The need of Virtual Extended LANs

ITNOG 10th May - Bologna

Andrea Rosano`

Contents

1

Technology trend

The major contributors of the technologic change

IOT

CLOUD

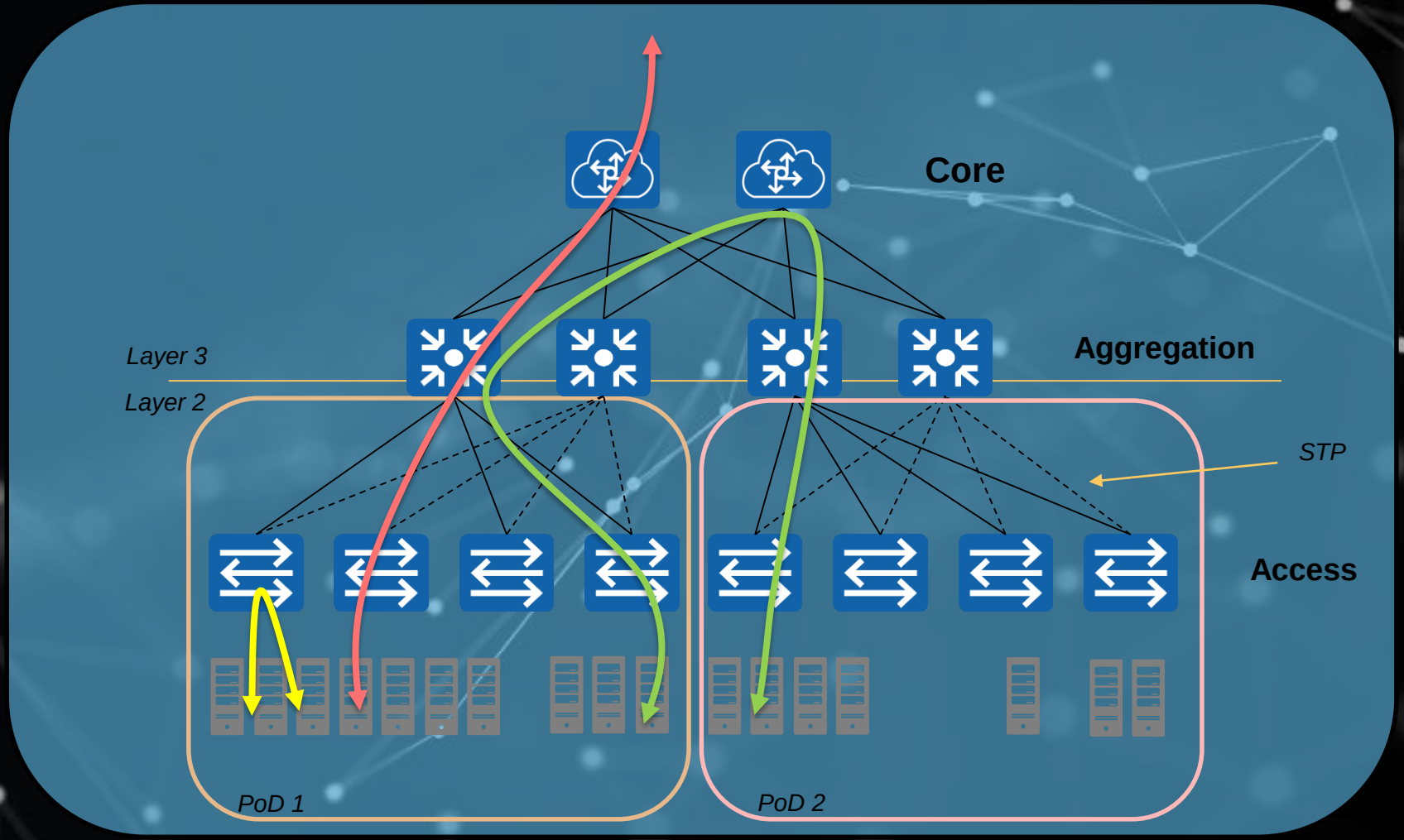
SDN

2

Data Center evolution

2.1

Legacy Data Center architecture



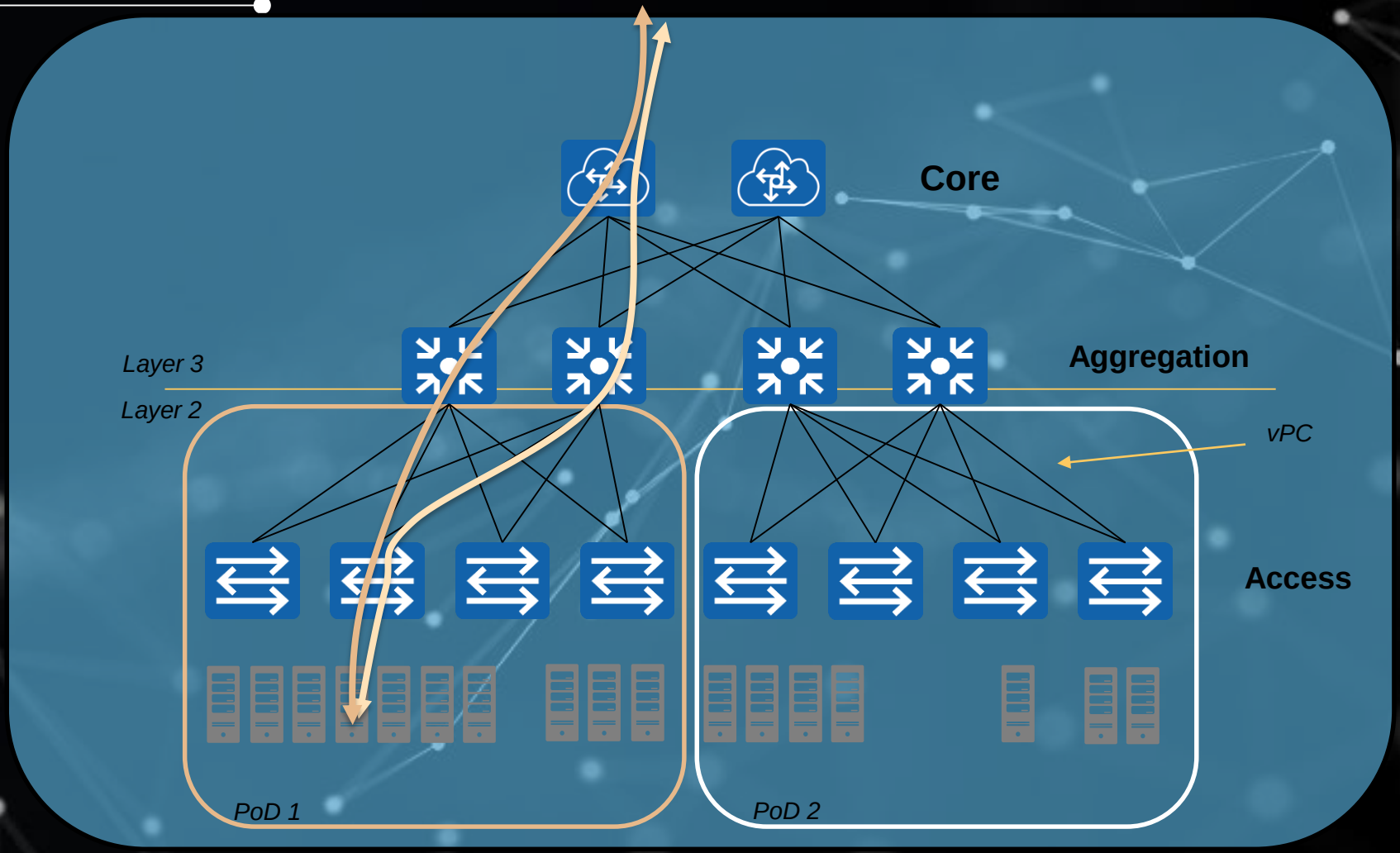
3 – Tier Architecture

- STP used to prevent loop in layer 2 network
- VLANs are extended within each PoD that servers can move freely within the pod without the need to change IP address and default gateway configurations.

- Spanning Tree Protocol cannot use parallel forwarding paths, and it always blocks redundant paths in a VLAN.

2.2

Legacy Data Center architecture



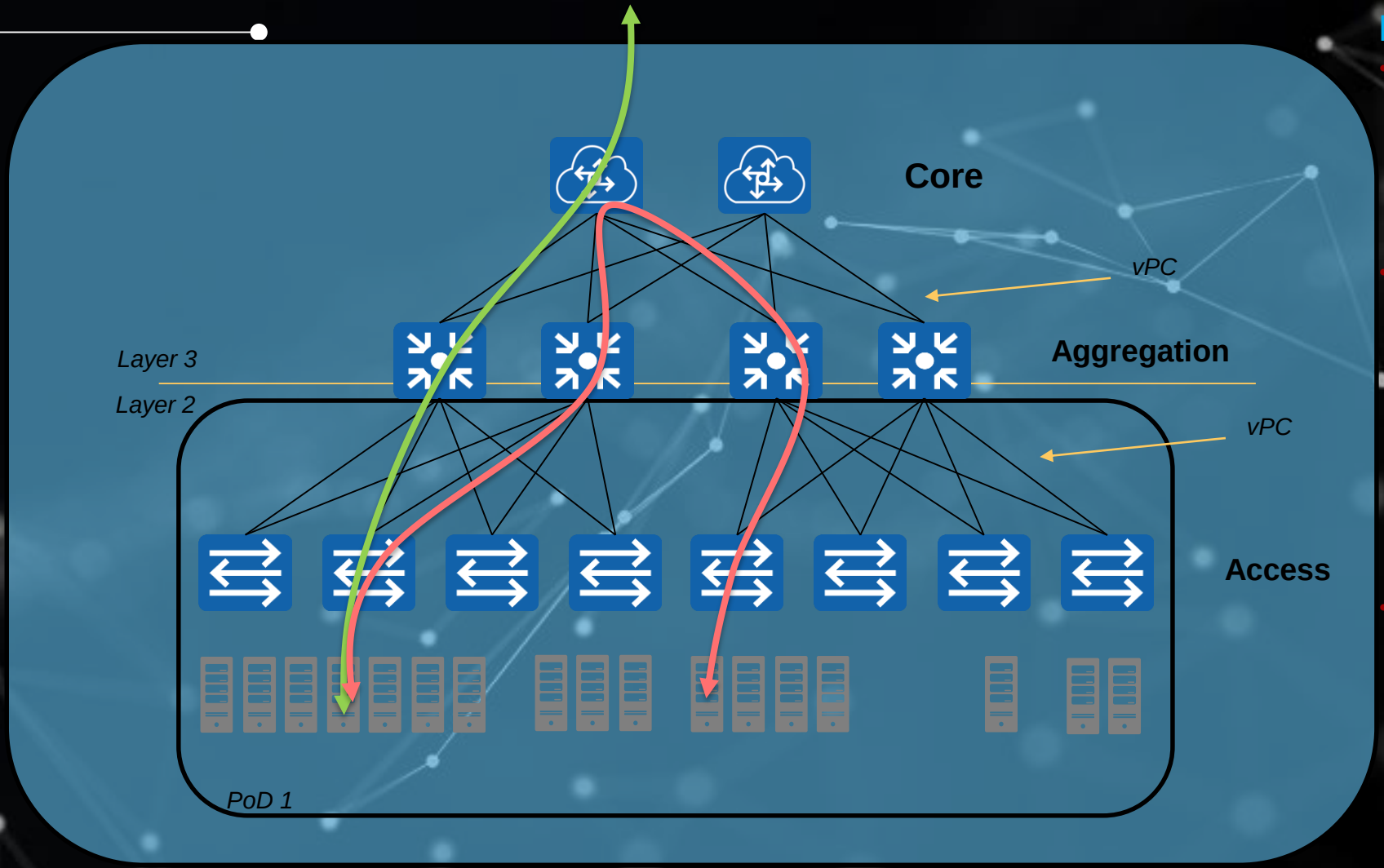
Virtual port channel (vPC)

- In 2010 Cisco introduced vPC to eliminate the spanning-tree blocked ports, provides active-active uplink from the access switches to the aggregation routers, and makes full use of the available

- vPC technology works well in a relatively small data center environment in which most traffic consists of northbound and southbound communication between clients and servers.

2.3

Legacy Data Center architecture



NFV needs

- Since 2003, with the introduction of virtual technology, the computing, networking, and storage resources that were segregated in pods in Layer 2.
- Need for a larger Layer 2 domain (Servers are virtualized into sets of virtual machines that can move freely from server to server without the need to change their operating parameters)

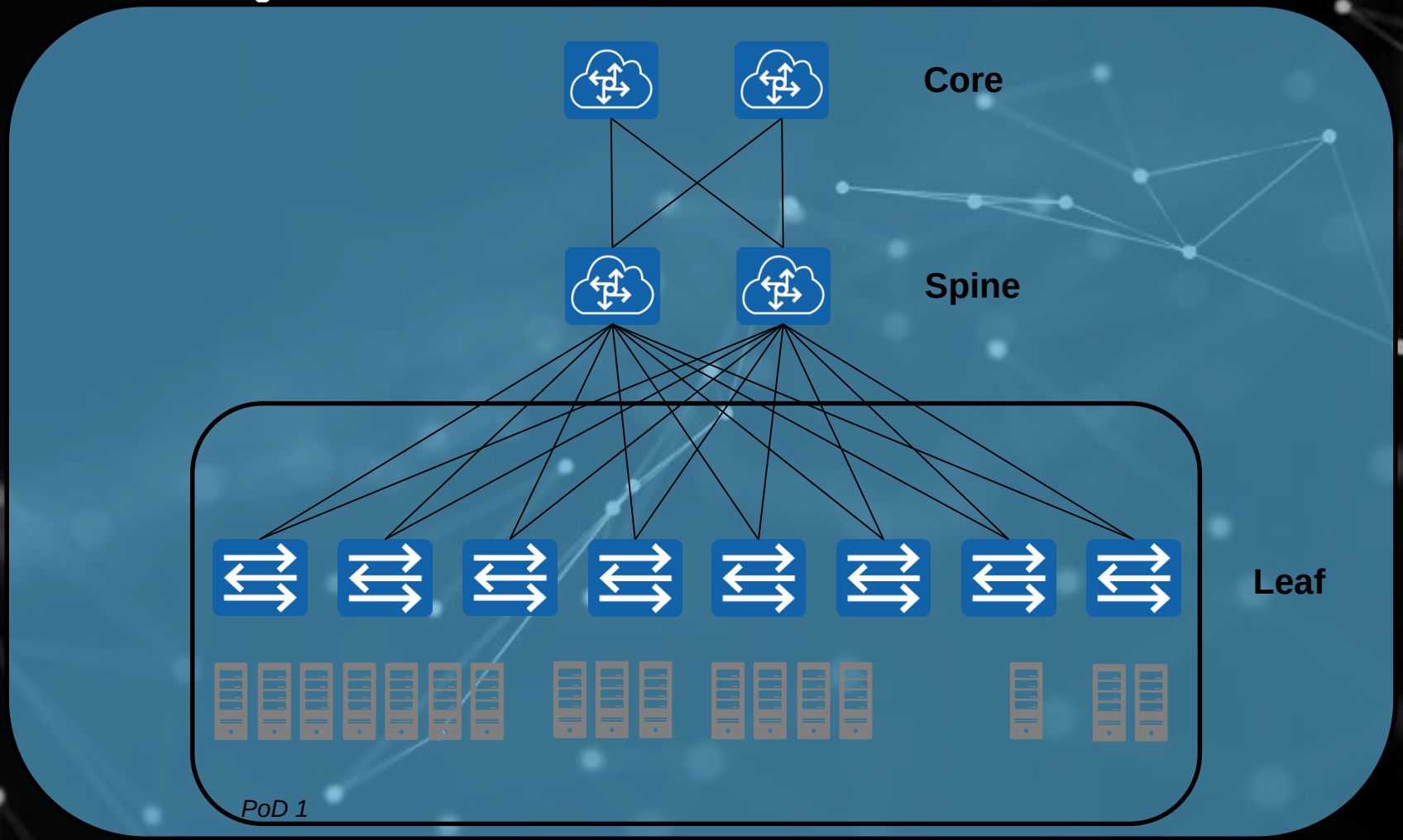
- vPC can provide only two active parallel uplinks, and so bandwidth becomes a **bottleneck** in a three-tier data center architecture

3

Leaf & Spine Architecture

3.1

Clos architecture



Spine & Leaf architecture (Clos)

- Every leaf switch connects to every spine switch in the fabric.
- The path is randomly chosen so that the traffic load is evenly distributed among the top-tier switches.
- If oversubscription of a link occurs (that is, if more traffic is generated than can be aggregated on the active link at one time), the process for expanding capacity is straightforward. An additional spine switch can be added.
- No matter which leaf switch to which a server is connected, its traffic always has to cross the same number of devices to get to another server (unless the other server is located on the same leaf).

- This approach keeps latency at a **predictable level** because a payload only has to hop to a spine switch and another leaf switch to reach its destination.

3.2

Clos architecture

All interconnection used

no need to use STP

All east-west traffic is equidistant

predictable latency

The architecture doesn`t solve L2 adjacency problem

Multitenant system is not easy to implement

Switch config. fixed

no network changes required for a dynamic server

4

VXLAN

4.1

VXLAN definition

is a Network Virtualization over Layer 3 (NVO3) technology that uses the MAC in User Datagram Protocol (MAC-in-UDP) mode to encapsulate packets.

MAC DA
specifies the destination MAC address of the next-hop device on the route to the destination VTEP.
MAC SA
specifies the source MAC address of the source VTEP that sends the packet.

IP SA
specifies the source IP address, which is the IP address of the source VTEP.
IP DA
specifies the destination IP address, which is the IP address of the destination VTEP.

VXLAN Network Identifier (24bits)
Identifies a VXLAN segment with up to 16M tenants
Users in different VXLAN segments cannot directly communicate at Layer 2

DestPort
specifies the destination UDP port number (4789).
Source Port
specifies the source port number. It is the hash value calculated using parameters in the inner Ethernet frame header.

4.2

VXLAN basic concept

L2 gateway
Similar to a Layer 2 access device on a traditional network, it allows tenant access to VXLANs and intra-subnet VXLAN communication in the same network segment.

Broadcast Domain (BD)
is used for broadcast domain division on a VXLAN.
On a VXLAN, to allow Layer 2 communication between VMs in a BD, VNIs and BDs are mapped in 1:1 mode

Network Virtualization Edge (NVE)
a network entity used to implement network virtualization functions. After packets are encapsulated and decapsulated through NVEs, a Layer 2 VXLAN can be established between NVEs over the basic Layer 3 network.

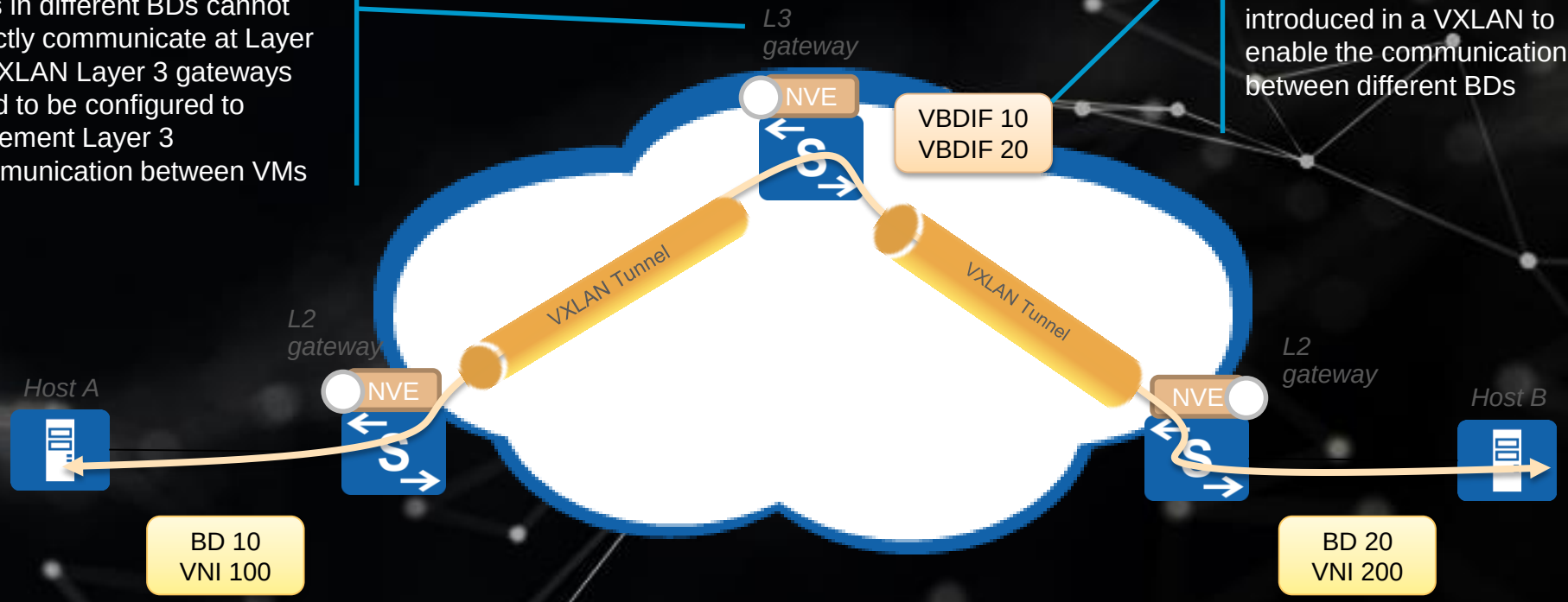
VXLAN Tunnel end point (VTEP)
encapsulates and decapsulates VXLAN packets. The source and destination IP addresses in a VXLAN packet are the IP addresses of the local and remote VTEPs, respectively.

4.3

VXLAN centralized gateway

L3 gateway
VMs in different BDs cannot directly communicate at Layer 2. VXLAN Layer 3 gateways need to be configured to implement Layer 3 communication between VMs

VBDIF interface
VBDIF interfaces are introduced in a VXLAN to enable the communication between different BDs



4.4

VXLAN distributed gateway

Distributed gateway

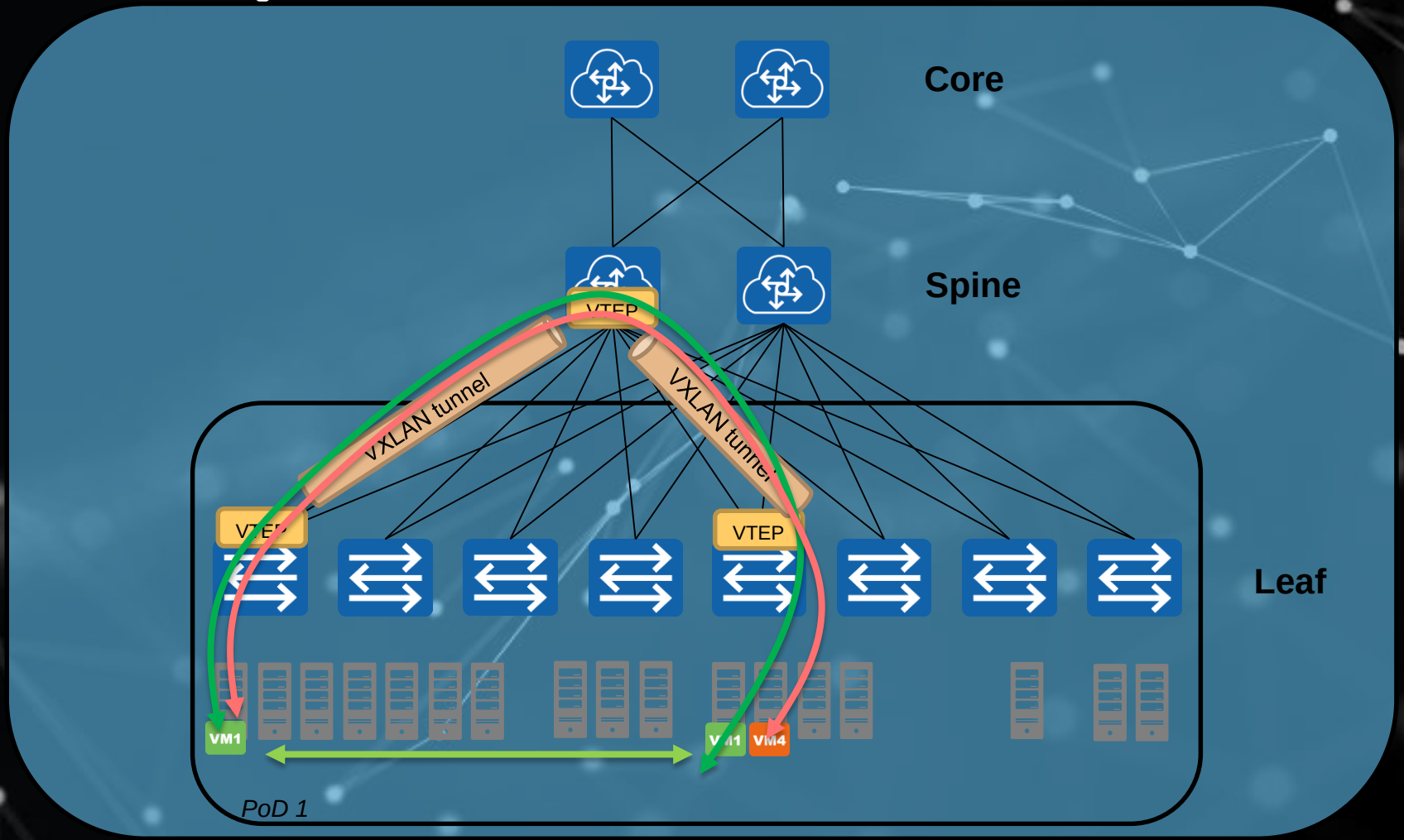
- One VTEP node can work as a VXLAN Layer 2 or 3 gateway, enabling flexible deployment.
- Unlike the centralized Layer 3 gateway which has to learn the ARP entries of all servers, the VTEP node only needs to learn the ARP entries of the connected server, solving the ARP entry problem of the centralized Layer 3 gateway and improving network scalability.

4.5

VXLAN & VLAN

Item	VLAN	VXLAN
Concept	Virtual local area network	Virtual extensible local area network
Implementation method	A physical LAN is divided into multiple BD geographically limited	Layer 2 network are not geo restricted. This allows large scalability
Capacity	12 bits used for VLAN ID with a maximum of 4096 number of VLANs	24 bits used for VNI with a maximum of 16 M of tenant.
Encapsulation mode	A VLAN tag is added to packets.	a VXLAN header, UDP header, IP header, and outer MAC header are added in sequence to an original packet.
Benefits	Limits broadcast domains: A broadcast domain is limited in a VLAN, which saves bandwidth and improves network processing capabilities.	Flexible network deployment: VXLANs are constructed over the traditional network. Technical advantage: VXLAN uses MAC-in-UDP encapsulation. Such encapsulation mode does not rely on MAC addresses of VMs, reducing the number of MAC address entries required on a large Layer 2 network

VXLAN in a Leaf & Spine architecture



Multi tenat DC

- VMs are uniquely identified by a combination of their MAC addresses and VNI. Thus it is acceptable for VMs to have duplicate MAC addresses, as long as they are in different tenant networks. This simplifies administration of multi-tenant customer networks for the Cloud service provider

- A multi-tenant cloud infrastructure is now capable of delivering “elastic” capacity service by enabling additional application VMs to be rapidly provisioned in a different L3 network which can communicate as if they are on a common L2 subnet

Benefits of VXLAN in a Leaf & Spine architecture

the VXLAN 24-bit VNI construct that enables **16 million** isolated tenant networks

VXLAN is a standard construct

multi-tenant cloud infrastructure VMs are uniquely identified by a combination of their MAC addresses and VNI. Thus it is acceptable for VMs to have duplicate MAC addresses

Overlay Networking overcomes the limits of STP and creates very large network domains where VMs can be moved anywhere

Overlay Networking can make hybrid cloud deployments simpler to deploy because it leverages the ubiquity of IP for data flows over the WAN

VXLAN is an evolutionary solution, already supported by switches and driven by software changes, not requiring "forklift" hardware upgrades thus easing and hastening the adoption of the technology

4.8

Huawei`s VXLAN-ready data center switches

Thank you

https://support.huawei.com andrea.rosano@huawei.com