ARISTA

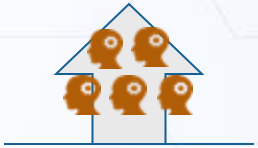
# Comfortable Complexity of Overlays

**Claudiu Captari**

Systems Engineering Manager, ANZ, Arista Networks

# Routing – Complexity is driving change…

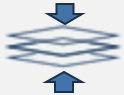
Multiple network domains
Different snowflakes



Bespoke skill sets to operate and troubleshoot each design


Spend increases to cover the multi domain
Expensive technology transitions across generations
Deploy and operate complexity

## A network transformation is required…



### Simplify

- Simplify the network Architecture
- Protocol Reduction
- Streamline HW and OS platforms

"Keep it **simple** and **consistent with a balanced approach to resiliency** " – to enable automation & performance at scale

### Iteration

- Repeatable design models
- Consistent skillset and operations
- Streamline HW platforms

*Power of **iteration** -Start somewhere and iterate to the next design as opposed to waiting for perfection – **evolution not revolution***

### Automation

- Common end-to-end APIs
- Service automation and orchestration
- Fine-grained telemetry

*Service velocity via **closed loop automation, orchestration** & fine-grain telemetry*

# What should be your strategy?

**Single Service Plane**

EVPN end-to-end from the Campus to the Data Center, across the WAN or securely across internet via Secure VPNs

Single consistent end-to-end operational model for Service delivery **– EVPN**

- Open, standards, No vendor lock-in to prevent end-to-end connectivity
- Convergence of skill-set across domains (WAN, Campus, DC) – Reduction, in OpEx cost
- Single operational model to troubleshoot, model and automate

**Any Transport**

Support for the appropriate, cost-effective transport. Campus, Branch, DC, WAN

Connectivity agnostic, allowing appropriate transport layer for the use case

- Cost sensitive high bandwidth - Campus and DC – **VXLAN**
- Security sensitive WAN – **VXLAN over IPSec**
- Bandwidth, connectivity restrictive WAN – **SR, SR-TE** or **SRv6**

**Futureproof**

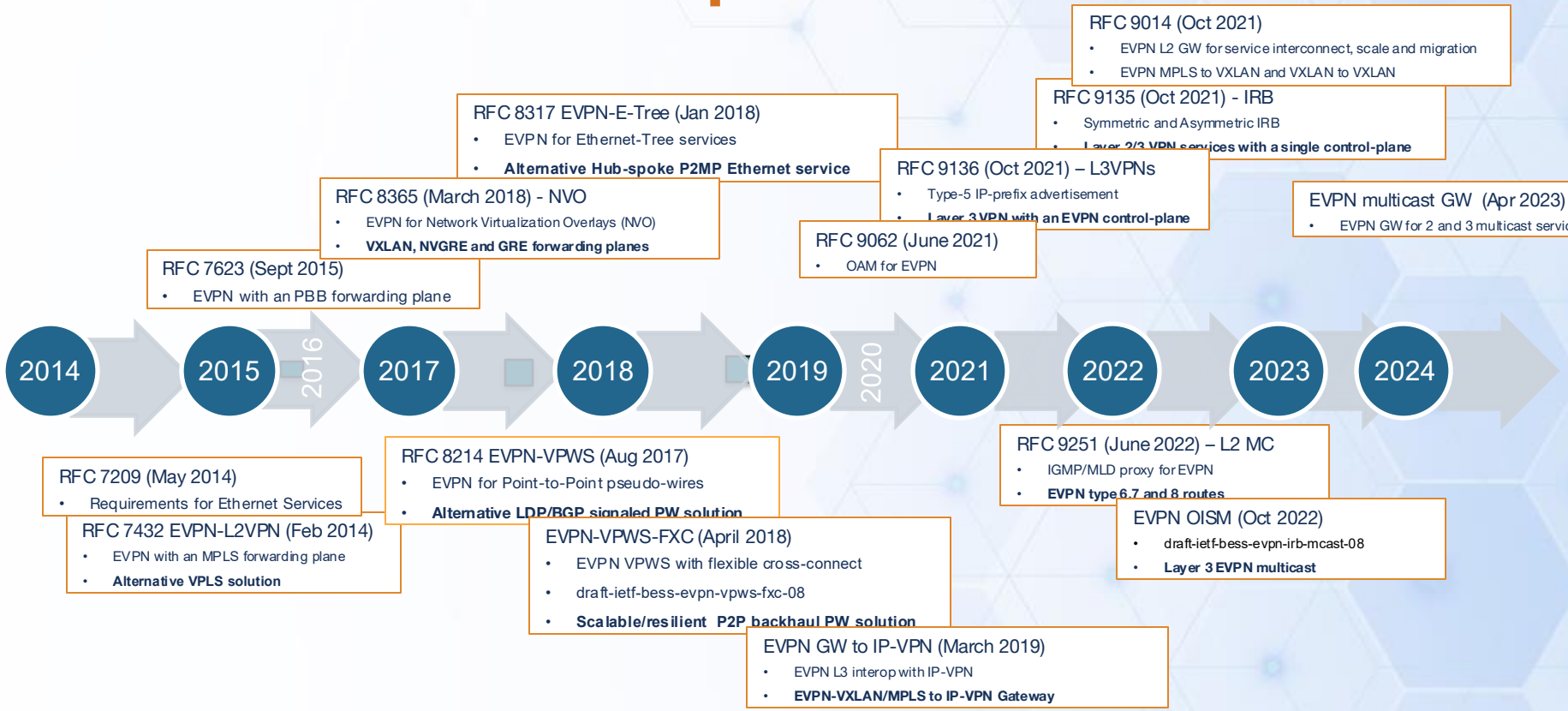
Zero technology lock-in provides incremental upgrades to next-Gen solutions i,e, SRv6

Enables, incremental network upgrade without a lock-in

- Ability to change transport and service layer at the customer's pace
- No need for a big bang approach
- No vendor lock-in to prevent end-to-end connectivity

# eVPN RFC Landscape



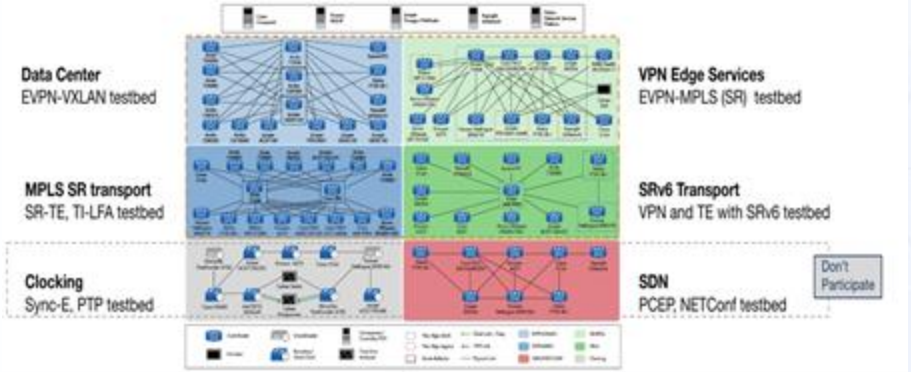

48 current active drafts

ARISTA

# European Advanced Networking Test Center (EANTC)

- **EANTC host an annual interop event over two weeks (Feb/March)**
  - Opportunity to do vendor interop testing of new and evolving IETF standards
  - Focus on next-generation standards for the DC and SPs (EVPN, SR, SRv6, SDN)
  - Currently only event of its type globally, so high level of interest across vendors
  - Completed tests validated by the EANTC, results published in their annual whitepaper
- **The test cases are broken across six separate testbeds:**
  - EVPN-VXLAN, EVPN-SR, MPLS-SR, SRv6, Timing/Clocking and SDN (PCEP)

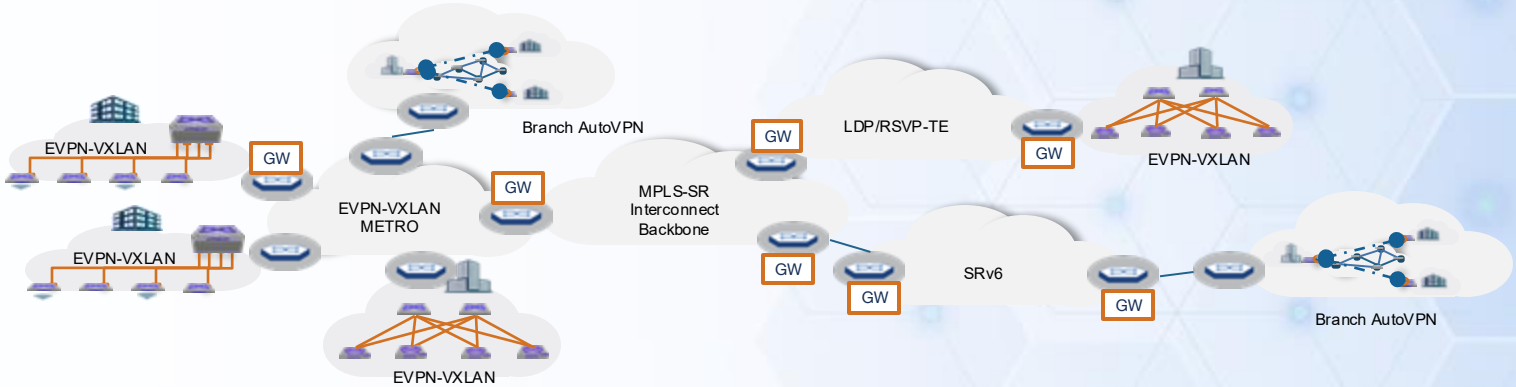

Testbeds

Participating vendors across the Testbeds

# A multi-domain blueprint

- What is required is a uniform service layer that allows end to end stitching across these different transport protocols. Meaning freedom of choice to pick the right transport for the right use case.
- EVPN gives us a uniform service layer that can exist in any domain with the ability to stitch services end to end, regardless of the underlying transport (IPv4/IPv6, MPLS, SR,VXLAN or even SRv6)
- EVPN GW for L2 and L3 services, provides the ability to seamlessly inter-connect domains regardless of forwarding plane : VXLAN, SR, MPLS, or SRv6
  - Allowing the appropriate transport for the use case
  - With support for both a IPv4 and IPv6 underlay



# Why EVPN…

- Why EVPN for service delivery…

| | |
|---|---|
| **Simplification** | **Protocol Reduction** – Single BGP AF for all Ethernet and IP VPN services removing the need for dedicated protocols per services, PW, VPLS, MVPN, IP-VPN |
| **Consistency** | **Repeatable model** – Consistent implementation and operational model for deploying any service type E-LINE, E-LAN, IP unicast/multicast enabling road to automation |
| **Resiliency** | **Flexible multi-homing** – Standard based consistent multi-homing procedures across all Ethernet and IP VPN services |
| **Flexible** | **Any encapsulation** – Support for multiple encapsulation models VXLAN/MPLS/SR and SRv6, allowing a consistent operation model from the WAN to the DC and Campus |
| **Skillset** | **Converged Teams** - One operational and implementation team for managing VPN services across both the WAN, DC, Campus infrastructure |


CapEx

The simplicity of EVPN removes the need for dedicated HW platforms for VPN services, service can now be achieved on a standard switch or router


OpEx

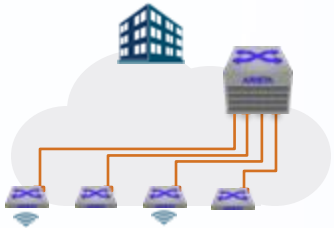
A consistent operational models for MPLS/SR and VXLAN means OpEX saving trough automation and the merging of teams

# Today's Transport Protocols and Domains..

### Enterprise Campus

Layer 3 design for scale and fault containment, overlay network for network/user segmentation

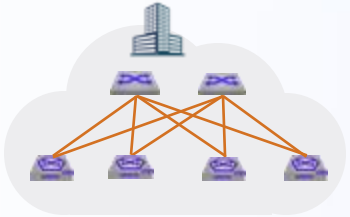


Transition to Layer 3 designs for scale/fault containment driving the adoption of EVPN → EVPN

Cost sensitive and bandwidth/latency not a major concern, driving a VXLAN transport → VXLAN

### Data Center

Larger scale design, requiring Layer 2 and 3 VPNs for both unicast and multicast services

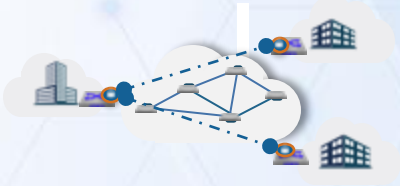


EVPN for L2 and L3 VPNs or BU and application separation → EVPN

No so cost sensitive but 100G/400G Leaf-Spine ECMP topologies, mean BW is freely available with fast failover - driving a VXLAN transport → VXLAN

### Branch

Large scale geographically disperse topology design, requiring secure VPNs services across all available branch connections
(IPsec, VXLANsec)

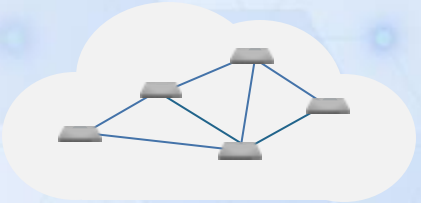


EVPN for site-to-site VPNs with Secure-EVPN for scalable P2MP key-exchange → EVPN, Secure-EVPN

DPS for agnostic 3rd-party forwarding, with IPsec encryption for data security → IPSec, DPS

### WAN/Backbone

MPLS , SR, SRv6 transport with Ethernet and IP. VPN services for site-to-site and region to region connectivity



Traditional MVPN and IPVPN solve to evolve. Major benefits of EVPN for Ethernet VPNs → IP-VPN, MVPN, L2 EVPN

Evolving trend to simplifies SR transport and potential future SRv6 transport → MPLS, MPLS-SR, SRv6

# Stitching multiple transport domains

## eVPN Gateway

# EVPN GW – Stitch domains and improve scale

### Standard based solution

- RFC 9014 & evpn-ipvpn-interworking
- Multiple encap support (VXLAN/MPLS/SR)
- Standards based Multi-homing support

### EVPN GW for Hierarchical scaling

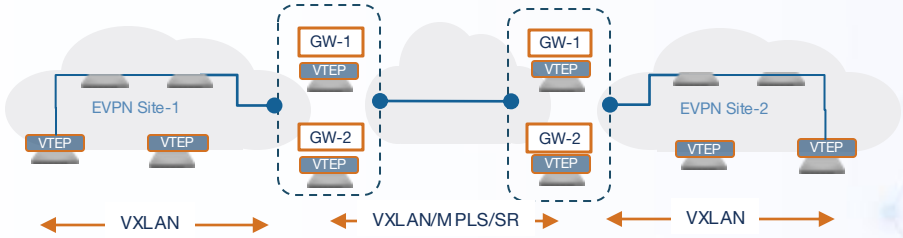
EVPN GW for scaling EVPN-VXLAN deployments inter-POD and intra-site (DCI) by introducing hierarchy

### Layer 3 interconnect

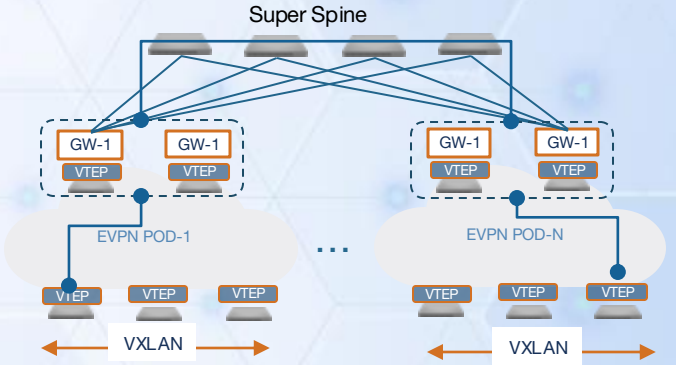
- Layer 3 (type-5) interconnect between domains
- Type-5 routes re-advertised with GW next-hop

### Scalable L2 interconnect

- GW scoping of Type 1,4 and 3 routes
- Flood-list scale with split-horizon forwarding of BUM traffic on GW
- Type-2 re-originated with GW next-hop



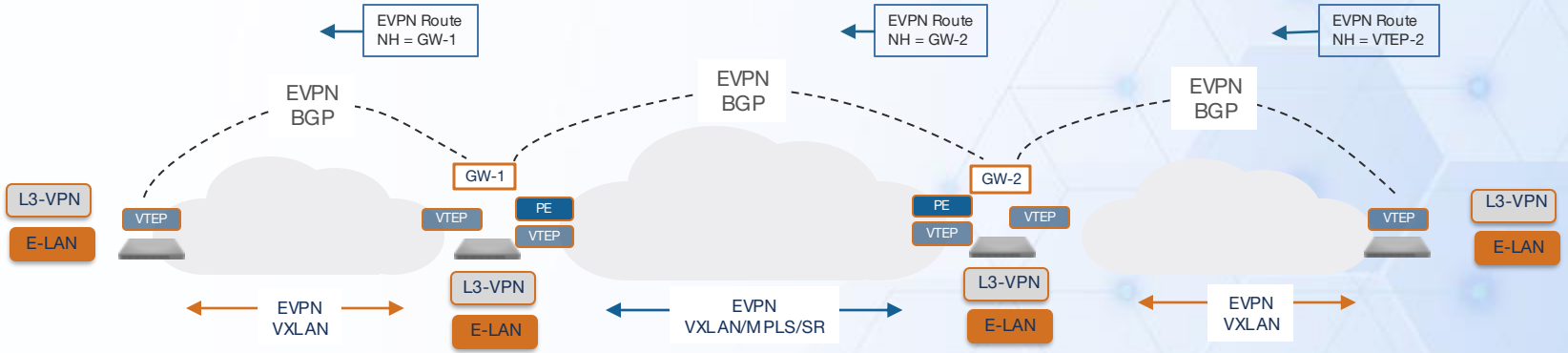

Interconnecting Domains




Interconnecting intradomain PODs for hierarchical scaling

# EVPN GW – EVPN VXLAN/MPLS GW

- EVPN Gateway Solution




## EVPN GW behavior

- PE/VTEP nodes EVPN peer with their local GW node via eBGP or iBGP
- GW node EVPN peer with the GW nodes in the remote domain via eBGP or iBGP
- Import received type-2 & 5 routes based on RT policy
- Export type-2 & 5 routes between domains based on RT policy
- When exporting between domains, new Next-hop, encap and label

## Benefits

- End-to-End Layer 2 and 3 connectivity regardless of interdomain encap
- Support L2 and L3 VPN between VXLAN VTEPs and MPLS PE nodes
- EVPN A-A for GW redundancy for L2 interconnect across domains
- Hierarchical flood-list for BUM traffic forwarding
- Reduction in EVPN state churn across domains

# Routing in the Data Center & Campus

## EVPN-VXLAN services

# Data Center and Campus needs…

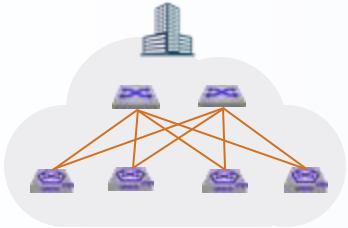
- All modern data centers and many campus designs require an overlay. EVPN-Vxlan gives the ability to stretch L2 / L3 services via an overlay in a mature scalable method.
- All modern hardware now supports Vxlan giving a wide choice of options and cost efficiency in the DC / Campus domain.
- The availability of cheap additional bandwidth and the wide use of CLOS type architectures alleviates the need for fast reroute technologies or traffic engineering.
- VXLAN operates with IPv6 underlay today if address exhaustion is concern

EVPN-VXLAN is correct solution for the Data Center and the Campus

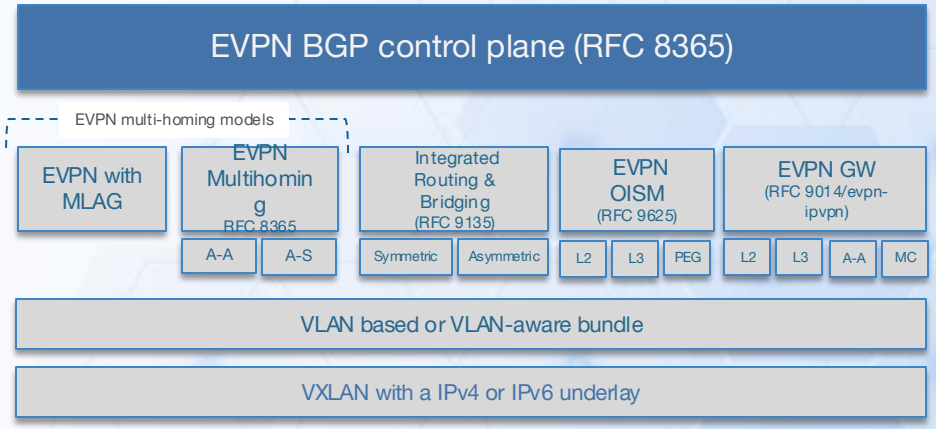
# Routing – EVPN Data Center & Campus

**Data Center & Campus**
Larger scale design, requiring Layer 2 and 3 VPNs for both unicast and multicast services



EVPN for L2 and L3 VPNs or BU and application separation → EVPN

Not so cost sensitive. 100G/400G Leaf-Spine ECMP topologies, mean BW is freely available with fast failover - driving a VXLAN transport → VXLAN

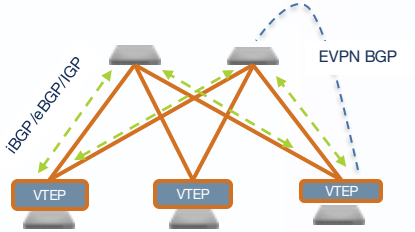



 | 

# Routing – EVPN Data Center & Campus

## Standard based IP fabric

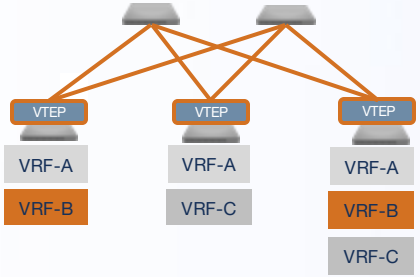
Flexible leaf-spine deployment models for all possible underlay and overlay routing protocol option




- eBGP over eBGP
- eBGP over iBGP supported
- iBGP over eBGP supported
- iBGP over IGP supported

## Network Segmentation

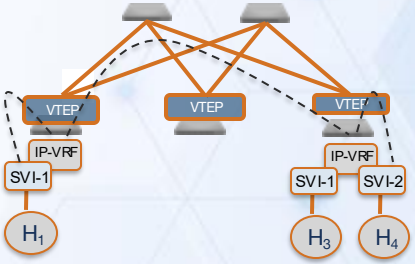
Support for both VLAN-based and VLAN aware service interfaces with VRF route leaking and RT-route constraint.




- VLAN based and aware-bundle
- EVPN RT import/export model
- VRF route leaking

## Any Endpoint Anywhere

Optimal first-hop routing with either symmetric IRB or asymmetric IRB models and support for resilient centralized IRB




- Symmetrical IRB routing
- Asymmetric IRB routing
- Centralized IRB routing

## Multi-Homing connectivity

Support for both standard based All-Active EVPN multi-homing and MLAG with EVPN across Campus and DC



- EVPN All-Active multi-homing
- EVPN Single-Active multi-homing
- Multi-chassis LAG (MC-LAG)

All design options with Consistent implementation

 | 

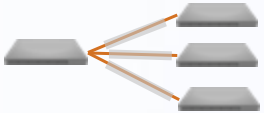
# Routing – EVPN Data Center & Campus

Evolving Trends

### Tenant Multicast Services

Requirement for VRF-aware multicast service for high BW, latency sensitive applications

EVPN-OISM



PEG | PEG-transit | PIM | IR

A-A | IPv4 | IPv6 | OISM-GW

### Collapsing the MPLS PE

Evolution of the Border leaf to MPLS PE for seamless layer 2 & 3 connectivity across MPLS and IP core.

Integrated Service GW


### Hierarchical POD designs

Improved EVPN scaling & fault-tolerance while maintaining on-demand Layer 2 and 3 services across domains

Hierarchal EVPN

A-A | MLAG | D-path | UMR

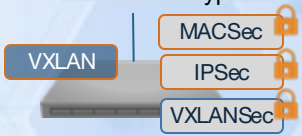
Route scaling | VXLAN | MPLS

### High Speed Encryption

In-flight encryption for both P2P and P2MP connections within the DC and across PODs and the WAN

TunnelSec

Line-rate VPN encryption




Differentiation – Marketing standards based consistent implementation of OISM and GW functionality

ARISTA

# Routing outside the DC

## eVPN services, MPLS-SR or SRv6 transport

# WAN Core needs

- Many WAN cores consist of partial mesh or ring topologies facilitating the need for fast reroute technologies and traffic engineering.
- MPLS and the various transport control planes available (LDP,SR,RSVP) give the options of fast reroute and traffic control that the WAN needs.
- SRv6 (Segment Routing over IPv6) evolves this concept by using the native IPv6 data plane, eliminating the need for MPLS. This allows the service functions to be encoded directly into the packet's IPv6 header.

EVPN over SR-MPLSv4 , SR-MPLSv6 or SRv6 in the WAN

# Why Segment Routing for transport …

| | |
|---|---|
| **Simplification** | **Protocol Reduction** - IGP used for segment distribution and TE, removes the need for additional dedicated protocols, LDP and RSVP-TE |
| **Scale** | **Improved scale** – Source based routing removing the need for state to be stored or signalled end-to-end, thus reducing hardware table resource utilisation |
| **Convergence** | **Fast Reroute** – FRR capabilities regardless of the topology with TI-LFA, delivering 50ms failover in the event of a link or node failure |
| **Traffic Engineering** | **Flexible TE** – Programmable stateless source-based TE forwarding for improved scale with Controller and Controller-less solutions (FlexAlgo and dynamic SR-TE) |
| **Skillset** | **Incremental Skillset** – MPLS-SR builds on the constructs of MPLS, which means existing skills are easily transferable at minimum cost and disruption<br>SRv6 leans heavily into IPv6, programmability, and automation |


CapEx

SR reduces CapEX spend with the ability to reuse existing MPLS hardware along with a reduction in resource requirements for new HW


OpEx

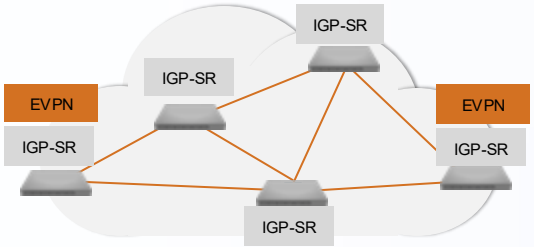
Protocol reduction with a incremental technology reduces OpEx cost by simplifying operations, skillsets and training costs

# Simplification with SR

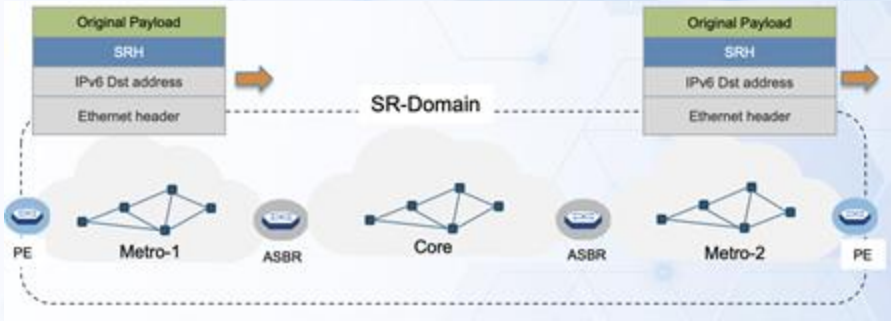
## The Segment Routing Choices

### MPLS-SRv4 or MPLS-SRv6




- SR SID programmed as an MPLS label
- Reuse existing MPLS IPv4 or IPv6 forwarding plane
- Reuse existing capable hardware
- Re-use existing BGP (label) control-plane
- Simplified Migration path no forklift upgrade
- **Pros: Incremental learning curve**

### SRv6




- No need for MPLS
- uses native IPv6 forwarding plane to convey routing instructions
- SR uSID programmed in the IPv6 Dest Address - SRH
- Likely requires a hardware platform upgrade
- New BGP control-plane to support uSID
- **Pros: Support IPv6 underlay, potential for service chaining and scale**

# Simplification with SR

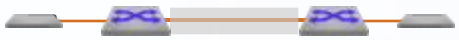
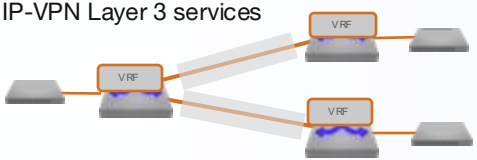
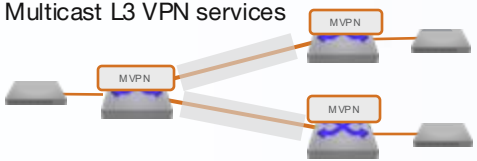
- The new use cases with SRv6

| Traffic-Engineering for AI | 5G and Network Slicing | Service Chaining |
|---|---|---|
| Workload steering for optimal BW utilization between GPUs of the AI fabric | New 5G rollout using an IPv6 underlay and SRv6 transport for scaling | New revenue streams by dynamically inserting service in-path |
| TE path calculated and SRv6 encapsulation done on the Server node | Multiple AS topology with IPv6 route summarization and UPA | Typically, Telco Cloud or Data Center deployment |
| Pure IPv4/IPv6 is encapsulated no need for VPNs | VPN traffic with requirement to support both Ethernet and IP services | VPN traffic with requirement to support both Ethernet and IP services |
| Leaf, Spine and Super-Spine nodes responsible for just Next-C-SID operation | Traffic Engineering for network slicing, External controller or FlexAlgo possible options | Traffic Engineering required with FlexAlgo |
| | | SRv6 proxy required if the Service node is not SRv6 aware. |

# eVPN services over SR

**Simplify, single BGP AF and EVPN control plane for all layer 2 and 3 services types**

| | Traditional MPLS Services | EVPN MPLS services |
|---|---|---|
| E-LINE (Point-to-Point) services  | **LDP/BGP signaled PW**<br>• LDP control-plane for signaling<br>• Type-4, Type-5 modes, Standby PW support<br>• PW coloring for traffic engineering | **EVPN VPWS (RFC 8214)**<br>• Consistent BGP EVPN Control-plane<br>• Auto-discovery via BGP EVPN (Type-1)<br>• Support for standards based A-A/A-S models |
| E-LAN (Point-to-multipoint) services  | **VPLS (RFC 7432)**<br>• LDP signaled VPLS support<br>• LDP signaled with BGP A-D<br>• Hierarchical VPLS, with standby PWs | **EVPN L2 VPNs (RFC 7432)**<br>• Consistent BGP EVPN Control-plane<br>• Control-plane learning with Type-2 routes<br>• Support for standards based A-A/A-S models |
| IP-VPN Layer 3 services   | **IP-VPN (RFC 4364)**<br>• BGP signaling and auto-discovery<br>• Control-plane learning via BGP<br>• Layer 3 only solution no l2 VPN support | **EVPN L3 VPNs (RFC 7432)**<br>• Consistent BGP EVPN Control-plane<br>• Control-plane learning with Type-5 routes<br>• Seamless integration with L2VPN models |
| Multicast L3 VPN services   | **NG MVPN (RFC 6513)**<br>• BGP signaling and auto-discovery RFC 6514<br>• P2MP transport with mLDP<br>• Default MDT phase 1 | **EVPN-Multicast (MPLS-OISM)**<br>• Consistent BGP eVPN control Plane<br>• LDP, RSVP-TE transport for P2MP LSPs |

# In summary …

Market Problem

CapEx

eVPN overlay reduces CapEX spend with the ability to unify multiple architectural domains and reuse software and hardware in a unified architecture

OpEx

Protocol reduction reduces OpEx cost by simplifying operations, skillsets and training costs

Solution

**Repeatable network architectures across domains**

*Power of **iteration** - scalable scale-out deployments*

**Protocol and state simplification**

"Keep it **simple** and **consistent with a unified protocol stack** " – to enable automation & performance at scale

**Network wide software API**

*Service velocity via **closed loop automation, orchestration** & fine grained telemetry*

ARISTA

# Thank You

www.arista.com