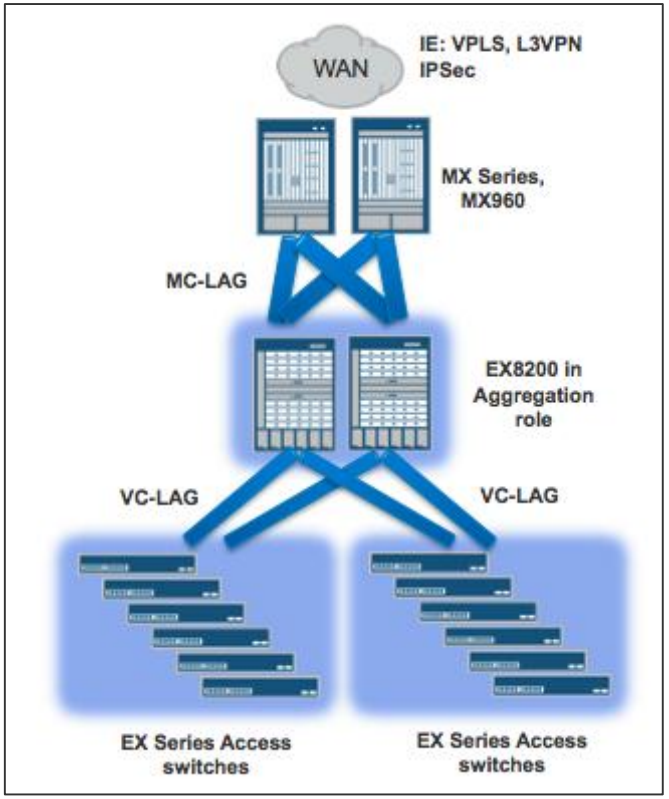
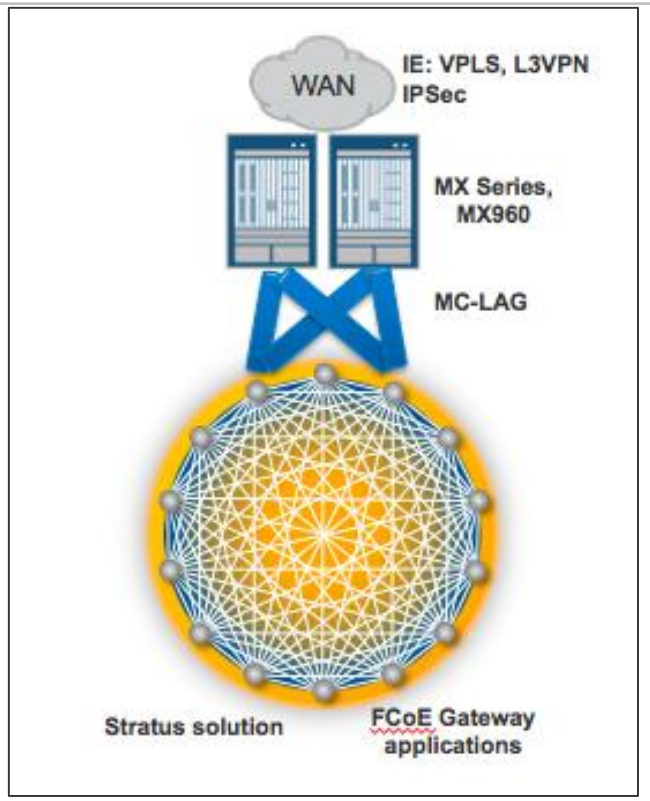
DATA CENTER INTERCONNECT

Prado Yang
System Engineer
2013/08/27

DATA CENTER EDGE FUNCTION





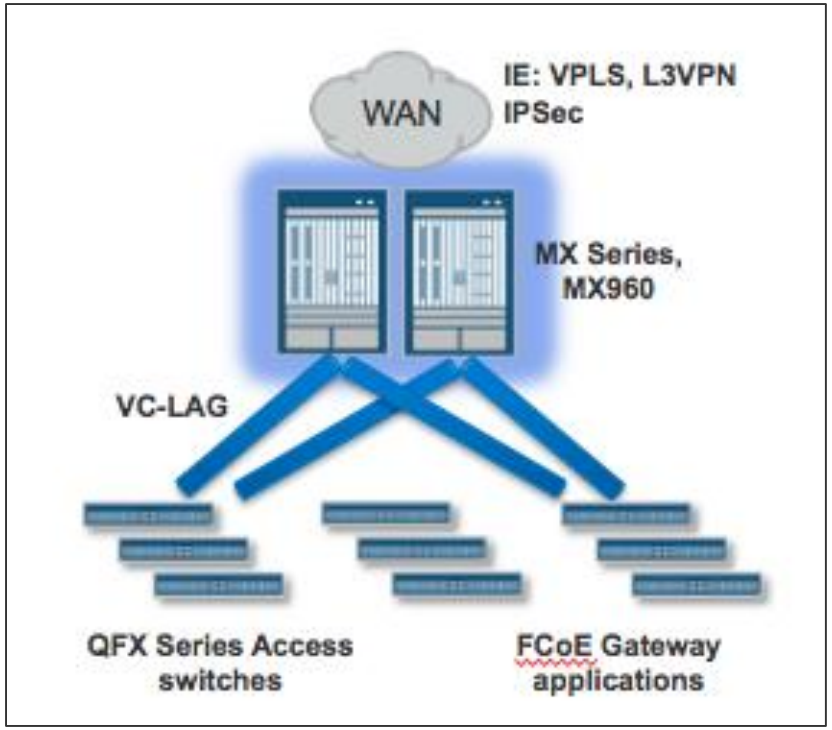
Data Center edge function + aggregation layer

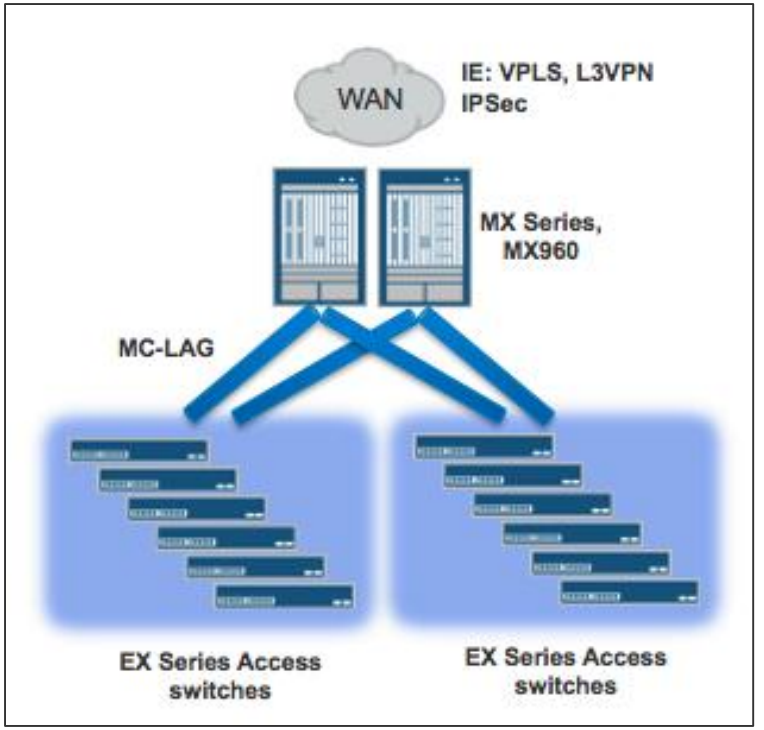
Use-case: Enterprise Data Center

Data Center edge function + single layer of LAN

Use-case: High Performance Data Center

COLLAPSED WAN AND AGGREGATION



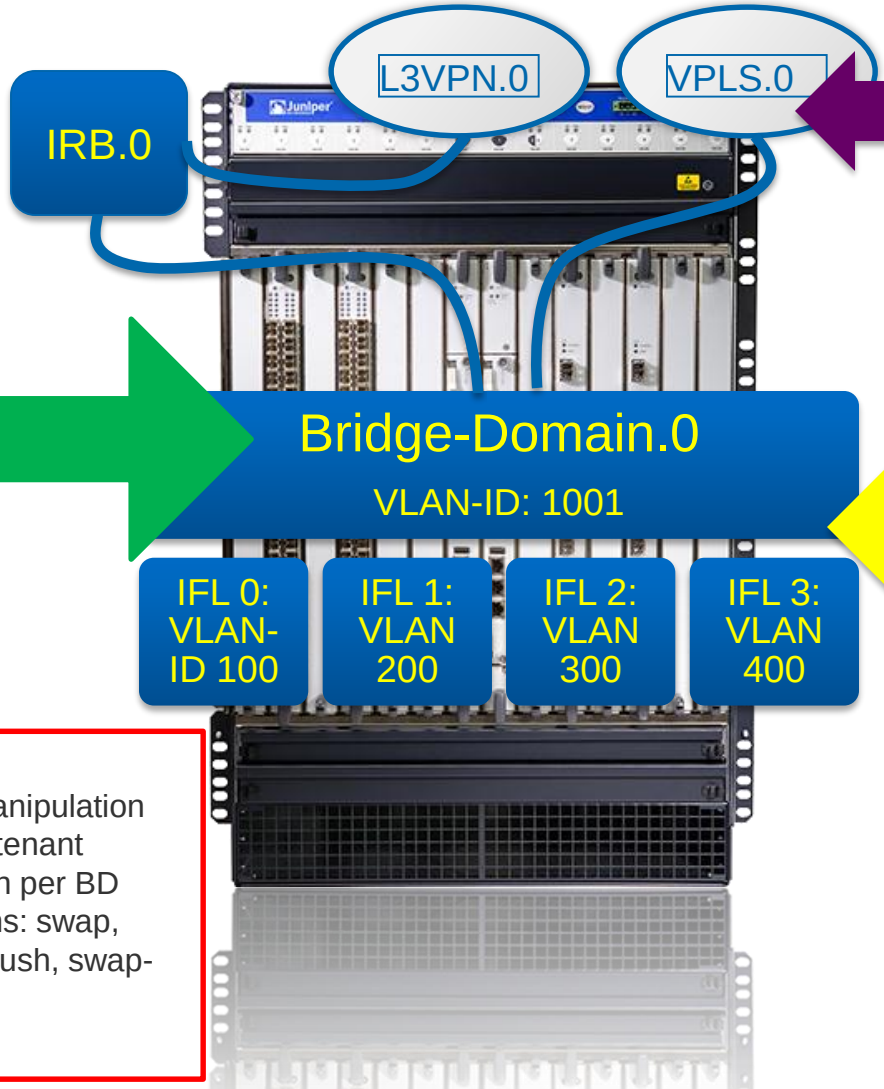


Collapsed WAN and aggregation solutions for small Data Center + access layers

Use-case: Low latency, financials, transaction

DATA CENTER LAN

L2 REFERENCE STYLE CONFIGURATION



WAN Instances Stitched per tenant

Bridge-Domain
L2 Flooding Domain
Typically one BD per cloud tenant
Assigned to tenant WAN instance
BD level VLAN tag preserved over WAN

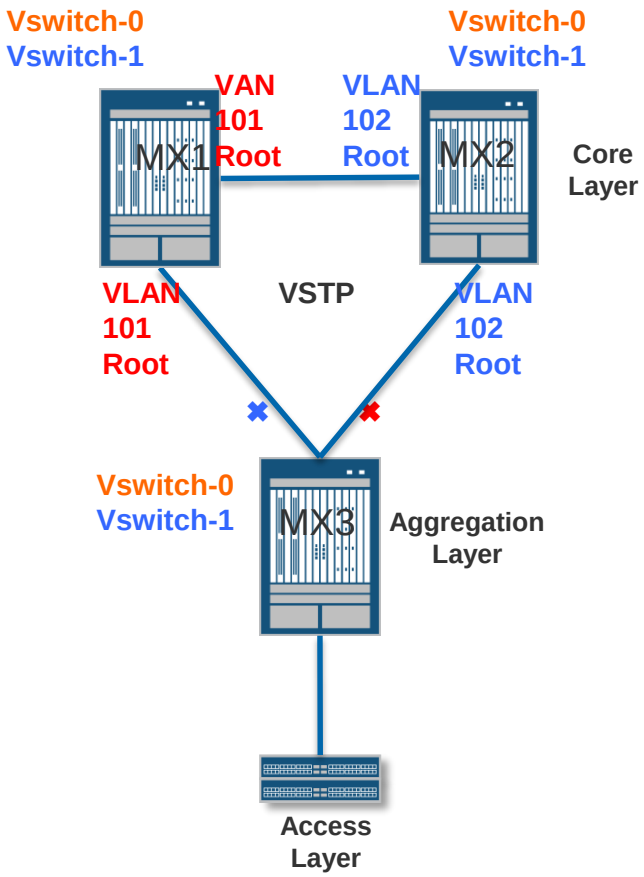
Multi-tenancy
Interface tags locally significant
IRB per tenant for L3 connectivity

Bridge Domain
Automatic port level VLAN manipulation
BD VLAN-ID used to identify tenant
4K VLAN / L2 learning domain per BD
Extensive VLAN manipulations: swap, pop, push, pop-swap, swap-push, swap-swap

L2 LOOP DETECTION CAPABILITIES 1/2

Active Loop Detection with xSTP

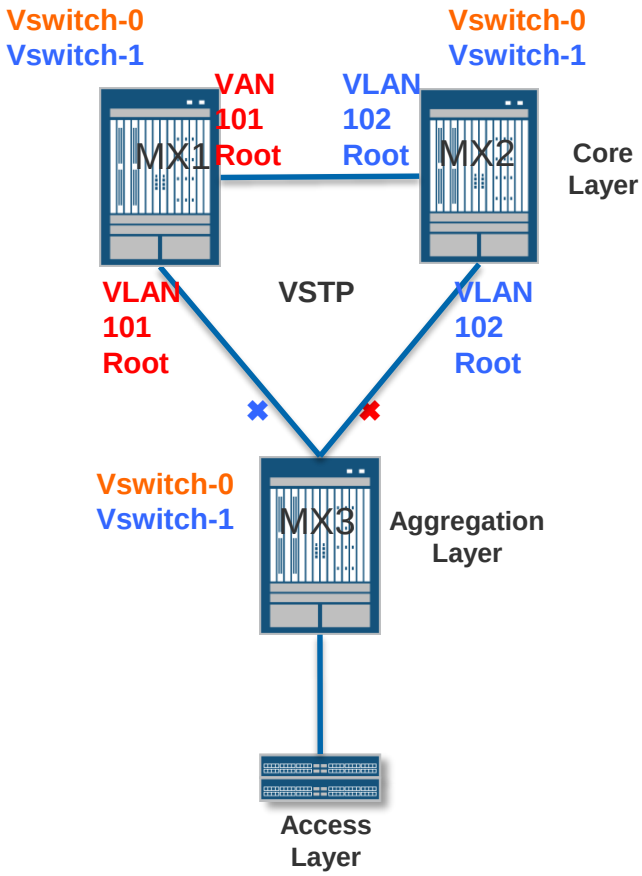
- STP flavors: **802.1D (802.1D-2004) STP, 802.1w RSTP, 802.1s MSTP, VSTP (compatible with PVST+ and RSTP+)**
- STP Protection: **BPDU Protection, Loop Protection, Root Protection**
- RSTP convergence rate: sub-second
  - 50 msec on point-to-point links
- Hello messages handled on line cards
  - Distributed processing for high scaling
- Separate xSTP instances with virtual-switch feature
  - @high scale with fully separate L2 domains
- NSR/NSB supported
- RL2GwP with MC-LAG A/A



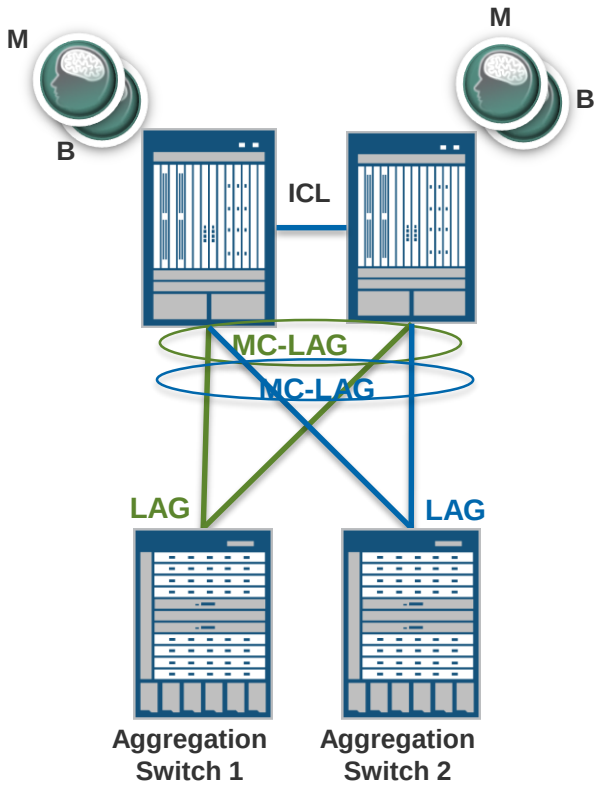
L2 LOOP DETECTION CAPABILITIES 2/2

Passive Loop Detection Capabilities

- MAC address tracking per IFL (ie: VLAN and VPLS)
- Isolation of the looping IFL, progressive actions leading to port blockage
- User configurable timers, block actions
- E-LAN



MULTI-CHASSIS LAG



Triangle and Square Topology options are supported

MC-LAG provides a single (virtual) LAG interface towards LAN

- LAG interface spread to 2 MX Series chassis
- Eliminates STP – Reverse L2 Gateway Protocol Support
- Active-Active and Active-Standby modes
- HA/load-balancing solution

Integrated Routing and Bridging (IRB)

- Same gateway MAC address across 2 MX Series switches → eliminates need for VRRP
- Essential for VM mobility

State replication between 2 independent MX Series platforms (MC-LAG Active-Active)

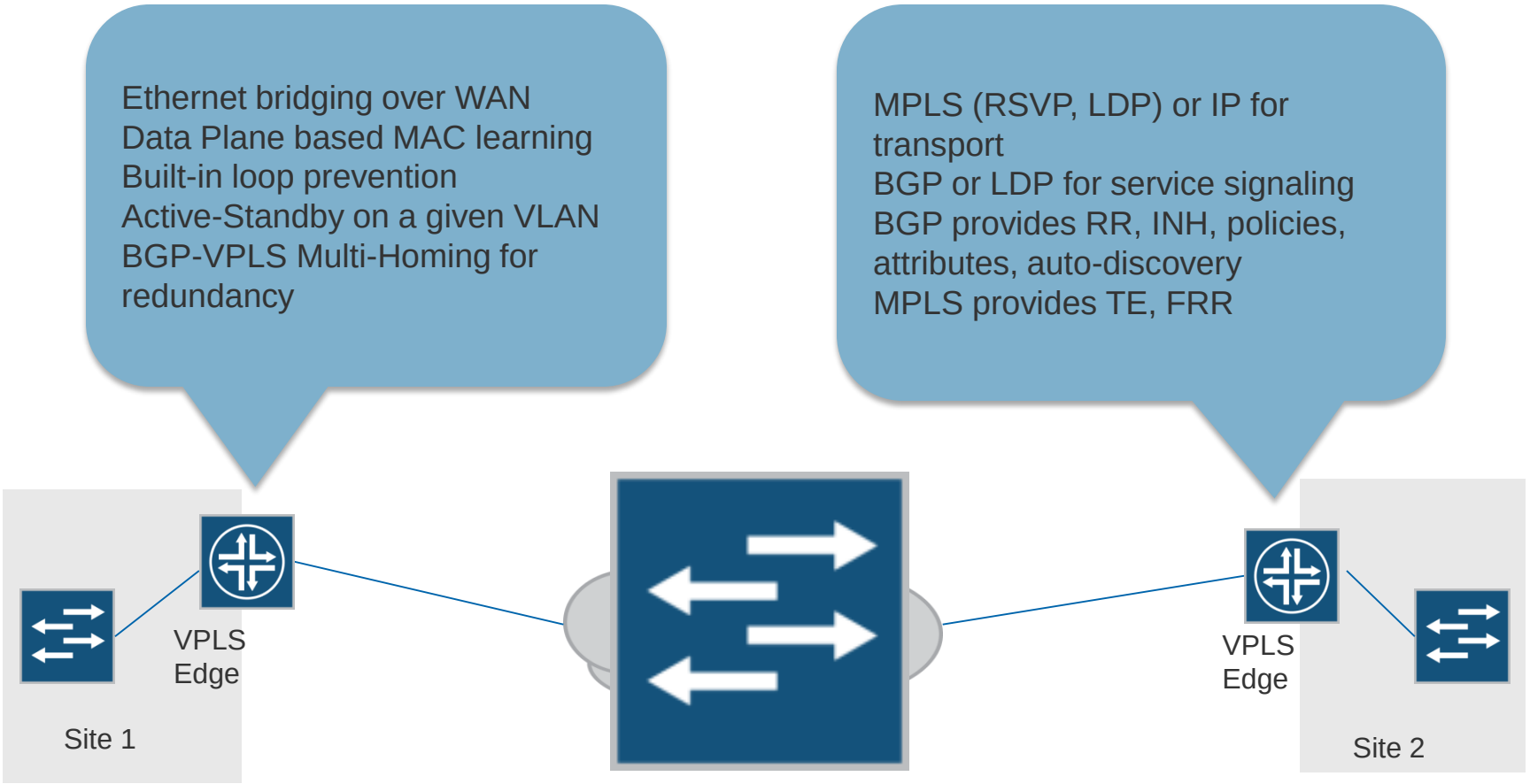
- L2, ARP, IGMP Snooping state replication
- LACP coordination

Support for bridged and routed setups

- L2 only, L2 & L3, L3 only options

DATA CENTER WAN

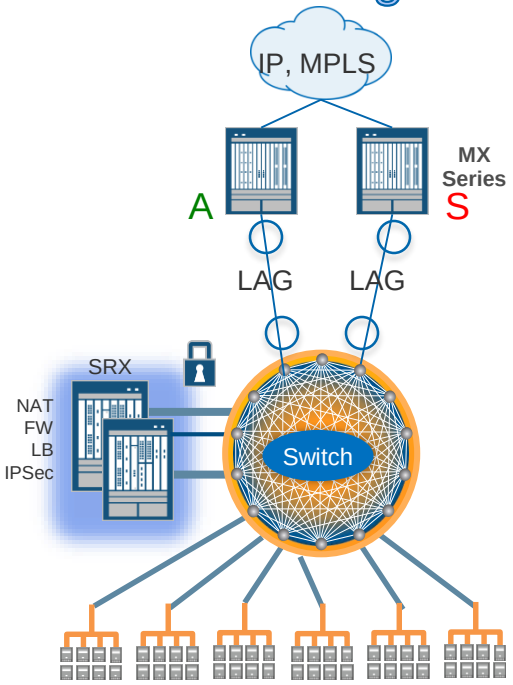
VPLS PROVIDES DATA CENTER CONNECTIVITY FOR L2 TRAFFIC



MULTIVENDOR – **PROVEN** – FAST RECOVERING – HIGH SCALE L2 CONNECTIVITY

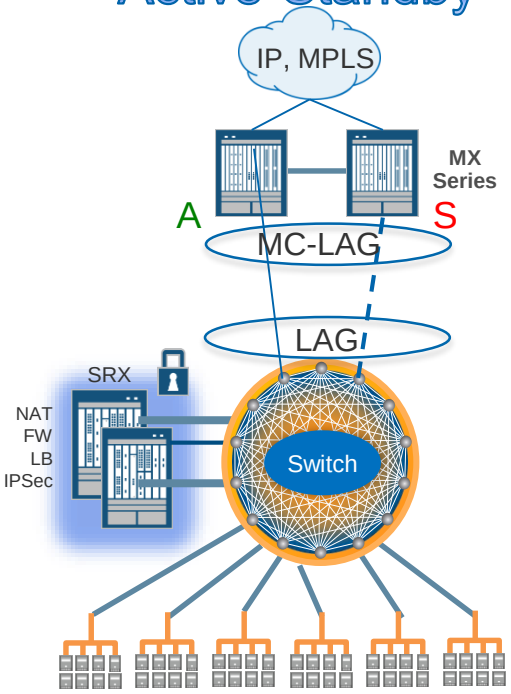
VPLS DEPLOYMENT OPTIONS WITH MIX – TODAY

VPLS Multi-Homing



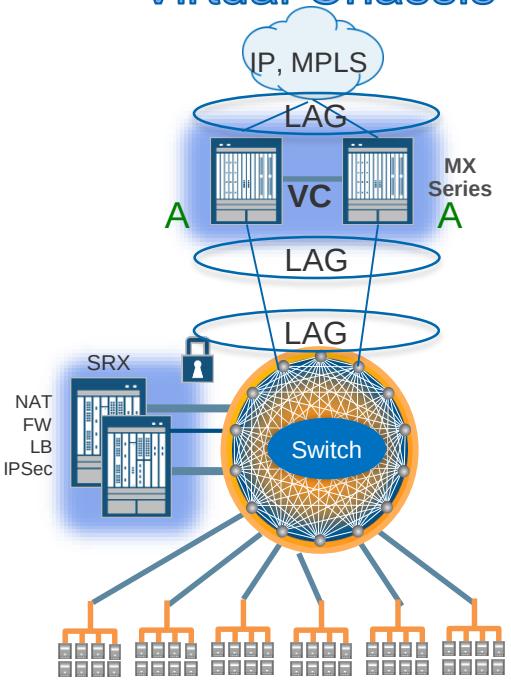
>1 VPLS devices
VPLS controlled Active-Standby
Per VLAN

VPLS with MC-LAG Active-Standby



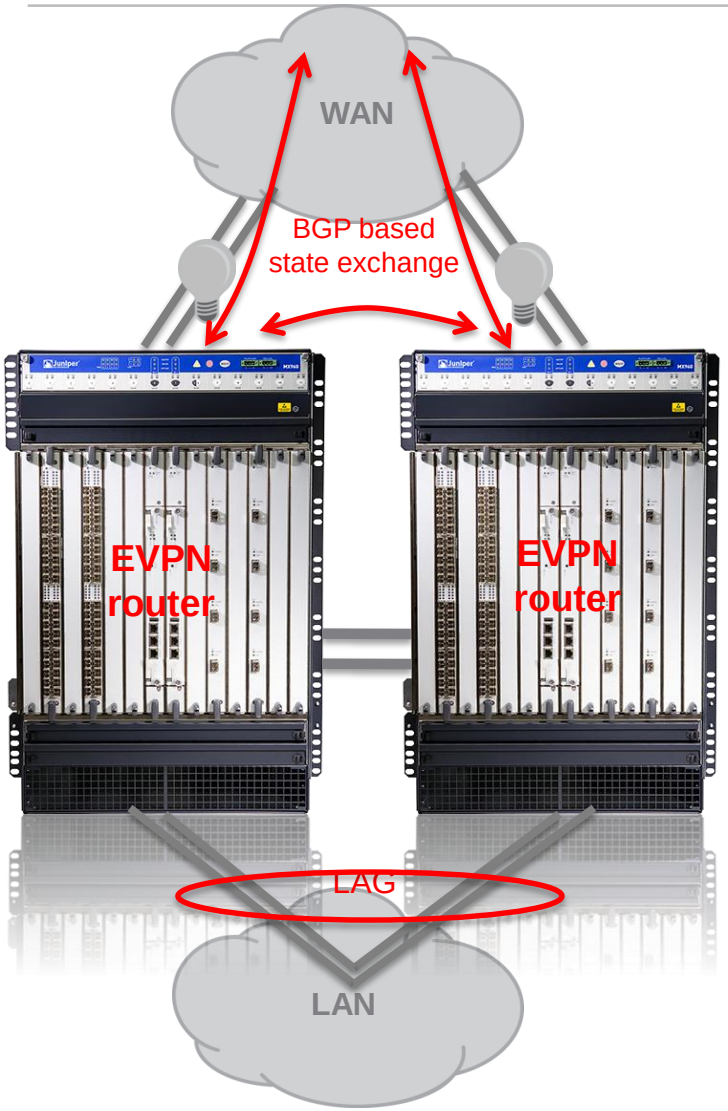
>1 VPLS devices
MC-LAG controlled Active-Standby on LAN
Per VLAN

VPLS with MX Virtual Chassis



One VPLS device
Active forwarding through all links of LAG

ETHERNET VPN INTRODUCTION



Ethernet-VPN a new standards based protocol

Inter-connects L2 domains over MPLS or IP backbone

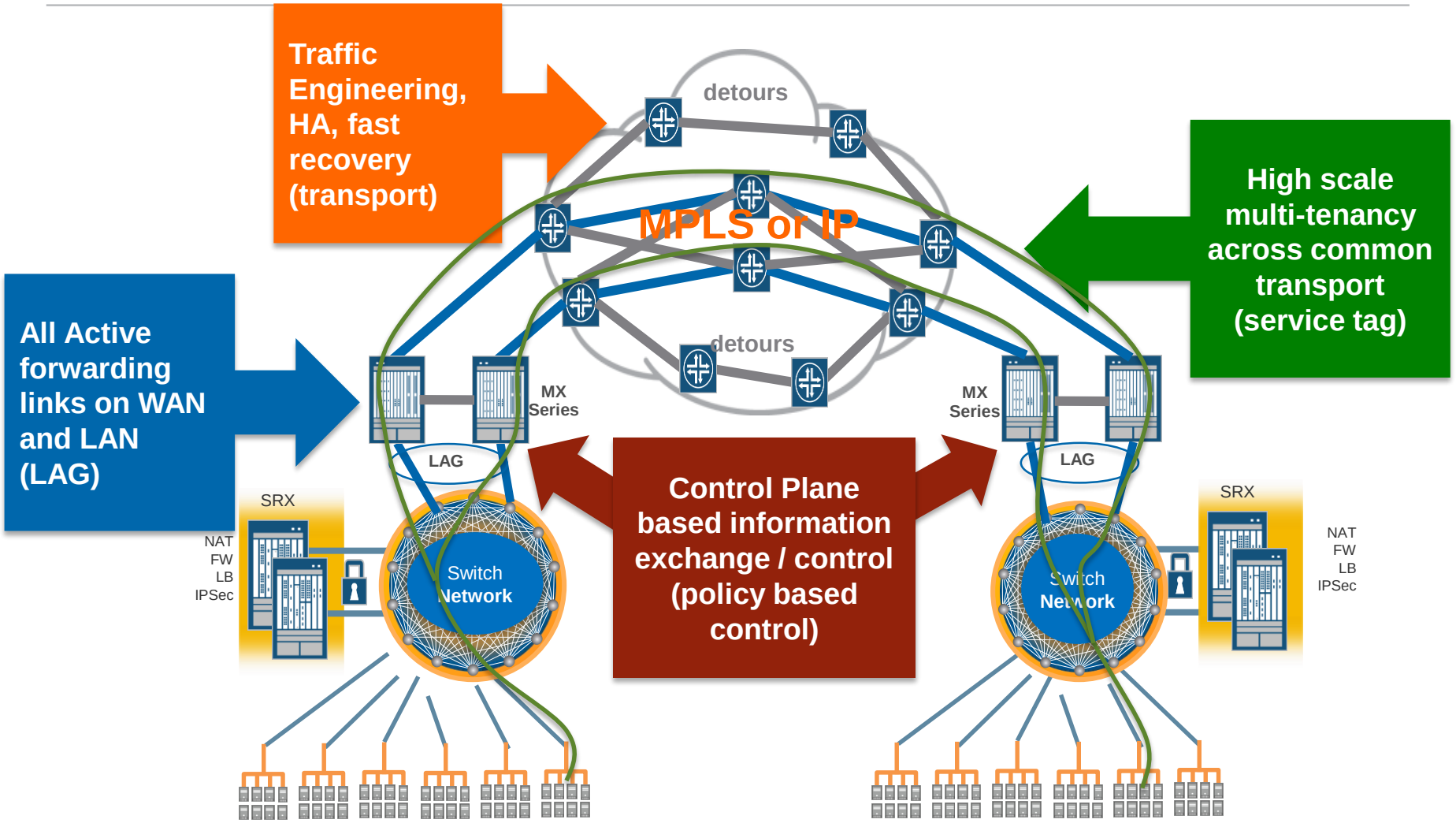
Multi-vendor initiative between JNPR, CSCO, ALU

Leverages BGP and MPLS strengths

Policy based learning, advertisement for controlled learning

L3 aware L2 solution

ADVANTAGES OF EVPN



EVPN L2 LOOP ELIMINATION CAPABILITIES

EVPN provides Active-Active and Active-Standby multi-homing options

Built-in L2 Loop Prevention capabilities

- Ethernet Segment Identifier (ESID)
  - Per VLAN / BD on the CE facing interface
  - Needed for all multi-homed deployments – to identify a (virtual) LAN instance
- Designated Forwarder Function
  - DF elected for a given ESID – designated L2-BUM authority
  - DF generates a (Split Horizon) MPLS label and distributes to all PEs – for each ESID
  - Non-DFs can send L2 BUM to MPLS network
    - Using this Split Horizon MPLS label – DFs identify own ESID and drop the packet
  - DFs send L2 BUM to MPLS network
    - Non-DFs drop the L2-BUM by default
- LAG
  - Required for Active-Active multi-homing
  - CE based loop prevention, single L2-BUM packet forwarding function

VPN POLICIES

EVPN provides several BGP based policies for better security and administrative capabilities

- Per customer MAC address discard or learning;
- Discard all or limit customer MAC learning;
- Per customer IP address discard or learning;
- Per site (ESI) based policies (discard or learn);
- BGP RT constraints for EVPN MAC routes
- At the Control Plane and at the Forwarding Plane
- EVPN will also have flooding on/off switch

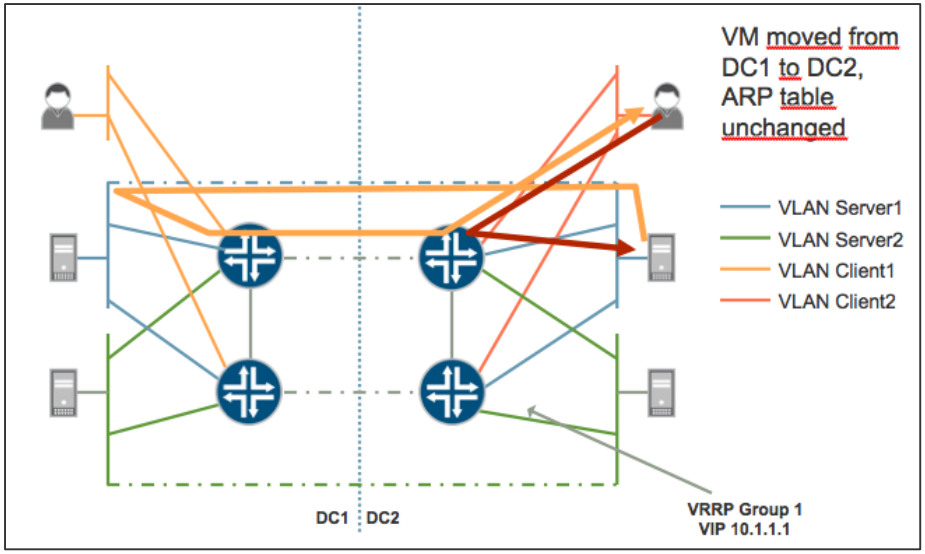
Addition, there is routing-instance level (ie: VPLS) L2-BUM rate limiting function

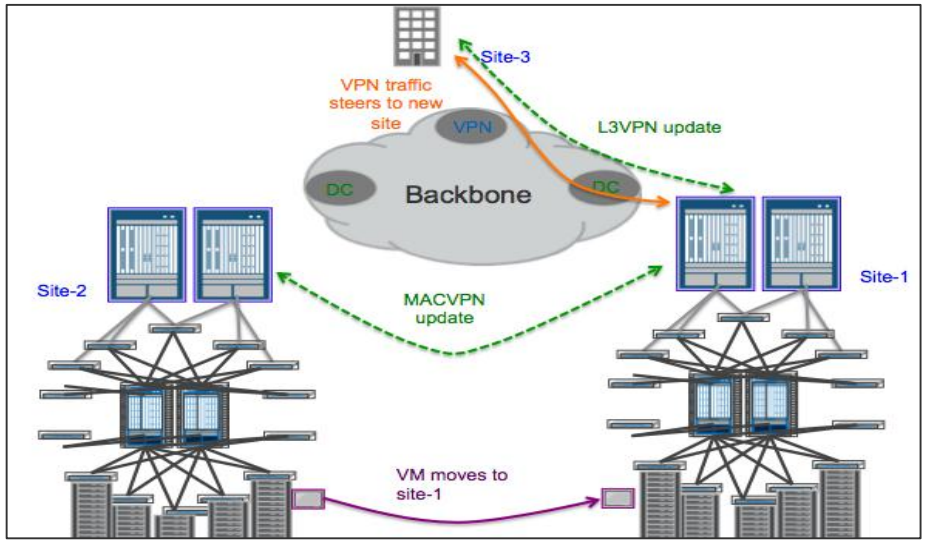
- Destination-mac-address: broadcast, multicast or unknown-unicast
- Action: rate-limit, filter or more

VM MOBILITY

CHALLENGES VM MOBILITY INTRODUCES

Challenges
L2 & L3 address no longer pinned to a site, interface
Fast convergence of network paths as VM moves
Ingress and Egress traffic convergence, optimization
Learning and information distribution control
L2 & L3 interaction for best user experience





SOLUTIONS FOR VM MOBILITY

Challenges	VPLS & L3VPN (TODAY)	EVPN & L3VPN (TOMORROW)
L2 & L3 address no longer pinned to a site, interface	L2: Split subnet supported by VPLS L3: Need provisioning help to advertise split subnet members	L2 and L3: Split subnet supported by EVPN and L3VPN
Fast convergence of network paths as VM moves	L2: MX implements integrated L2 and ARP learning (DP) Fast convergence through flooding	L2: MX implements integrated L2, ARP, L3 advertisement (DP, CP) Convergence through flooding and CP announcement
Ingress and Egress traffic convergence, optimization	Ingress, egress L2: automatic Ingress L3: provisioning based Egress L3: VRRP leverage	Ingress, Egress, L3, L3 automatic
Learning and information distribution control	L2: DP based learning no advertisement L3: BGP policies	L2 and L3: BGP policies
L2 & L3 interaction for best user experience	Limited	Full

PBB-EVPN COMPARISON ANALYSIS

WHAT IS PBB-EVPN?

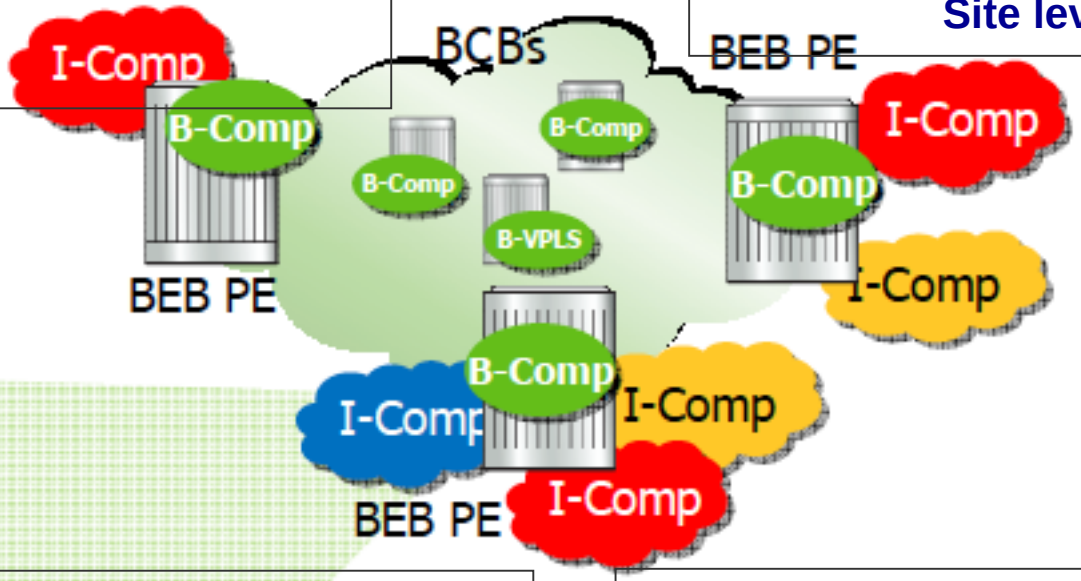
Latest draft: draft-ietf-l2vpn-pbb-evpn-04

Technology derived from PBB, and PBB-VPLS

Based on BGP

Uses Service-ID (I-SID) per tenant identification

Site level policy control

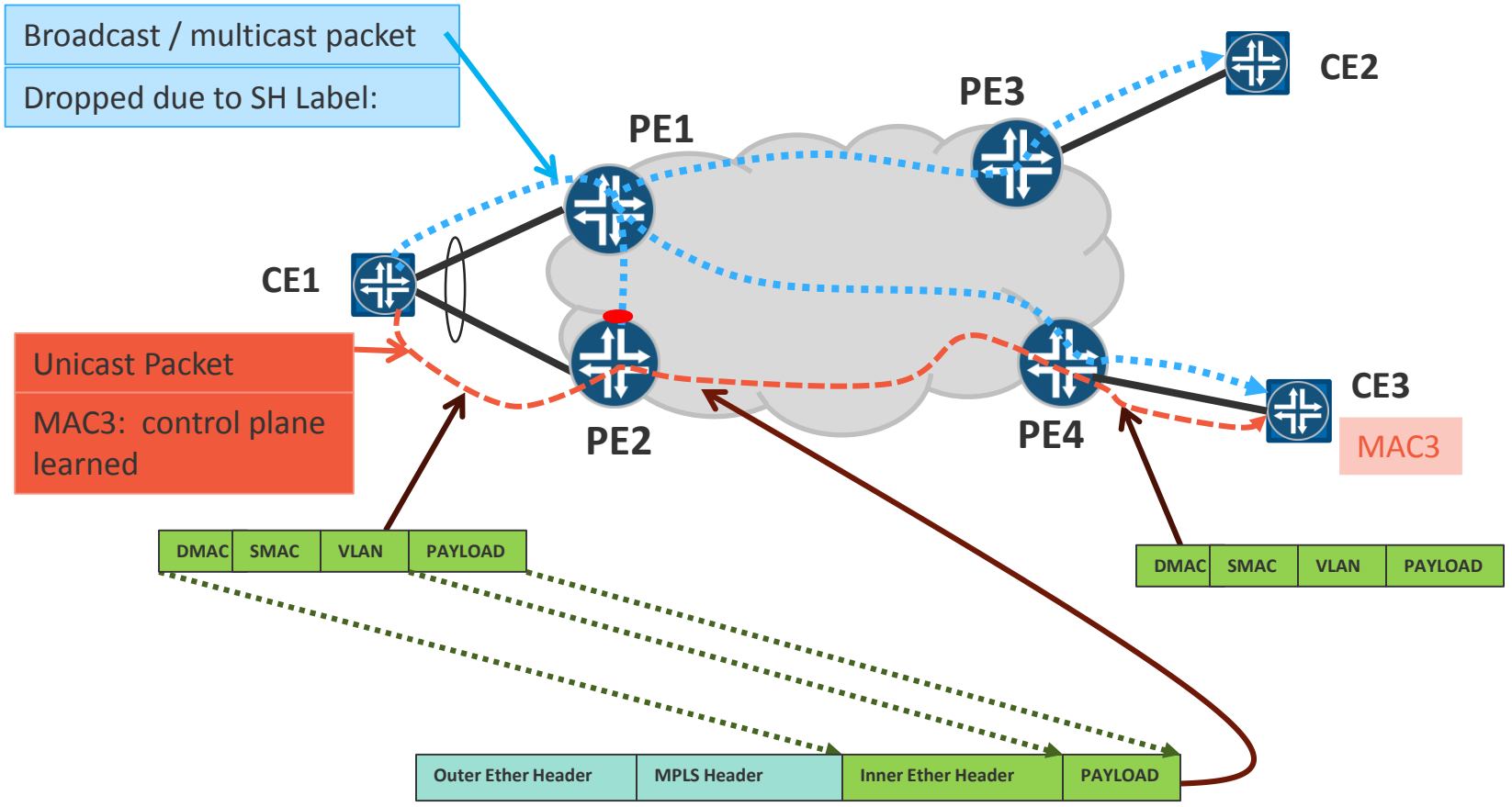


MAC in MAC encapsulation

Mostly advertises B-MACs, not C-MACs (Client MACs) – C-MACs are learned through data plane (*)

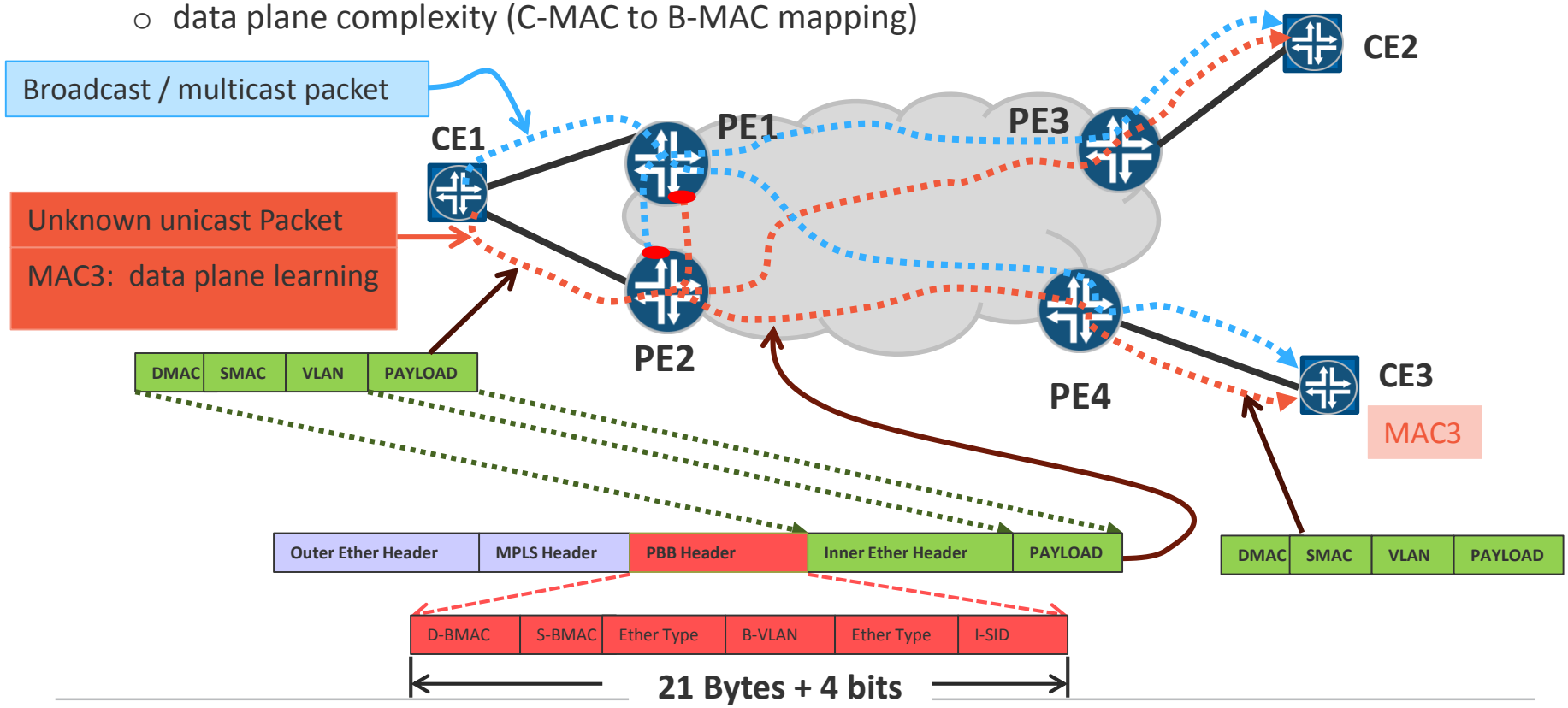
Claims high scale through not learning C-MACs across Control Plane

LIFE OF A PACKET WITH EVPN



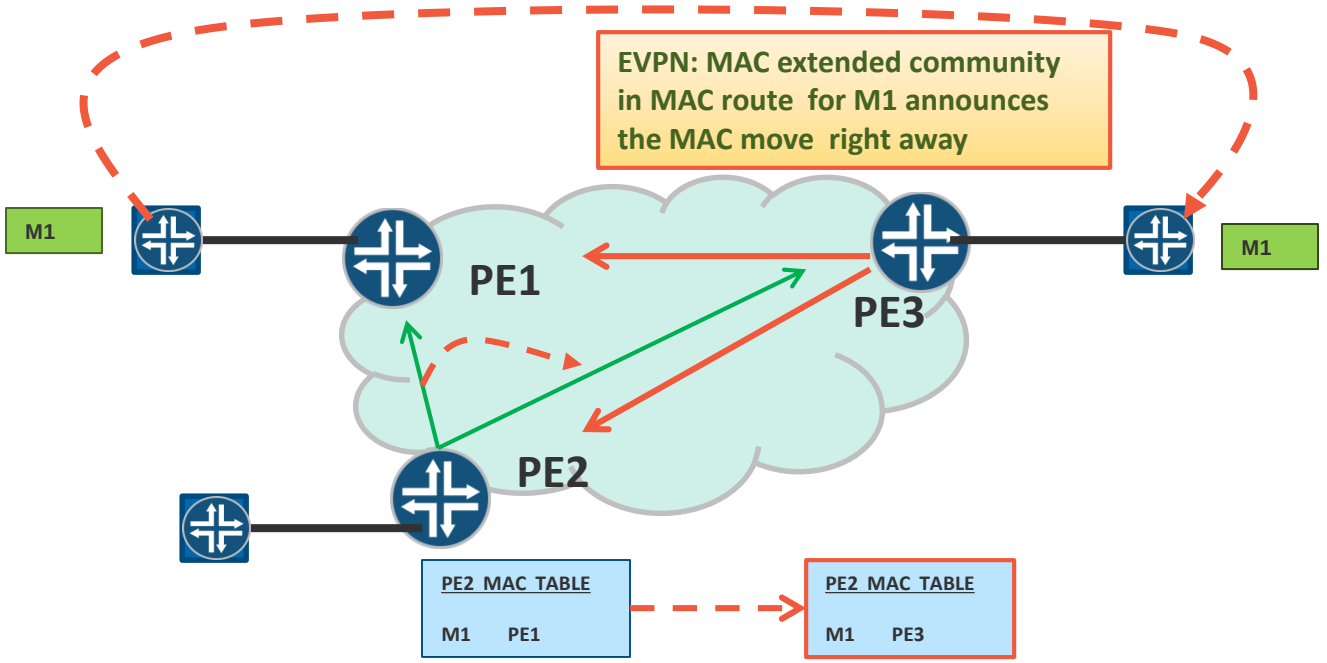
LIFE OF A PACKET WITH PBB-EVPN

- Extra PBB header encapsulation
  - waste network bandwidth
  - higher chances of packet fragmentation exceeding MTU
  - OAM complexity due to extra layer
  - data plane complexity (C-MAC to B-MAC mapping)



CUSTOMER MAC MOBILITY

- EVPN: Uses MAC extended community to announce the C-MAC move through the control plane. As soon as one of the PEs detects the C-MAC move, it informs all other PEs about it through the control plane.
- PBB-EVPN: Uses data plane learning & flooding to solve MAC mobility, causing unnecessary flooding and potentially L2 loops in the network



POLICY BASED MAC LEARNING

EVPN provides several BGP based policies for better security and administrative capabilities

- Per customer MAC address discard or learning;
- Discard all or limit customer MAC learning;
- Per customer IP address discard or learning;
- Per site (ESI) based policies (discard or learn);
- BGP RT constraints for EVPN MAC routes

PBB-EVPN only provides per site based policy.

PBB-EVPN HYPES AND REALITY

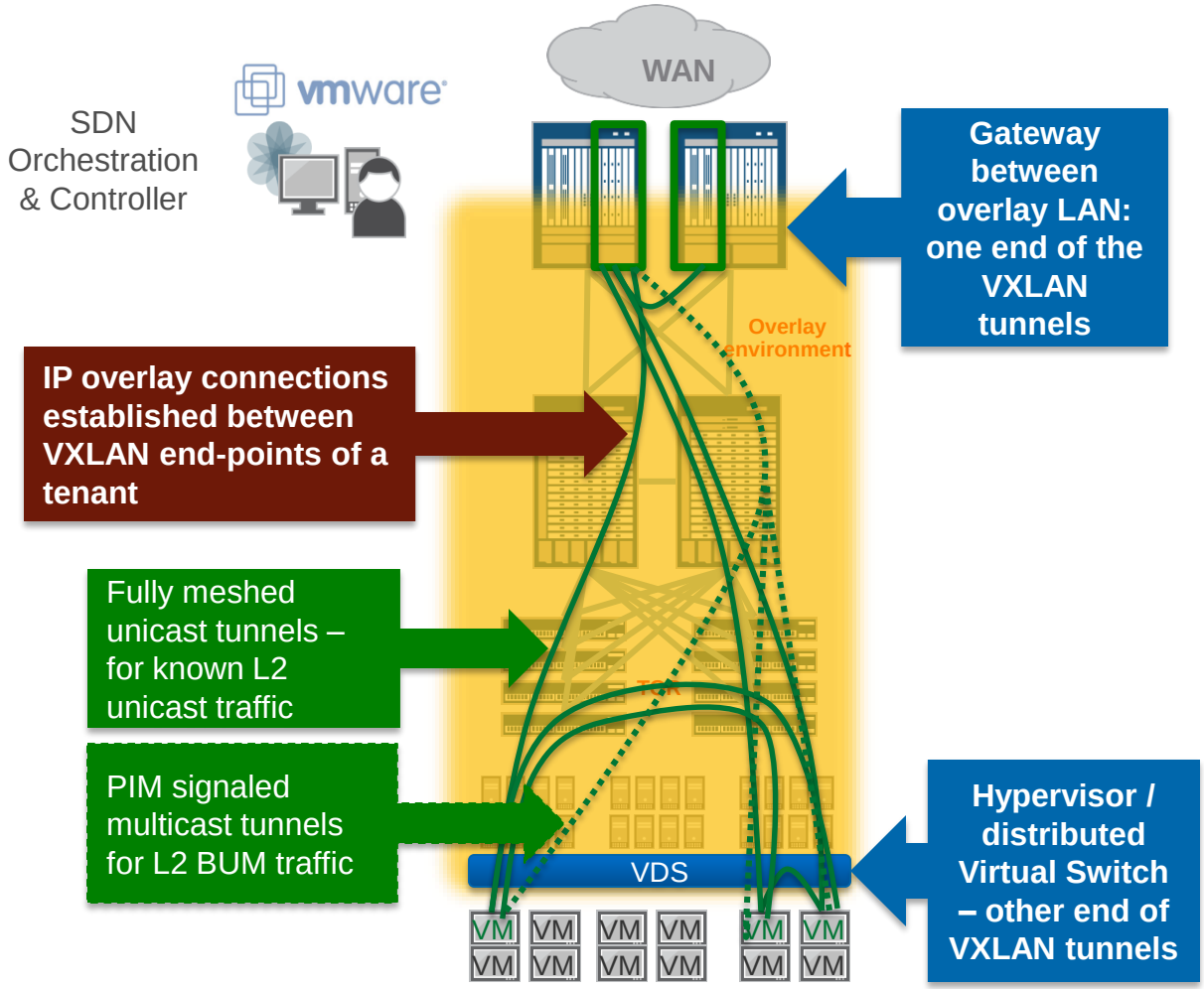
1. PBB-EVPN provides better MAC scaling
   - Reality: Attempts to solve control plane scaling (RIB) by adding more complexity in data plane (costlier resource)
   - The forwarding path (FIB) requirements are same for EVPN & PBB-EVPN
2. PBB-EVPN supports confinement of C-MAC Learning
   - Reality: C-MAC learning procedures are based on data plane flooding for unknown unicast MACs (uncontrolled MAC learning)
   - EVPN uses BGP policies to solve the problem with full control to service provider
3. PBB-EVPN supports C-MAC mobility with MAC summarization
   - Reality: Rely on data plane learning to solve MAC mobility causing unnecessary flooding and potentially L2 loops
   - EVPN solves the problem for both MAC and IP as it carries them in the BGP control plane

EVPN TECHNOLOGY COMPARISON

Requirement	VPLS	PBB-EVPN	EVPN
PE Auto-discovery	✓	✓	✓
Access multi-homing (Single Active)	✓	✓	✓
All-active multi-homing (flow based load balancing)		✓	✓
Fast convergence on failure	✓	✓	✓
MAC mobility	✓	✓	✓
Control plane scaling of large number of MACs	(No CP)	✓	✓
Policy up to MAC/IP address level granularity (enable flexible topologies)			✓
Avoid flooding of unknown unicast MAC and ARP			✓
Faster convergence based on local repair during failures			✓
Optimized VM (MAC+IP) mobility & inter-subnet routing			✓
Ease of configuration, operation and manageability			✓
Per service COS			✓

OVERLAY CONNECTIVITY FOR LAN

VXLAN OVERLAY GATEWAY ON MX



Virtual eXtensible Local Area Network (VXLAN)

L2 connections within IP overlay

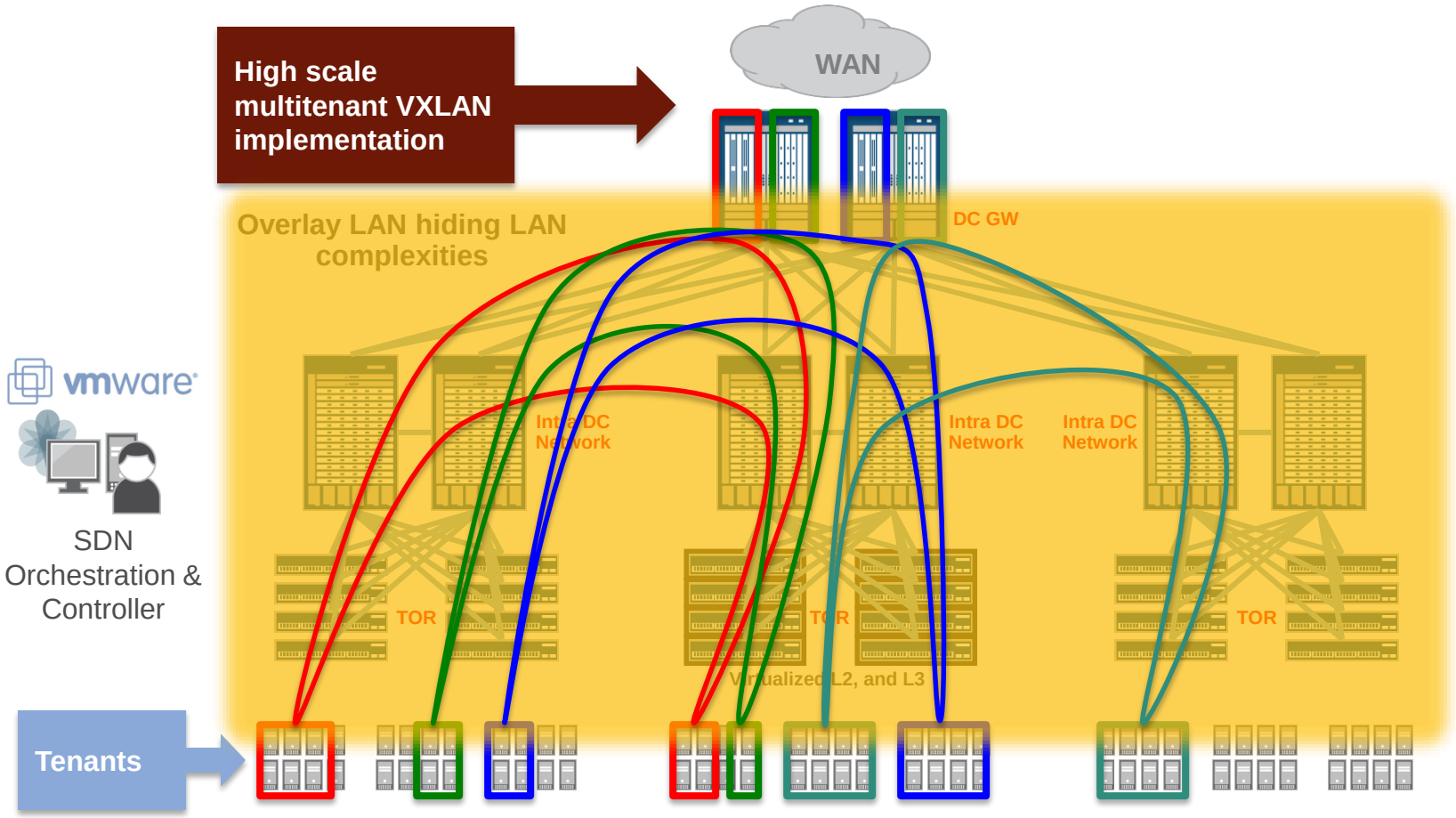
- **Unicast & multicast**

Delivering a solution agnostic of Ethernet switch layers

Allows flat DC design w/out boundaries

Providing network design simplicity and elasticity

VXLAN MULTITENANT IMPLEMENTATION



Hypervisor based environment, terminating VXLAN tunnels on MX and on virtual-switch of servers
Each tenant having its own virtual network slice
Connecting cloud assets on LAN with customers coming from WAN

VXLAN MODES

Data Plane Based

Network Orchestrator

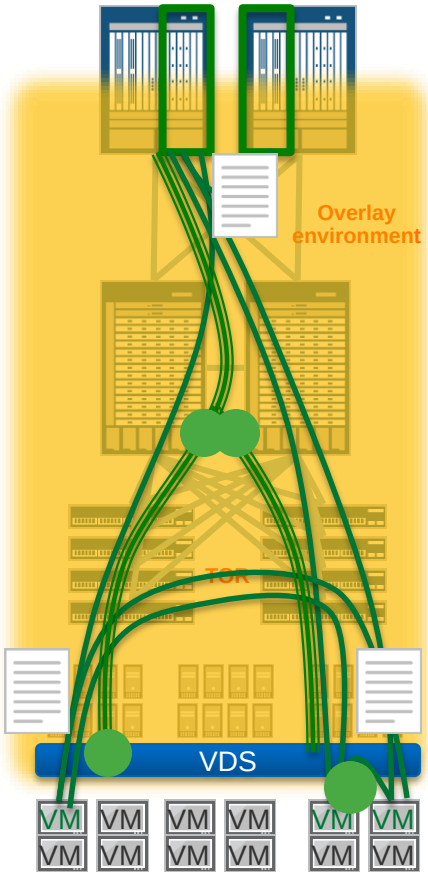
VXLAN IETF draft based

Inline – Data-Plane based learning and forwarding

Uses IP multicast for learning and L2-BUM traffic forwarding

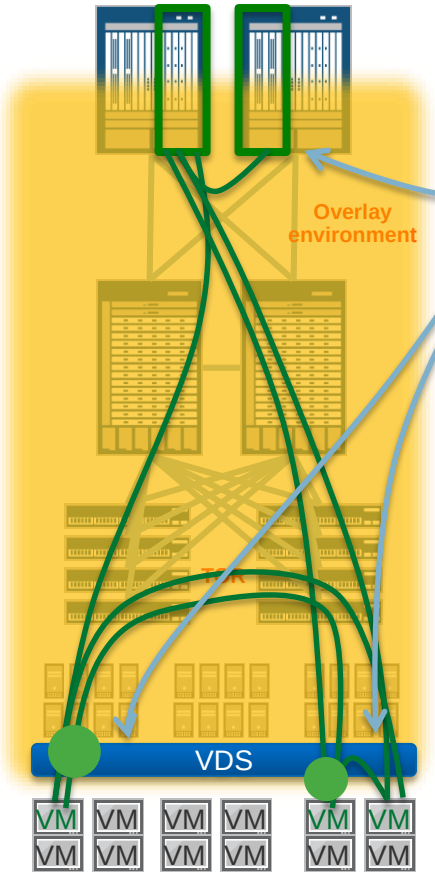
Builds P2P tunnels dynamically

Minimal configuration required



Control Plane Based

Vmware NVP or JNPR vContrail SDN Controller



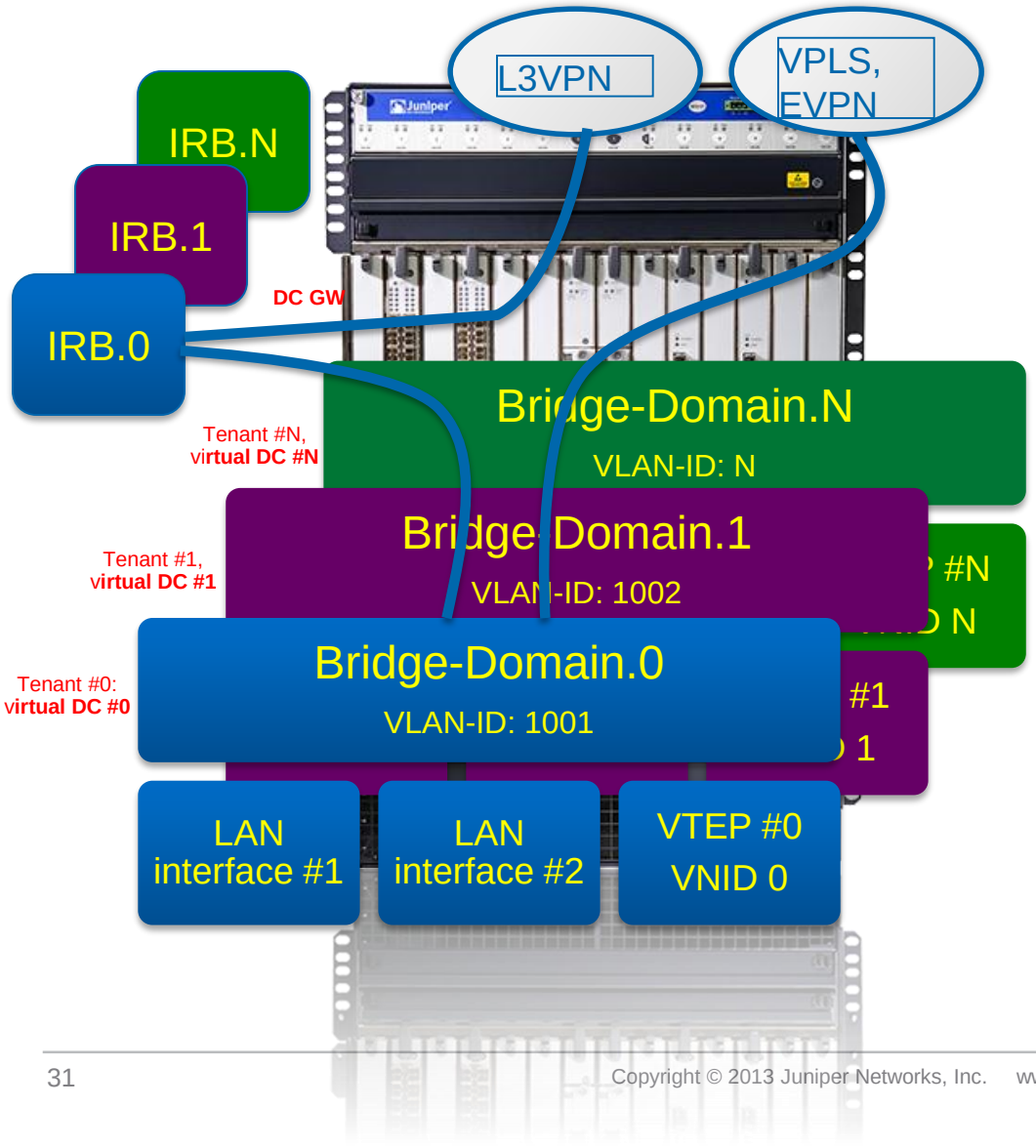
SDN controller providing forwarding info – L2 & ARP

Combination of Data-Plane and Control Plane based learning and forwarding

P2P tunnels built by the controller

API includes CP and MP information

VXLAN PART OF UNIVERSAL GATEWAY FUNCTION ON MX



High scale multi-tenancy

- VTEP tunnels per tenant
- P2P, P2MP tunnels

Tie to full L2, L3 functions on MX

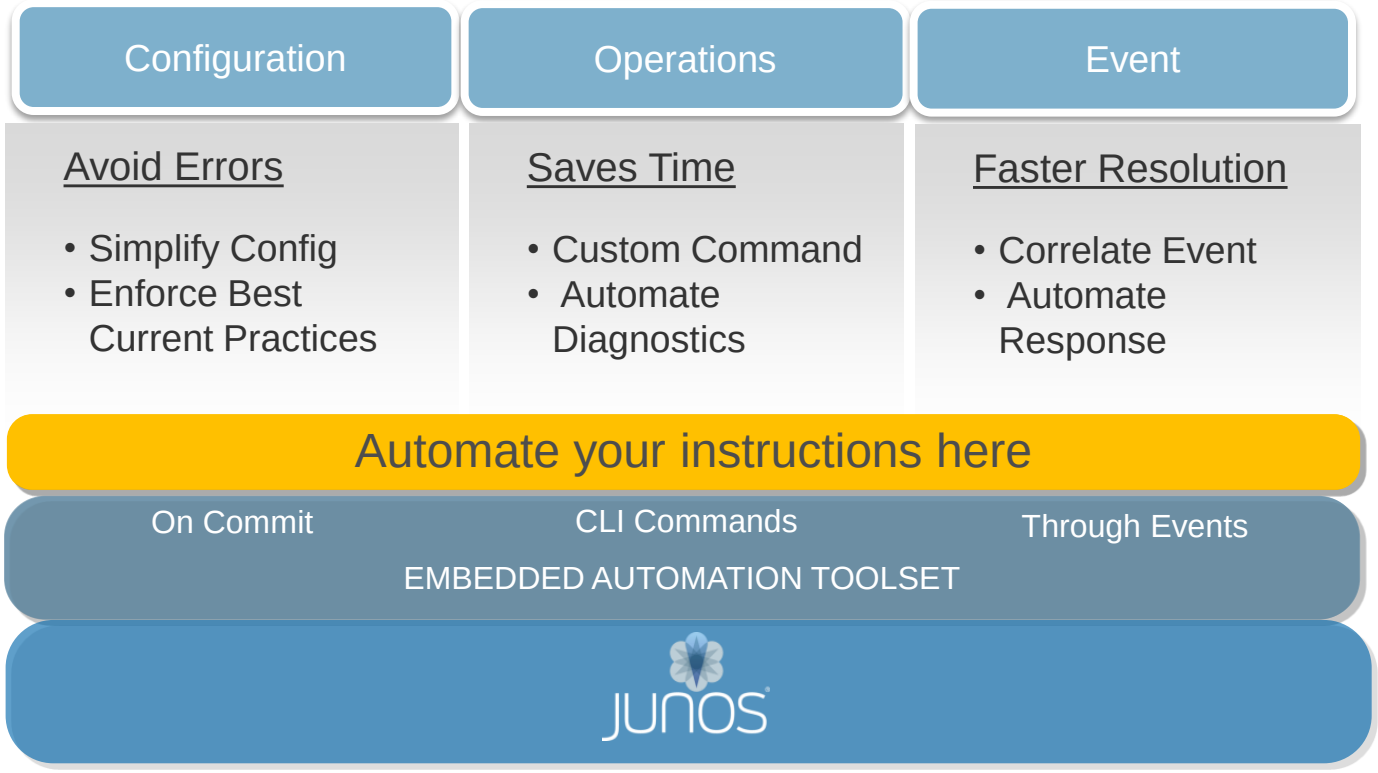
- Unicast, multicast forwarding
- IPv4, IPv6
- L2: Bridge-domain, virtual-switch

Gateway between LAN, WAN and Overlay

- Ties all media together
- Giving migration options to the DC operator

PLATFORM MANAGEABILITY

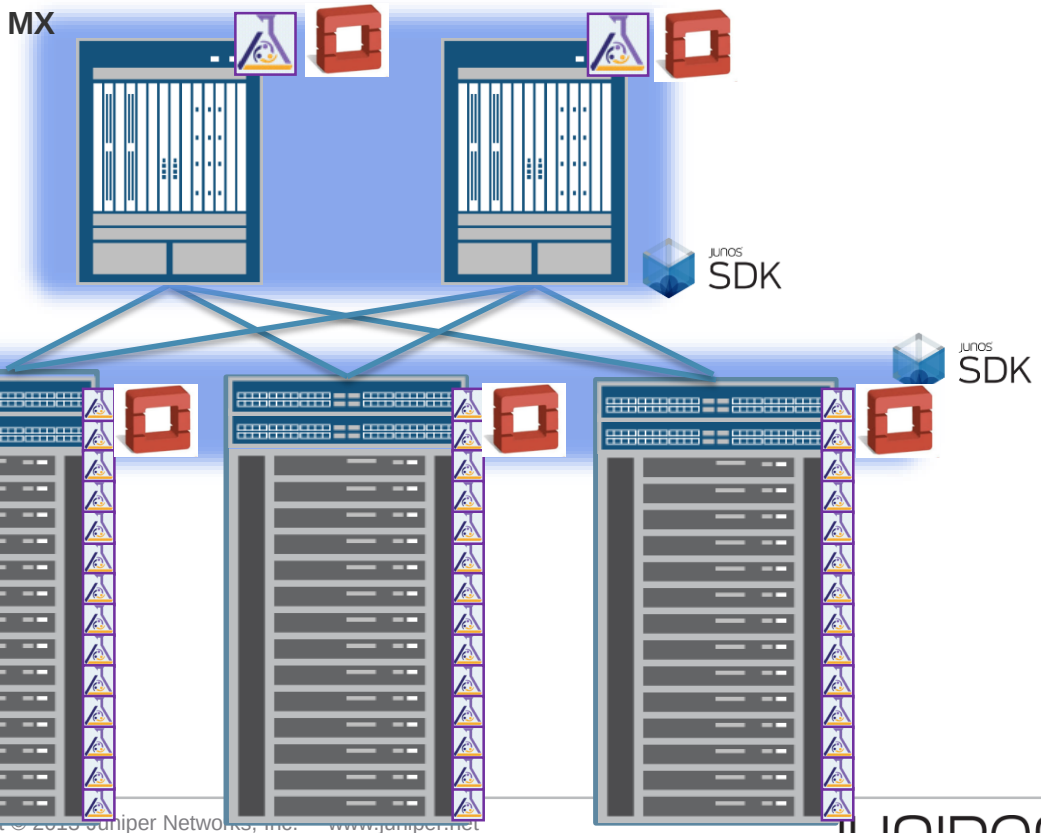
PROGRAMMABILITY WITH AUTOMATION
PROGRAMMABLE CONFIGURATION TEMPLATE



- ✓ Create user-defined service template in CLI
- ✓ Automatically build the required interface (API) with the underlying platform-specific implementation
- ✓ Provisioning system invokes an uniform interface (API) via NETCONF by passing mandatory/optional service parameters

PLATFORM API'S FOR ORCHESTRATION

- Automate orchestration of network along with VMs
- Network abstraction enabling faster resolution to change requests
- Workflow for the network – Layer 2, Layer 3 interfaces, Bridge Domains



Native Puppet Agent running on Junos1

Native Openstack Agent on Junos

junos SDK NETCONF API enables programmatic configuration changes and operational management

everywhere