A Software-defined Metro Fabric Architecture using Disaggregated Switching and VXLAN

NANOG 85

Marco Pessi, Global Technical Architect, Pluribus Networks
marco@pluribusnetworks.com

07-JUN-2022

Transport Fixed Broadband IoT Distributed DC

TDM DWDM Routing & Switching LISP VxLAN

TL1 Stacking & Virtual Chassis Automation SDN DPU

Marco Pessi, Global Technical Architect, Pluribus Networks

Agenda

- Metro Ethernet Architectures, Past to Present
- New Approaches based on Commodity Switching + VXLAN
- Management and Control Plane Options
- Deployment Example
- Summary

Metro Ethernet Network

Architectures and Services: a Brief History

- L2VPN MEF services
- Metro Ethernet network architectures: from pure L2 to IP/MPLS
- *Is there a better way?*

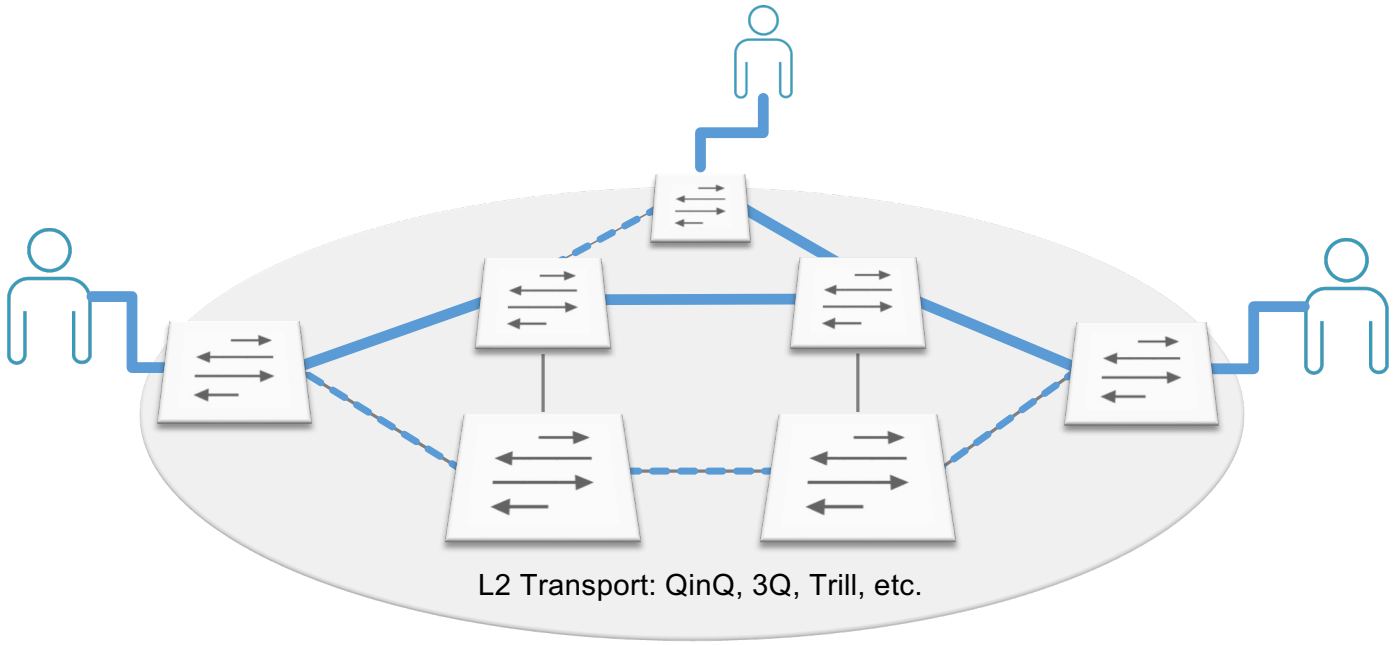
Traditional L2 Fabrics

What do we like/dislike?

+ Relatively Simple to Manage / Automate

− Multi-Pathing Inefficient

− Geographical Scale & Tenant Scale

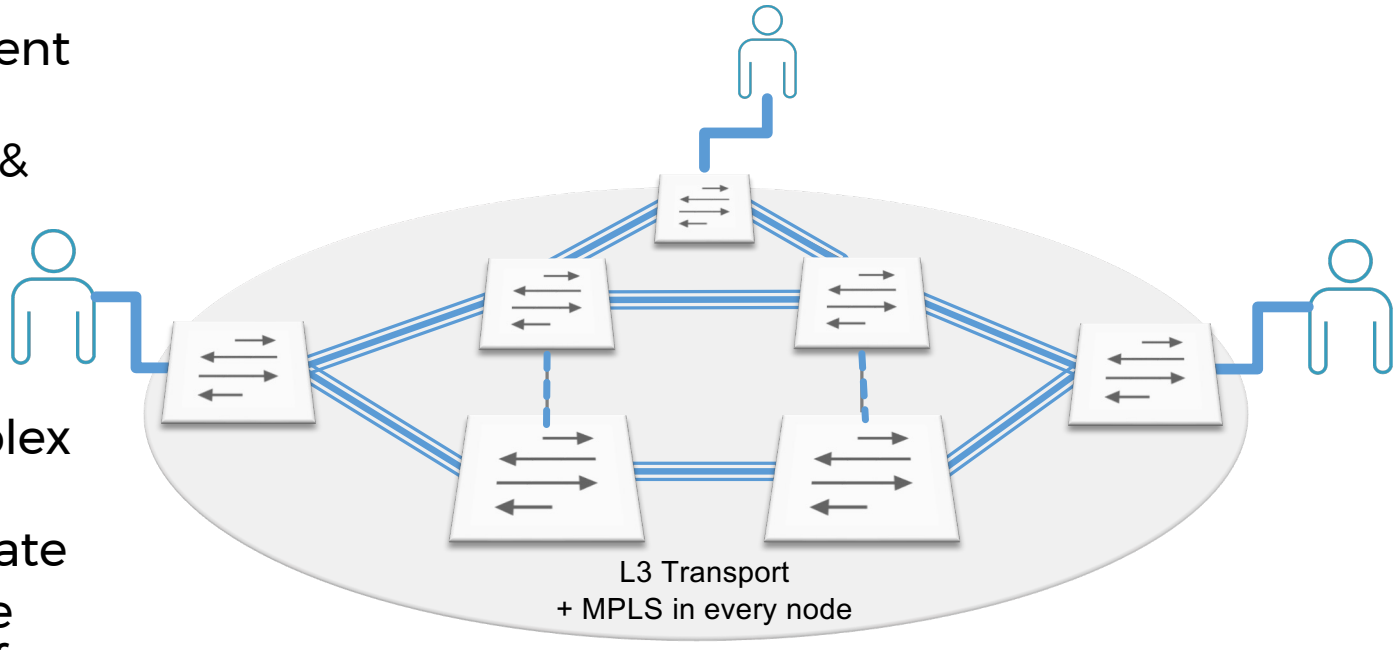


IP/MPLS Fabrics

What do we like/dislike?

+ Multi-Pathing Efficient (Best Path & ECMP)
+ Geographical scale & “theoretical” tenant scale

– Every node must participate in complex MPLS control plane
– Difficult to orchestrate
– Heavy control-plane burden limits use of lower-cost switching and/or tenant scale



Is there a better way?

What about VXLAN and EVPN?

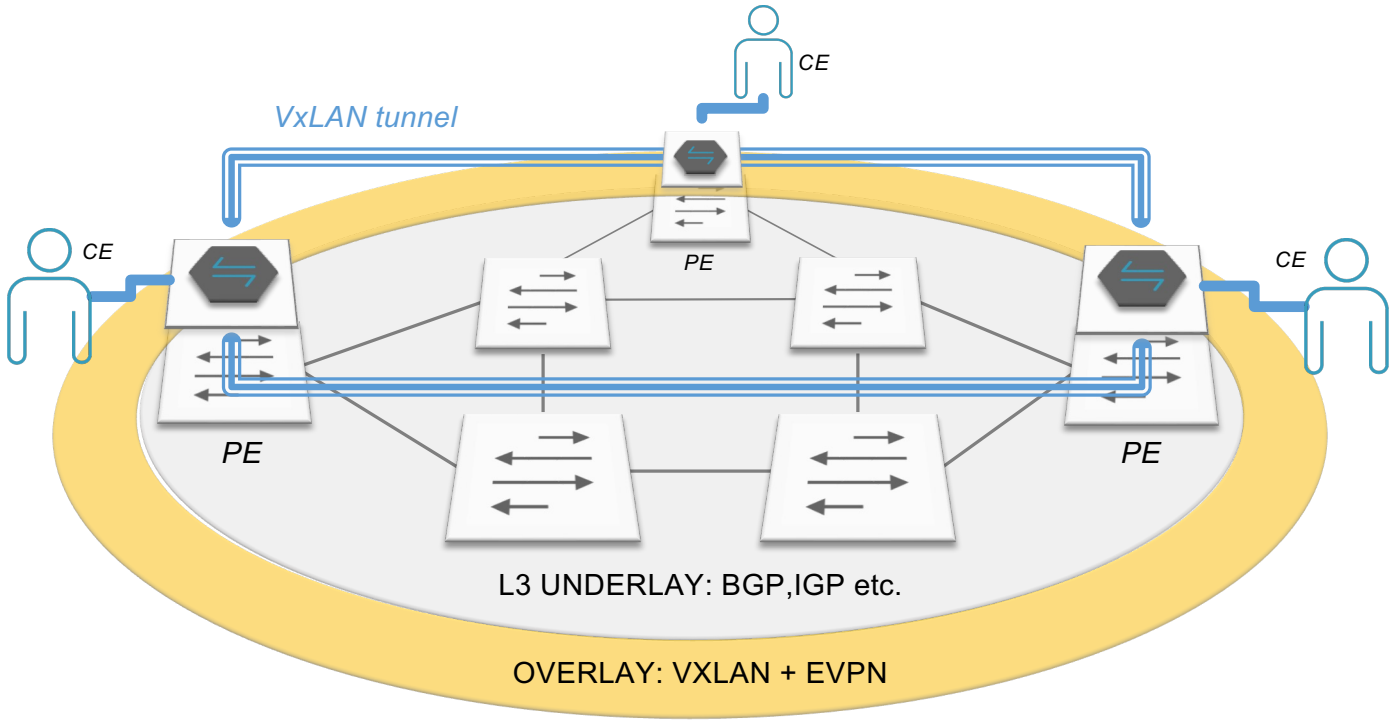
Overlay | Underlay Abstraction with VXLAN & EVPN

OVERLAY

- Daily provisioning

UNDERLAY

- Day 1 provisioning



Q. Is this really a “new” idea?
A. No. We’ve been talking about it for a few years.

R. Steenbergen at NANOG 72 (Feb ‘18)

Conclusions:
- VXLAN is simpler, lighter weight than prior approaches, but it needs a control plane
- SDN with centralized controller doesn’t scale
- BGP EVPN can scale for metro/WAN

R. Dhople at Brazil IX (Dec ‘19)

Conclusions:
- VXLAN over IP can handle all ISP traffic, incl. IPTV, mobile backhaul, MEF services, etc.
- Better utilization, resilience than G.8032 rings
- Narrowing gap on MPLS-equivalent capabilities

So where are the widescale deployments? Something is still missing...

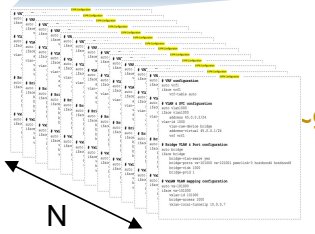
BGP EVPN Configuration

L3 Service Configuration, N-switch Network

EVPN Configuration

```
# VRF configuration
auto vrf1
iface vrf1
    vrf-table auto

# VLAN & SVI configuration
auto vlan1000
iface vlan1000
    address 45.0.0.2/24
vlan-id 1000
    vlan-raw-device bridge
    address-virtual 45.0.0.1/24
    vrf vrf1

# Bridge VLAN & Port configuration
auto bridge
iface bridge
    bridge-vlan-aware yes
    bridge-ports vx-101000 vx-101001 peerlink-3 hostbond4 hostbond5
    bridge-vids 1000
    bridge-pvid 1

# VxLAN VLAN mapping configuration
auto vx-101000
iface vx-101000
    vxlan-id 101000
    bridge-access 1000
    vxlan-local-tunnelip 10.0.0.7

[# RD/RT …]
```

~30 lines per service per switch



N=32 :
~900 lines of config
32 ssh sessions

- BGP EVPN service configuration can be very time-consuming and error-prone
- Operational scale requires some type of automation
- No obvious answer, so network teams are going in different directions
  - Ansible
  - Python / home-grown scripting
  - 3rd party software packages
  - …
  - Static VXLAN configuration 😉

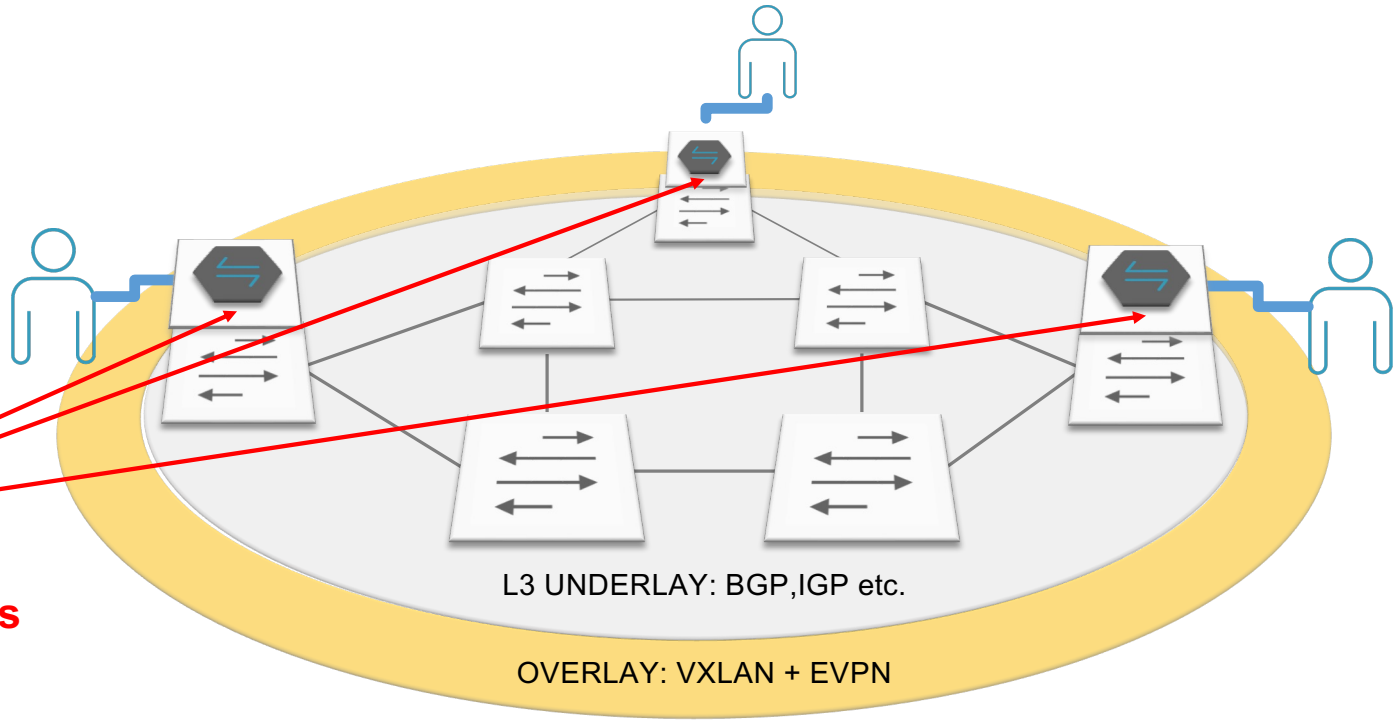
NANOG™

Overlay | Underlay Abstraction with VXLAN & EVPN

Operational Scale Analysis

- Provisioning Blocks

Service Definition



High Code
Independent config states
Not synchronized

Overlay | Underlay Abstraction with VXLAN & EVPN

Operational Scale Analysis

- Provisioning Blocks

Service Definition

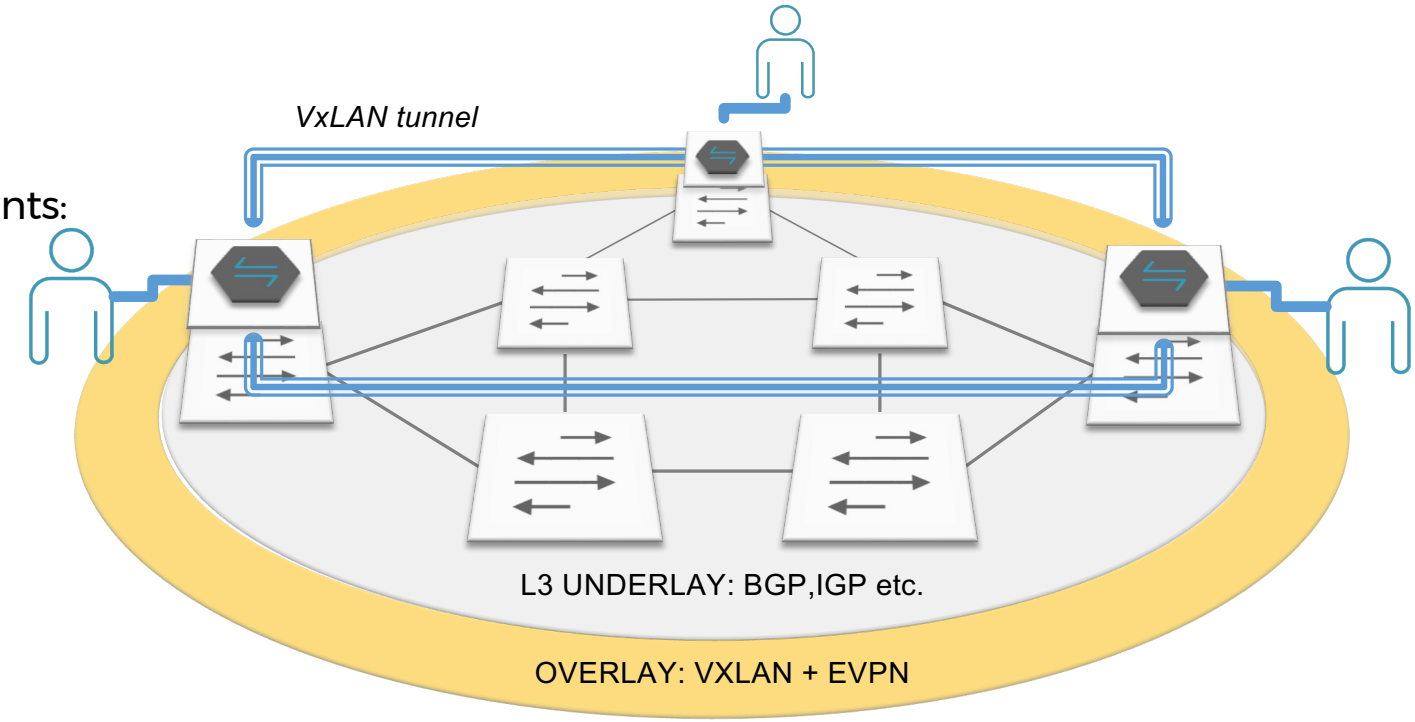
3 independent components:

- Local service
- VxLAN data plane
- EVPN control plane

- Operational Effort

1 Service for N sites:

Total = 3N



Is there a way to improve EVPN?

What about VXLAN [and EVPN] and SDN?

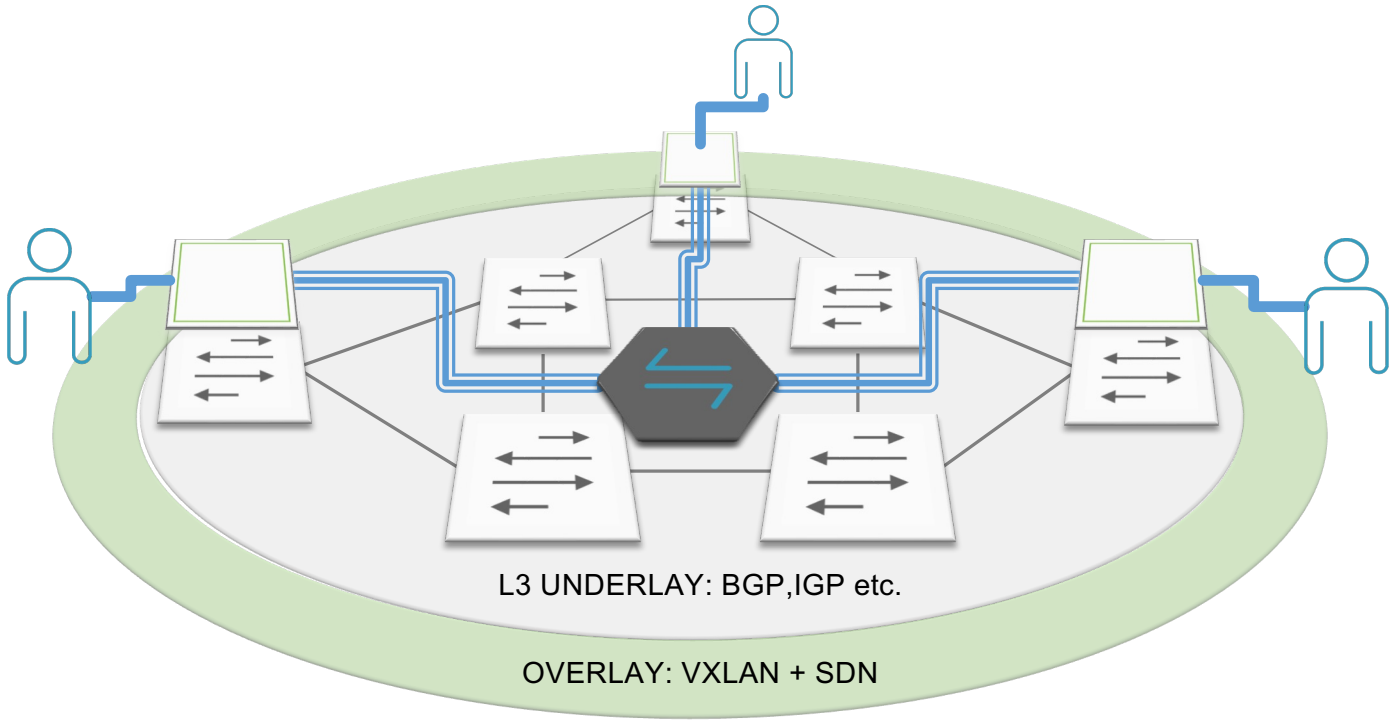
Overlay | Underlay Abstraction with VXLAN & SDN

OVERLAY

- Daily provisioning

UNDERLAY

- Day 1 provisioning



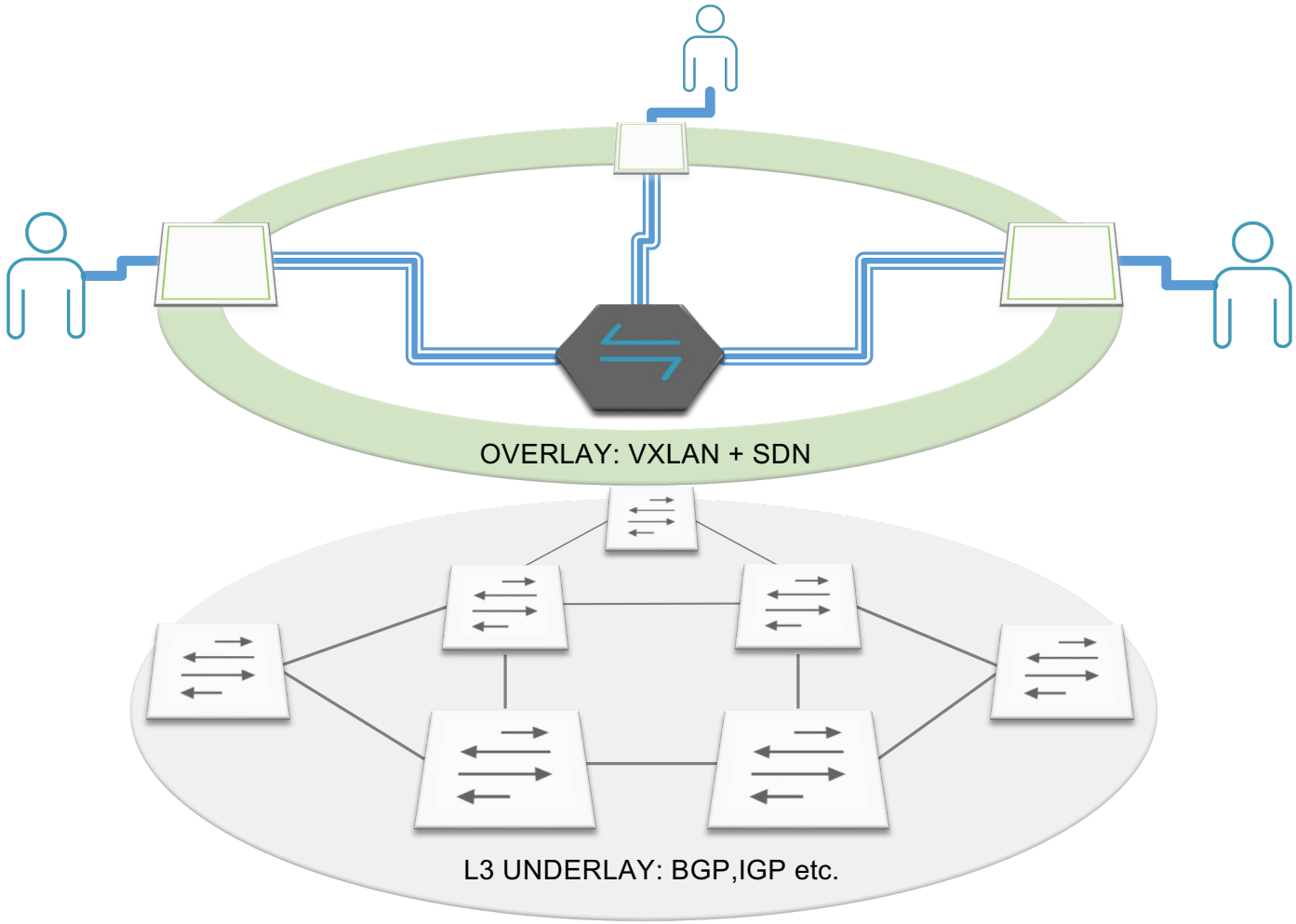
Overlay | Underlay Abstraction with VXLAN & SDN

OVERLAY

- Daily provisioning

UNDERLAY

- Day 1 provisioning



Overlay | Underlay Abstraction with VXLAN & SDN

Operational Scale Analysis

- Provisioning Blocks

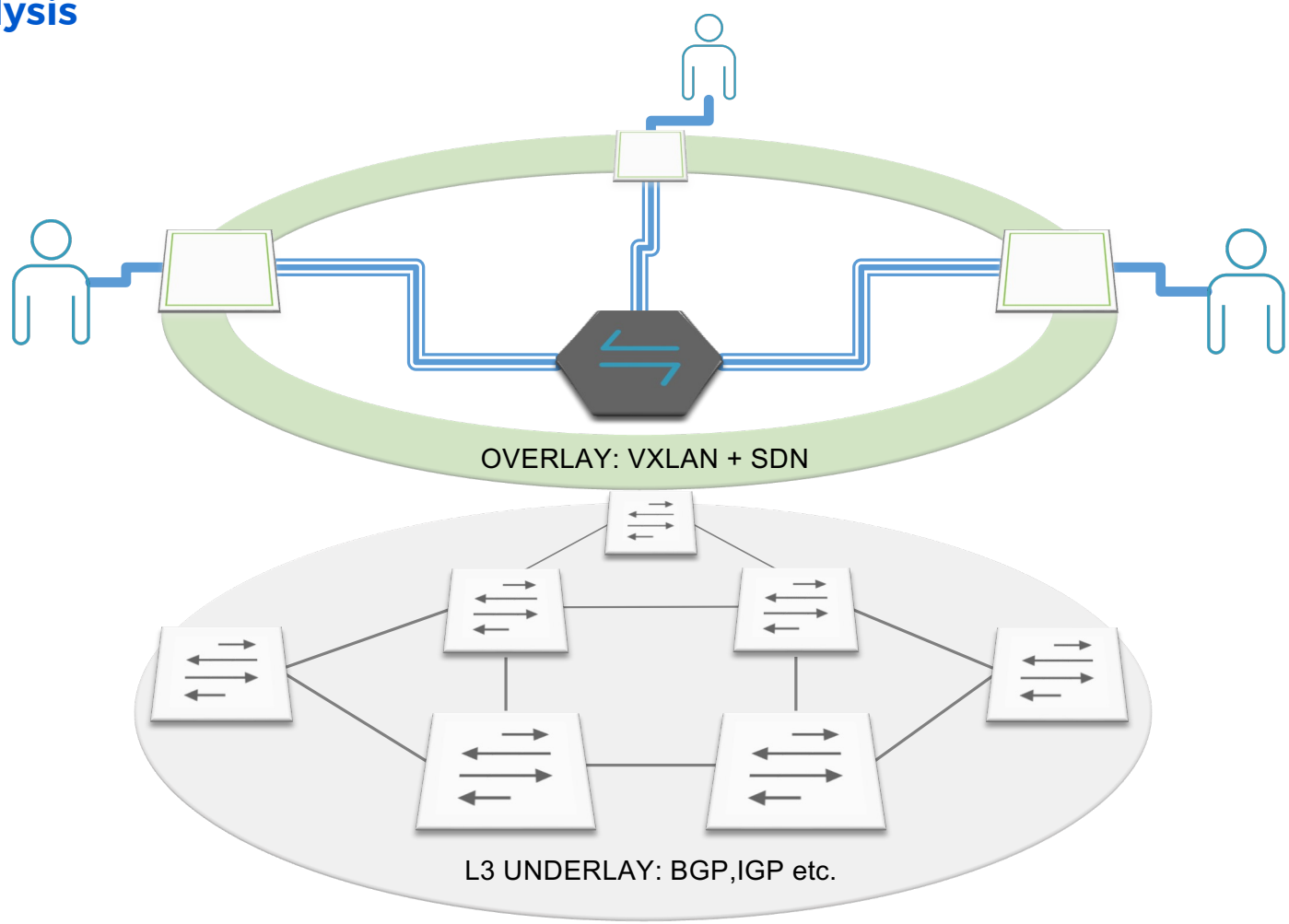
Service Definition

- Operational Effort

1 Service for N sites:

Service Definition = 1

Total = 1



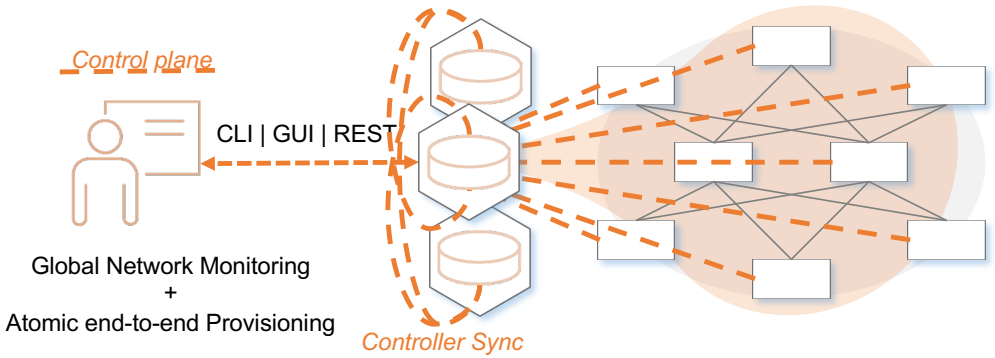
VXLAN Management & Control Plane Options

	Network Scale	Operational Scale
BGP EVPN	Yes	No
Centralized SDN	No	Yes
Distributed SDN (w/ or w/o EVPN)	Yes	Yes

SDN: Centralized vs. Distributed Controller

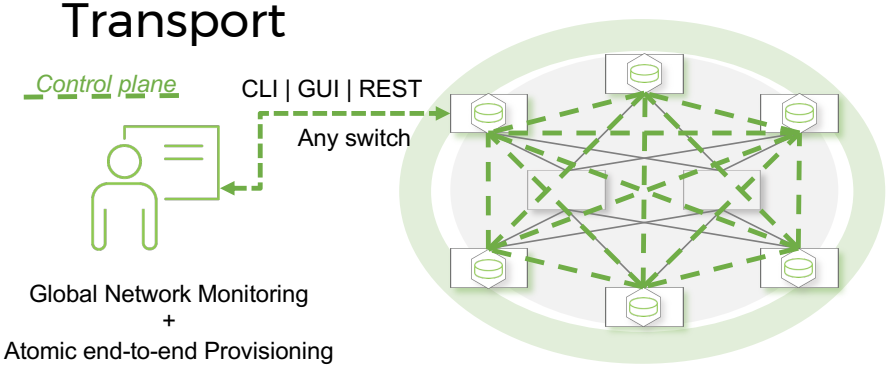
Centralized Controller:

- Needs external controllers → cost, scale & resilience issues
- Requires OOB management network
- Limited topology flexibility



Distributed Controller:

- No external controller
- Switch/network drives scale/resilience
- Use In-band or OOB management network
- Any topology | overlay w/ open Transport



Questions: what network transports the control-plane?
What happens when the network spans multiple sites (and/or not leaf-spine topology)

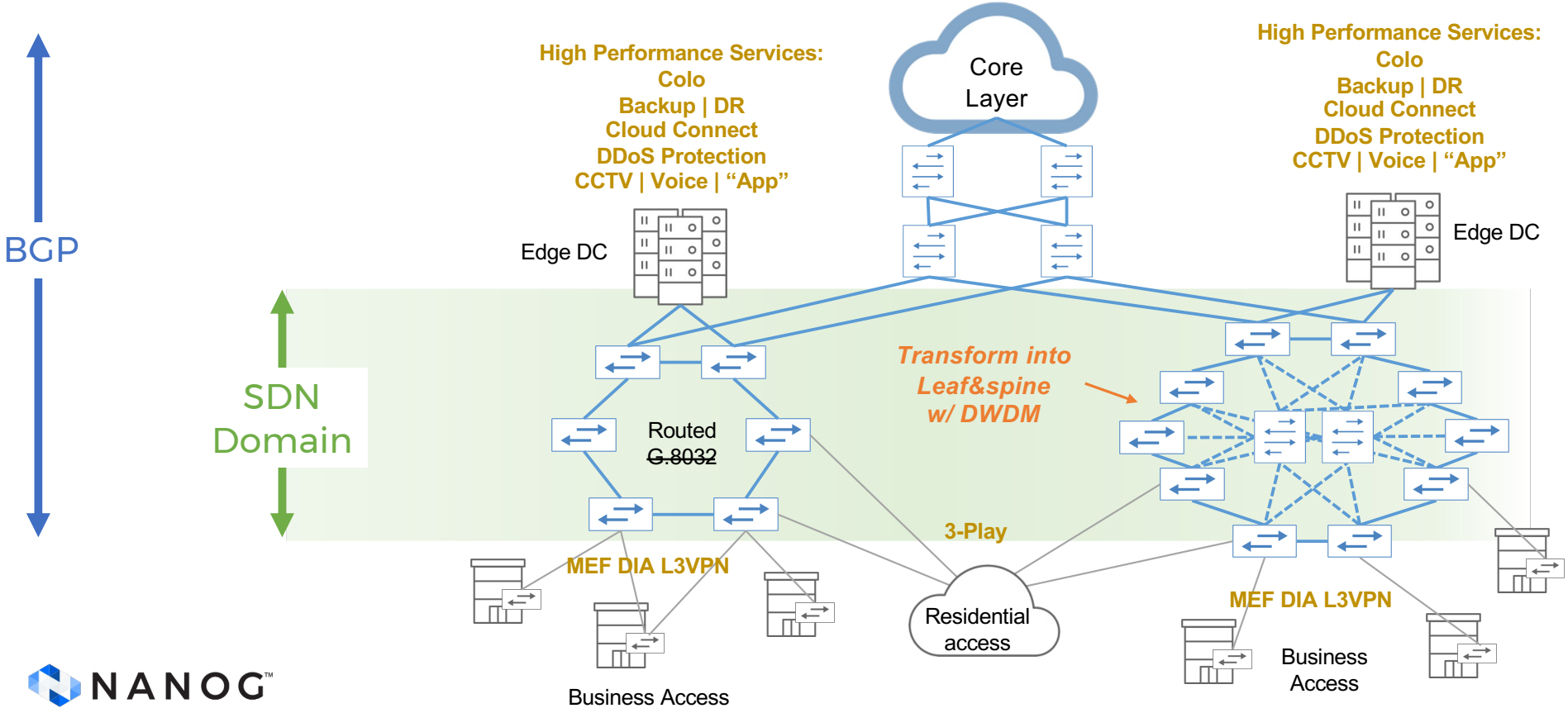
Disaggregated Switching Platforms

- Data-center class switching silicon (e.g. Broadcom Trident 3/4) delivers everything you need:
  - Multi-terabit capacity
  - Table scale & flexibility for multi-tenant L2/L3 services
  - Hardware VXLAN Tunnel End Points (VTEPs) for single-pass line-rate encapsulation
  - Dynamic/Resilient Load Balancing
  - Line-rate telemetry
  - etc., etc.
- Bottom line: No need for custom silicon or deep buffers*

** Assuming reasonable topologies + link capacities*

- *In the data center, we used leaf-spine, non-blocking Clos architectures with rich interconnectivity*
- *Metro topologies vary widely but can achieve similar result, leveraging metro optical DWDM (especially with emerging pluggable 400ZR optics)*

xSP metro network + services fabric
Deployment Example



NANOG™

Summary

- Fully functional, high-performance Metro Ethernet networks can be built using disaggregated commodity switching + VXLAN over IP
- BGP EVPN control plane can provide network scalability, but needs some type of automation layer to achieve operational scalability
- Distributed SDN control plane overcomes scaling and resilience issues of a centralized SDN controller architecture, provides a viable alternative or complement to BGP EVPN with potential for simpler operations

Thank you

07-JUN-2022

NANOG™