Data Center Interconnect MPLS L2VPN Solutions

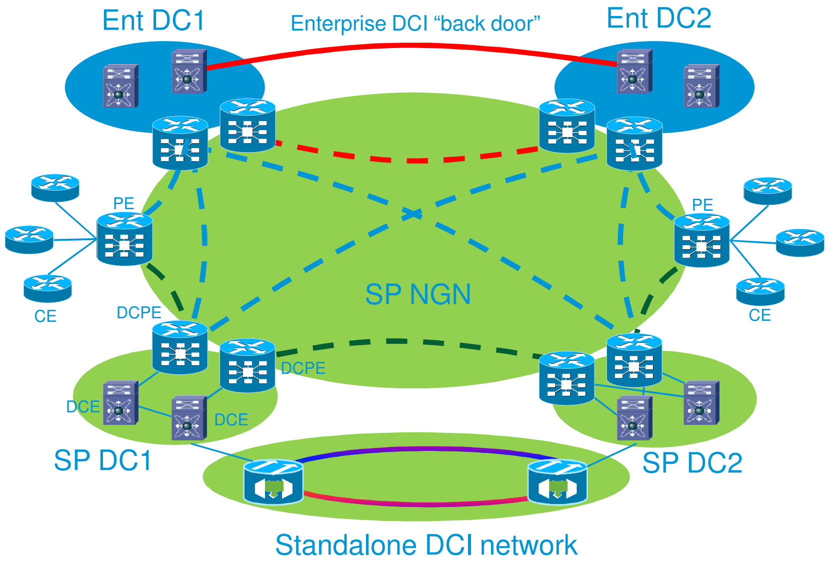
Yves Hertoghs, yves@cisco.com

Distinguished Consulting Engineer

SP Multitenant DCI: Baseline Use Cases and Requirements

- Use Cases:
  - Virtual Machine Mobility at L2 and/or L3
  - Server Clustering at L2 and/or L3
- Scales to the level required for SP virtual private cloud
  - 100s of thousands of MAC addresses per data centre
  - Thousands of tenants ; potentially more than 4K service instances
  - 10s of data centres
- Optimally forward unicast and multicast
  - Shortest path
  - Loop free
  - Avoiding duplicates
- Is resilient to all single element failures, i.e. in both NGN and DC
- Provides control plane isolation between DCs
- Fast to converge
- Uses network resources efficiently
  - All connections active with load balancing
  - Flood minimisation
- Easy to manage and operate
- Open standards based or clear track to standardisation
- Integrates with SP NGN, whilst honouring any administrative boundaries between DC and NGN, including DC connectivity across multiple AS'es
- Supports geo-redundant PEs, i.e Enterprise DCI "back door"
- Is DC transparent
  - works for plain old spanning tree 802,1Q environment (Normalized DCI Handoff)
  - interworks with other DC technologies (Seamless DCI Handoff)

SP Managed Data Center Interconnect Solutions



- NGN Based DCI Interconnection models:
  - Enterprise to Enterprise (E2E)
  - Enterprise to Service Provider (E2SP)
  - Service Provider to Service Provider (SP2SP)

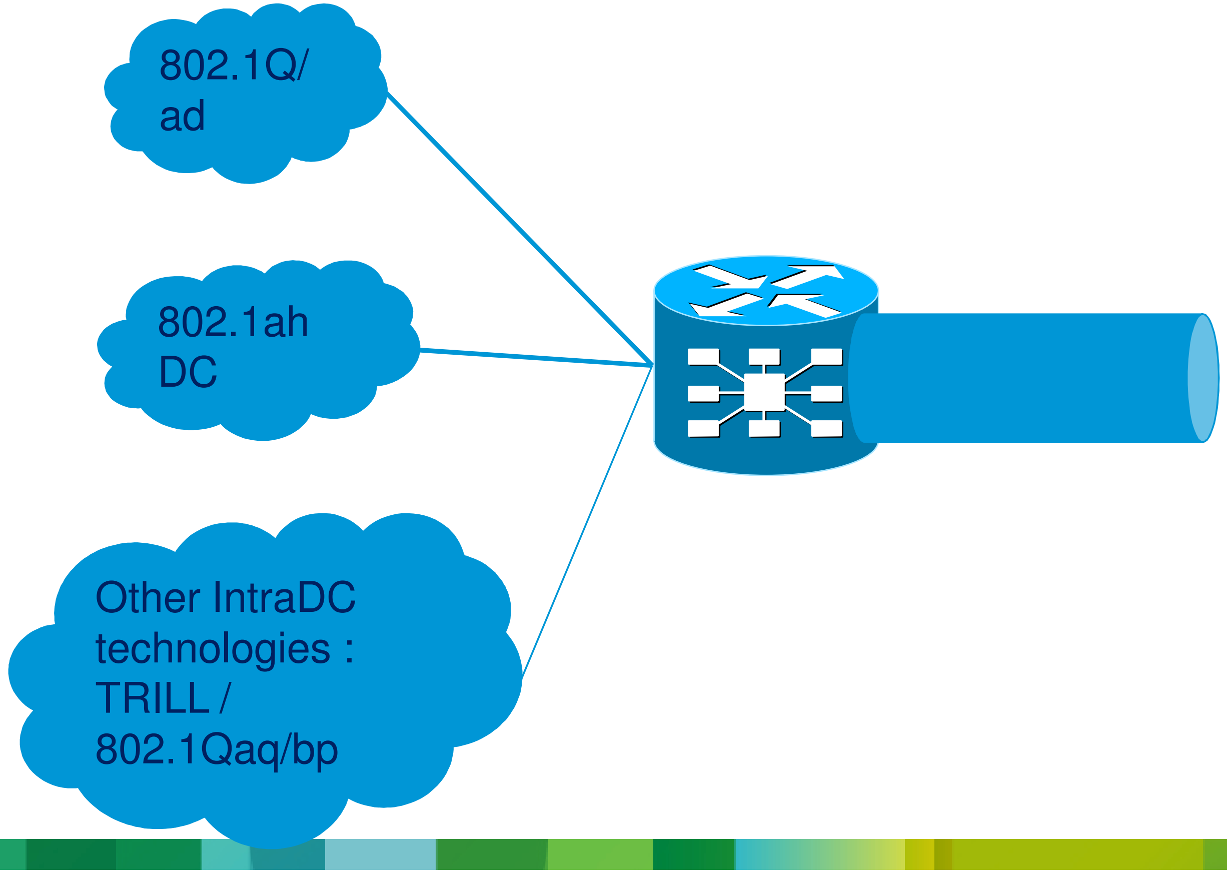
- Standalone DCI network provides interconnection between main SP DCs
  - Owned by SP DC team
  - Addresses SP2SP only
  - Very high bandwidth – packet / optical solution likely the most cost effective

- DCI Requires Technology Evolution in Data Center and SP NGN for:
  - Multihoming
  - Scale (MAC-addresses, Number of Service Instances
  - Loadbalancing
  - Optimal Forwarding
  - Multicast optimization
  - Multitenancy

Data Center Interconnect: Layer 2 Extension Technology

- The SP managed Data Center Interconnect solution will simultaneously cater for :
  - L3 adjacencies: technologies such as MPLS-VPNs will be used
  - L2 adjacencies: L2VPN technologies such as:
    - Virtual Private LAN Service (VPLS)
      - The best available option in shipping code
      - Does not meet some of the data center interconnect requirements for large SP Multitenant Deployment options
    - Ethernet-VPN (E-VPN) / Provider Backbone Bridging Ethernet VPN (PBB-EVPN)
      - New technologies to meet all of the large SP multitenant data center interconnect requirements
        - http://tools.ietf.org/html/draft-ietf-l2vpn-evpn
        - http://tools.ietf.org/html/draft-ietf-l2vpn-pbb-evpn

Towards a common DCI Handoff ?



- Is DCI a UNI or NNI ?

  All Service Instances remapped to 802.1q VLANs

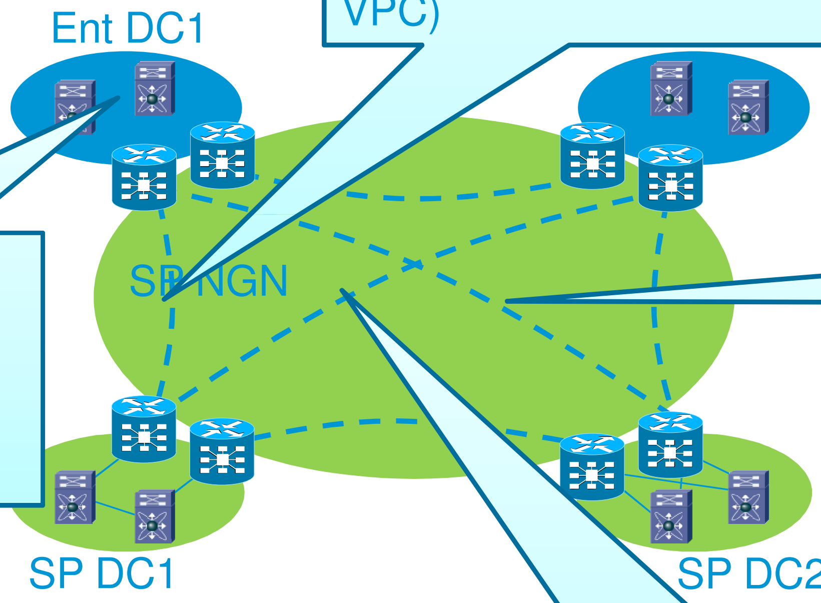
  or end to end (assumes other encapsulation inside DC)

- Is there a Control Plane inside the Data Center?

  Control Plane interworking considerations

  IGP in DC ; BGP across DCs ?

Multi-Homing DC's and Loadbalancing/Resilience across VPLS



Signalling Active PE : Inter Chassis Control Protocol (ICCP,) or BGP
Or PE Clustering (VSS, nV, VPC)

Multi Chassis Link Aggregation (MC-LAG)
Multi Chassis Automatic Protection Switching (MC-APS)

Flow Aware Transport: FAT-PW

Selecting correct Pseudo-Wire: Pseudowire Redundancy
Or BGP

VPLS constraints

- Not optimal with multicast

  Enhancements are maturing (using Label Switched Multicast with VPLS instead of ingress resplication)

- No active/active dual-homing per flow

  Per VLAN is possible

- Does not hide customer mac-addresses
- PW scaling
- Handoff scaling and Service Instance Scaling

  4k services per physical interface

  000's of VSI's (hardware limitations)

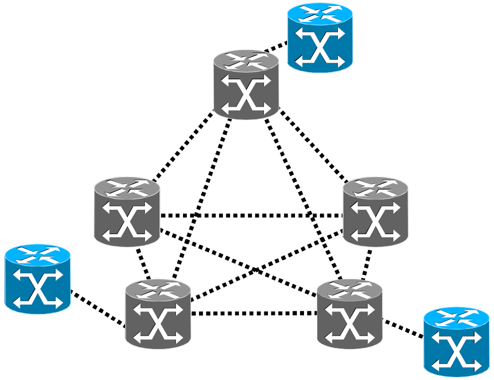
Scaling VPLS : PBB-VPLS

- VPLS current challenges

  MAC-Address Scalability at the PE

  Service Instance Scaling

  Limits DCI handoff to 4K services per interface

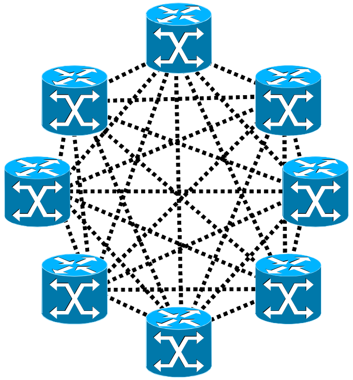


- Approach:

  Use Provider Backbone Bridging (PBB)/802.1ah with VPLS

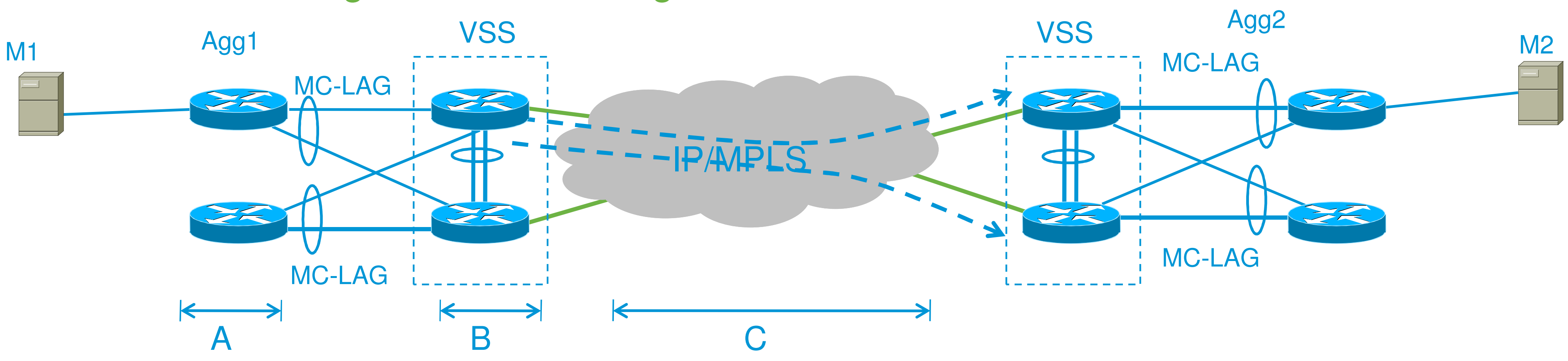
  Hides Customer MAC-Addresses

  Described in http://tools.ietf.org/html/draft-ietf-l2vpn-pbb-vpls-pe-model and http://tools.ietf.org/html/draft-ietf-l2vpn-pbb-vpls-interop



Cisco A-VPLS : VPLS w/ MC-LAG & Fat-PWs

Advanced 3-stage Load-Balancing



- Flow Aware Transport (FAT) Pseudo-wires as in RFC6391
- A: Aggregation switch performs EtherChannel flow-based hashing (on L2/L3/L4) & elects a link towards VSS switch (e.g. Cat6000).
- B: VSS performs flow-based hashing (L2/L3/L4) to select outbound ECMP link. Optionally inserts FAT-PW Flow Label (to be used in C).
- C: P nodes in MPLS core perform Loadbalaning over ECMP using Flow Label.

Note: Load-balancing decisions in A, B & C are independent.

Evolving Requirements for L2VPN

1. All-active Redundancy
   - Flow Based Load Balancing
   - Flow Based Multi-pathing
   - Geo-redundancy and Flexible Redundancy Grouping
2. Simplified Provisioning and Operation
   - Core Auto-Discovery
   - Access Multi-homing Auto-Discovery
   - New Service Interfaces
3. Optimal Multicast with LSM
   - P2MP Trees
   - MP2MP Trees
4. Fast Convergence
   - Link/Port/Node Failure
   - MAC Mobility

5. Scalable for SP virtual private cloud service:
   - Support O(10 Million) MAC Addresses per DC
   - Confinement of C-MAC Learning
6. Seamless interworking between TRILL / 802.1aq / 802.1Qbp and MST / RSTP
   - Guarantee C-MAC Transparency on PE
7. Fast Convergence
   - Avoiding C-MAC Flushing

Underline: Addressed by VPLS

🟨 Addressed by E-VPN & PBB-EVPN

🟩 Addressed by PBB-EVPN

What is Ethernet-VPN (E-VPN)

At a glance

- Treat MAC addresses as routable addresses and distribute them in BGP
- When multiple PE nodes advertise same MAC, create multiple adjacencies in forwarding table
- When forwarding traffic for a given unicast MAC DA, use hashing (L2/L3/L4) to pick one of the adjacencies
- MP2MP or P2MP LSPs for Multicast Traffic Distribution
- MP2P (like L3VPN) LSPs for Unicast Distribution
- NO FULL MESH of PW's !!!

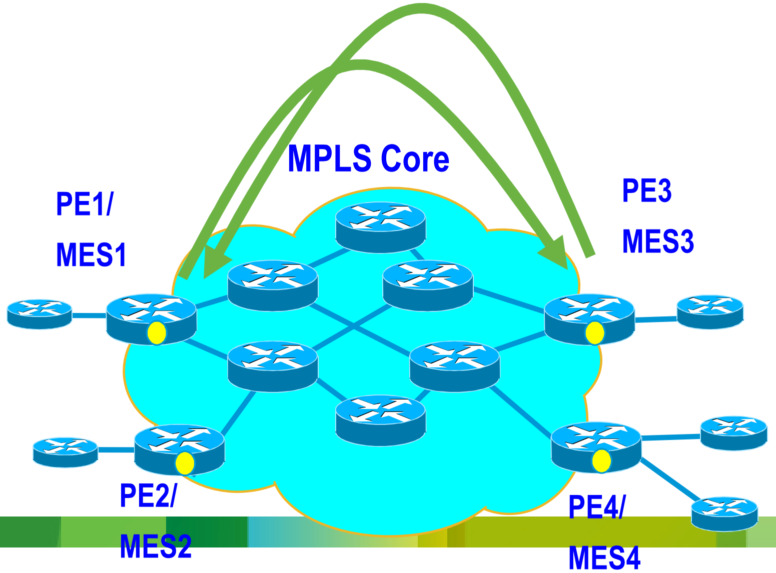
From PE1

iBGP L3-NLRI:

- next-hop: n-PE1
- <C-IP1, L1>

iBGP L2-NLRI

- next-hop: n-PE1
- <C-MAC1, L2>



From PE3

iBGP L3-NLRI:

- next-hop: n-PE3
- <C-IP5, L1>

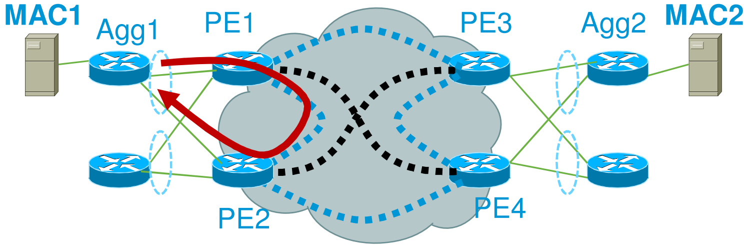
iBGP L2-NLRI

- next-hop: n-PE3
- <C-MAC3, L2>

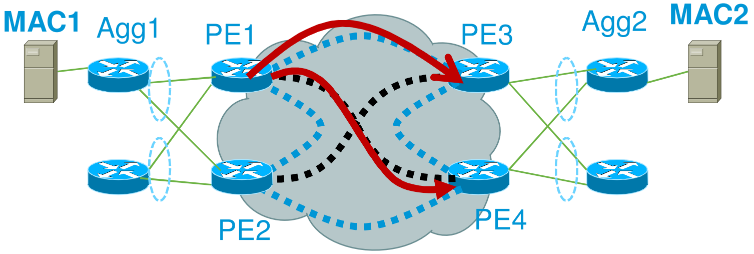
It looks easy but not so fast !

- In the shown example, how do we ensure that

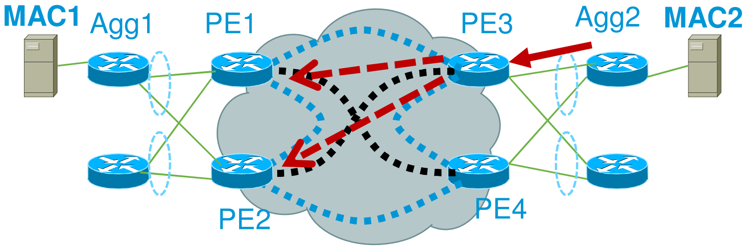
ARP broadcast packet doesn’t get loopback to the originating Agg device (Agg-1) :
Split Horizon for ESI



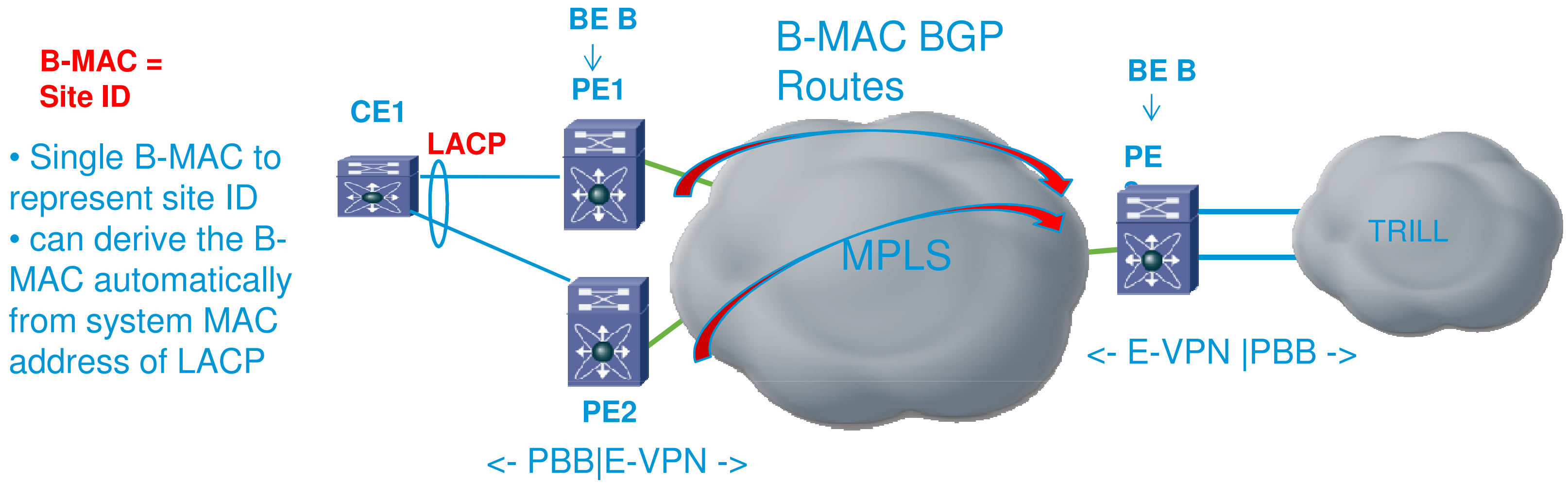
Either PE3 or PE4 forward the broadcast frame to the far-end dual-homed device (Agg-2)
Designated Forwarder Selection



When PE3 wants to forward a packet with destination address MAC1, it needs to send it to both PE1 and PE2 even though it only learned MAC1 from PE1
Aliasing



Provider Backbone Bridging E-VPN (PBB-EVPN)


B-MAC =
Site ID
• Single B-MAC to represent site ID
• can derive the B-MAC automatically from system MAC address of LACP
BE B
PE1
CE1
LACP
PE2
<- PBB|E-VPN ->
B-MAC BGP
Routes
MPLS
BE B
PE
TRILL
<- E-VPN |PBB ->

- Advertise local B-MAC addresses in BGP to all other PEs that have at least one VPN in common just like E-VPN
- Build a forwarding table from remote BGP advertisements just like E-VPN (e.g., association of B-MAC to MPLS labels)
- PEs perform PBB functionality just like PBB-VPLS

  C-MAC learning for traffic received from ACs and C-MAC/B-MAC association for traffic received from core

PBB-EVPN Main Principles

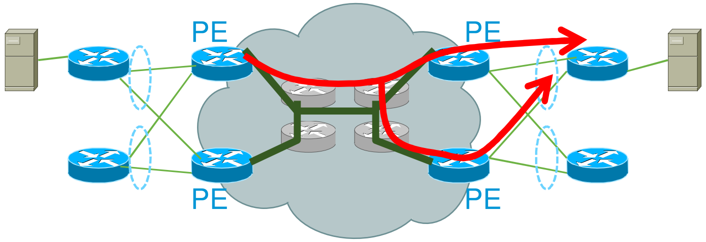
- DF Election with VLAN Carving

  Prevent duplicate delivery of flooded frames.

  Uses BGP Ethernet Segment Route.

  Performed per Segment rather than per (VLAN, Segment).

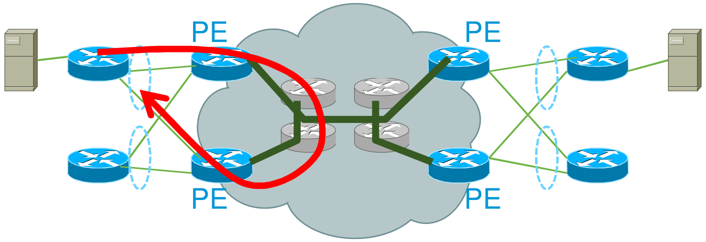
  Non-DF ports are blocked for flooded traffic (multicast, broadcast, unknown unicast).



- Split Horizon for Ethernet Segment

  Prevent looping of traffic originated from a multi-homed segment.

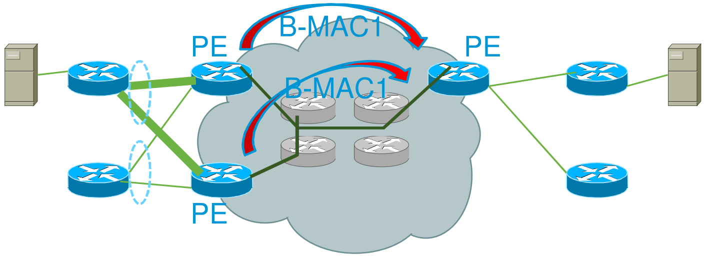
  Performed based on B-MAC source address rather than ESI MPLS Label.



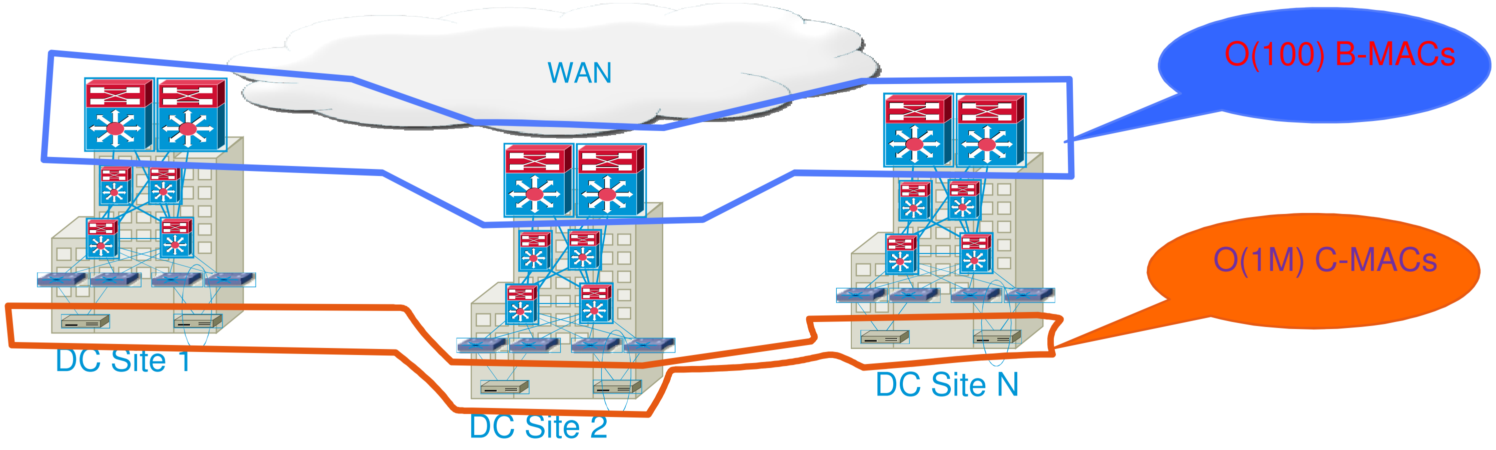
- Aliasing

  PEs connected to the same multi-homed Ethernet Segment advertise the same B-MAC address.

  Remote PEs use these MAC Route advertisements for aliasing load-balancing traffic destined to C-MACs reachable via a given B-MAC.



Advantages of PBB-EVPN : MAC Address Scalability



1. BGP MAC Advertisement Route Scalability

   Multiple orders of magnitude difference between C-MAC & B-MAC addresses

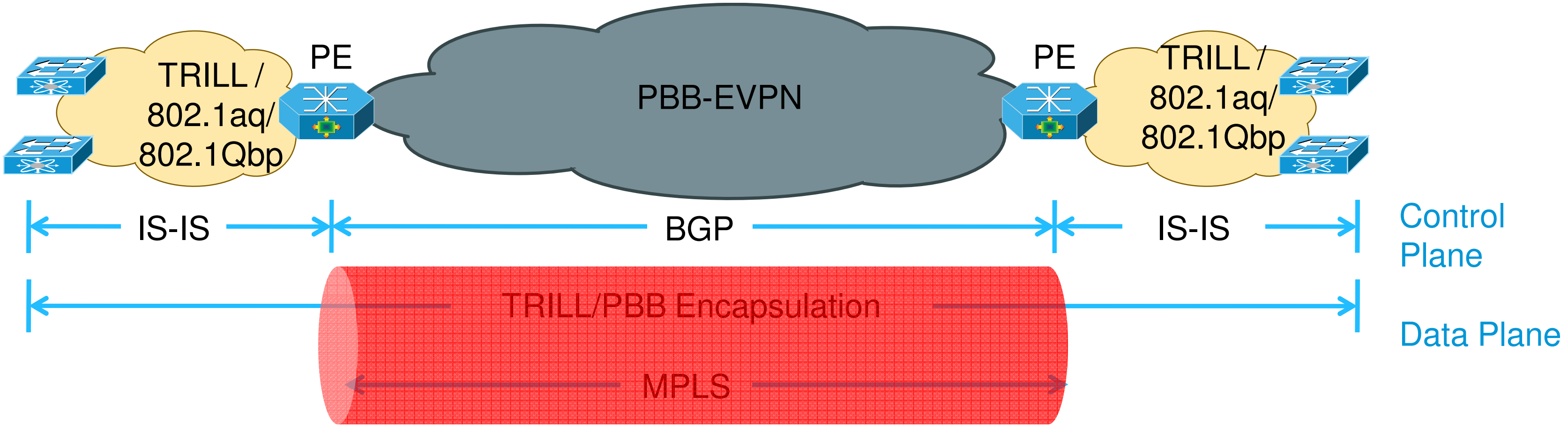
2. C-MAC Address Confinement

   With data plane C-MAC learning, C-MACs are never in RIB and are only present in FIB for active flows

   Whereas, with control plane C-MAC learning, C-MACs are always in RIB and maybe also in FIB

Advantage: IntraDC Interworking

TRILL / IEEE 802.1aq / 802.1Qbp



- End-to-end tunneling of C-MAC addresses thus avoiding data-plane termination and C-MAC learning by PE.
- Control plane isolation between different TRILL / IEEE 802.1aq/ 802.1Qbp islands.

Comparison DCI MPLS solutions

Characteristics	Legacy VPLS	Cisco's A-VPLS	E-VPN	PBB-EVPN
Flow-based Load Balancing	No	Yes	Yes	Yes
Flow-based multi-pathing	No	Yes	Yes	Yes
Geo redundant group & opt. unicast	No	No	Yes	Yes
Flexible redundancy grouping	No	No	Yes	Yes
MAC Scaling	No	No	No	Yes
MP2MP MDT support	No	No	Yes	Yes
P2MP MDT support	No	No	Yes	Yes
Fast convergence upon AC failure	No	Yes	Yes	Yes
Flow-based or VLAN-based LB for MHN	No	Yes	Yes	Yes
Minimal configuration	No	Yes	Yes	Yes
Auto detect of MHN/MHD for flow-based LB	No	No	Yes	Yes
Scaling MPLS Core – full-mesh	No	No	Yes	Yes

Thank You!

CISCO