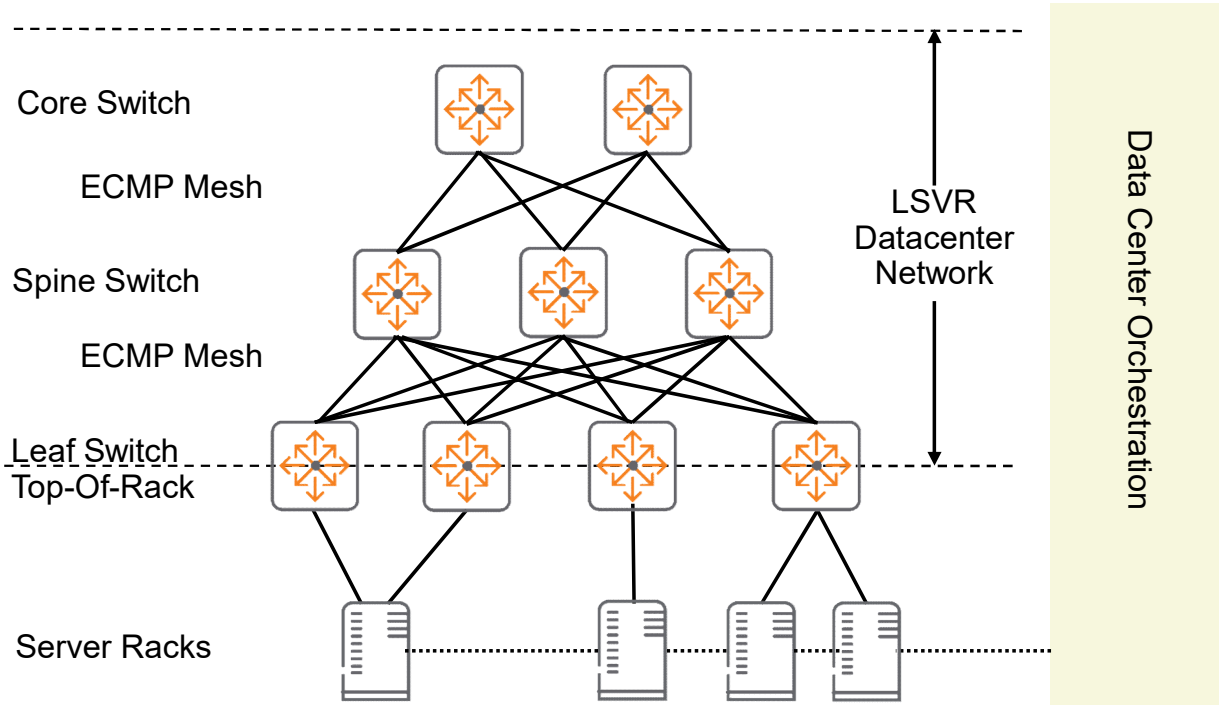
Data Center Server to Network Address Discovery

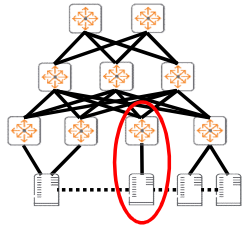
aruba
a Hewlett Packard Enterprise company

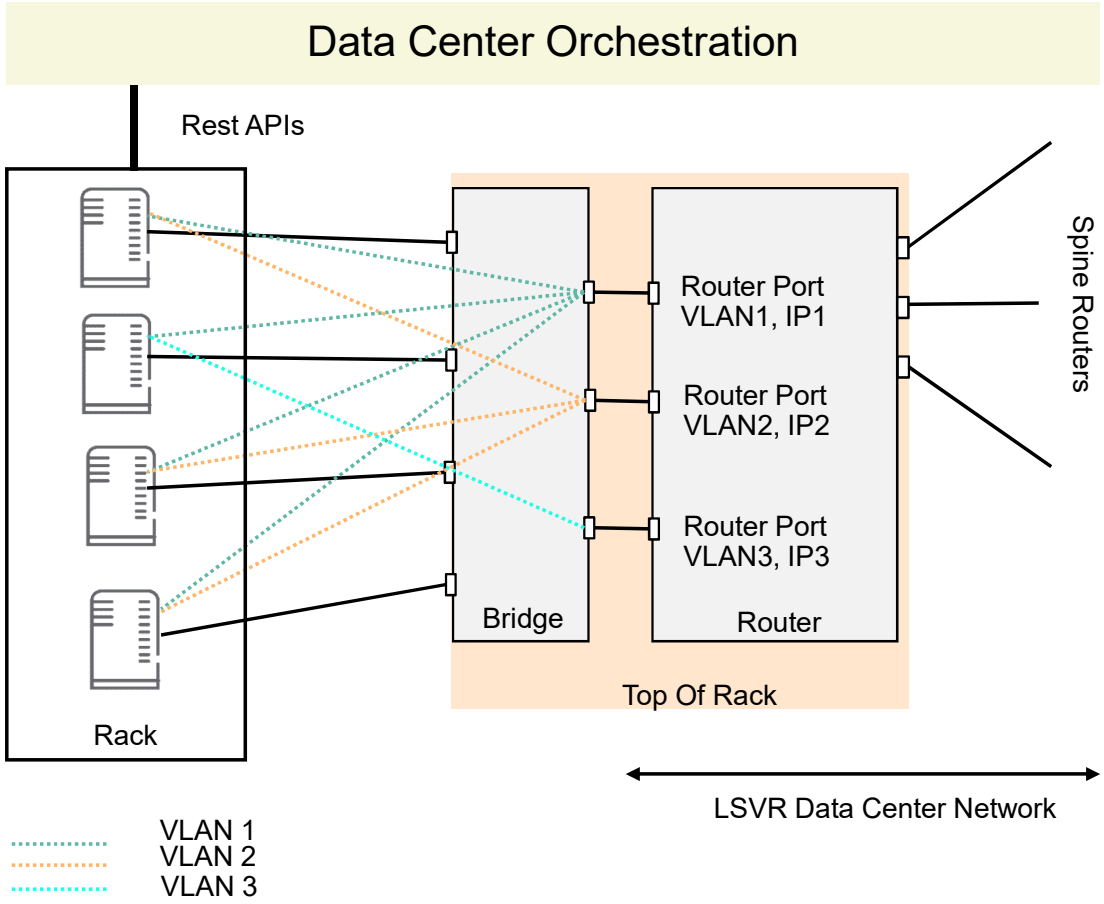
Datacenter Network Using LSVR



- Most datacenters are configured as 2-3 layer Clos networks using ECMP for distribution over the mesh and LAGs/M-LAGs for server attachment
- Typically these networks provide an IPv4/IPv6 topology organized with ToR and Spine switches within Pods (around 8-128 racks)
- Servers at the network edge manage virtual and tenant networks which are encapsulated into the IP packets for transmission over the data center
- The orchestrator controls the creation of the virtual and tenant networks along with coupling to services

Typical Server and Switch Rack Configuration

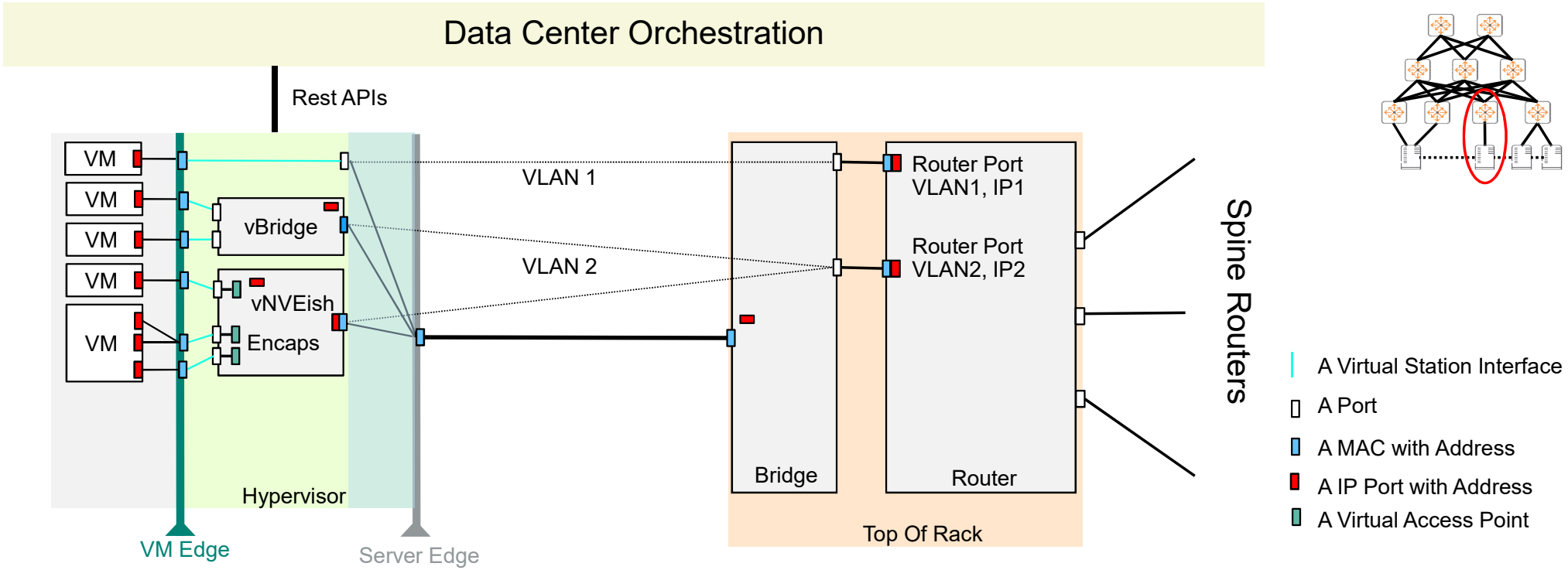




- Here the Bridge portion of the Top Of Rack Switch couples physical ports to each server in the rack
- Over the Bridge Ports VLANs are distributed to each server
- For each VLAN within the rack an IP subnet is assigned
- Each router port in the Top Of Rack is coupled to a single VLAN which is mapped onto an IP subnet
- Protocols within the switch (in this case LSVR) advertise the subnets available within the rack to the rest of the network

aruba
a Hewlett Packard Enterprise company

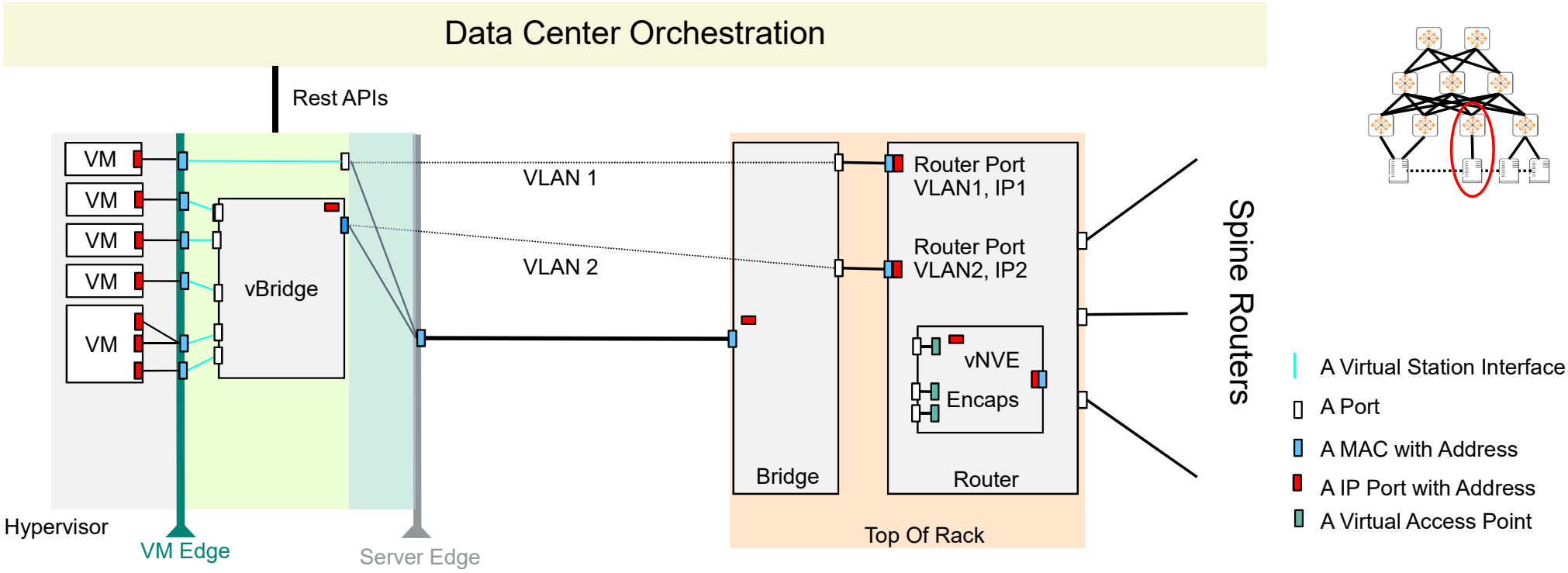
Server Network Interfaces – Virtual Machines (i.e. VMWare)



- **Virtual Station Interface (VSI, defined in IEEE Std 802.1Q-2018): is an internal LAN which connects between a virtual NIC and a virtual Bridge Port**
- **Virtual Access Point (VAP): A logical connection point on the Network Virtualization Edge (NVE) for connecting a Tenant System to a virtual network**
- **DC network is a simple IP underlay network. For scaling L3 encapsulations are supported using "NVE like" procedures within the server controlled by Data Center Orchestration**

aruba
a Hewlett Packard Enterprise company

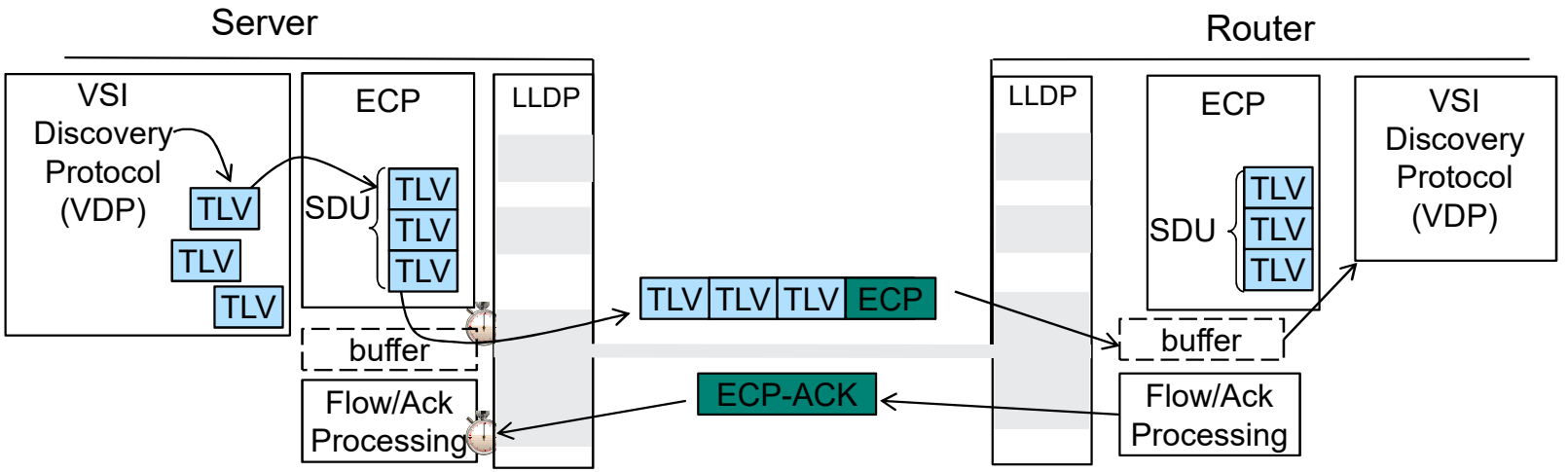
Server Network Interfaces – Virtual Machines Split NVE

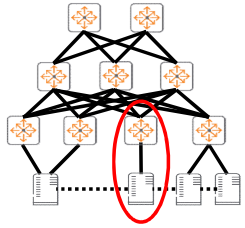


- **In the split Network Virtualization Edge the NVE is moved out of the server and into the DC network requiring the network to become aware of overlays**
- **Split NVE is rarely (if ever) done because it makes coupling with DC Orchestration much less straight forward**
- **In the split NVE case the DC network takes on the encapsulation. This requires co-ordination between the server and network to form the encapsulations**

aruba
a Hewlett Packard Enterprise company

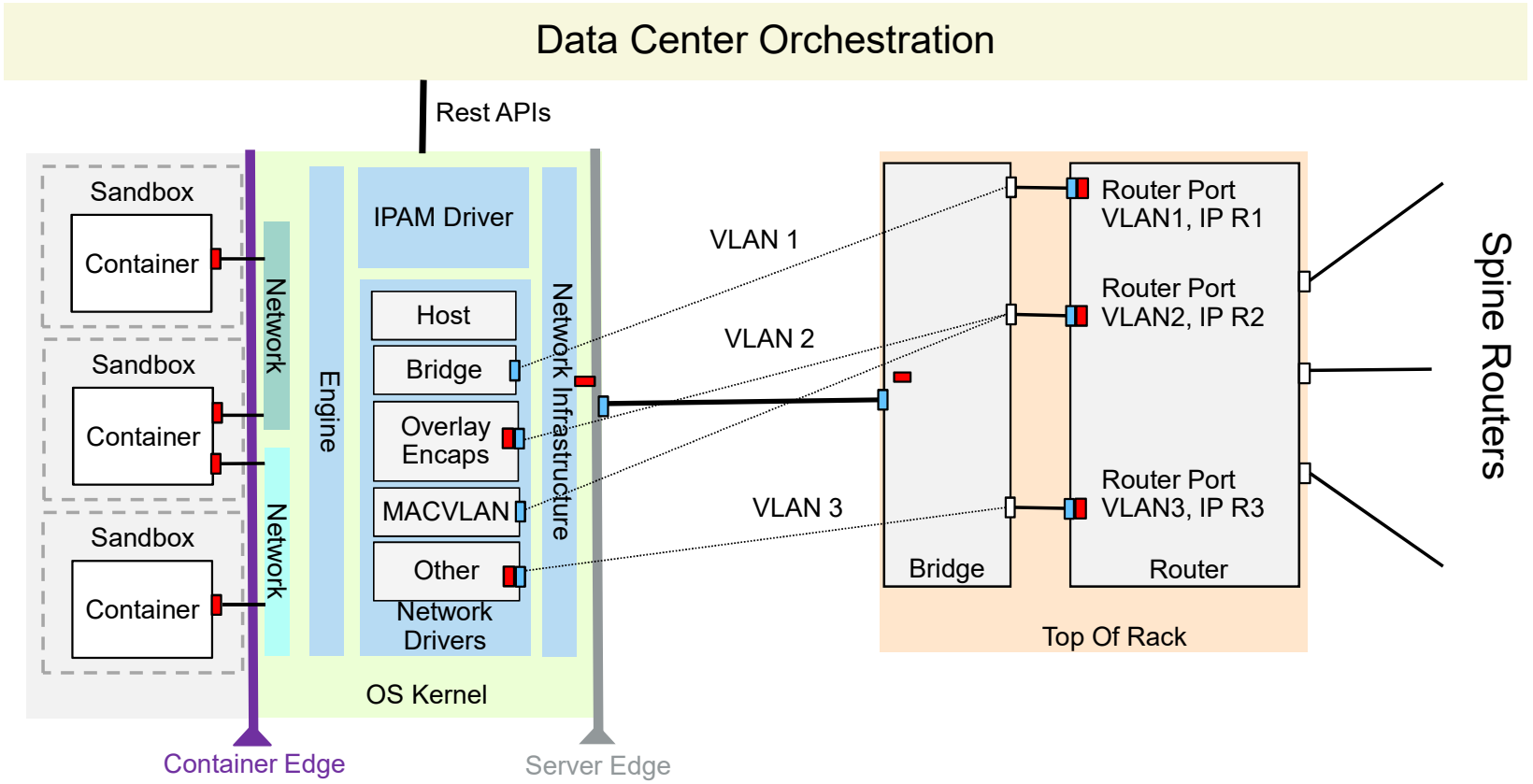
IEEE VSI Discovery and Configuration(VDP) Protocol for Split NVE

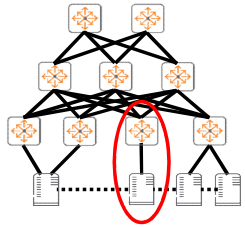




- IEEE Std 802.1Qcy-2019, VDP Extension to Support NVO3, amends IEEE Std 802.1Q-2018 clauses 40, 41, 43 to include support for IPv4 and IPv6 split NVEs. This standard was done in conjunction with IETF NVO3 group and is fully approved as of March 2019.
- Allows discovery of Server, VM, and Container virtual interfaces along with the associated MAC, IPv4 addresses, IPv6 addresses, service classes, managers, profiles, and VPNs.
- Provides quite operation, reliable delivery, flow and congestion control, livelyness, and VM mobility state
- Supports scaling to an unlimited number of virtual station interfaces
- Implementation uses a stack with VDP as the application, ECP transport (others could be used), and LLDP for configuration

Server Network Interfaces – Containers (i.e. Docker)

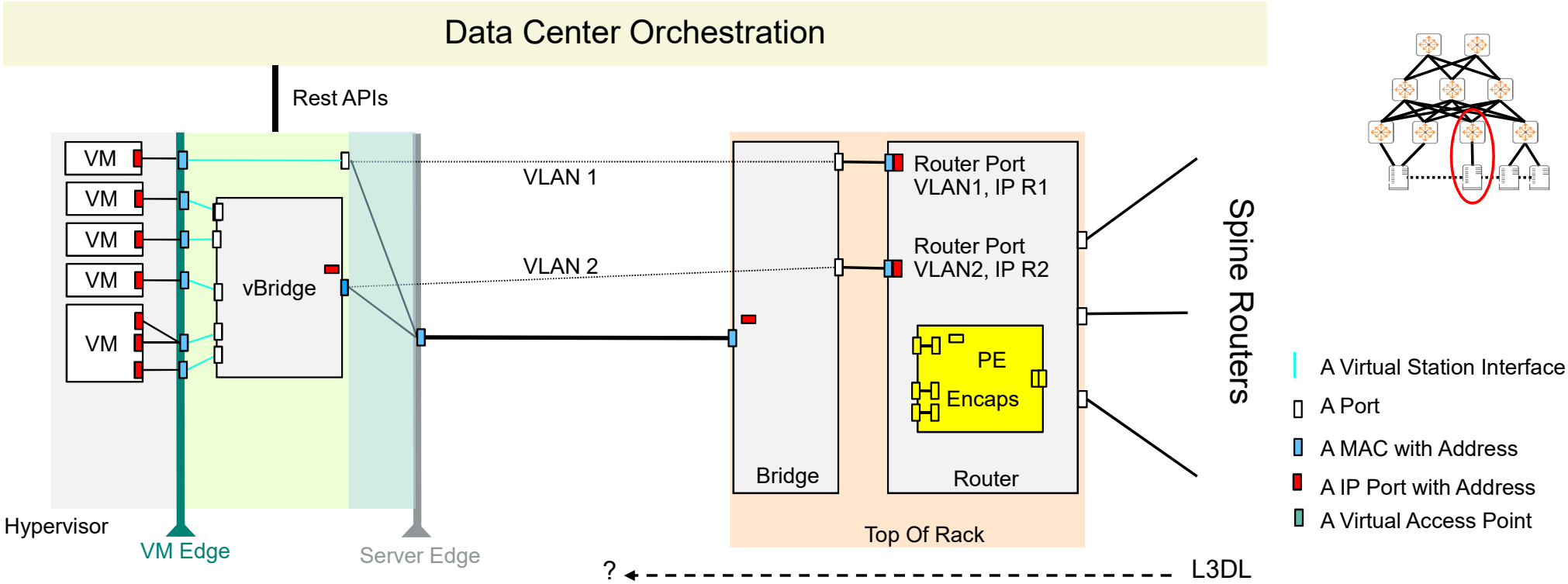




- **Container Solutions use Linux Namespaces and Groups to isolate containers**
- **These solutions provide a variety of network connections, though use an overlay for large scale datacenters**
- **DC network is a simple IP network. For scaling L3 encapsulations are supported using "NVE like" procedures within the server controlled by Data Center Orchestration**

aruba
a Hewlett Packard Enterprise company

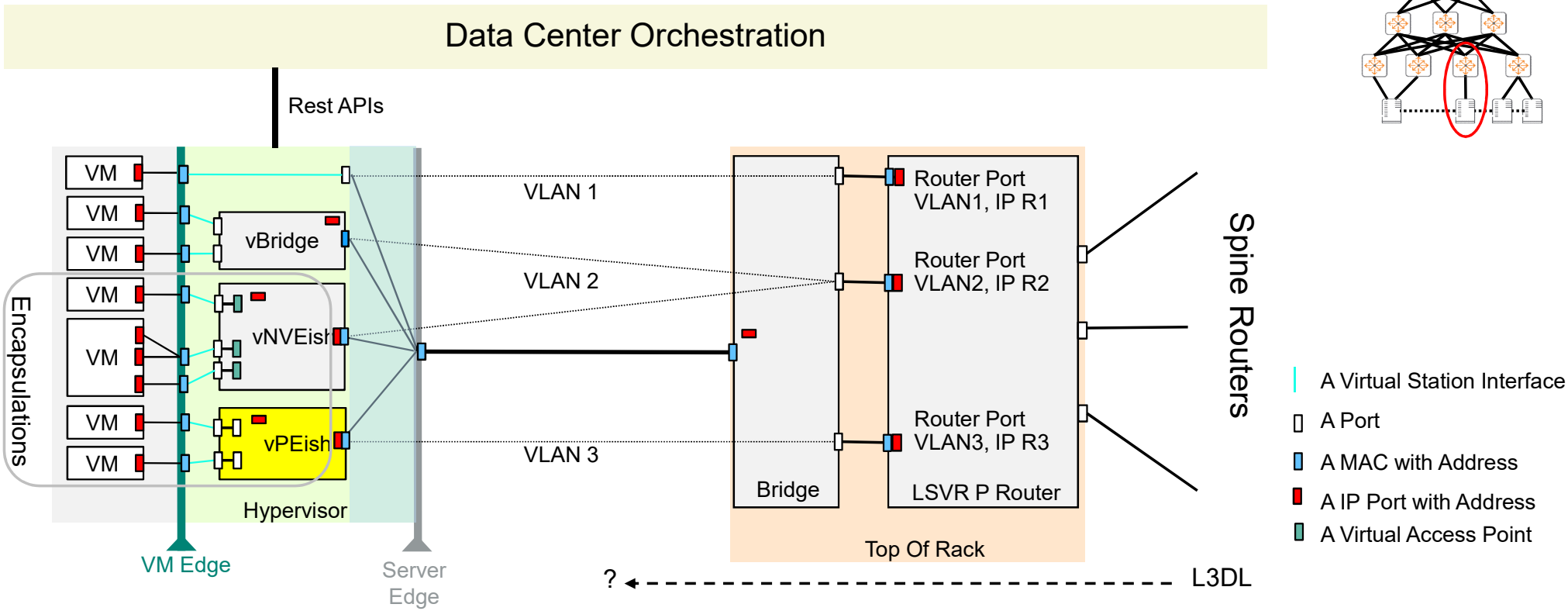
Were To Place a PE Within the Data Center? (split PE case)



- **Placing the PE at the Top Of Rack would be consistent with WAN implementations, however the Data Center is different because the servers are the centrally controlled resource.**
- **It is likely a PE would be implemented in the hypervisor or the OS layers of the server under the control of the Data Center Orchestration since this allows configuration from the Orchestrator**
- **In the split PE case the IEEE VDP protocol (Std 802.1Qcy) could be used to co-ordinate addressing between the server and the PE**

aruba
a Hewlett Packard Enterprise company

Likely DC PE Implementation – Virtual Machines

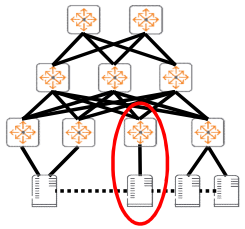


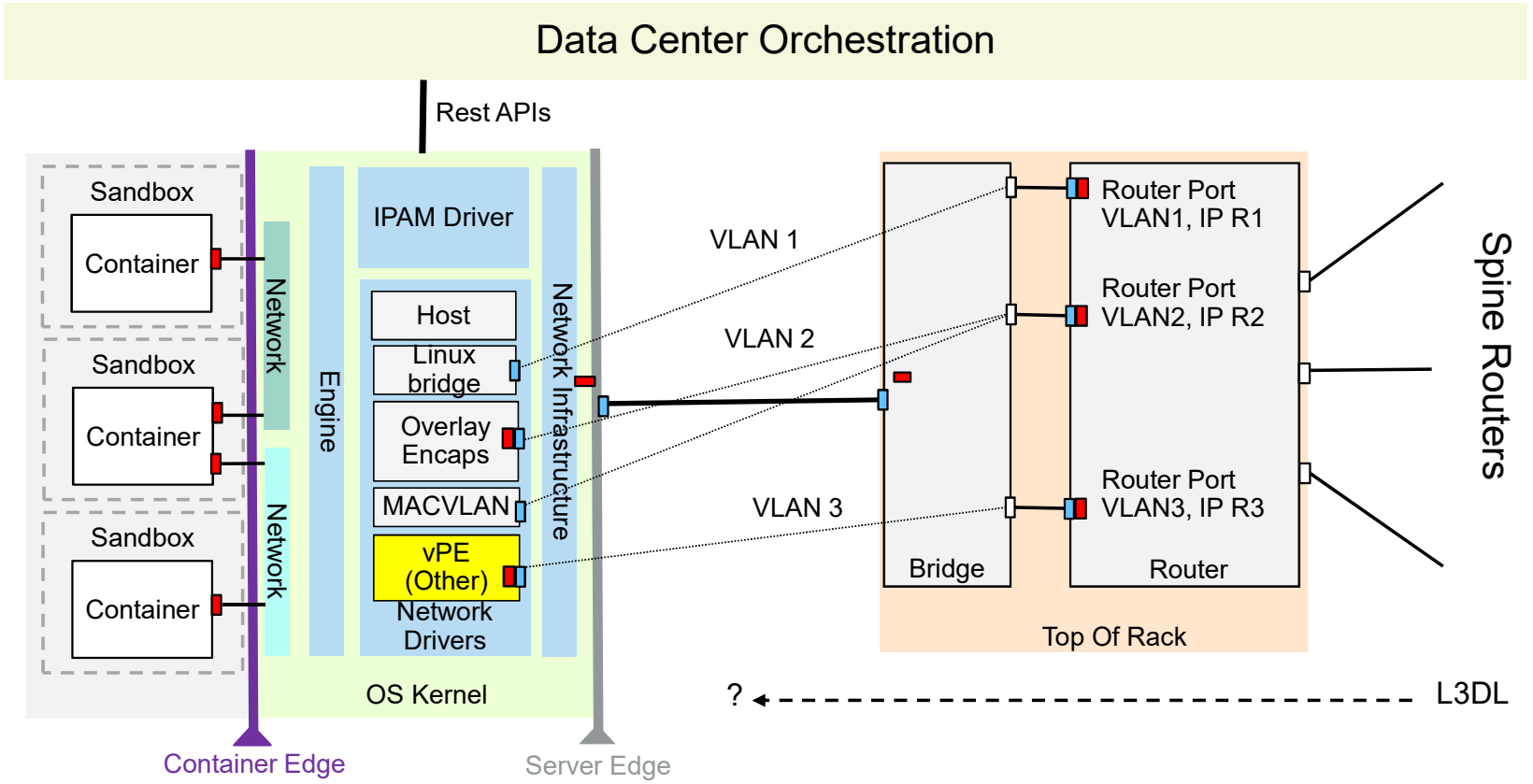
– **Here we add a vPE to the hypervisor within the server. The vPE addressing is controlled by the Data Center Orchestration.**

– **The DC network is a simple IP or MPLS network used to forward between the vPEs.**

aruba
a Hewlett Packard Enterprise company

Likely DC PE Implementation – Containers



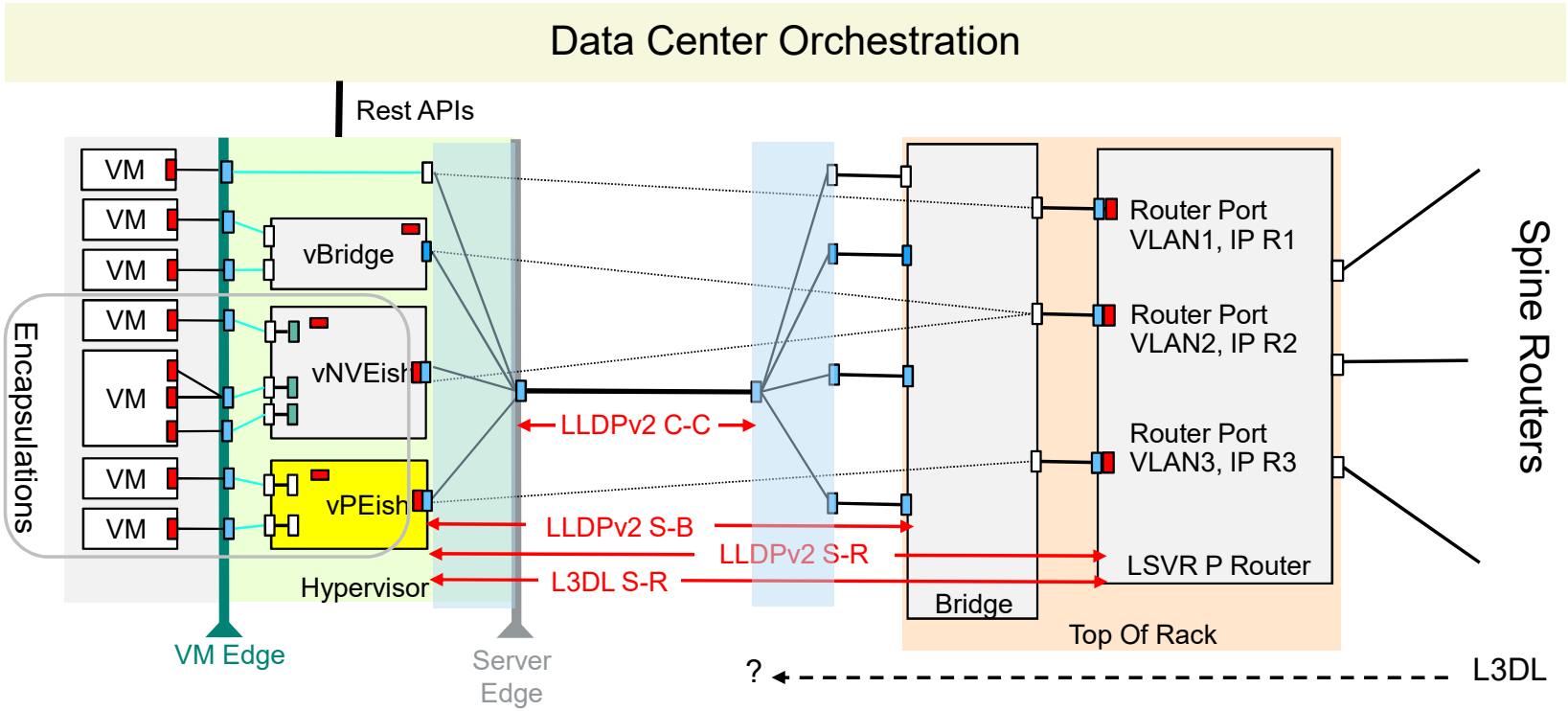


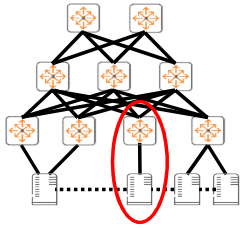
– **Here we add a vPE as a new Network Driver (plug-in) within the OS kernel. The vPE is controlled by the Data Center Orchestration.**

– **The DC network is a simple IP or MPLS network used to forward between the vPEs**

aruba
a Hewlett Packard Enterprise company

Discover Protocol Termination Points for LLDPv2





C-C: Server S-Channels
S-B: Server to Bridge
S-R: Server to Router

- **Currently LLDPv2 is specified to operate at two levels within a Server. These are between the Server and the adjacent Top Of Rack switch (S-B) and over an S-Channel to a Virtual Edge (PE-B).**
- **The IETF L3DL protocol is specified to operate between end system ports (PR-R). LLDPv2 could also take this path by choosing a destination MAC that passes through Bridges rather than contained at Bridges**
- **For the typical case where there are no other Bridges except those embedded in the Server and ToR it is un-necessary to pass LLDP through the Bridge layer. Instead, the Router control plane just needs an API to the LLDPv2 database.**

aruba
a Hewlett Packard Enterprise company

Conclusions

- Servers co-ordinate with data center Orchestration to get address and network assignments
- The Orchestration configures virtual overlay networking for VMs and Containers using encapsulation protocols
- The encapsulation protocols used by Orchestration to support virtual network overlays are implemented as software in the server's hypervisor or OS kernel
- Virtual Network support for VMs and Containers both assume the data center network is a simple IP underlay without any knowledge of the virtual networks the Orchestrator overlays on them
- In the event we have a split NVE (or PE) the VDP protocol (IEEE Std 802.1Qcy-2019) is already standardized to allow Virtual Network and Address co-ordination.

Recommendations

– We already have a protocol for discovery of server, VMs, Containers within data center servers which is IEEE Std 802.1Qcy-2019 (VDP).

– The VDP protocol fills the IETF requirements for end system discovery(except for security), as well as providing solutions for VPN assignment, traffic class, profiling, and VM/Container motion which the IETF has yet to consider.

– The VDP protocol could be secured at the link level using MAC Sec which is a well understood and worked out protocol.

– VDP is too much protocol for common LSVR deployments since typically all we need is to identify a router at the other end of a link.

– A better solution for LSVR L2 discovery would be the use of LLDP (v2) along with CFM (or other link liveliness protocol) for router to router link discovery.

– In the rare event the LSVR router is hosting an NVE (i.e. PE) which needs address co-ordination with servers located on an attached subnet LLDP can be used to request VDP support.

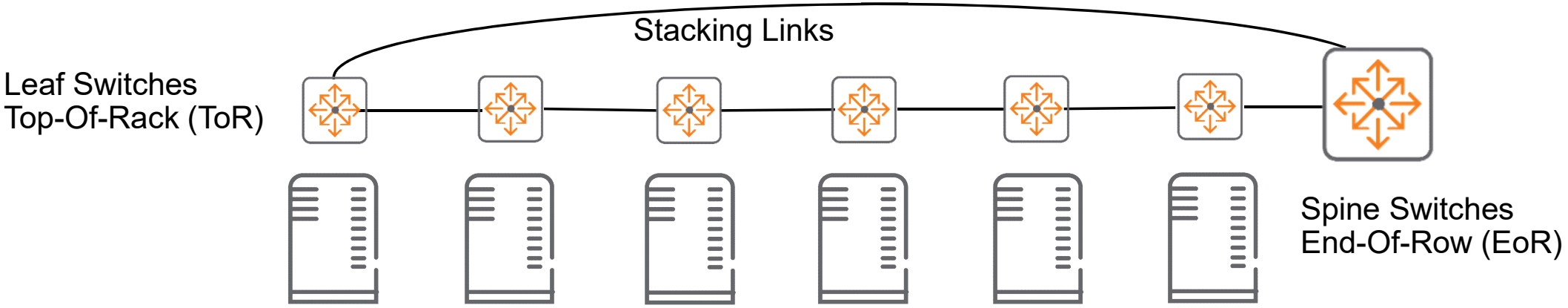
– To secure LLDP(v2) we could use MAC Sec rather than supporting signatures for specific LLDP TLVs.

Thank You

aruba
a Hewlett Packard
Enterprise company

Backup Slides

Alternate Small Datacenter



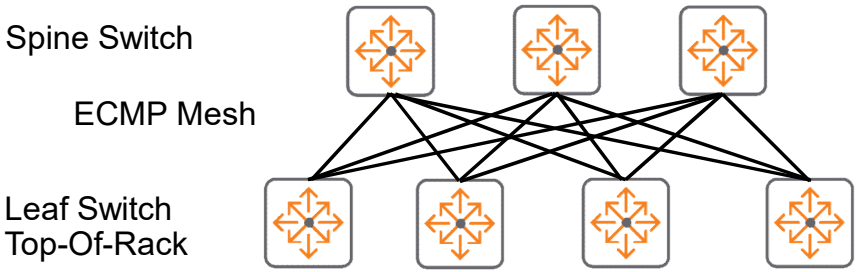
– Here the link rate between switches is lower and max hop count is higher than the Clos tree, however this is a low cost structure for a small datacenter composed of a handful of racks

NVO3 Underlays: VxLAN, VxLAN-GPE, GENEVE, GUE

UDP Encapsulations

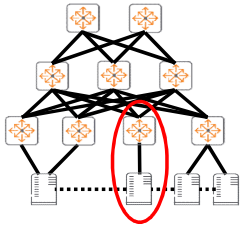
Outer Ethernet Headers	
Outer IP Headers	
Source Port	Destination Port
UDP Length	UDP Checksum

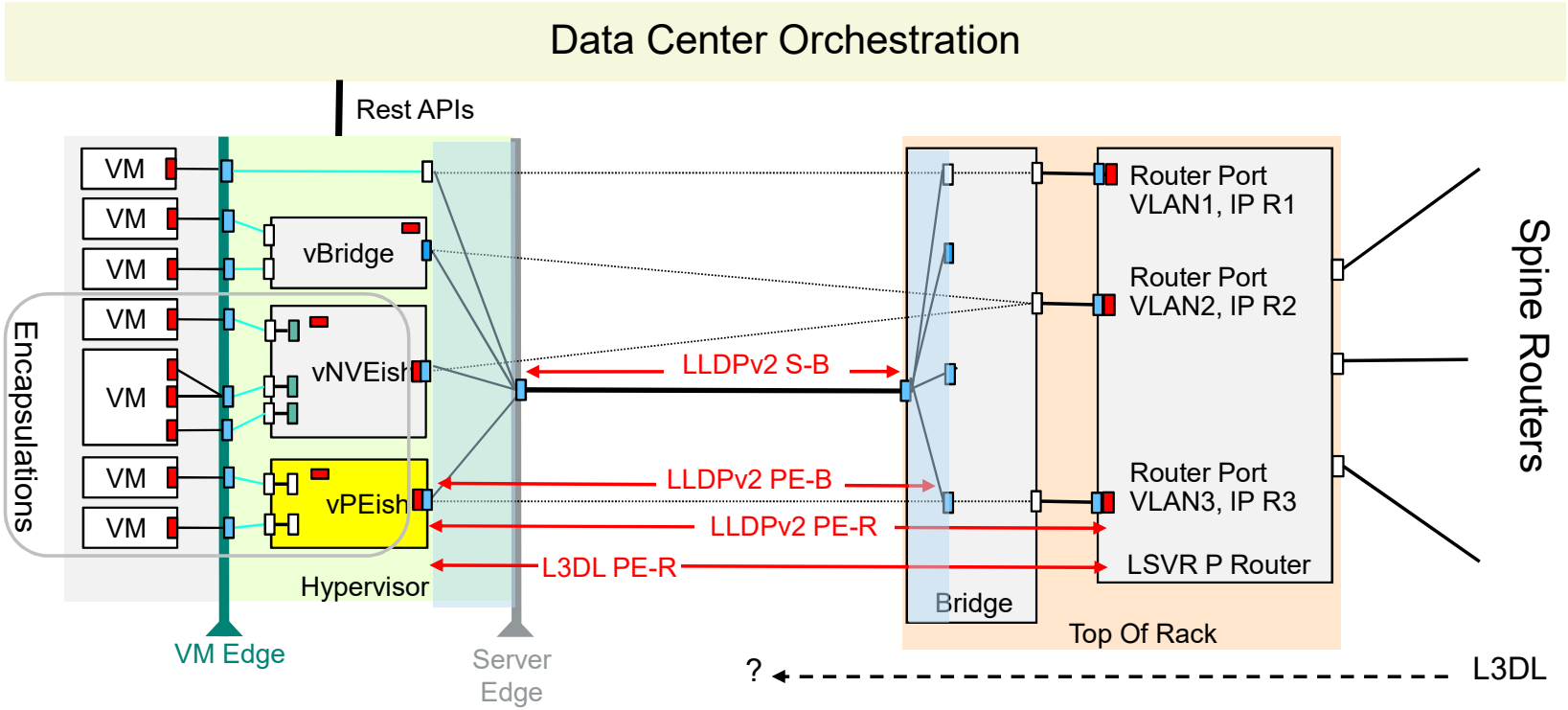
Clos Network



- Like VxLAN all the new encapsulations use IP/UDP for the outer underlay addresses
- The UDP Destination Port determines the encapsulation
  - 4789 for VxLAN, 4790 for VxLAN-GPE, 6080 for GUE, 6081 for Geneve
- The UDP Source Port is used for ECMP entropy
- For Data Center Applications the Underlay may terminate at ToR, NIC or Server Software, however termination at server software is by far the prevalent technique

Discover Protocol Termination Points for LLDPv2





- **Currently LLDPv2 is specified to operate at two levels within a Server. These are between the Server and the adjacent Top Of Rack switch (S-B) and over an S-Channel to a Virtual Edge (PE-B).**
- **The IETF L3DL protocol is specified to operate between end system ports (PR-R). LLDPv2 could also take this path by choosing a destination MAC that passes through Bridges rather than contained at Bridges**
- **For the typical case where there are no other Bridges except those embedded in the Server and ToR it is un-necessary to pass LLDP through the Bridge layer. Instead, the Router control plane just needs an API to the LLDPv2 database.**