

Virtual Private Networks

Eduardo Grampín Castro grampin@fing.edu.uy Universidad de la República Uruguay

based on several industrial presentations

Outline



- L3 MPLS VPN
 - Review
- L2 VPN
 - VXLAN
 - EVPN
 - BGP control plane
 - Use cases

What is a Virtual Private Network?

- UNIVERSIDAD DE LA REPUBLICA URUGUAY
- VPN is a set of sites or groups which are allowed to communicate with each other
- VPN is defined by a set of administrative policies
 - Policies established by VPN customers
 - Policies could be implemented completely by VPN service providers
- Flexible inter-site connectivity
 - Ranging from complete to partial mesh
- Sites may be either within the same or in different organizations
 - VPN can be either intranet or extranet
- Site may be in more than one VPN
 - VPNs may overlap
- Not all sites have to be connected to the same service provider
 - VPN can span multiple providers

Traditional vs L3 VPNs



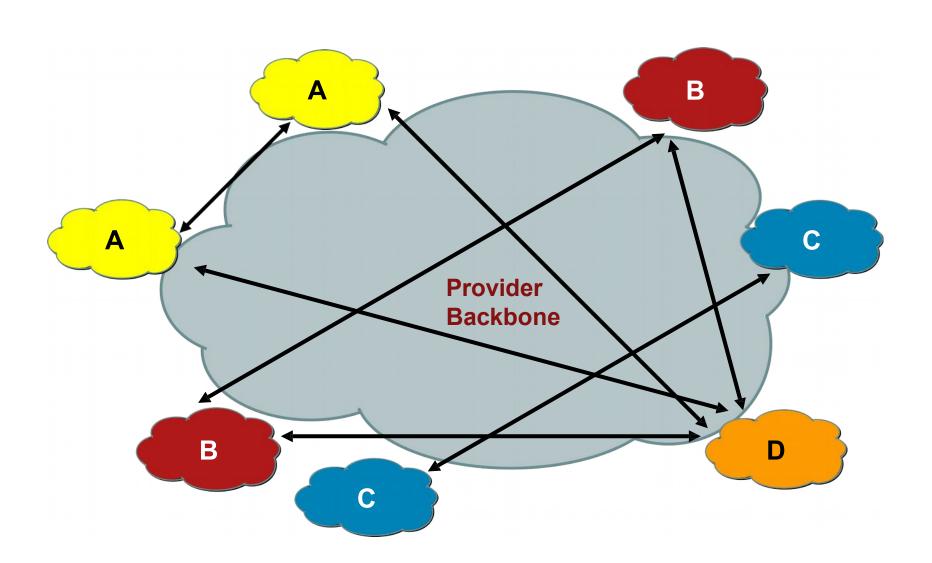
- Traditional VPNs (wo. MPLS)
 - Customer endpoints (CPE) connected via Layer 2 such as Frame Relay DLCI, ATM VC or point-to-point connection
 - Provider network is not responsible for distributing site routers as routing relationship is between the customer endpoints
 - Provider will need to manually fully mesh end points if any-to-any connectivity is required

Layer 3 VPN

- Customer end points peer with providers' routers at L3
- Provider network responsible for distributing routing information to VPN sites
- Don't have to manually fully mesh customer endpoints to support any-toany connectivity

VPN model









how to



checkmate VPNs in three moves:
(1) make PEs reachable each other using IP,
(2) use BGP for announcing customer prefixes, and
(3) use MPLS for tunnels inside the backbone

MPLS VPN: I.cittadini, m.cola, g.di battista, m.patrignani

Implementation

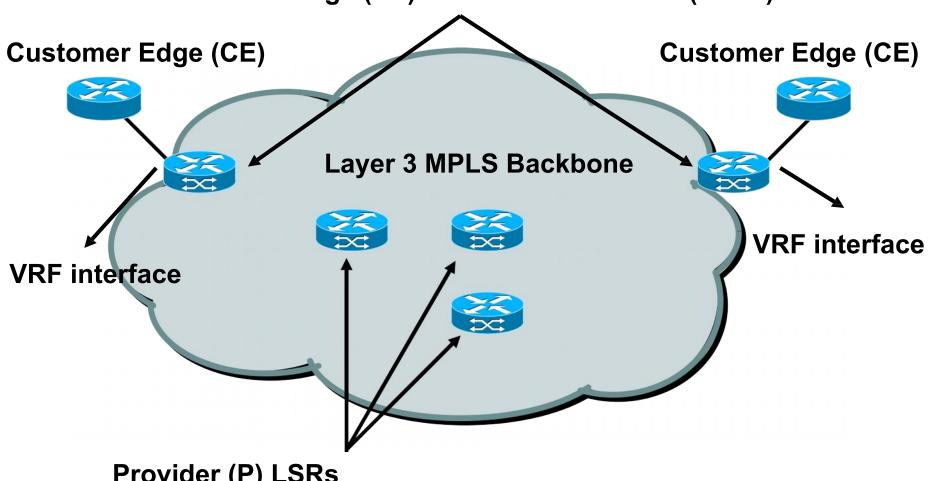


- Control Plane
 - BGP
- Data Plane
 - IP/MPLS using IGP and MPLS signaling (e.g. LDP)

MPLS VPN Architecture



Provider Edge (PE) Label Switch Router (LSRs)



Provider (P) LSRs



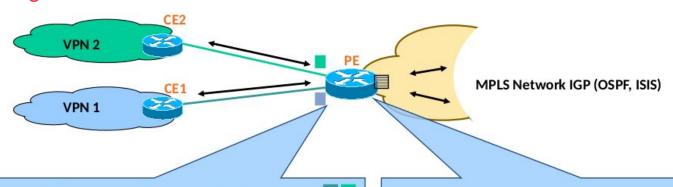
- MPLS framework (labels) in the core
 - IGP (any)
 - LDP or MPLS Traffic Engineering
- VRF (Virtual Routing/Forwarding) context to keep VPNs seperate
 - VRF on PE interface towards CE

VRF knowledge only needed on edge routers

- VRF routing table
- RD attached to prefixes to make VPN prefixes unique
 - RD is 64 bits
 - RD allows for overlapping VPN prefixes
- Route targets (ext BGP community) attached to VPN prefixes to allow prefixes to be imported/exported to VPNs
- BGP in the core to advertise VPN prefix and VPN label to all Provider Edge (PE) routers



Separate Routing Tables at PE



Customer Specific Routing Table

- Routing (RIB) and forwarding table (CEF) dedicated to VPN customer
 - VPN1 routing table
 - VPN2 routing table
- Referred to as VRF table for <named VPN>

IOS: "show ip route vrf <name>"
IOS-XR:"sh route vrf <name> ipv4
NX-OS: "sh ip route vrf <name>"

Global Routing Table

- Created when IP routing is enabled on PE.
- Populated by OSPF, ISIS, etc. running inside the MPLS network

IOS: "show ip route"

IOS-XR:"sh route ipv4 unicast"

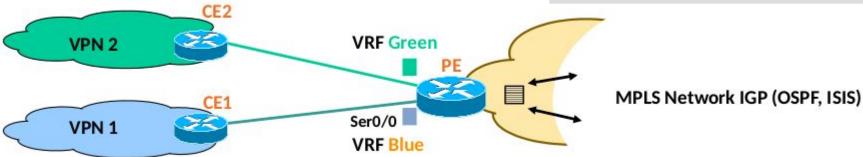
NX-OS: "sh ip route"



Virtual Routing and Forwarding Instance

- What's a Virtual Routing and Forwarding (VRF) ?
 - Representation of VPN customer inside the MPLS network
 - Each VPN is associated with at least one VRF
- VRF configured on each PE and associated with PE-CE interface(s)
 - Privatize an interface, i.e., coloring of the interface
 - No changes needed at CE

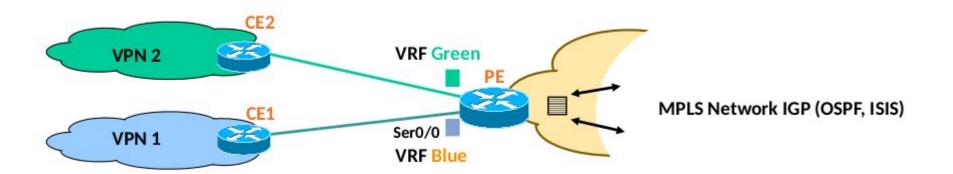
IOS_PE(conf)#ip vrf blue
IOS_PE(conf)#interface Ser0/0
IOS_PE(conf)#ip vrf forwarding blue





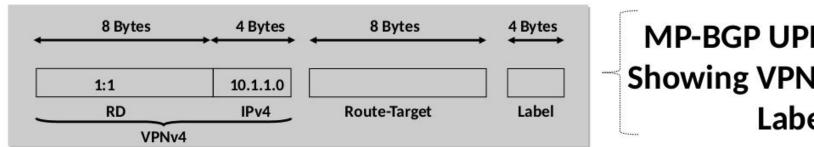
Virtual Routing and Forwarding Instance

- PE installs the internal routes (IGP) in global routing table
- PE installs the VPN customer routes in VRF routing table(s)
 - VPN routes are learned from CE routers or remote PE routers
 - VRF-aware routing protocol (static, RIP, BGP, EIGRP, OSPF) on each PE
- VPN customers can use overlapping IP addresses
 - BGP plays a key role. Let's understand few BGP specific details..





Control Plane = Multi-Protocol BGP (MP-BGP)

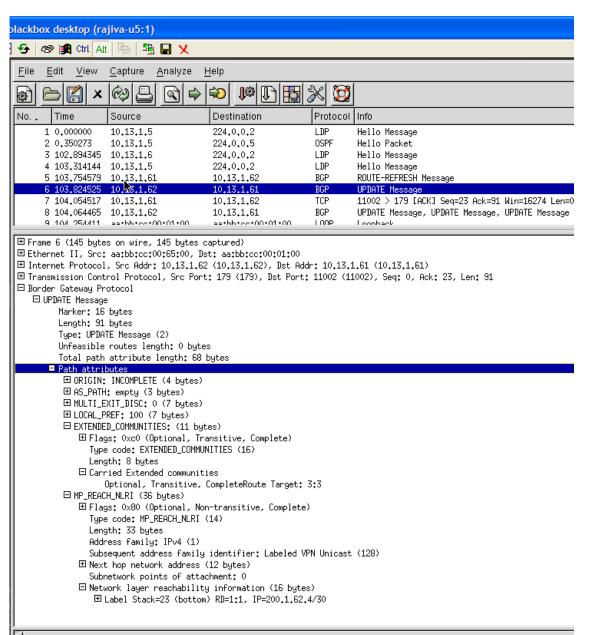


MP-BGP UPDATE Message Showing VPNv4 Address, RT, Label only

- MP-BGP Customizes the VPN Customer Routing Information as per the Locally Configured VRF Information at the PE using:
 - Route Distinguisher (RD)
 - Route Target (RT)
 - Label



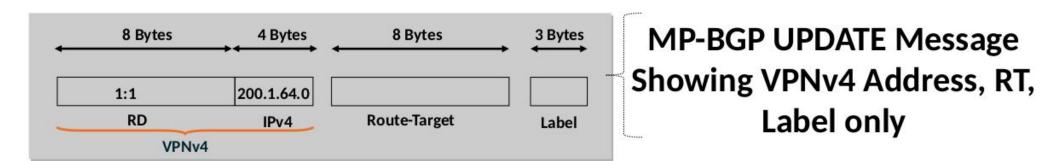
MP-BGP UPDATE Message Capture



- Visualize how the BGP UPDATE message advertising VPNv4 routes looks like.
- Notice the Path Attributes:
 - Route Target = 3:3
 - VPNv4 Prefix 1:1:200.1.62.4/30; Label = 23



Route-Distinguisher (rd)

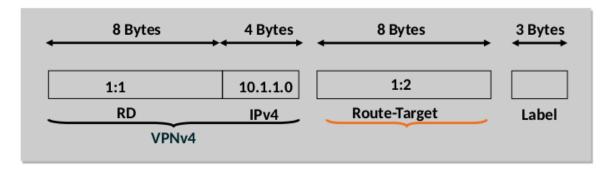


- VPN customer IPv4 prefix is converted into a VPNv4 prefix by appending the RD (1:1, say) to the IPv4 address (200.1.64.0, say) => 1:1:200.1.64.0
 - Makes the customer's IPv4 address unique inside the SP MPLS network.
- Route Distinguisher (rd) is configured in the VRF at PE
 - RD is not a BGP attribute, just a field.

```
IOS_PE#
!
ip vrf
green
 rd 1:1
!
```



Route-Target (rt)



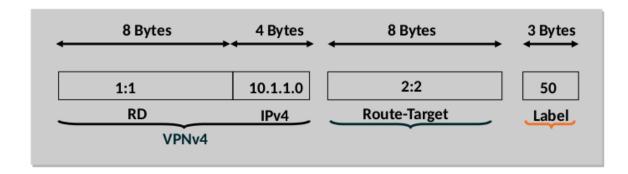
- Route-target (rt) identifies which VRF(s) keep which VPN prefixes
 - rt is an 8-byte extended community attribute.
- Each VRF is configured with a set of route-targets at PE
 - Export and Import route-targets must be the same for any-to-any IP/VPN
- Export route-target values are attached to VPN routes in PE->PE

MP-iBGP advertisements

```
!
ip vrf green
  route-target import 3:3
  route-target export 3:3
  route-target export 10:3
!
```



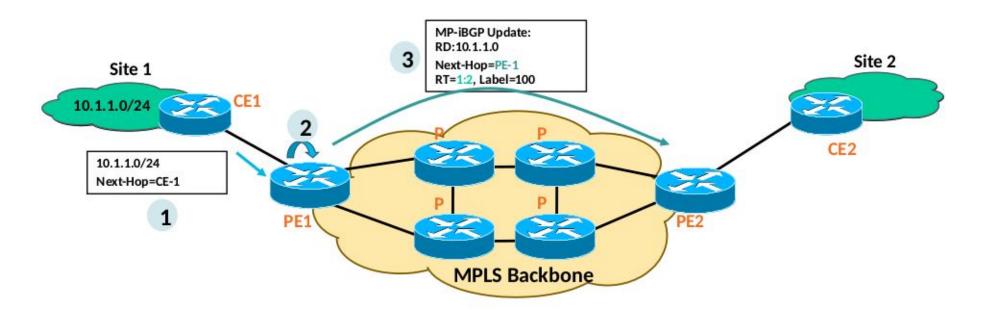
Label



- PE assigns a label for the VPNv4 prefix;
 - Next-hop-self towards MP-iBGP neighbors by default i.e. PE sets the NEXT-HOP attribute to its own address (loopback)
 - Label is not an attribute.
- PE addresses used as BGP next-hop must be uniquely known in IGP
 - Do not summarize the PE loopback addresses in the core

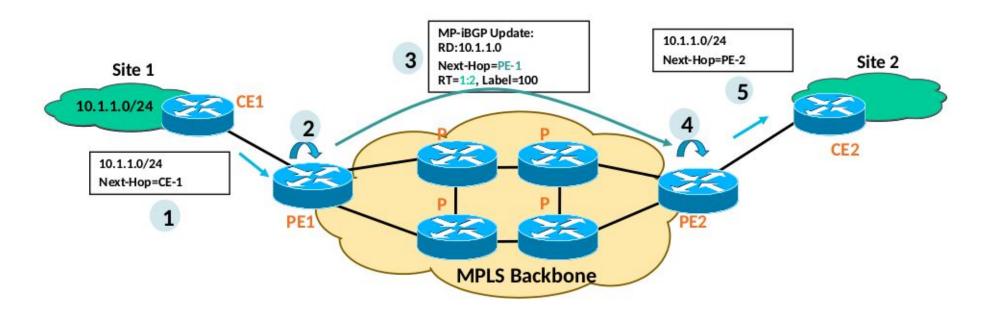
```
IOS_PE#
!
ip vrf green
  route-target import 3:3
  route-target export 3:3
  route-target export 10:3
!
```





- PE1 receives an IPv4 update (eBGP/OSPF/ISIS/RIP/EIGRP)
- PE1 translates it into VPNv4 address and constructs the MP-iBGP UPDATE message
 - Associates the RT values (export RT =1:2, say) per VRF configuration
 - Rewrites next-hop attribute to itself
 - Assigns a label (100, say); Installs it in the MPLS forwarding table.
- PE1 sends MP-iBGP update to other PE routers

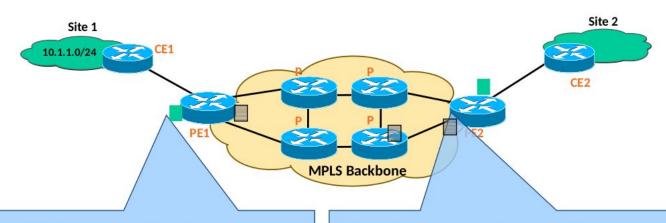




- PE2 receives and checks whether the RT=1:2 is locally configured as 'import RT' within any VRF, if yes, then
 - PE2 translates VPNv4 prefix back to IPv4 prefix
 - Updates the VRF CEF Table for 10.1.1.0/24 with label=100
- PE2 advertises this IPv4 prefix to CE2 (using whatever routing protocol)

MPLS VPN Forwarding Plane





Customer Specific Forwarding Table

- Stores VPN routes with associated labels
- VPN routes learned via BGP
- Labels learned via BGP

IOS:show ip cef vrf <name>
NX-OS: show forwarding vrf <name>
IOS-XR: show cef vrf <name> ipv4

Global Forwarding Table

- Stores next-hop i.e. PE routes with associated labels
- Next-hop i.e. PE routes learned through IGP
- Label learned through LDP or RSVP

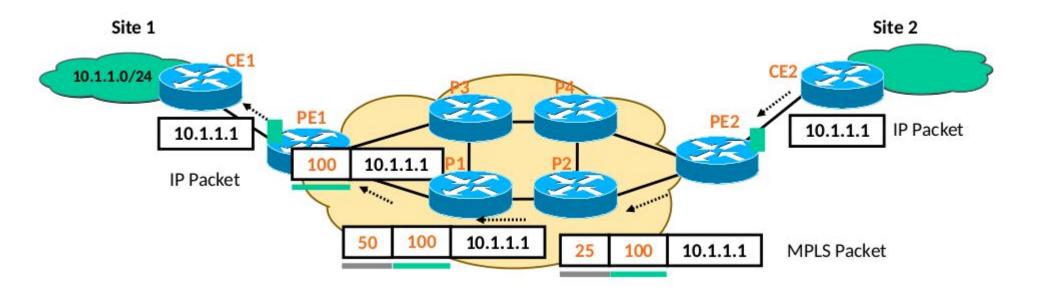
IOS: show ip cef

NX-OS: show forwarding ipv4

IOS-XR: show cef ipv4

MPLS VPN Forwarding Plane





- PE2 imposes two labels (MPLS headers) for each IP packet going to site2
 - Outer label is learned via LDP; Corresponds to PE1 address (e.g. IGP route)
 - Inner label is learned via BGP; corresponds to the VPN address (BGP route)
- P1 does the Penultimate Hop Popping (PHP)
- PE1 retrieves IP packet (from received MPLS packet) and forwards it to CE1.

MPLS VPNs

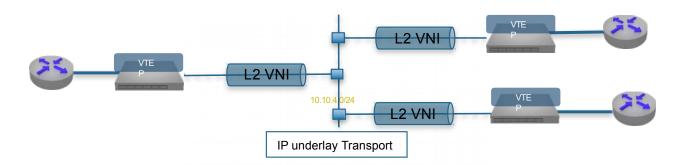


- Successful model, has been widely implemented
- Layer 2 BGP/MPLS VPNs:
 - Psewdowires
 - VPLS
- Datacenter world: VXLAN
 - MAC in IP encapsulation
 - Scaling problems → BGP to the rescue (?)



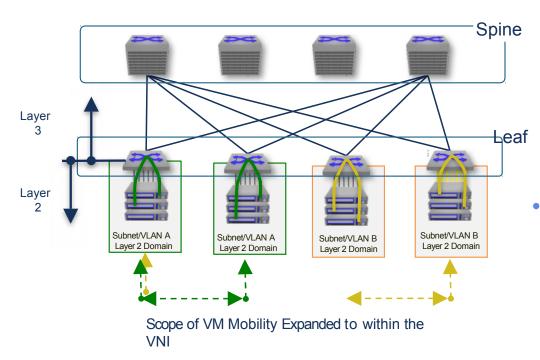
Introducing VXLAN

- Layer 2 "Overlay Networks" on top of a Layer 3 network
 - "MAC in IP" Encapsulation
 - Layer 2 multi-point tunneling over IP UDP
 - Transparent to the physical IP underlay network
 - Provides Layer 2 scale across the Layer 3 IP fabric





Data Center – Layer 3 Overlay Architectures

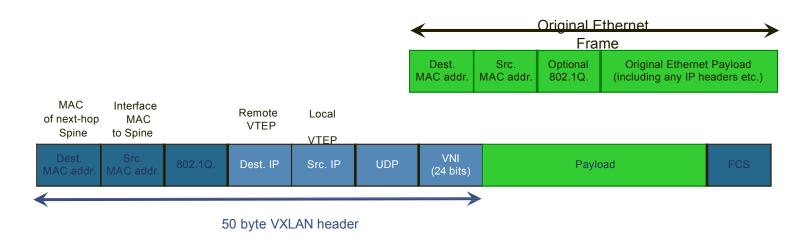


- Virtual eXtensible LAN (VXLAN)
 - IETF framework proposal, co-authored by:
 - >> Arista
 - >> VMware
 - >> Cisco
 - >> Citrix
 - >> Red Hat
 - >> Broadcom
- Enables Layer 2 interconnection across Layer 3 boundaries
 - Transparent to the physical IP network
 - Provides Layer 2 scale across the Layer 3 IP fabric
 - Abstracts the Virtual connectivity from the physical IP infrastructure
 - Enables Vmotion, etc. across IP fabrics



VXLAN Encapsulated Frame Format

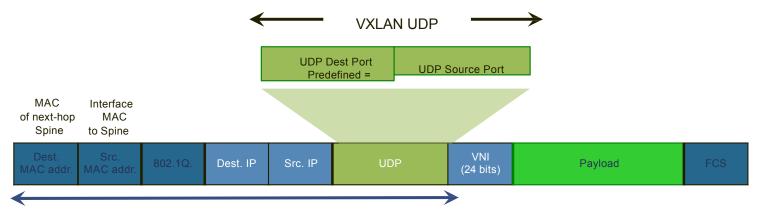
- Ethernet header uses local VTEP MAC and default router MAC (14 bytes plus 4 optional 802.1Q header)
- The VXLAN encapsulation source/destination IP addresses are those of local/remote VTI (20 bytes)
- UDP header, with SRC port hash of the inner Ethernets header, destination port IANA defined (8 bytes)
 - Allows for ECMP load-balancing across the network core which is VXLAN unaware.
- 24-bit VNI to scale up to 16 million for the Layer 2 domain/ vWires (8 bytes)





VXLAN Encapsulated Frame Format

- To provide Entropy across a multi-path ECMP underlay network
 - UDP source port created from a Hash of the inner frame
 - What fields are hashed from the inner is not defined in the standard
 - Silicon vendor, will define the level of Entropy that can be achieved
 - UDP destination port, predefined in the standard as 4789

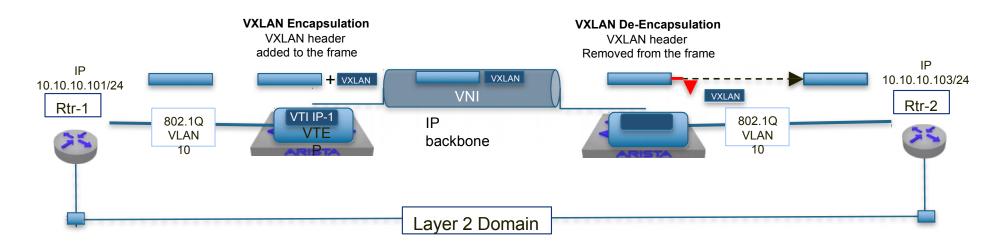


Source Port: It is recommended that the UDP source port number be calculated using a hash of fields from the inner packet - one example being a hash of the inner Ethernet frame's headers. When calculating the UDP source port number in this manner, it is RECOMMEND that the value be in the dynamic/private port range 49152-65535 [RFC6335].



VXLAN Terminology

- Virtual Tunnel End-point (VTEP).
 - Entry point for connecting nodes into the VXLAN overlay network.
 - Responsible for the encap/decap with the appropriate VXLAN header.
- Virtual Tunnel Identifier (VTI)
 - An IP interface used as the Source IP address for the encapsulated VXLAN traffic
 - IP address residing in the underlay network
- Virtual Network Identifier (VNI)
 - A 24-bit field added within the VXLAN header.
 - Identifies the Layer 2 segment of the encapsulated Ethernet frame

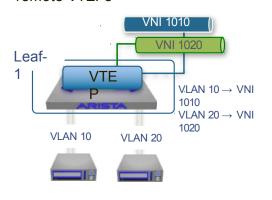




VXLAN Terminology - VLAN service interfaces

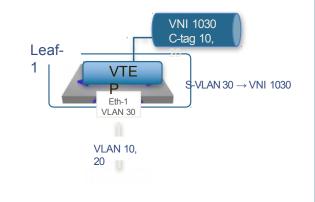
VLAN to VNI mapping

- One to One mapping between VLAN ID and the VNI
- Mapping is only locally significant,
- VLAN ID not carried on VXLAN encap frame
- Allows VLAN translation between remote VTEPs



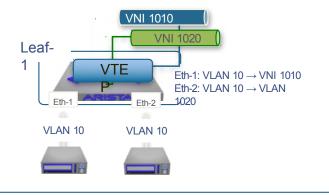
S-LAN to VNI mapping

- Mapping of the outer S-Tag to a single VNI
- Inner C-Tags are transported within a single VNI
- The inner VLAN ID are carried on VXLAN encap frame
- Ability to transport all customer VLANs across a single VXLAN point to point link



Port + VLAN to VNI mapping

- Mapping traffic to a VNI based on a combination of the ingress port and it VLAN- ID
- The VLAN ID is not carried on VXLAN encap frame
- Provides support for overlapping VLANs within a single VTEP to be mapped to different VNIs





VXLAN Control Plane Options

The VXLAN control plane is used for MAC learning and packet flooding

- Learning what remote VTEP a host resides behind
- Allowing the mapping of remote MACs to their associated remote VTEP
- Mechanism for forwarding of the Broadcast and multicast traffic within the Layer 2 segment (VNI)

Controller Model

- State learning driven by third-party controller
- OVSDB or OpenStack ML2 plugin for orchestration
- Data Center virtualization and Orchestration focus







IP Multicast Control Plane

- VTEP join an associated IP multicast group(s) for the VNI(s)
- Unknown unicasts forwarded to VTEPs in the VNIs via IP multicast
- Flood and learn and requires IP multicast support in the underlay
- Limited deployments

Head-End Replication (HER)

- BUM traffic replicated to each remote VTEPs in the VNIs
- Unicast Replication carried out on the ingress VTEP
- MAC learning still via flood and learn, but no requirement for IP multicast

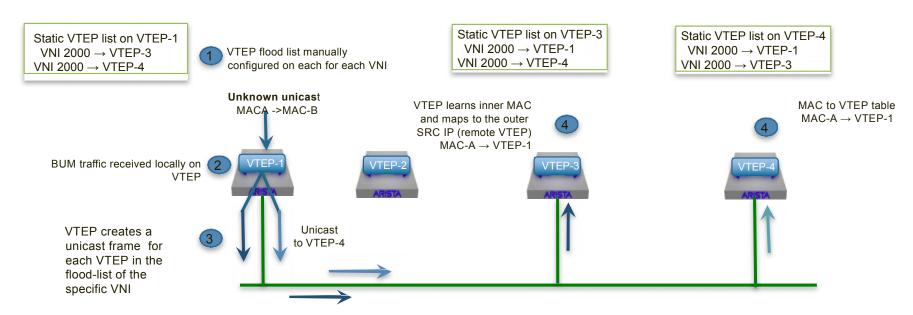
EVPN Model

- BGP used to distribute local MAC to IP bindings between VTEPs
- Broadcast traffic handled via IP multicast or HER models
- Dynamic MAC distribution and VNI learning, configuration can be BGP intensive



VXLAN Control plane – HER

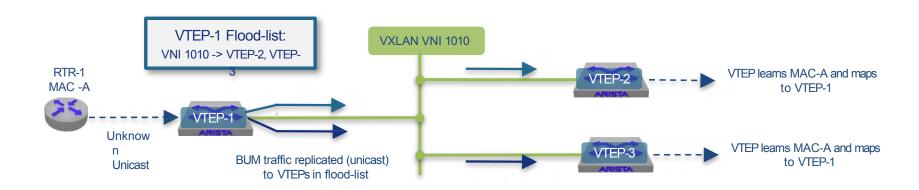
- Head-end Replication operation
 - Each VTEP is configured with an IP address "flood list" of the remote VTEPs within the VNI
 - Any Broadcast/Multicast or Unknown traffic is then replicated to the configured VTEPs in the list
 - Remote VTEPs receiving the flooded traffic learn inner source MAC from the received frame
 - VTEP's creating a remote MAC to outer SRC IP (VTEP) mapping for the entry





VXLAN Control Plane - HER

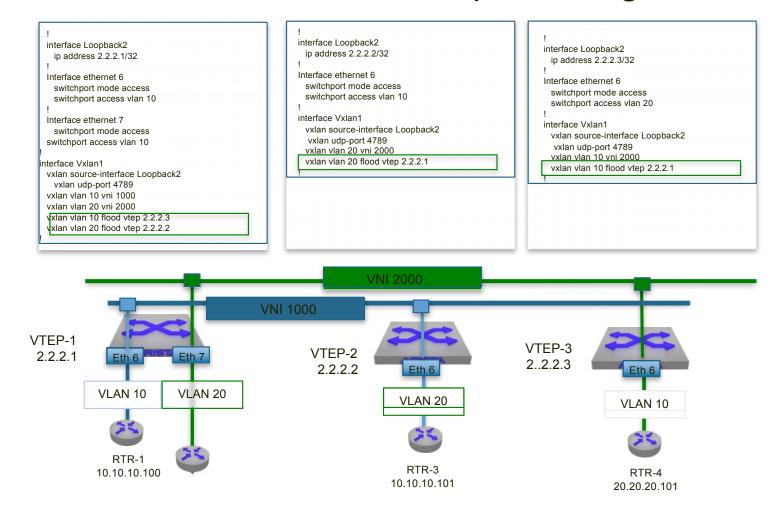
- Head-end Replication operation
 - Each VTEP is configured with an IP address "flood list" of the remote VTEPs within the VNI
 - Any Broadcast/Multicast or Unknown traffic is then replicated to the configured VTEPs in the list
 - Remote VTEPs receiving the flooded traffic learn inner source MAC from the received frame
 - Creating a remote MAC to outer SRC IP (VTEP) mapping for the entry



Flood list requires provisioning, MAC learning via flood and learn



VXLAN Control Plane – HER, simple config





What is Ethernet VPN (EVPN) - Standard body for EVPN

- EVPN Standard RFC 7432
 - Specifics an BGP EVPN control plane with a MPLS data plane
 - BGP control plane, new address family to advertise MAC/IP and IP prefixes.
 - Previously known as draft-ietf-l2vpn-evpn
 - Multi-vendor authors involving vendors and operators : ALU, Cisco, Juniper, AT&T, Bloomberg and Verizon
- Proposal for EVPN with Network Virtualisation Overlay (NVO)
 - Same EVPN control plane with a VXLAN Data plane (NGRE, MPLSoGRE)
 - Draft-ietf-bess-evpn-overlay



For the EVPN Data Plane, currently 1 standard (MPLS) and 2 proposals (NVO and PBB)

RFC 8365: A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)



Ethernet VPN

- EVPN, MP-BGP control-plane for delivering L2 and L3 VPN services with VXLAN
 - Evolution from the flood-learn mechanism of traditional L2 VPN (VPLS) service
 - Abstracts the (MP-BGP) control-plane from the (VXLAN/MPLS/PBB) forwarding plane
 - MP-BGP control plane to advertise host MAC and IP addresses and IP prefixes
 - Allows within a single MP-BGP control, L2 VPNs (hosts addresses) and L3 VPNs (IP prefixes).
- Potential use cases
 - Network virtualization (overlay) services for stretching Layer 2 connectivity
 - Integration of Layer 2 and Layer 3 VPN services in the overlay
 - Data Center Interconnect (DCI)
 - Internet Exchange Points (IXPs)



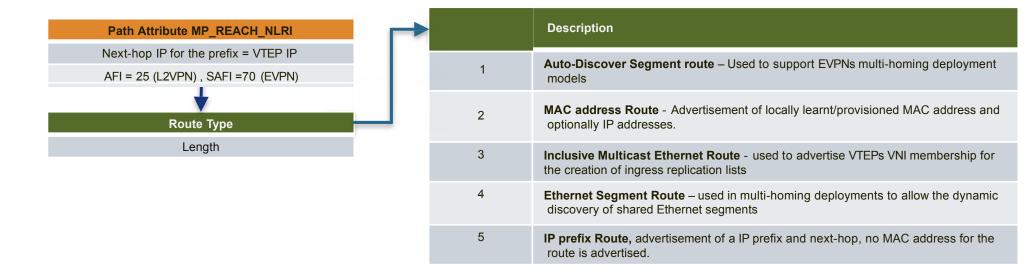
EVPN Operation

- EVPN is built on Multi Protocol BGP
 - Introduction of a new EVPN address family
 - >> Address Family Identifier 25 (Layer 2 VPN) subsequent AFI 70 (EVPN)
 - >> Advertisement of host MAC/IP binding and IP prefixes
 - Distribution of Layer 2/3 information allows support for integrated bridging and routing in VXLAN overlay networks.
 - Utilises Layer 3 VPN concepts of Route-distinguishers and Route Targets
 - >>> Providing support for multi-tenant VXLAN overlays
 - >> Support for over-lapping IP address spaces between tenants
 - Multiple tenant's NLRI information carried within a single shared BGP session,
 - >> NOT BGP session per tenant

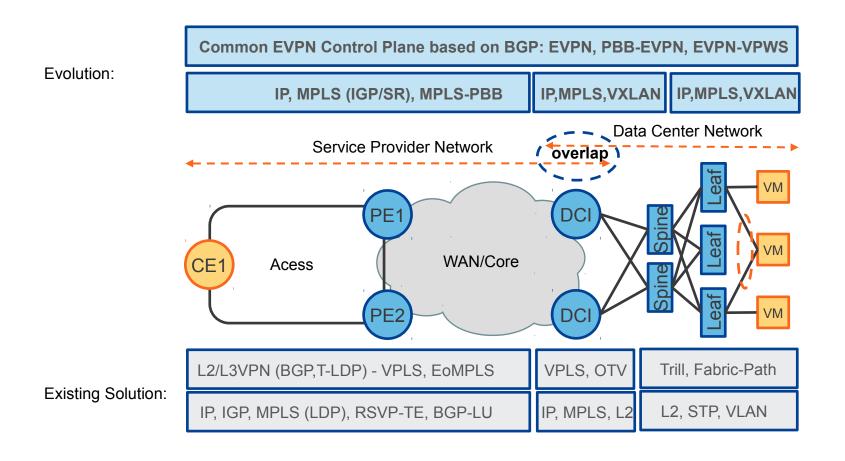


EVPN Operation – Route Types

- The new EVPN NLRI defines five route types
 - Not all route type are mandatory, specific support will be based on the vendors implementation
 - Next hop (VTEP IP address) for the route is contained in the MP_REACH_NLRI path attribute

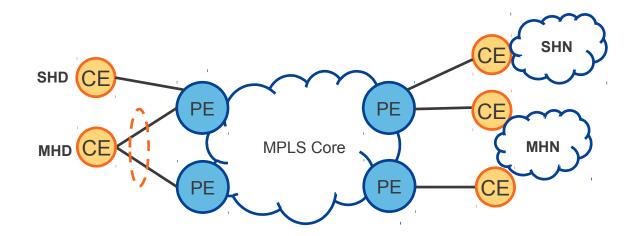


Summary: EVPN - End-to-End Control-Plane



Ethernet Segment





- Ethernet Segment is a 'site' connected to one or more PEs.
- Ethernet Segment could be a single device (i.e. CE) or an entire network.
 - Single-Homed Device (SHD)
 - Multi-Homed Device (MHD)
 - Single-Homed Network (SHN)
 - Multi-Homed Network (MHN)
- Uniquely identified by global Ethernet Segment Identifier (ESI).

Use cases

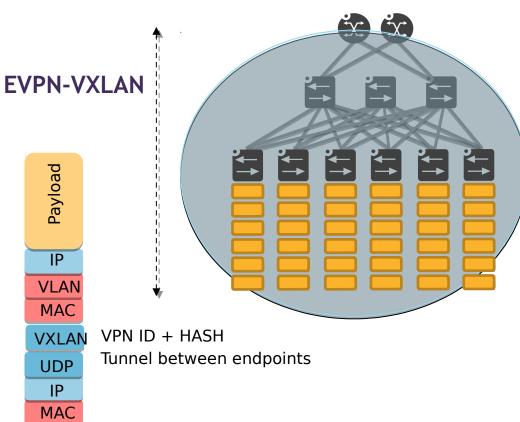


Data Center

ᄝ

Fabric





VXI AN

Cloud computing and NFV are shifting DC networks to SDN-based DCs where only VXLAN and EVPN provide the required capabilities

Legacy DC networks can't cope with 10,000s of dynamic hosts/VMs

Required EVPN features

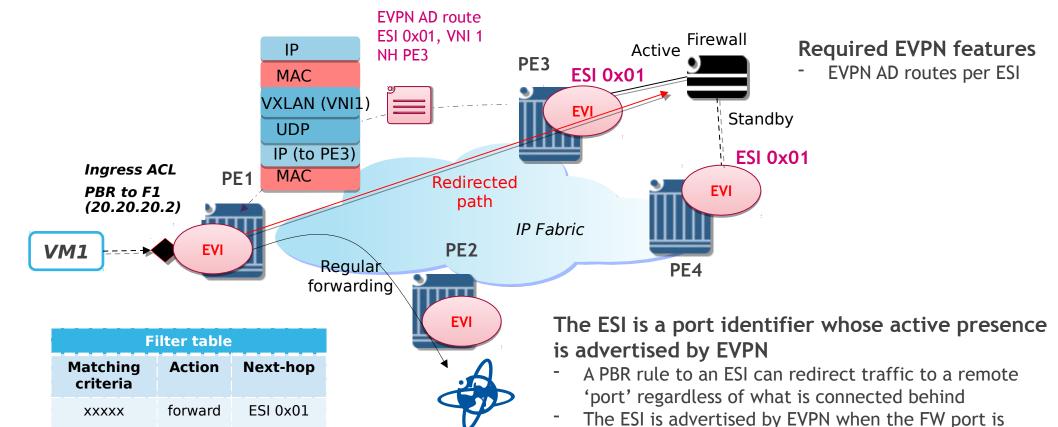
- EVPN provides L2/L3 connectivity for 1,000s of tenants in the DC
- The IP fabric can also be extended to the WAN for DC interconnect
- MAC mobility, proxy-ARP/ND, MAC protection, unknown flooding suppression, inter-subnet forwarding

VXLAN data plane provides the required scalability, performance and simplicity

- De-facto standard with assisted hardware in servers
- ECMP and fast resiliency
- Loop-free forwarding for L2
- Shortest path between any 2 endpoints

Policy Based Routing / service chaining (NFV)





active and withdrawn when the port goes inactive Active-active redirect is also possible (re-using the

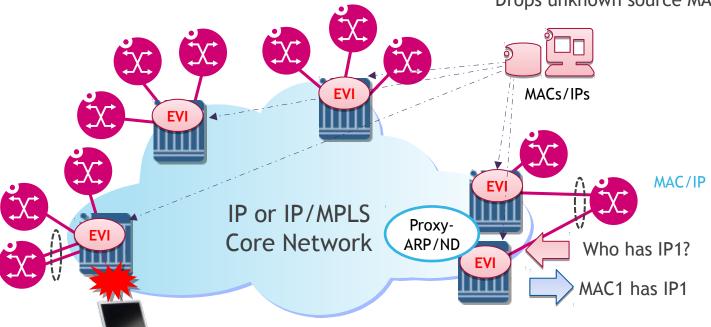
aliasing concept)





Static MAC/IP provisioning of the router interfaces for maximum security

- Suppresses unknown and ARP/ND flooding
- Drops unknown source MACs



ARP

SPOOFER

EVPN required features

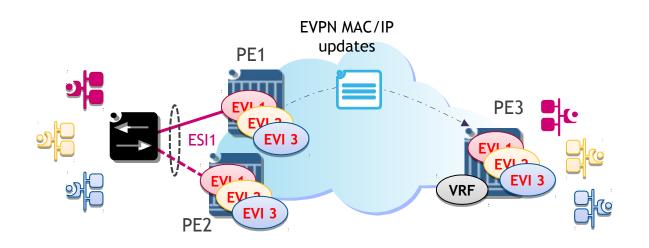
- L2 interconnection over a VXLAN or MPLS peering fabric
- Proxy-ARP/ND and unknown/ARP/ND suppression
- MAC duplication, MAC protection
- Anti-spoofing operation

Dynamic ARP/ND learning of proxy-ARP/ND entries for easy provisioning, minimum flooding and anti-spoofing monitoring

- Dynamic learning of ARP/ND entries is possible
- Anti-spoofing monitors hosts claiming the same IP
 - If a duplicate is detected, an alarm is triggered and MAC/IPs put in holddown mode
 - An option to inject an anti-spoof mac is possible too

Provider-provisioned VPNs Layer-2 and Layer-3 services





EVPN provides layer-2 and layer-3 services

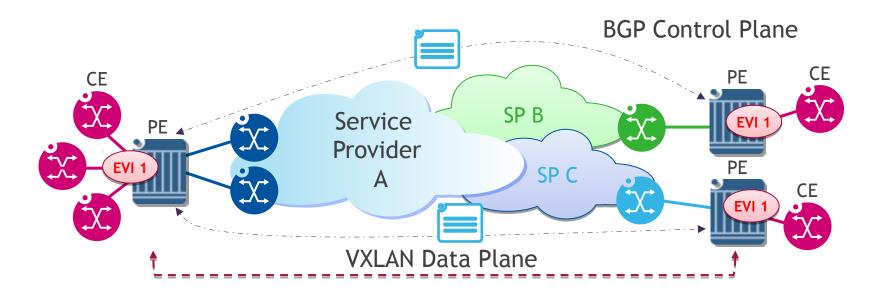
- Both services are provided through the same logical AC to the customer
- One VPN technology for both services, no need for multiple protocols
- VXLAN or MPLS data planes are possible

Required EVPN features

- IP-prefix advertisement and intersubnet forwarding
- All-active multihoming for link utilization
- Single-active multihoming for better determinism
- PBB-EVPN for large layer-2 VPNs

Enterprise-provisioned overlay VPNs





EVPN-VXLAN works over any IP service to provide a flexible Layer-2 and Layer-3 VPN

- Just requires IP connectivity between the sites, no need to run any MPLS or special configuration by the IP service provider
- Service Provider is transparent to EVPN
- EVPN overlay is transparent to service providers

VPN routing between endpoints can be controlled with BGP (ipv4) and routing policies to service providers

Routing and MAC/IP advertisements within EVPN controlled via iBGP (evpn) between PEs

References



[Roma3] Giuseppe Di Battista, Infrastrutture delle Reti di Calcolatori: Reti locali virtuali. Online: http://www.dia.uniroma3.it/~impianti/HomePage18-19/index_irc.html

[RFC4364] RFC4364 BGP/MPLS IP Virtual Private Networks (VPNs). E. Rosen, Y. Rekhter. February 2006. (Obsoletes RFC2547) (Updated by RFC4577, RFC4684, RFC5462). Status: PROPOSED STANDARD, DOI: 10.17487/RFC4364

[TREX2017] Ralf Korschner, VXLAN/EVPN in a Nuttshell. Online: http://www.trex.fi/2017/Ralf-Korschner-VXLAN-EVPN-in-a-Nuttshell.pdf

[BR2014] Ethernet VPN (EVPN) - Casos de Uso e Aplicação. Alexandre Silvestre, November 2014. Online: http://ix.br/pttforum/8/doc/06-EVPN_Use_Case_Reviewv1.pptx

[RFC7348] RFC7348 Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks. M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, C. Wright. August 2014. Status: INFORMATIONAL, DOI: 10.17487/RFC7348

[RFC7432] RFC7432 BGP MPLS-Based Ethernet VPN. A. Sajassi, Ed., R. Aggarwal, N. Bitar, A. Isaac, J. Uttaro, J. Drake, W. Henderickx. February 2015. Status: PROPOSED STANDARD, DOI: 10.17487/RFC7432

[RFC8365] RFC8365 A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN). A. Sajassi, Ed., J. Drake, Ed., N. Bitar, R. Shekhar, J. Uttaro, W. Henderickx. March 2018. Status: PROPOSED STANDARD, DOI: 10.17487/RFC8365