Advanced MPLS Deployment in Enterprise Networks

BRKIPM-3014

Patrice Bellagamba

Cisco Networkers 2007

HOUSEKEEPING

- We value your feedback, don't forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a 'No Smoking' venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

Benefits of a self-owned MPLS network

- IP Virtualisation
- Build an SP class network
- Agile handling of “Merge & Acquisition”
  - Cost
  - Ownership
- Datacenter consolidation
  - Front-end access Virtualisation
  - Network services Centralization
  - VLAN extension thru MAN/WAN
- Segmentation of the Enterprise
  - Increase Security (Closed Users Groups)
  - Worm mitigation thru isolation

Advanced MPLS for Enterprise
Agenda

This session is a companion of the ‘Architecture MPLS for enterprise’ breakout, its technical focus is:

- MPLS L3 VPN

  Branches virtualization:

  2547 over DMVPN

- MPLS L2 VPN

  Use VPLS to extend L2 between Data Center

  Without Spanning-tree extension

These are brand new solutions for Enterprise

MPLS L3 VPN
Branches virtualization

Service-provider's MPLS IP-VPN is an attractive solution for many remote sites

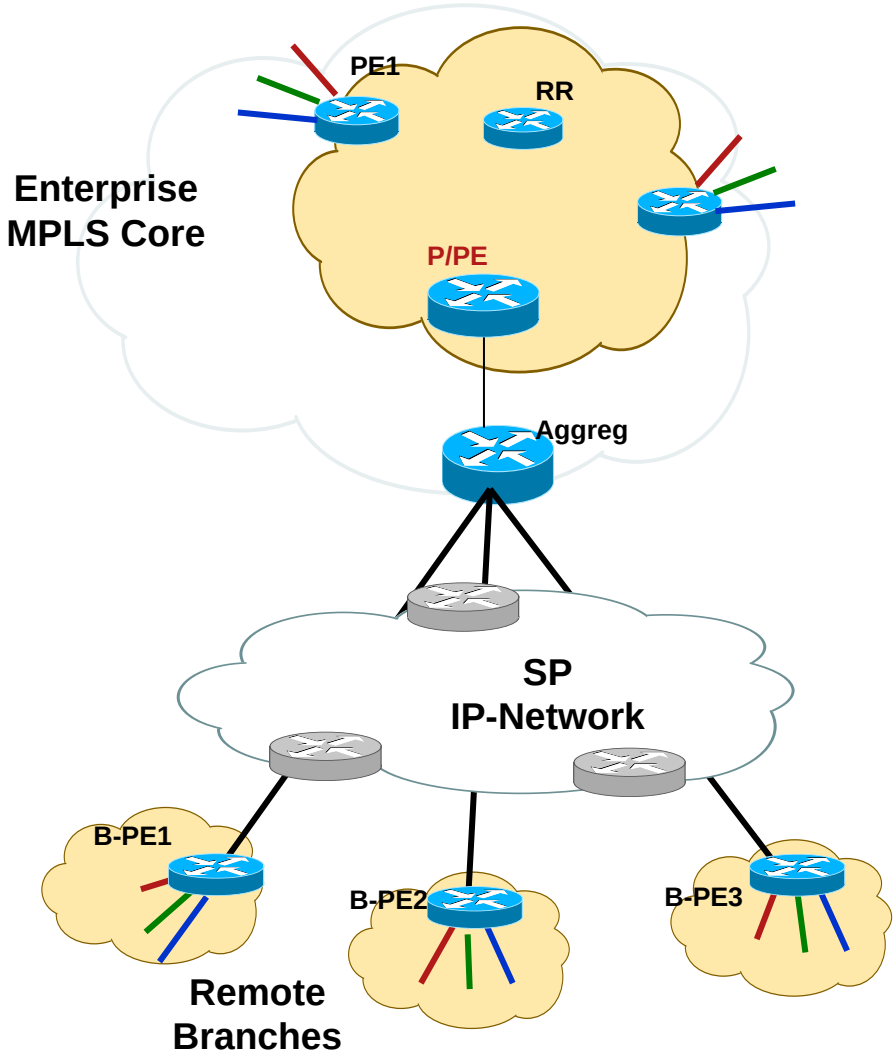
- Any Local-loop (aka xDSL, Ethernet plus any legacy links)
- Cost (Shared backbone)
- Any to any
- As secured as any L2 (Virtualization)

Alternatives to extend VRF concept over SP-owned IP-VPN:

- Multi-VRF CE
- CsC (Carrier supporting Carrier)
- Tunneling over IP
  - VRF-aware IP tunneling
  - 2547 over DMVPN

Remark: RFC 4364 replace & obsolate 2547, but function is still referenced as 2547oDMVPN on documentation

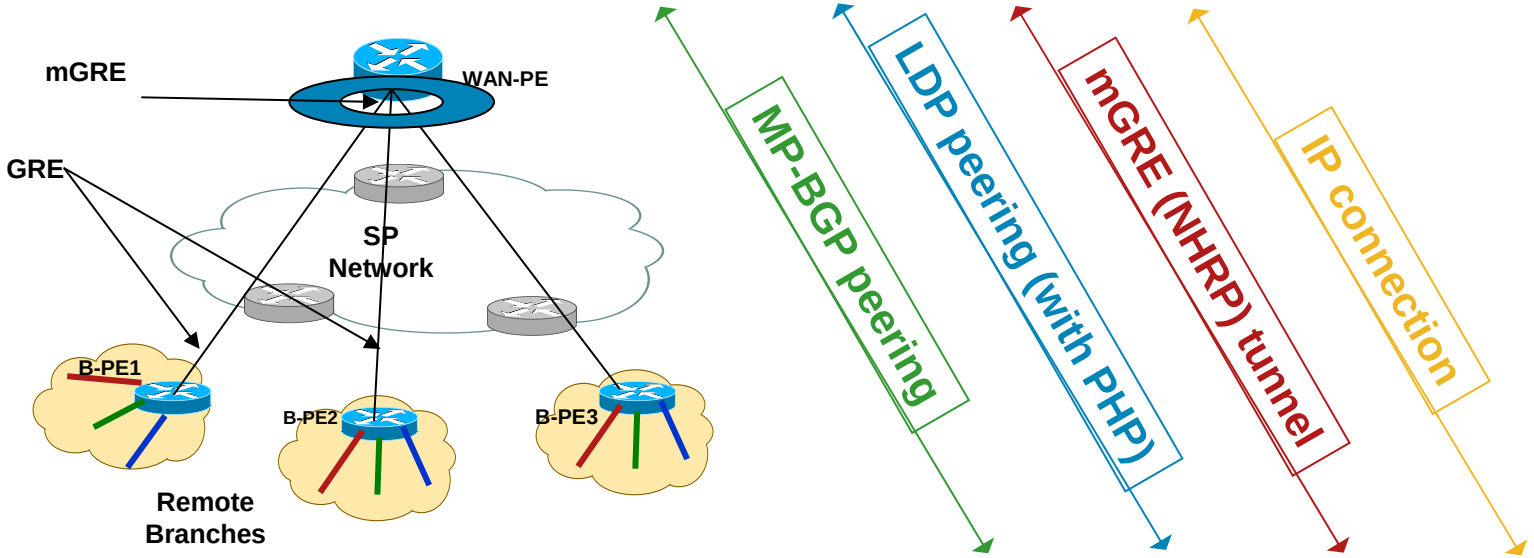
Extend Virtualization down to branches



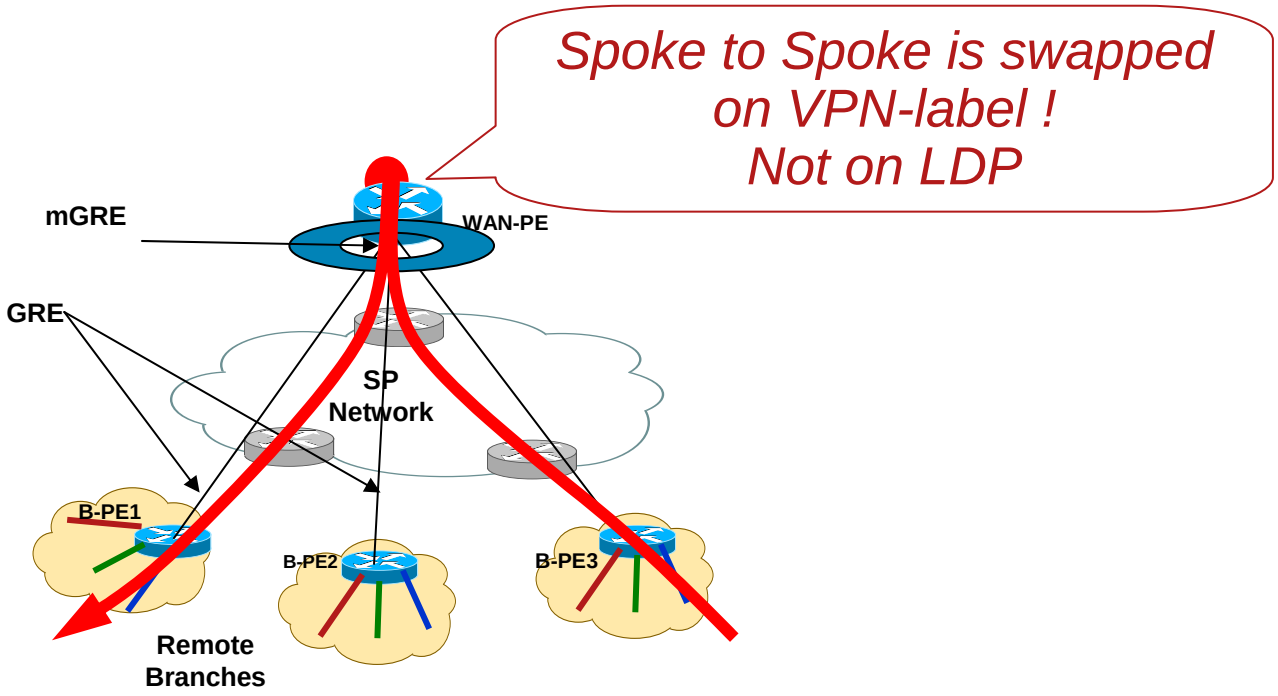
- L2 connected branches
  - L2/L3 VPN extension
  - Using dedicated AS for WAN
    Scalability
    eMP-BGP peering aka option b)

- L3 connected branches
  - SP IP-VPN offer
  - or Internet

2547oDMVPN concept

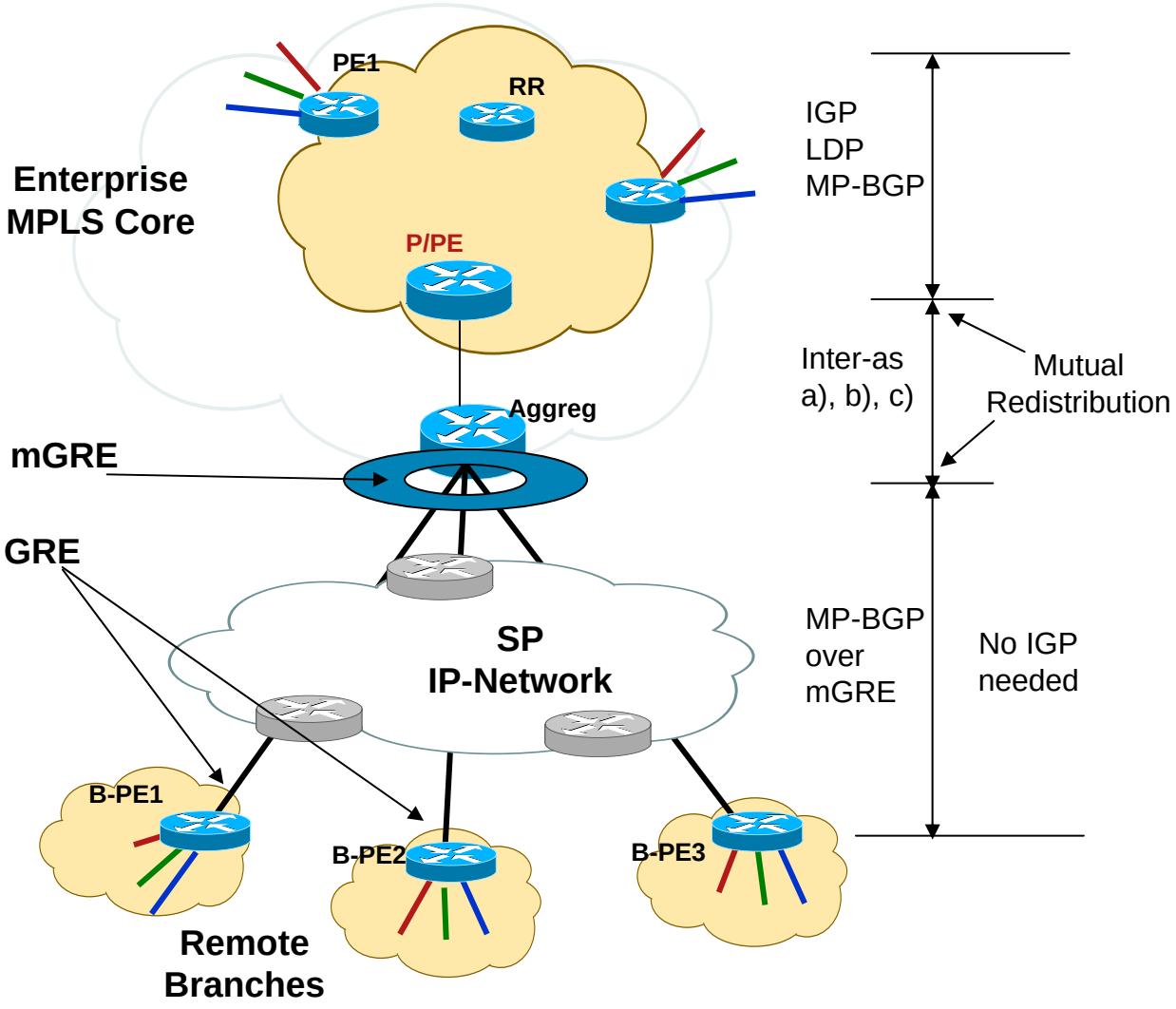


2547oDMVPN concept



No direct Spoke to Spoke connection thru mGRE

Virtualized Branch Aggregation
2547oDMVPN - Hub & Spoke



Available in
12.4(11)T
FCS - 11/06

- Hub is a PE and RR for the DMVPN network
- No direct spoke-to-spoke tunnels
- No IGP over the tunnels needed – scales better

Platforms

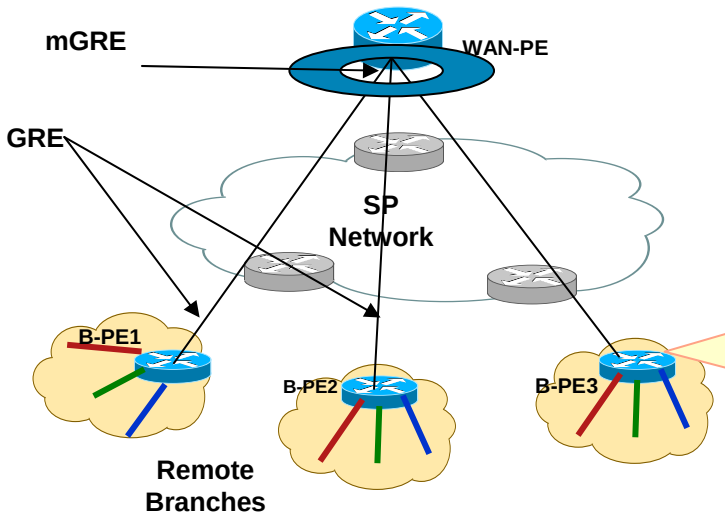
- 7200 (hub) – PE
- ISR (spoke)

2547oDMVPN - Hub & Spoke

Step 1: LDP over mGRE

```
interface Tunnel1
 ip address 10.0.0.1 255.255.255.0
 ip nhrp authentication cisco
 ip nhrp map multicast dynamic
 ip nhrp network-id 1
 mpls ip
 tunnel source fa0/0
 tunnel mode gre multipoint
 !(tunnel protection ipsec profile prof)
```

- Establish an mGRE tunnel
  - Hub&Spoke
  - No direct Spoke to Spoke
- Enable LDP over mGRE
  - Penultimate Hop Popping will force Null-label



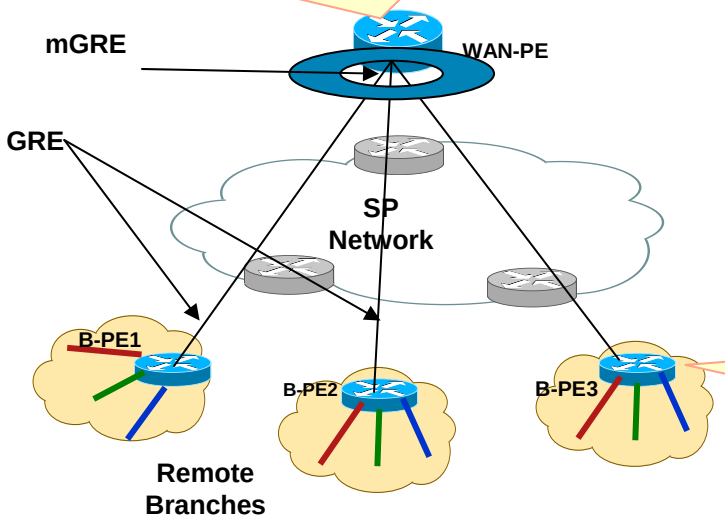

```
interface Tunnel1
 ip address 10.0.0.11 255.255.255.0
 ip nhrp authentication cisco
 ip nhrp map multicast dynamic
 ip nhrp map 10.0.0.1 172.0.0.1
 ip nhrp map multicast 172.0.0.1
 ip nhrp network-id 1
 ip nhrp nhs 10.0.0.1
 mpls ip
 tunnel source Ethernet0/0
 tunnel mode gre multipoint
 !(tunnel protection ipsec profile prof)
```

2547oDMVPN - Hub & Spoke

Step 2: MP-iBGP Hub&Spoke

```
router bgp 1
 address-family vpnv4
  neighbor SPOKE activate
  neighbor SPOKE send-community extended
  neighbor SPOKE route-reflector-client
  neighbor SPOKE route-map NEXTHOP out
   …
exit-address-family

route-map NEXTHOP permit 10
 set ip next-hop 10.0.0.1
```



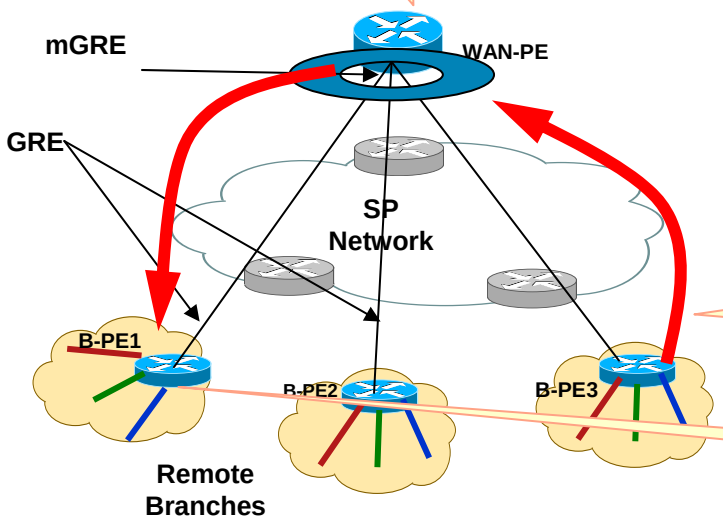
- Exchange Routes thru MP-BGP
  - MPLS L3-VPN down to branches
- Hub is next-hop-self
  - ➔This forces hub to allocate a local VPN label
- Hub is RR
  - Allows Spoke to Spoke communication thru Hub

```
router bgp 1
 address-family vpnv4
 neighbor HUB activate
 neighbor HUB send-community extended
 exit-address-family
```

2547oDMVPN - Hub & Spoke

Step 2: MP-iBGP Hub&Spoke

```
VPNv4: 1:2:192.168.12.0/24
  mpls labels in/out 21/17
  NH: 10.0.0.12
```



- Exchange Routes thru MP-BGP
  - MPLS L3-VPN down to branches
- Hub is next-hop-self
  - ➔This forces Hub to allocate a local VPN label
- Hub is RR
  - Allows Spoke to Spoke communication thru Hub

```
VPNv4: 1:2:192.168.12.0/24
  mpls labels in/out 17/aggregate(red)
  NH: 0.0.0.0
```

```
VPNv4: 1:2:192.168.12.0/24
  mpls labels in/out nolabel/21  NH: 10.0.0.1
```

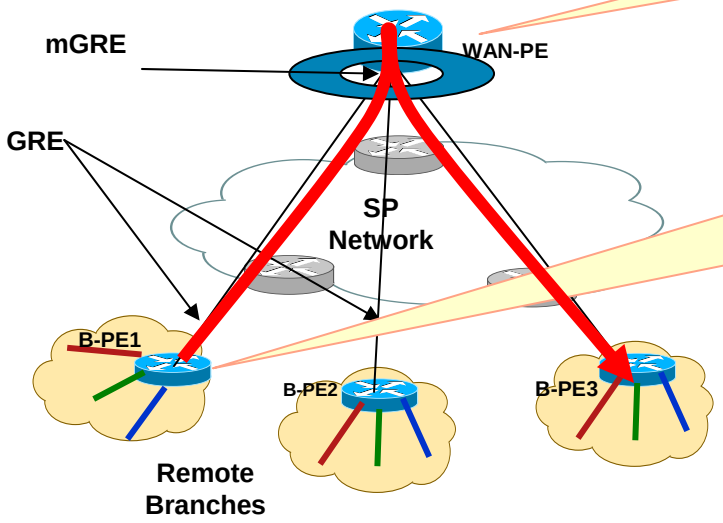
2547oDMVPN - Hub & Spoke

MPLS label swapping over mGRE

```
sh mpls forw
Local Out      Prefix          Out   Next Hop
tag   tag                      intf
21    17  192.168.12.0/24[V] Tu1   10.0.0.12
```

Hub swaps directly on the VPN labels
LDP labels are Null

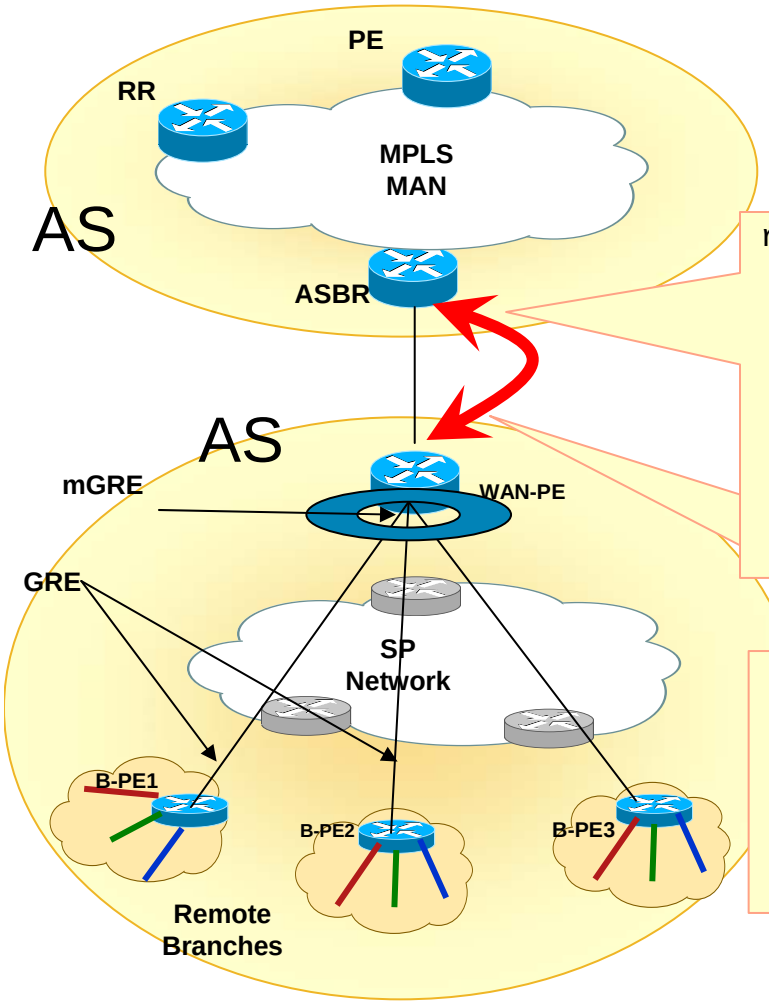
```
debug mpls pac
 MPLS packet debugging is on
  MPLS: Tu1: recvd: CoS=0, TTL=255, Label(s)= 21
  MPLS: Tu1: xmit:  CoS=0, TTL=254, Label(s)= 17
```




```
sh ip cef vrf red 192.168.12.0
 via 10.0.0.1, 0 dependencies, recursive
    next hop 10.0.0.1, Tunnel1 via 10.0.0.1/32
    tag rewrite with Tu1, 10.0.0.1, tags imposed: {21}
```

2547oDMVPN - Hub & Spoke

Step 3: Inter-AS connection



- Exchange Routes thru MP-BGP
  - Here in Inter-AS Option b)
- ASBR is Next-hop-self toward RR
  - Redistribute connected method can be used also

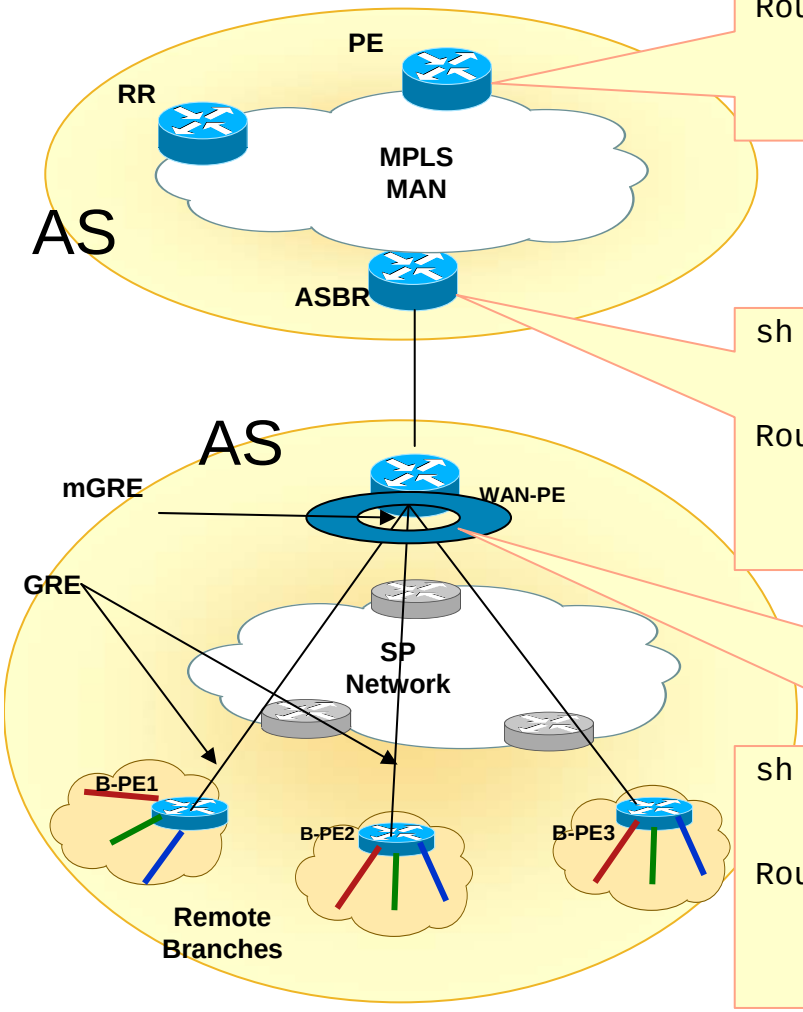
```
router bgp 4
 neighbor HUB remote-as 1
 address-family vpnv4
 neighbor HUB activate
 neighbor HUB send-community extended

 neighbor RR route-map NEXTHOP-SELF out
```

```
router bgp 1
 neighbor ASBR remote-as 4
 address-family vpnv4
  neighbor ASBR activate
  neighbor ASBR send-community extended
```

2547oDMVPN - Hub & Spoke

Step 3: Inter-AS connection




```
sh ip bgp vpnv4 all label
   Network          Next Hop      In label/Out label
Route Distinguisher: 1:2 (red)
   192.168.5.0      0.0.0.0         20/aggregate(red)
   192.168.11.0     172.0.1.4       nolabel/29
```

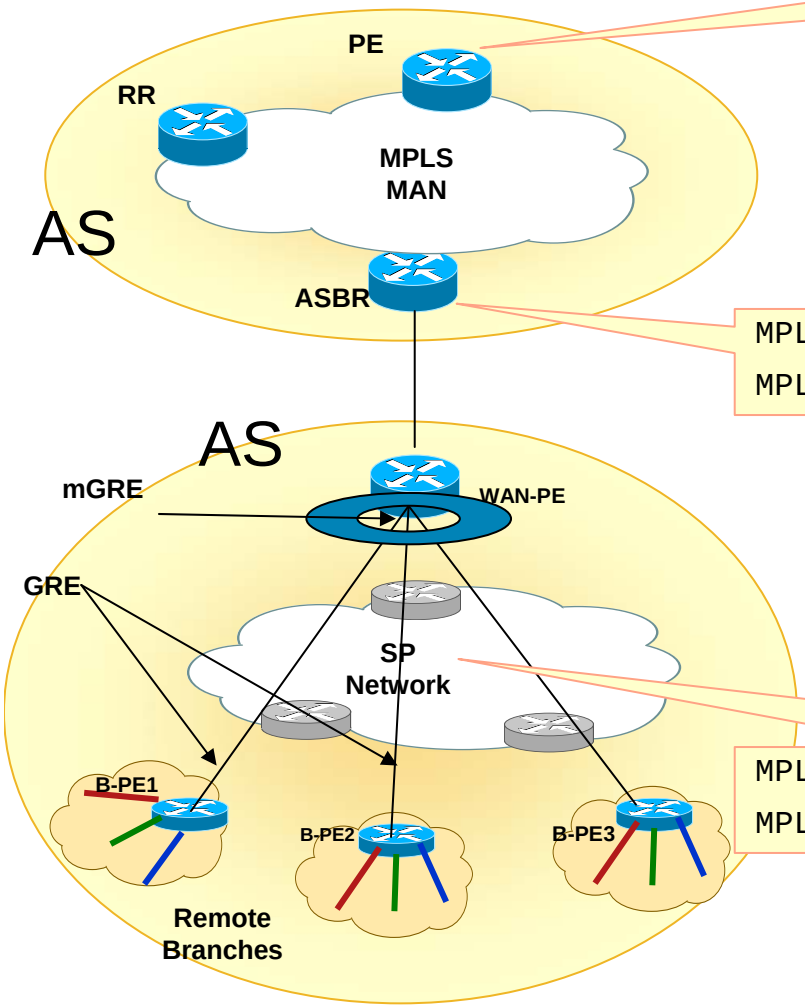
```
sh ip bgp vpnv4 all label
   Network          Next Hop      In label/Out label
Route Distinguisher: 1:2 (red)
   192.168.5.0      10.9.9.5        32/20
   192.168.11.0     172.0.1.1       29/27
```

```
sh ip bgp vpnv4 all label
   Network          Next Hop      In label/Out label
Route Distinguisher: 1:1 (blue)
   192.168.5.0      172.0.1.4       33/32
   192.168.11.0     10.0.0.11       27/20
```

2547oDMVPN - Hub & Spoke

Step 3: Inter-AS connection

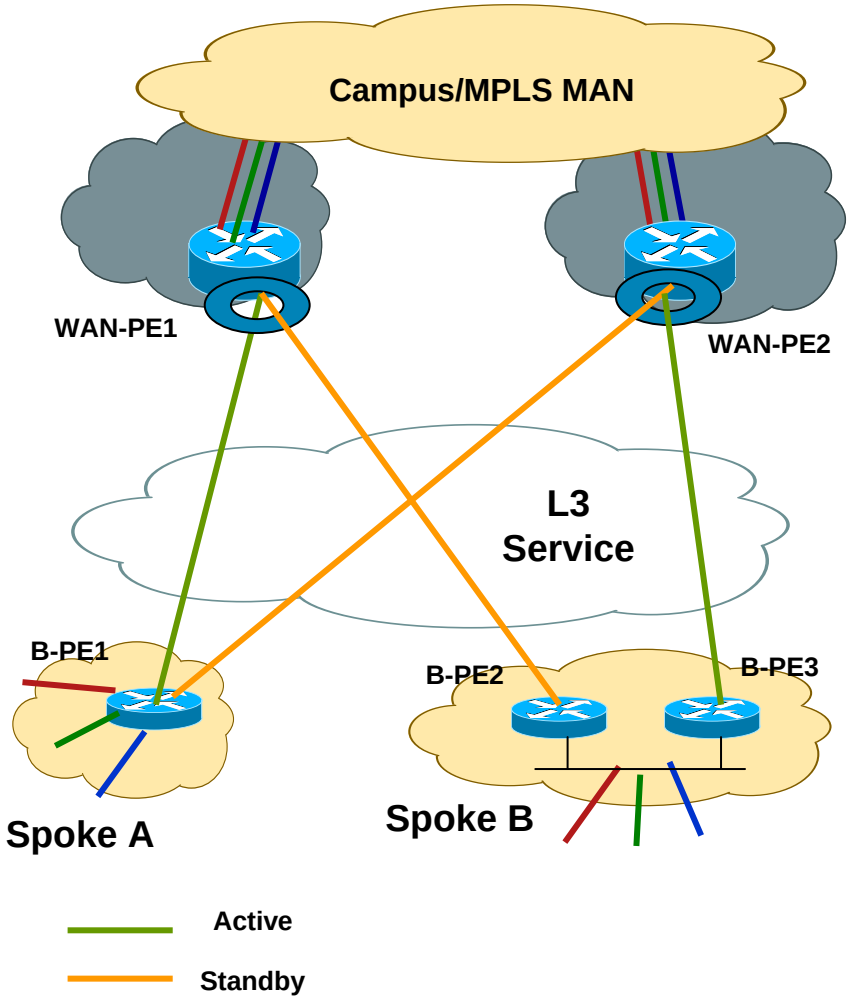
```
ping vrf red 192.168.11.2
  !!!!!
```




```
MPLS: Et0/0: recvd: CoS=0, TTL=255, Label(s)=29
MPLS: Et1/0: xmit: CoS=0, TTL=254, Label(s)=27
```

```
MPLS: Et1/0: recvd: CoS=0, TTL=254, Label(s)=27
MPLS: Tu1: xmit: CoS=0, TTL=253, Label(s)=20
```

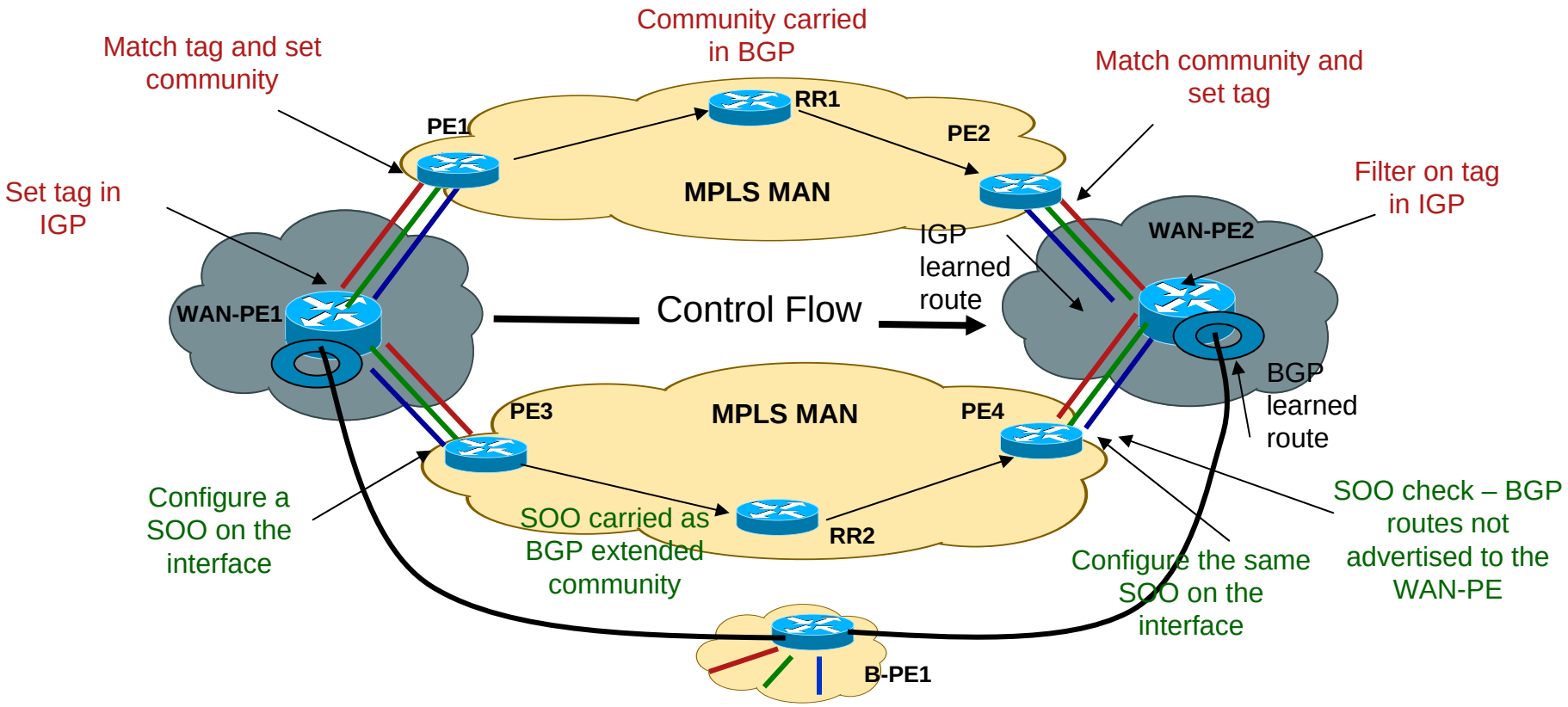
2547oDMVPN - Hub & Spoke
Redundancy Options



- Use dual/multiple hubs
  - Each is a RR
  - No BGP peering mandatory between them
- Use dual tunnels at the spokes
  - In active/standby mode
  - Control the activeness by changing the metric of the VPNv4 routes advertised
- Similar models can be used when Internet is used as a backup

2547oDMVPN - Hub & Spoke
Loop Prevention

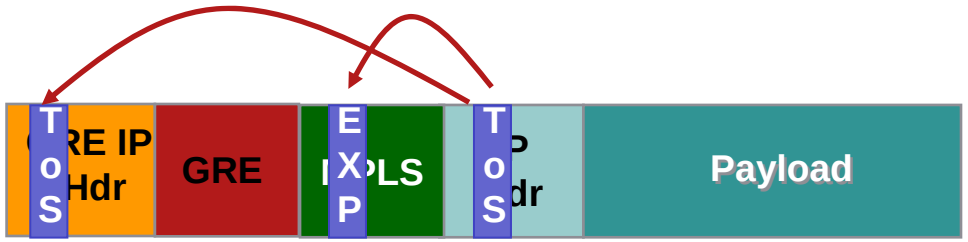
- **Active/active tunnel can create routing loops when going across MPLS MAN**
- **The IGP learned route over MPLS MAN may get preferred at the hub**
- **Filter using IGP tags and BGP communities**
- **Use BGP SOO to prevent route advertisement**



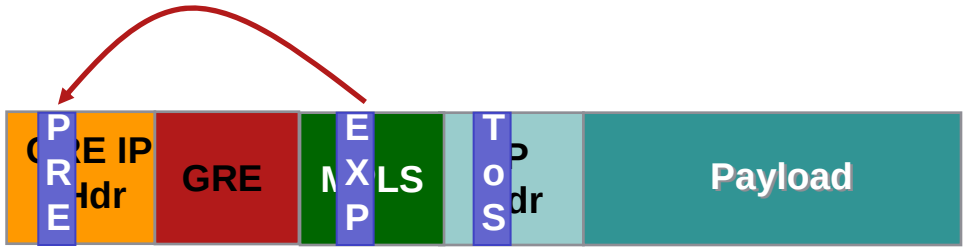
2547oDMVPN - Hub & Spoke
Implementing QoS

- Normal IP/GRE/DMVPN based QoS recommendations apply
- The MPLS VPN label is never exposed at the outgoing interface
- Per spoke QoS at the hub not supported

Hub ←→ Spoke



Spoke ←→ Hub ←→ Spoke



2547oDMVPN - Hub & Spoke
Enabling Multicast

- Use Multicast VPNs (MVPN) across DMVPN cloud
  - MVPN is using Multicast-GRE encapsulation
- Within the VRF:
  - – RP can reside at the hub closer to the sources
  - – RP should be reachable by each spoke from within the VRF
- In global space:
  - – Make the hub a RP

Caveat

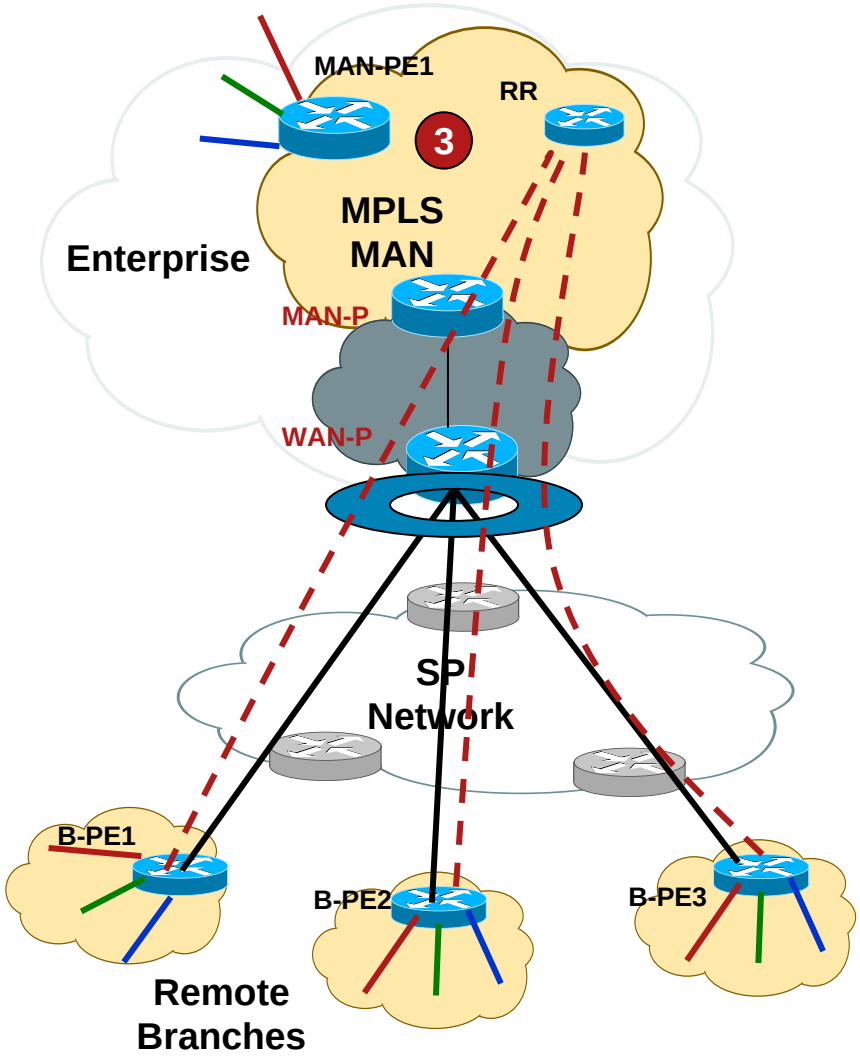
- Hub has to do replication for all the spokes

Restrictions

- Spokes can not have the sources

2547oDMVPN - Hub & Spoke

Hub as a 'P'



- Performs label switching of packets hub <-> spoke and spoke <-> spoke
- Global IGP extended to the spokes for RR and MAN PE reachability
- Spokes become MP-BGP clients of MAN RR

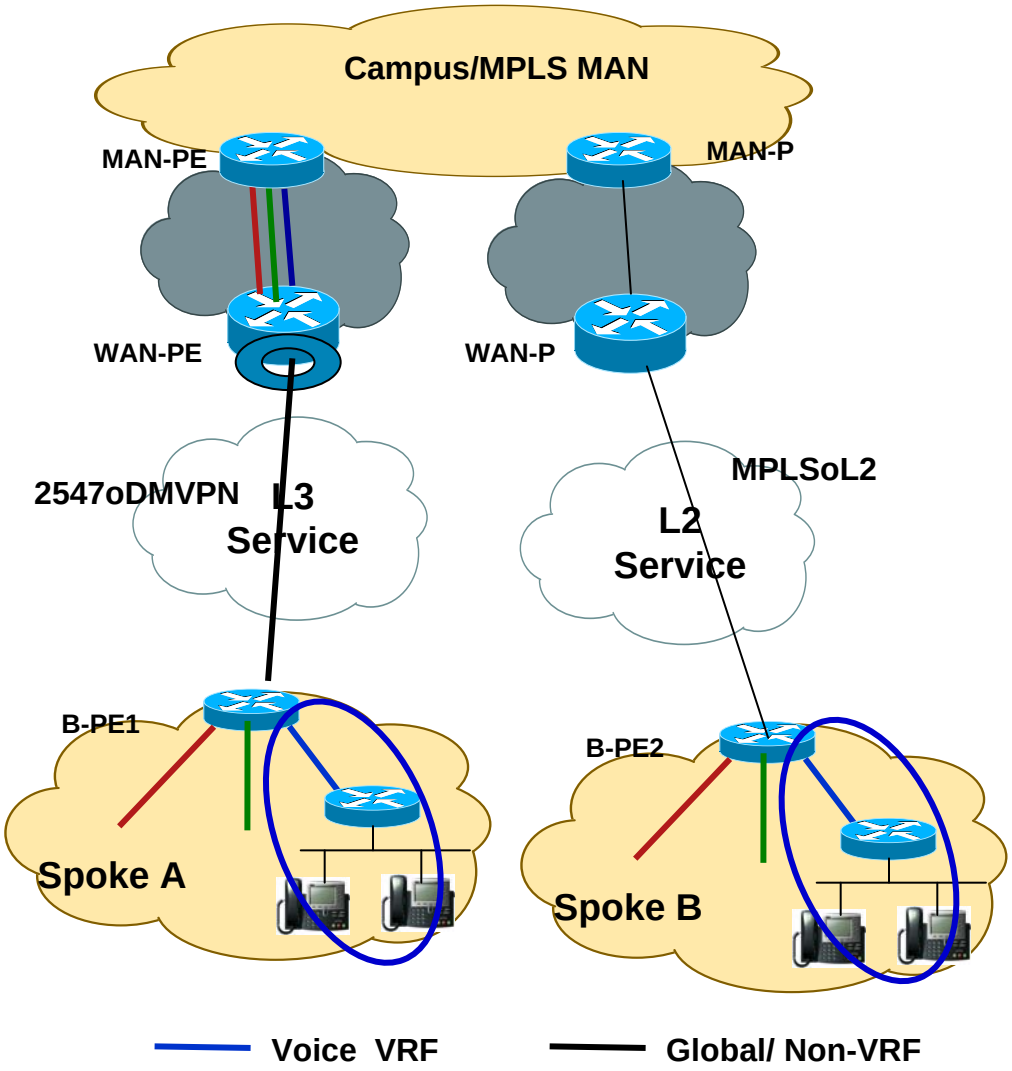
Advantages

- No VPN info required at hub
- Creates a truly integrated MAN/WAN MPLS network

Status

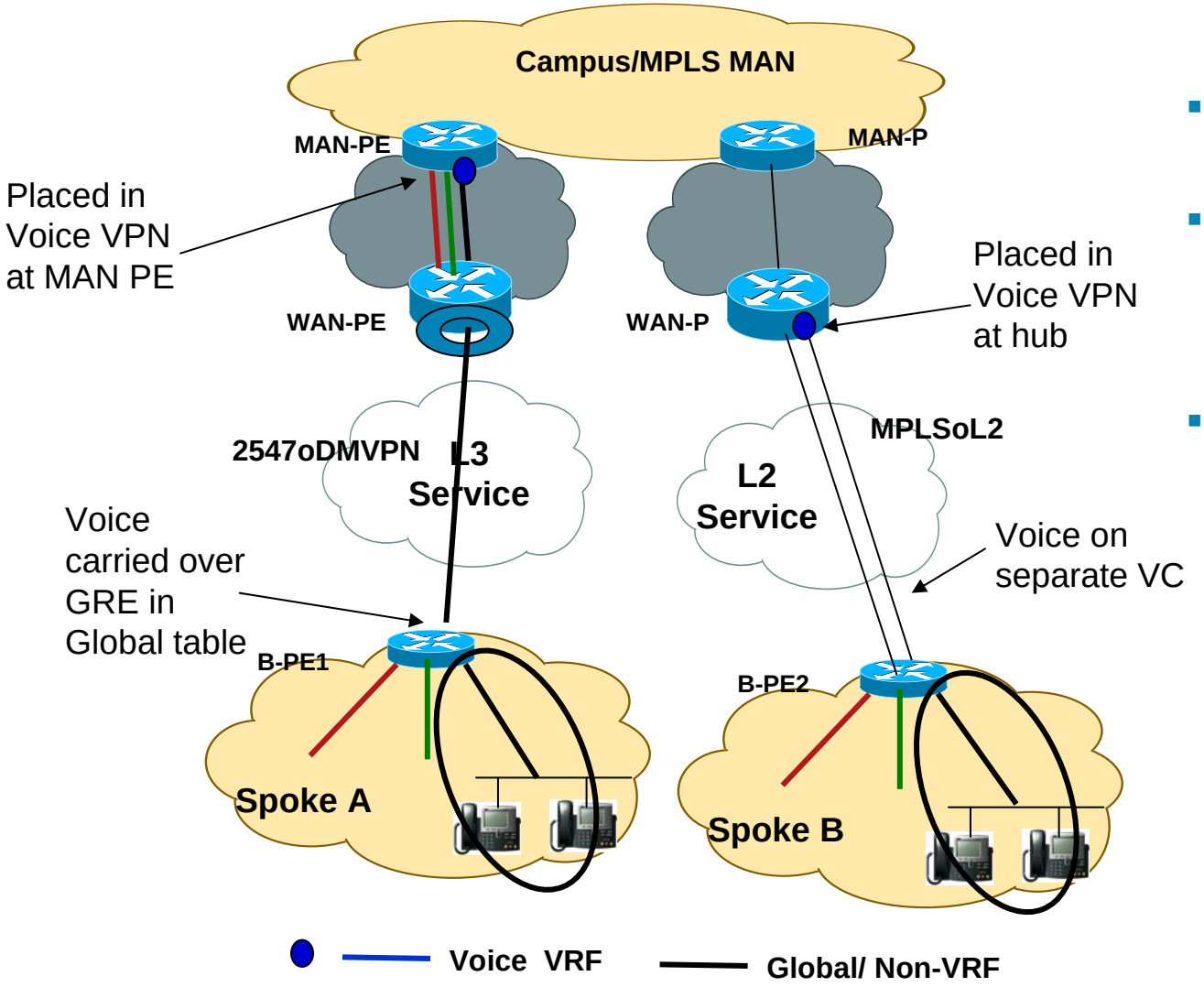
- Requires LDP and tag-switching support on mGRE

Voice and VPNs
Voice in a VRF at the Branch



- Voice traffic placed in VRF at the branch
- Transported to the hub via MPLS
- Requires dual routers at each branch
  - VRF-aware Integrated Voice services is not supported
- Expensive option – may be useful for large branches only

Voice and VPNs
Voice in Global



- Voice remains in global table at the branches
- Can be put into the Voice VPN at the hub depending on the overall voice architecture
- Adds more complexity to WAN virtualization

Advanced MPLS for Enterprise
Agenda

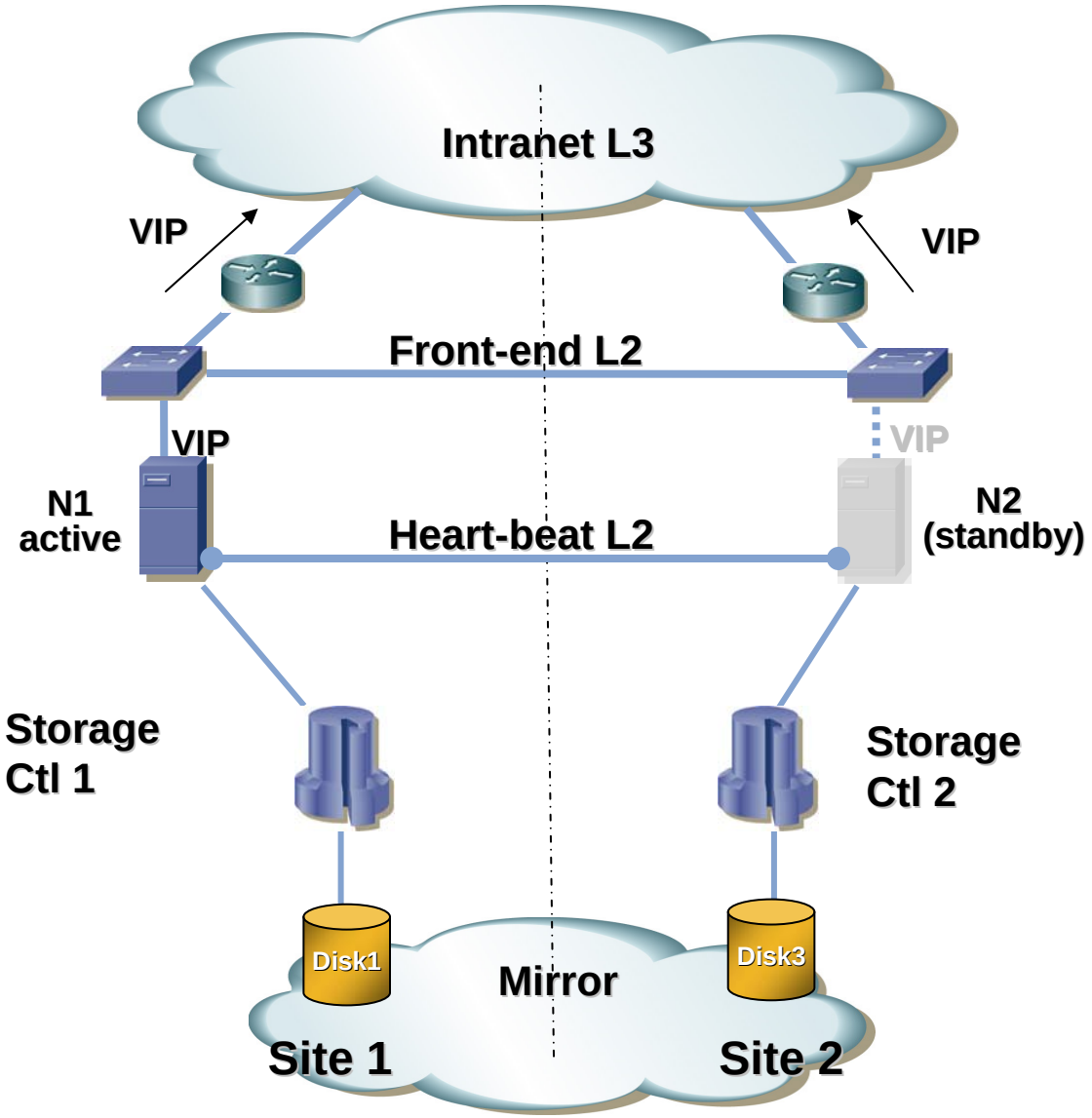
This session is a companion of the ‘Architecture MPLS for enterprise’ breakout, its technical focus is:

- MPLS L3 VPN
  - Branches virtualization:
    - 2547 over DMVPN
- MPLS L2 VPN
  - Use VPLS to extend L2 between Data Center
    - Without Spanning-tree extension

Extended L2 between Data-Centers
Motivations

- **Migration purposes**
  - **Legacy Applications where the IP parameters can not be easily modified.**
  - **Move a portion of the farm**
- **Geoclusters or geographically dispersed HA Clusters**
  - **Heartbeat**
  - **VIP**
- **Geographically dispersed Network Services**
  - **Statefull Failover**
  - **Conns and Sticky Replication**

Cluster Geographically Dispersed



Requirements:

* L2 node to node
  (VIP + Bck-Up HB)
* 100% resilient
  (no split support)

L2 extension options:

* Dedicated Fiber
  - xWDM
  - Ethernet
* Dedicated Infrastructure
  (Octopus – WRiedel)
* Intranet L2/L3-Mix
* Intranet MPLS
  - EoMPLS
  - VPLS
* Intranet IP
  - L2TPv3
  - …

Why L2-Core using MPLS

- Two main improvements can be done using MPLS
  - Core Spanning-tree suppression
    - Core links are protected via MPLS L3 convergence
      - ➔ *Stability & Fast-convergence including FRR*
  - Inter-DC Spanning-tree suppression
    - Introducing some complexity into the core architecture may lead to loop-free interconnection
      - ➔ *each DC STP will be isolated from each others*

Spanning-tree and Extended L2 concerns

Main concerns with redundant L2 vlan extension

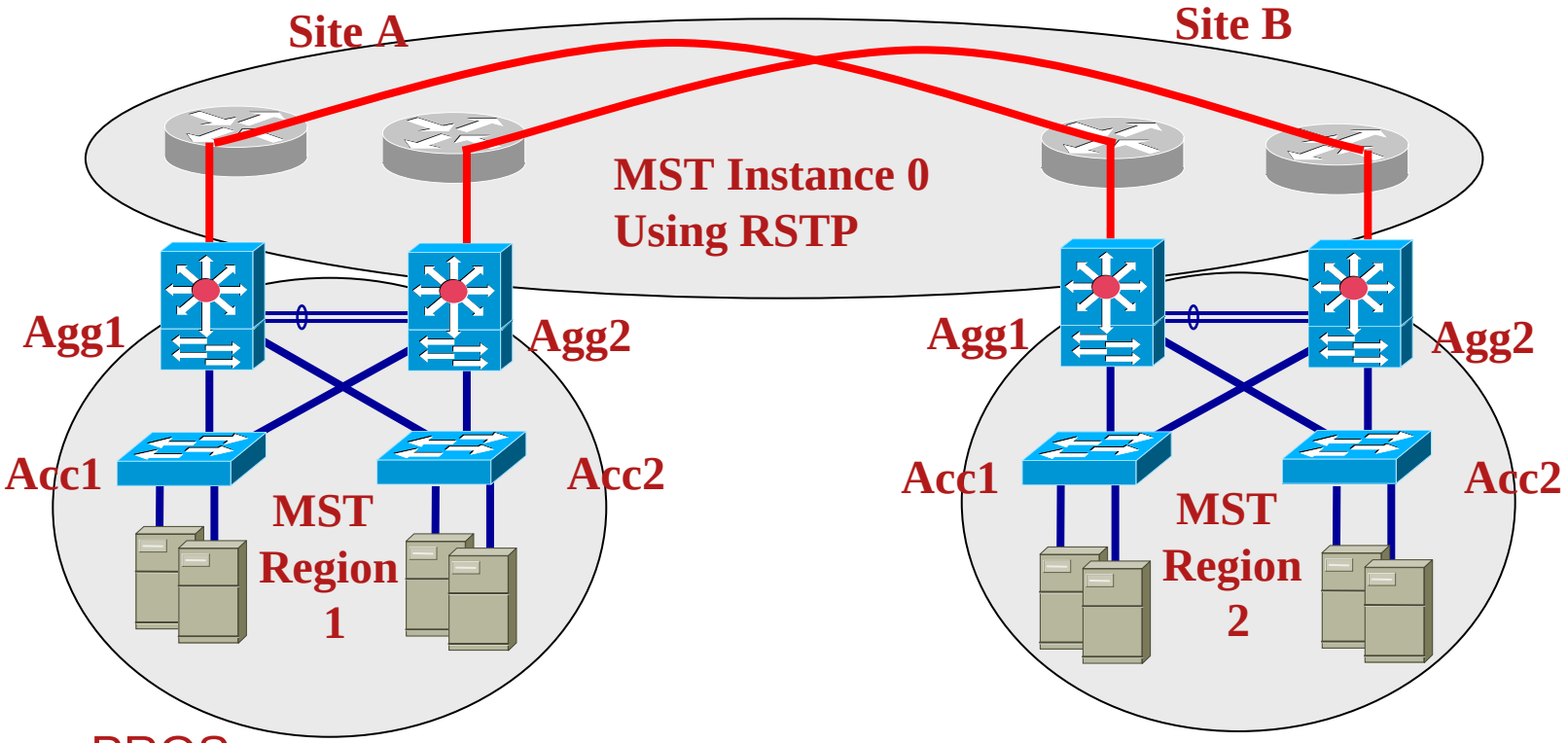
1. Spanning-tree architecture becomes complex and fragile when diameter / topology becomes complex
2. STP convergence on one DC affects all other DCs
3. Goal is to avoid STP interconnection between Data-Centers
   - ➔ But, STP is enabled to prevent any loop, thus when suppressing STP, architecture must ensure loop-free forwarding
4. Broadcast Storm beyond STP domain
   - ➔ Isolating STP doesn't mean blocking storm

MPLS technologies to extend L2

- **EoMPLS**
  - **Technology:**
    - •Cross-connect port to port thru PW (Pseudo-Wire)
  - **Main positionning:**
    - •Extend a few VLAN between two DC
    - •Not adapted to multiple sites interconnection
    - •Low cost solution
      - Using the DC Aggregation switches as EoMPLS PE (PFC controlled)
      - But, usage of a L2 dedicated device would improve operational separation
- **VPLS**
  - **Technology**
    - •Bridge VLAN over PW (Pseudo-Wire)
  - **Main positionning :**
    - •Create a new concept of « L2-Core »
      - ➔ Multiple sites interconnection
    - *Intended to allow extension of required L2 VLAN between sites*

L2 extension Loop Prevention

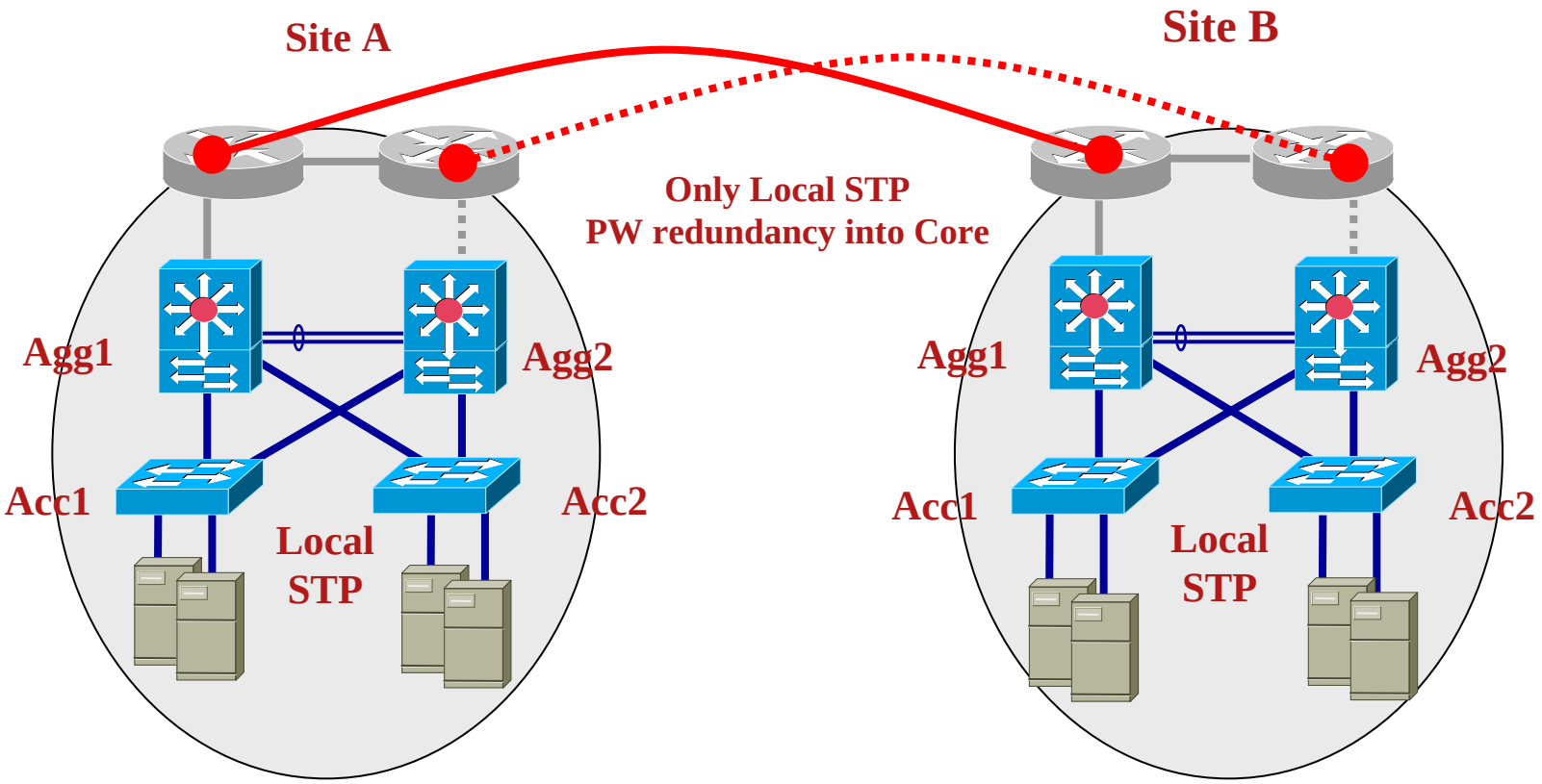
MST over EoMPLS



- PROS:
  - DC STP convergence is local
  - RSTP Fast convergence for Core
- CONS:
  - Impact all aggregation switches
  - Instance 0 exists on all ports
  - Any non-MST switch belong to Instance-0

L2 extension Loop Prevention

Anycast-PW or PW-Redundancy over VPLS



- PROS:
  - Total DC STP independance
  - No MST
- CONS:
  - Requires N-PE boxes
  - More complex MPLS core (TE / Anycast-PW)

PE choice

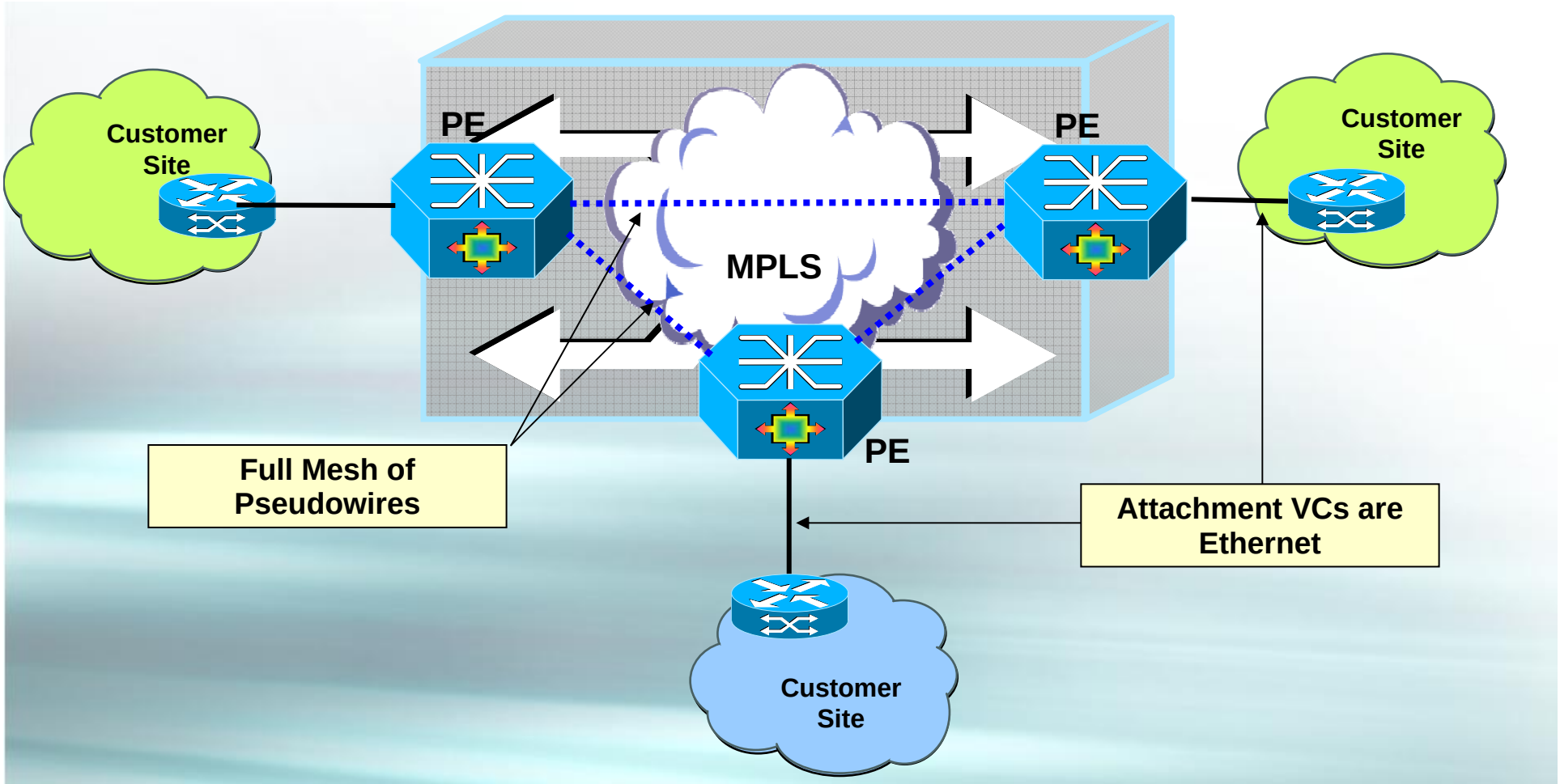
- **In MPLS L3VPN there are two options:**
  - Push PE functions into Aggregation switch
    - Catalyst 6500
  - Install an independent PE in DC core
    - Cisco 7600
- **EoMPLS:**
  - If a L3 PE is in front-end of the Aggregation switches
    - then EoMPLS per port into the PE is best choice
    - Cisco 7600
  - If the aggregation switch is the PE
    - then EoMPLS with loopback cable to isolate both function is best
    - Catalyst 6500
- **VPLS**
  - A dedicated N-PE is a must
  - Cisco 7600 with SIP is required and adds value

SIP or no-SIP

- Standard LAN card (PFC switched):
  - MPLS label switching
  - MPLS L3 VPN (VRF)
  - EoMPLS
    - Port mode
    - Sub-interface mode
  - Ingress dual-rate policer
  - Simple QoS
    - (no per port shaping, a few CoS)
  - Traffic-Engineering
  - FRR (but at 200/500ms today)
  - IGP fast-convergence (200/500ms)

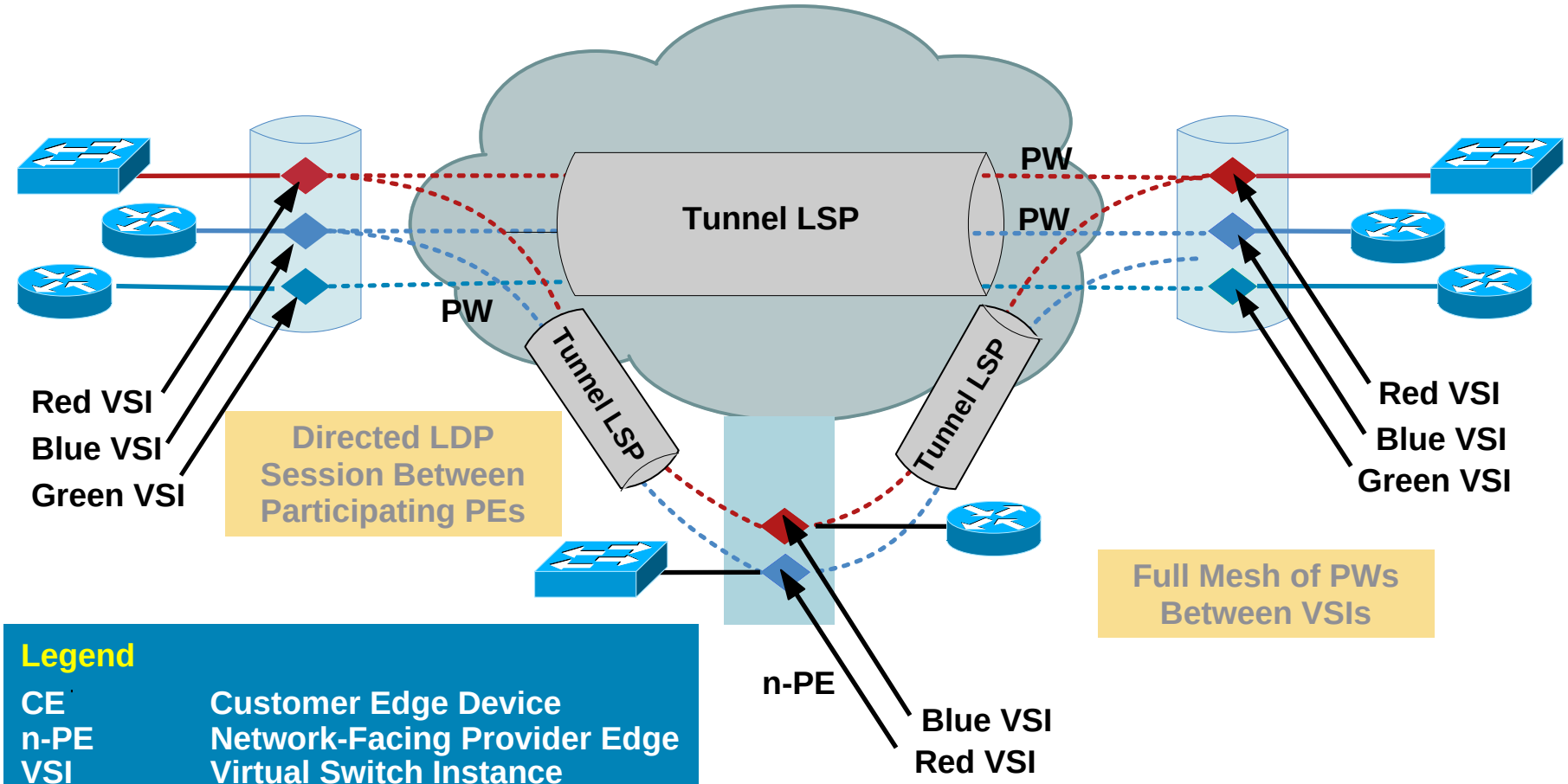
- SIP card: adds-on
  - L2VPN local-switching
  - EoMPLS internal VLAN mode
  - **VPLS** (VFI)
  - Core link shaping
    - (mandatory when buying sub-rate from SP)
  - Sophisticated QoS
    - (hierarchical shaping, 64 CoS)
  - 50ms **FRR**
  - MPLS over GRE

VPLS Reference Model



A full mesh of Pseudowires (PWs) is used to connect all Provider Edge (PE) devices which support a given VPLS VPN

VPLS Components



Legend

CE	Customer Edge Device
n-PE	Network-Facing Provider Edge
VSI	Virtual Switch Instance
PW	Pseudowire
Tunnel LSP	Tunnel Label Switch Path That Provides PW Transport

VPLS: Layer 2 Forwarding Instance *"VFI"*

A Virtual Switch MUST operate like a conventional L2 switch!

Flooding / Forwarding:

- MAC table instances per customer and per customer VLAN (L2-VRF idea) for each PE
- VSI will participate in learning, forwarding process
- Uses Ethernet VC-Type defined in pwe3-control-protocol-xx

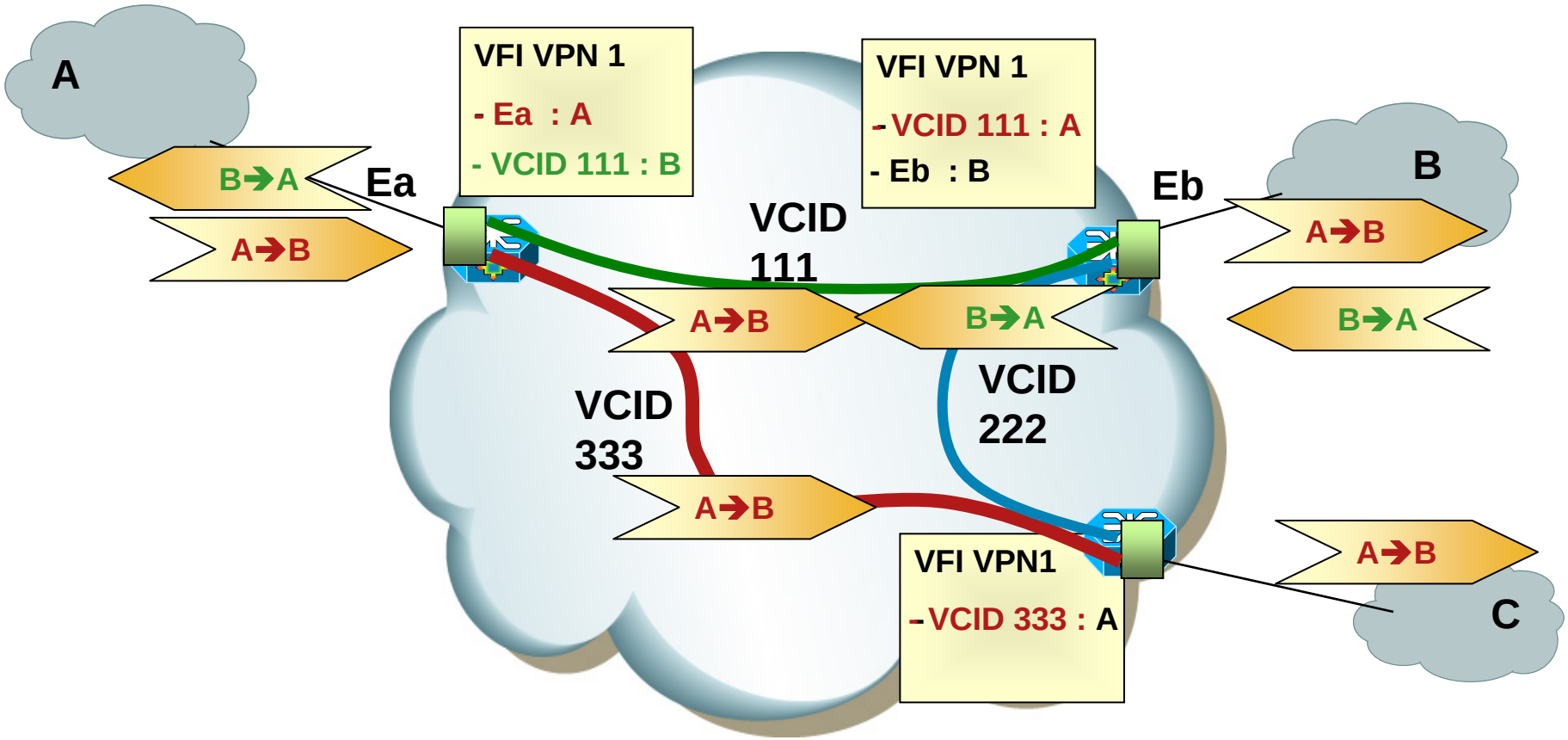
Address Learning / Aging:

- Self Learn Source MAC to port associations
- Refresh MAC timers with incoming frames
- New additional MAC TLV to LDP

Loop Prevention:

- Create partial or full-mesh of EoMPLS VCs per VPLS
- Use "split horizon" concepts to prevent loops
- Announce EoMPLS VPLS VC tunnels

VPLS L2signalling and forwarding
aka Transparent-Bridging



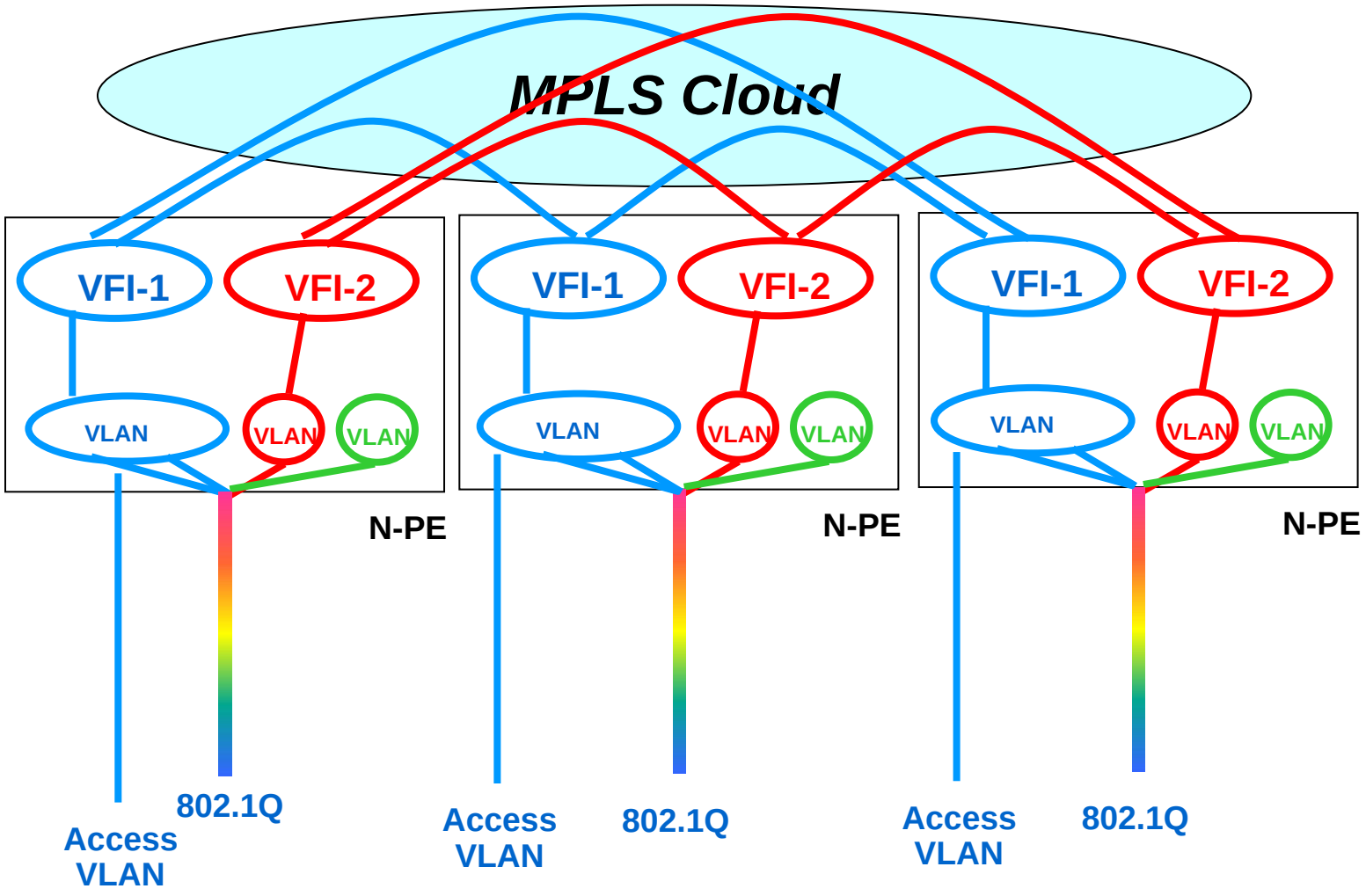
VPLS design concepts

PW (Pseudo-Wire) full-mesh

Cloud Split-Horizon

Virtual Forwarding Interface

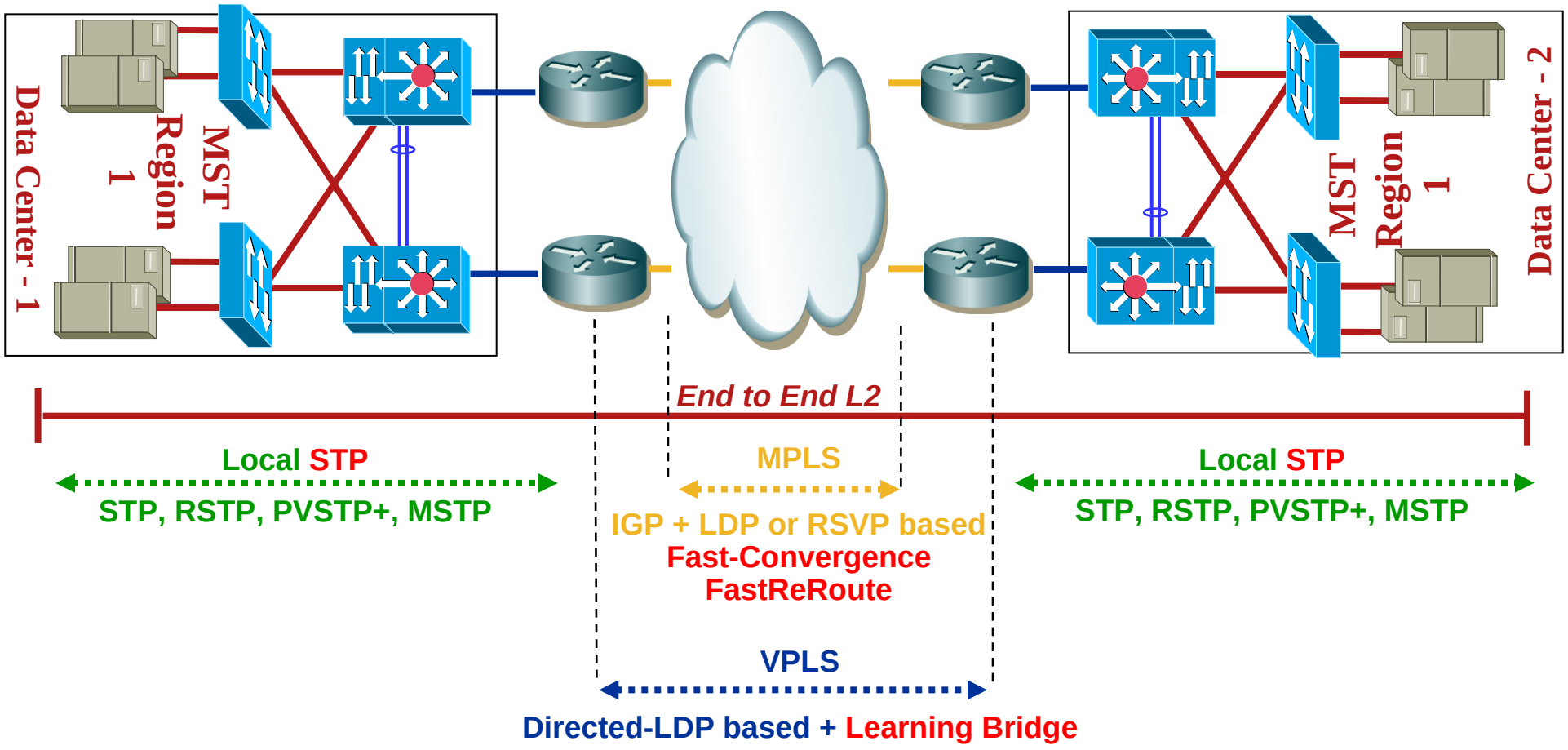
Xconnect



Scalability

- **Example for Cisco 7600-Sup720B**
  - **Up to 4000 VSIs are supported.**
  - **Up to 60 remote peers per VSI**
  - **Up to 30,000 total virtual circuits**
  - **For example, if number of VSIs is limited to 1K and number of peers per VSI is limited to 30, then the full 30,000 peer limit can be supported.**
  - **Up to 900 directed LDP sessions have been tested. It is not recommended that this many be used in a single chassis.**
  - **Up to 4k attachment VCs, which is limited by 4k VLAN on the box [PW treated as AC]**
  - ***Number of mac addresses is also limited (32K), but remember that here VPLS is intended for Server to Server connection, never for PC to Server which should go thru L3-switches first.***

VPLS Convergence



VPLS Split-Horizon

- A packet will never be bridged from a PW to an other PW in the VFI
- Assuming PW full-mesh in a VFI:

  Full reachability

  Core link back-up

  No core L2 loop

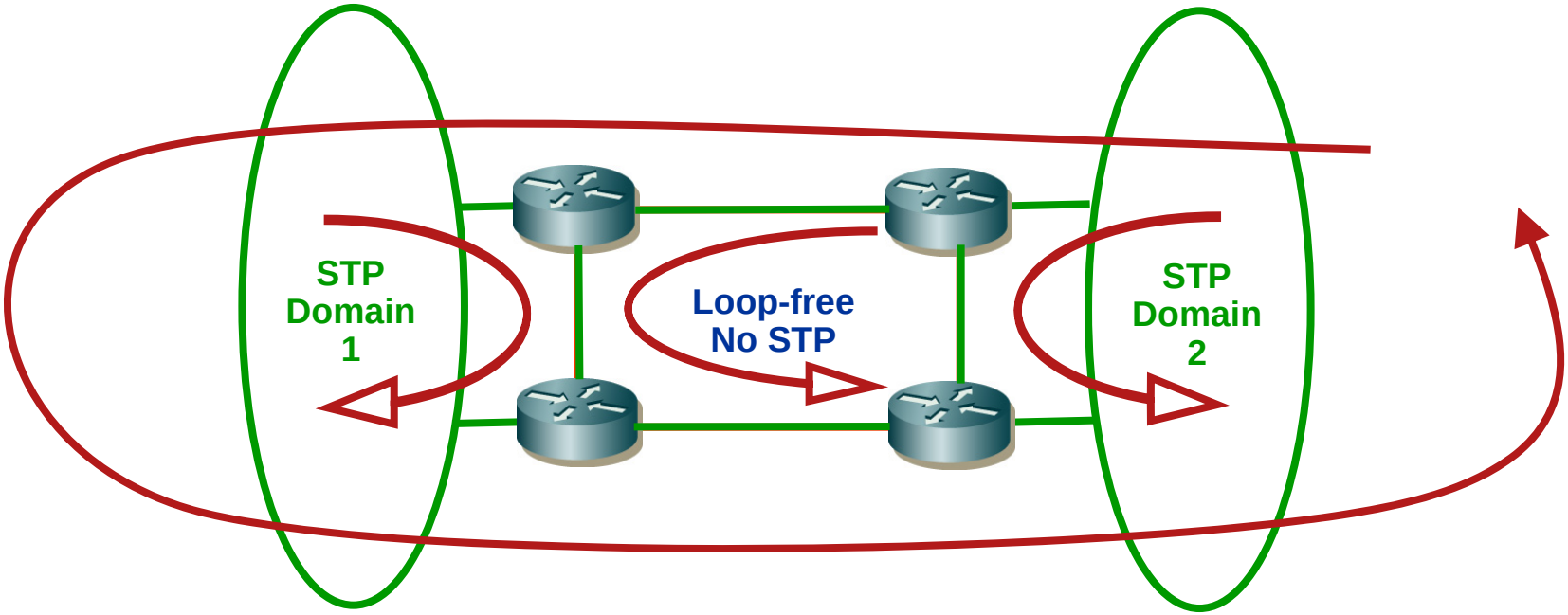
  ➔ No need for a loop prevention core STP

Remark:
Split-Horizon does not protect against loops
on L2 parallel networks built for edge N-PE protection

VPLS implementation versus STP

- VPLS may work in two modes:
  1. STP transparency with extension
     - Core is tunneling BPDU (plain or QinQ)
     - Core is not L2 loop-free
     - End to End STP is preventing loops
  2. STP isolation
     - Core is filtering BPDU
     - Core & DC to DC must be L2 loop-free
     - DC independance / Small STP size
     - This is one important goal for customers
     - Mandatory VPLS-PE ! cannot be the aggregation switch
     - More complex with QinQ

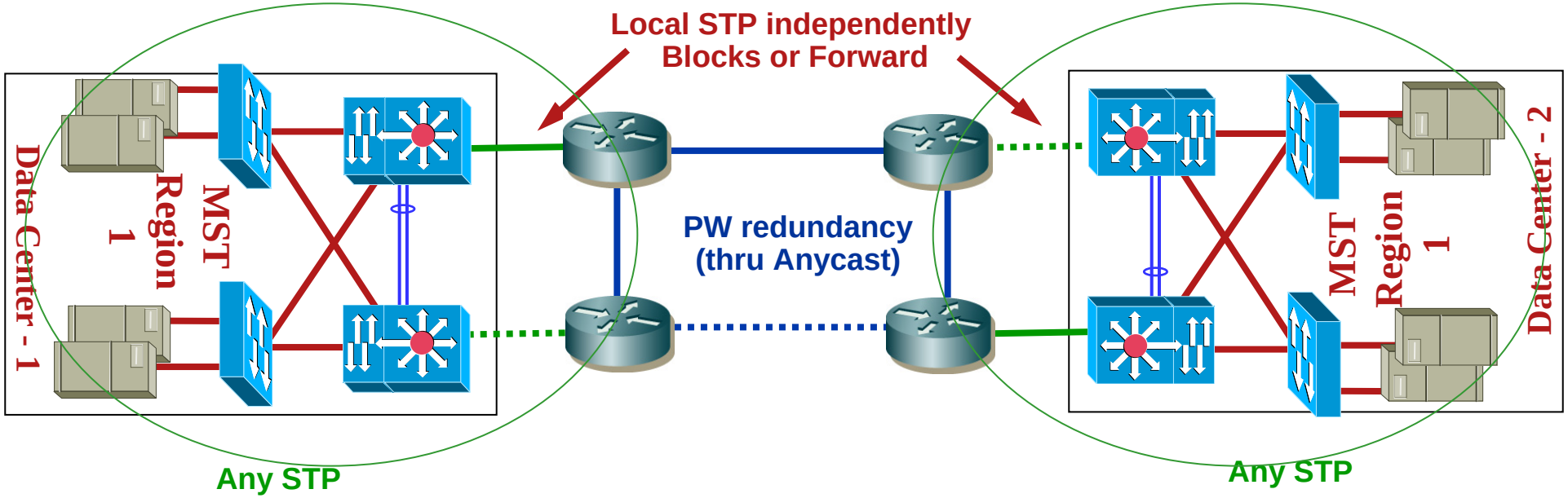
Loop-free interconnection with STP isolation
Problematic



- The role of an end-to-end STP is to suppress any loop
- Without STP loops are created when L2 redundant path.
- Loops means
  - change topology
  - risks of broadcast storms !!

Loop-free interconnection

Anycast PW



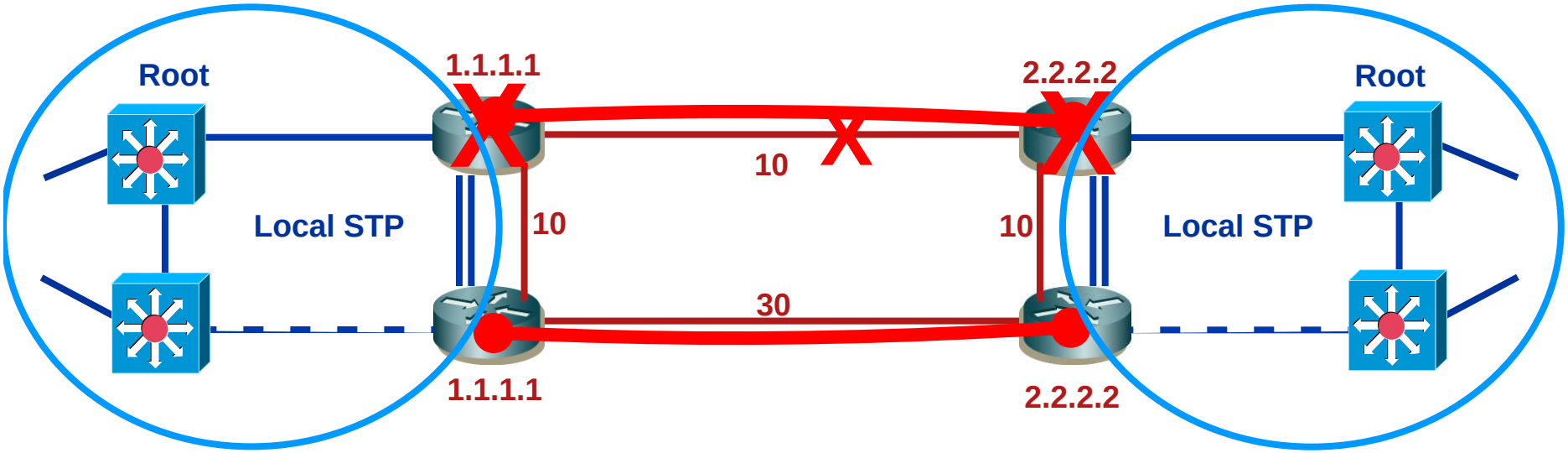
Each site STP is independent from the other

Concept of back-up N-PE and back-up PW

Implementation with no impact on local STP

Independance between DC designs & operations

Anycast PW
Concepts



Anycast concept:

LDP Router-ID is duplicated into back-up N-PE

Notes:

- **PW must be stitched to physical topology**
  - **➔ Traffic-Engineering may adjust this**
- **Link core back-up is directed-LDP detection driven**
  - **➔ TE-FRR may resolve this**
- **Edge links are RSTP protected**

Anycast PW

why do we need to include N-PEs into local STP

Local DC is dual attached to local N-PEs

One only PW is active at a time

➔No loop

But N-PEs must be included into local STP

in order to protect against local links failure!

Anycast PW
Configuration

```
interface Loopback98
 ip address 10.98.65.1 255.255.255.255
 isis circuit-type level-1

interface Vlan5
 ip address 10.10.52.11 255.255.255.0
 xconnect vfi VFI-Anycast

mpls ldp router-id Loopback98 force

mpls ldp neighbor 10.98.65.4 targeted ldp
```

```
interface POS6/0/0
 isis metric 30
```

Same identical loopback on Primary & Backup N-PE on a DC (do not forget to advertise it thru ISIS / LDP)

Xconnect the VLAN to the VFI

Which becomes the LDP router-id

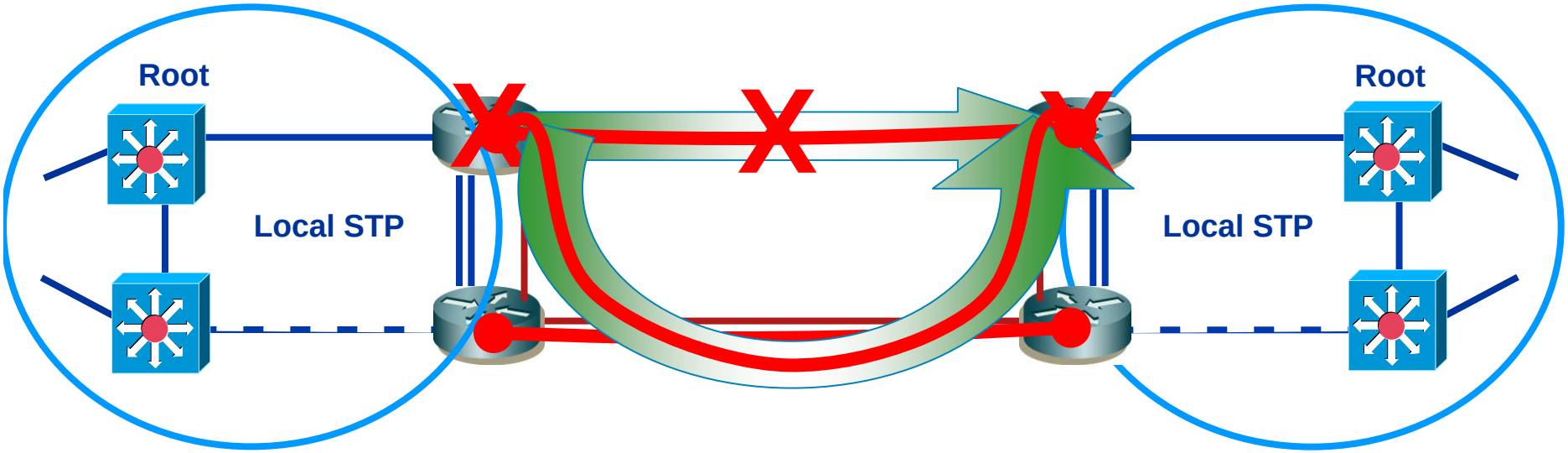
And peers to the other DC

Inter backup nodes link is less prefered

Back-up nodes can peer only between them when primary nodes cannot establish connection

➔ No L2 loop can occur

Anycast PW with Traffic-Engineering



Anycast with TE concepts:

- **LDP Router-ID is duplicated into back-up N-PE**
- **TE is assuring the back-up thru alternate path**

Notes:

- **PW do not need to be stitched to physical topology**
- **Link core back-up is RSVP-TE protected**
- **Edge links are RSTP protected**

Option 5: Anycast PW in square topology
Configuration

```
mpls traffic-eng tunnels
mpls traffic-eng reoptimize events link-up

interface POS6/0/0
 mpls traffic-eng tunnels
 ip rsvp bandwidth 100000

router isis
 metric-style wide
 mpls traffic-eng router-id Loopback99
 mpls traffic-eng level-1

interface Loopback99
 ip address 10.99.65.1 255.255.255.255
 isis circuit-type level-1
```

Enable TE globaly, into ISIS and also on every core links

Set-up only into the primary router an other loopback than the LDP'one (do not forget to advertise it thru ISIS / LDP)

First establish a standard MPLS TE core

Load Repartition

1. ECMP load balance L2 labelized packet base on the two outer labels of the stack:
   - Aka Destination N-PE / VFI-ID
   - May be not the best way to equaly balance traffic
2. Build a symetric primary VFI on back-up N-PE
   - Use PVST+ to have VLAN select which N-PE is root
   - Xconnect-VFI these VLAN
   - Be careful to never xconnect the same VLAN on both N-PE
3. Use TE to balance VLAN
   - Create a TE tunnel on back-up path
   - Have the VFI prefering the alternate TE tunnel
   - Aka load balance selectively VLANs on pathes

Use TE to balance VLAN

```
interface Tunnel1
 ip unnumbered Loopback99
 mpls ip
 tunnel destination 10.99.65.4
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng forwarding-adjacency
 isis metric 2 level-1
 tunnel mpls traff path-option 1 explicit name LB
!
ip explicit-path name LB enable
 exclude-address 10.169.14.4

pseudowire-class VPLS-Tunnel-1
 encapsulation mpls
 preferred-path interface Tunnel1
!
l2 vfi VFI-Alternate manual
 vpn id 98
 neighbor 10.98.65.4 pw-class VPLS-Tunnel-1
```

Create a TE tunnel on back-up path

Have the VFI using the alternate TE tunnel

Option 5: Anycast PW in square topology
Configuration

```
interface Tunnel14
 ip unnumbered Loopback99
 mpls ip
 tunnel destination 10.99.65.4
 tunnel mode mpls traffic-eng
 tunnel mpls traff path-option 1 dynamic

 tunnel mpls traff forwarding-adjacency
 isis metric 2 level-1
```

As TE tunnel are unidirectional, do not forget to create one TE into both Primary N-PE

TE tunnel is considered a a link between Pes, this overcome the limitation to have physical topology equal to L2-PW

Then create TE tunnel between Primary N-PE

Is TE mandatory?
Or can we get rid of it?

- PW Anycast requires PEs to be full-meshed
- Link failure without PW failure requires PEs to be full-meshed

- In many cases, this is no physically possible
  - ➔TE solves this
  - ➔ but is not mandatory

- Load balancing is well managed per TE
- FRR (sub-50ms) is TE

Protection against potential loop

- With L2 network, potential loop may occur on configuration error (human mistake, hardware issue..).
- The STP loop should affect only one building block of a single DC design
- PW and Remote N-PE (therefore remote DC) must be protected against broadcast storm
- ISIS/LDP/RSVP control packet would rather be protected against broadcast storm
- QoS Policer may be used to protect the L2-Core and therefore the remote DC from braodcast storms
  - Protocol Independent MAC ACL

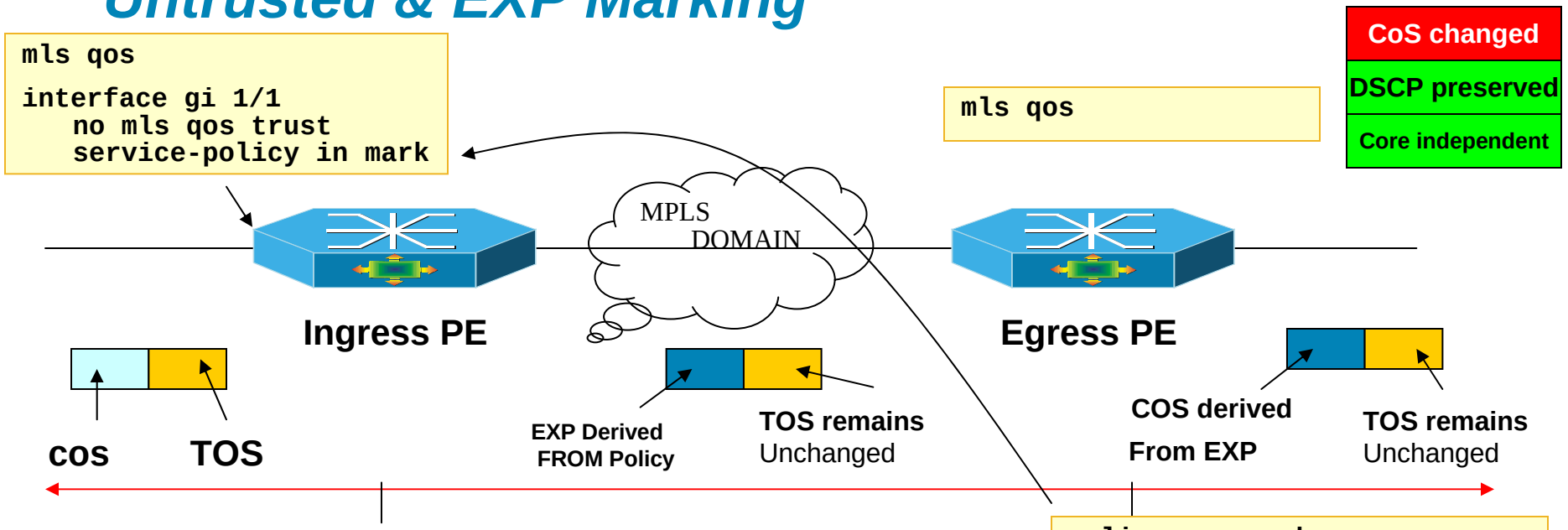
    inbound rate limiter applied to the xconnect port
  - VLAN-based rate limiter

QoS

- QoS will be used in DiffServ mode into Core
  - To protect signaling (ISIS/LDP/RSVP control packet)
- To differentiate L2 traffic from L3
  - Dedicated EXP for L2 flows
- To differentiate some L2 traffic from other L2
  - Protected some important VLANs
- Per traffic-class L2 differentiation is also possible
  - In one PW, have a concept of Gold/Silver/Bronze
  - Marking being done at ingress
- Beware of a max of 8 CoS

Cisco 7600 PFC3 based QoS

Untrusted & EXP Marking



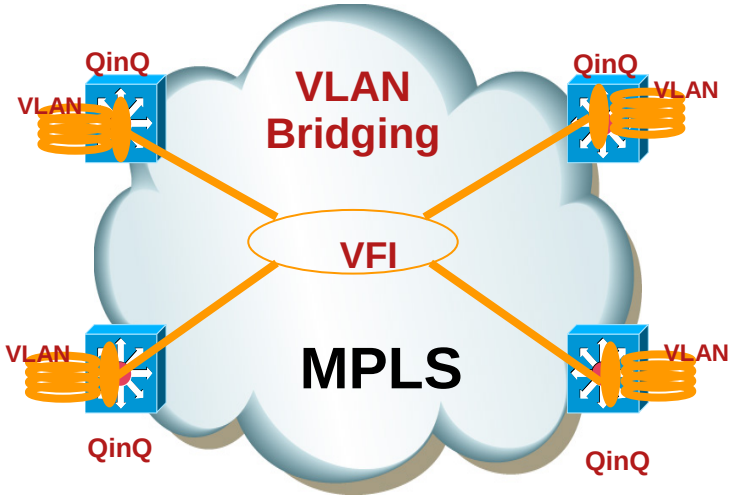
- Ingress PE
  - Internal DSCP is determined by Input Marking
  - Internal DSCP is copied to EXP during imposition
- Egress PE
  - EXP bits are copied to Internal DSCP
  - Internal DSCP is used to rewrite CoS

Conclusion on VPLS

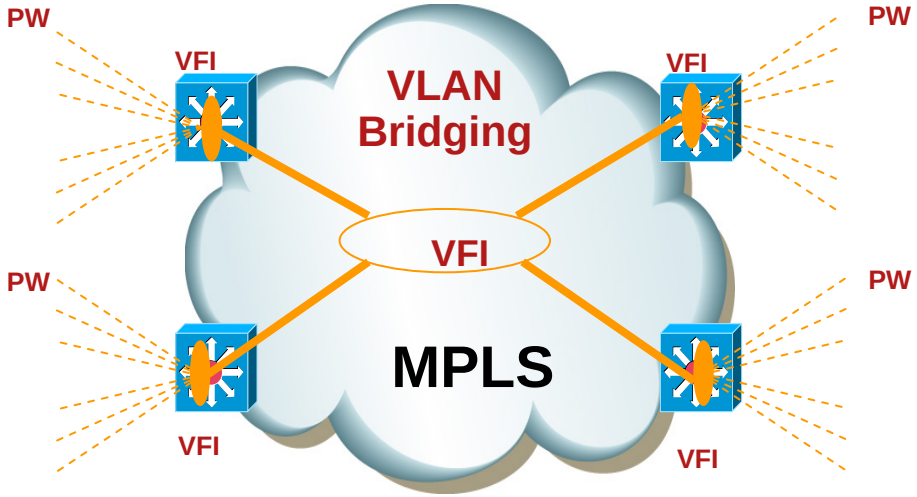
- VPLS is bringing a brand new paradigm in the inter-DC VLAN interconnection
  - Link/Node failure L3 protected
  - Spanning-tree isolation
- A more complex core allows a simple & scalable DC interconnect
- VPLS with Anycast-PW is a good solution for a small amount of VLAN
- This solution is new, but already deployed with easy success
- Technology evolution will add easiness
- H-VPLS is bringing scalability

L2 Extension
Scaling with HVPLS

Hierarchical-VPLS (H-VPLS)
2 standard approaches



H-VPLS with bridge-group domain at access



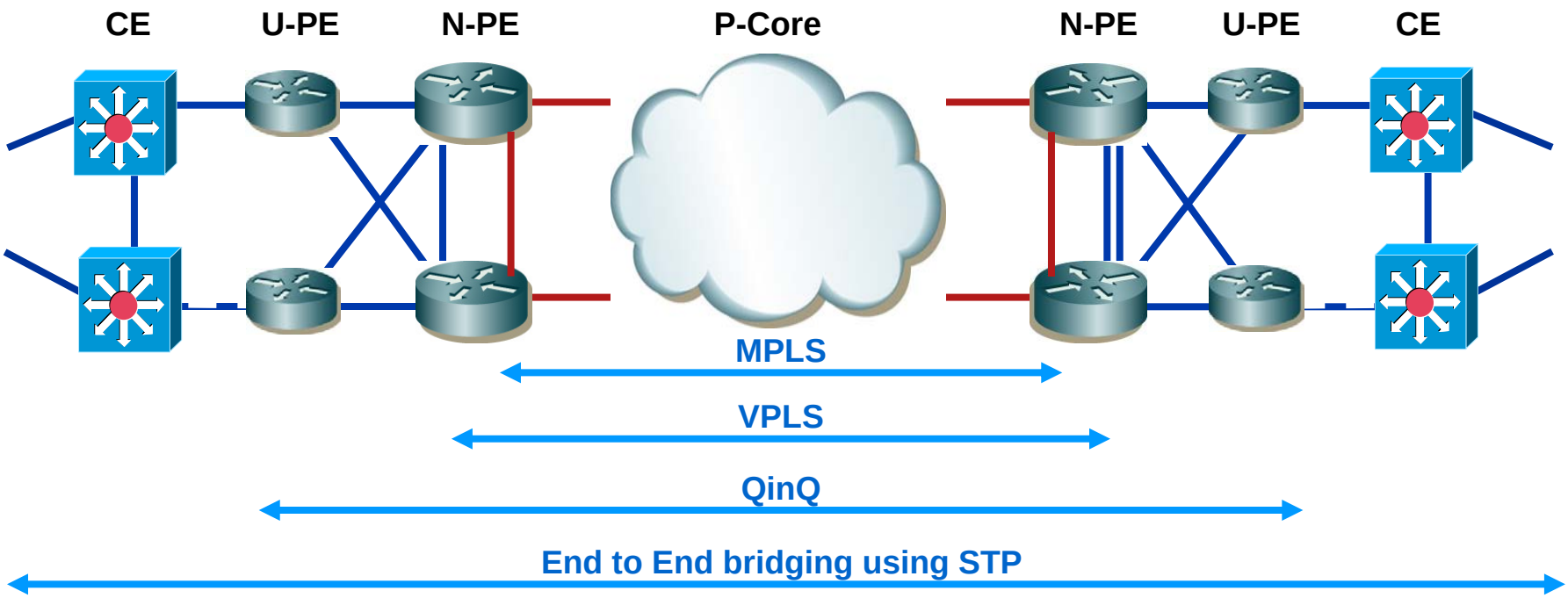
H-VPLS using PW EoMPLS at access

H-VPLS devices role

- **U-PE: User facing PE**
  - QinQ encapsulation
    - usually BPDU tunneling (L2PT)
  - EoMPLS point to point encapsulation
    - Per port or per VLAN
- **N-PE: Network facing PE**
  - VFI hosting
  - PW full-mesh with split-horizon toward all other N-PE with same VFI
  - Xconnect Core-VLAN to VFI
  - Xconnect EoMPLS edge to VFI without split-horizon

H-VPLS using QinQ standard layers and devices

Technology overview

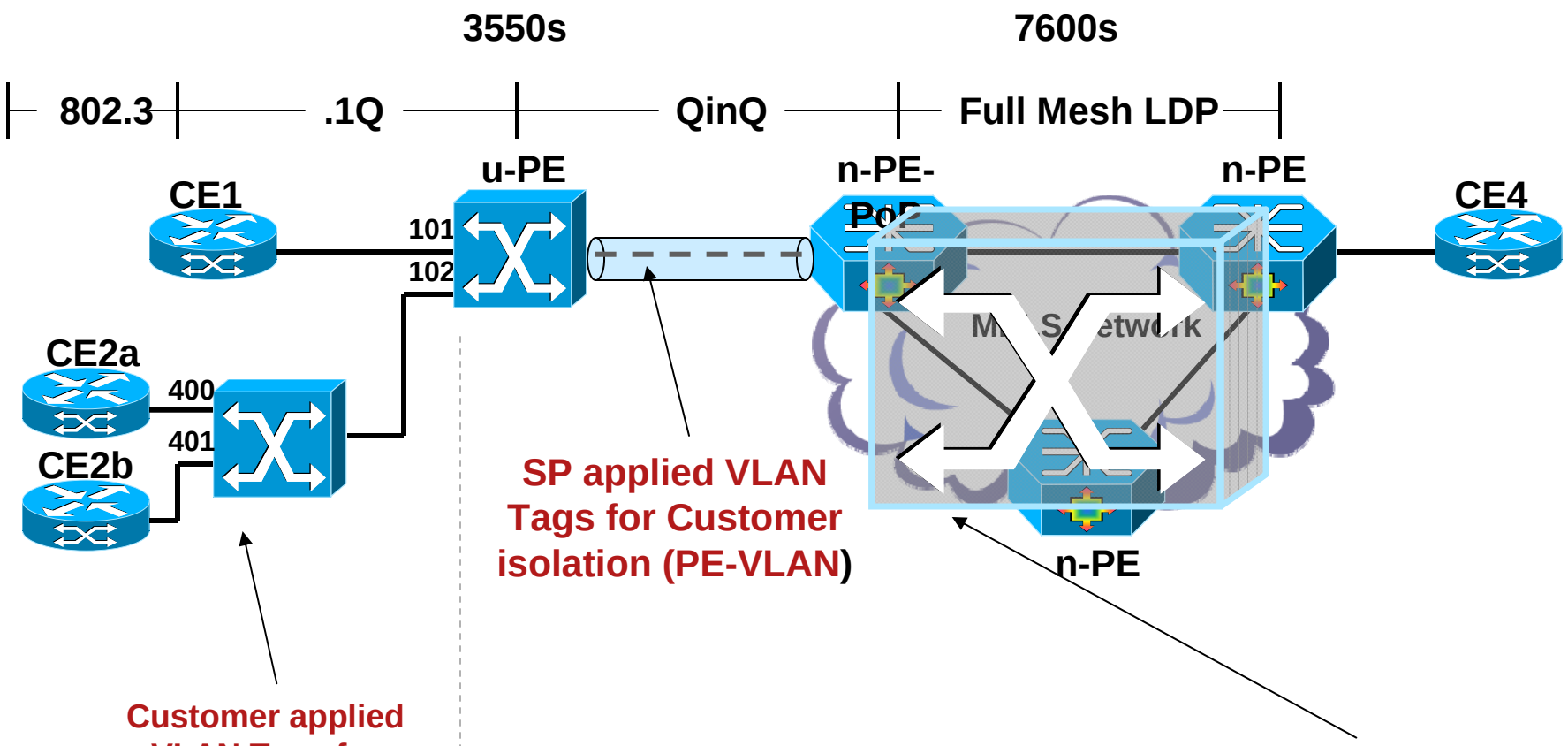


In standard H-VPLS:

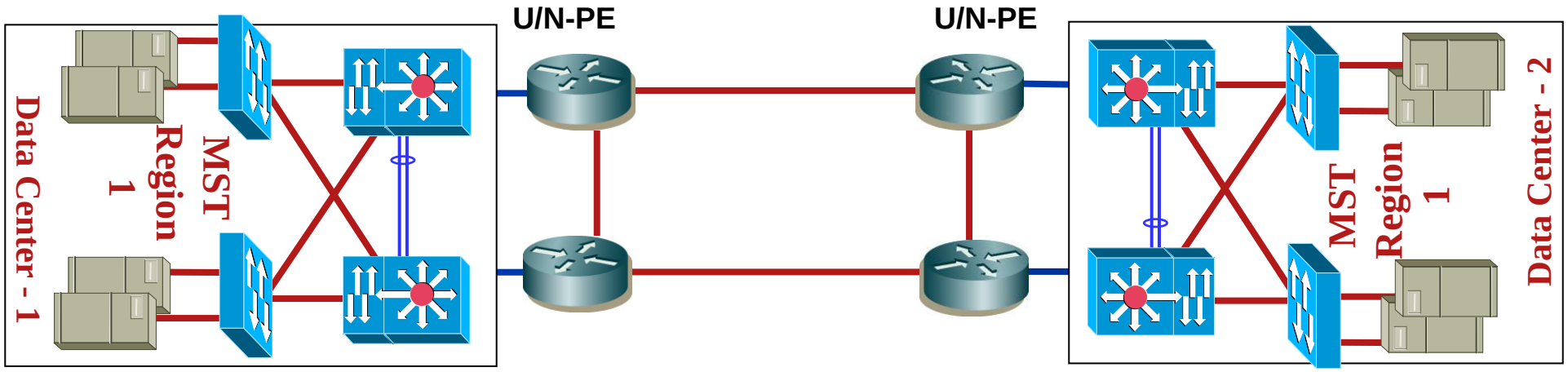
- **Edge STP is tunneled from end to end**
- **U-PE is isolated from N-PE**
- **Integrated U-PE into N-PE is supported**

H-VPLS using QinQ
Data encapsulation

Technology overview



H-VPLS using QinQ applicability to DC interconnect

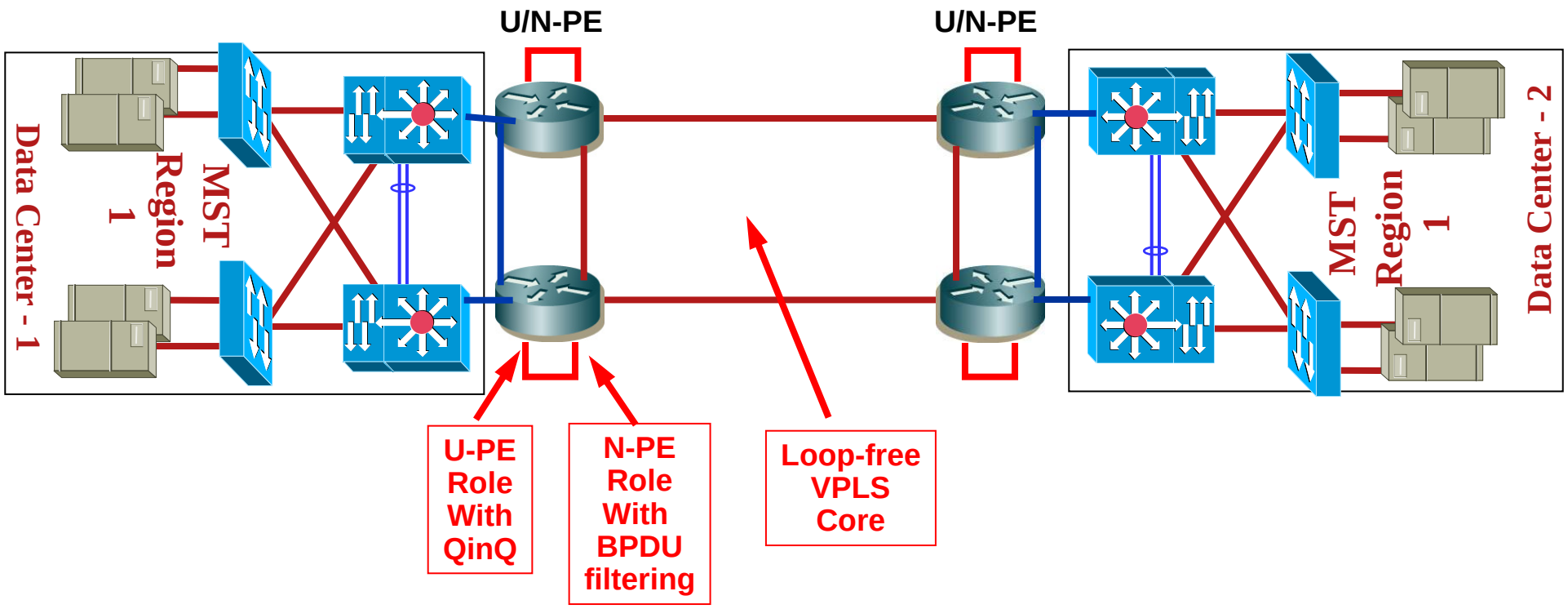


PE may play at the same time U-PE & N-PE role

But, this will imply End-to-End STP tunneling !

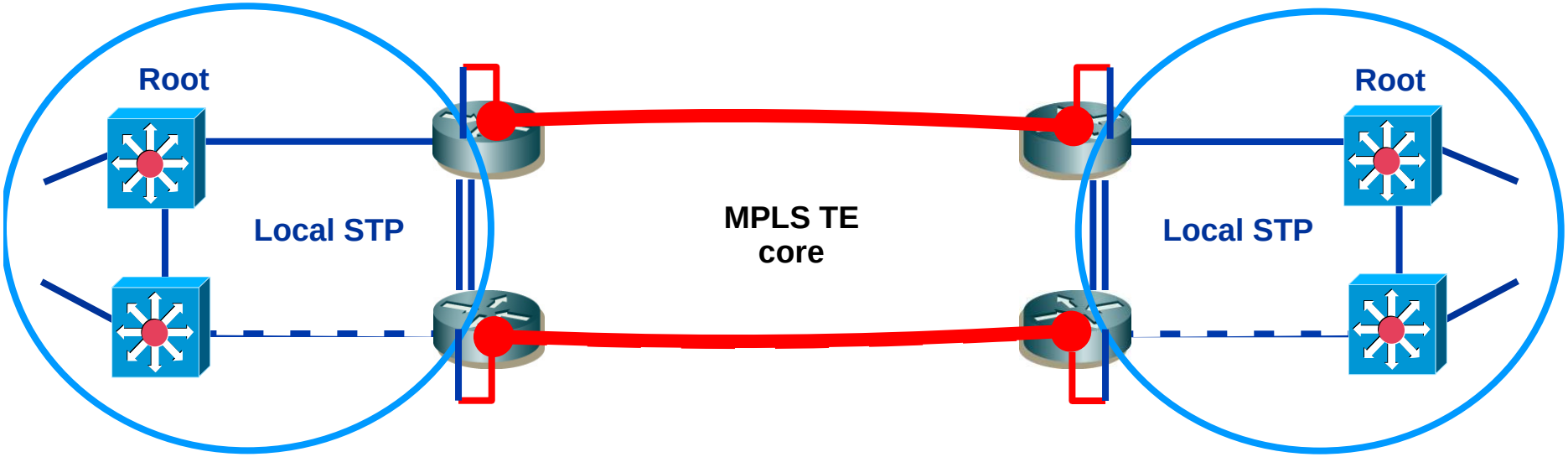
Else, …

H-VPLS using QinQ for DC interconnect



In order to decouple U-PE & N-PE roles into the same box
An external loopback cable may be used
QinQ over VPLS then supports STP isolation !
Requires dedicated VLAN numbers for QinQ core labels

H-VPLS with QinQ over PW Anycast STP isolation using Loopback cable



1. Every local links including inter N-PE are loop-protected using local STP
2. With PW Anycast, inter-PE link has no critical role, just backup path
3. Local STP is NOT tunnelled toward other sites
4. Every local VLAN is tunnelled into Core-VLAN
5. Core-VLAN is xconnected to VFI
6. Core VPLS is loop-free using PW-Anycast approach over TE

H-VPLS over PW Anycast
Configuration

```
interface Vlan601


interface Vlan98
 xconnect vfi VFI-Anycast
```

VLANs to be transported do not need anymore to be xconnected

```
interface GigabitEthernet3/36
 switchport
 switchport access vlan 98
 switchport mode dot1q-tunnel
 no cdp enable
 spanning-tree bpdufilter enable
 spanning-tree cost 1000
 lan-name Loopback
```

Core-VLAN is the only one to be xconnected
With U-PE & N-PE fusion, this VLAN must have a dedicated number

QinQ on one side of the loopback cable

STP is filtered only before beeing sent to VFI

```
interface GigabitEthernet3/35
 switchport trunk allowed vlan 1,601,602
 ...
```

VLAN to be transported
Core VLANs numbers must not be set at edge

Conclusion
for L2 Data-Center interconnection

- Segregate the different Applications using Layer 3
- Avoid Extending L2 VLAN if it's not required by the Application
- If Extended L2 VLAN is required:
  - dedicate the L2 for the specific Application and keep it isolated from other Application via L3 Network
  - Do not propagate the same STP outside your local DC
  - Police & Rate limit the traffic per VLAN ➔ prevent broadcast storm
- Prefer L3 Fast-Convergence and MPLS FRR with TE to make a single L2-VPN Pseudowire
  - The Physical layer becomes logically fully resilient
  - Physical link failure becomes transparent for L2
- VPLS and Split-Horizon assure a fully resilient Loop-Free Network without the need to deploy STP.

Meet the Experts
IP and MPLS Infrastructure Evolution

- Andy Kessler
  Technical Leader
- Beau Williamson
  Consulting Engineer
- Benoit Lourdelet
  IP services Product manager
- Bertrand Duvivier
  Consulting Systems Engineer
- Bruce Davie
  Cisco Fellow
- Bruce Pinsky
  Distinguished Support Engineer






Meet the Experts
IP and MPLS Infrastructure Evolution

- Gunter Van de Velde
  Technical Leader
- John Evans
  Distinguished Systems Engineer
- Oliver Boehmer
  Network Consulting Engineer
- Patrice Bellagamba
  Consulting Engineer
- Shannon McFarland
  Technical Leader





Meet the Experts
IP and MPLS Infrastructure Evolution

- Andres Gasson
  Consulting Systems Engineer

- Steve Simlo
  Consulting Engineer

- Toerless Eckert
  Technical Leader

- Dino Farinacci
  Cisco Fellow & Senior Software Engineer

Recommended Reading

BRKIPM - 3014

- Continue your Networkers learning experience with further reading from Cisco Press.
- Visit the on-site Cisco company store, where the full range of Cisco Press books is available for you to browse.