NOKIA

EQUINIX

Segment Routing for DCI

Pete Moyer, Principal Consulting Engineer - Nokia
Muhammad Durrani, Chief Network Architect - Equinix

Agenda

Pete

- Refresher of Segment Routing
- Why SR for DCI?
- Why EVPN in DC?

Muhammad

- Unified Packet Fabric (UPF) architecture & services
- EVPN over SR for L2 & L3 use cases

Background of Segment Routing

- The SPRING working group will define procedures that will allow a node to *steer* a packet along an *explicit route* using information attached to the packet and *without the need for per-path state* information to be held at transit nodes.
- The initial data planes that will be considered are MPLS and IPv6.

  https://datatracker.ietf.org/wg/spring/about/

- Leverages source routing paradigm
- Label stack (aka: SID-list) populated at ingress headend node; transit nodes pop

Segment Routing for DCI Use Case

- While DCI (Data Center Interconnect) could be simple p2p circuits, these use cases are based on diverse paths being available amongst several or more DC locations.
- The services and applications running over this diverse topology in these use cases require traffic-engineering capabilities.
- An additional requirement for these use cases is to avoid complex signaling protocols, such as RSVP-TE. Drives the solution towards segment routing.
  - SR-TE

Specifics of Segment Routing for DCI

- Well known traffic-engineering options are available in SR
- IGP floods link-state plus SR parameters; TED is built
- SR-TE paths calculated based on link characteristics and other parameters
  - SRLG, node-SID, adjacency-SID, BW, delay, etc
  - Link-coloring like capability is available for path diversity
- SR-TE path is determined, label stack is created & populated
- FRR capability is available
  - LFA provided by IGP
  - 50ms
- What about the control plane?
  - Controllers are good …

SR data plane for DCI

- MPLS has been deployed for decades
  - Operators fully understand MPLS; similar experience with IGPs and BGP
  - Must ask oneself “why not MPLS?” 🙂
- What about IPv6 (SRv6)?
  - Capable, interoperable, functional, etc
  - See first bullet …

What about the DC overlay?

- EVPN provides both L2 and L3 services in a unified control and forwarding plane.
  - EVPN can be used for p2p and/or p2mp services
- EVPN leverages a widely deployed and well known protocol; BGP.
  - This makes EVPN highly scalable
  - After all, doesn't BGP "fix everything?" 🙂
- EVPN route types for these use cases.
  - Type 1 Ethernet auto-discovery for fast convergence and multi-homing of CE devices
  - Type 2 MAC/IP binding for control plane learning of MAC addresses
  - Type 4 Ethernet segment identifier for multi-homing and DF election
  - Type 5 IP Prefix advertisement
- VxLAN data plane
  - While maturing, has not *yet* reached feature parity with MPLS
  - While supported by network vendors, has not *yet* displayed *wide scale* interoperability

Why EVPN w/MPLS in DC?

- See previous slide …
- But what about ToR/Leaf support of MPLS?
- This use case leverages a “smart edge” initiative
  - L3 @ leaf
  - Leaf network elements are capable devices (aka: routers)
- This use case not reliant on NFV/VNFs
  - Could still deploy VNFs
  - Could leverage VxLAN from VNF to ToR/leaf; *stitch* to EVPN-MPLS for DCI
- Result: EVPNoSR-TE/MPLS provides *right* foundation for these use cases

Why Unified Packet Fabric ?

- Reduce OPEX and CAPEX cost
- Architecture Optimization
- Product and Services consolidation over common Infrastructure
- Simplify day to day operations
- Simplify control and data plane across products
- Reduce service delivery time
- Simplify network monitoring tools
- Network Infrastructure readiness for 5G

Control & Data Plane – Why EVPN / SR?

Why EVPN ?	Why Segment Routing ?
All-in-one VPN technology - Unified control plane supports multiple data plane encapsulations (VXLAN and MPLS)	**Lower CAPEX :** • Better utilization of network bandwidth due to packet based path optimization using ECMP unlike circuit based path optimization with RSVP-TE by making optimal use of available BW • Eliminate need to use end to end signaling protocol maintain flow state in the network relaxes memory / CPU requirements on Edge/core devices gives flexibility to use cheaper white label or merchant silicon in the core
Provides wide range of services such as E-LAN, E-Line, E-TREE, L3VPN, DCI, DC-overlay, IRB etc without running separate control plane protocols	**Lower OPEX :** • Due to technology simplicity troubleshooting is comparatively simpler then traditional RSVP-TE deployments
	Advance Value added service offering: • SR supports Class of Service-based TE (CoS) where one can define per-flow CoS policies and encode a segment to fulfill the CoS demands. RSVP-TE has failed to provide this level of granular control due to scalability issues • Path Computation Element (PCE) support enables an agile WAN-SDN use cases . SR with SDN Platform can be used to provision TE tunnels automatically and provide value-added services such as bandwidth management, bandwidth calendaring, and bandwidth on-demand. • 5G roll-out will drive significant investment in the network infrastructure to support new requirements such as network slicing – specific slices include encrypted, low latency and high bandwidth slices.

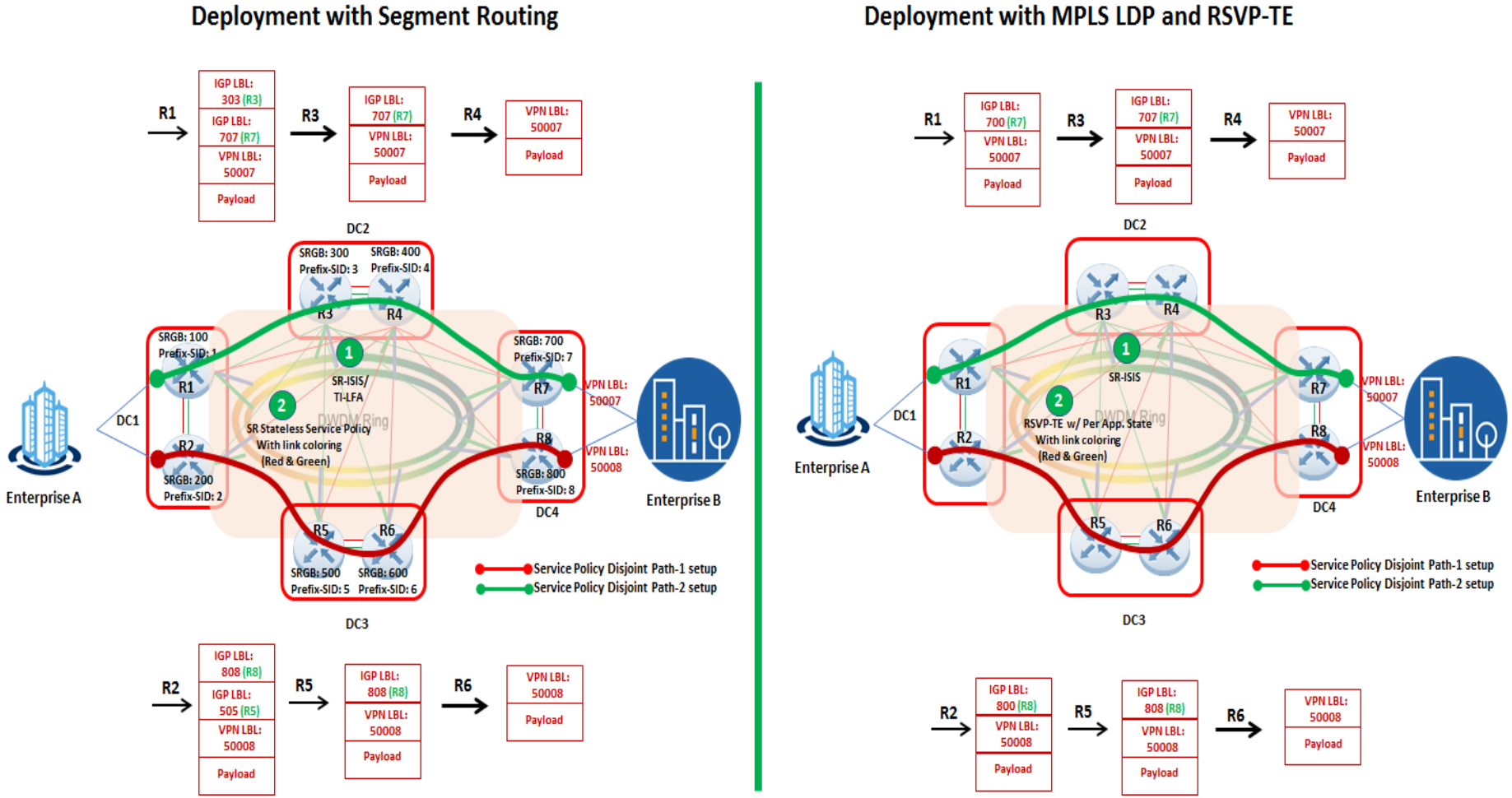
Use Case-1: Low Speed Metro connect services

1G & 10G protected and unprotected optical services over packet switching network

Design Requirement	Segment Routing	MPLS Traffic Engineering
Low CAPEX Cost	Mix of silicon rather then custom only	Complex nature of technology demands custom silicon to meet scaling needs
Low OPEX Cost	• EVPN with Segment routing ISIS/TI-LFA • SR helps simplify network and eliminates use of complex protocols thus helps ease troubleshooting	• EVPN with IGP-ISIS/LDP/RSVP-TE/FRR • RSVP-TE complicates network requires expert to understand complex RSVP control plane for troubleshooting
Service Types: Dual Diverse and Single Protected	Link coloring Service Policy	Link coloring Stateful Service policy
Convergence : 50ms with link or Node failure	TI-LFA	Fast Reroute
Transparently tunnel all layer 2 PDUs	L2 Transparency	L2 Transparency
Link loss forwarding: CE link will go down upon remote or local PE link failure	EVPN route Type-1 support	EVPN route Type-1 support
ECMP to efficiently utilize infrastructure BW	Segment routing is Packet optimized hence inherently supports ECMP	Technology is circuit optimized hence for every ECMP Path need one policy
End to End path latency and jitter visibility to customers	CFM on VPWS end points & Y.1731 over CFM	CFM on VPWS end points & Y.1731 over CFM

Use Case-1: Low Speed Metro connect services

1G & 10G protected and unprotected optical services over packet switching network



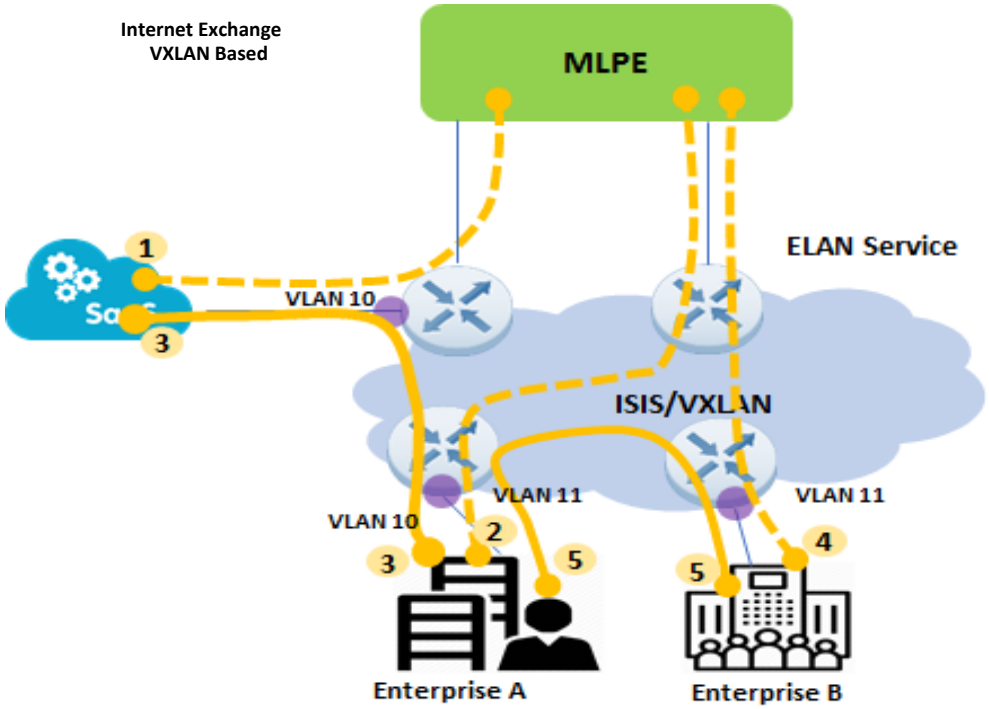
Use Case-2: IXP Public and Private Peering

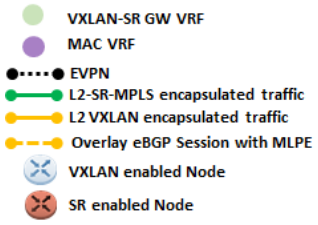
Internet Exchange migration from legacy VLL/VPLS/LDP based network to EVPN / Segment routing

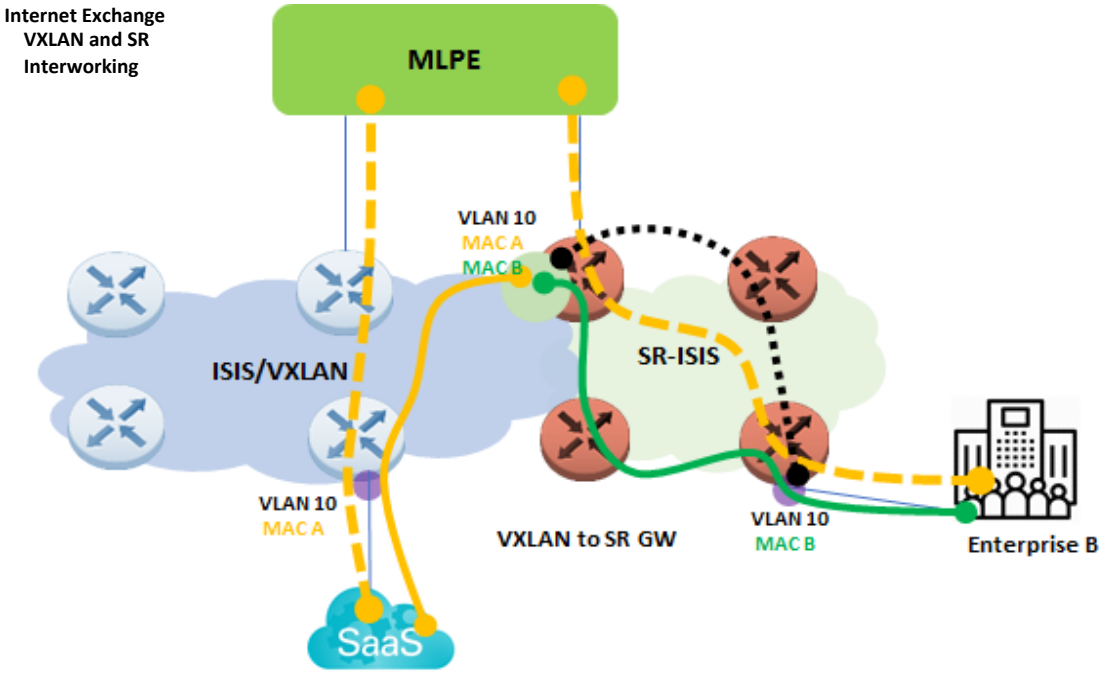
Design Requirement	Segment Routing	Traditional MPLS
Low OPEX and CAPEX Cost	Mix of silicon rather then custom only	Complex nature of technology demands custom silicon to meet scaling needs
Service Types: multi-lateral & Bi-lateral Public and Private Peering	Public Peering: Single Layer 2 Bridge Domain (ELAN) Private Peering: P2P VLAN based Bridge Domain (EVPL)	Public Peering: Single Layer 2 Bridge Domain (ELAN) Private Peering: P2P VLAN based Bridge Domain (EVPL)
Convergence : 50ms with link or Node failure	TI-LFA	RSVP-TE/Fast Reroute
Implement methods to avoid flood and learn layer 2 MAC learning	EVPN route Type 1, 2 and 4 support	EVPN route Type 1, 2 and 4 support
Service Migration / Co-existance	SR-VXLAN interworking/GW function	MPLS – VXLAN GW function
ECMP to efficiently utilize infrastructure BW	Segment routing is Packet optimized hence inherently supports ECMP	Technology is circuit optimized hence for every ECMP Path need one policy
Peering Security to avoid DOS attacks/BGP prefix hijack attacks etc ..	RPKI Origin Validation , BGPSEC , BGP FlowSpec, Per-Peer control Plane Policers	RPKI Origin Validation , BGPSEC , BGP FlowSpec, Per-Peer control Plane Policers
BGP monitoring	BGP Mon along with sampling and traffic to AS mapping	BGP Mon along with sampling and traffic to AS mapping

Use Case-2: IXP Public and Private Peering

Internet Exchange migration from legacy VLL/VPLS/LDP based network to EVPN / Segment routing







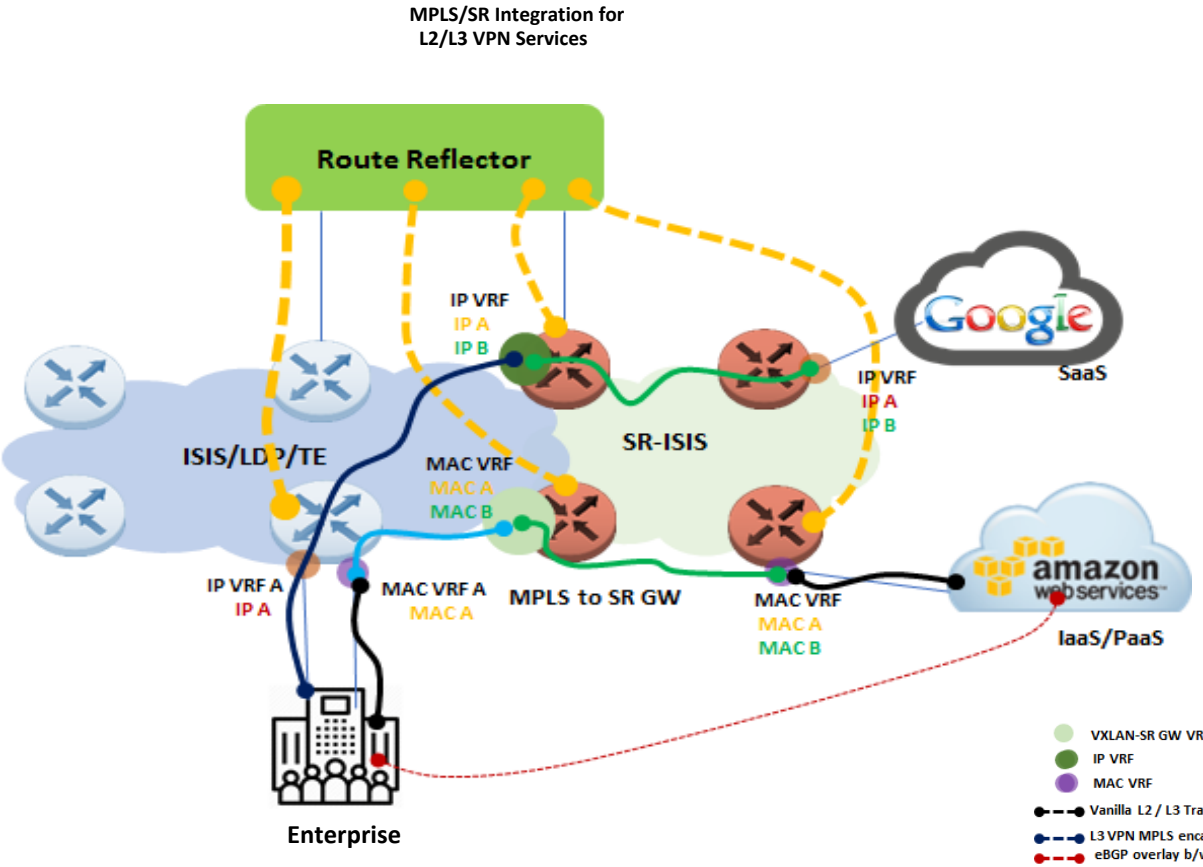
Use Case-3: Enterprise Cloud Exchange Services

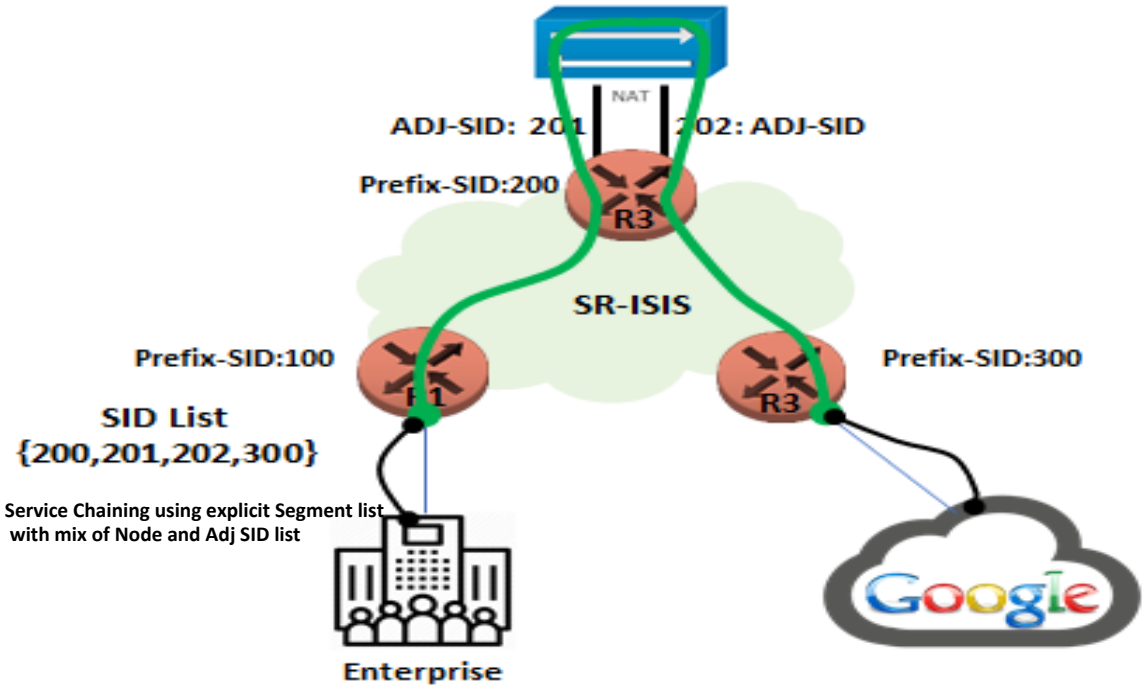
IaaS/PaaS & SaaS Service consumption by Enterprise customer using Equinix private exchange

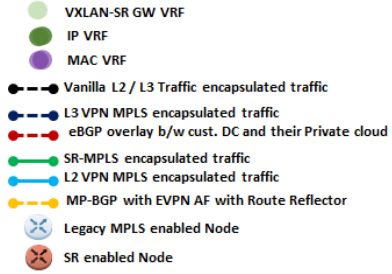
Design Requirement	Segment Routing	Traditional MPLS
Low OPEX and CAPEX Cost	Mix of silicon rather then custom only	Complex nature of technology demands custom silicon to meet scaling needs
Service Types-1: Layer 3 CSP SaaS offering **Service Types-2:** Layer 2 enterprise subscribing layer 3 SaaS services **Service Types-3:** Layer 3 Enterprise with Private IP address range should be able to subscribe L3 SaaS **Service Types-4:** Traffic between Enterprise private cloud should be able to traverse via FW	**Layer 3 CSP (SaaS):** Hub and Spoke L3VPN Services – **EVPN route type 1 , 2 and 5** **Layer 2 CSP (IaaS/PaaS):** P2P Layer 2 VPN Services – EVPN route type 1 , 2 and 4 **Layer 2 Enterprise subscribing Layer 3 SaaS:** IRB functionality is required to terminate L2 into L3 **Service Chaining:** Chain traffic via FW / NAT when traffic traverse between customer private clouds	**Layer 3 CSP (SaaS):** P2P & Hub and Spoke L3VPN Services – **EVPN route type 1 , 2 and 5** **Layer 2 CSP (IaaS/PaaS):** P2P Layer 2 VPN Services – EVPN route type 1 , 2 and 4 **Layer 2 Enterprise subscribing Layer 3 SaaS:** IRB functionality is required to terminate L2 into L3 **Service Chaining:** Enterprise Private IP range will be NAT'ed to Public subnet range or FW when traffic traverses between public clouds
50ms convergence upon link and node failure	TI-LFA	IP-FRR or MPLS FRR
Service Chaining enablement	Stateless Service Policy – Binding SID	a) Stateful MPLS Traffic Engineering tunnels + Stitching b) LDP LSPs with NAT service card ** To Support L2 and L3VPN requires with Chaining requires both LDP and RSVP TE
ECMP to efficiently utilize infrastructure BW	Solution should not be circuit Optimized but Segment optimized	Technology is circuit optimized hence for every ECMP Path need one policy
Class of service based Traffic engineering	SR supports Class of Service-based TE (CoS) where one can define per-flow CoS policies and encode a segment to fulfill the CoS demands.	RSVP-TE has failed to provide this level of granular control due to scalability issues
Legacy and Segment routing integration	SR to MPLS Gateway Function	SR to MPLS Gateway Function

Use Case-3: Enterprise Cloud Exchange Services

IaaS/PaaS & SaaS Service consumption by Enterprise customer using Equinix private exchange







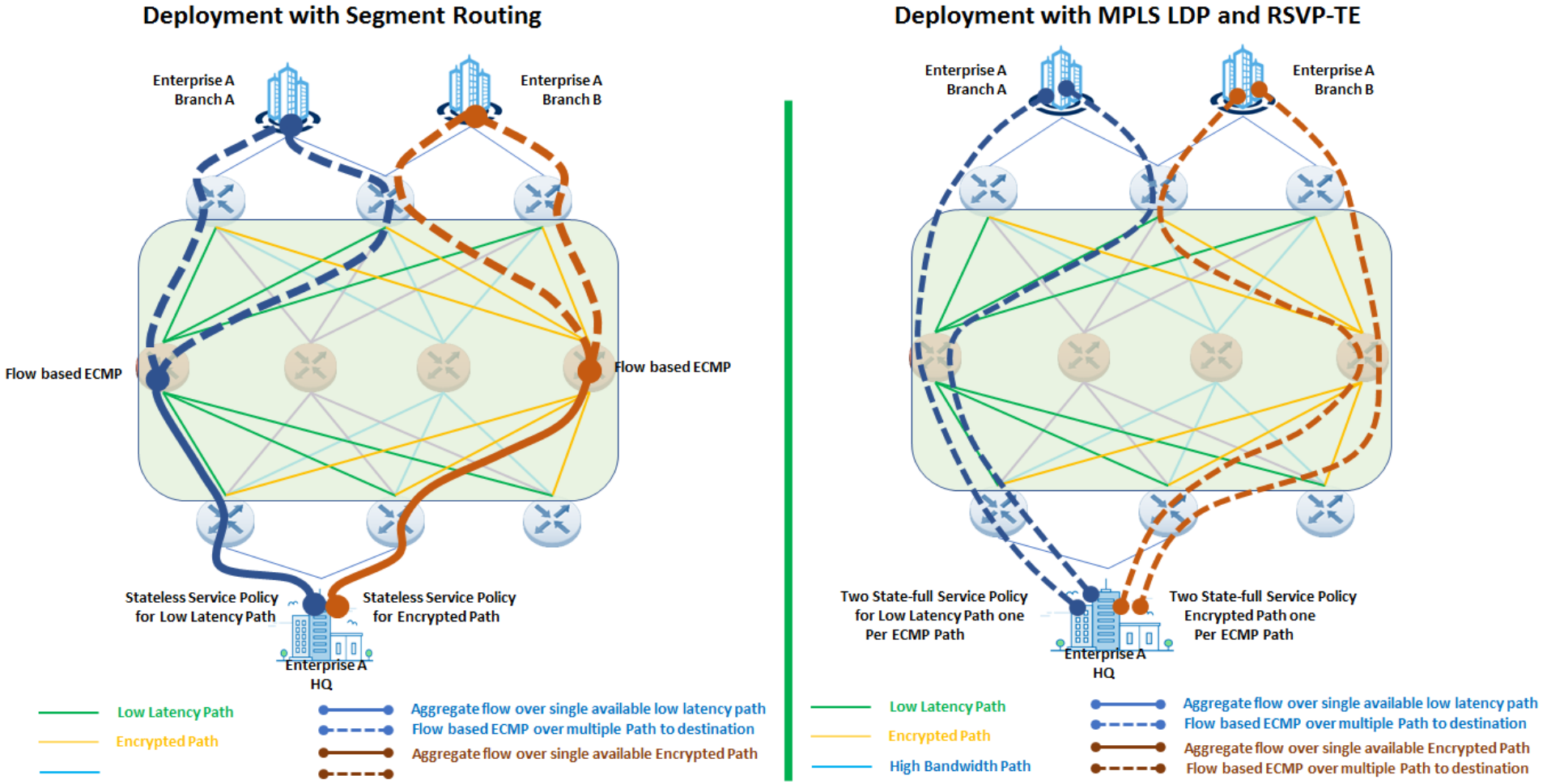
Use Case-4: Application SLA based Path selection

Application SLA based Service offering over Equinix Private exchange

Design Requirement	Segment Routing	Traditional MPLS
Low OPEX and CAPEX Cost	EVPN with Segment routing ISIS **(LOW OPEX)** Complete solution can be deployed with Merchant silicon only **(LOW CAPEX)**	EVPN with MPLS-LDP/RSVP-TE **(High OPEX)** Mix of silicon rather then custom only **(HIGH CAPEX)**
Service Type-1: Traffic steering to encrypted network slice **Service Types-2:** Traffic steering to low latency network slice **Service Types-3:** Traffic steering to high bandwidth network slice **Service Types-4:** Service Chain traffic between Enterprise private cloud should be able to traverse via FW	**Service Type-1:** Create end to end MACSEC or edge supported IPSEC enabled network slice , create stateless service policies with Binding SID to enable application steering into the encrypted network slice **Service-Type-2:** Create network slice with low latency path, create stateless service policies with Binding SID to enable application steering into low latency network slice **Service-Type-3:**Create network slice with high bandwidth path, create stateless service policies with Binding SID to enable application steering into high bandwidth network slice **Service-Type-4:** Create multiple service policies – from client to service farm, from service farm (NAT/FW) to server and stitch them via Binding SID for end to end flow	**Service Type-1:** Create end to end MACSEC or edge supported IPSEC enabled network slice , create state-full RSVP Traffic-engineering tunnels maps tunnel interface to application traffic to steer traffic into the encrypted network slice **Service-Type-2:** Create network slice with low latency path, create state-full RSVP Traffic-engineering tunnels maps tunnel interface to application traffic to steer traffic low latency network slice **Service-Type-3:**Create network slice with high bandwidth path, create state-full RSVP Traffic-engineering tunnels maps tunnel interface to application traffic to steer traffic into high bandwidth network slice **Service-Type-4:** Create multiple TE tunnels – from client to service farm, from service farm (NAT/FW) to server and stitch TE tunnels for end to end flow
50ms convergence upon link and node failure	TI-LFA	FRR with Link and Node Protection
Real time latency measurement	MPLS LSP Ping and trace-route , Link OAM and Y.1731	MPLS LSP Ping and trace-route , Link OAM and Y.1731
ECMP to efficiently utilize infrastructure BW	Solution should not be circuit Optimized but Segment optimized	Technology is circuit optimized hence ECMP cant be used

Use Case-4: Application SLA based Path selection

Application SLA based Service offering over Equinix Private exchange



Summary

- EVPN overlay in DC combined with SR-TE for DCI is technologically ready and makes sense for these use cases.
  - Leverages well known and widely deployed protocols; BGP for control plane and MPLS for data plane
- Equinix UPF architecture collapses multiple service networks onto cohesive infrastructure.
  - Reduces complexity for operations
  - Provides high customer QoE
    - QoS and TE for multiple services
    - Protection and fast reroute services maintained
  - Provides control to enable service chaining without adding complexity
- New services, such as virtualization and data analytics, being pursued

NOKIA

EQUINIX

Questions?

Pete Moyer <pmoyer@nokia.com>
Muhammad Durrani <mdurrani@equinix.com>