You make **possible**

How to design networks for transport-optimized platforms (NCS5500/NCS5xx) leveraging new advanced technologies

Guillermo Trueba – Technical Solutions Architect

BRKSPG-3292

Barcelona | January 27-31, 2020

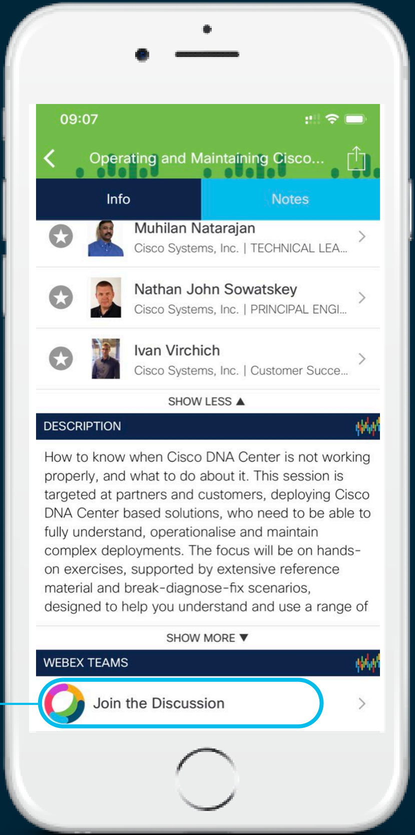
Cisco Webex Teams

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



09:07

Operating and Maintaining Cisco...

Muhilan Natarajan
Cisco Systems, Inc. | TECHNICAL LEA...

Nathan John Sowatskey
Cisco Systems, Inc. | PRINCIPAL ENGI...

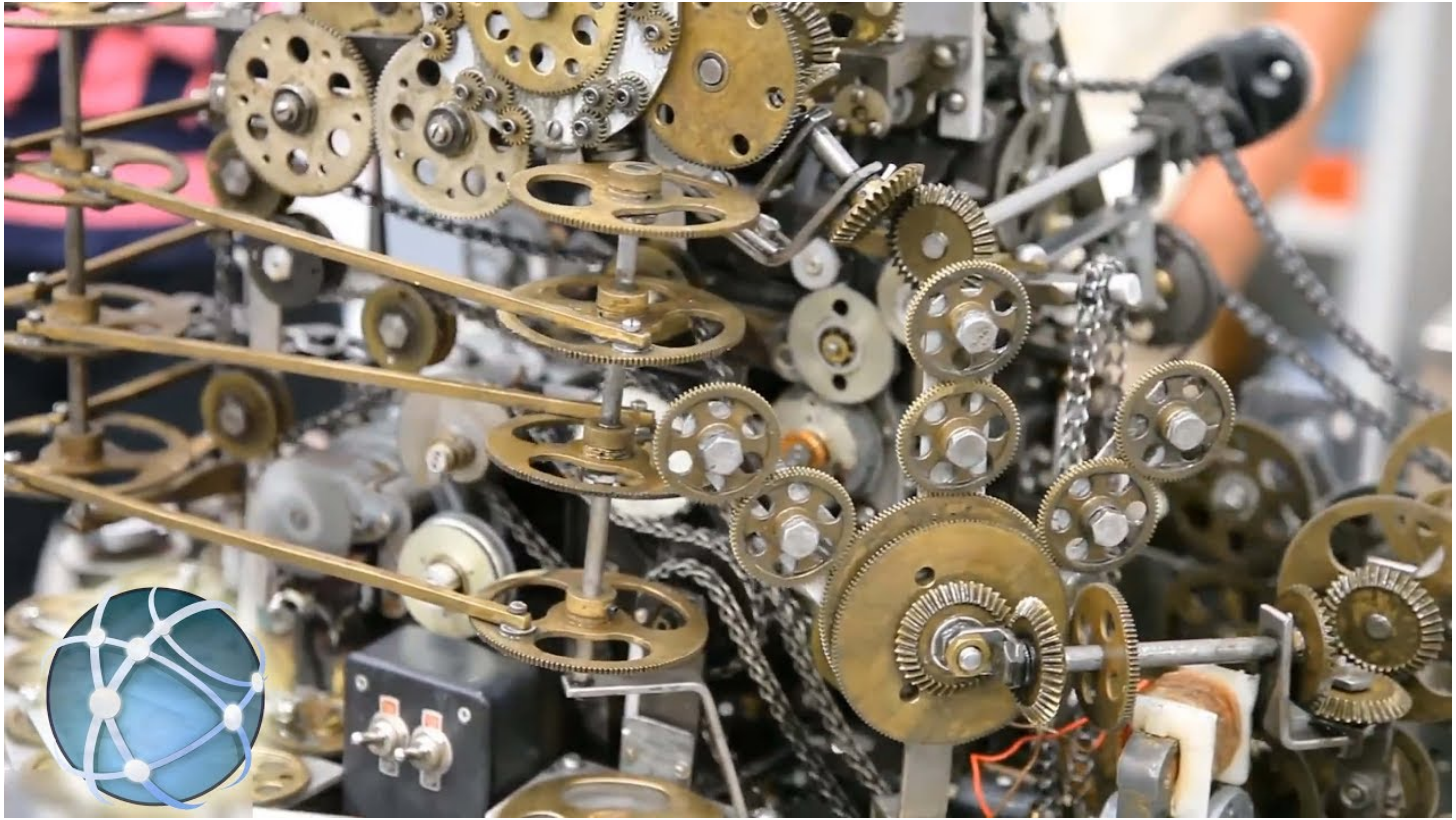
Ivan Virchich
Cisco Systems, Inc. | Customer Succe...

DESCRIPTION

How to know when Cisco DNA Center is not working properly, and what to do about it. This session is targeted at partners and customers, deploying Cisco DNA Center based solutions, who need to be able to fully understand, operationalise and maintain complex deployments. The focus will be on hands-on exercises, supported by extensive reference material and break-diagnose-fix scenarios, designed to help you understand and use a range of

WEBEX TEAMS

NCS inside?



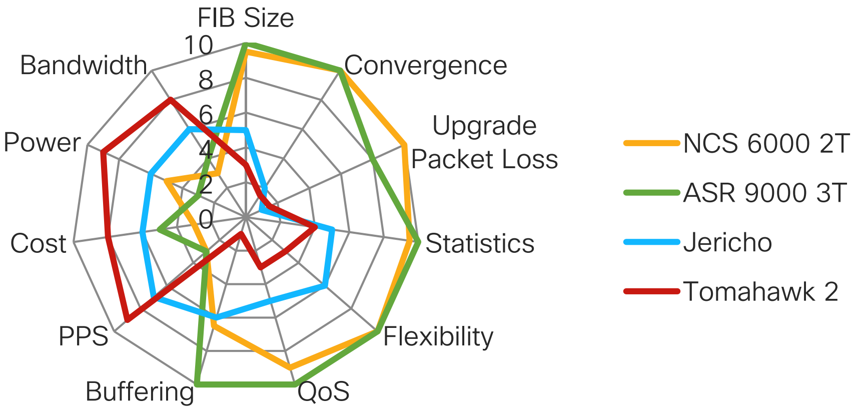
AGENDA

- Broadcom: Setting the Expectations
- Optimize MPLS resource availability with SR
- First hop L2 and L3 redundancy with EVPN
- Centralization of L3 services
- Relevant architectures for BNG transport
- Conclusions

Broadcom: Setting the Expectations

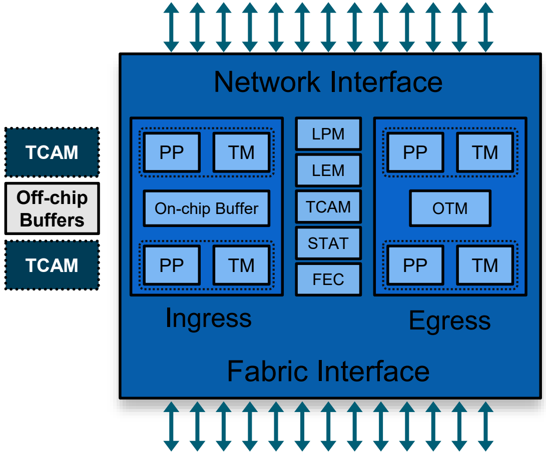
Custom vs Merchant

- Merchant optics
  - QSFP28 meets all requirements
- Custom and merchant fabrics
  - Ethernet and cell-based fabrics
- Merchant forwarding processors today
  - High pps/bw & low flexibility/buffers available from Broadcom (XGS line – 3.2T)
  - Medium pps/bw/flexibility/stats & deep buffers available from Broadcom (DNX line – 900G)
  - Low pps/bw & high features/FIB/buffers/flexibility from EZChip (NP-5c)



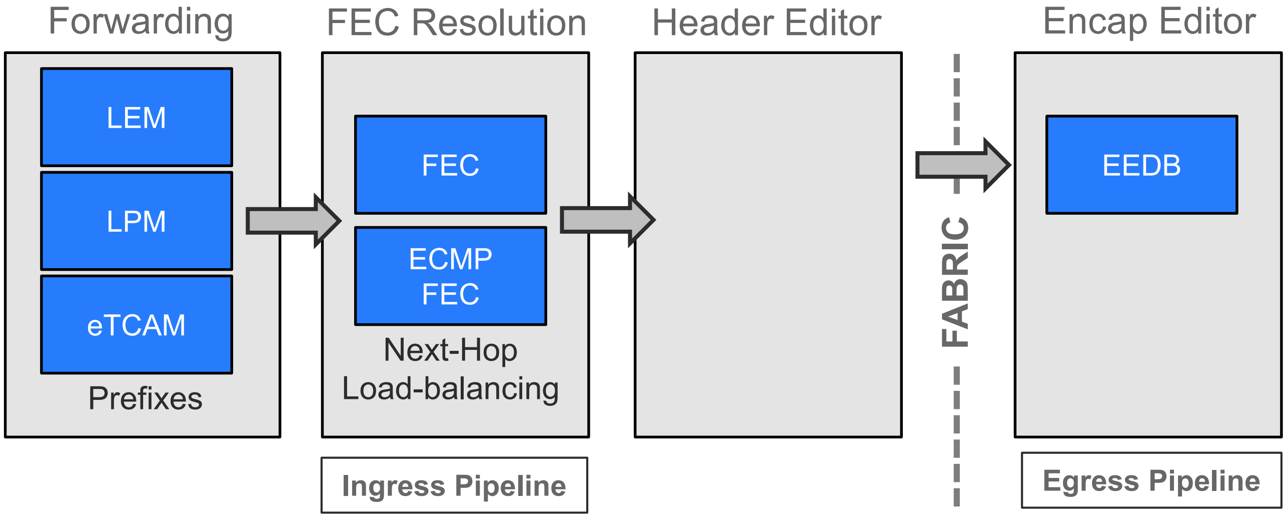
NCS 5500 Forwarding ASIC Detail
Jericho+ ASIC (BCM88680)

- Integrated Forwarding and Fabric Interface
  - 28 nm @835 Mhz – one packet per clock cycle
- Two packet processing cores (PP)
- 900G/835 Mpps ASIC
- On-chip resources
  - Small internal buffers (16MB) & iTCAM
  - Route table memory (up 1M LPM entries)
- Expansion via off-chip resources
  - Deep GDDR5 packet buffers external packet buffers
  - Optional eTCAMs for route/ACL scale (4M+ prefixes)
- Ingress/Egress Traffic Managers
  - 96k Virtual Output Queues



NCS 5500 CEF: what resources to monitor

- Prefix lookup points to FEC Entry
- FEC Entry contains Egress Interface and pointer to EEDB (encapsulation entry)
- EEDB indicates the encapsulation for the packet (ARP,GRE, MPLS,…)



NCS 5500 & NCS 500 Forwarding ASICs

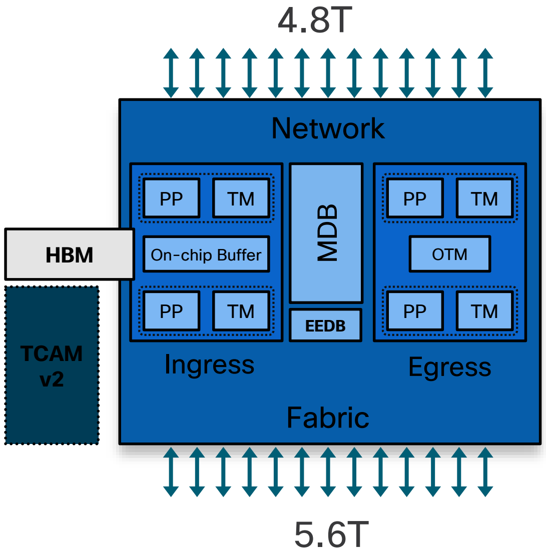
	NCS5502 Jericho 88675	NCS560/5501 Qumran-MX 88375	NCS55A1-24H NCS55A1-48Q6H Jericho+ 88680	NCS5500 Jericho+ 88681,88683	NCS540 Qumran-AX 88470
ASIC technology	28nm and 25G SerDes				
Packets / Second	720 Mpps		835 Mpps		300 Mpps
Network interface	720G	800G	900G		640G
Fabric interface	900G	N/A	1200G		N/A
LPM/KAPS	256K v4 or 64K v6		1M v4 or 256K v6	256K v4 or 64K v6	128K v4 or 32K v6
LEM	750K				250K
External TCAM	2M IPv4		3M to 4M IPv4		N/A
EEDB Entries	96K		112K		88K
FEC	128K				64K
ECMP-FEC	4K				
ISEM/ESEM	64K				32K
Statistics	256K				64K

NCS 5500 Forwarding ASIC Detail

Jericho2 ASIC (BCM88690)

- 16 nm @ 1GHz per core
- 2BPPS packet forwarding
- 4.8 Tbps packets forwarding
- 5.6 Tbps fabric bandwidth
  - 53.125 Gbps SERDES
- 8GB HBM shared between cores
- 32MB OCB (16MB assigned for each core
- eTCAM

  OP2 (4M+ extra v4 pfx + stats)

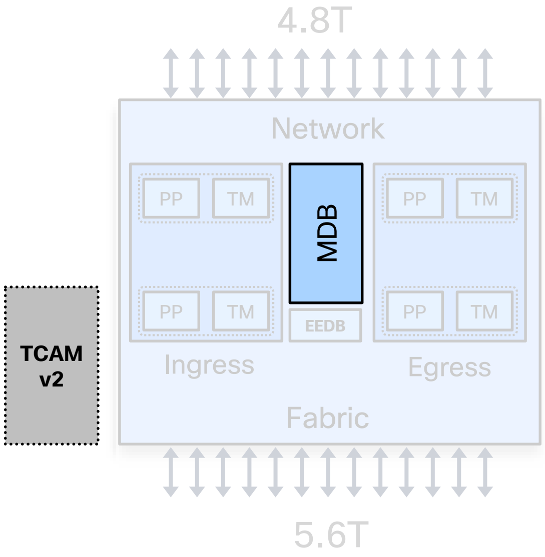


NCS 5500 Jericho2

	NCS5500 Jericho 88675	NCS560/5501 Qumran-MX 88375	NCS55A1-24H Jericho+ 88680	NCS5500 Jericho+ 88681,88683	NCS5500 Jericho2 BCM88690
ASIC technology	28nm and 25G SerDes				
Packets / Second	720 Mpps		835 Mpps		2Bpps
Network interface	720G	800G	900G		4.8Tbps
Fabric interface	900G	N/A	1200G		5.6Tbps
LPM/KAPS	256K v4 or 64K v6		1M v4 or 256K v6	256K v4 or 64K v6	1.8M v4 or 900K v6
LEM	750K				900K
External TCAM	2M IPv4		3M to 4M IPv4		3M to 4M IPv4
EEDB Entries	96K		112K		384K
FEC	128K				378K
ECMP-FEC	4K				32K
ISEM/ESEM	64K				112K
Statistics	256K				384K

NCS 5500 Jericho2 Allocation Profiles with J2 native mode

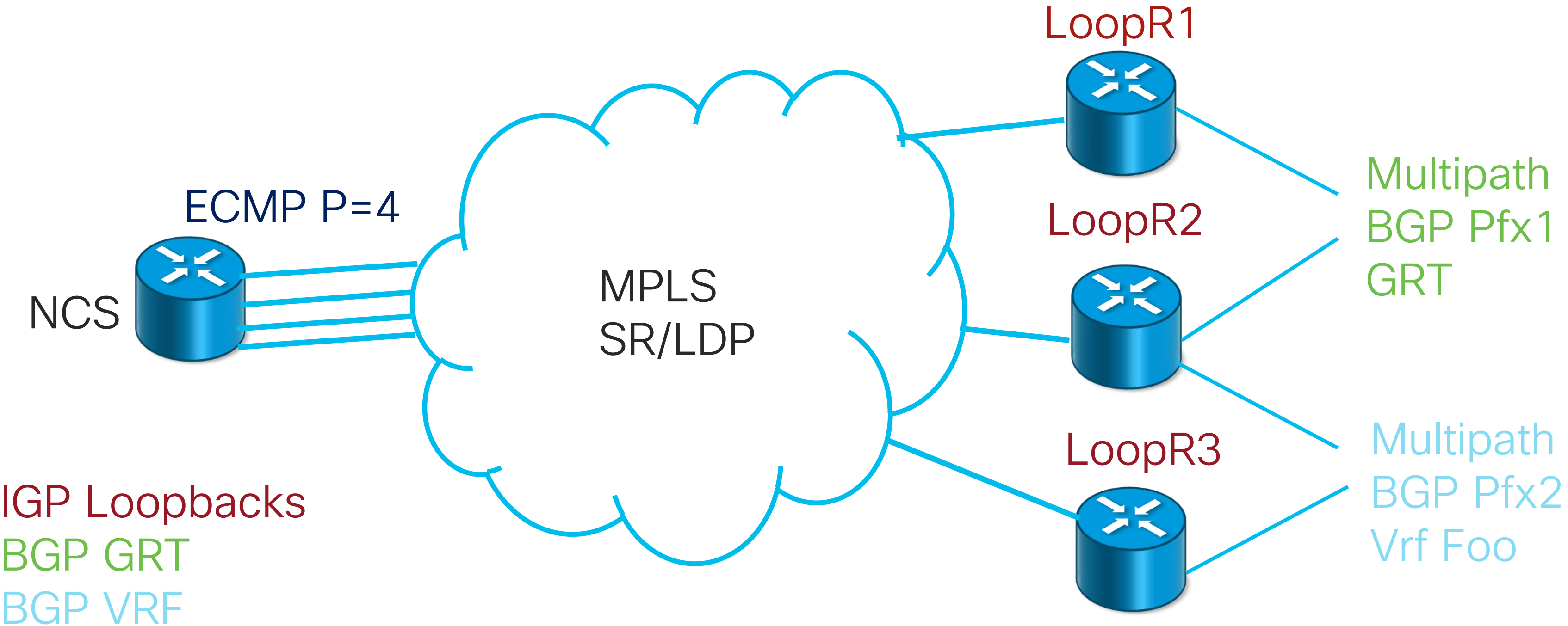
- MDB (Modular Database), configurable instead of fixed memory allocation → Profiles
- OP2 eTCAM
  - -Routing tables
  - -Statistics extension



	Balanced	L2 XL	L3 XL	IP+MPLS	Ext-KBP
FEC	204K	153K	613K	230K	768K
EEDB	144K	168K	176K	144K	512K

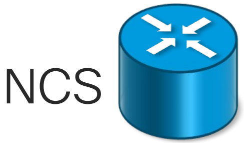
Optimize MPLS resource availability with SR

MPLS Topology



N1 IGP Loopbacks
N2 BGP GRT
N3 BGP VRF
P Paths

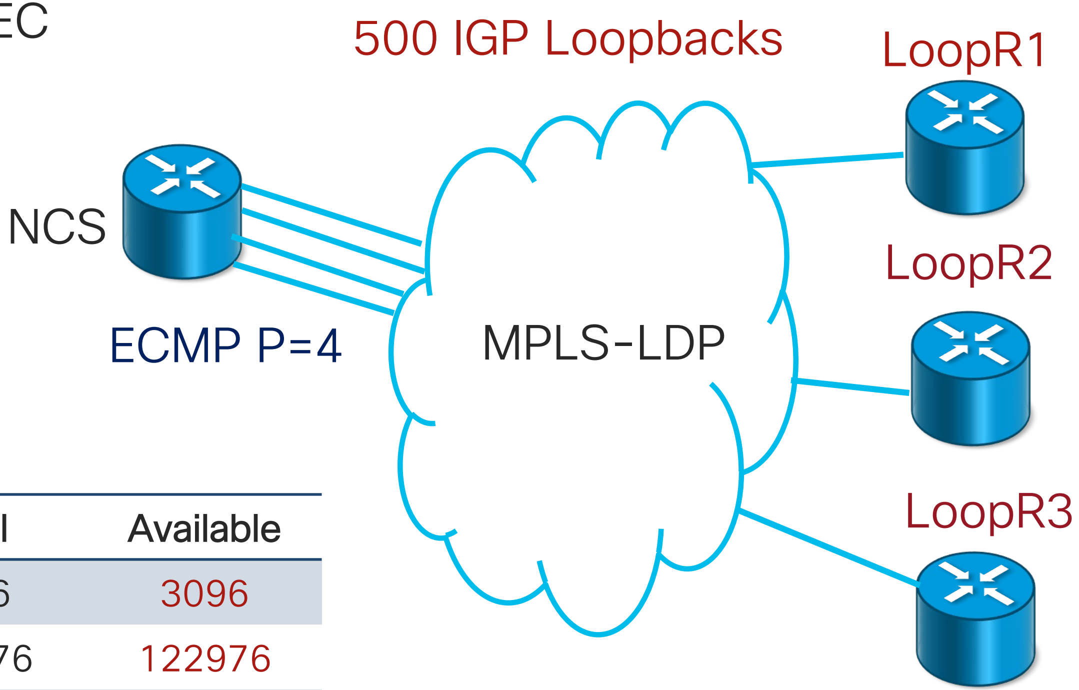
MPLS-LDP Resource Calculation



	Total	Available
ECMP-FEC	4096	4096
FEC	126976	126976
EEDB (MPLS)	80000	80000

MPLS-LDP Resource Calculation

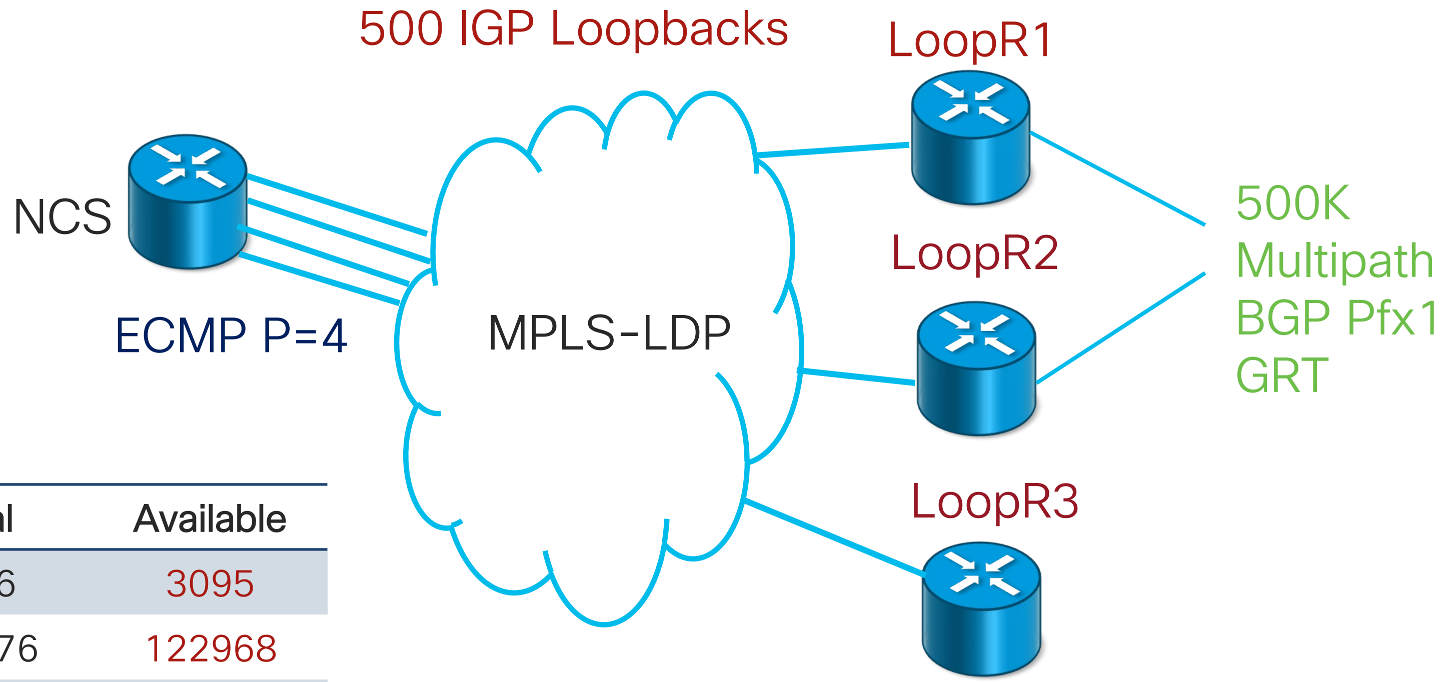
2 * 500 ECMP-FEC
2 * 500 * 4 FEC
2 * 500 * 4 EEDB



	Total	Available
ECMP-FEC	4096	3096
FEC	126976	122976
EEDB (MPLS)	80000	76000

MPLS-LDP Resource Calculation

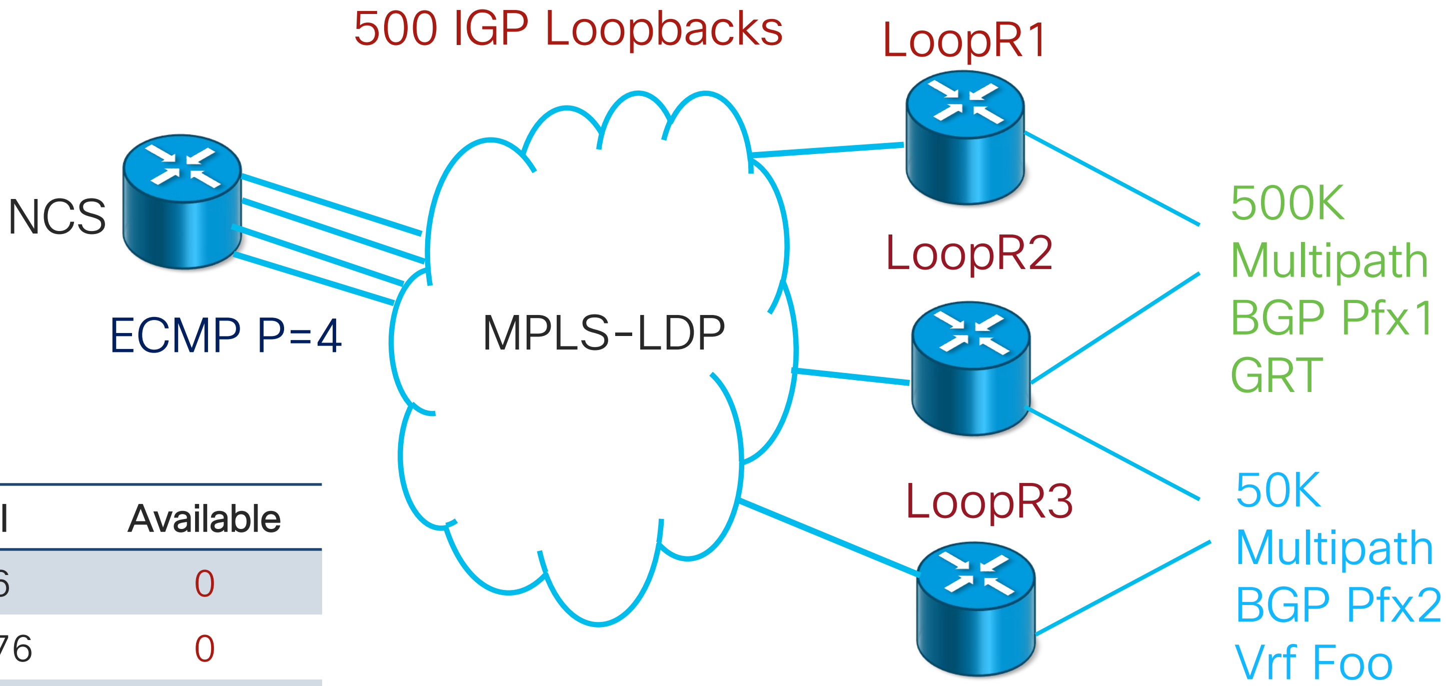
1 ECMP-FEC
1x2 FEC
1x2 EEDB



	Total	Available
ECMP-FEC	4096	3095
FEC	126976	122968
EEDB (MPLS)	80000	75992

MPLS-LDP Resource Calculation

50K ECMP-FEC
50K * 2 FEC
50K * 2 EEDB

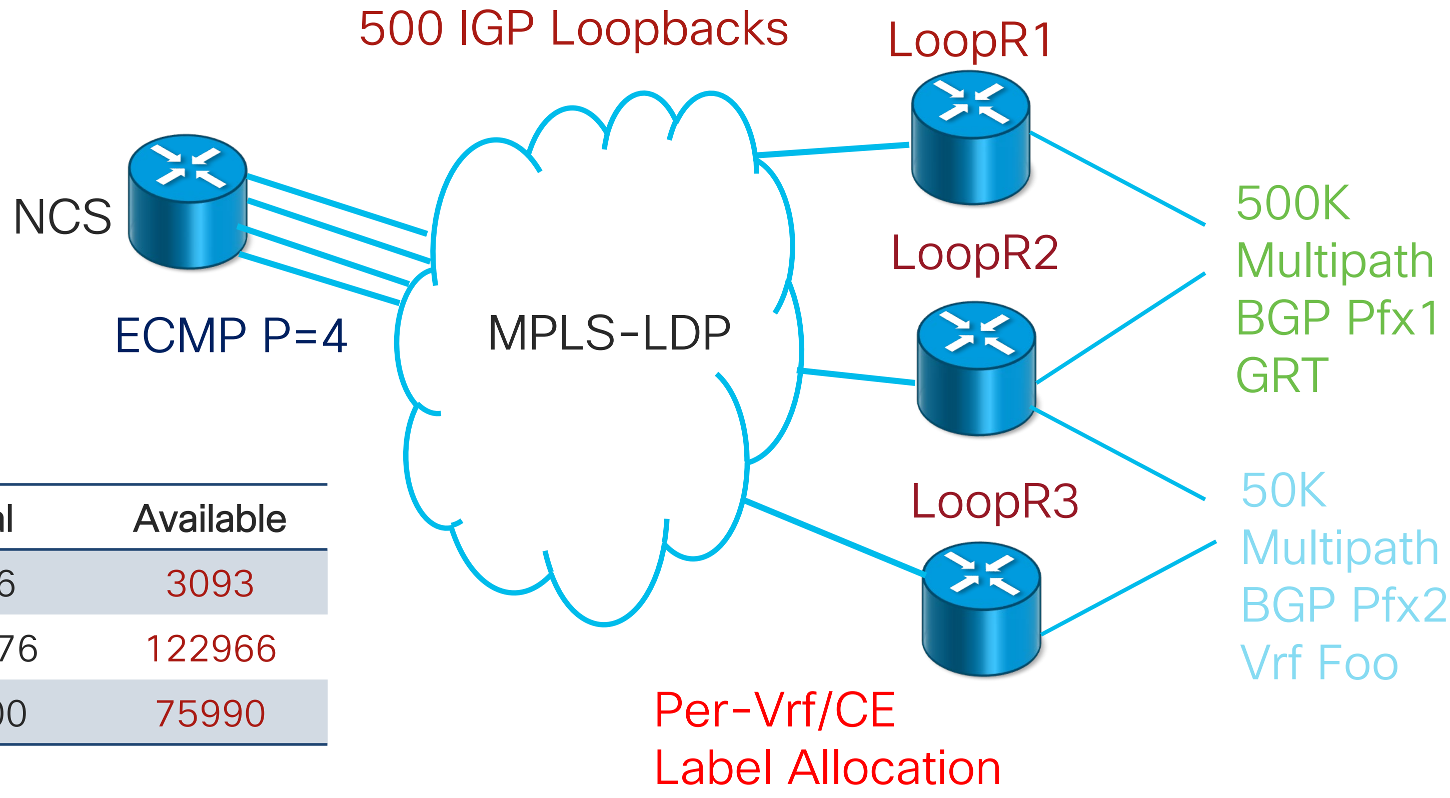


	Total	Available
ECMP-FEC	4096	0
FEC	126976	0
EEDB (MPLS)	80000	0

Per-Prefix
Label Allocation

MPLS-LDP Resource Calculation

1 ECMP-FEC
1 x 2 FEC
1 x 2 EEDB



	Total	Available
ECMP-FEC	4096	3093
FEC	126976	122966
EEDB (MPLS)	80000	75990

Resource Calculation for LDP

v4&v6 consume the same except LPMx2

- For N1 IGP prefix with LDP it will be consumed:
  - 2 * N1 LEM/LPM ; MPLS2MPLS and IP2MPLS ; 3 * N1 if mix IP path
  - 2 * N1 ECMP-FEC ; 3 * N1 if mix IP path
  - 2 * N1 * P FEC ; 3 * N1 * P if mix IP path
  - 2 * N1 * P EEDB

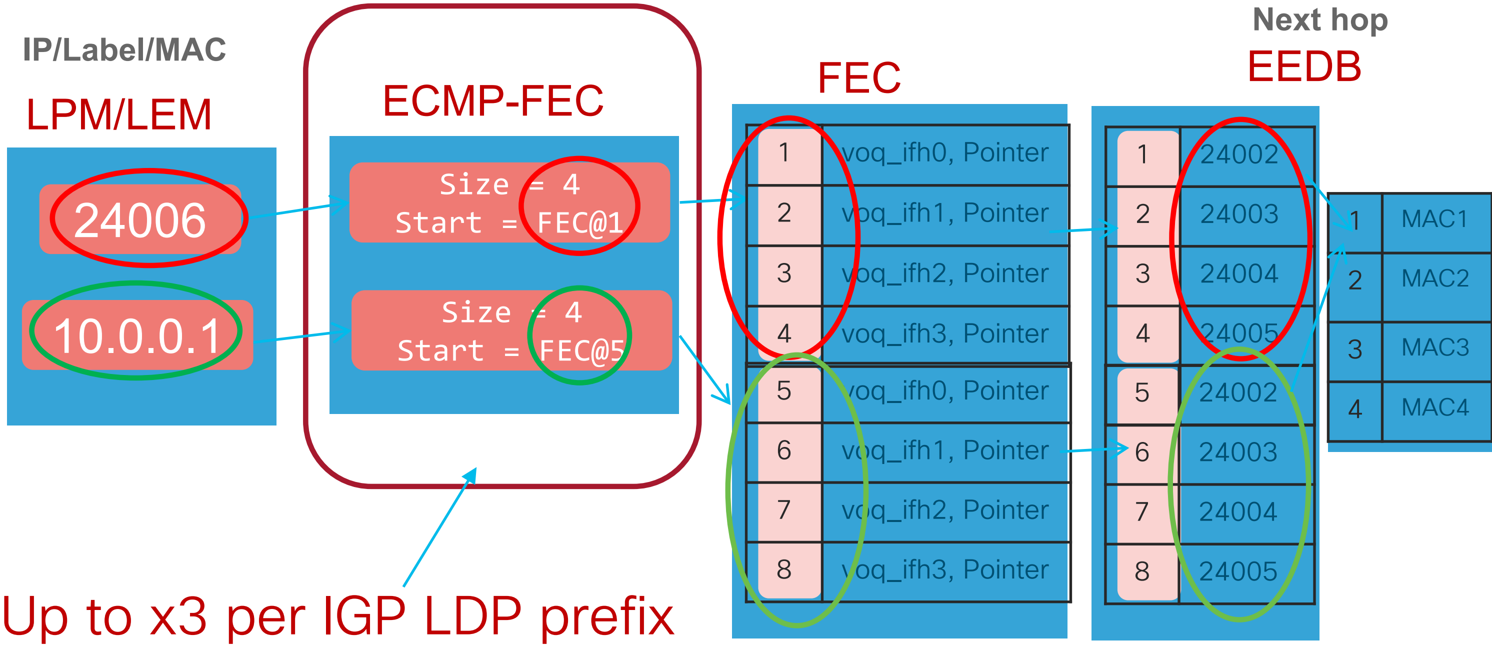
- For N2 BGP GRT and N3 BGP VRF prefixes with multipath to Z PEs:
- N2 + N3 LEM/LPM
- 1 (Group NH PEs) + N3 ECMP-FEC (if per-prefix used, 1 (Group PEs) if per ce/per vrf)
- Z * #NH + N3 * Z FEC (if per-prefix used, Z if per ce/per vrf)
- Z * #NH + N3 * Z EEDB (if per-prefix used, Z if per ce/per vrf)

IGP Prefixes with different label per interface

Remote R1

Local Label	Out Label	OIF
24006	24002	if0
	24003	if1
	24004	if2
	24005	if3

Prefix	Out Label	OIF
10.0.0.1	24002	if0
	24003	if1
	24004	if2
	24005	if3



FEC

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer
5	voq_ifh0, Pointer
6	voq_ifh1, Pointer
7	voq_ifh2, Pointer
8	voq_ifh3, Pointer

Next hop EEDB

1	24002
2	24003
3	24004
4	24005
5	24002
6	24003
7	24004
8	24005

1	MAC1
2	MAC2
3	MAC3
4	MAC4

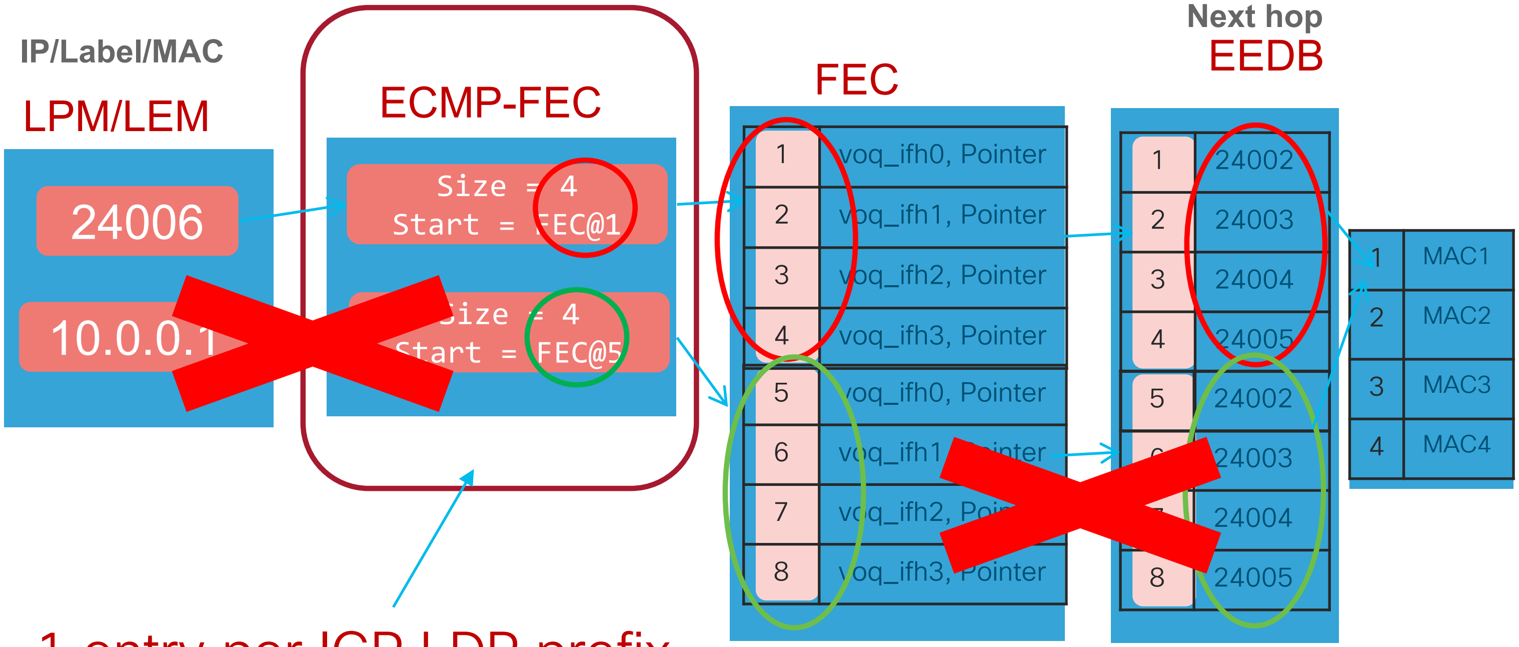
Up to x3 per IGP LDP prefix

IGP Prefixes with different label per interface- Optimized for LSR

Remote R1

Local Label	Out Label	OIF
24006	24002	if0
	24003	if1
	24004	if2
	24005	if3

Prefix	Out Label	OIF
10.0.0.1	24002	if0
	24003	if1
	24004	if2
	24005	if3



FEC

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer
5	voq_ifh0, Pointer
6	voq_ifh1, Pointer
7	voq_ifh2, Pointer
8	voq_ifh3, Pointer

EEDB

1	24002
2	24003
3	24004
4	24005
5	24002
6	24003
7	24004
8	24005

Next hop

1	MAC1
2	MAC2
3	MAC3
4	MAC4

1 entry per IGP LDP prefix

LDP Optimizations

- NCS device is only doing LSR role (no IP2MPLS)
- No services configured: L3VPN, L2VPN, BGP-LU
- All paths are labelled.
- We can then collapse the 2-3 entries into just one for swap case saving ECMP-FEC, FEC and EEDB.
- Convergence is also a benefit.

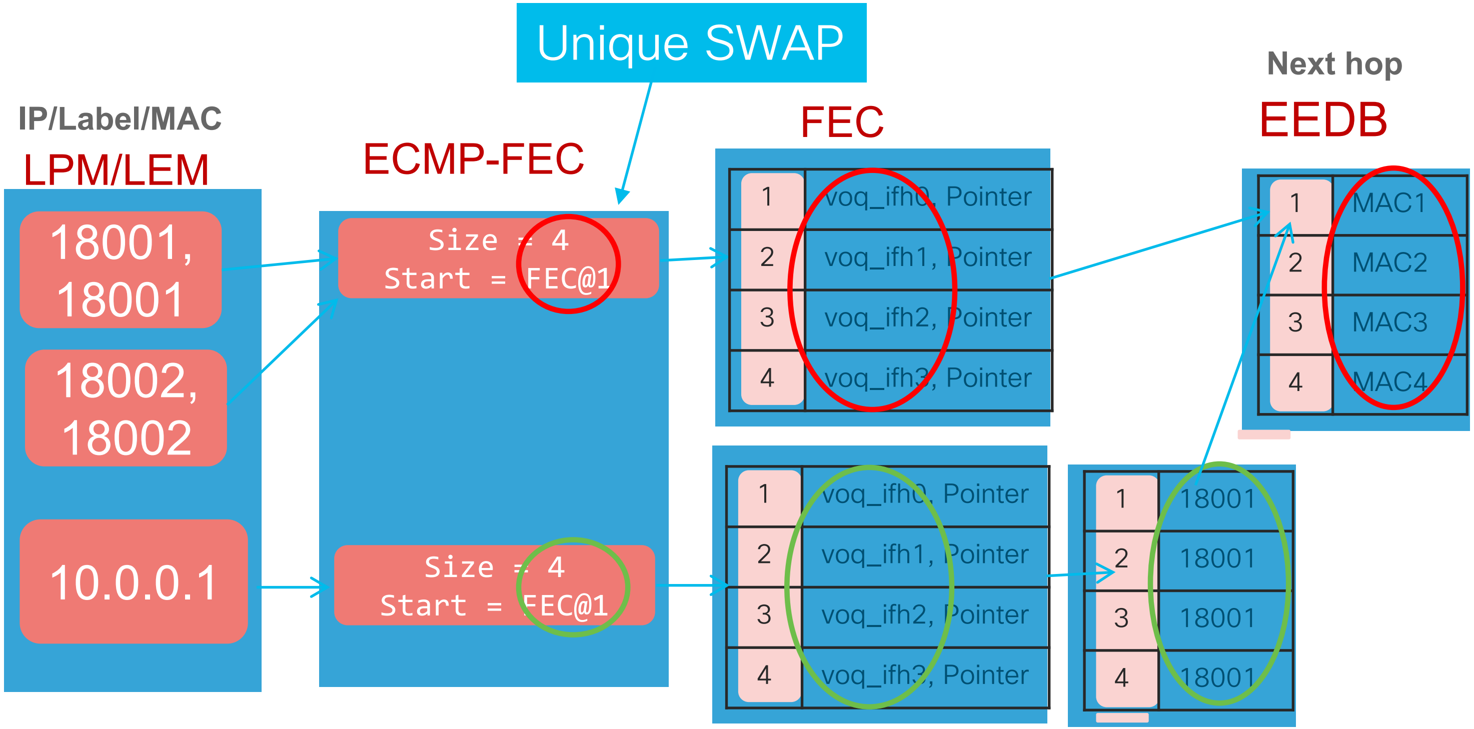
- CLI “hw-module fib mpls ldp lsr-optimized”
  - 2 * N1 LEM/LPM -> N1
  - 2 * N1 ECMP-FEC -> N1
  - 2 * N1 * P FEC -> N1 * P
  - 2 * N1 * P EEDB -> N1 * P
  - Support up to 3.3K LDP prefixes

IGP Prefixes with same label per interface - SR

Remote R1

Local Label	Out Label	OIF
18001	18001	if0
	18001	if1
	18001	if2
	18001	if3

Prefix	Out Label	OIF
10.0.0.1	18001	if0
	18001	if1
	18001	if2
	18001	if3



Unique SWAP

IP/Label/MAC

LPM/LEM

18001, 18001

18002, 18002

10.0.0.1

ECMP-FEC

```
Size = 4
Start = FEC@1
```

```
Size = 4
Start = FEC@1
```

FEC

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer

1	18001
2	18001
3	18001
4	18001

Next hop

EEDB

1	MAC1
2	MAC2
3	MAC3
4	MAC4

SR Gains out of the box

- ECMP-FEC push entry remains the same.
- Savings in ECMP-FEC SWAP entry that will be shared by all LEM entries.
- This will also make us save FEC and EEDB entries.
- All services can run in this mode.

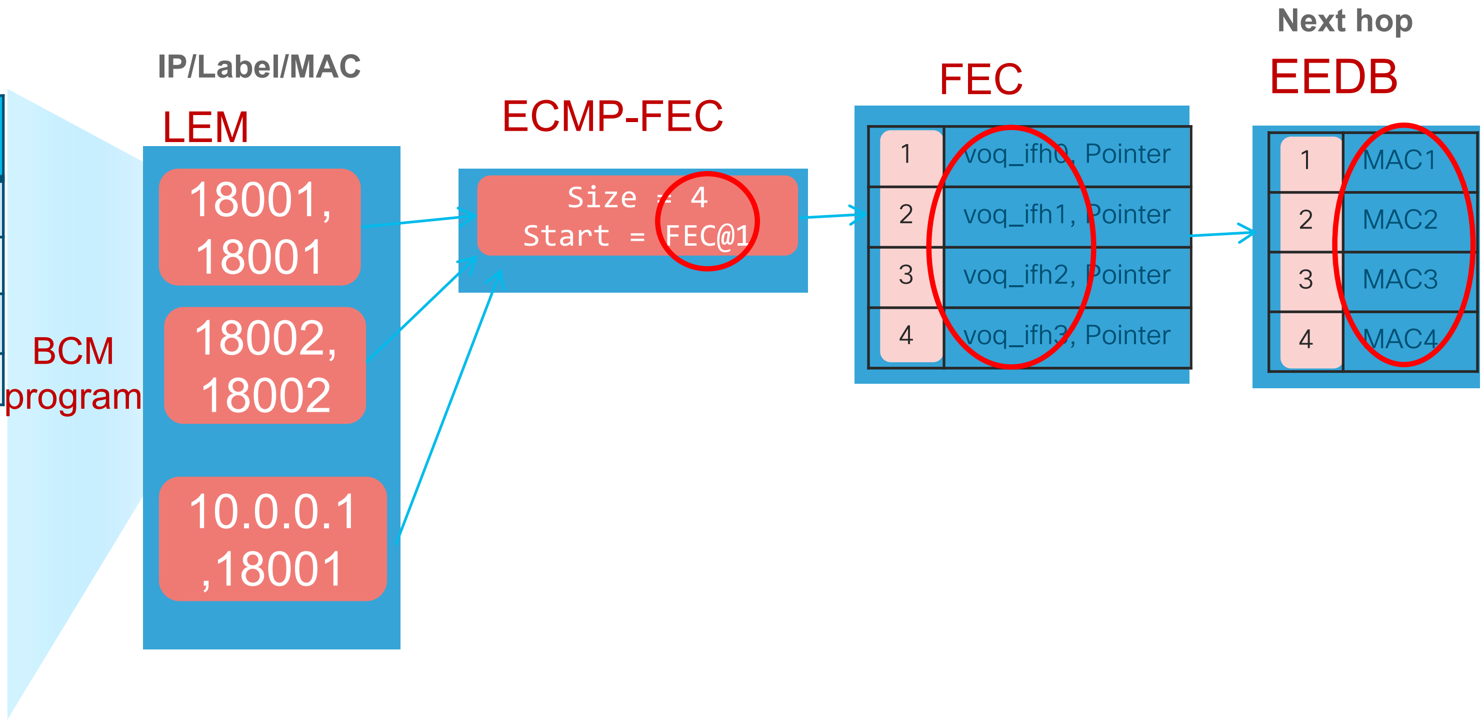
- 2 * N1 LEM -> No change
- 2 * N1 ECMP-FEC -> N1
- 2 * N1 * P FEC -> N1*P
- 2 * N1 * P EEDB -> N1

IGP Prefixes with same label per interface - SR Optimized

Remote R1

Local Label	Out Label	OIF
18001	18001	if0
	18001	if1
	18001	if2
	18001	if3

Prefix	Out Label	OIF
10.0.0.1	18001	if0
	18001	if1
	18001	if2
	18001	if3



FEC

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer

EEDB

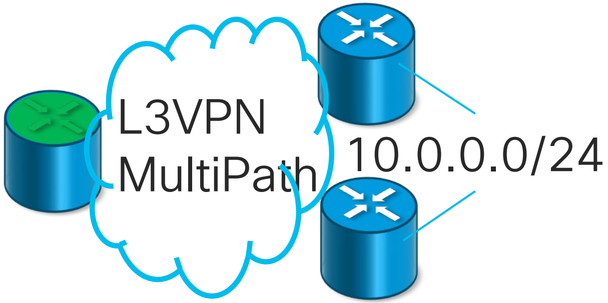
1	MAC1
2	MAC2
3	MAC3
4	MAC4

SR Optimizations

- NCS device is only doing LSR role and IP2MPLS (Only for IGP IPv4 /32 in LEM)
- No services configured: L3VPN (7.1.1) , L2VPN, BGP-LU, 6PE, 6vPE
- All paths are labelled.
- ECMP entries can be collapsed in 1.
- FEC/EEDB entries are saved compared to LDP.
- Convergence is also a benefit.

- CLI "hw-module fib mpls label lsr-optimized" for IP2MPLS (MPLS2MPLS default)
  - 2 * N1 LEM -> No change
  - 2 * N1 ECMP-FEC -> 1
  - 2 * N1 * P FEC –> P
  - 2 * N1 * P EEDB -> P

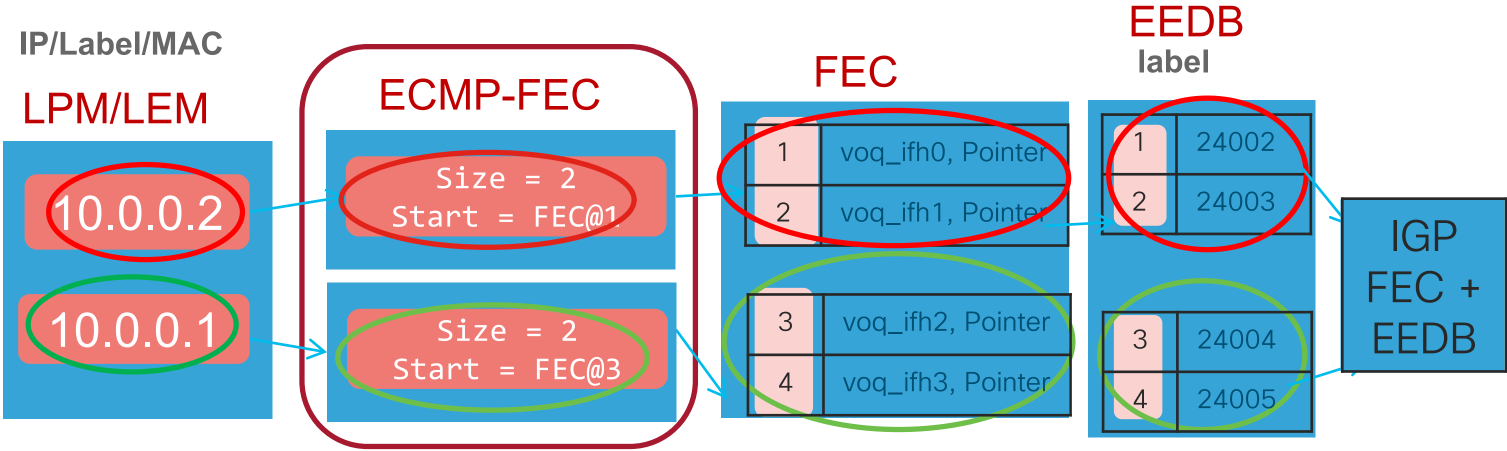
L3VPN – per prefix label allocation



Remote R1

Local Label	Out Label	OIF
10.0.0.2	24002	if0
	24003	if1

Prefix	Out Label	OIF
10.0.0.1	24004	if0
	24005	if1



ECMP-FEC:

```
Size = 2
Start = FEC@1

Size = 2
Start = FEC@3
```

FEC:

1	voq_ifh0, Pointer
2	voq_ifh1, Pointer
3	voq_ifh2, Pointer
4	voq_ifh3, Pointer

EEDB label:

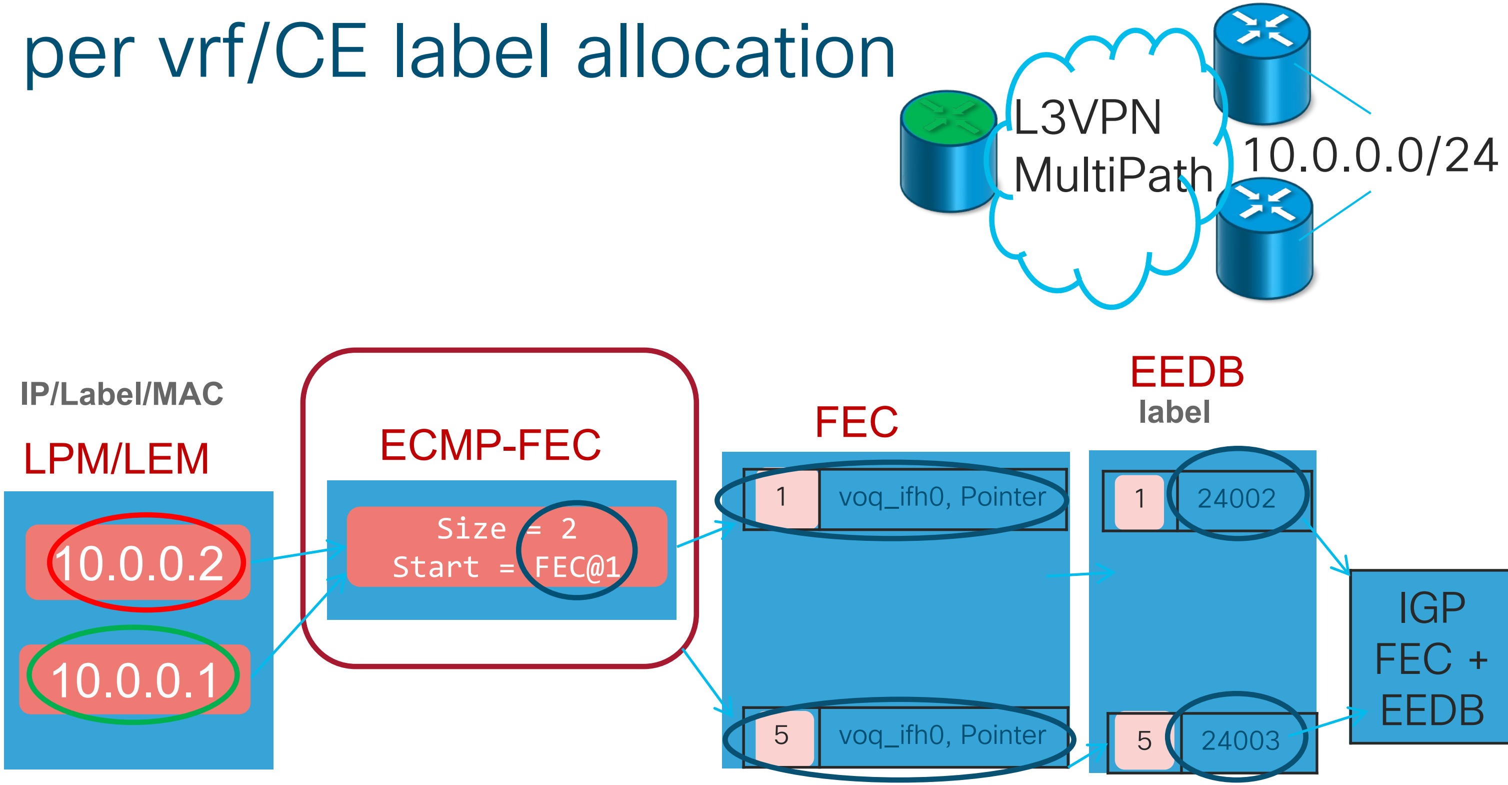
1	24002
2	24003
3	24004
4	24005

L3VPN – per vrf/CE label allocation

Remote R1

Local Label	Out Label	OIF
10.0.0.2	24002	if0
	24003	if1

Prefix	Out Label	OIF
10.0.0.1	24002	if0
	24003	if1

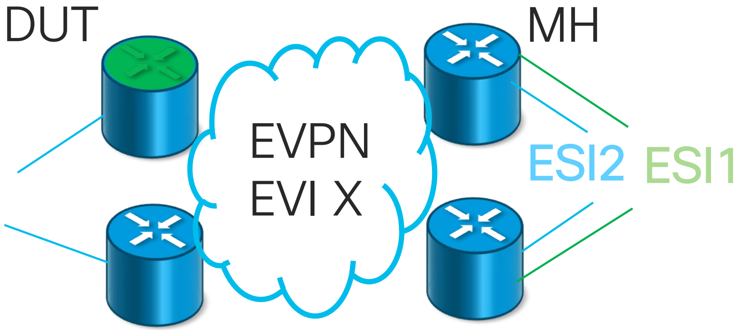


A Word on Protection

- Protection (Ti-LFA, TE-FRR, BGP PIC…) will double the number of FEC and EEDB consumed.
- Need to be careful understanding actual implementation.
- HW based BGP PIC Edge:
  - Only 1 active and 1 backup path
  - “cef encap-sharing disable” needs to be configured so no matter what allocation mode is used, per-prefix behaviour is seen.
  - Up to 24K protected VPNv4 prefixes.

Mcast and EVPN

- Each L3 mcast route consumes a LPM (iTCAM for L2) and a FEC
- FEC entries are not reused even when outgoing interface is same. Careful dimensioning is required as multicast and unicast flows shared the same FEC pool.



- Per remote EVI + ESI will need:
  - Unicast-> ECMP FECx1 + FEC x2 + EEDB x 2 (per remote EVPN MH pair)
- Per remote EVI:
  - BUM -> FEC x 1 + EEDB x1 (per remote peer including MH pair)

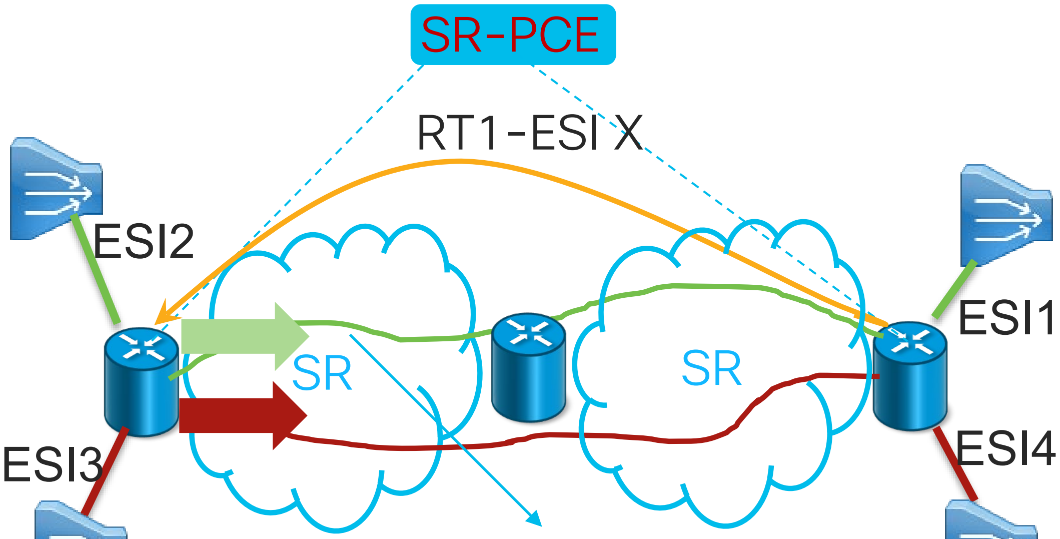
Resource Calculation

```
show controllers npu resources all location all
```

	LPM/LEM	ECMP-FEC	FEC	EEDB	
LDP	2*N1	2*N1	2*N1*P	2*N1*P	All Services
LDP with Optimizations	N1	N1	N1*P	N1*P	Only LSR
SR	2*N1	N1	N1*P	N1	All Services
SR with optimizations	2*N1	1	P	P	LSR and IP2MPLS, L3VPN
BGP-LU	2*N1	2*N1	2*N1*#NH (ABR)	2*N1*#NH (ABR)	No Multipath for services

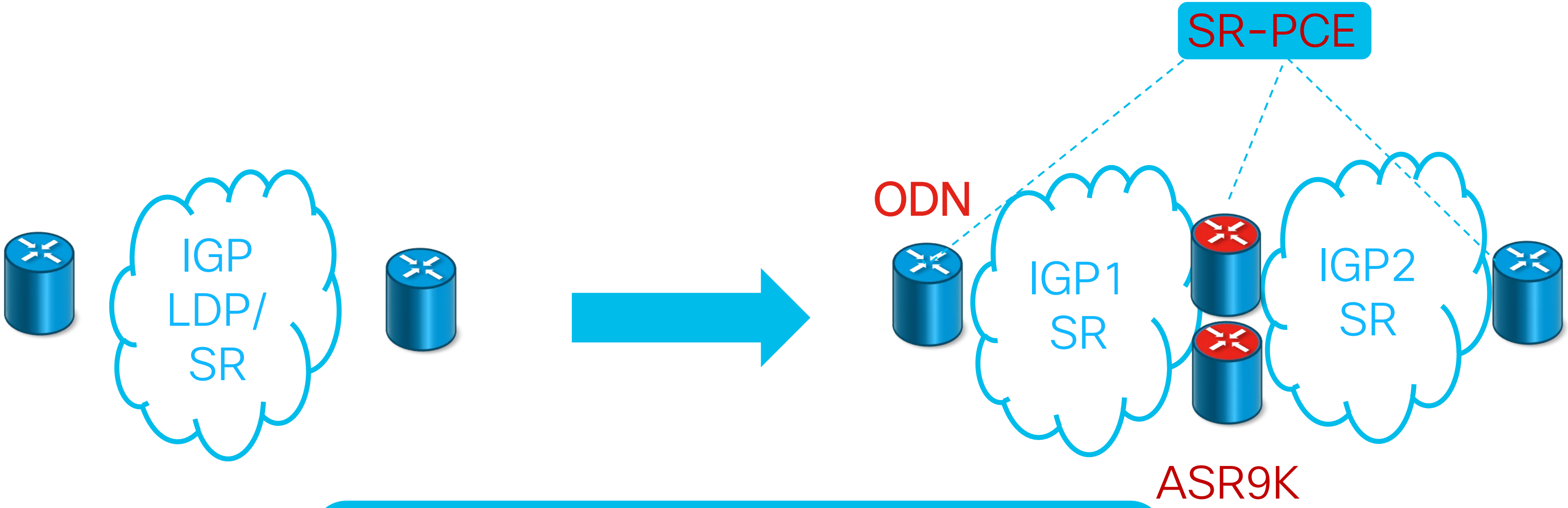
Services over SRTE – On Demand Next Hop

- P2P EVPN VPWS Single homed based on Route-type 1 (ESI).
- Also supported L3VPN, L3 GRT, 6PE, 6vPE.
- EVPN ELAN on 7.2.1
- MH options for P2P and ELAN on 7.2.1
- Classification on RT2 and RT5 roadmap item.



Disjoint Paths, minimize cost, delay….

Split IGP: Save Resources – SR-PCE+ODN



Savings ECMP/FEC/EEDB
Depends on #Remote PEs we need to connect for Services - Dynamic

Tactical Approach – Label Filtering

- LDP prefix filtering (ie. Do not learn output labels for remote prefixes we are not interested in).
- We remove all memory structures.
- Saves ECMP-FEC, FEC and EEDB.

mpls ldp
label accept for prefix-acl from ip-address

- Label allocation filtering (ie. Do not assign local label to remote prefixes if we do not need to perform SWAP).
- Saves SWAP entry ECMP-FEC, FEC and EEDB.
- Useful for IGP prefixes that we are not LSR for and BGP-LU remote prefixes (if we are not ABR)
- label local allocate for *prefix-acl*

```
router bgp X
    address-family ipv4 unicast
    allocate-label route-policy  pol
    route-policy pol
     if source in (0.0.0.0) then pass
    Else drop
```

ECMP behaviour change (6.6.2/7.0.1)

- When optimizations are not sufficient or redesign not possible, ECMP FEC dependencies can be eliminated by setting ISIS “maximum-paths 1”
  - Single different output interfaces chosen for each prefix
    - ECMP FEC usage = 0
    - Load balancing is fair, but ECMP won’t be done in HW
  - Only available for ISIS, plan for OSPF in 7.3.1

- Alternatively, similar behavior can be achieved for both OSPF/ISIS using:
  - “hw-module fib dlb level-1” enable
    - From release 6.6.2
    - No SRTE
    - No BGP PIC
- In 7.1.1, removes prefixes used for services as always using same link.

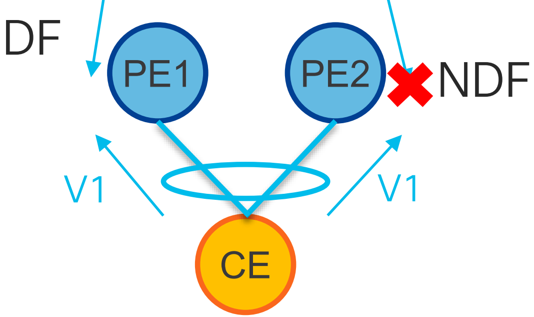
Key Take Aways

- Beware of BGP Multipath/ECMP. ECMP-FEC is the most precious resource.
- Try to restrict IGP LDP prefixes in the domain only allocating labels to loopbacks and stitch domains with BGP-LU/Controller
- Always use per-vrf/per-ce label allocation mode for L3VPN
- Move to SR for better resources allocation
- Use max-paths 1 when still scaling is not achievable.
- Be careful with redundancy implementation (ie. BGP Pic Edge, TI-LFA..) as it multiplies FEC&EEDB resources x 2.

First hop L2 and L3 redundancy with EVPN

EVPN – MH load-balancing modes (1)

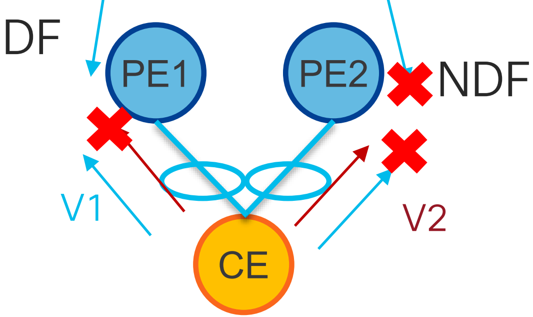
All-Active (per flow)



Single LAG at the CE
VLAN goes to both PE
NDF blocks egress BUM
Traffic hashed per flow

Benefits: Bandwidth, Convergence

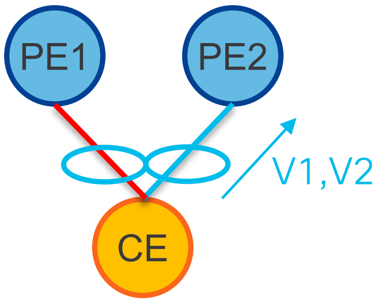
Single-Active (per VLAN)



BD + int/LAGs at the CE
VLAN active on single PE
NDF blocks all ingress traffic
and BUM egress traffic.
Traffic hashed per VLAN

Benefits: Billing, Policing

Port-Active (per port)

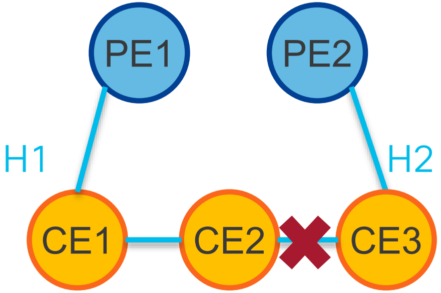


CE with S/A or A/A LAG options.
Port active on single PE
Backup port down or LACP OOS
Traffic hashed per port

Benefits: Protocol Simplification,QOS

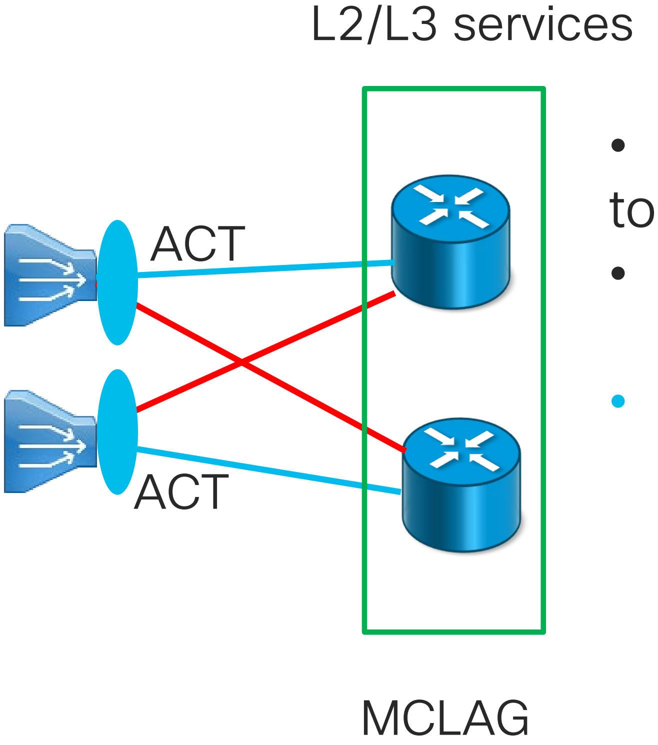
EVPN – MH load-balancing modes (2)

Single-Flow-Active (access L2 GW)



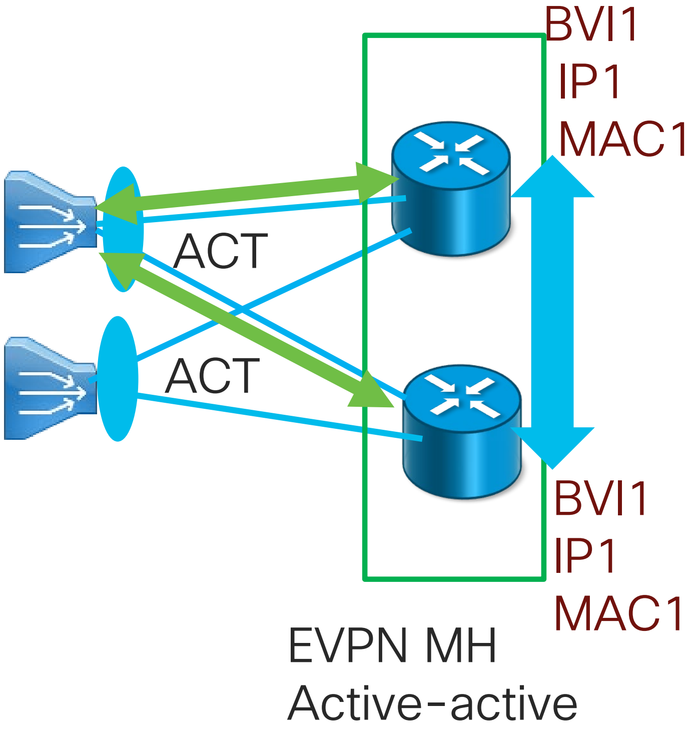
Single LAG at the CE
VLAN goes to both PE
Access takes care of L2 loop
Benefits: Legacy support for STP, MSTAG, G.8032. Faster convergence

MCLAG Common Design – First ask



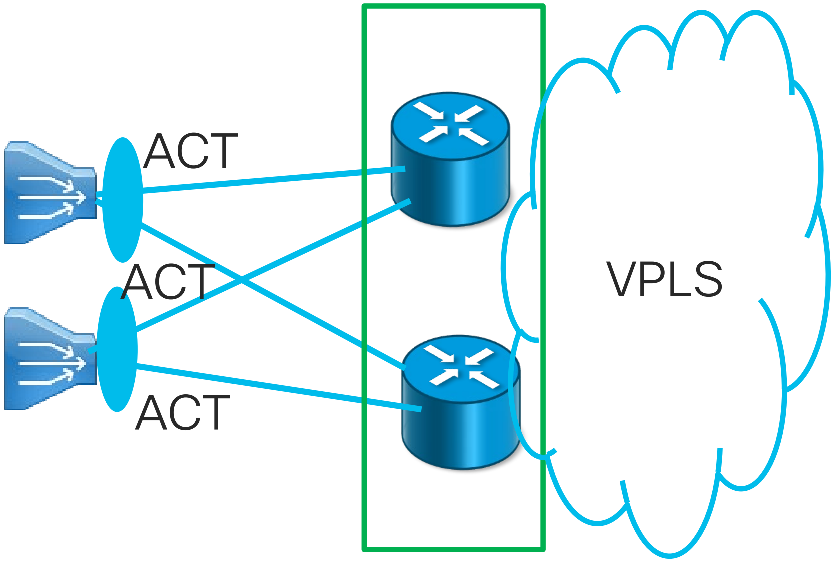
- Need to provide redundant POA to access devices → MCLAG
- MCLAG devices do L2 and L3 services.
- NCS has taken EVPN approach instead of MCLAG to provide FHRP, in a more scaled and flexible manner.

EVPN for redundant Access (L2 + L3)



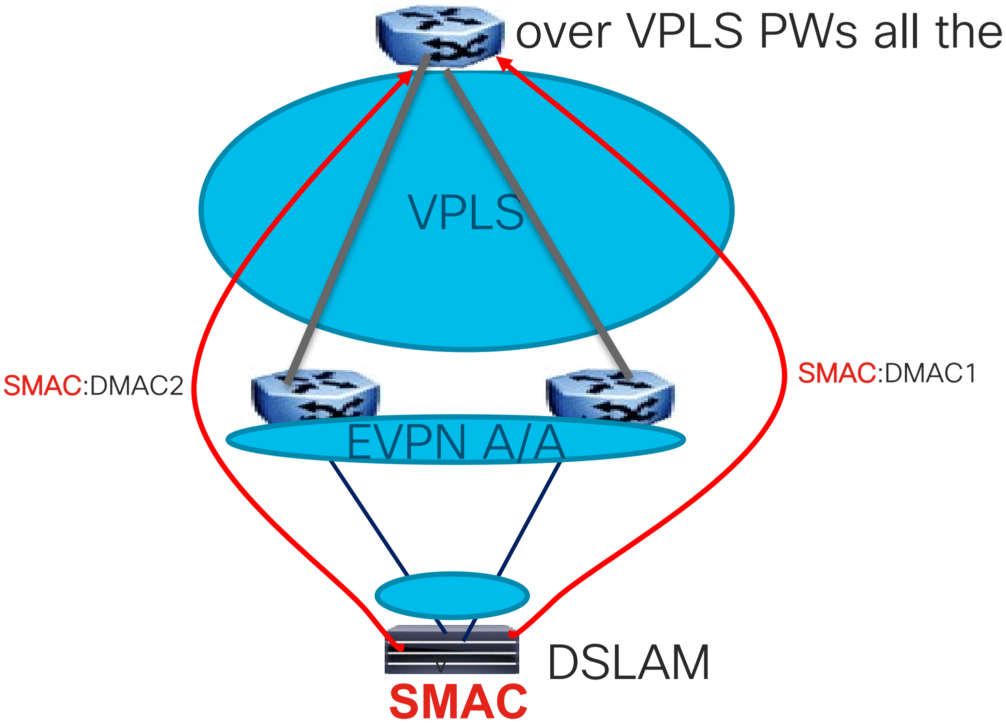
- EVPN MH in active-active mode
- There is load balancing in both directions.
- L2 subinterface + BD with BVI provides L3
- BVI with same IP and MAC in both devices (Anycast Gateway).
- EVPN does the ARP and IGMP sync for failover convergence.
- L2 subinterface + BD and remote EVPN peers provide L2.

MCLAG Common Design – Second ask



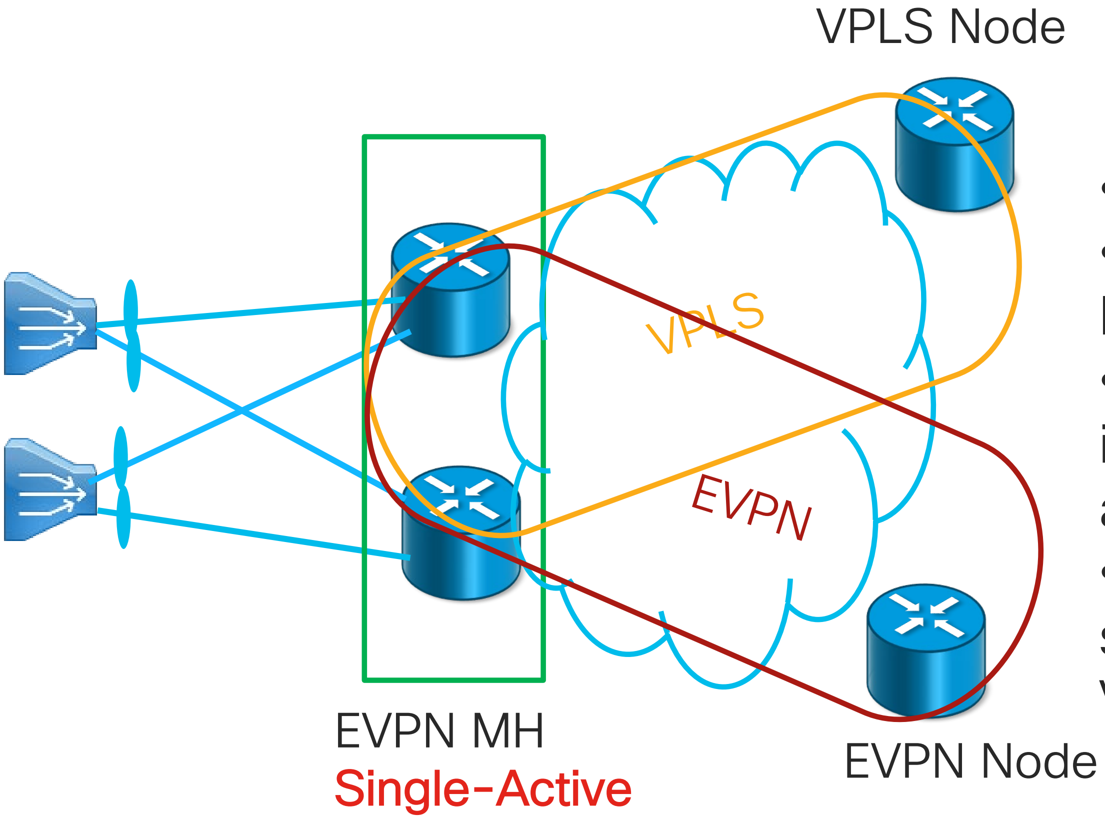
- While EVPN functionally replaces VPLS in the core, VPLS might still be required to interoperate with legacy devices
- EVPN MH A/A cannot be used due to VPLS protocol limitation.

Why not EVPN Active-Active for better LB?



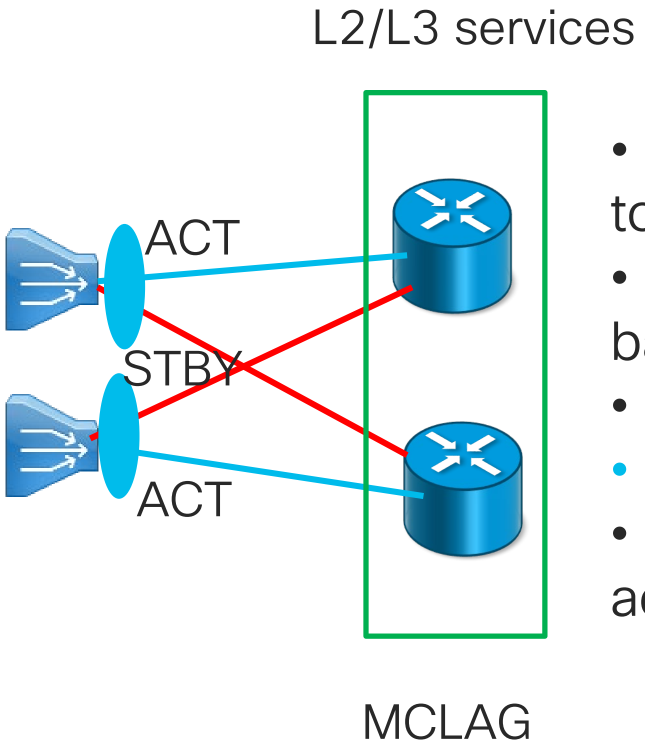
Device will see SMAC flip over VPLS PWs all the time.

EVPN-VPLS Migration



- EVPN to simulate MCLAG
- EVPN-VPLS seamless Integration feature
- Auto-detects if EVPN is enabled on a given BD for a given PE
- EVPN MH S/A to make sure only one PE is seen in VPLS world.

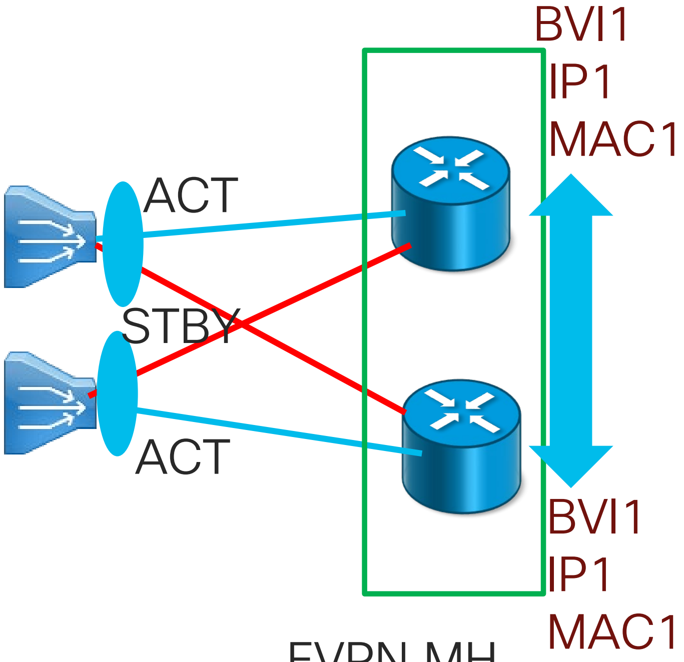
QOS accuracy – Third ask



- Need to provide redundant POA to access devices → MCLAG
- One port is active and the other is backup to ensure correct QOS.
- MCLAG devices do L2 and L3 services.
- NCS does not support MCLAG
- EVPN can provide this redundant access and other variants more optimized

EVPN MH Port-active

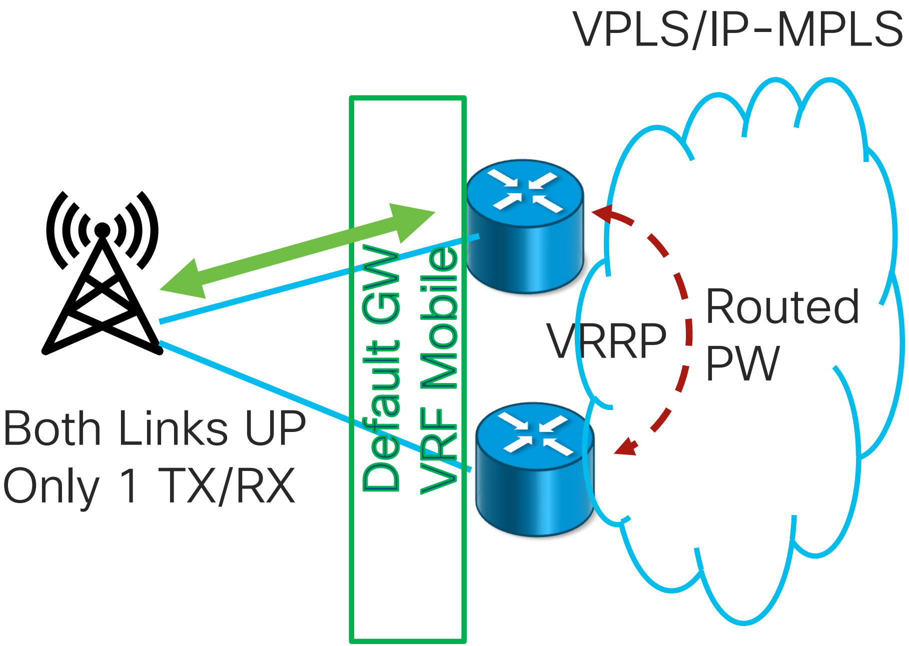
XR 7.1.2



EVPN MH
Port-active

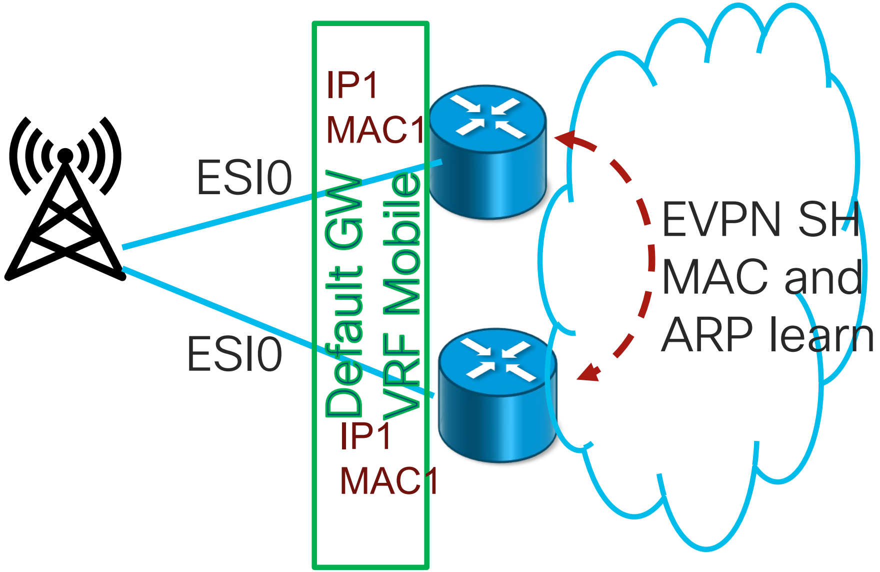
- EVPN MH in port-active mode
- One port is active and the other is backup for a given access device
- BVI with same IP and MAC in both devices for routed traffic (Anycast Gateway).
- EVPN does the ARP and IGMP sync for failover convergence.
- QOS is accurate as traffic only flows across one link.

VRRP Specific Design



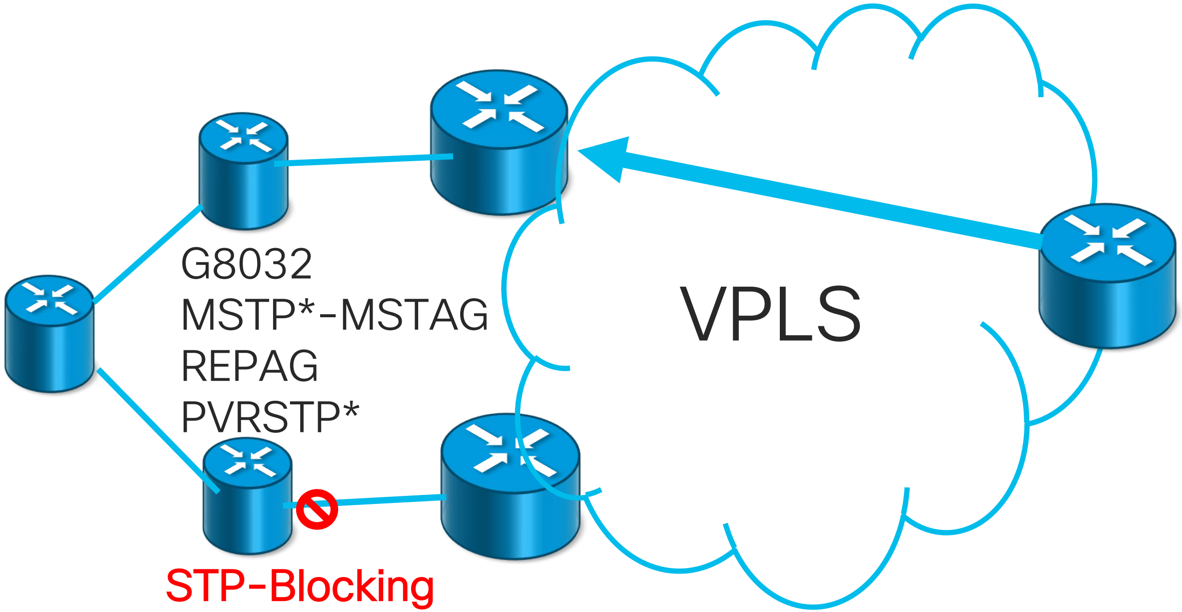
- VRRP to provide FHRP to eNODEb
- Routed PW to be able to send and receive VRRP hellos.
- Traffic in L3VPN.
- NCS55XX does not support VPLS Routed PW (7.1.1 but without VRRP support)
- Scale of VRRP is 255 VRID per System “hw-module vrrpscale enable” (6.6.1)

VRRP Alike Function with EVPN



- EVPN SH to be able to learn MAC from active link.
- No VRRP needed as same BVI IP/MAC on both NCS55XX
- Use the same custom MAC1 on BVIs to scale (7 custom max per Box)
- Core Isolation feature to bring down access port when core links fail.
- Scale of 1250 BVI in 7.0.1.
- If eNODEb can TX and RX on both links use EVPN MH A/A.

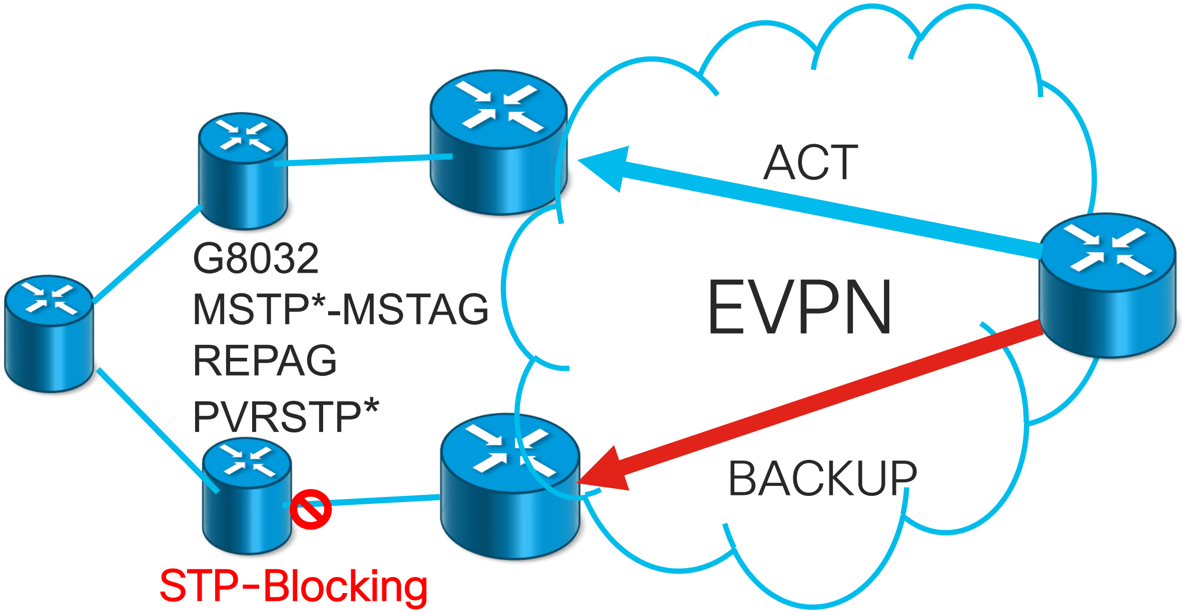
VPLS with Ring topologies



- Supported on NCS
- Care with BD scaling and convergence

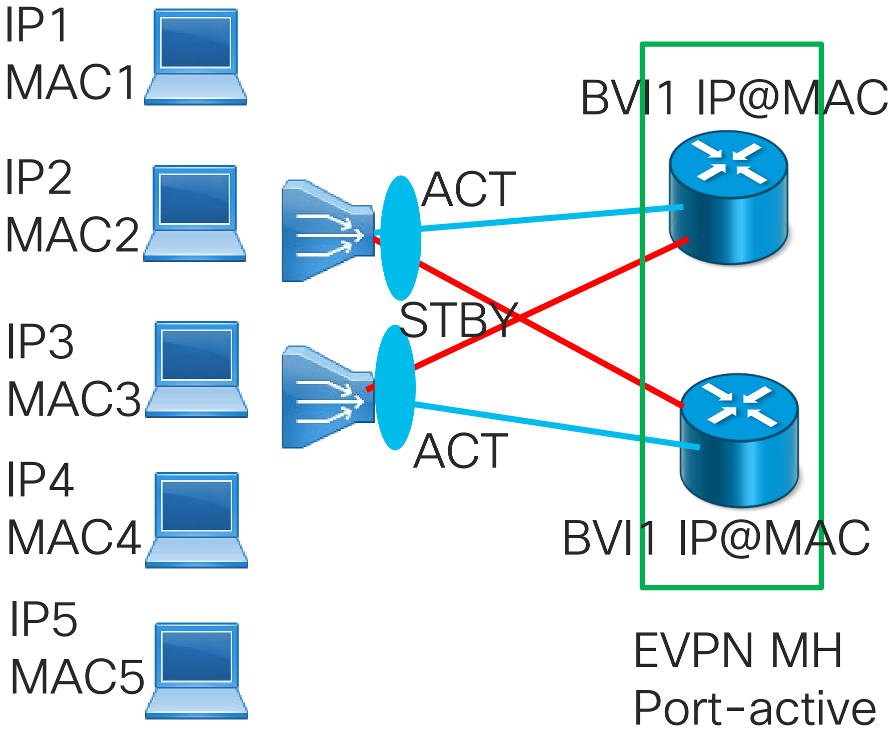
* In 7.1.1

EVPN with Ring topologies



- Plan for 7.4.1
- Will add MH Single-Flow-active mode to improve convergence time (AC down and PE down) as MACs are sync'ed accross PEs and learned as ACT/Backup on remote PEs

Scaling Concerns



- NCS Will learn MAC addresses from local users.
- Actual scale is 64K per BD and 128K per box. Plan to increase to 128K/256K
- If not enough, need to change to EVPN VPWS design → **More later**
- If doing L3, NCS Will learn ARP entries from local users.
- ARP scale 6144 per box. Increased to 30K in 6.6.1 on BVI.
- If not enough, need to provide centralized L3 services → **More later**

A Word on Statistics

- BCM chipsets offer limited number of statistics counters.
- The consequence is two fold:
  - Number of stats punches to be provided on a packet: 2 per direction on J+ (interface+subinterface+QOS+deny ACL)
    - hw-profiles to assign them differently

```
 RP/0/RP0/CPU0:(config)#hw-module profile stats ?
acl-permit          Enable ACL permit stats.
enh-sr-policy        Enable Enhanced_SR_Policy_Scale stats profile counter.
ingress-sr          Enable ingress SR stats profile counter.
qos-enhanced         Enable enhanced QoS stats.
tx-scale-enhanced  Enable enhanced TX stats scale (Non L2 stats)
```

  - Number of overall stats counters: J+ has 128K counters per core
  - **Solution**: Use platform with external FPGA as NCS560 or 55A2-MOD-SE/NC55-MOD-SE (Still to be supported by SW), or J2 (eTCAM)

Key Take Aways

- MCLAG can be replaced by EVPN
- If no strict QOS is required, MH Single-Active/Active-Active can be used to achieve per EVI/Flow Load balancing.
- Legacy VPLS integration is possible, but forces us not to use MH Single

Active/Port-Active for now.

- VRRP function can be provided with EVPN technology.
- MAC scale can be a concern and if so, a shift to EVPN VPWS is required.
- ARP scale can also be a concern and if so, a shift to centralized L3 services is required.

Demo

Centralization of L3 services

Complex asks for transport optimized platforms

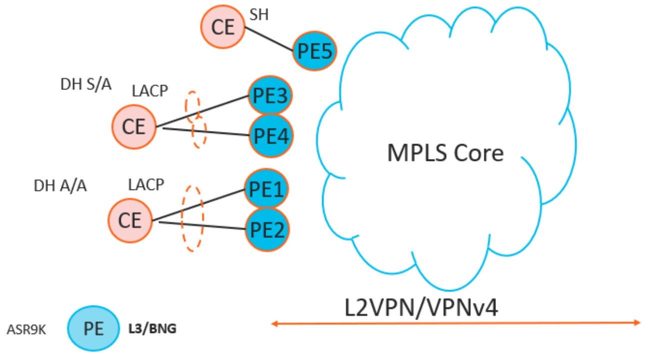
- Customer wants a 4 level QOS hierarchy to provide device-link-subscriber-traffic class QOS
- Policy-map with more tan 8 class-maps.
- Egress clasification of COS/DSCP.
- Customer wants QOS pmap to share a given number of subinterfaces (aka

SPI- Shared policy instance).

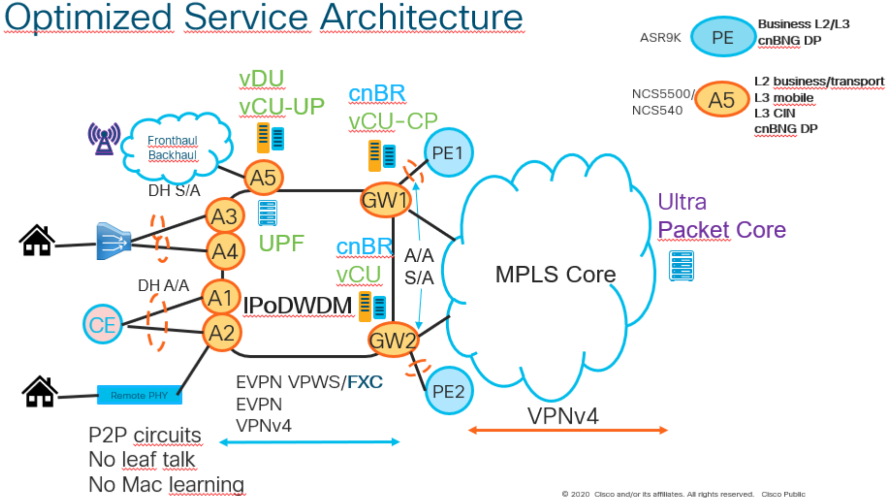
- We want IP services done with PWHE construct.
- Scale, Scale, Scale!!!!

Optimizing Service Location

Distributed PE architecture



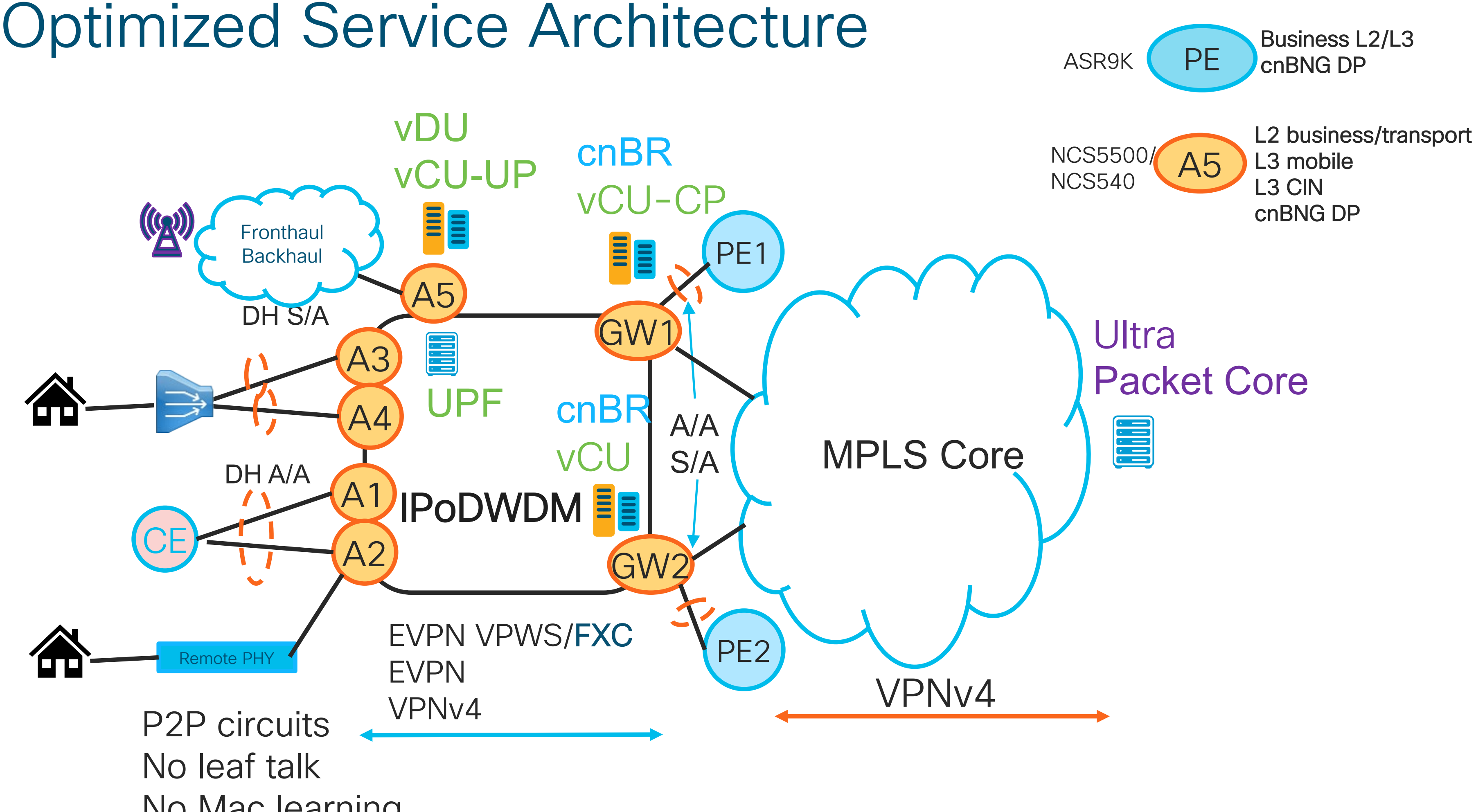
Optimized Service Architecture



Why not optimize services placement?

- ‘Cost optimized’ NCS family in Access/aggregation for EVPN transport with efficient load balancing and lower signalling overhead and ASR9K when needed.
- Lower Power footprint, for 100s/1000s of sites.
- Resources are provisioned and consumed only where needed, using ASR9K for complex and scaled features.
- SW Essentials will suffice. No need to provide >8 VRF licenses.
- QOS on Service PE with higher scale/functionality and downstream traffic will be shaped before entering aggregation network (5:1 ratio). Provides QOS accuracy.
- No scaling combinations in Access devices and fully utilize ASR9K as Service PE.
- Customer routing simplified with Anycast GW.

Optimized Service Architecture



ASR9K PE Business L2/L3 cnBNG DP

NCS5500/ NCS540 A5 L2 business/transport L3 mobile L3 CIN cnBNG DP

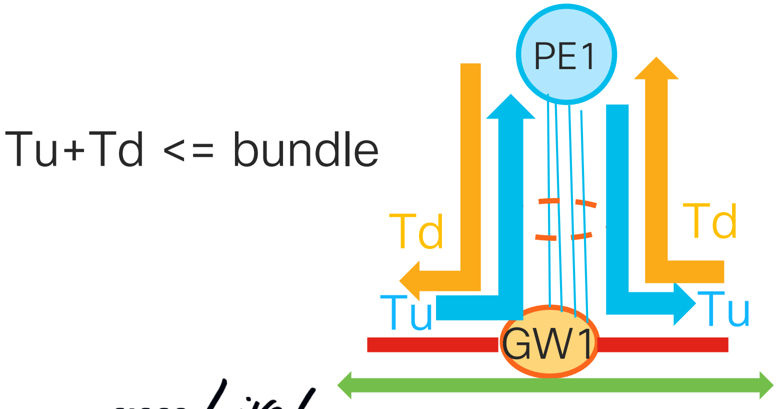
EVPN VPWS/**FXC**
EVPN
VPNv4

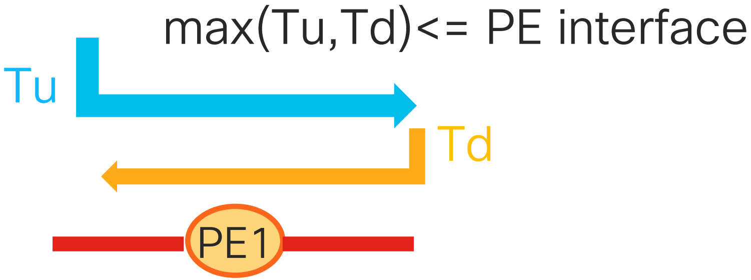
VPNv4

P2P circuits
No leaf talk
No Mac learning

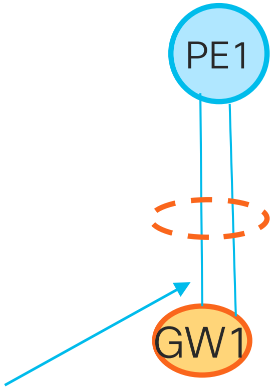
Why router on-a-stick bundle

- Effective when only a subset of traffic being transported needs L3 services.
- Provides an easy control of core/Access failure logic (if bundle fails, both fail).
- If allows more traffic through the box in case of assymetric flows (normal case). Transport network needs to provide this additional BW also.





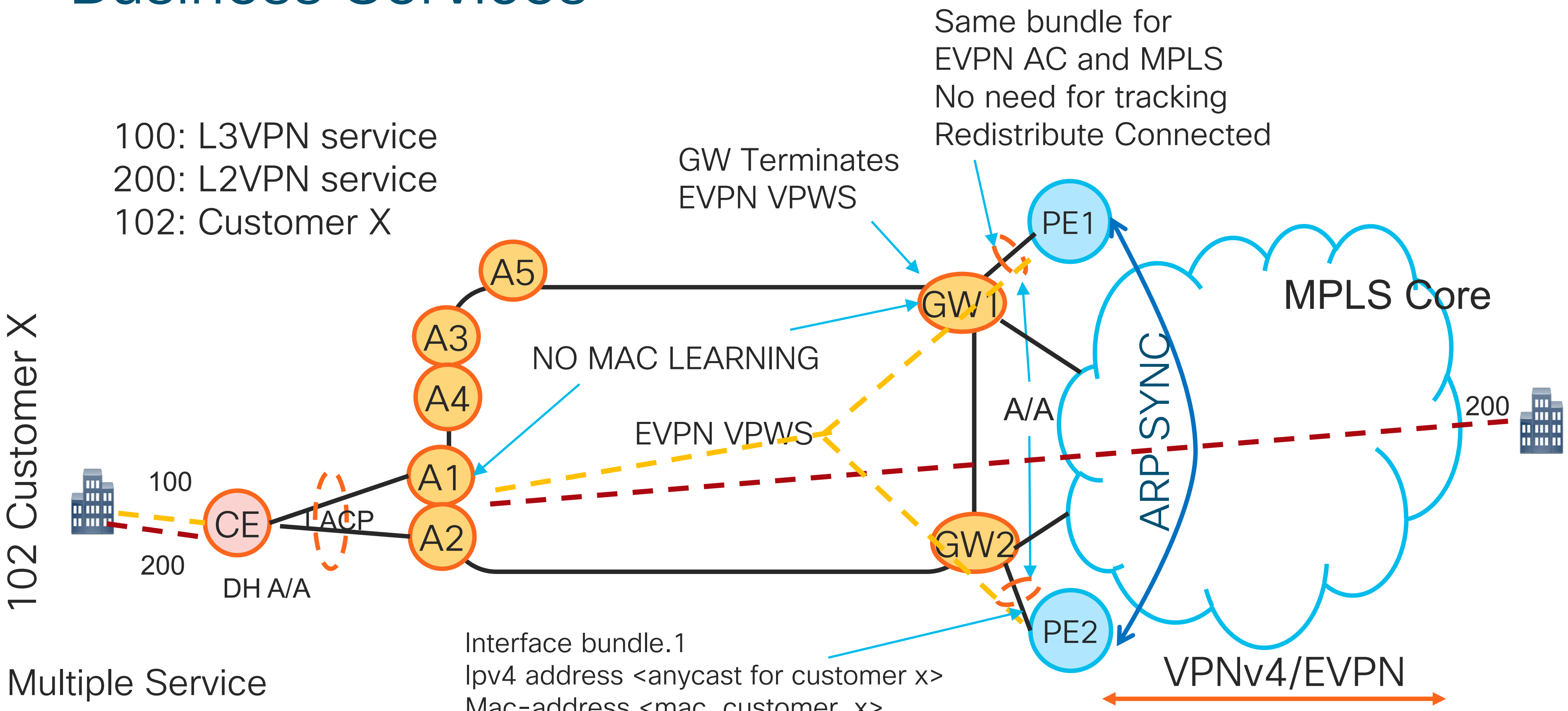
Single Bundle Solution



```
Int bundle X.1 l2transport
Encapsulation dot1q 100
Int bundle X.2
Ipv4 address Z
Mpls
Int bundle X.2
```

- Bundle needs to support subinterfaces with L2transport services (EVPN handoff) and L3 MPLS services (PE mpls exit point)
- Both ASR9K and NCS5500 support this scenario as GW.
- Routing is simple in this case as just redistributing static is enough.
- If bundle goes down, there is no blackholing as both Access and Core subinterfaces go down.

Business Services



100: L3VPN service
200: L2VPN service
102: Customer X

GW Terminates
EVPN VPWS

Same bundle for
EVPN AC and MPLS
No need for tracking
Redistribute Connected

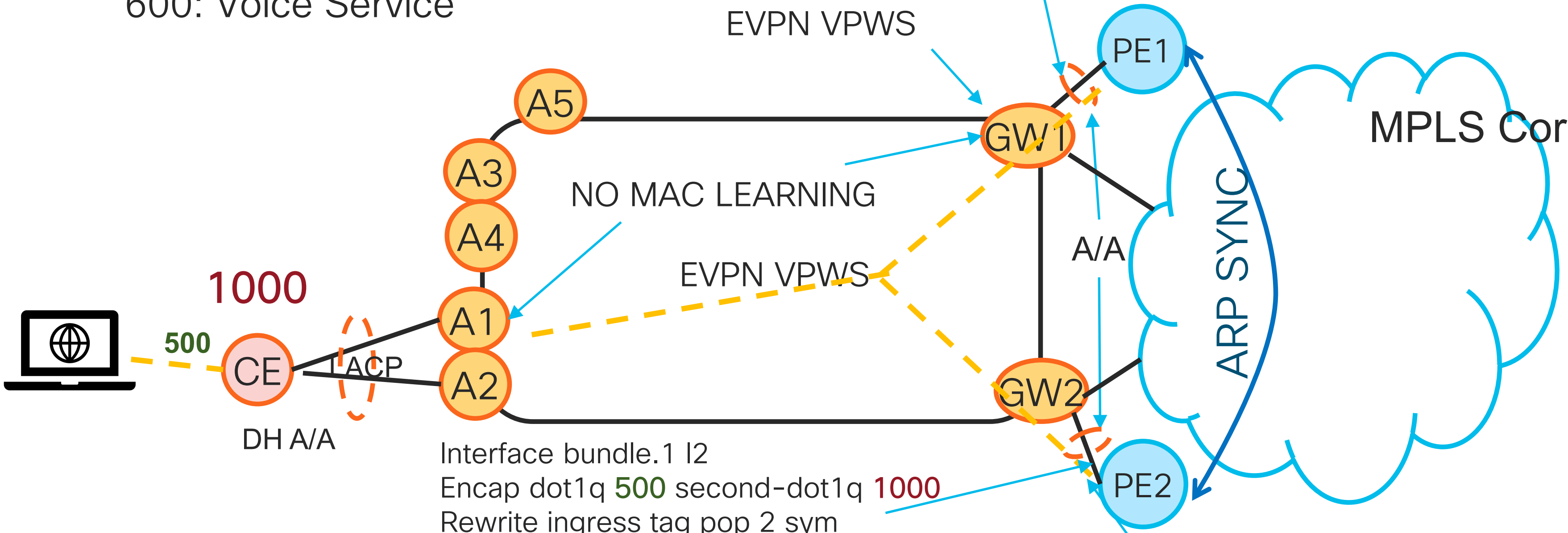
NO MAC LEARNING

Multiple Service
Per Ethernet Segment

```
Interface bundle.1
Ipv4 address <anycast for customer x>
Mac-address <mac_customer_x>
Encap dot1q 100 second-dot1q  102
Vrf customer_X
service-policy output <customer x>
```

Residential Services (no QOS)

500: Internet Service
600: Voice Service



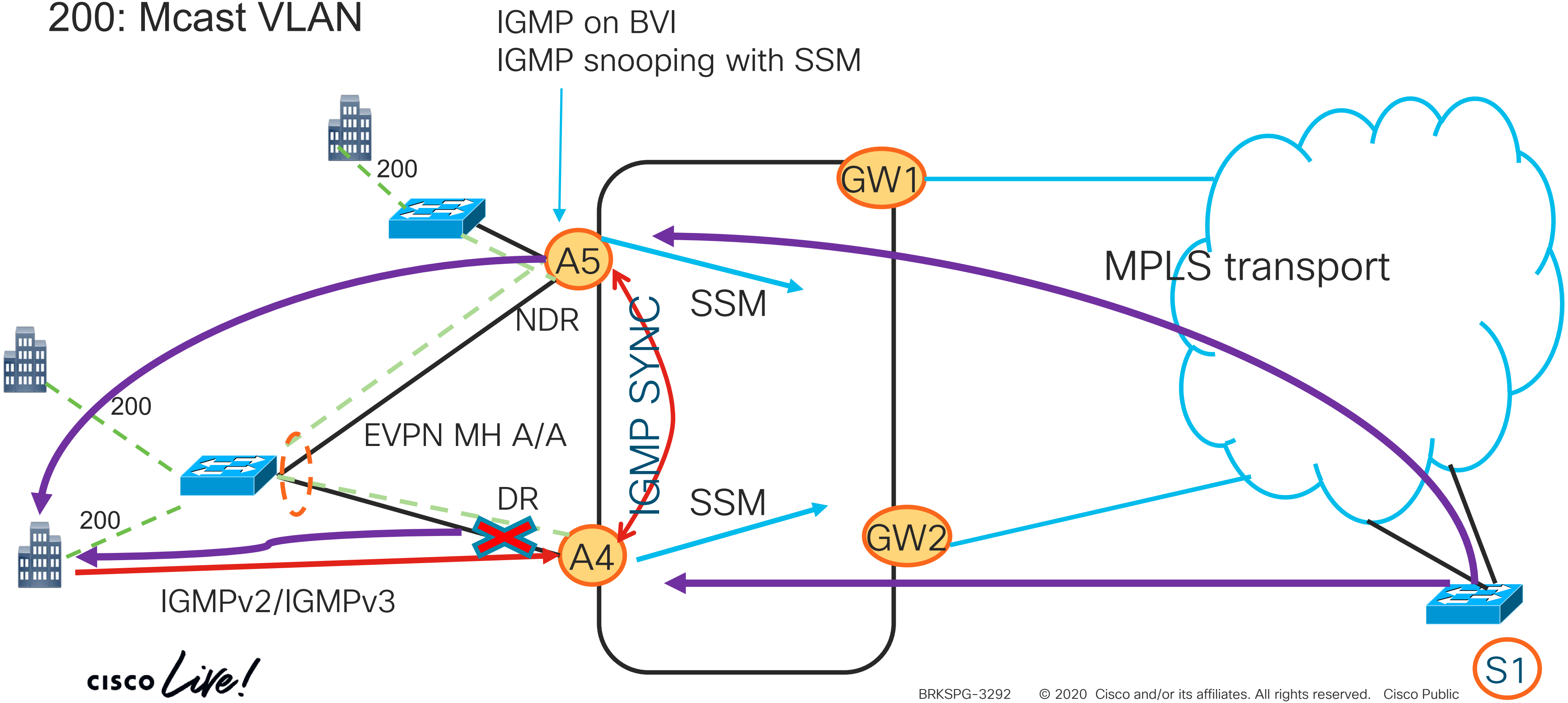
GW Terminates
EVPN VPWS

Same bundle for
EVPN AC and MPLS
No need for tracking
Redistribute Connected

```
Interface bundle.1 l2
Encap dot1q 500 second-dot1q 1000
Rewrite ingress tag pop 2 sym
L2vpn
Bridge-domain X
Interface bundle.1
Routed Interface BVI1
```

```
Interface BVI1
Ipv4 address <Residential big subnet>
Mac-address <residential MAC>
Vrf <Residential>
```

Multicast

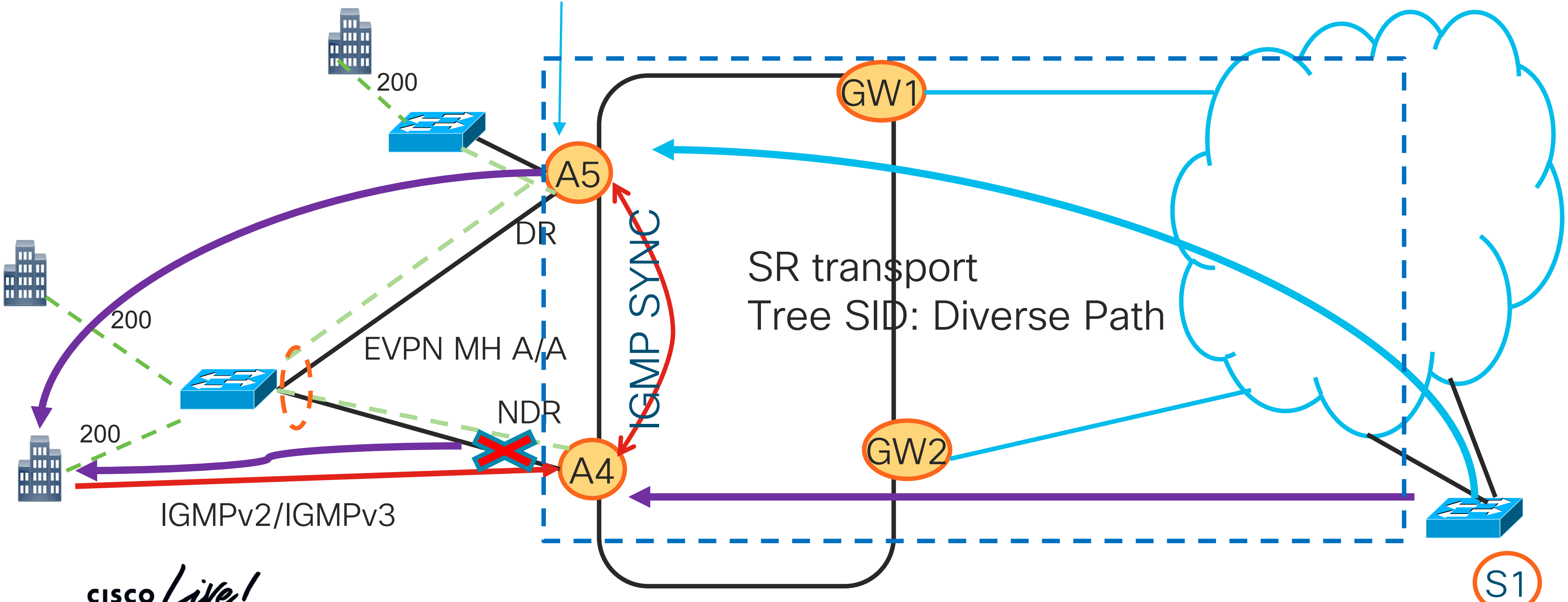

200: Mcast VLAN
IGMP on BVI
IGMP snooping with SSM
200
200
200
A5
GW1
NDR
SSM
MPLS transport
IGMP SYNC
EVPN MH A/A
DR
SSM
A4
GW2
IGMPv2/IGMPv3
S1

Multicast + Diverse Path

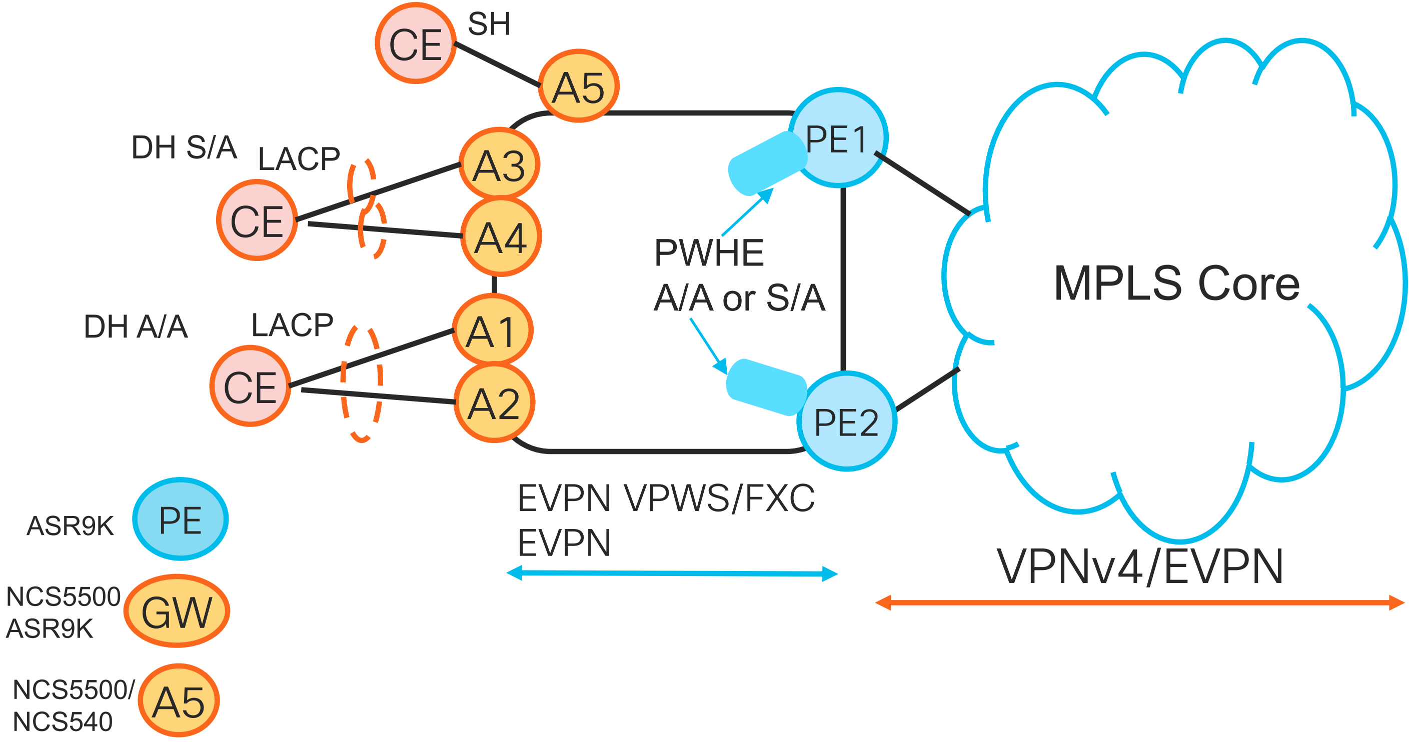
TREE-SID (7.0.1): PCE multi-IGP
- RP and S behind Root if ASM
- IPv4 only, no FRR/TILFA
- Manual config. BGP for source/RX discovery 7.3.1

200: Mcast VLAN

IGMP on BVI
IGMP snooping with SSM



Future Solution – EVPN Headend



Key Take Aways

- Objective of the design is optimize service location to bring on efficiency.
- NCS platform will perform a subset of these services.
- Centralization may bring cost savings, simplification and efficiency in QOS design.
- Achieves optimal positioning of NCS55xx for L2 transport and some L3 and ASR9K for scaled/complex L3 services

Relevant architectures for BNG transport

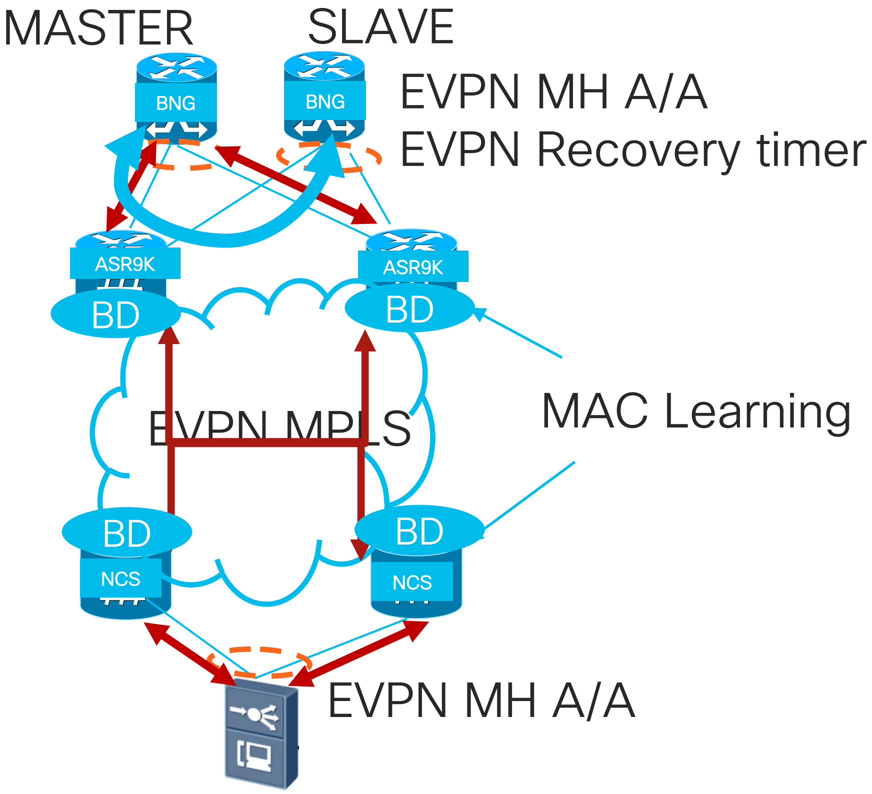
BNG Transport Service

- Centralized BNG needs traffic to be backhauled to POPs where BNG device resides.
- NCS55XX high density allows connectivity for many Access devices.

20 OLTs x 4K subs per OLT= 80,000 MAC addreses.

- NCS55XX MAC scale is 64K MAC addresses per BD and 128K MAC per box.
- Technologies that need MAC learning may prove to be challenging for certain designs and customers.
- Need to re-think with P2P technology backhaul.

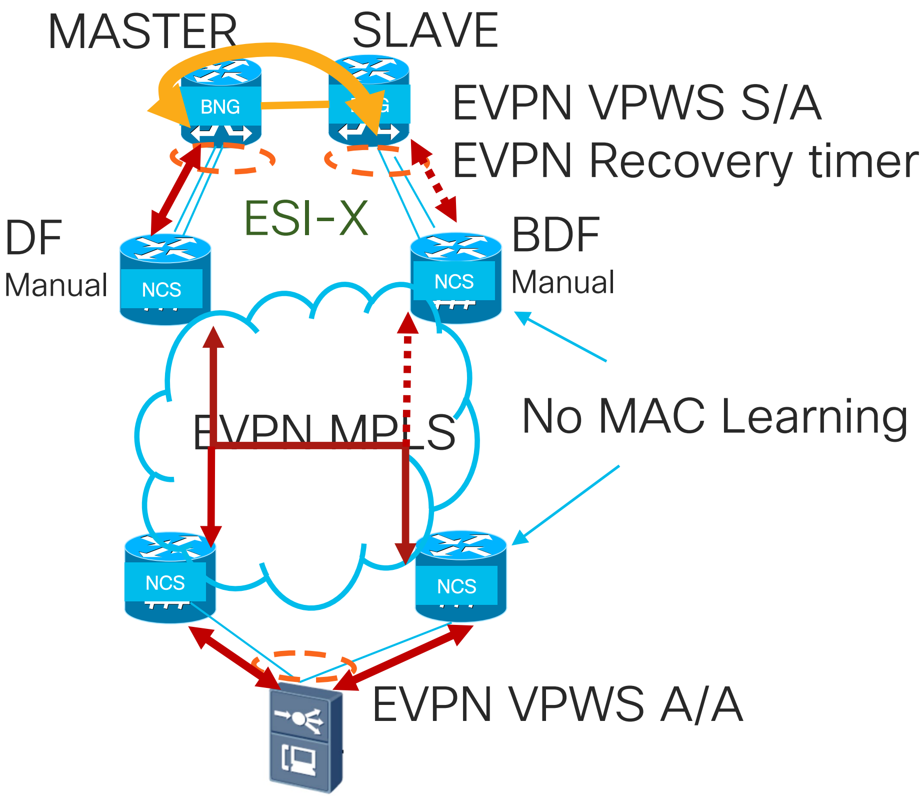
EVPN MH Active/Active – Simple Design



- All EVPN in Multihoming Active-active-mode.
- High link efficiency as load balancing occurs per Flow in all parts of the network.
- BNG keepalives share the same link (can be accross same VLAN or different).
- NCS in aggregation may not hit scale limit but the devices connecting to BNGs will do (use ASR9K with 2M MAC)

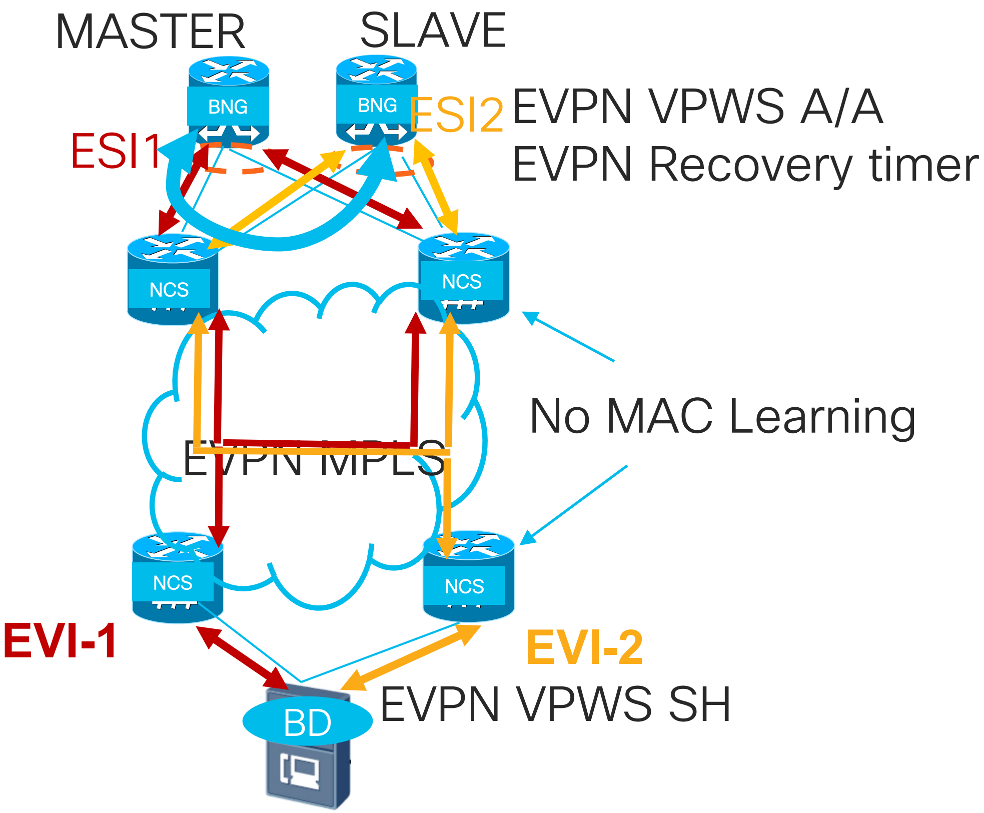
EVPN VPWS Single Active – OLT bundle

EVPN VPWS S/A 7.1.1



- Access devices will be in A/A mode, so that LB is achieved.
- BNGs will be in S/A so all traffic is directed to Master BNG.
- Additional link needed for BNG keepalive.
- In case DF link or Master BNG fails, backup DF to Slave BNG takes over.

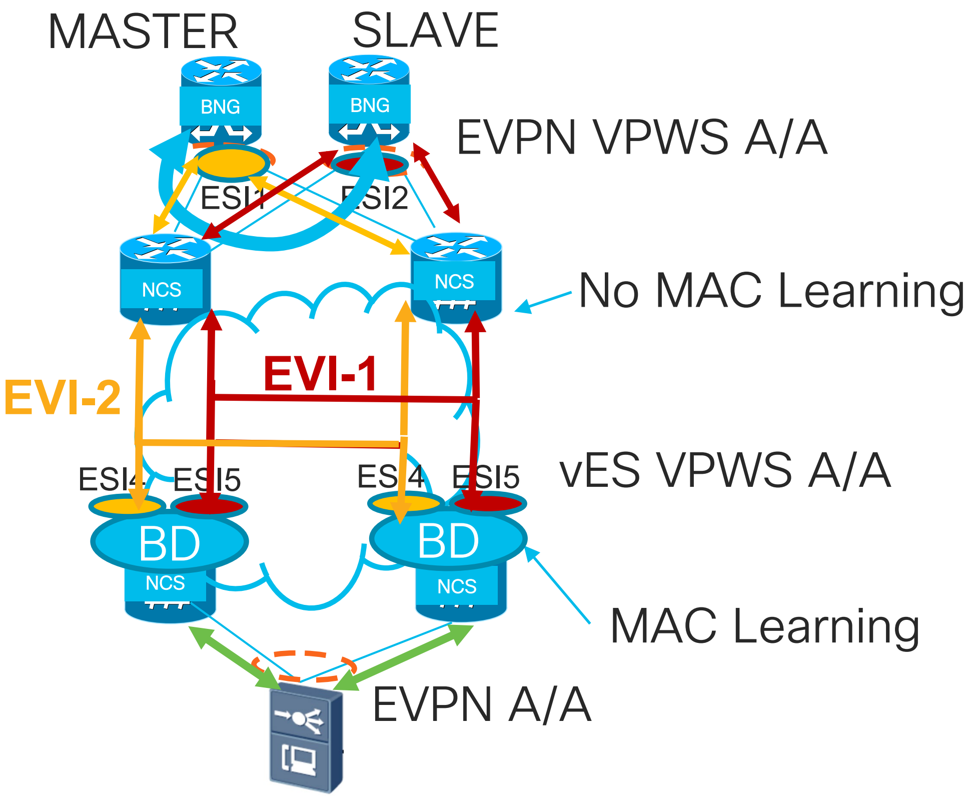
EVPN VPWS A/A– No bundle in OLT



- Access devices will do MAC learning so will point to the active Master BNG MAC.
- No LB done in Access. Need more VLANs to achieve this.
- BNGs will be in A/A so all so traffic is LB towards them.
- Master BNG will have ESI1 and Slave ESI2. No additional link needed for BNG comm (EVPN on subint)
- If Master BNG fails, OLT will learn BNG MAC across other link.

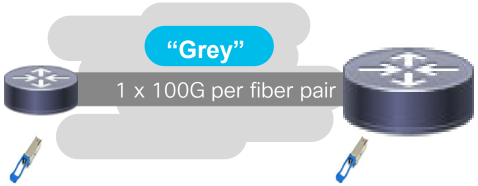
EVPN VPWS Active-Active - vES

EVPN vES A/A roadmap



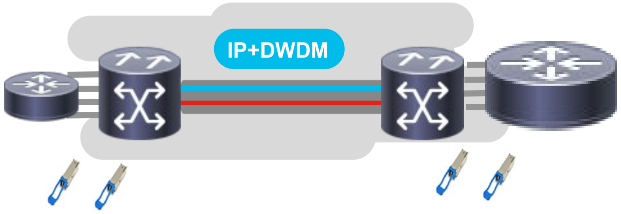
- All in A/A mode, so that LB is achieved everywhere.
- Aggregation devices will do MAC learning and have vES VPWS pointing to the active BNG.
- NCS connecting to BNGs do not need to do MAC learning.
- If Master BNG fails, aggregation will learn new BNG MAC accross remaining EVI.
- No separate link for BNG keepalives.

Transport Options for BNG backhaul



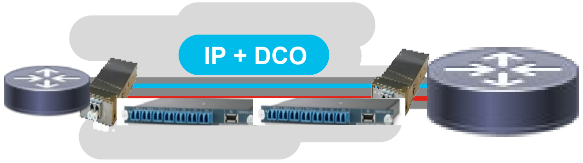
- Regular Optics
- Simpler management (No Two Networks)

- Distance Limitations (<40Kms)
- Single 100G per fiber



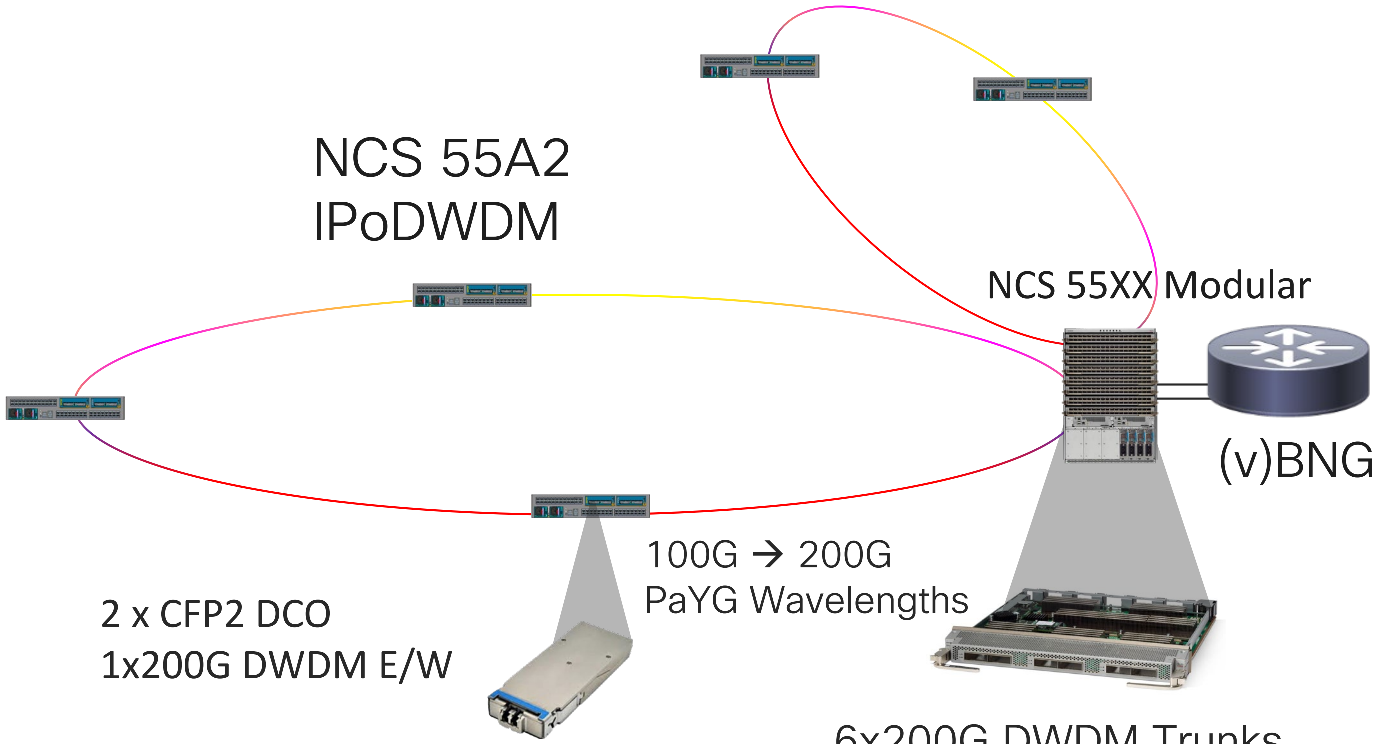
- No distance limitations (Amplifiers required in case of large distances)

- Two disparate networks: Additional Capex + Opex
- Inter-operability (IP + Optical)



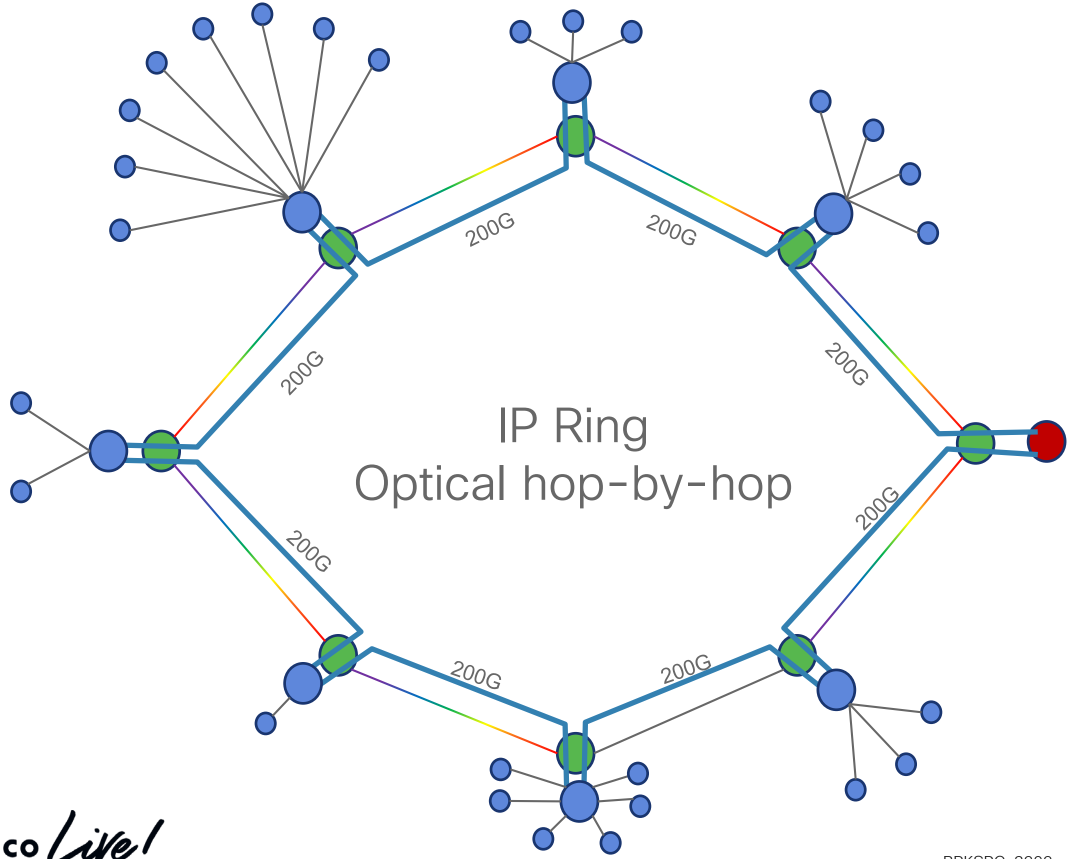
- DWDM on the DCO to carry IP traffic up to 80Kms, passive Mux/Demux.
- Simpler management, IOS-XR (No Two Networks), Fewer devices
- Deployment flexibility (PAY as you Grow 100GE→ 200GE, DCO optics where required)
- 45% TCO savings

Product Setup For Ring Aggregation



6x200G DWDM Trunks

Single Ring Topology



200G Ring BW

1 Shared 200G IPoDWDM

200G/N Gbps per 55A2

🔴 ASR 9000 BNG

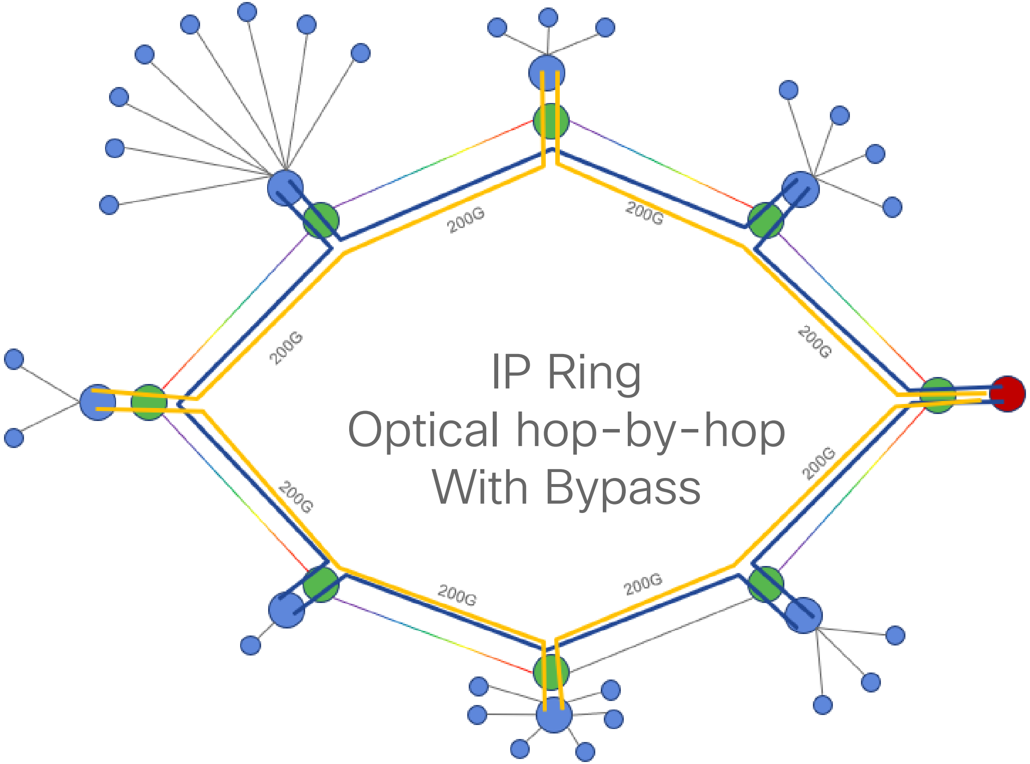
🔵 NCS 55A2

🟢 Passive mux/demux

Optical Fiber

200G IPoDWDM

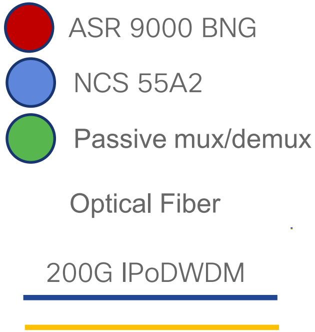
2 Sub-Rings Topology



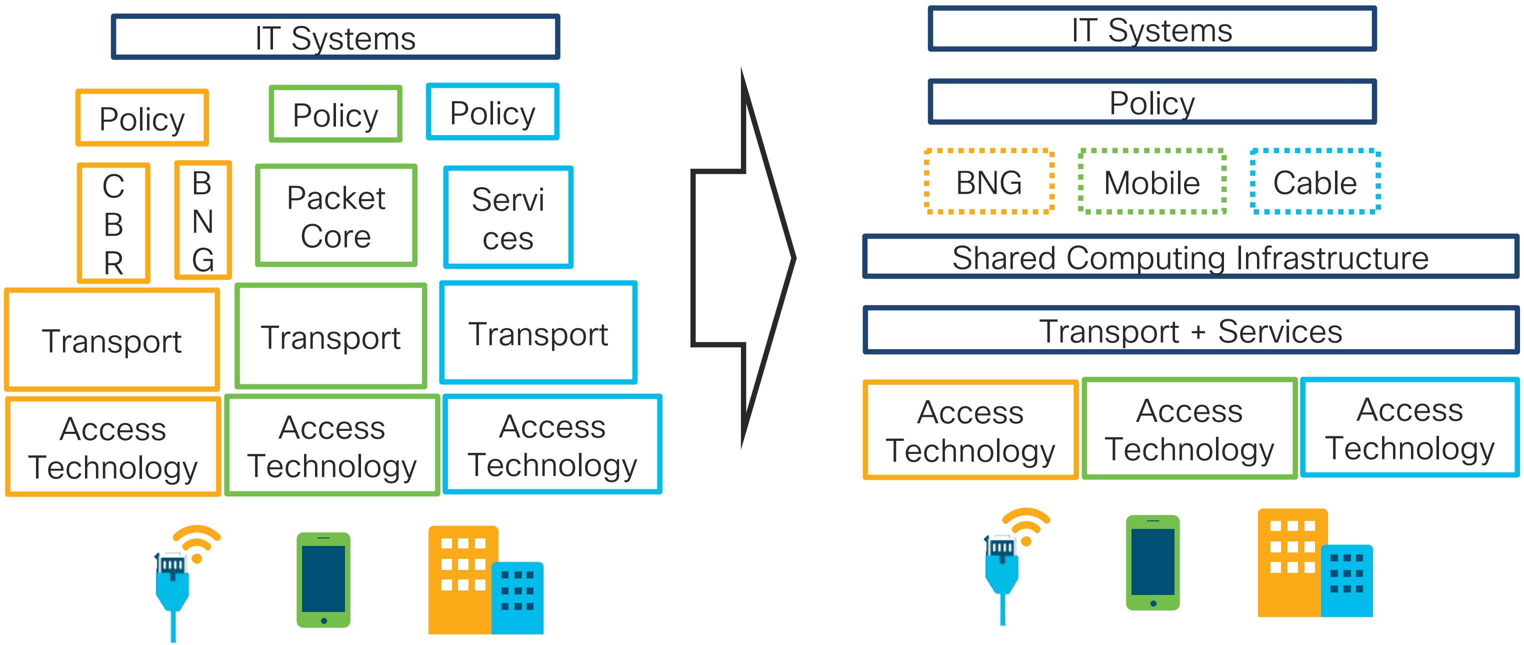
400G Ring BW

2 Shared 200G IPoDWDM

400G/N Gbps per 55A2



Subscriber Convergence Path (cnBNG)



Key Take Aways

- NCS5500 has a decent MAC scale that will take care of most customers BNG transport designs.
- EVPN VPWS is a transport option to avoid MAC learning in NCS5500 devices.
- With EVPN VPWS supported feature set, there is no complete load balancing across all parts of the network
- DCO CFP2 based optics are a cost effective option to provide long distance BNG transport
- cnBNG will bring CUPS approach integrating mobile, cable and wireline with different DP options.

NCS5500 is an optimized transport platform

- Some feature gaps or scalability may difficult design.
- SR together with ODN will solve FEC/EEDB shortage.
- EVPN will provide FHRP and MCLAG functionalities.
- If L3 services do not scale, centralize them on ASR9K
- BNG transport with EVPN P2P is the most efficient design.

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.

Continue your education

- Demos in the Cisco Showcase
- Walk-In Labs
- Meet the Engineer 1:1 meetings
- Related sessions

Thank you

You make **possible**