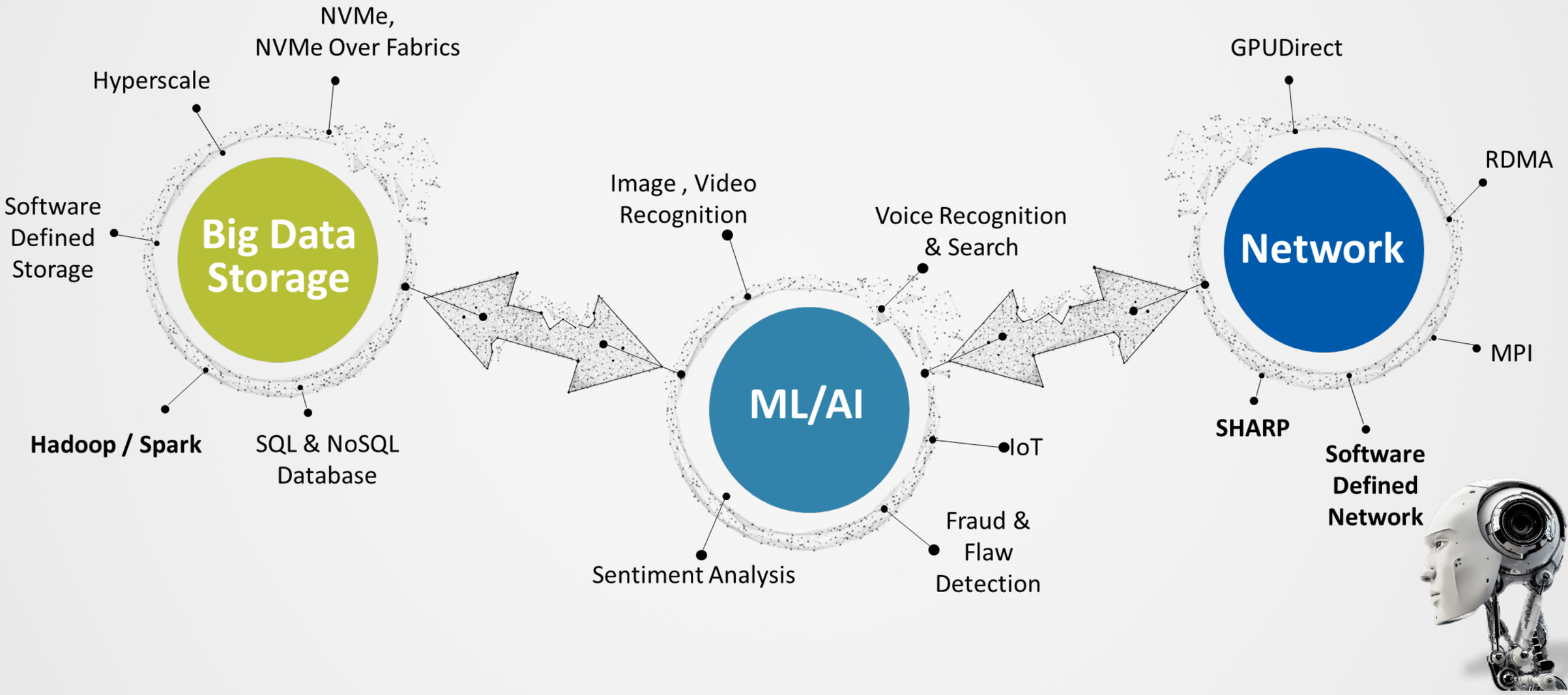
ML/AI Cloud Data Center Evolution

Shaowen Ma, APAC Product Director

March 2019

AI/ML powered by innovation from Network and Storage

AI business grow from $8B in 2016 to $47B by 2020*



Enabling Efficient Machine Learning Platforms

- OCP Big Sur Artificial Intelligence Platform
- Real Time Fraud Detection
- Machine Learning System with 400Gb/s
- 18X Speedup For Image Recognition
- 4X Speedup For Image Recognition
- Data Analytics Image Recognition
- World Record For Data Sort, 3X Faster

Allowing Machine learning to Perform Critical & Real Time Decisions

Switch Architect in 2019

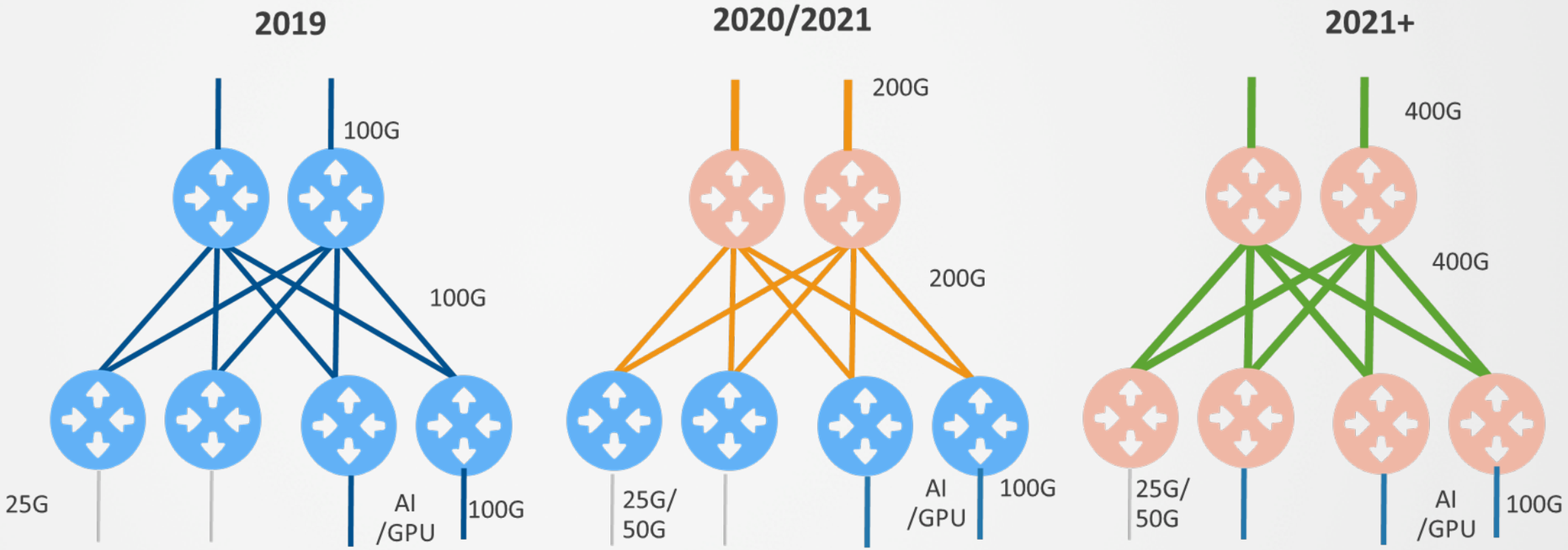
New Hardware in DC

SSD

GPU

TSU

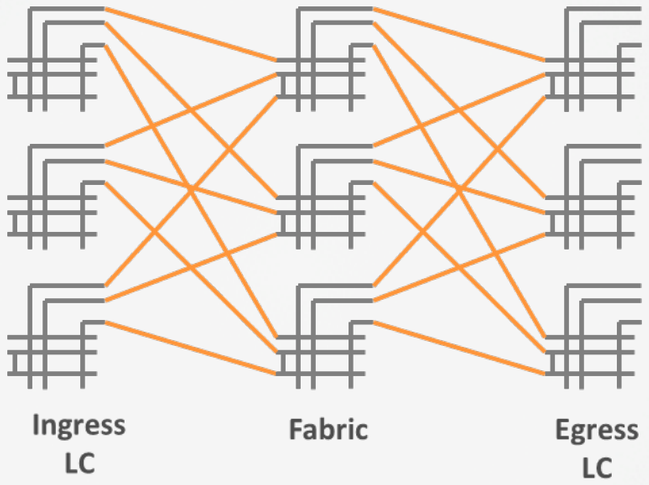
Bandwidth Adoption



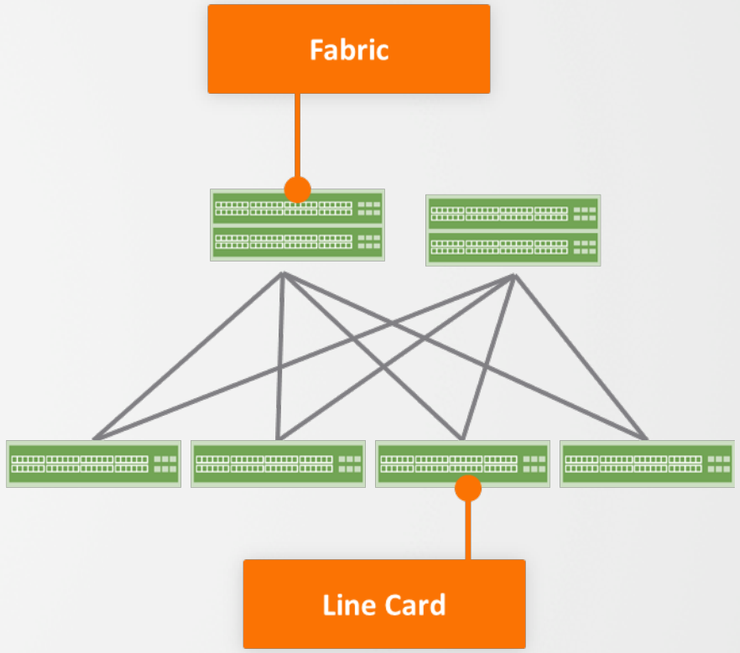
Reuse Spine as Leaf as bandwidth increase?

Modern DC select CLOS Pizzabox Architecture

Facebook Backpack
4 x 32 x 100GE + 8 x Fabric

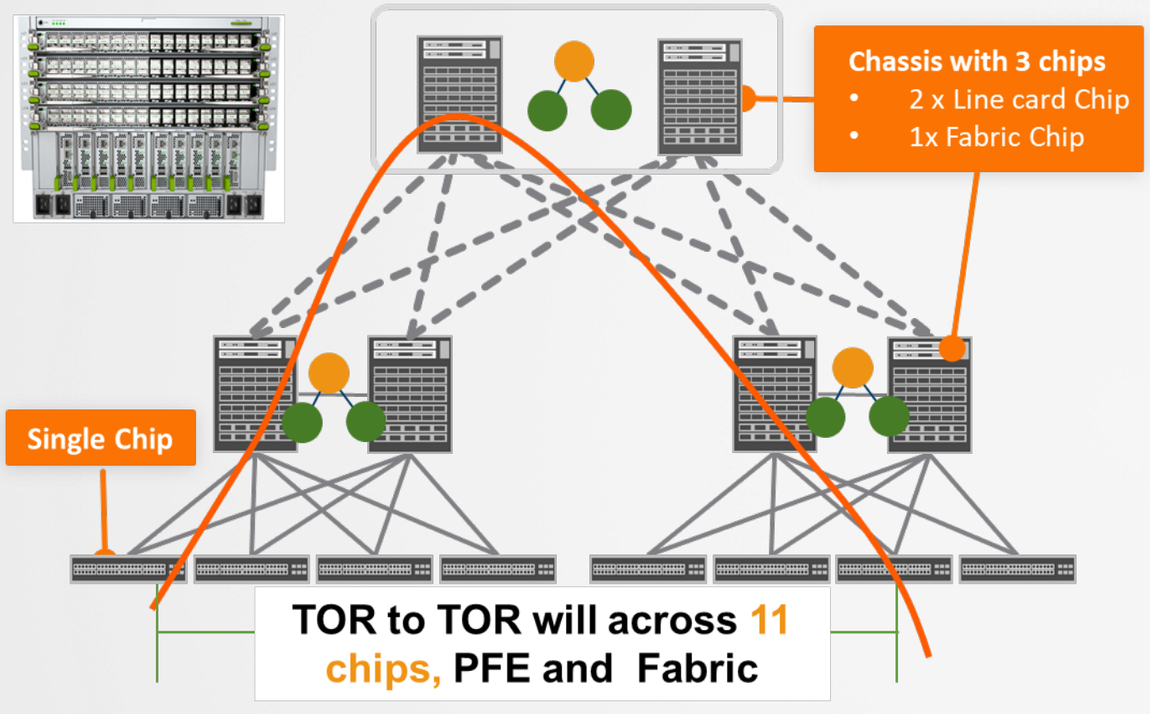


Charles CLOS - 1953

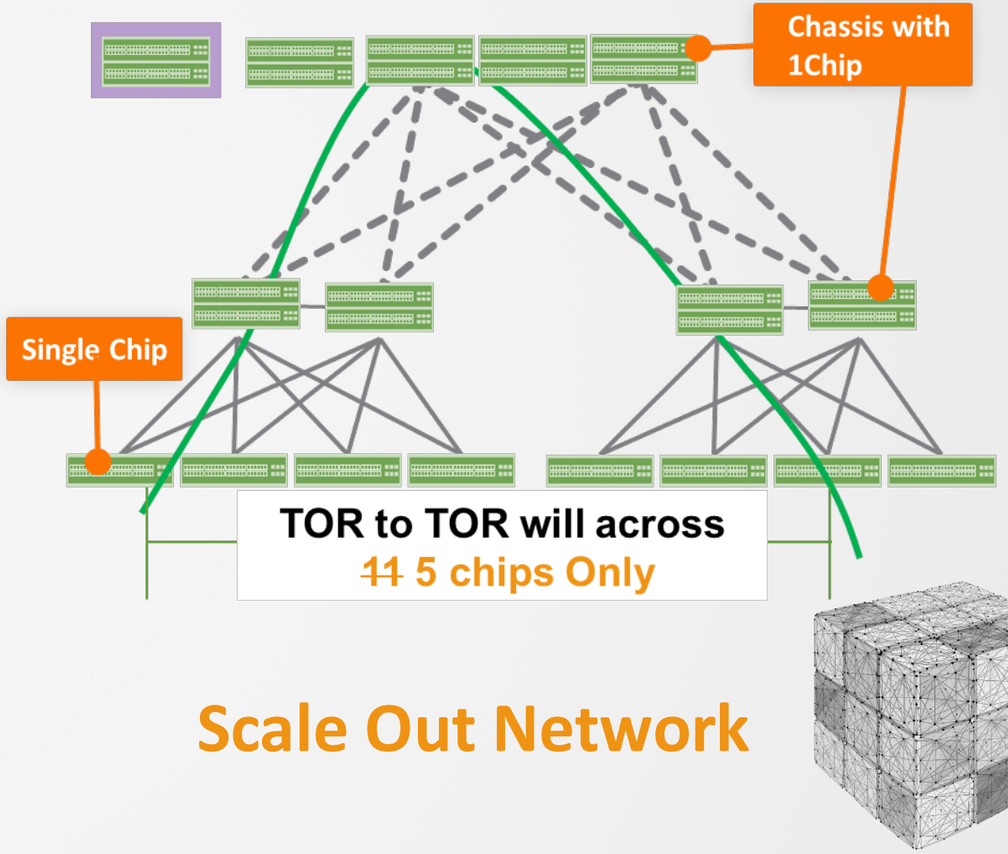


CLOS Switch

Scale Out DC architect with Single Chip Box



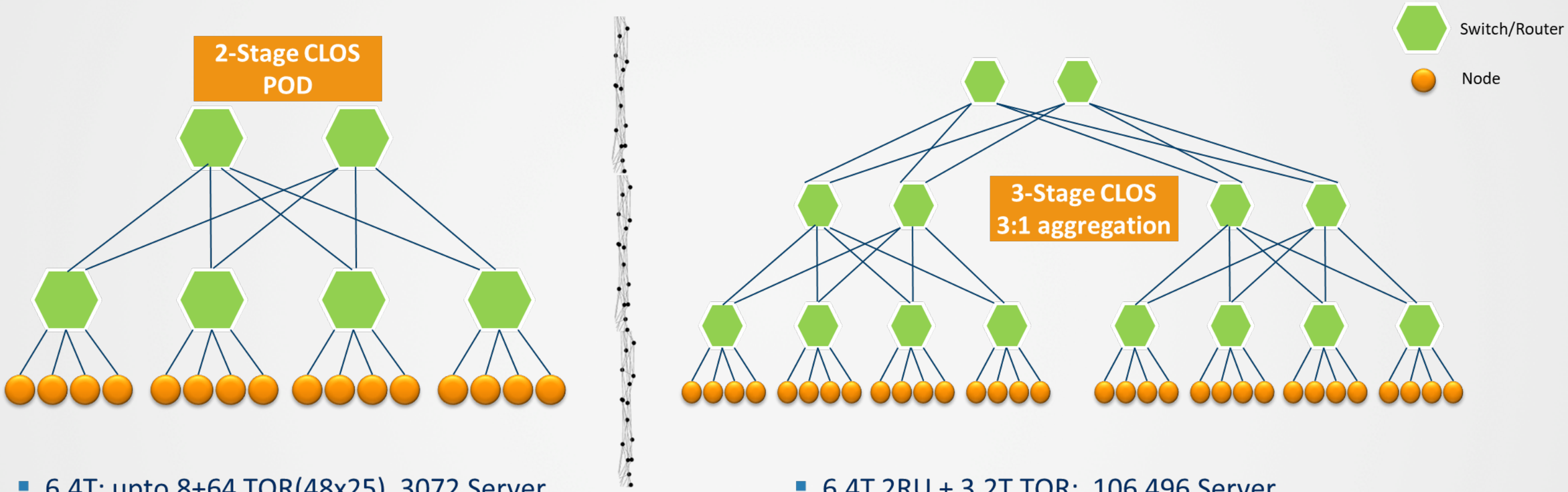
Scale Up Network



Scale Out Network

Better COST, Better Power Consumption, Better Scale Out

2/3 Stage CLOS with 6.4T/12.8T Chips



- 6.4T: upto 8+64 TOR(48x25), 3072 Server

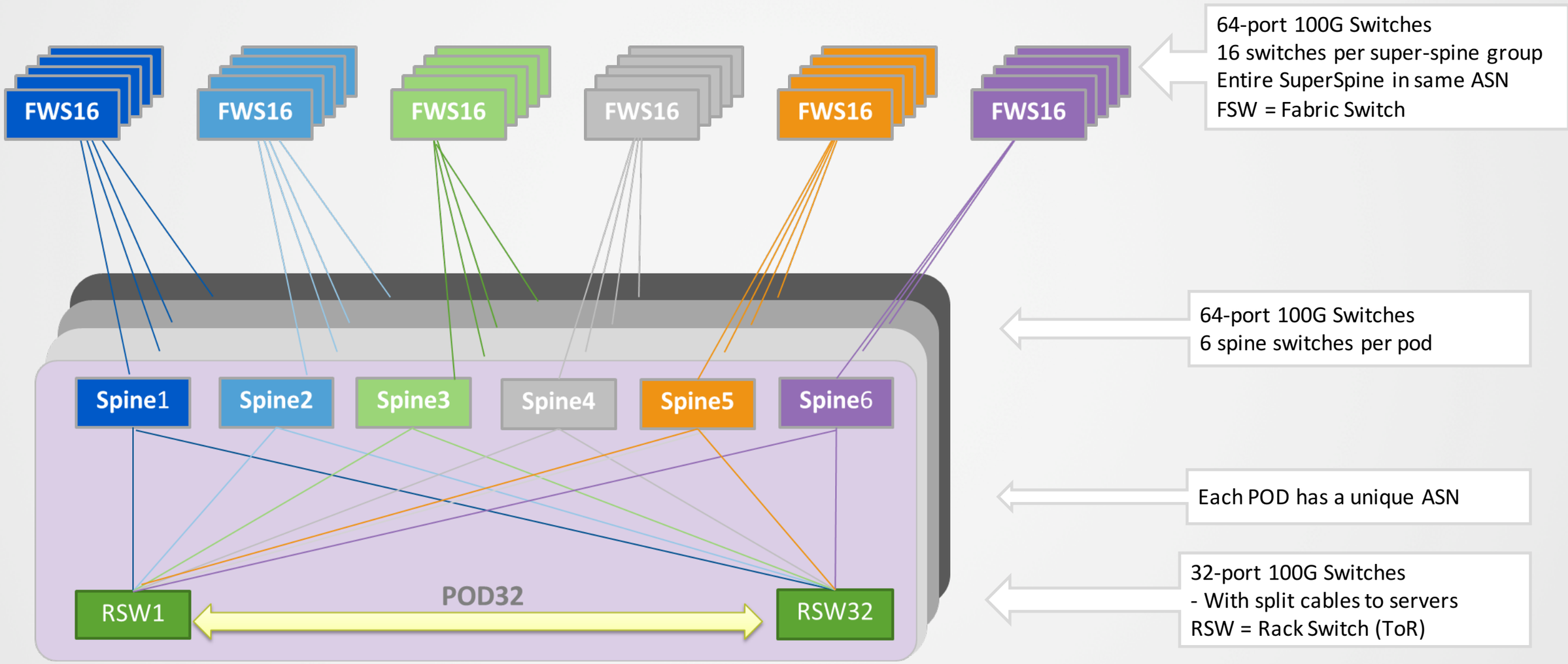
For Most Private DC/Cloud

- 6.4T 2RU + 3.2T TOR: 106,496 Server

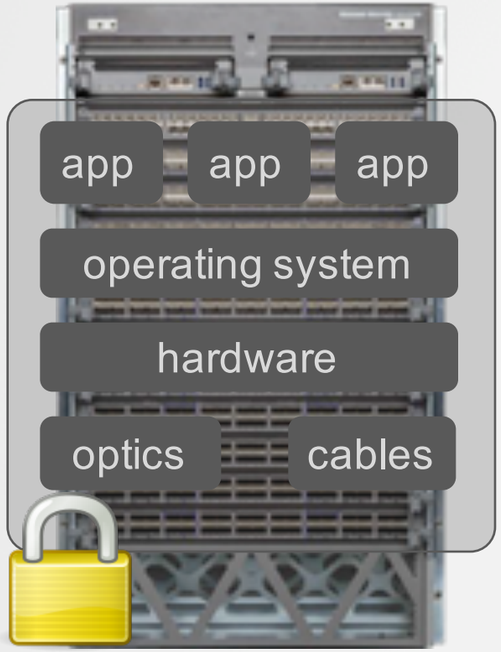
For Public Cloud

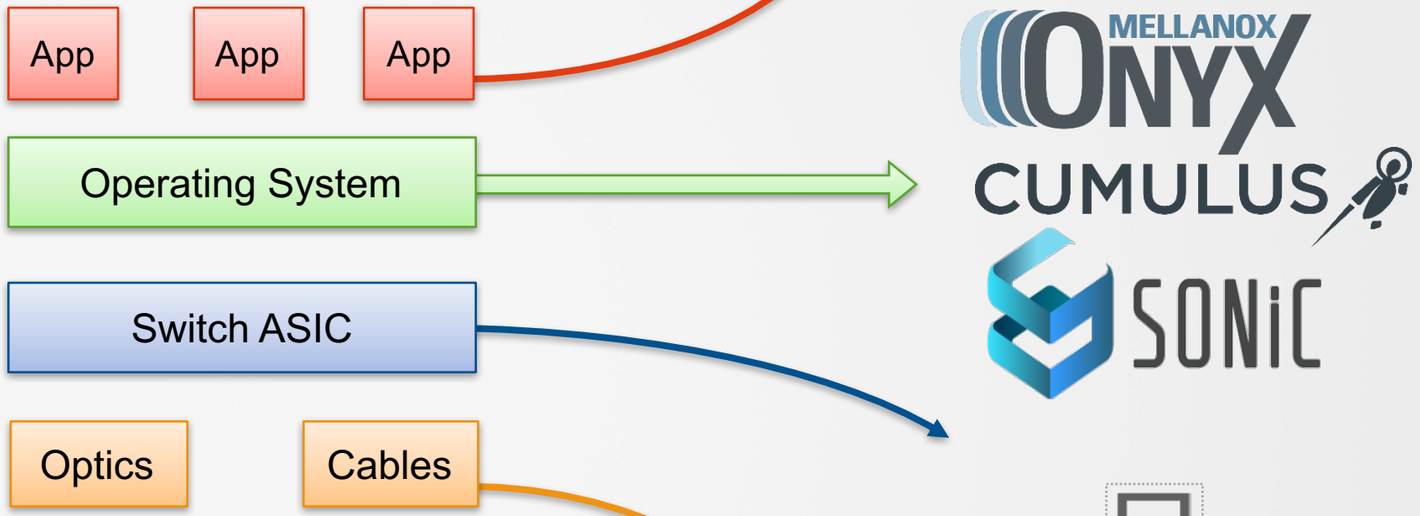
US Tier 2 Hyper scale Data Center Fabric

32 Pod x 32 TOR x 26 interface= 26,624 x100GE = 106,496 x 25G Server



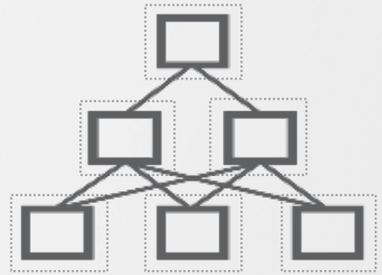
Network Disaggregation





Ganglia vmware NSX openstack CFEngine sFlow ANSIBLE WORKS collectd

MELLANOX ONYX CUMULUS SONiC



Leaf/Spine
CLOS Network

Mainframe-like Networks:

- Vendor lock-in
- Higher switch prices
- Higher support prices
- Slow feature development

Open Networking Platforms:

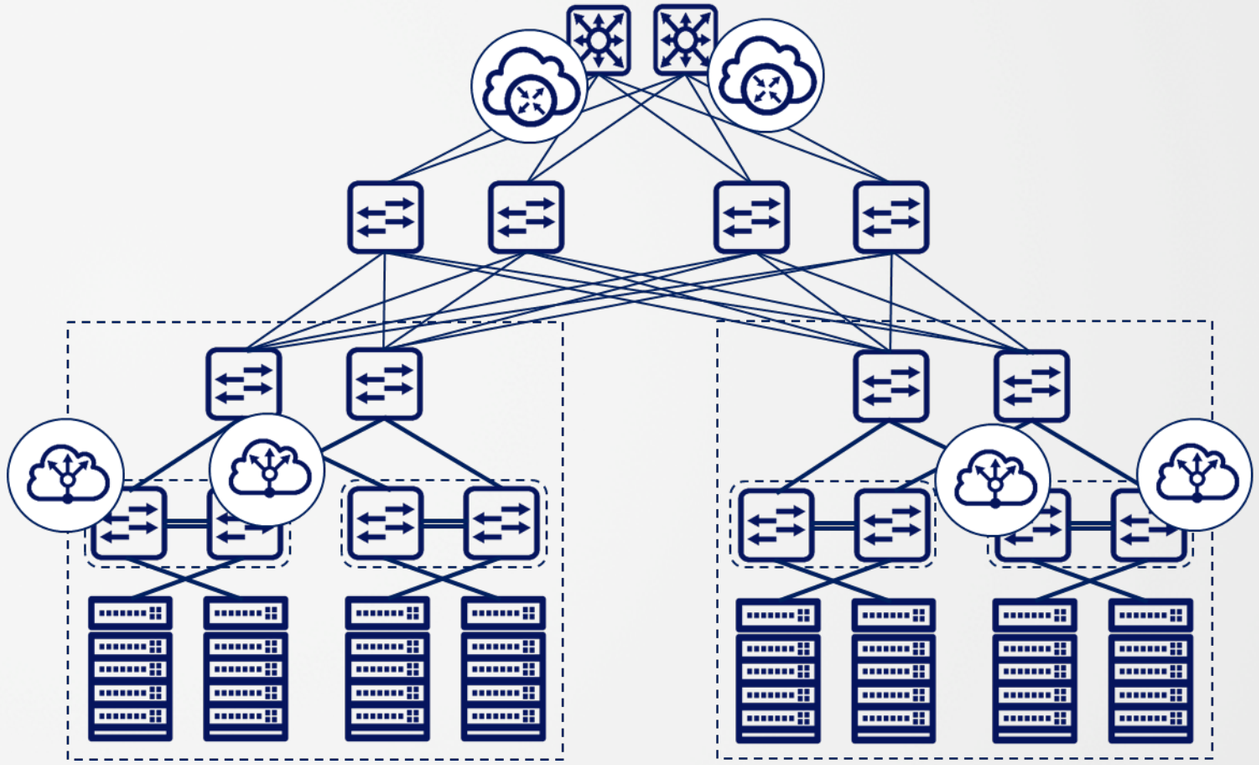
- Best of breed hardware
- Best of breed software
- Rapid deployment

Reliable Network Foundation - BGP Innovation

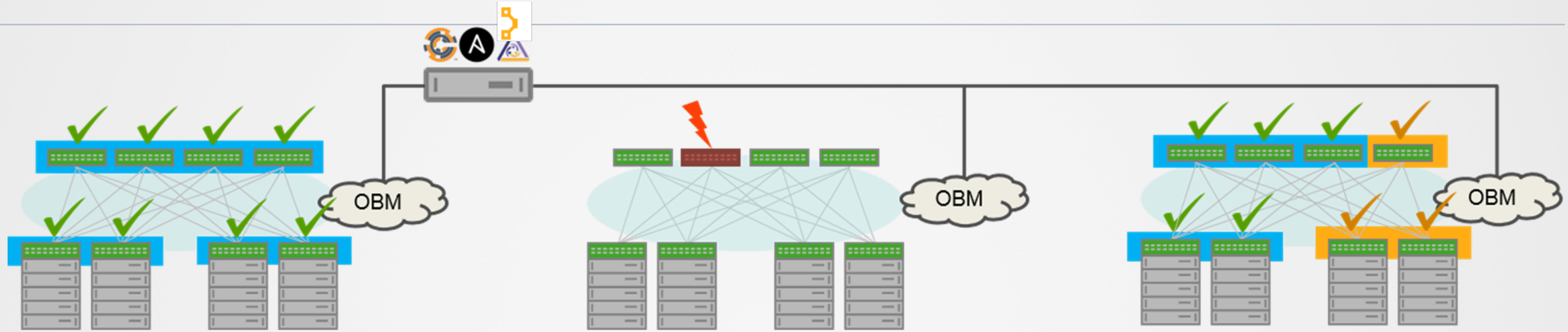
2

- BGP Based Datacenter
  - RFC 7938
- BGP Unnumbered/Linklocal
  - Easy. No /30 or /31 link IPs
  - Traceroute support
  - Standards based. RFC 5549
  - Unnumber/Link Local BGP session

```
router bgp 65000
 neighbor swp1 remote-as external
 neighbor swp2 remote-as external
 neighbor swp3 remote-as external
 neighbor swp4 remote-as external
 neighbor swp48 remote-as internal
```



Automation Use Cases: Infrastructure as Code



Rapid Provisioning

- Weeks and months now take seconds and minutes
- Pods of equipment can be stamped out in multiple locations

Hot Swap the Whole Switch

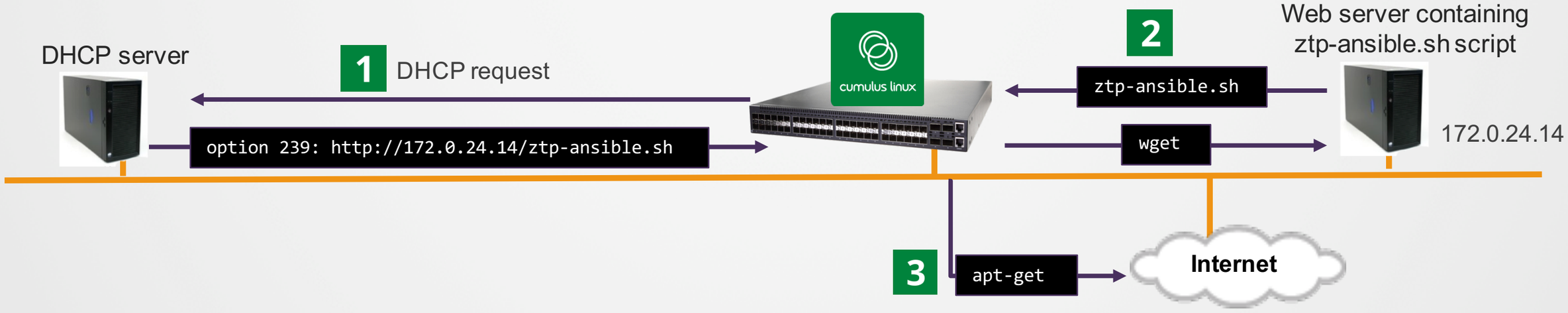
- The whole switch can be replaced and provisioned
- Configuration lives in Git instead of on the box

Configuration Management

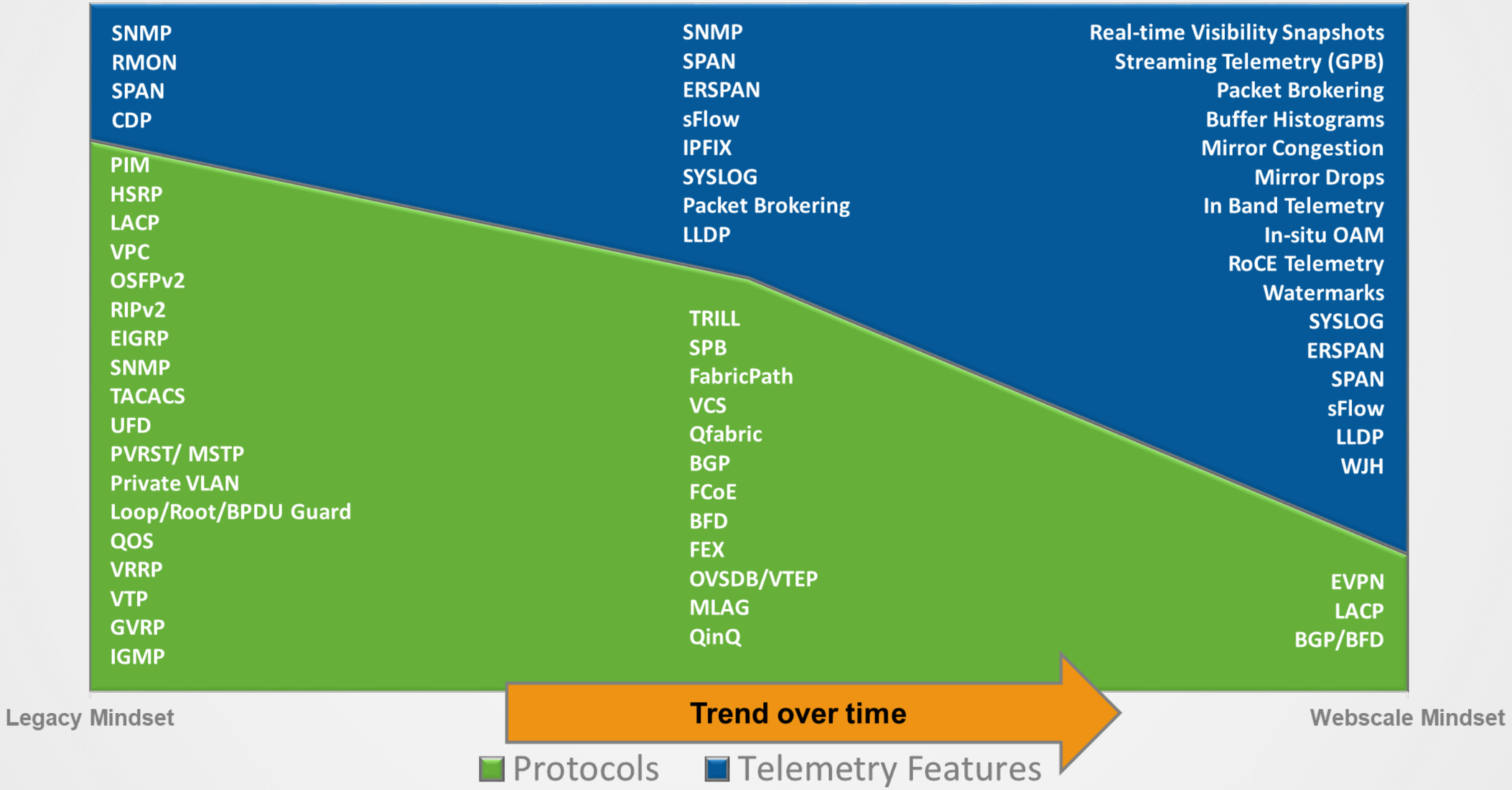
- Policy enforcement from central source of truth
  - Users
  - Permissions

Zero Touch Provisioning: Setting the Stage for Automation

1. Switch boots Cumulus with DHCP (default behavior)
   DHCP server responds with option 239 and URL location of ZTP script
2. Switch downloads ZTP script from specified webserver
   (Executes as root)
3. Script installs the SSH public keys (Ansible)



Best in Class Telemetry



WJH – How Does It Work?

1. SDK generates:
WJH messages

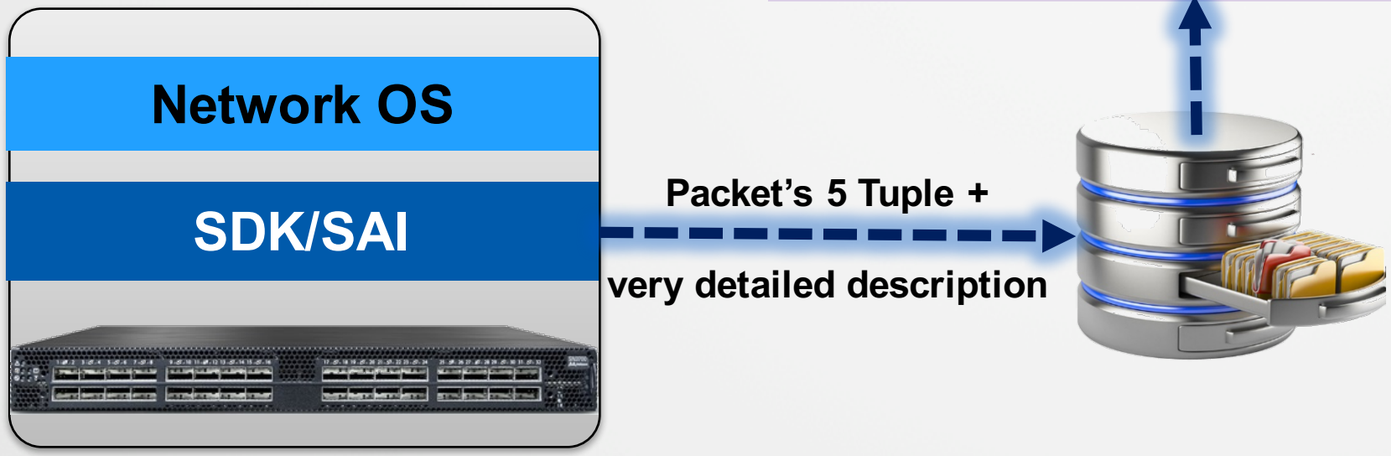
2. Agent collects the data:
Streams to a Database

3. Presentation layer shows:
What Just Happened

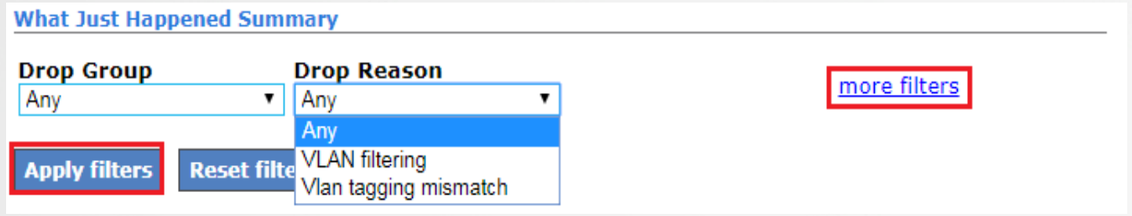
The Important Questions

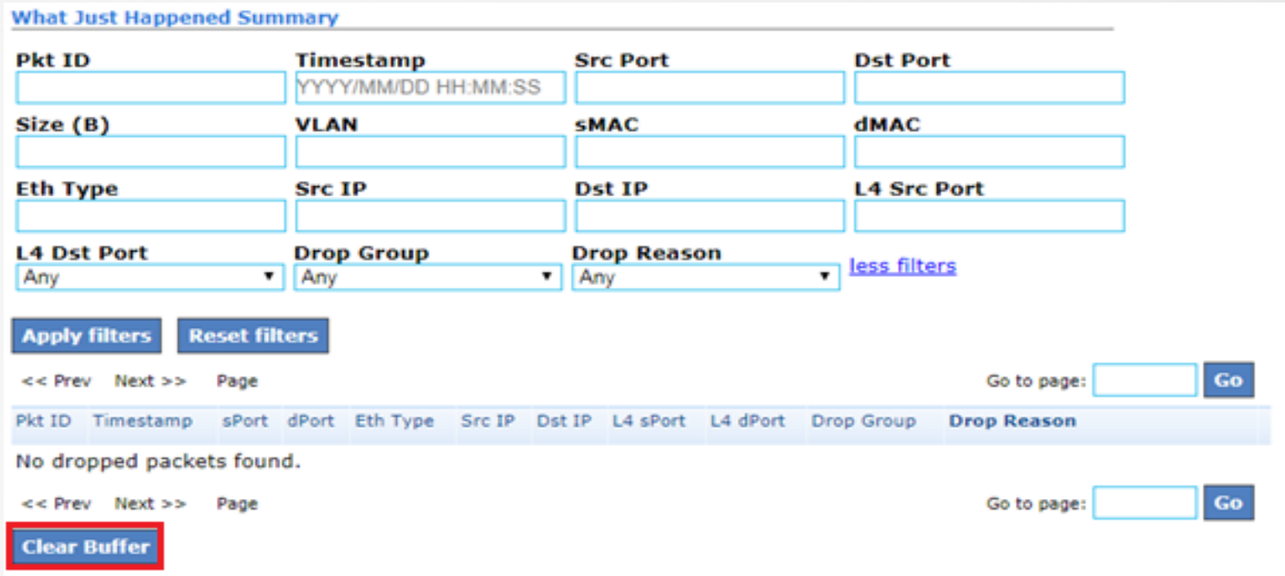
- ✓ WHO is being impacted
- ✓ WHEN it happened
- ✓ WHAT is causing the problem
- ✓ WHERE is the problem
- ✓ WHY it is happening

Root Cause + how to fix it



WJH packets





<< Prev Next >> Page [1] 2 Go to page: Go

Pkt ID	Timestamp	sPort	dPort	Eth Type	Src IP	Dst IP	L4 sPort	L4 dPort	Drop Group	Drop Reason
1	2019/01/08 14:24:38.997	eth1/9	N/A	IPv4	1.1.100.100	1.1.100.200	60	80 (http)	Forwarding	**Vlan tagging mismatch**
2	2019/01/08 14:24:38.997	eth1/9	N/A	IPv4	2.2.2.2	2.2.2.1	60	22 (ssh)	Forwarding	**VLAN filtering**
3	2019/01/08 14:24:38.997	eth1/9	N/A	IPv4	1.1.100.100	1.1.100.200	60	80 (http)	Forwarding	**Vlan tagging mismatch**
4	2019/01/08 14:24:38.997	eth1/9	N/A	IPv4	2.2.2.2	2.2.2.1	60	22 (ssh)	Forwarding	**VLAN filtering**
5	2019/01/08 14:24:38.997	eth1/9	N/A	IPv4	1.1.100.100	1.1.100.200	60	80 (http)	Forwarding	**Vlan tagging mismatch**

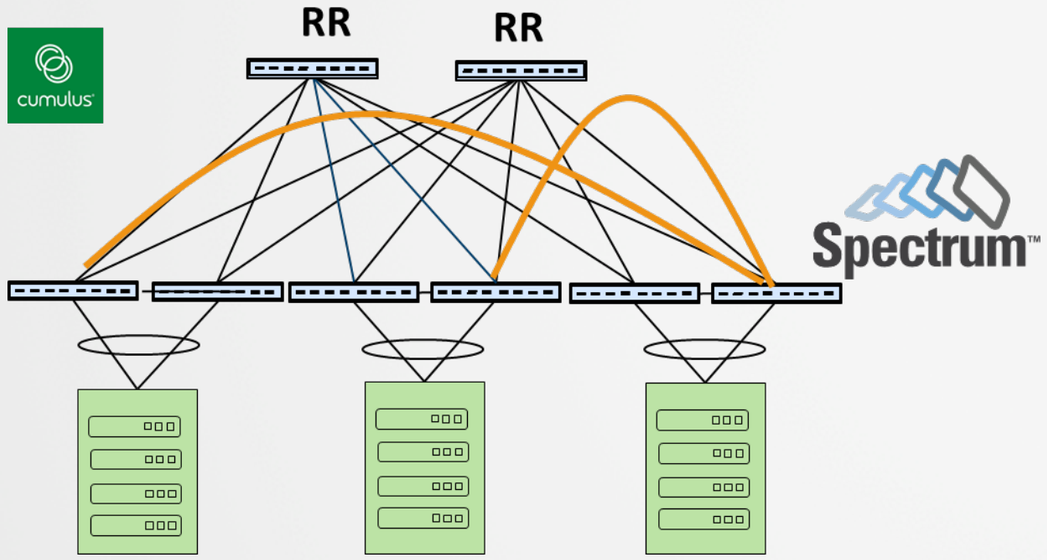
Regenerating Issues is Painful

With WJH

Without WJH

Controller-less SDN With Cumulus & Spectrum

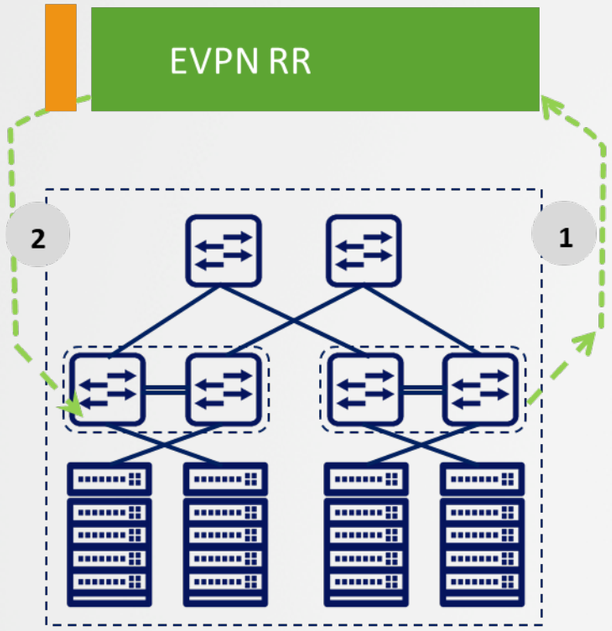
BGP-EVPN+VxLAN



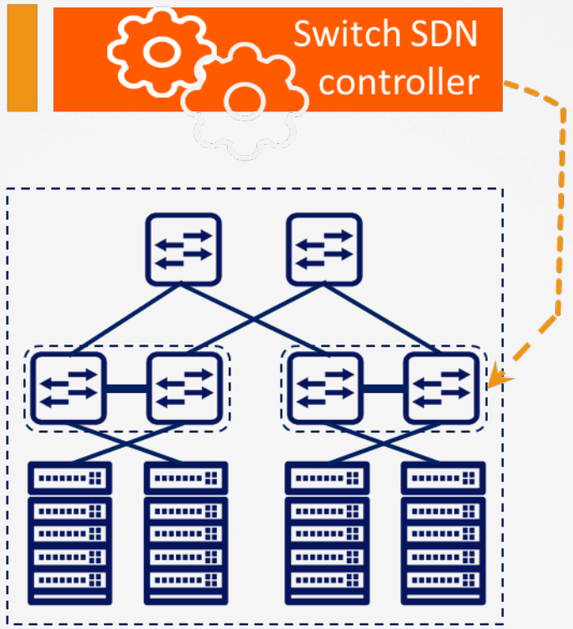
- Same protocol (BGP) that drives the Internet extended to support VXLAN overlays
- RFC based standard
  - No vendor lock-in
  - No hypervisor restriction
- Controller less orchestration of tunnels
  - Better horizontal scale
  - No controller license $$$s
- Multi-tenancy and workload flexibility
  - Tenant isolation
  - Any workload and network on any leaf

Open Standards	Scale Out Architecture	Physical Layer Abstraction	Total Host & IP Mobility

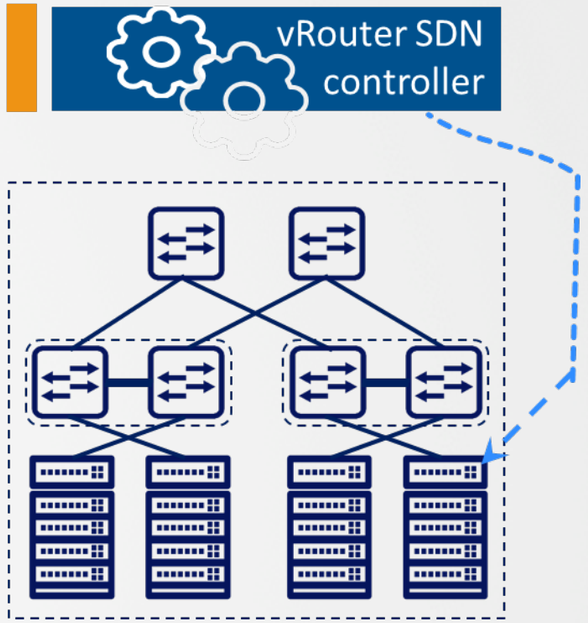
Controller-less vs Controller SDN



- Prefix/MAC Advertise by BGP
- Any Vendor Router/Switch can support
- Standard BGP EVPN Message
- Work for Host and Switch
- Very big scale, Very easy Inter-DC



- Prefix/MAC Advertise by Switch SDN
- Cisco/BigSwitch or Netconf etc
- Proprietary Message, not easier for multi-vendor.
- Dedicate algorithm Propagate Prefix/Mac
- Smaller scale, Not easy Inter-DC

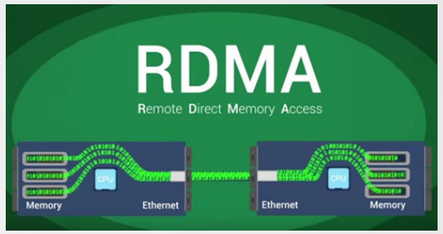


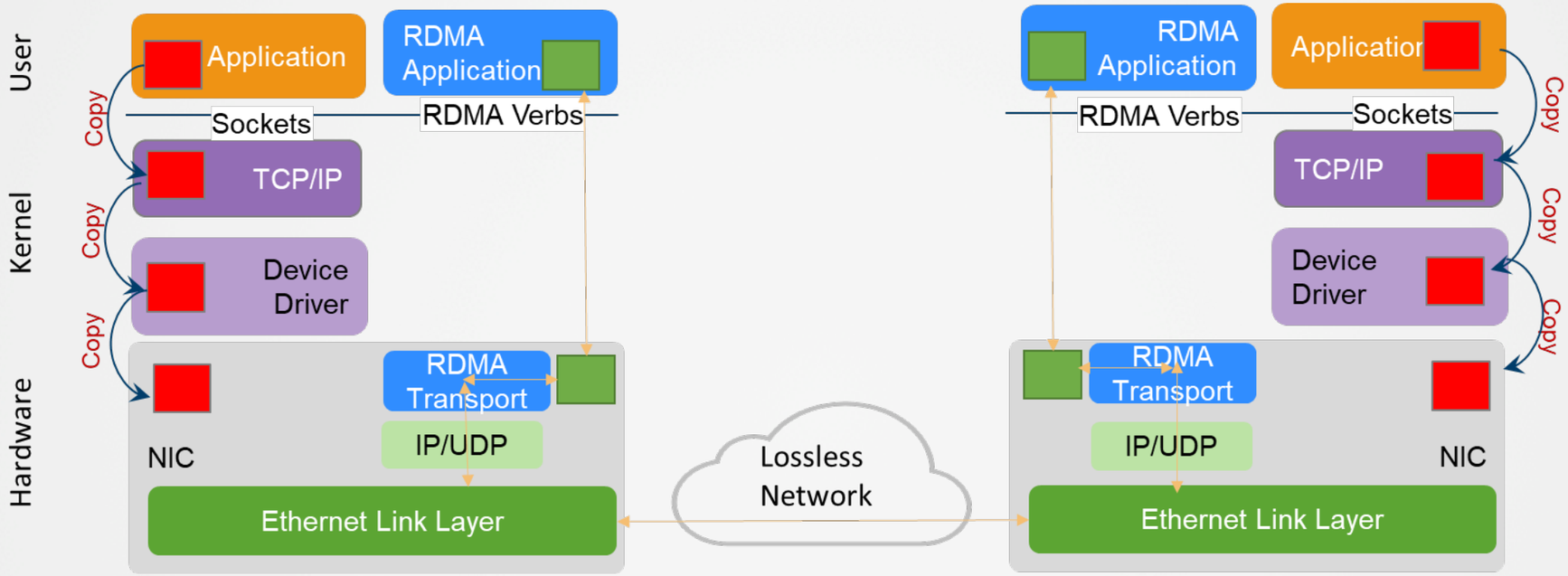
- Prefix/MAC Advertise by vRouter SDN
- Vmware NSX/Nuage/Contrail
- Proprietary Message, not easier for multi-vendor.
- Controller simulate a RR.
- Smaller scale, Not easy Inter-DC

RDMA In Cloud

- Enable RDMA applications to run on cloud
  - Scientific
  - HPC
  - Machine Learning and AI
  - Data bases
- Basis for DPDK applications
  - Telco and NFV
- Accelerate cloud infrastructure
  - VM migration over RDMA
  - Message queue over RDMA (e.g. gRPC)
- Accelerate cloud storage
  - iSER
  - NVMf
  - Ceph over RDMA

RDMA Zero buffer copy





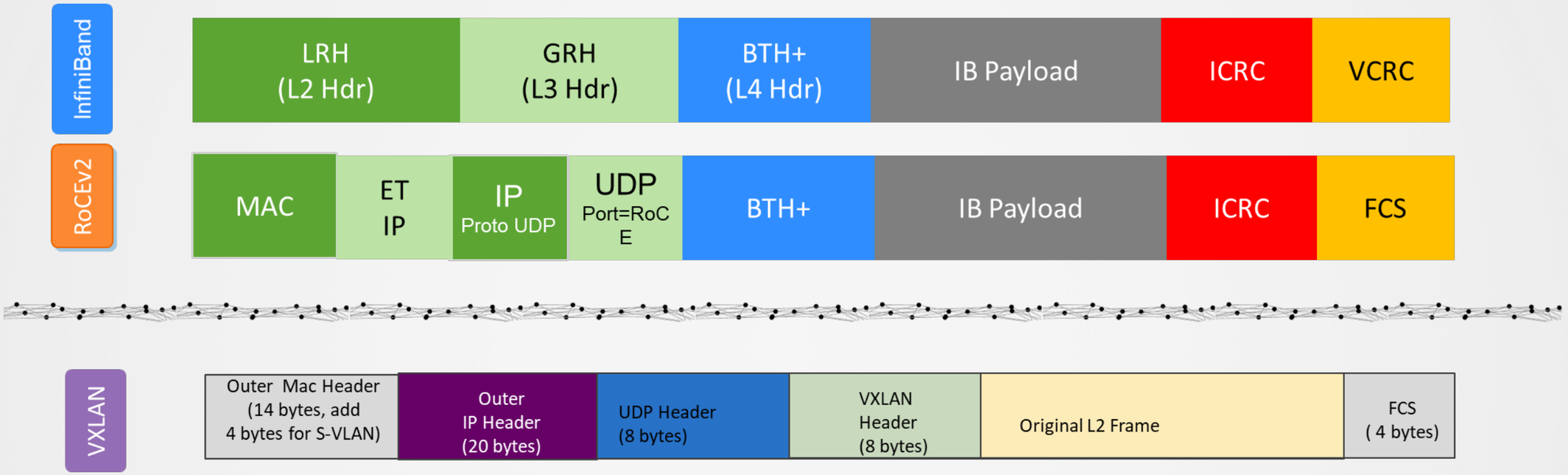
What is RDMA?

- Direct Memory Access from Memory of one computer to another without involving OS
- Transport for Compute-Storage and Compute and Compute
- Bypasses OS and TCP/IP stack, saves CPU cycles
- Low Latency , high throughput and low CPU utilization transport.

Why?

- CPU is an expensive elements in the Data Center, its utilization should be maximized
- Real time applications require low latency for consistence response
- The move to SSD has made Latency a factor in storage

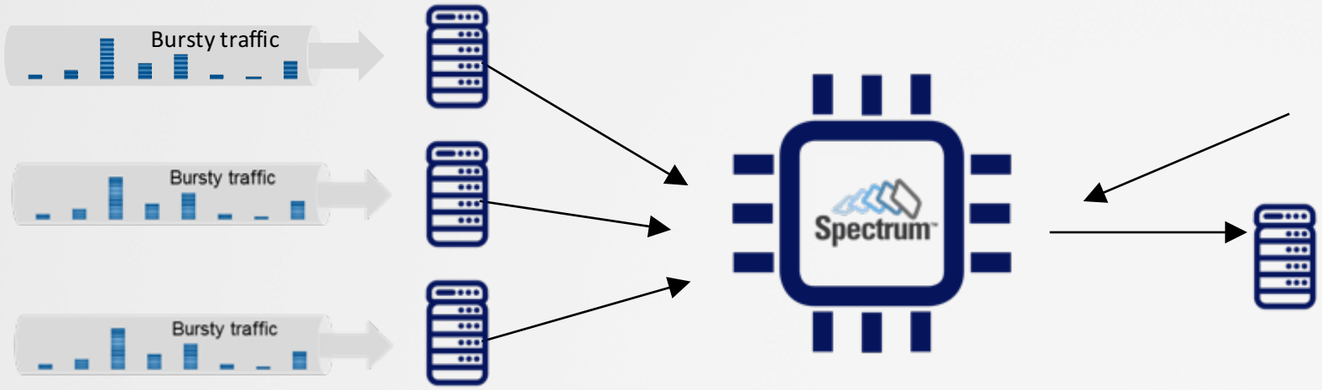
RoCEv2 vs. VXLAN



- RoCE is open source and a formal InfiniBand Trade Association (IBTA) standard.
- The original implementation of RoCE, known as version 1 does not span across IP subnets. RoCE version 2 enables communication across IP subnets.
- { RoCEv2: IB over UDP } | { VXLAN: MAC over UDP }

Fully Shared Buffer is Superior

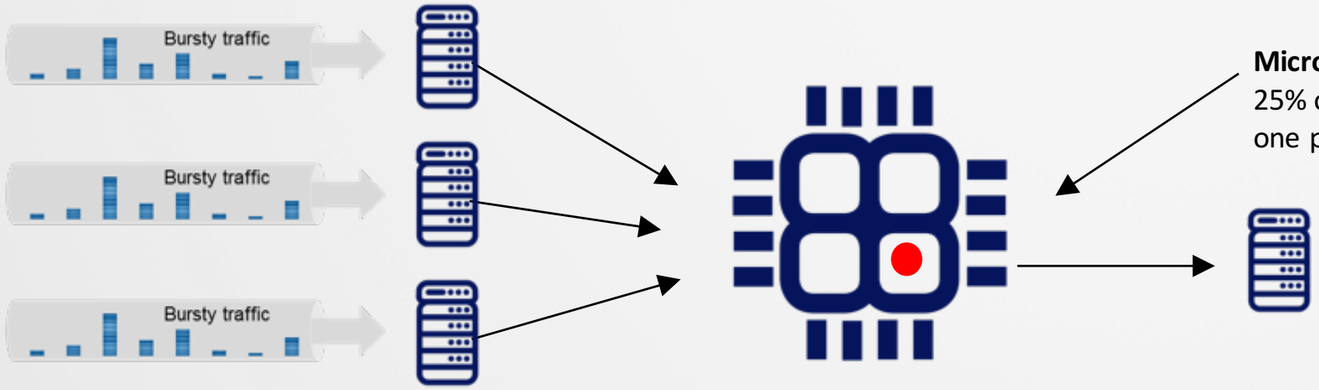
Fully Shared Packet Buffer



Microburst occurs here with 16MB of packet buffers to absorb burst

Superior Micro Burst Performance Spectrum's Fully Shared Buffer Provides **4X effective buffer size**!

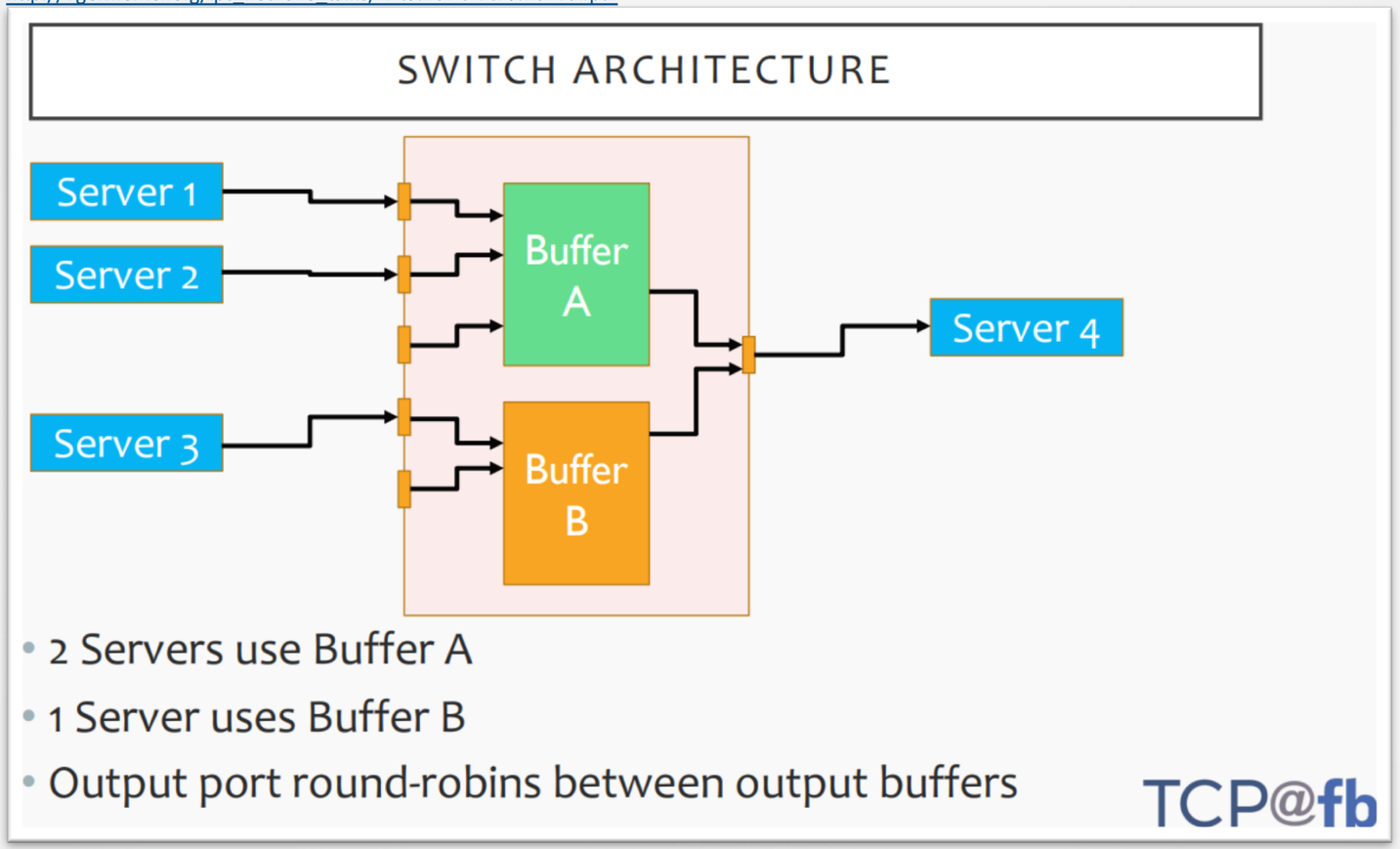
Competitor's multi-core based buffer scheme



Microburst occurs here, but only 25% of buffers are available to any one port

Facebook Found Fairness Failures

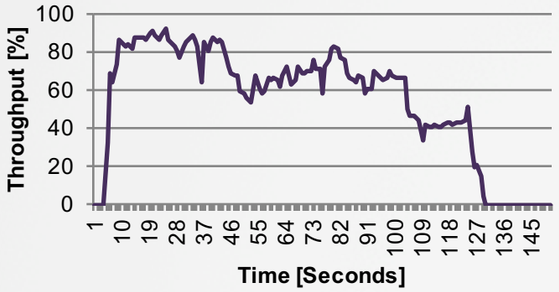
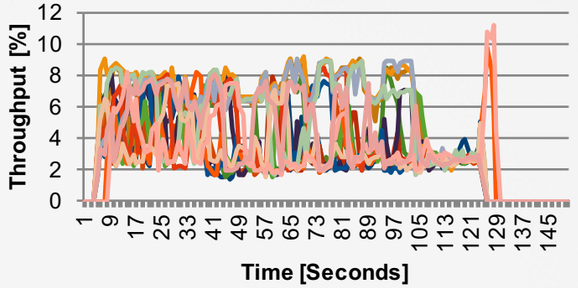
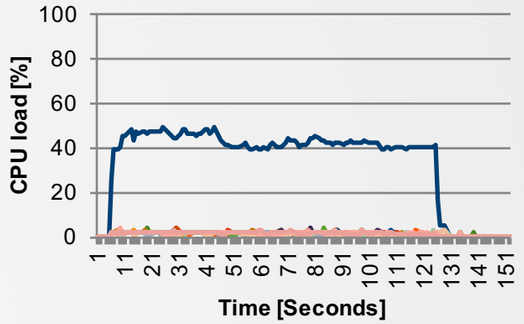
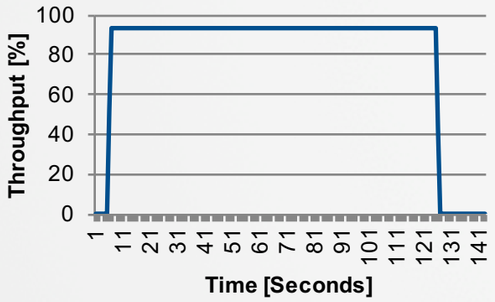
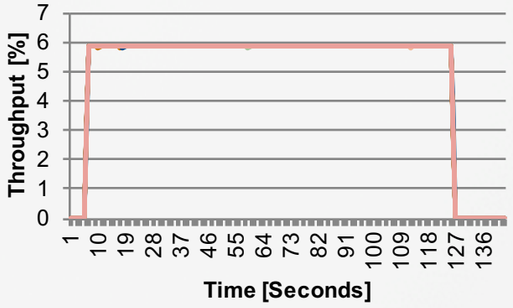
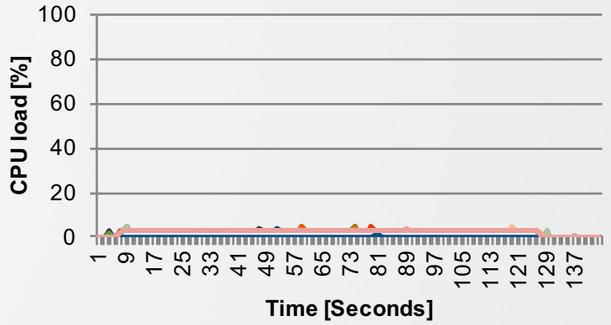
http://vger.kernel.org/lpc_net2018_talks/LPC%20DC-TCP%20Eval.pdf



Issue: Unfairness between flows

- With only 2 flows, one flow would get much higher link utilization: 23Gbps vs. 0.5 Gbps

RoCE with ECN-only vs. TCP

	Total Throughput	Throughput per Sender	CPU Utilization
TCP			
RoCE™			

2X Better Throughput

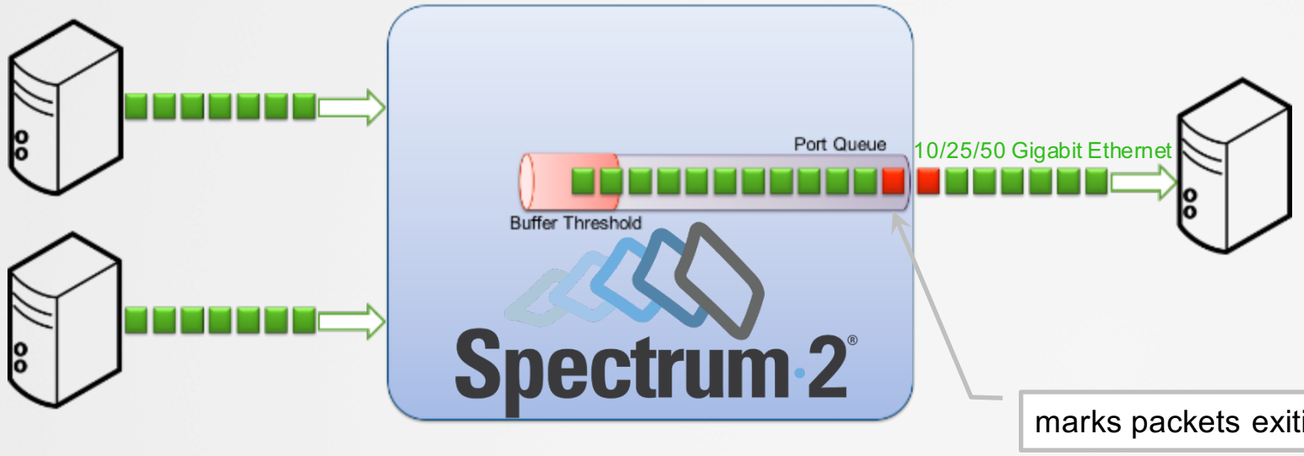
Better Fairness and Consistency

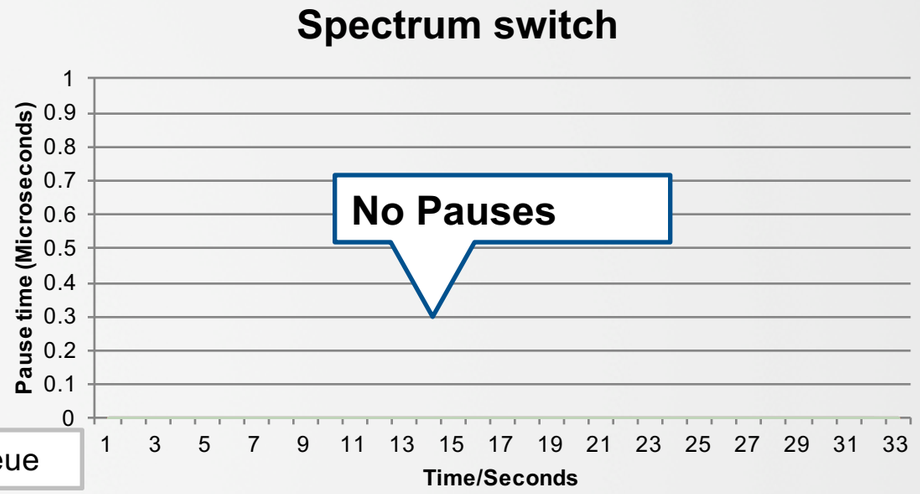
Close to 0% CPU Utilization!

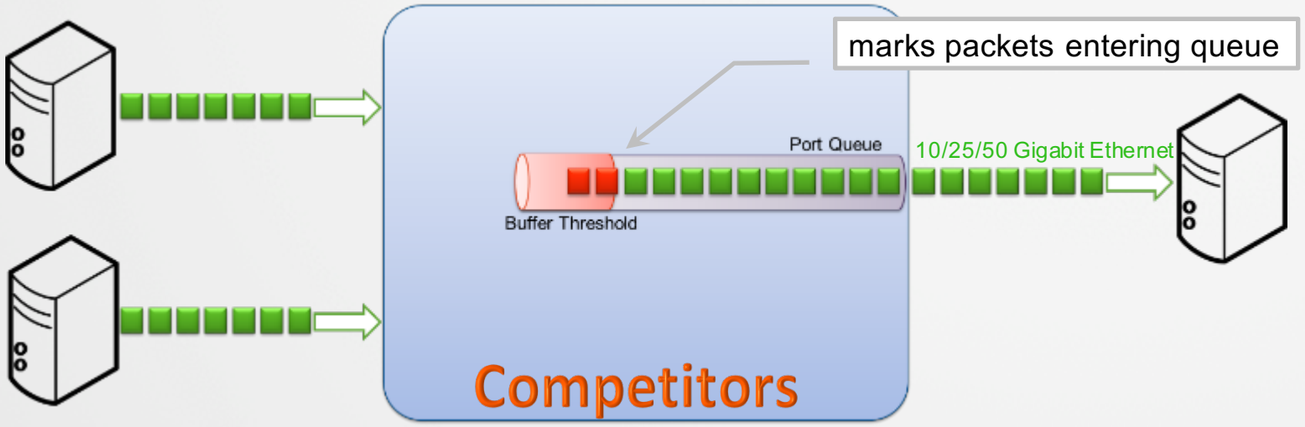
16 hosts to 1 host, 64 QPs per Sender

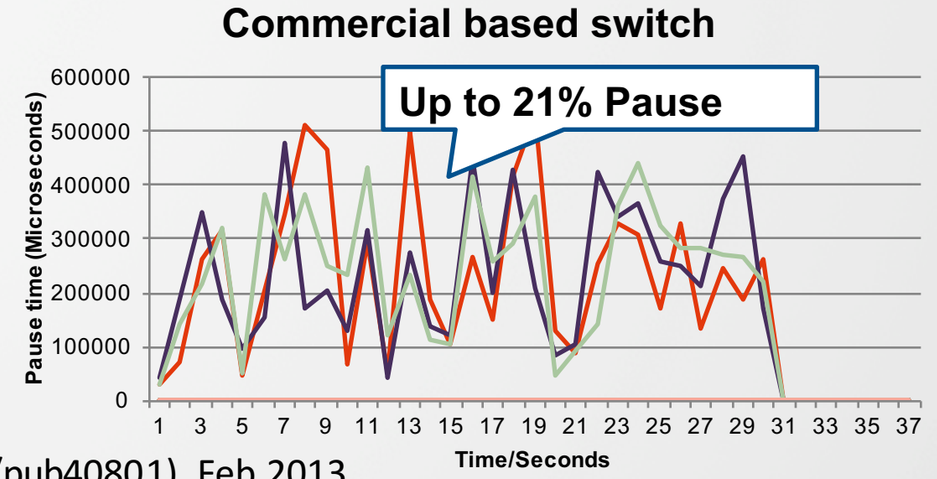
Increasing RoCE Performance

Fast Congestion Notification ECN





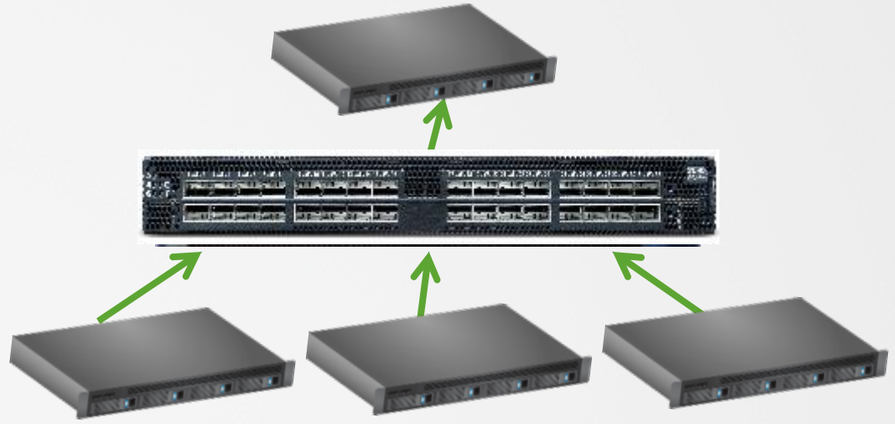




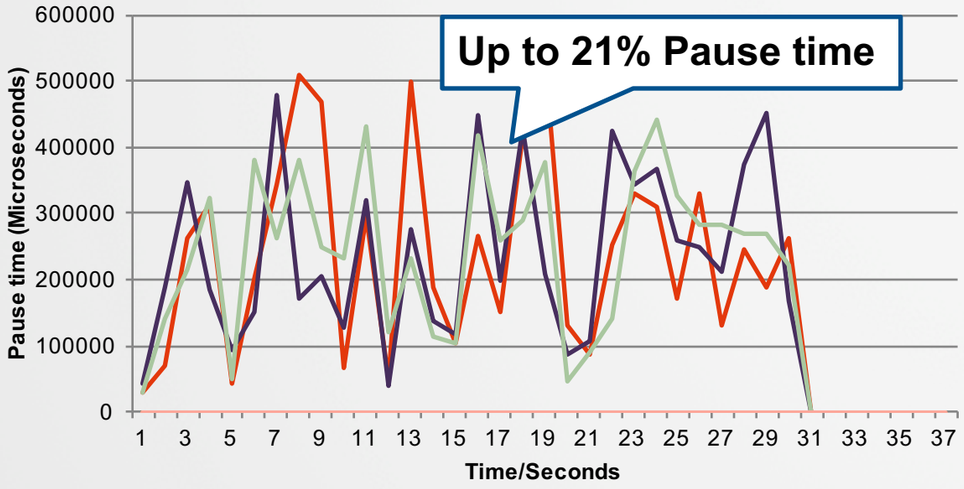
Chip design Based on Google paper "Tail at Scale" (https://ai.google/research/pubs/pub40801) Feb 2013

Best Congestion Management For RoCE

- Configuration
  - 4 hosts connected to 1 switch in a star topology
  - ECN enabled, PFC enabled
  - 3 sources to 1 common destination
- Results
  - Tomahawk sends pauses to hosts, no pauses sent by Spectrum

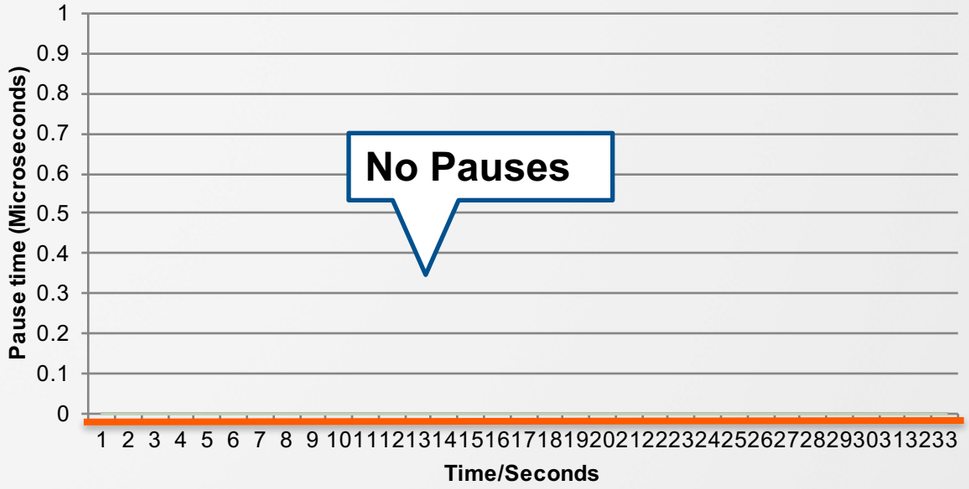


Tomahawk based switch



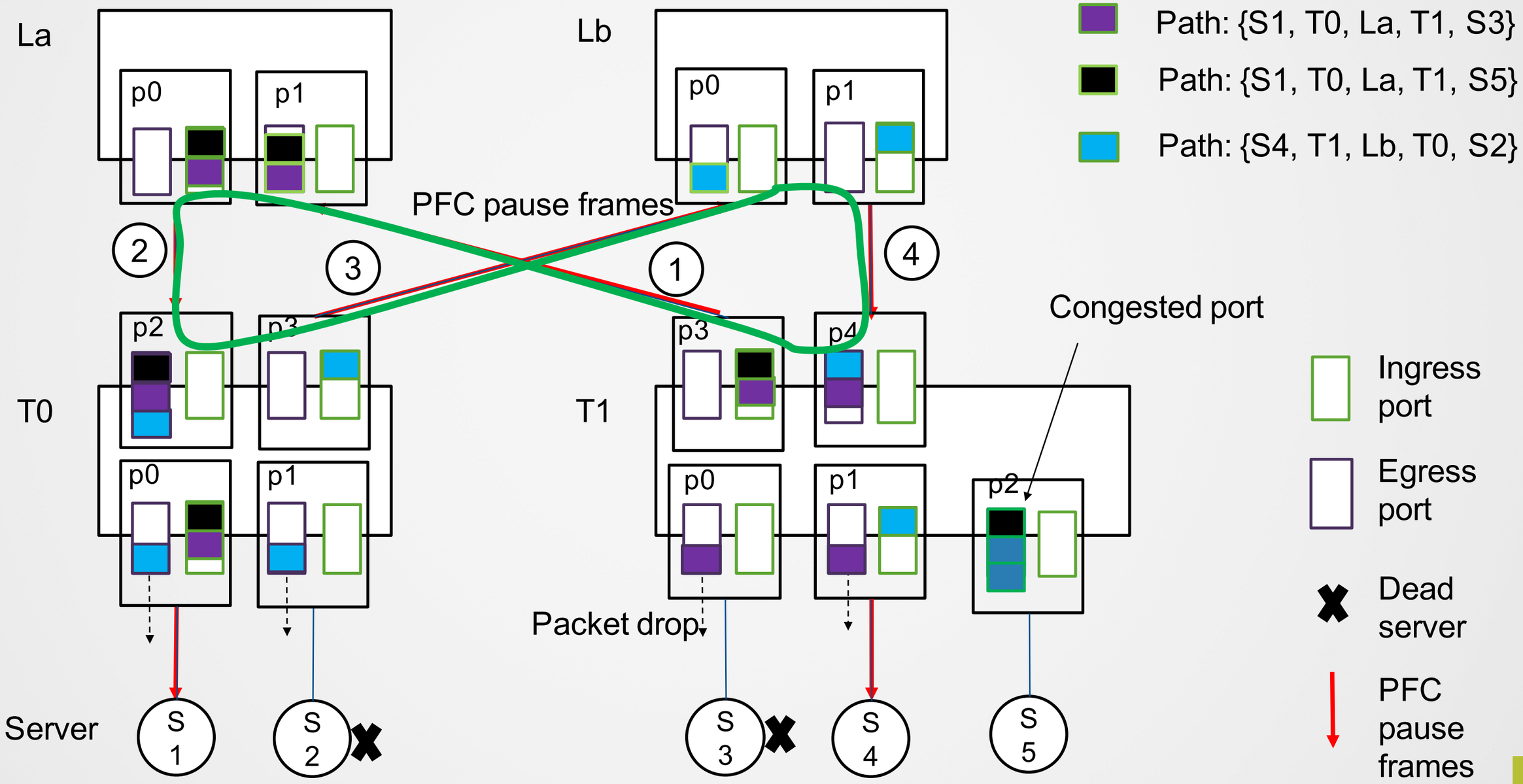
VS

Spectrum switch

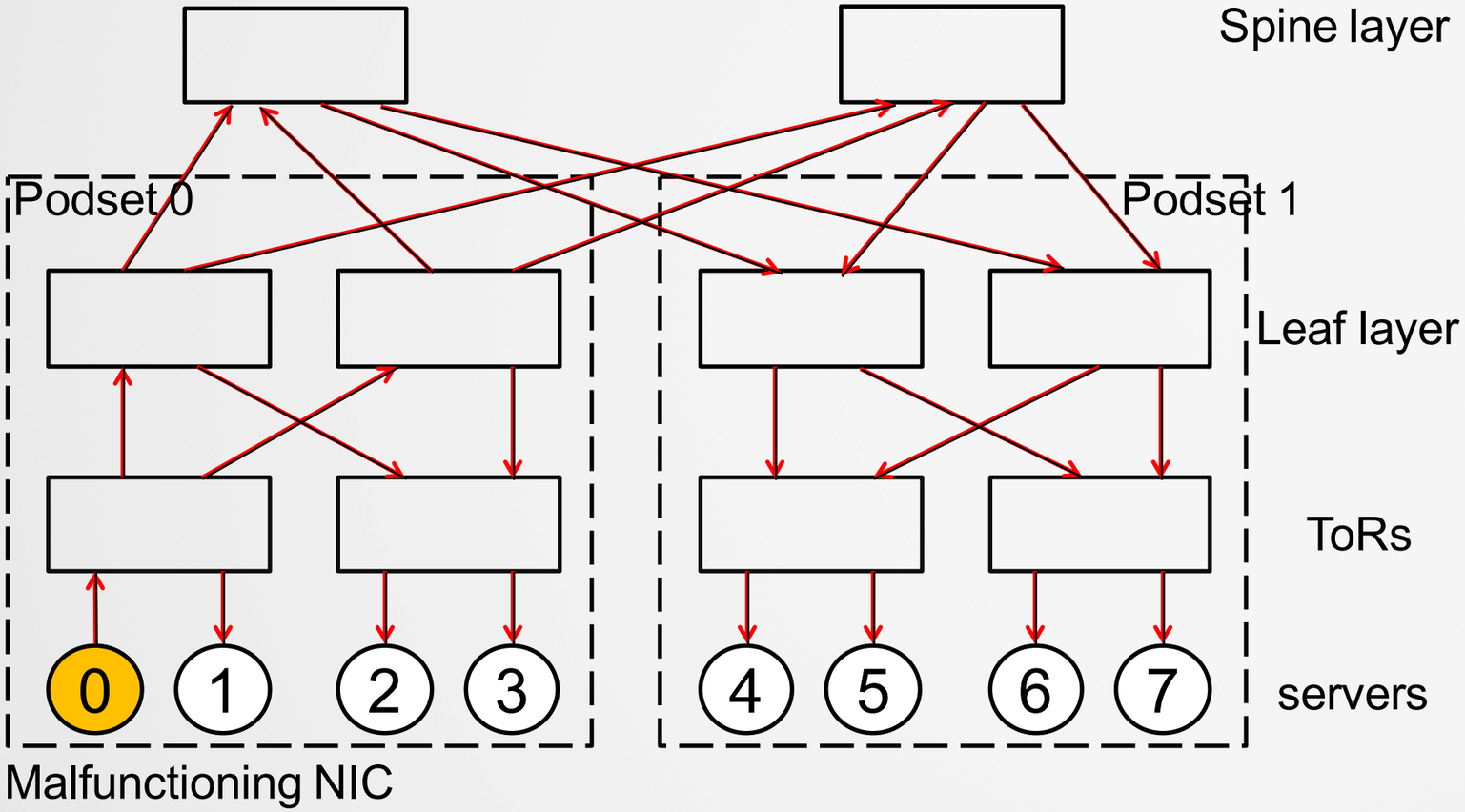


Q: Why should I care about congestion management?
A: Poor congestion management creates pauses which then result in bandwidth degradation

RoCEv2 PFC deadlock Issue

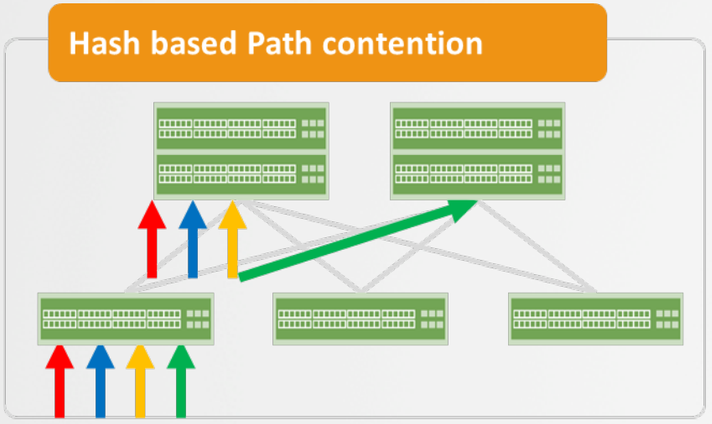


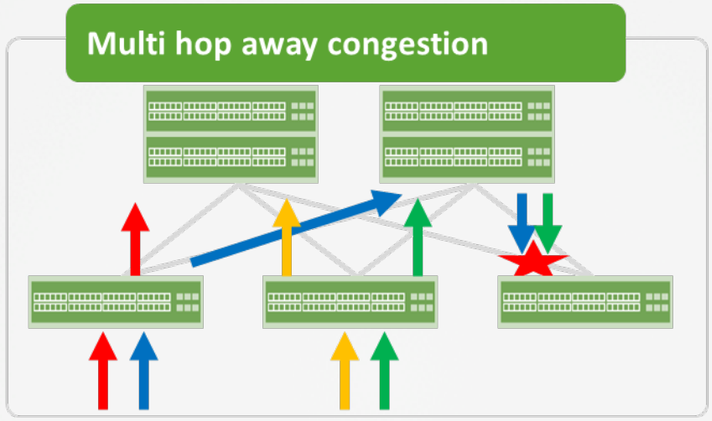
NIC PFC pause frame Storm

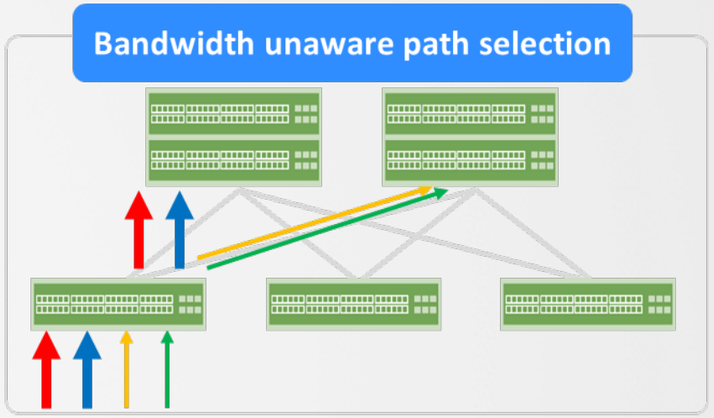


- A malfunctioning NIC may block the whole network
- PFC pause frame storms caused several incidents
- Solution: watchdogs at both NIC and switch sides to stop the storm

Adaptive Routing and Notification







FlowBender: Flow-level Adaptive Routing for Improved Latency and Throughput in Datacenter Networks

Abdul Kabbani
Google Inc.
Mountain View, CA, USA
akabbani@google.com

Balajee Vamanan
Purdue University
West Lafayette, IN, USA
bvamanan@ecn.purdue.edu

Jahangir Hasan
Google Inc.
Mountain View, CA, USA
jahangir@google.com

Fabien Duchene
Universite catholique de Louvain
Louvain-La-Neuve, Belgium
fabien.duchene@uclouvain.be

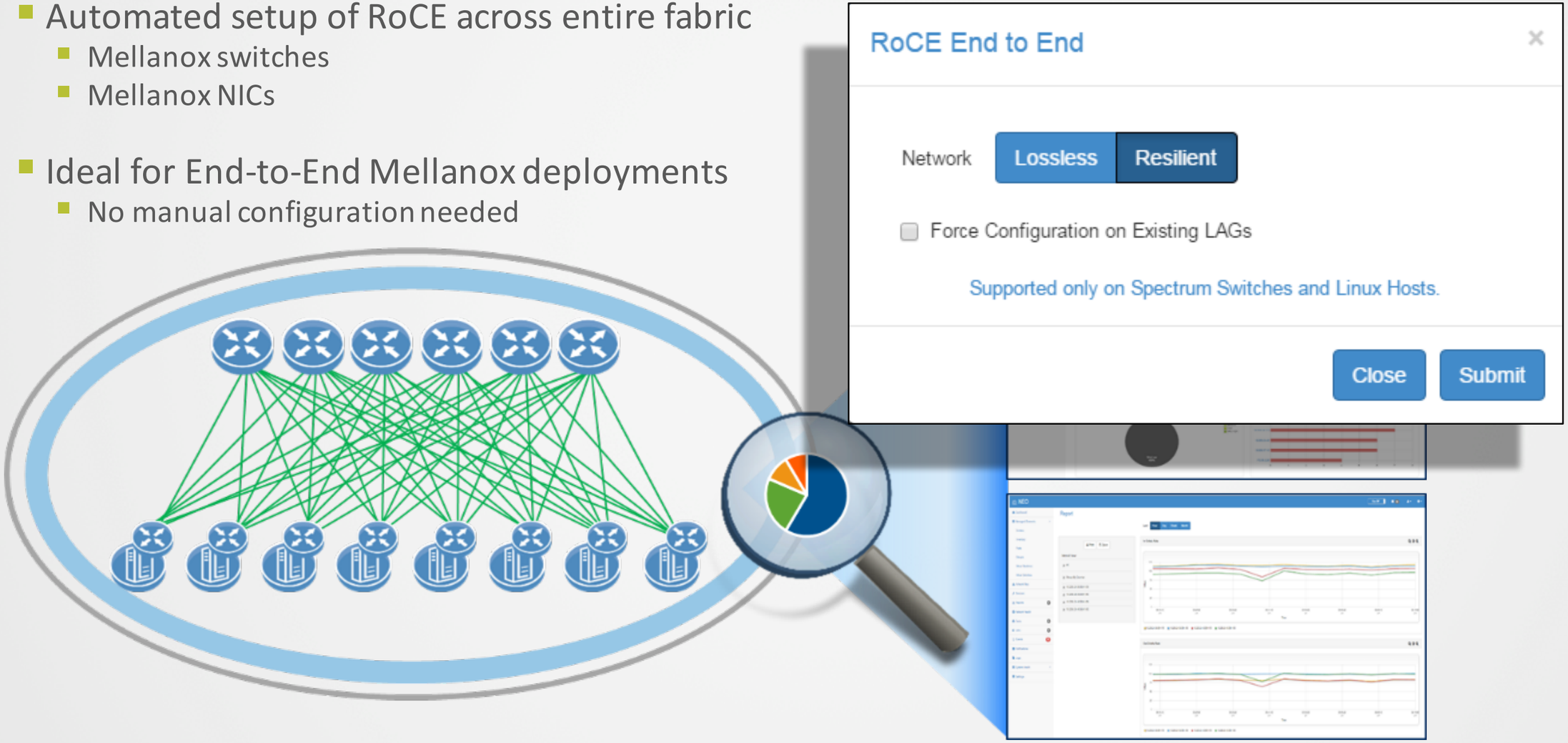
Dec 2014, Google

AR challenges:

- Prevent packet reorder
- Identify the elephant flows
- Build in New Chips

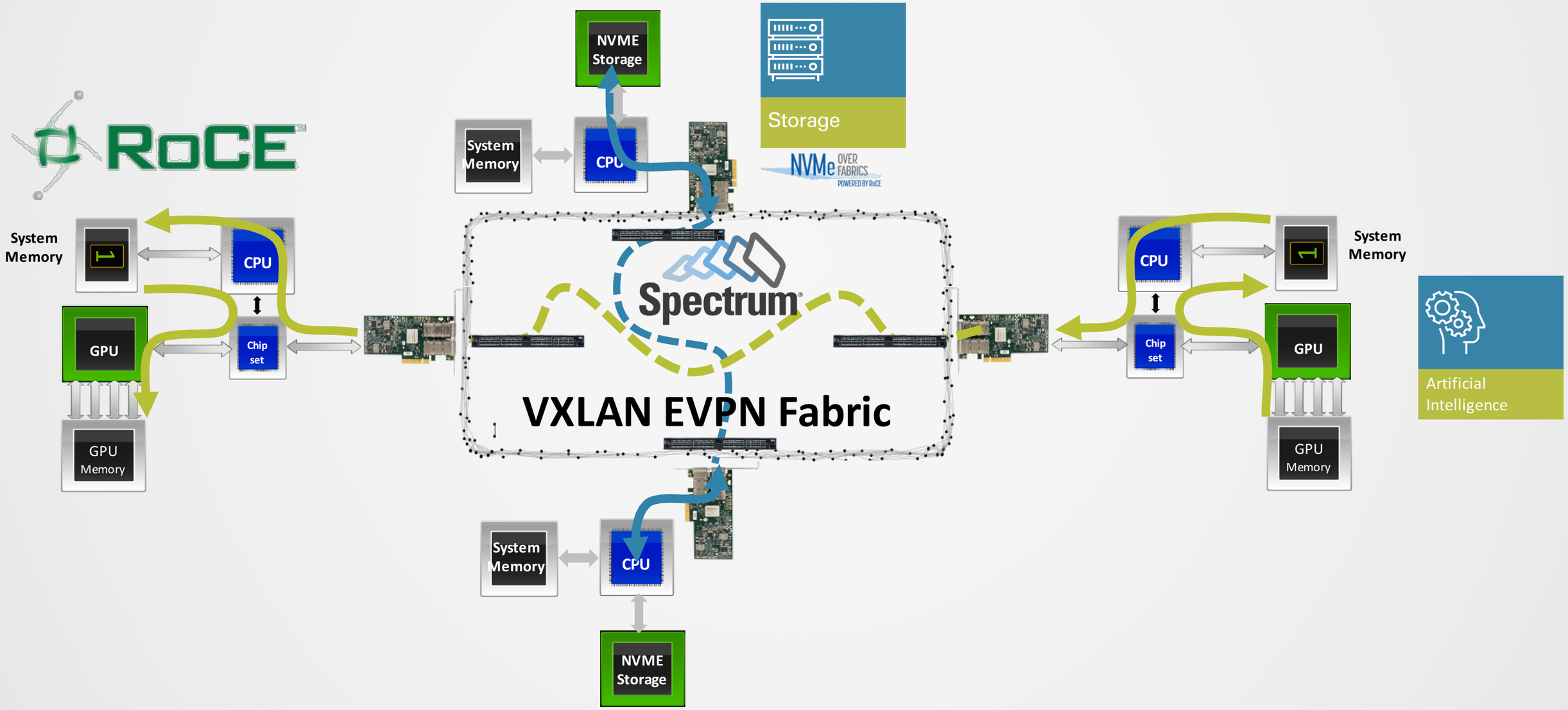
NEO Simplifies RoCE Provisioning

- Automated setup of RoCE across entire fabric
  - Mellanox switches
  - Mellanox NICs
- Ideal for End-to-End Mellanox deployments
  - No manual configuration needed



High Performance Cloud

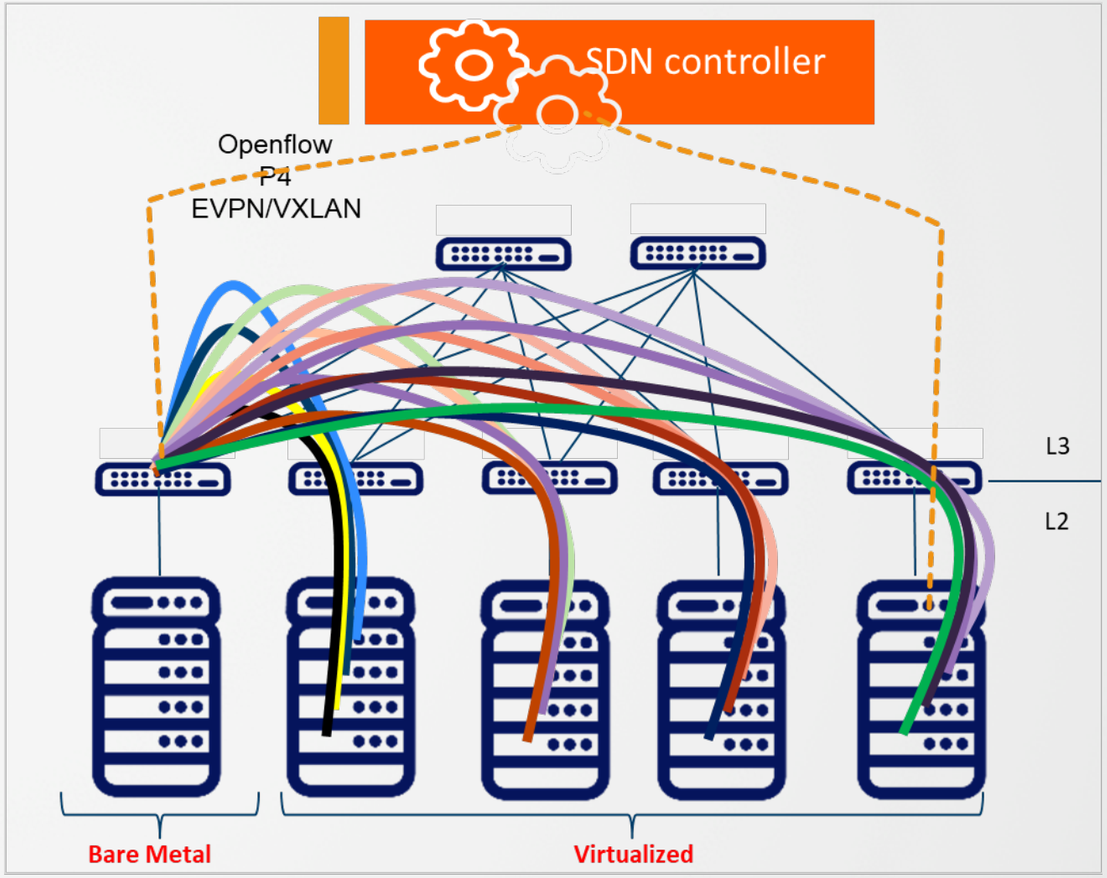
RoCE-ready VXLAN with GPU direct and NVMeoF



Practically Unlimited Hardware

MAC/IP/VXLAN Tunnel Scale, and Fast Program

- Program Hardware like Software
  - high speed with ATCAM, Bloom filter in Hardware table. Target for 30K+ entry/second
  - Much better performance 3.2T/6.4T/12.8T+
- Massive Scale for MAC/IP/VXLAN tunnels
  - 512K shared table, can ext to 2M
  - For ACL/FIB/Tunnel or any table.
- Network Reachable information distribution
  - EVPN/VXLAN with BGP for VTEP.
  - SDN Controller collect information on OVS and Switch and mapping IP/MAC to each
  - Flexible Control between OVS/Switch, Micro Segmentation, VPC security group etc.



512K shared table, 100K+ to 512K VXLAN Tunnels

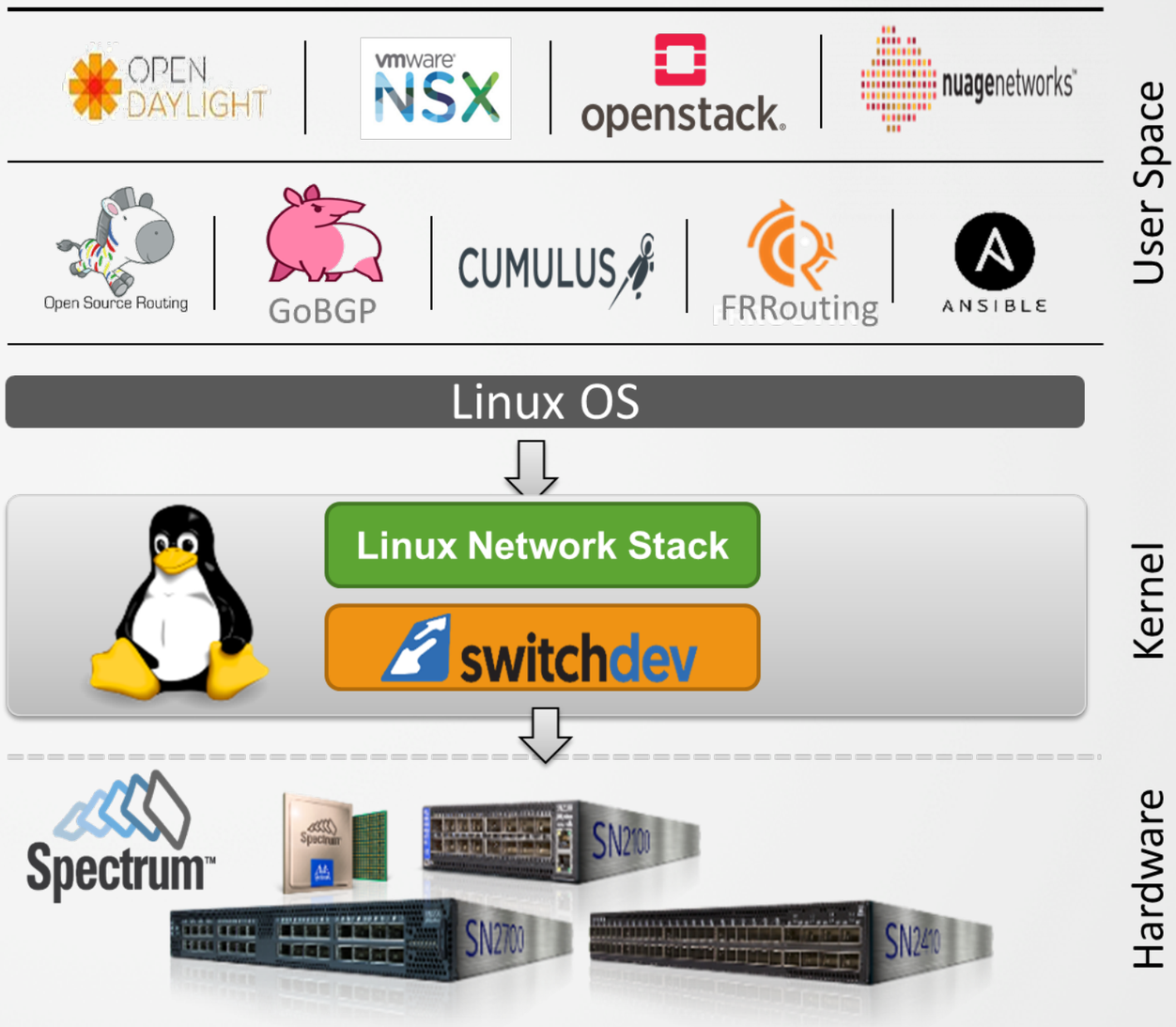
Spectrum™

Linux Switch Innovation

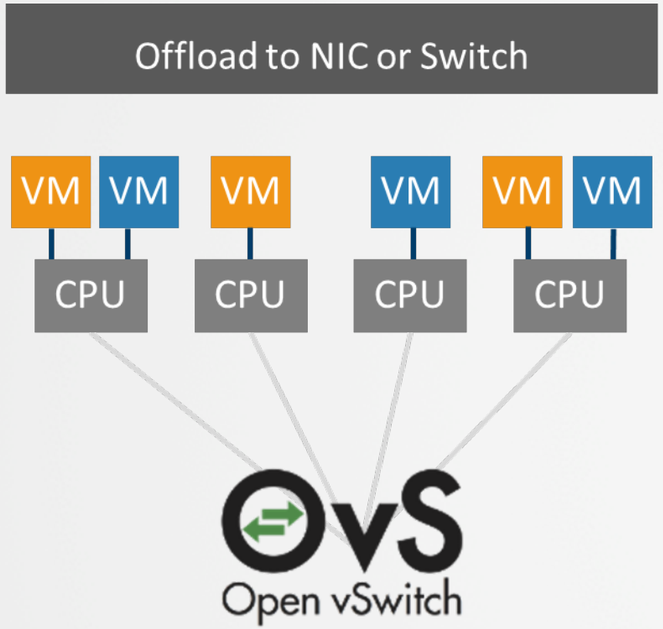
only Mellnaox can support

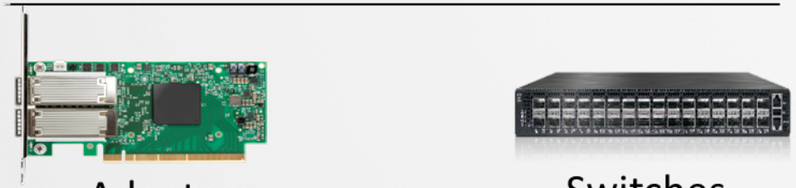
- Working with any Linux
- Working with any opensource Routing and system software
- Working with any SDN controller
- Offloading the Linux network stack into switching ASICs
  - L2 bridging
  - L3 routing (IPv4, IPv6, ECMP, etc.)
  - Match-action (a.k.a ACLs)
  - Encapsulations (IP tunnels, VxLAN)
  - Mirroring
  - OVS(talking to popular SDN)

https://github.com/Mellanox/mlxsw/wiki



OVS Offload to NIC or Switch



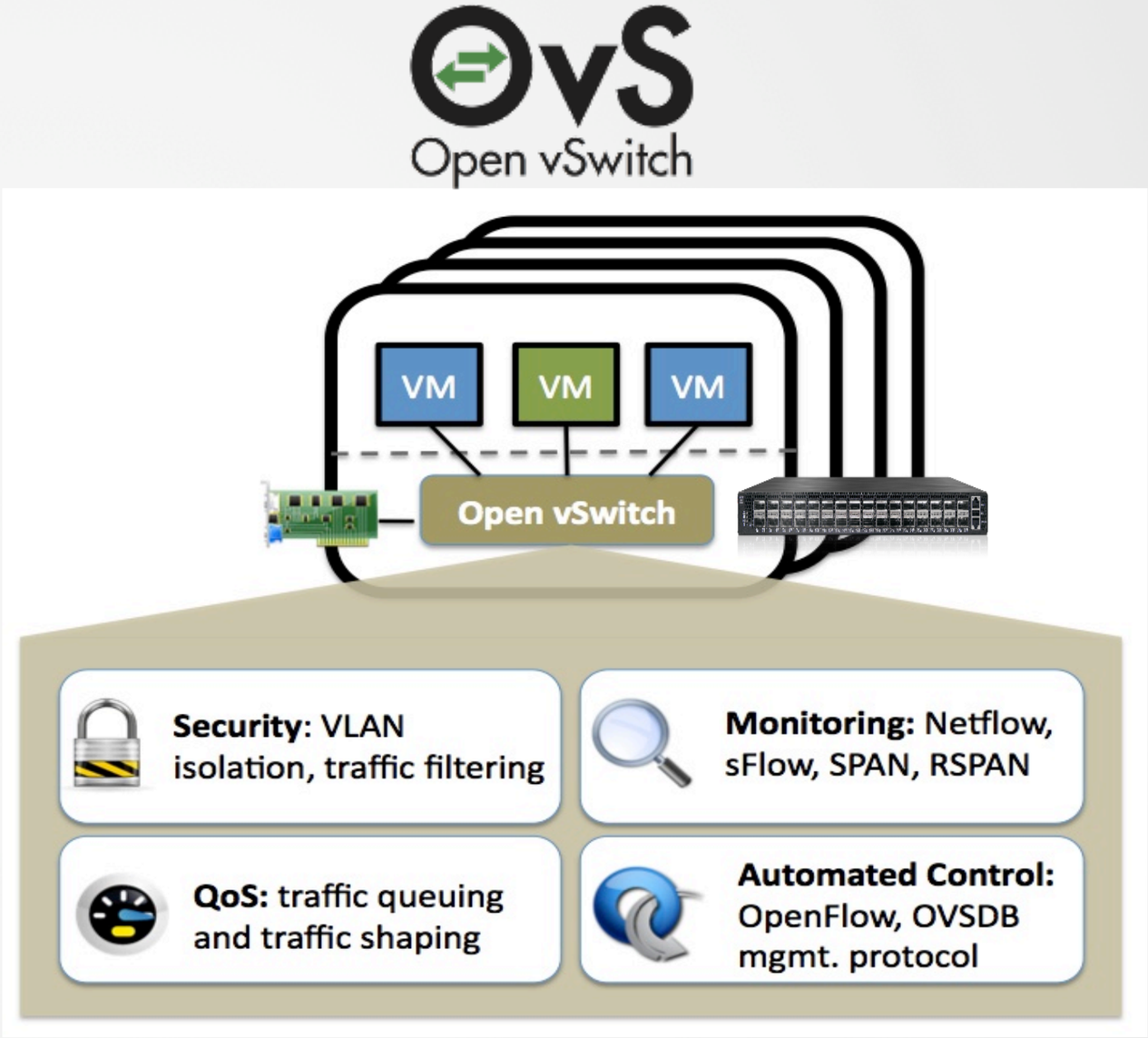


Adapters

Switches

Ipsec, FW

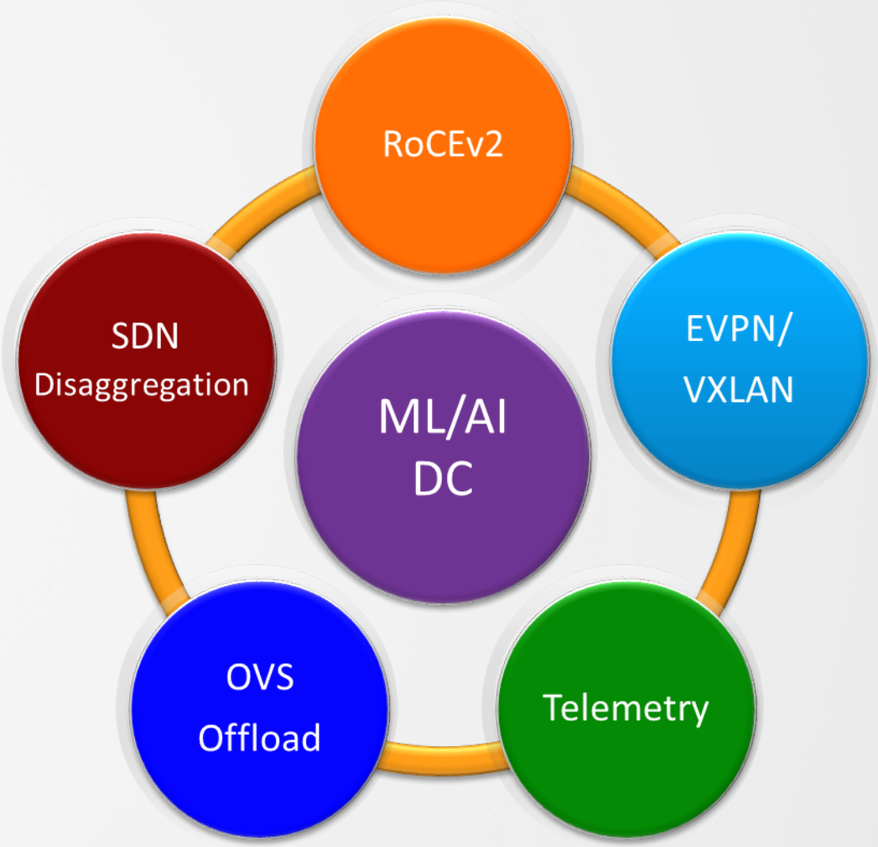
VXLAN/NAT etc



https://github.com/Mellanox/mlxsw/wiki

Summary

- **RDMA critical for New DC Evolution**
  - Machine Learning and AI, Facial/Voice/Image Recognition
  - Ultra Low latency, 6.4T TOR feature, Best Scale
- **SDN and Disaggregation**
  - Openflow for special use case
  - Sonic/Linux switch based disaggregation
- **OVS offload**
  - Linux Switch with OVS talking to a lot of SDN controller.
  - 3.2T/6.4T/12.8T SDN Switch
- **EVPN/VXLAN**
  - Controller Less L2/L3 with QinVNI DCI and Ansible Automation
- **Telemetry**
  - What Just Happen
  - Inband Telemetry



Thank You