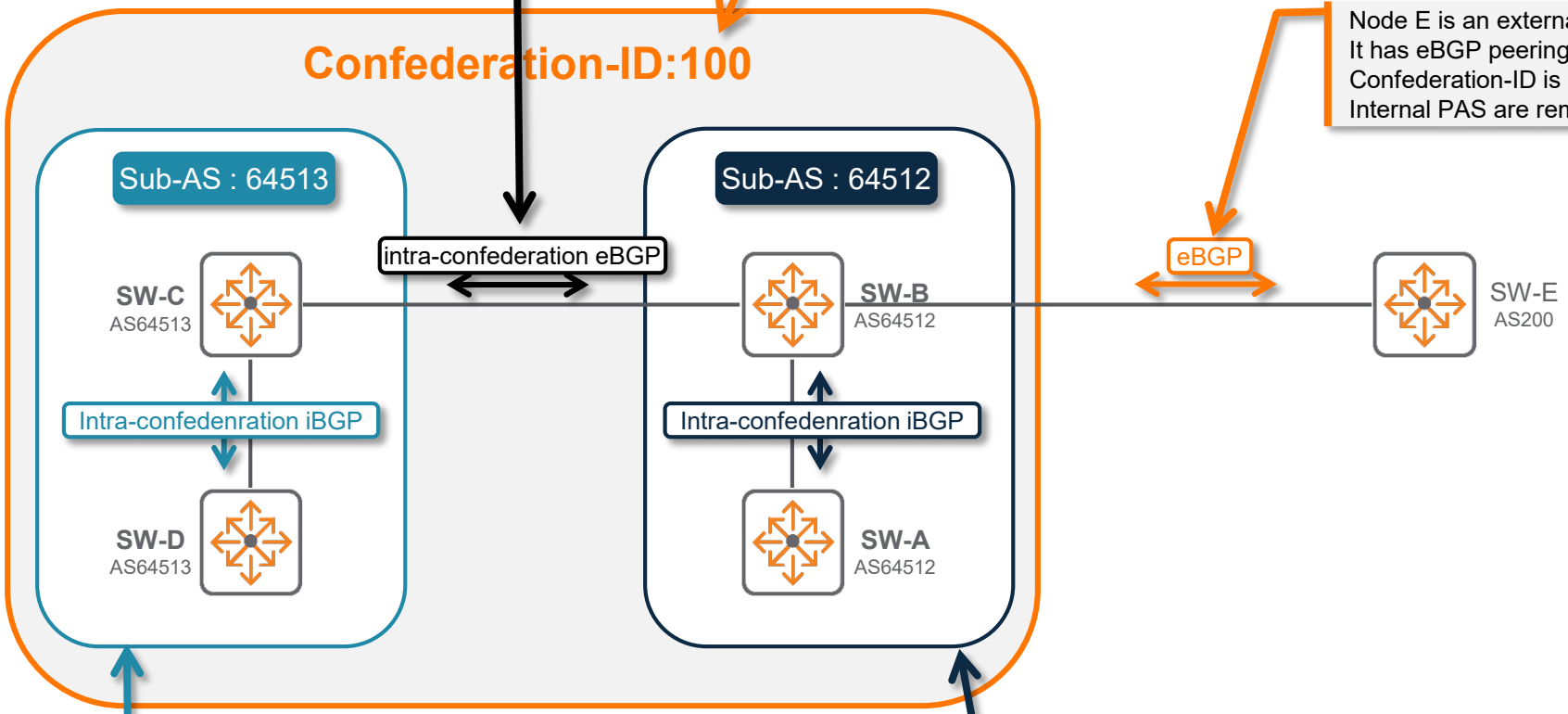
Overview

- BGP **Confederation** breaks up an Autonomous System with a very large number of BGP speakers into smaller **Sub-Autonomous Systems**.
- Each **Sub-Autonomous System** is uniquely identified within the confederation AS by a Sub-AS number also referred as **Member-AS number**.
- Typically, sub-AS numbers are taken from the **private AS numbers** range between 64512 and 65535. These sub-AS are not exposed outside the confederation.
- The connection **between Sub-AS** is always **eBGP** (Since they are identified by a unique ASN) referred as **intra-confederation eBGP**.
- Within a sub-AS the same iBGP full mesh requirement exists. (Route Reflector can be used here).
- The sub-AS numbers are removed when the route is advertised out of the confederation.
- To the outside world, **the confederation appears as a single AS**, identified by the **Confederation-ID**.

BGP Confederation

The peering between node B and C is eBGP inside confederation 100.

All the nodes inside confederation should be configured with same Confederation-ID.



Node E is an external BGP peer with AS200.
It has eBGP peering with node B with AS100.
Confederation-ID is used as remote-as on external nodes.
Internal PAS are removed.

Node C and D are iBGP peers inside Sub-AS 64513.
Sub-AS 64513 is part of confederation 100.

Node A and B are iBGP peers inside Sub-AS 64512.
Sub-AS 64512 is part of confederation 100.

aruba
a Hewlett Packard
Enterprise company

Use Cases

BGP confederation

Pros and Cons

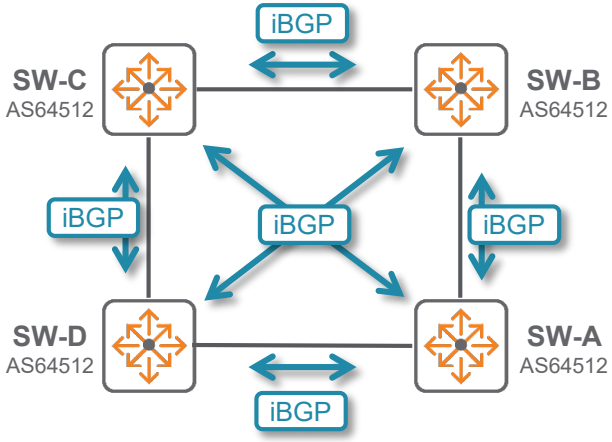
- **Pros:**
  - Suitable for large networks and more precisely when routing design is driven by the topology complexity.
    Very diverse and complex topologies will lead to use Confederation, whereas Route-reflector would be a good fit for simple hub & spoke traditional topologies.
  - No need to expose the internal topology of the divided autonomous system.
  - In comparison to full-mesh, significant reduction in the total number of intra-domain BGP connections.
  - For outside world, it looks like a single AS.
  - RR can also be deployed to address the full mesh iBGP requirement within sub-AS.

- **Cons:**
  - It may increase the complexity of routing policy based on AS_PATH information.
  - It may increases the maintenance overhead in planning and deploying multiple PAS.
  - Not suitable for small networks. It is a complex architecture and has configuration overhead.
  - **Confederation feature must be supported on all nodes inside the Confederation.**

BGP architectures

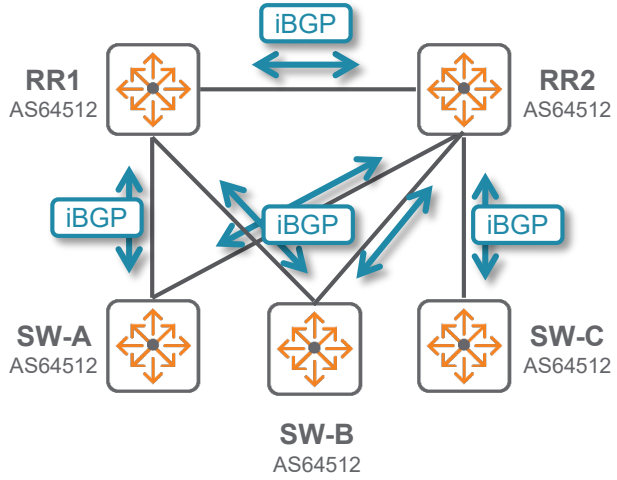
Full-Mesh

- Small network



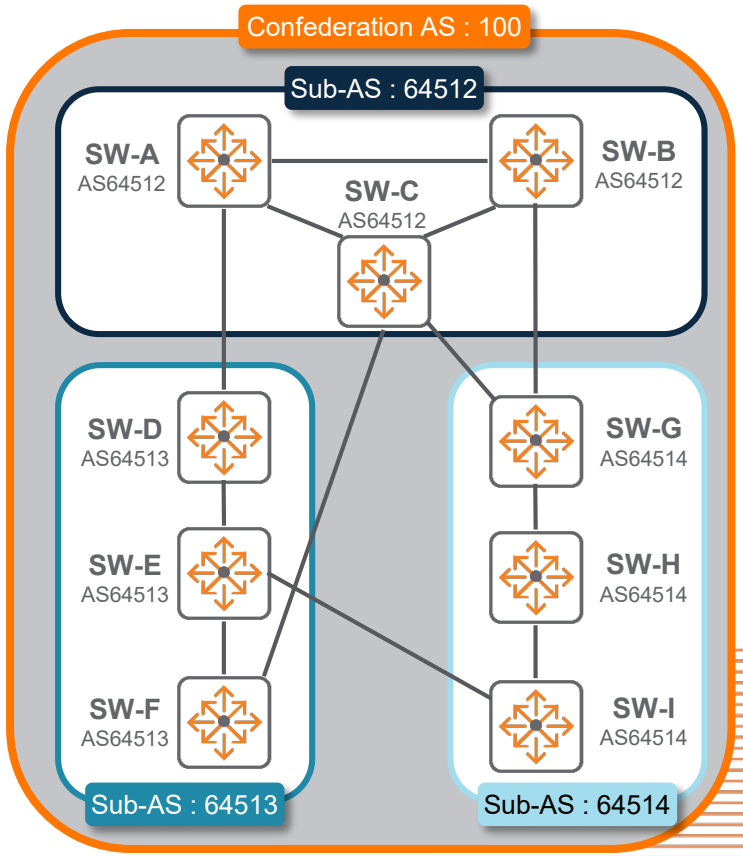
Route-Reflector

- Ideal for Hub & Spoke topology



Confederation

- Partially mesh, complex topology

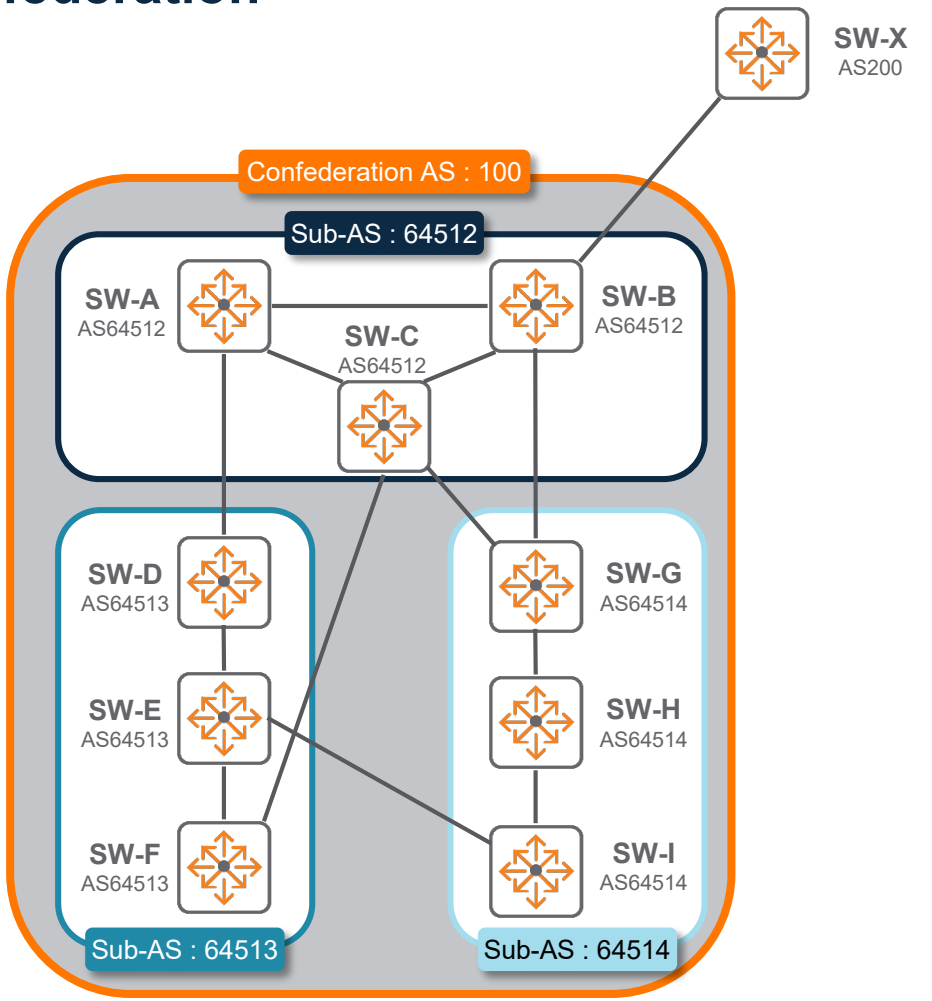


For best HA:
it is recommended that iBGP topology follows the physical topology

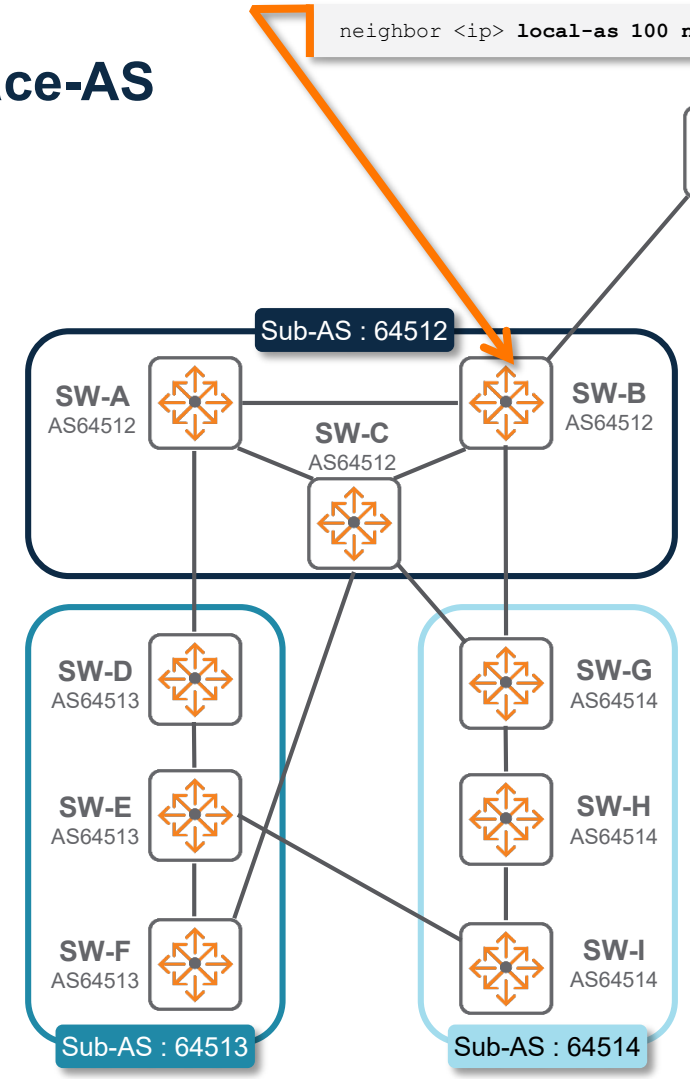
BGP architectures

Confederation or bunch of PAS with replace-AS ?

Confederation



Replace-AS



BGP architectures

Confederation or bunch of PAS with replace-AS ?

Confederation

- Confederation feature must be supported on all nodes inside the Confederation.
- Configuration is uniform and peering to external AS from any node automatically removes PAS.
- **Local-preference, MED and next-hop attributes are maintained across Sub-AS.**
- In case of AS (Confederation ID) change, all nodes must be reconfigured: disruptive change.
- It is highly recommended that all nodes of the same Confederation share the same IGP (OSPF).

Replace-AS

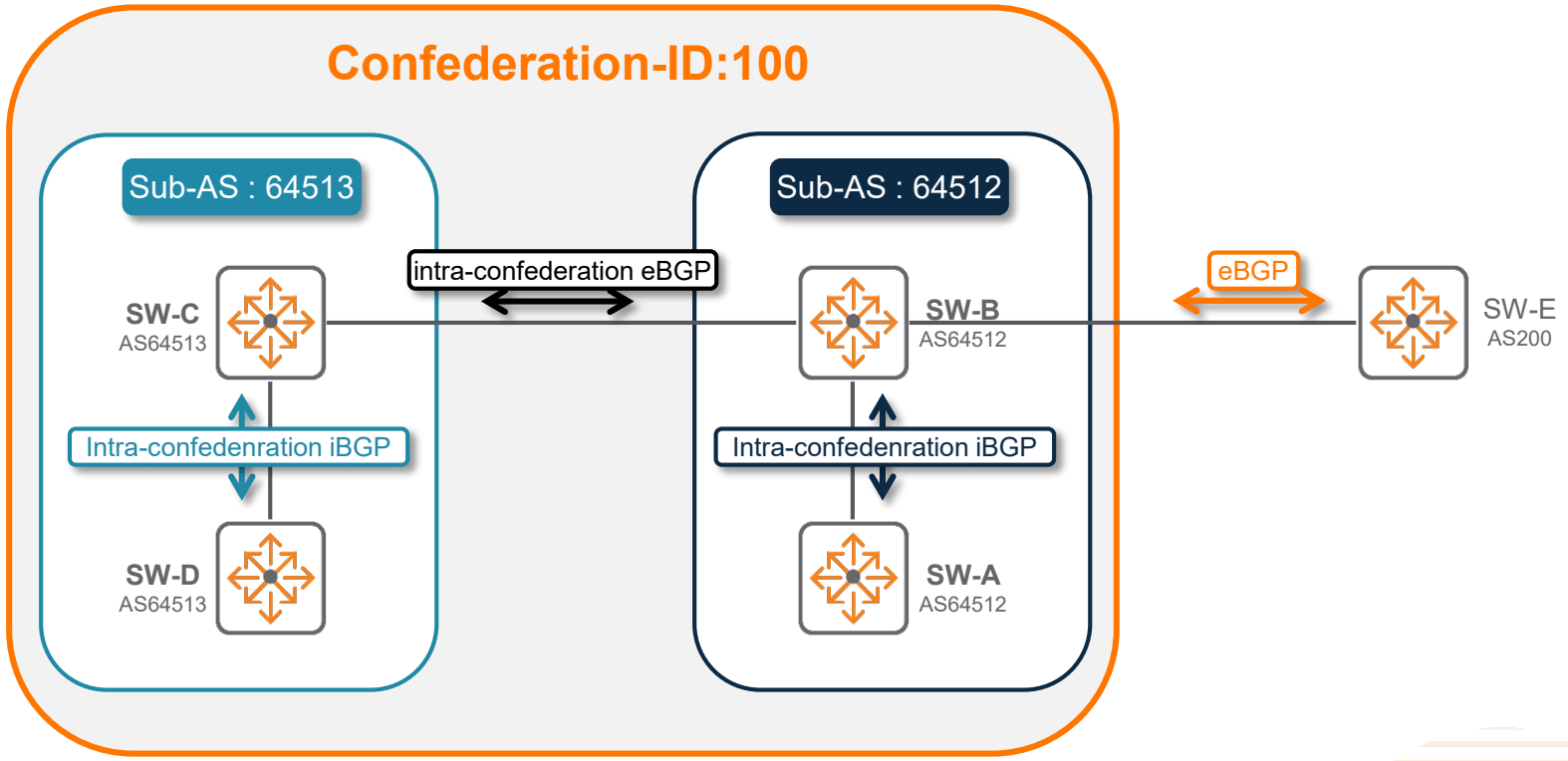
- No dependency on confederation feature being supported (more interoperable design).
- Specific configuration is required on the node peering to external AS.
- **Local-preference and next-hop attributes are reset crossing PAS.**
- In case of AS change, only the node(s) peering with external AS must be modified. No disruption inside the bunch of PAS domains.
- IGP domain is usually mapped to the PAS domain.

aruba
a Hewlett Packard
Enterprise company

Details / Caveats

BGP Confederation

- iBGP network is divided into two Sub-AS 64512 and 64513
- Both the Sub-AS are part of a confederation-100
- No change in the external BGP configuration.
- Only 3 connections required within confederation 100.



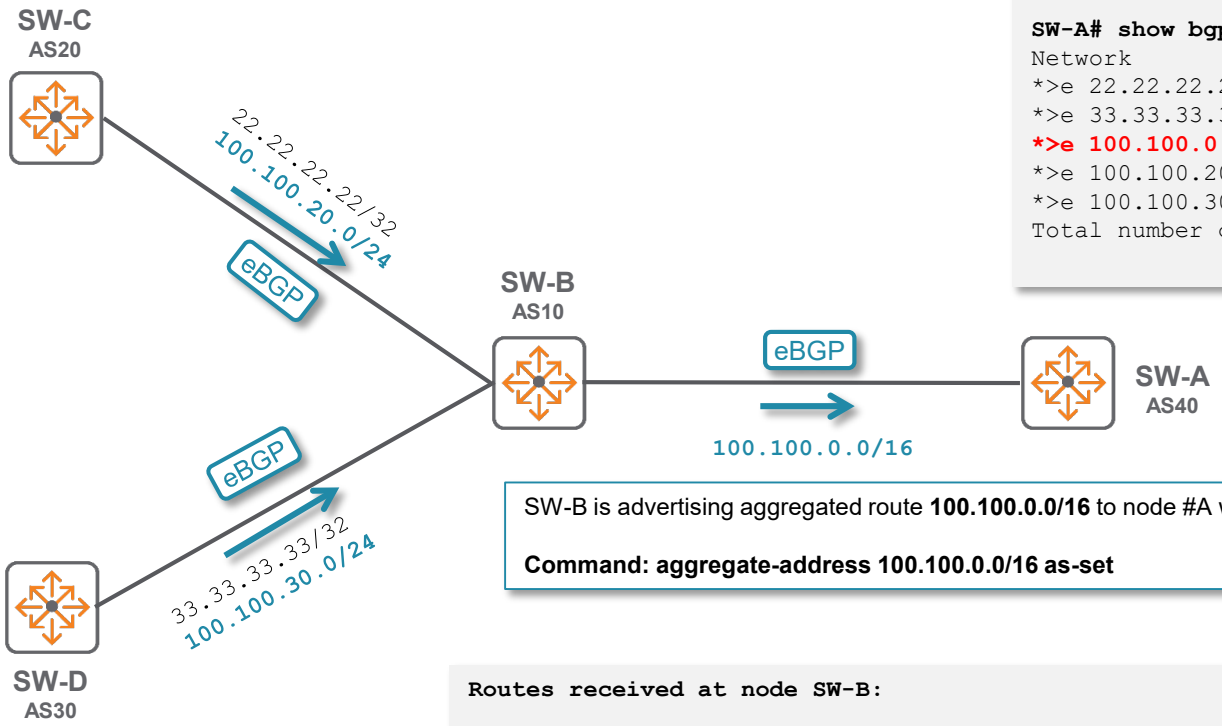
BGP Confederation

AS_PATH attribute in Confederation

- AS_PATH is a well-known mandatory attribute. This attribute identifies the autonomous systems through which routing information carried in the UPDATE message has passed.
- AS_PATH attribute is composed of a sequence of AS path segments. (32 AS-paths max in 10.6)
- Each AS path segment is represented by a TLV :
  - <**path segment type**, **path segment length**, **path segment value**>.
- BGP defines the below path segment types:
  - **AS_SET**: unordered set of autonomous systems that a route in the UPDATE message has traversed.
  - **AS_SEQUENCE**: ordered set of autonomous systems that a route in the UPDATE message has traversed.
- **Two additional segment types are added for BGP Confederations:**
  - **AS_CONFED_SEQUENC**E: ordered set of Member Autonomous Systems in the local confederation that the UPDATE message has traversed.
  - **AS_CONFED_SET**: unordered set of Member Autonomous Systems in the local confederation that the UPDATE message has traversed.

AS_SET and AS_SEQUENCE

AS_CONFED_SET and AS_CONFED_SEQUENCE are used inside confederation



Aggregated route 100.100.0.0/16 at node SW-A with as-set:

SW-A# show bgp all

Network	Nexthop	Metric	LocPrf	Weight	Path
*>e 22.22.22.22/32	10.10.10.2	0	100	0	10 20 i
*>e 33.33.33.33/32	10.10.10.2	0	100	0	10 30 i
***>e 100.100.0.0/16**	**10.10.10.2**	**0**	**100**	**0**	**10 (20,30) i**
*>e 100.100.20.0/24	10.10.10.2	0	100	0	10 20 i
*>e 100.100.30.0/24	10.10.10.2	0	100	0	10 30 i

Total number of entries 5

Aggregated route is received with as-set (20,30).
Other routes have as-sequence.

SW-B is advertising aggregated route **100.100.0.0/16** to node #A with as-set.

Command: aggregate-address 100.100.0.0/16 as-set

Routes received at node SW-B:

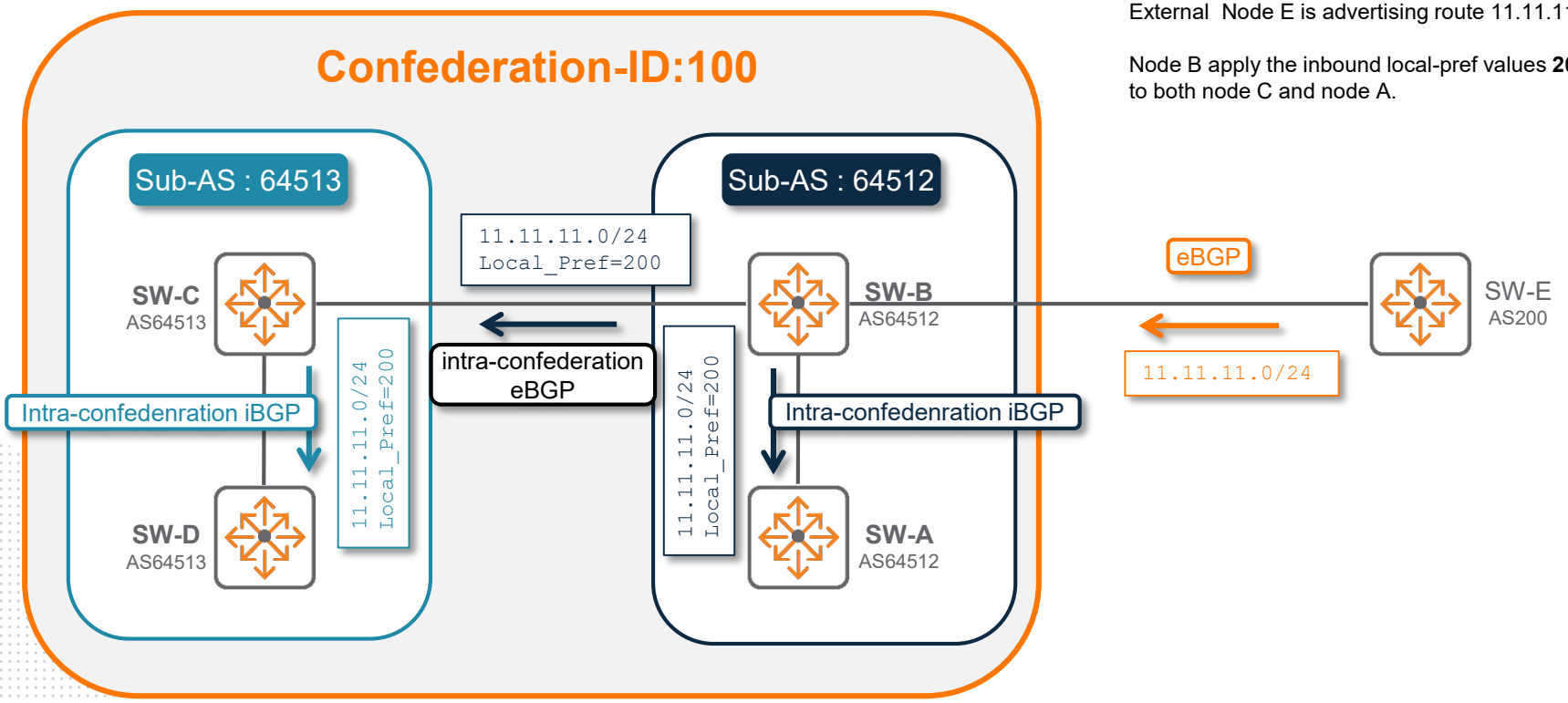
SW-B# sh bgp all

Network	Nexthop	Metric	LocPrf	Weight	Path
*>e 22.22.22.22/32	20.20.20.1	0	100	0	20 i
*>e 33.33.33.33/32	30.30.30.1	0	100	0	30 i
*>e 100.100.20.0/24	20.20.20.1	0	100	0	20 i
*>e 100.100.30.0/24	30.30.30.1	0	100	0	30 i

Total number of entries 4

Local Preference value inside confederation

- The LOCAL_PREF path attribute is always advertised to iBGP peers **and to intra-confederation eBGP peers**.
- It is never advertised to eBGP peers external to the confederation.
- Local-preference attribute is maintained across Sub-AS in the confederations which helps in routing design for best path calculation.



Node A, B, C, and D are part of confederation 100.
External Node E is advertising route 11.11.11.0/24 to B.

Node B apply the inbound local-pref values **200** which will be advertised to both node C and node A.

BGP Confederation

Caveats

- Confederation applies to BGP global configuration: it applies to all AFs and all VRFs, similar to the AS number given to the BGP process
- Not tested for EVPN AF (only IPv4 and IPv6 AF were tested).
- Dotted format (4-byte AS) for Confederation ID is supported as well as for Member-AS number.
- Any confederation ID change will trigger BGP session reset.
- All nodes in the Confederation MUST be configured with the same Confederation ID. If a node is misconfigured with a different Confederation number, the BGP session will not establish.
- Standard BGP scale applies:
  - Number of PAS (1024)
  - AS-path length (32)

aruba
a Hewlett Packard
Enterprise company

Configuration

Feature/Solution configuration

- Configure confederation identifier. This will be externally visible ASN for the current Autonomous system.

```
switch(config-bgp)# bgp confederation ?
  <1-4294967295>      Set the identifier for the confederation.
  peers               Peers for this sub AS.
switch(config-bgp)# bgp confederation 100
```

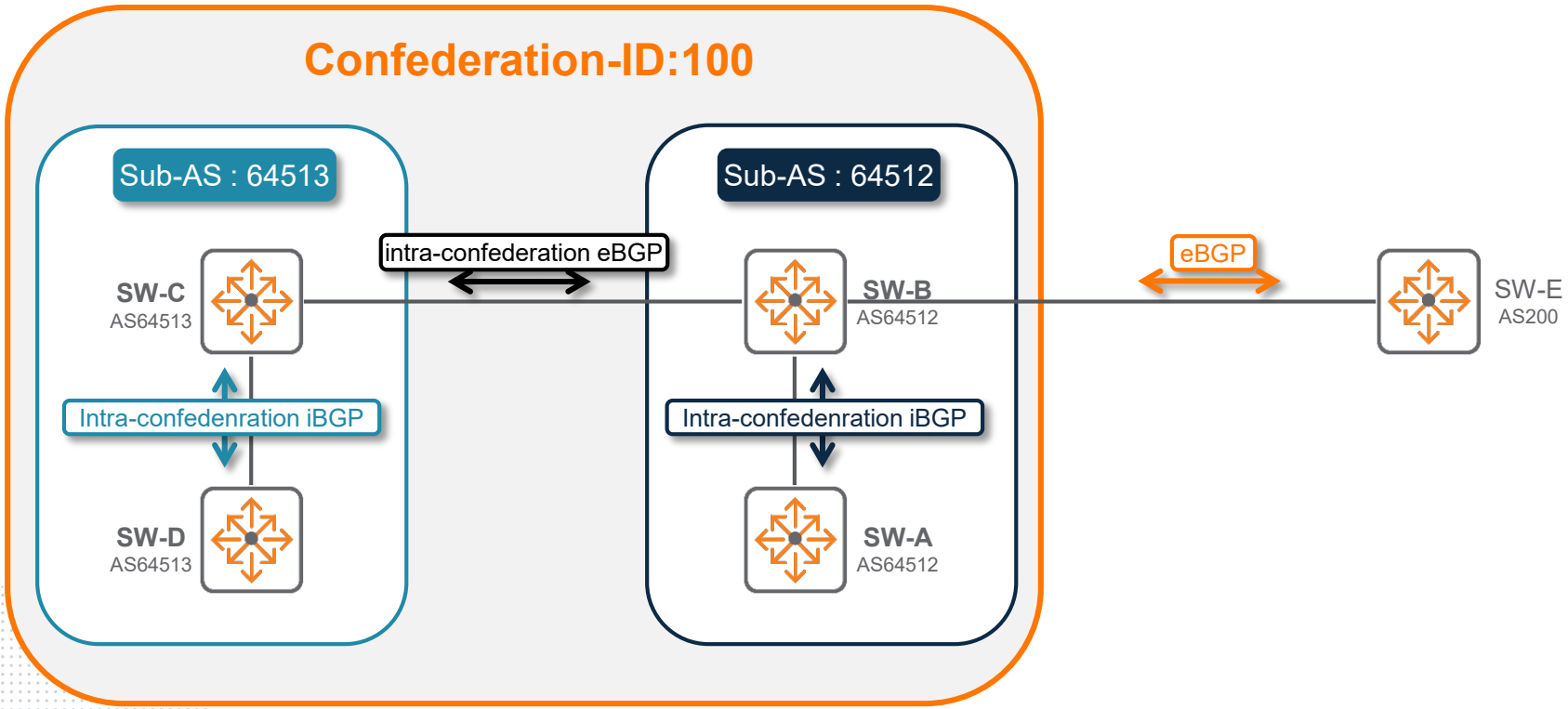
- Configure confederation peers. Configure sub ASes within the same confederation to establish an eBGP/iBGP membership inside bgp confederations.

```
switch(config-bgp)# bgp confederation peers ?
  WORD              The range for confederation peer number as an integer value between <64512-65535>
  <cr>
switch(config-bgp)# bgp confederation peers 64512 64513
```

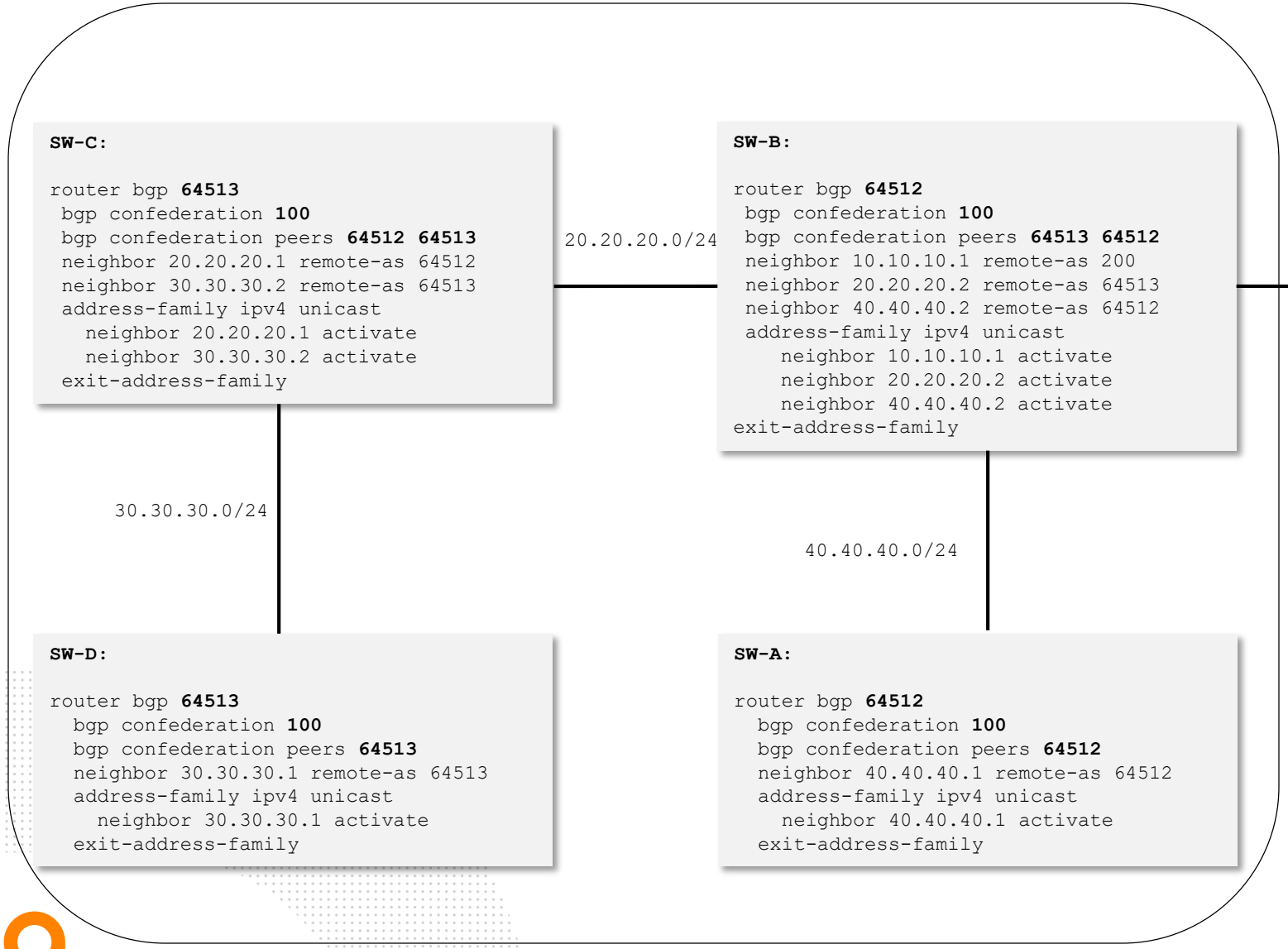
- Configure bgp bestpath med confed: To compare identical routes received from different confederation peers during the best path selection process and to select the route with the lowest MED value as the best path

```
switch(config-bgp)# bgp bestpath med ?
  confed  Configure MED comparison among paths from confederation peers
(config-bgp)# bgp bestpath med confed
```

A simple BGP Confederation topology with config



Configuration of all the nodes:



Best-path inside Confederation

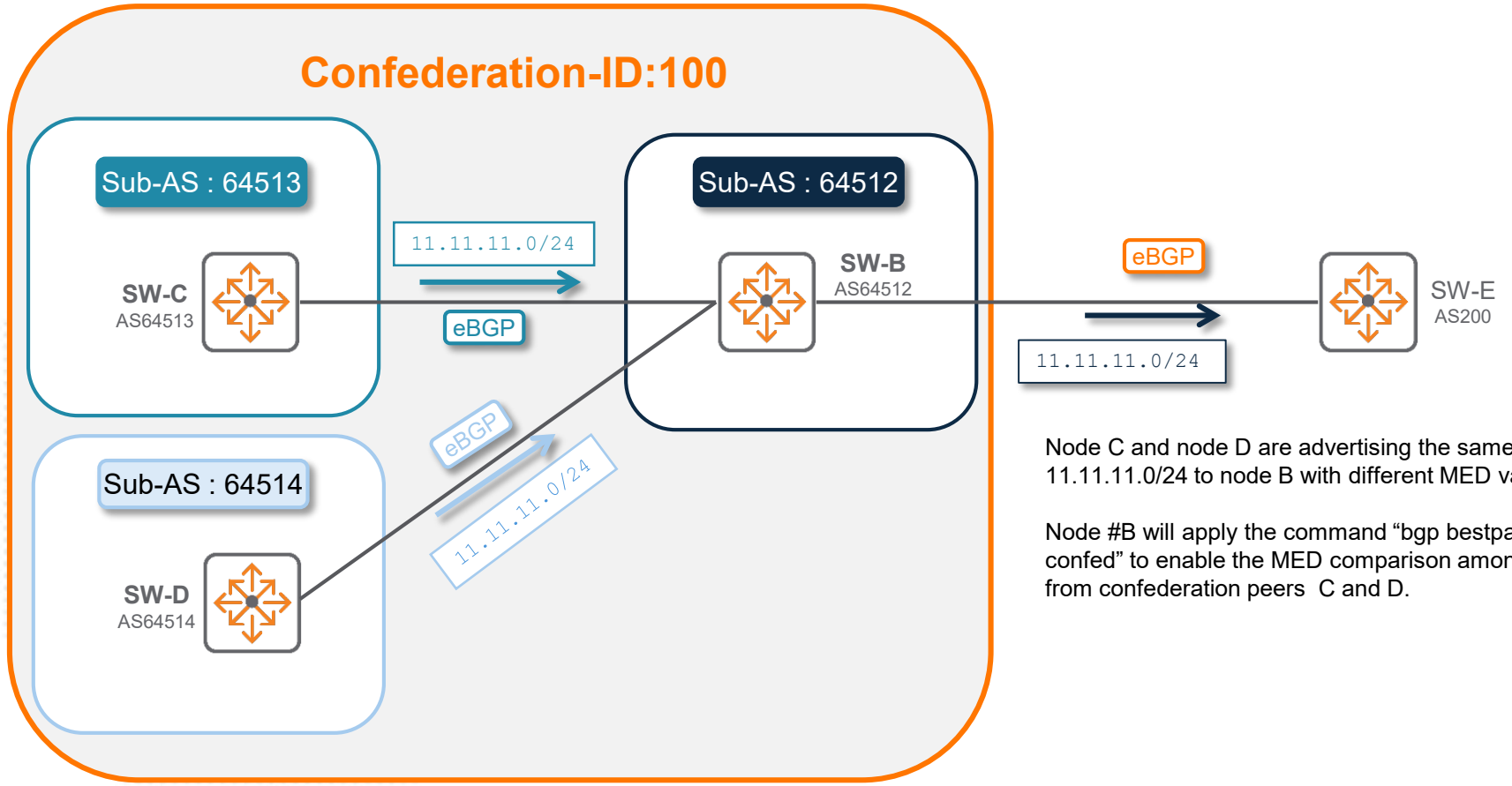
Influencing best path calculation inside confederation

- By default MED values are only compared among routes advertised by the same AS.
  The same applies inside confederation.

- "**bgp bestpath med confed**" command is used for forced evaluation of the MED attributes inside confederation.
  - Analogy with **always-compare-med** command use for eBGP routes coming from different ASN.
  - MED values are compared among routes from multiple BGP confederation peers.
  - MED values inside confederation are compared only if no external ASN (to the confed) are present in the AS path.
  - BGP will prefer route with lowest MED.
  - Applying configuration will reset all the peers. The behavior is identical as other bgp bestpath commands (like bgp deterministic-med, always-compare-med).
  - Even if this command applies to the local node, it is strongly recommended that all BGP nodes in a given Confederation share the same configured setting for "bgp bestpath med confed". (to avoid potential routing loop due to different route selection algorithm).

BGP bestpath med confed topology

All the nodes are configured with the same Confederation Identifier but with different sub-as.



Node C and node D are advertising the same route 11.11.11.0/24 to node B with different MED values.

Node #B will apply the command “bgp bestpath med confed” to enable the MED comparison among paths from confederation peers C and D.

aruba
a Hewlett Packard
Enterprise company

Best Practices

Feature/Solution Best Practices

- Private AS number range <64512-65535> should be used for sub-ASes.
- Should be used only when iBGP network has large number of speakers.
- RR can be deployed to address the full mesh iBGP requirement within sub-AS.

aruba
a Hewlett Packard
Enterprise company

Troubleshooting

Confederation Troubleshooting

Configuration:

```
router bgp 64512
    bgp confederation 100
    bgp confederation peers 64513
    bgp bestpath med confed
    neighbor 10.10.10.2 remote-as 64513
    address-family ipv4 unicast
        neighbor 10.10.10.2 activate
    exit-address-family
```

Check the configured Confederation ID:

```
switch# sh bgp all summary
VRF : default
BGP Summary
-----------
 Local AS               : 64512        BGP Router Identifier  : 10.10.10.1
 Peers                  : 0            Log Neighbor Changes   : No
 Cfg. Hold Time         : 180          Cfg. Keep Alive        : 60
 Confederation Id       : 100
```

```
R4# show bgp all
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

VRF : default
Local Router-ID 100.100.20.1

Address-family : IPv4 Unicast
------------------------------
    Network           Nexthop         Metric     LocPrf     Weight Path
*>e 1.1.1.1/32        10.10.10.1      0          100        0       [64512],200 i
```

Check the configured Confederation peers:

```
Switch# show bgp all neighbors

BGP Neighbor 10.10.10.2 (External)
    Description          :
    Peer-group           :

    Remote Router Id     : 0.0.0.0          Local Router Id     : 10.10.10.1
    Remote AS            : 64513            Local AS            : 64512
    Remote Port          : 0                Local Port          : 0
    State                : Idle             Admin Status        : Up
    Conn. Established    : 0                Conn. Dropped       : 0
    Passive              : No               Update-Source       :
    Cfg. Hold Time       : 180              Cfg. Keep Alive     : 60
    Neg. Hold Time       : 0                Neg. Keep Alive     : 0
    Up/Down Time         : 00h:00m:00s      Alt. Local-AS       : 0
    Local-AS Prepend     : No
    BFD                  : Disabled
    Password             :
    Last Err Sent        : No Error
    Last SubErr Sent     : No Error
    Last Err Rcvd        : No Error
    Last SubErr Rcvd     : No Error

    Graceful-Restart     : Enabled          Gr. Restart Time    : 120
    Gr. Stalepath Time   : 300              Remove Private-AS   : No
    TTL                  : 1                Local Cluster-ID    :
    Weight               : 0                Fall-over           : No
    Confederation-Peers  : Yes
```

Confederation Troubleshooting

AS Numbers in the show bgp route command

```
R4# show bgp all
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

VRF : default
Local Router-ID 100.100.20.1

Address-family : IPv4 Unicast
------------------------------
    Network           Nexthop         Metric     LocPrf     Weight Path
*>e 1.1.1.1/32        10.10.10.1      0          100        0      [64512],200 i
```

This is the internal AS numbers that this routes has traversed inside confederation.

This is the external AS number where the route is originated.

Confederation Troubleshooting

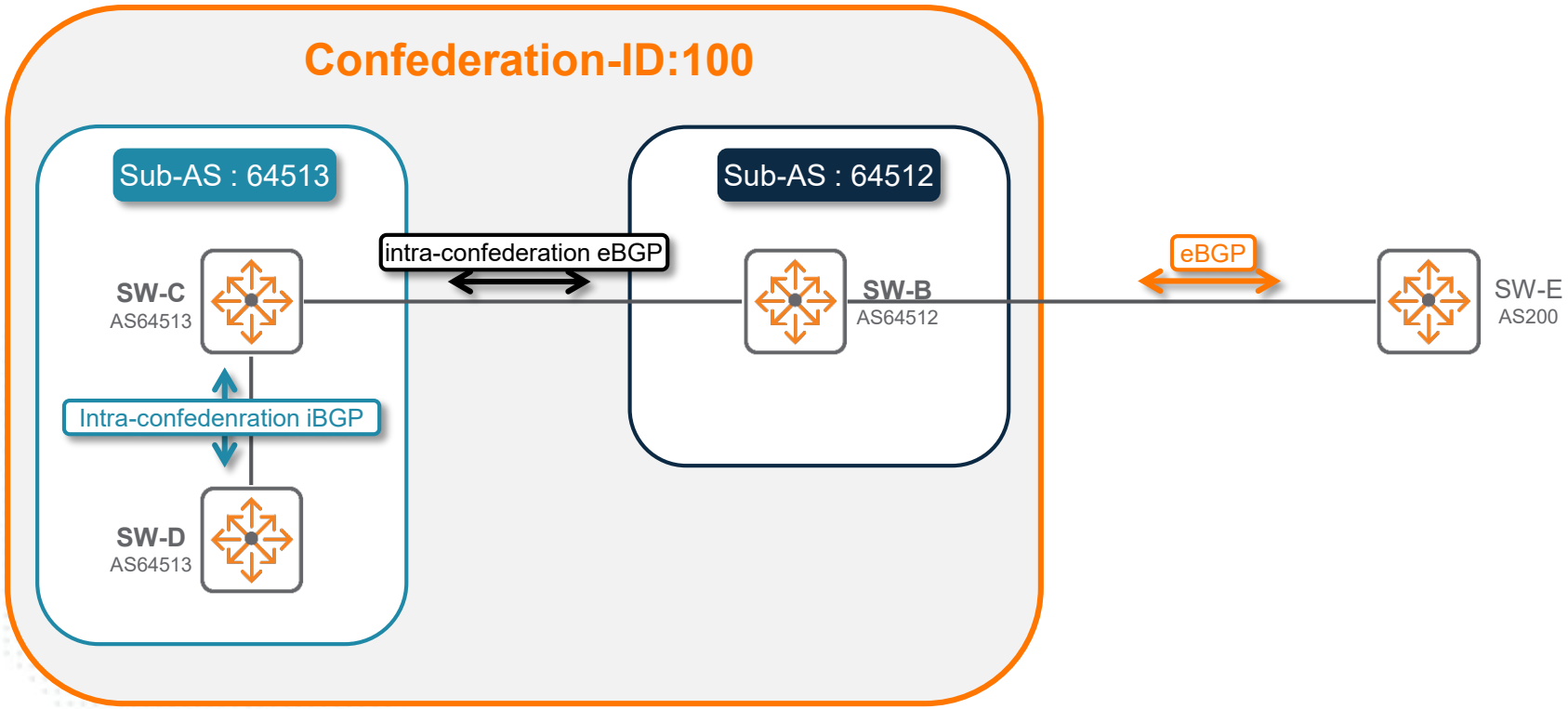
MIB Dumps and Logs

- Below MIB tables should be verified for the configured Confederation values. These table are captured as part of “diag bgp dump mib” command.
  - python /etc/mib.py get localhost bgpPeerTable
  - python /etc/mib.py get localhost bgpRmEntTable

- Fastlogs: ovs-appctl -t hpe-routing fastlog show bgp_dump
  - BGP Confederation ID is set to: **100**
  - BGP Confederation-peers set to **true** for**: 30.30.30.1**
  - BGP MED Confed configured/unconfigured

Demo

Demo BGP Confederation topology with config



Demo BGP Confederation topology with config

Confederation node D
==========

```
hostname D
interface 1/1/1
    no shutdown
    ip address 30.30.30.2/24
    ip ospf 1 area 0.0.0.0
interface loopback 1
    ip address 3.3.3.3/32
    ip ospf 1 area 0.0.0.0
!
!
!
!
!
router ospf 1
    area 0.0.0.0
router bgp 64513
    bgp confederation 100
    bgp confederation peers 64513
    neighbor 2.2.2.2 remote-as 64513
    neighbor 2.2.2.2 update-source loopback
1
    address-family ipv4 unicast
        neighbor 2.2.2.2 activate
        redistribute connected
        network 3.3.3.3/32
    exit-address-family
!
https-server vrf mgmt!
```

Confederation node C
==========

```
hostname C
interface 1/1/1
    no shutdown
    ip address 20.20.20.2/24
    ip ospf 1 area 0.0.0.0
interface 1/1/2
    no shutdown
    ip address 30.30.30.1/24
    ip ospf 1 area 0.0.0.0
interface loopback 1
    ip address 2.2.2.2/32
    ip ospf 1 area 0.0.0.0
!
!
router ospf 1
    area 0.0.0.0
router bgp 64513
    bgp confederation 100
    bgp confederation peers 64512 64513
    neighbor 1.1.1.1 remote-as 64512
    neighbor 1.1.1.1 update-source
loopback 1
    neighbor 3.3.3.3 remote-as 64513
    neighbor 3.3.3.3 update-source
loopback 1
    address-family ipv4 unicast
        neighbor 1.1.1.1 activate
        neighbor 3.3.3.3 activate
        redistribute connected
    exit-address-family
!


!
```

Confederation node B
==========

```
hostname B
interface 1/1/1
    no shutdown
    ip address 10.10.10.2/24
    ip ospf 1 area 0.0.0.0
interface 1/1/2
    no shutdown
    ip address 20.20.20.1/24
    ip ospf 1 area 0.0.0.0
interface loopback 1
    ip address 1.1.1.1/32
    ip ospf 1 area 0.0.0.0
!
!
route-map rm1 permit seq 10
     set local-preference 200
!
router ospf 1
    area 0.0.0.0
router bgp 64512
    bgp confederation 100
    bgp confederation peers 64513
    neighbor 2.2.2.2 remote-as 64513
    neighbor 2.2.2.2 update-source
loopback 1
    neighbor 10.10.10.1 remote-as
200
    address-family ipv4 unicast
        neighbor 2.2.2.2 activate
        neighbor 10.10.10.1 activate
        neighbor 10.10.10.1 route-
map rm1 in
        redistribute connected
    exit-address-family
!
```

External node E
==========

```
hostname E
!
ssh server vrf mgmt
vlan 1
interface mgmt
    no shutdown
    ip dhcp
interface 1/1/1
    no shutdown
    ip address 10.10.10.1/24
interface loopback 1
    ip address 11.11.11.11/24
!
!
router bgp 200
    neighbor 10.10.10.2 remote-as
100
    address-family ipv4 unicast
        neighbor 10.10.10.2
activate
        network 11.11.11.0/24
    exit-address-family
!
```

aruba
a Hewlett Packard Enterprise company

Thank you

vincent.giles@hpe.com