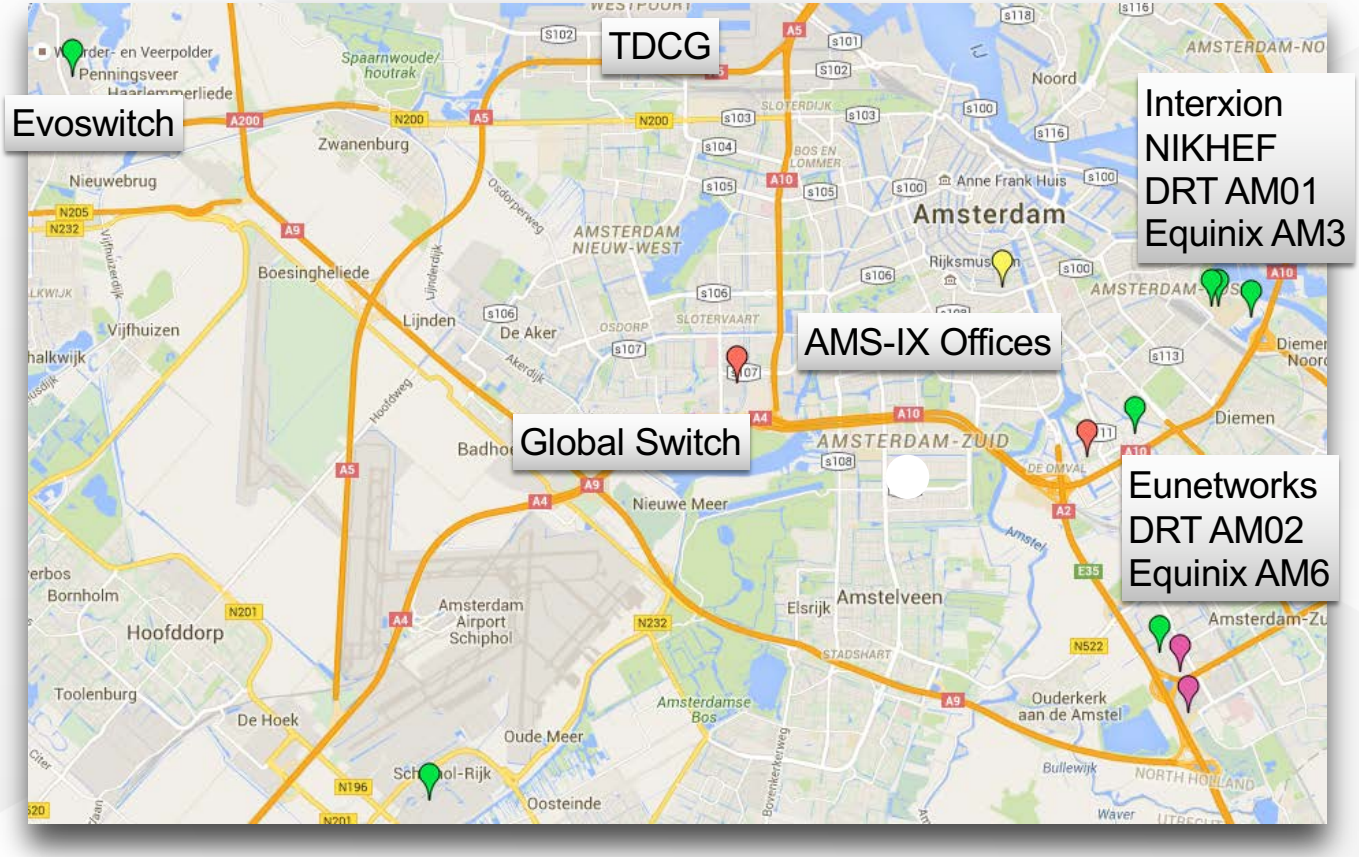
Embracing Open:
The AMS-IX Journey to Open Networking

Bart Myszkowski
Maxx Cherevko
NANOG 76
Washington, DC
10-12 June 2019

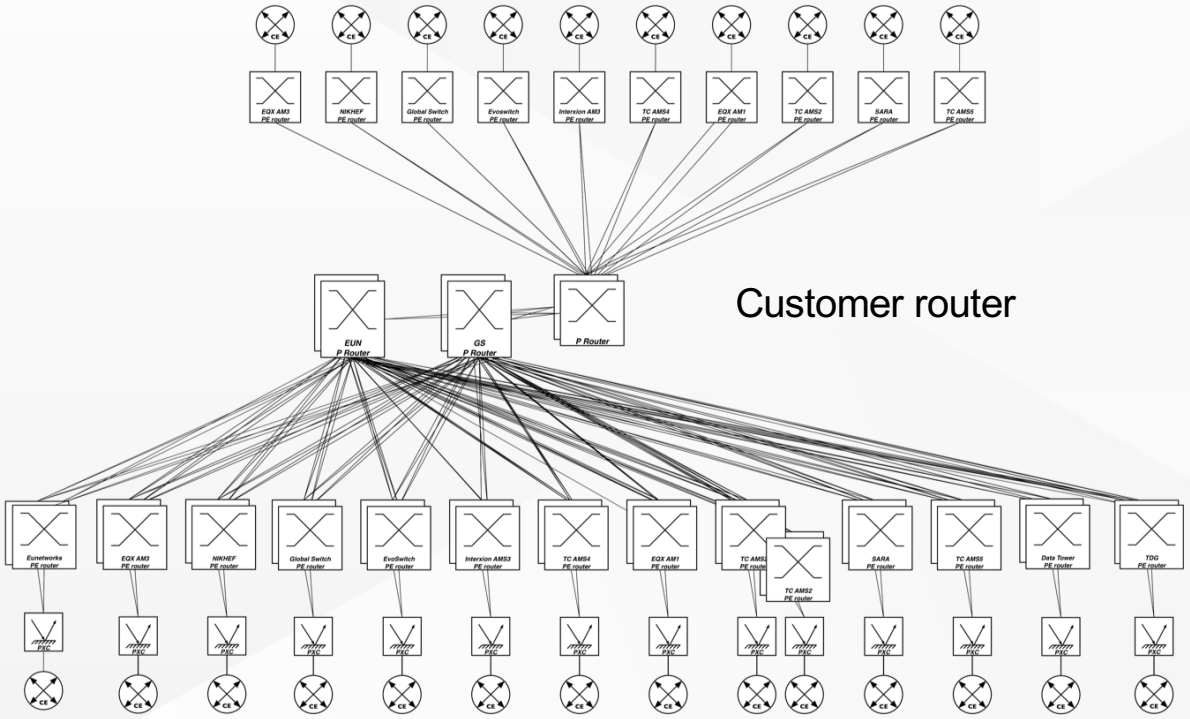
Embracing Open Networking Outline

- AMS-IX introduction
- Network overview and “before” state
- Upgrade motivations, options
- Why we chose open networking
- Open network fabric technology
- Network “after” state
- Experience and lessons learned

AMS-IX in Amsterdam:



AMS-IX Amsterdam Platform



Customer router

Low Speed access

Core or Spine

High Speed access

Customer router

Optical access

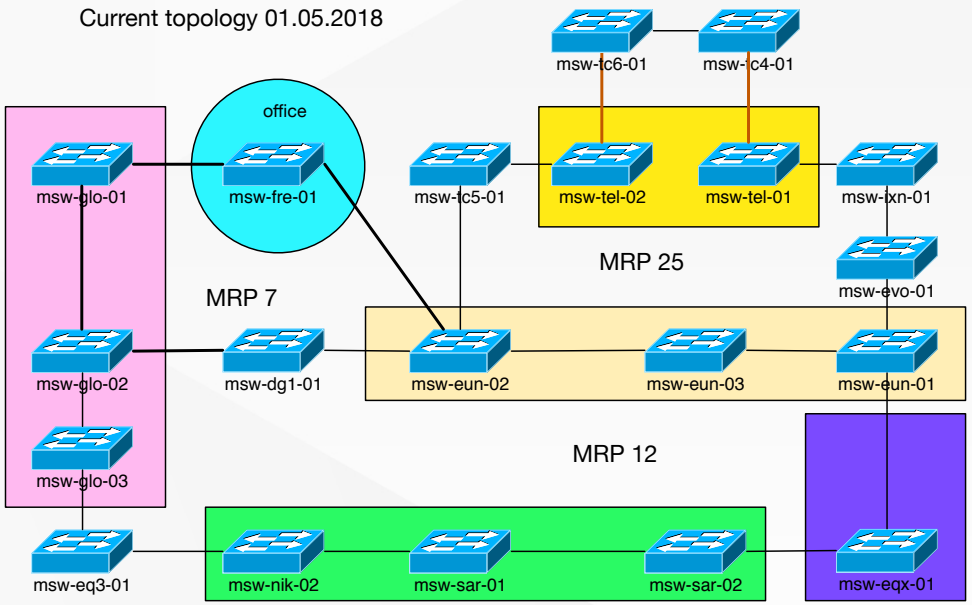
AMS-IX Around the world

AMS-IX management network

- Gives us access to our production equipment (SLX, MLX, DWDMs, PXCs, TS etc.)
- Servers, load-balancers, firewalls, PTP devices, NIDs
- VM/SAN replication
- Monitoring system relies on management network
- Access to the Internet from office/sites

“Before” network set-up

- **Scale**
  - 22 switches, 15 geographically separate locations, 463 ports in use in NL
  - 10 switches on remote locations (CHI, BAY, HK, CW, NY)
- **Equipment in use:**
  - Foundry/Brocade FCX, FES, FGS, ICX (Ruckus)
- **Topology/protocol:**
  - Ring topology: 3 rings connected by 17 dark fibers
  - MRP (metro ring protocol) L2 resilience protocol

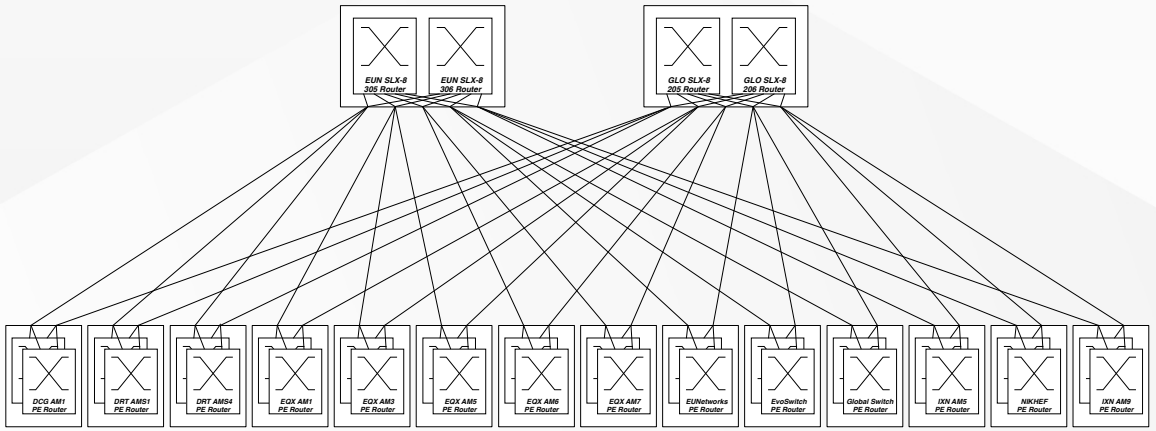


“Before” network issues

- Easy to create a loop/outage
- Inefficient link utilization, some bandwidth bottlenecks
- Ring isolation in case of double fiber cut or issue with MRP
- Different switches with different software versions, challenging to manage
- Some of the switches will be end-of-life soon
- Fiber cost: Management network (17 dark fibers) completely separate from production network (30 dark fibers + DWDM)

Fiber connectivity solution: re-use current production DWDM set-up

- **Use existing DWDM muxes on production fibers to support new channels/wavelengths to connect the management network**
- **Eliminate rings, move to fully redundant leaf-spine topology**
- **Eliminate separate management network fibers, reduce cost**



Switching upgrade goals

- Make environment homogeneous (same HW/SW)
- Higher speed for VM moving, NAS/SAN cluster replication
- More redundant topology
- Easier management
- Better visibility

Where to go?

Technology? Pure L2, TRILL, eVPN, VxLAN etc.

Brand? Cisco, Juniper, Brocade, Arista, Huawei etc.

Hardware? Branded or baremetal

Software? Open source or branded

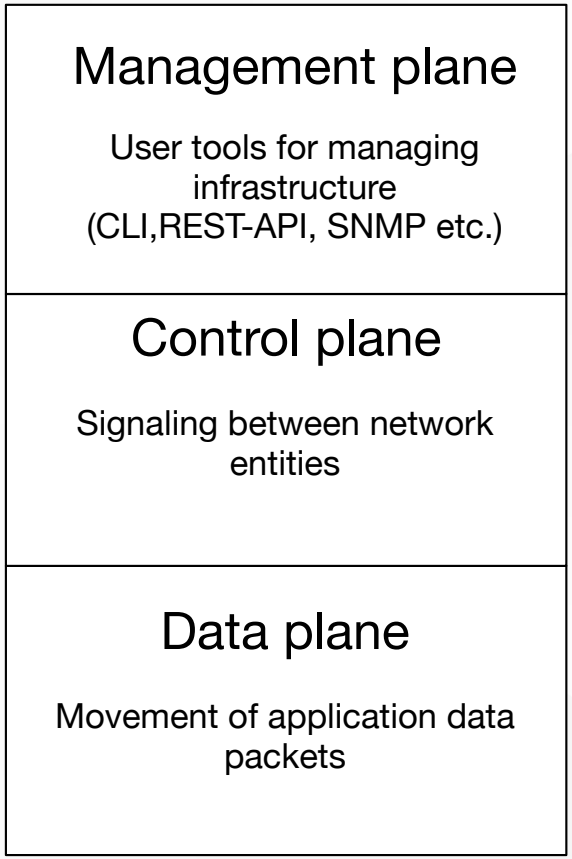
Advantages of open network: bare metal + software

- Decoupling hardware from software on network equipment (same as we have on servers now)
- Ability to change OS or hardware any point of time (like we do with Linux Debian → CentOS)
- New players appeared on the market with newest software features (Pluribus Networks, Cumulus, BigSwitch, IPinfusion etc.)
- Ability to use free OPX (openswitch.net) project

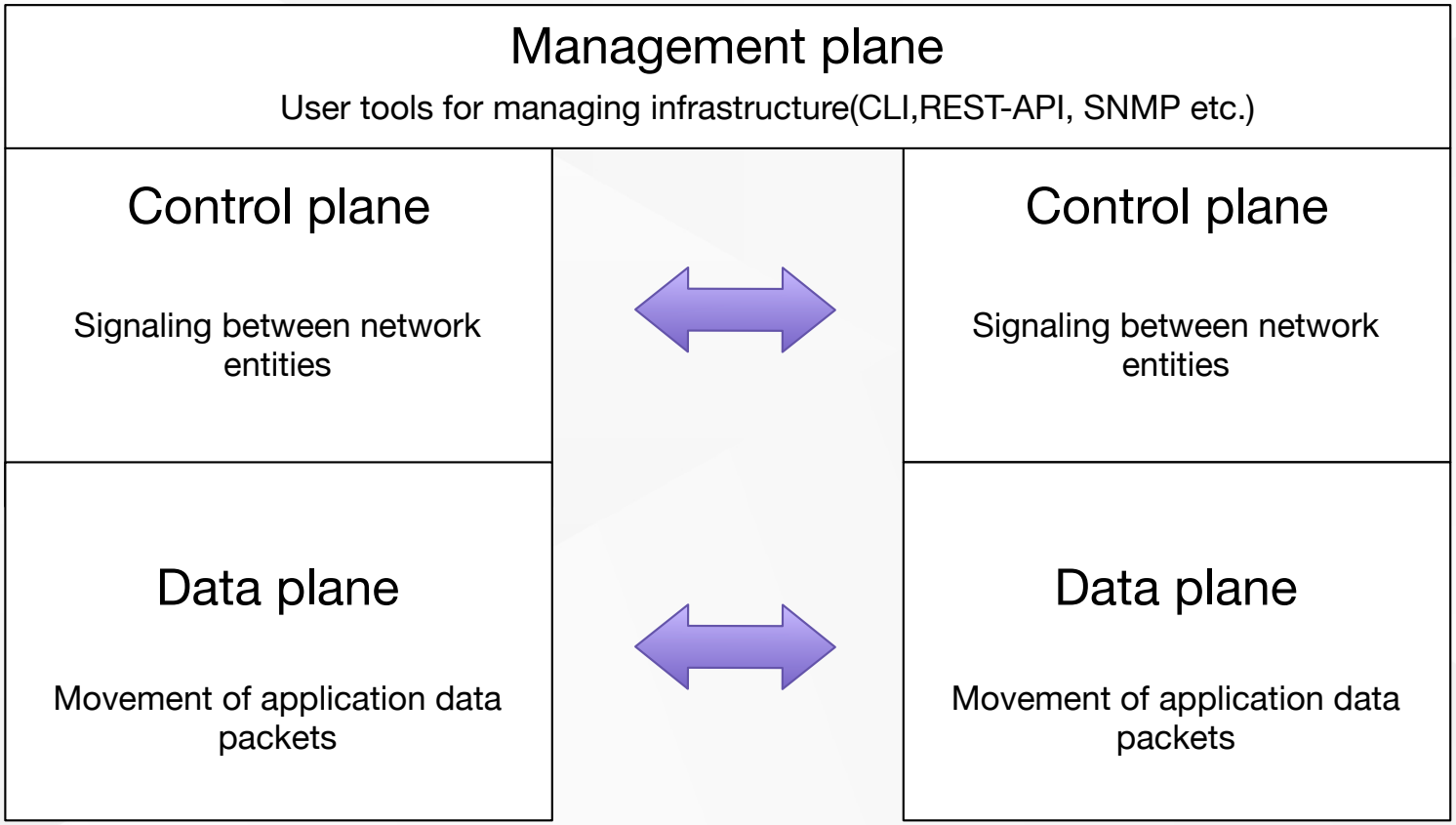
Other decision considerations for open network

- **HW/SW maturity**
  - White box HW standardized in OCP, used for years in hyperscale DCs
  - NOS SW also in wide use, supports all the L2/L3 protocols and features that we need
- **Support**
  - Larger vendors now offering open networking with full support
- **Manageability**
  - Newer SDN approach actually provides better manageability than traditional systems

Classic switch design

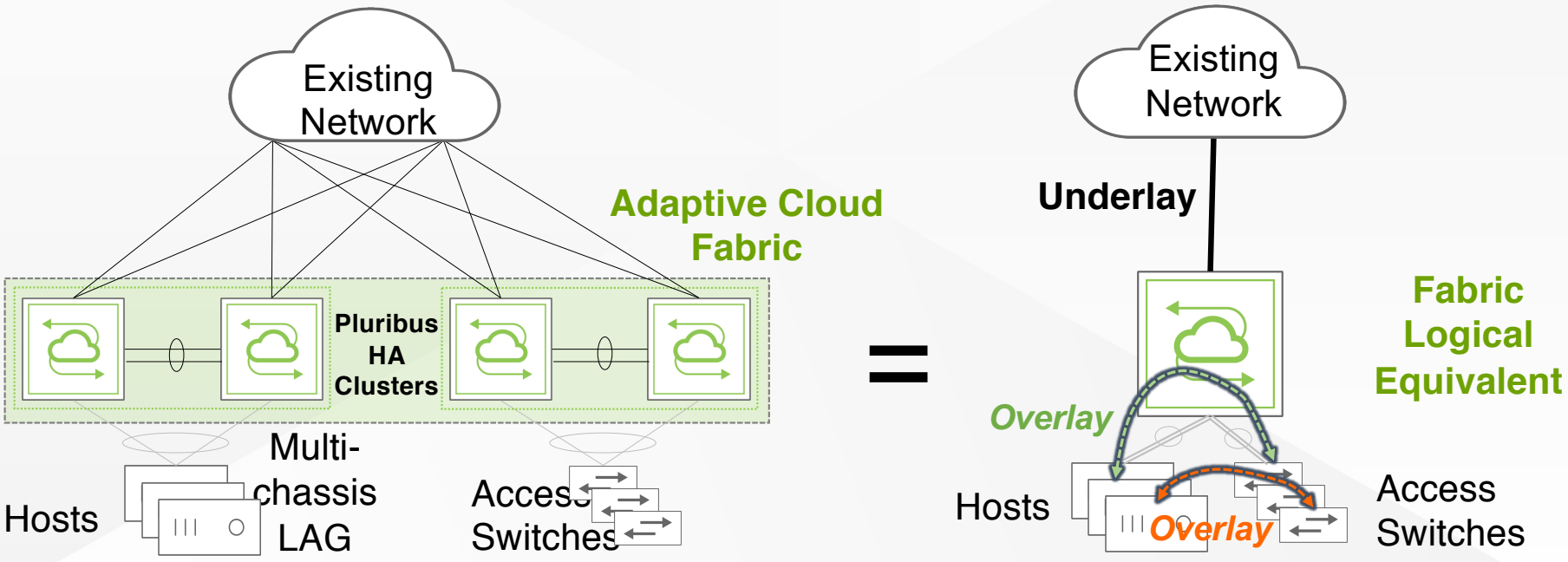


Pluribus distributed SDN fabric concept



Fabric logical view

- Multiple geographically distributed sites act as one programmable entity
- Deploy network services as “fabric object” which updates all switches in fabric



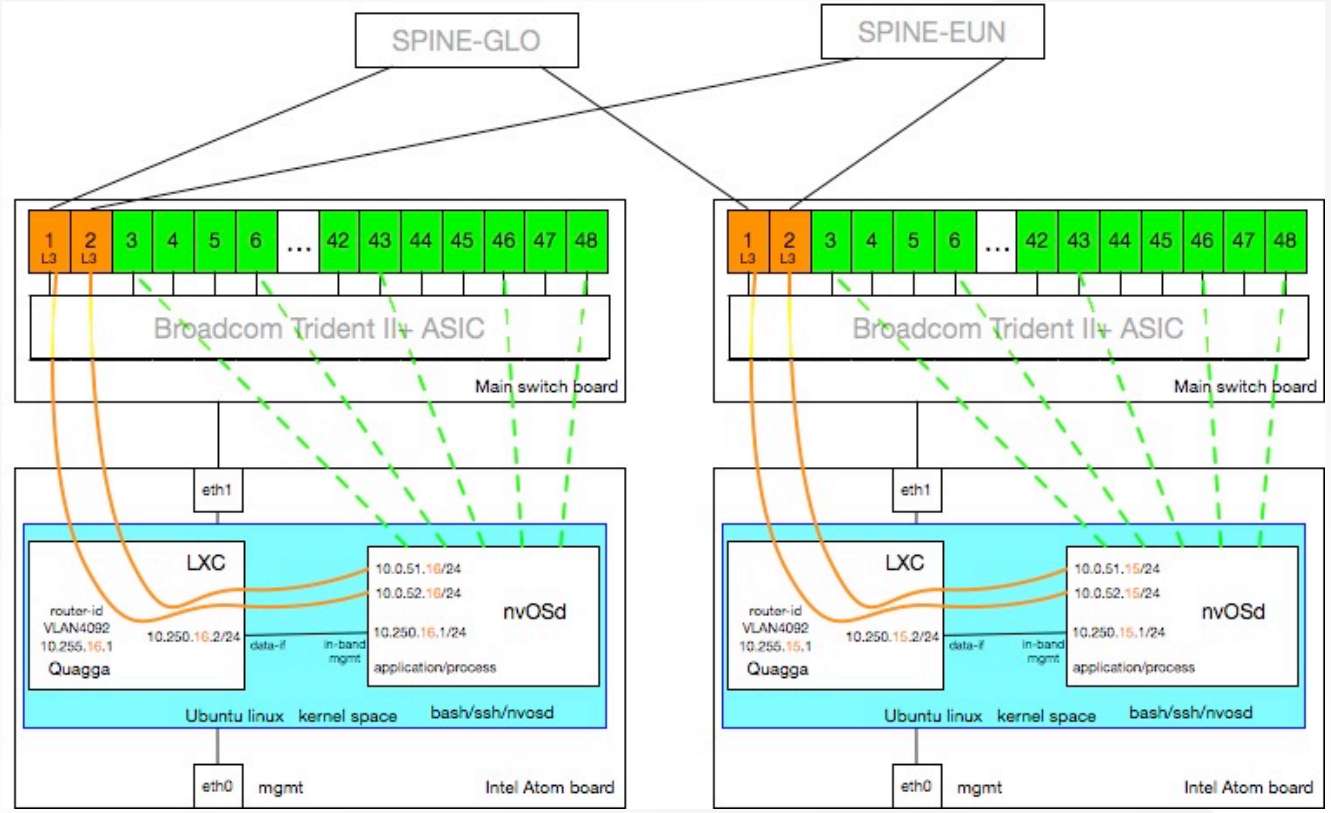
Geographically dispersed sites

Building a fabric with VxLAN

- VxLAN enables L2 network over L3 underlay (with OSPF)
- Use all available links
- Traffic is load balanced using ECMP over all backbone links
- MC-LAG for critical servers/NAS
- Loop-free
- Enables network segmentation for application isolation

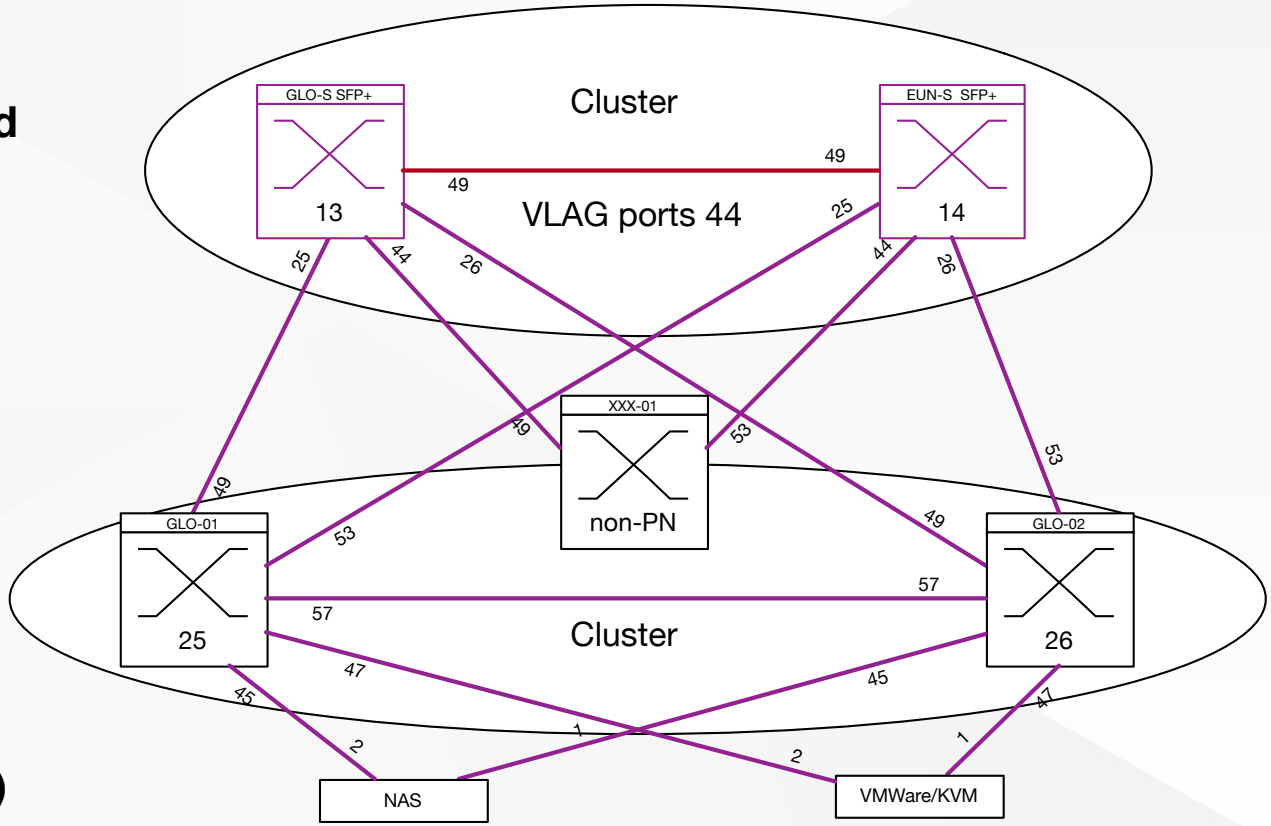
Open switch configuration

- **Switching ASIC connects at high speed to CPU (e.g. Intel)**
- **L2/L3 protocols run in Linux containers**



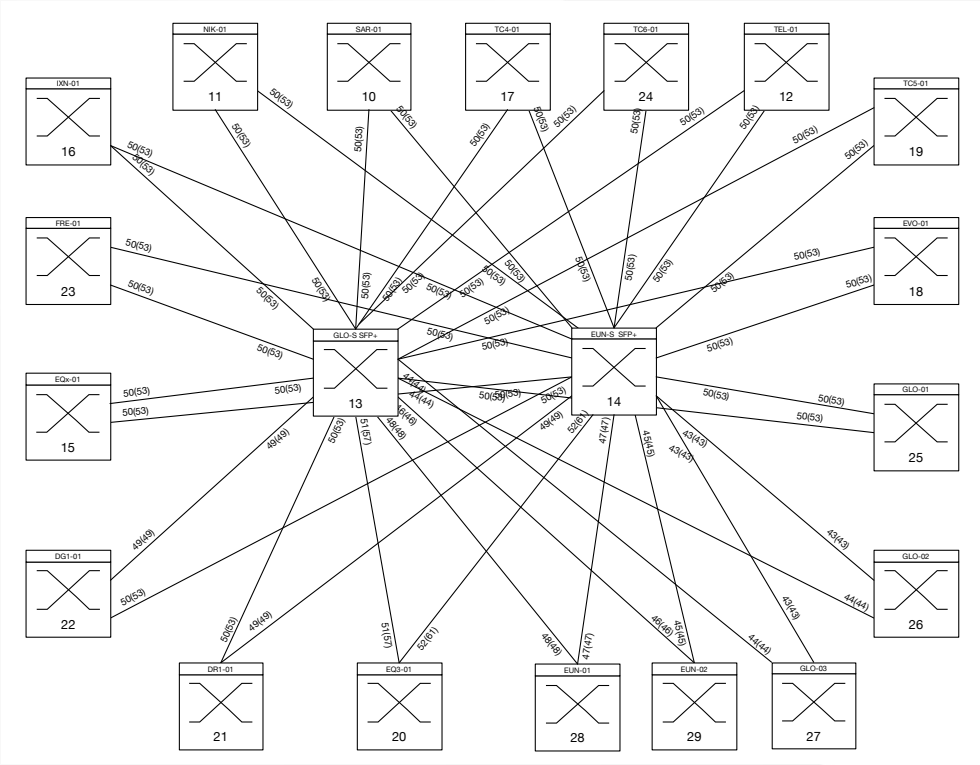
MC-LAG redundant connections

- **Two switches configured as a cluster support redundant connections to avoid downtime during maintenance or device/link failure**
- **Spine cluster enables redundant leaf connections**
- **Leaf cluster used where needed for critical infrastructure (e.g. NAS, production web servers)**



New AMS-IX management network (“after”)

- **Geographically distributed fabric built on standard OSPF underlay**
- **Loop-free ECMP/BFD for efficient multi-pathing**
- **No STP, fast reconvergence**
- **No controller = no split brain, resilient**
- **vLAG for critical servers I NAS**
- **Improved visibility**



Experience to date

- **Best result of adopting new open network approach with fabric concept = simpler management**
  - Whole network visibility and monitoring
  - Automation / reduced manual operations steps, e.g. one step to configure new L2VPN across multiple sites
  - Segmentation / isolation of different applications is built in, managed at fabric level
- **Lower HW costs also a plus**

Thank you!

Questions, suggestions or remarks?