VXLAN EVPN Multi-Site Design and Implementation

East Carolina University, Greenville, NC

Term Paper for Course ICTN 6880 Advanced Topics in Information Infrastructure Design

By Vinay Sawant

Date: 10/23/2019

Abstract

VXLAN (Virtual Extensible Local Area Network) is the future of networking. Since the introduction of VXLAN from last three to four years into network infrastructure, almost all organizations are gradually migrating their network infrastructure to VXLAN. VXLAN is a standard based overlay protocol which is used to extend layer 2 network over traditional layer 3 underlay network. There are various ways of VXLAN implementation depending on your organization's network infrastructure needs and requirements. For example, VXLAN within same datacenter, Multi-Pod VXLAN, and Multi-Site VXLAN. Customer, Network designers and Network Engineers are still learning and exploring the different options and capabilities of VXLAN and trying to figure out how exactly VXLAN is going to help them solving their problem. This research paper mainly focuses on multi-site VXLAN design, configuration, and implementation using BGP (Border Gateway protocol) EVPN (Ethernet Virtual Private Network). VXLAN is a standard based protocol and it can interoperate with various different manufacture's networking product which supports VXLAN. Research in this paper also includes information about VXLAN implementation between different vendors and how does that work and if there are any challenges interacting with multi-vendor environment. This paper also includes brief introduction of VXLAN, why we need VXLAN, what are the different challenges in VXLAN implementation and what is the future of VXLAN in networking infrastructure. Based on the research done in this paper, it is observed that VXLAN is being accepted and implemented globally in many different organizations and it is going to be more popular and have lots of new features, implementation types as we go further.

Introduction

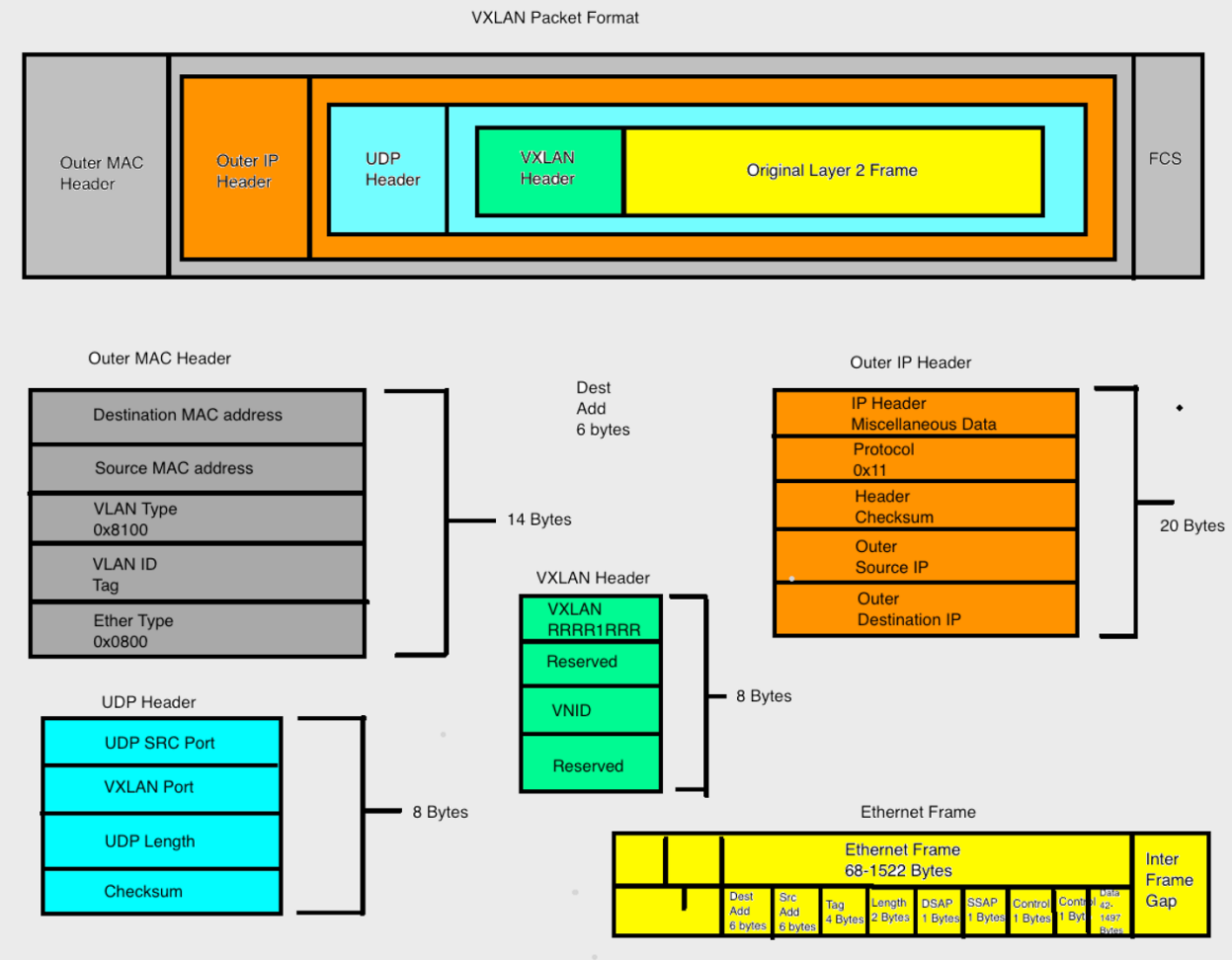
In this era of Internet of Things (IoT), where everything is getting digitized, smart homes devices are increasing at very faster rate and every day more and more people are adding smart devices to their homes. Automatic and self-driving cars are already available, and a new concept is coming out called automated homes. To handle all these new technologies, IT infrastructure needs to be able handle all the load which will need thousands of servers, very big data storages in exabytes, a very high bandwidth, and low latency devices (Singh, Jain, Satish Babu, 2017). To be able to support this requirement, IT infrastructure is evolving towards virtualization and cloud-based infrastructure. There are many different technologies that makes virtualization and cloud-based infrastructure possible. Virtual Extensible Local Area Network (VXLAN) is one of the technologies that supports it and plays important role. The new era of virtualization and cloud-based infrastructure changes the way applications are being designed and implemented. It also changes the way network infrastructure and data centers are being designed. In the new form of data center architecture design, we need flexibility, resiliency, multitenancy capacity and excellent performance and capacity. (Gustavo, Salazar, Edison, Naranjo, & Luis, 2018). Multitenancy enables logical isolation of shared virtual compute, storage, and network resources. Multitenancy is the fundamental technology that cloud use to share IT resources cost efficiently and securely (Del Piccolo, Amamou, Haddadou, & Pujolle, 2016). Virtualization at data center level provides foundation for offering application services and resources for multiple tenant or customers over shared infrastructure. The shared IT infrastructure may not be possible with traditional VLAN and IP routing-based networks and it will be very difficult to manage and design such network. Virtualization helps achieve this goal (Singh, Jain, Satish Babu, 2017). Virtualization adds functionalities such as, remote OS install, access to server console, reboot of frozen server,

possibility of server snapshot for backup, upgrade without shutting down servers, etc. (Del Piccolo, Amamou, Haddadou, & Pujolle, 2016). In traditional form of Networking we have Local Area Network (LAN) which is local to a site or a building or data center. The IP addressing for these local sites will fall in same subnet. For example, all devices at site A may use subnet 10.10.10.0/24, site B may use subnet 10.10.11.0/24 etc. These various LAN are connected using Wide Area Network (WAN). Now when a node in a site A moves to site B, we need to re IP that node to a new IP address that belong to new site. This is ok for small scale network. If the network is very big and if the client are virtual machines and if you want to VMOTION them from one site to another site across WAN and if we want to keep the same IP after VMOTION to new site, we will have to extend our Layer 2 domain. But, making your network flat layer 2 design might not be feasible and advisable. That is where requirement of extending your later 2 domains over Layer 3 boundaries comes into play. There are many technologies available for extending layer 2 domain over Layer 3 boundaries, such as Cisco's Overlay Transport Virtualization (OTV) with Location Identifier Separation Protocol (LISP), Data Center Interconnect (DCI) etc. LISP reorganize network address by two separate address space, virtual address space and physical address space. When a host is moved from one site to another site the LISP mapping service will maintain virtual address space and upgrade physical location address space (Dai, Xu, Xu, Huang, & Qin, 2017). At this time Data Center Interconnect (DCI) are through IP or MPLS VPN (Virtual Private Network) over layer 2 is achieved by Virtual Private Lan Services (VPLS) or Ethernet Over MPLS (EoMPLS) (Singh, Jain, Satish Babu, 2017). The main problem with all these techniques is scalability and interoperability with different vendors. That is where Virtual Extensible Local Area Network (VXLAN) helps to solve that problem. Other such protocols are NVGRE and STT that can do L2-in-L3 tunneling (Dai, B., Xu, Y., Xu, G., Huang, B., & Qin, P. (2017).

VXLAN Basics

VXLAN enables us to extend layer 2 segment over Layer 3 network using VXLAN encapsulation (MAC in UDP). VXLAN is defined in RFC 7348 which is an overlay technology developed to carry layer 2 ethernet frames over traditional IP network. VXLAN solves the problems related to inefficient layer 2 segmentation with traditional VLANs (Gustavo, Salazar, Edison, Naranjo, & Luis, 2018). It provides layer 2 abstraction over leaf and spine architecture and is widely used in multi-tenant environment and is advocated by IEEE for building overlay network on the top of existing layer 3 network (Wang & Lin, 2019). VXLAN uses UDP port 4789 (Ricart-Sanchez, Malagon, Salva-Garcia, Perez, Wang, & Alcaraz Calero, (2018). To form a VXLAN packet, a VXLAN header is attached to the original ethernet frame. Then this entire frame with VXLAN header is encapsulated in UDP header. This UDP encapsulated frame is now put into Outer IP header. Then Outer MAC header is attached to outer IP header and then the frame is forwarded into VXLAN enabled fabric. This frame is carried over VXLAN fabric and delivered to another site. At other side all the various headers are removed and the packet if forwarded towards local lan. To understand this VXLAN encapsulated frame structure better, it will be good to have took at the VXLAN packet format shown below.

Figure 1: VXLAN Packet Format



VXLAN header is attached to Ethernet header. The main content of VXLAN header is VXLAN Network Identifier (VNID).

VXLAN Network Identifier (VNID)

VNID is similar to Virtual LAN (VLAN) id but VXLAN filed is 24 bits and VLAN ID filed is 12 bits. VLAN ID field is a part of TAG field in Ethernet Frame whereas VNID is part of VXLAN

header. The 24 bits VXLAN field allows us to create 16 million VXLAN segments whereas 12 bits VLAN id let us create 4096 VLANs (Del Piccolo, Amamou, Haddadou, & Pujolle, 2016).

VXLAN Tunnel End Point (VTEP)

When an ethernet frame needs to traverse VXLAN fabric, it is encapsulated using VXLAN encapsulation and send over the VXLAN tunnel. When the VXLAN encapsulated packet reaches remote site, the VXLAN encapsulation is removed and normal ethernet frame is forwarded down to the local LAN. This function of encapsulating and decapsulating VXLAN frames is done by a device called VXLAN Tunnel End Point (VTEP) (Wang & Lin, 2019).

VXLAN Layer 2 Gateway

Layer 2 gateway Function of VXLAN is for communication between traditional 802.1q tagged VLAN and associated VXLAN VLAN. When VXLAN Layer 2 gateway receives tagged packet, which is going into VXLAN fabric, layer 2 gateway will encapsulate it into VXLAN segment. Also, other way round when VXLAN encapsulated frame is received by layer 2 gateway it will decapsulate it and map to associated traditional VLAN (Dhodapkar, 2019).

VXLAN Layer 3 Gateway

Layer 3 gateway function is to route packet between one VXLAN segment into another VXLAN Segment. When a VXLAN encapsulated packet is received from VXLAN fabric and when it want to go to a VLAN not mapped to that ingress VXLAN VLAN, layer 3 VXLAN gateway functionality is used (Dhodapkar, 2019).

Underlay Network

Underlay network is the traditional IP based network using either BGP, OSPF, or EIGRP. VXLAN Tunnels are built on this underlay network.

Overlay

Overlay is the VXLAN tunnel that is built over traditional IP underlay network. Overlay network technology has been in use in various for many years to overcome the constraints and limitation of physical and traditional networks without replacing the network infrastructure already in place. Overlay based network are proved to be useful in many ways to support various network infrastructure requirements such as multicast, traffic engineering, resilient network (Rodriguez-Natal, Paillisse, Coras, Lopez-Bresco, Portoles-Comeras, & Cabellos-Aparicio, 2017).

Leaf

Leaf switch is the edge switch which connects traditional network with VXLAN Network. This device also functions as VTEP. To provide high availability network design, two leaf switches work together to form a cluster type of setting and the host can be dual connected to these two leaf switches using Link aggregation control protocol (LACP) and network interface card bonding (Wang & Lin, 2019).

Spine

Spine switch is the inside the core of the IP network. All the leaf switches connect to spine switches. The spine can be also configured as route reflector in case of iBGP designs.

VXLAN Control Plane

VXLAN needs some kind of mechanism to exchange control plane information. VXLAN control plane information includes discovering remote VTEP, exchanging host reachability information, exchanging broadcast, multicast, and unknown unicast traffic. VXLAN control plane information can be exchanged in two different ways. First method is by flooding and learning the information and second method is by exchanging this information in a controlled fashion using Multi-Protocol Border Gateway Protocol Ethernet Virtual Private Network (MP-BGP EVPN).

VXLAN Flood and Learn

Flood and Learn VXLAN implementation method are one of the data plane learning method for VXAN. In this method, broadcast, multicast and unknown unicast is flooded to associated multicast group. This method was not scalable and not efficient enough.

MP-BGP EVPN

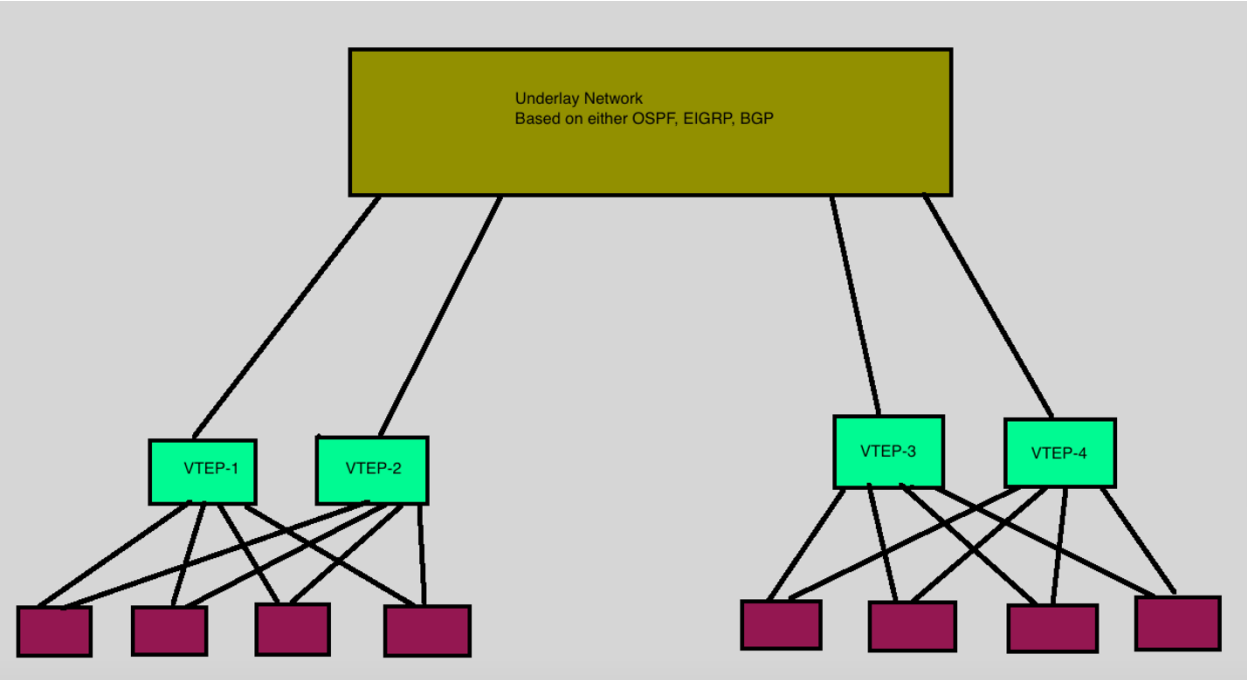
MP-BGP EVPN based method is efficient, scalable, standard based and more popular. In MP-BGP EVPN method, VXLAN control plane related updates are exchanged using MP-BGP L2VPN (Layer 2 Virtual Private Network) Address Family. EVPN MP-BGP based control plane helps to exchange protocol information distribution, discovery of VTEPs, end device discovery and connectivity (Gustavo, Salazar, Edison, Naranjo, & Luis, 2018). It is important to understand the concept of Overlay Network and Underlay Network. Underlay is basically traditional network which every organization has in place. That underlay network can be based of EIGRP, OSPF, ISIS, or BGP. VXLAN overlay Tunnel is formed over existing underlay network infrastructure between VTEP. VXLAN network can be connected with external non VXLAN based network using a VTEP that is acting as Border Leaf.

VXLAN Design Options

VXLAN can be deployed as a single pod, multi-pod, or Multi-site. Single Pod design is simplest and basic form of VXLAN implementation. In single pod design, entire data center network is put together in a single VXLAN Fabric. In multi-pod design, data center may be divided into multiple pods and those pods are connected together using VXLAN fabric. These pods can be local to a site or data center, can be dispersed over different data centers, or can be geographically separate locations. They are put together in single VXLAN fabric domain which will have different pods. Third design option is multi-side VXLAN. In this design every site will have their own VXLAN domain and then those two sperate VXLAN domains are connected together to form a big and integrated multi-site VXLAN fabric. Following figure 2. shows how single pod, multi-pod, and multi-site VXLAN domains looks like.

VXLAN Single Pod Design

Figure 2: VXLAN Single Pod

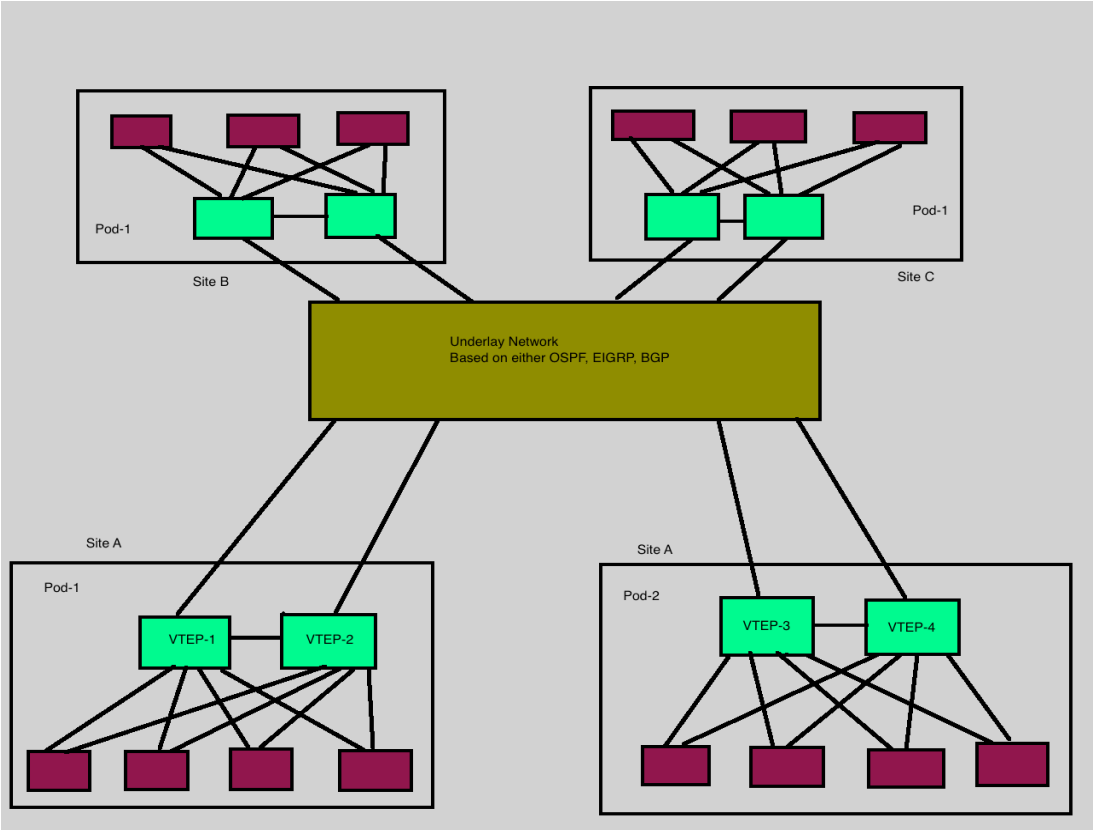


VXLAN single point of delivery (POD) design works better for smaller network and at the same time provide sufficient number of possible nodes. It is a single VXLAN fabric which is a typical small to mid-size network which has access layer, distribution layer and core layer which is a single domain within a datacenter. If the company is small size, it makes more beneficial to put them all in single pod. Also, devices belong to one type of business or devices performing similar types of tasks can be put together in same pods.

VXLAN Multi-Pod

For mid-size to large size company who has many different business units or many different network section or pods, can be separated into different pods and then they can be connected together in same VXLAN Domain using VXLAN Multi-Pod design. Multi-pod design is a form of modular design where you can keep different pods separate based on their function and then connect them together. All the benefits of modular network design apply to multi-pod design like failure in one domain stays local to that domain, ease of management etc. Following figure 3 shows how a multi-pod VXLAN design may look like.

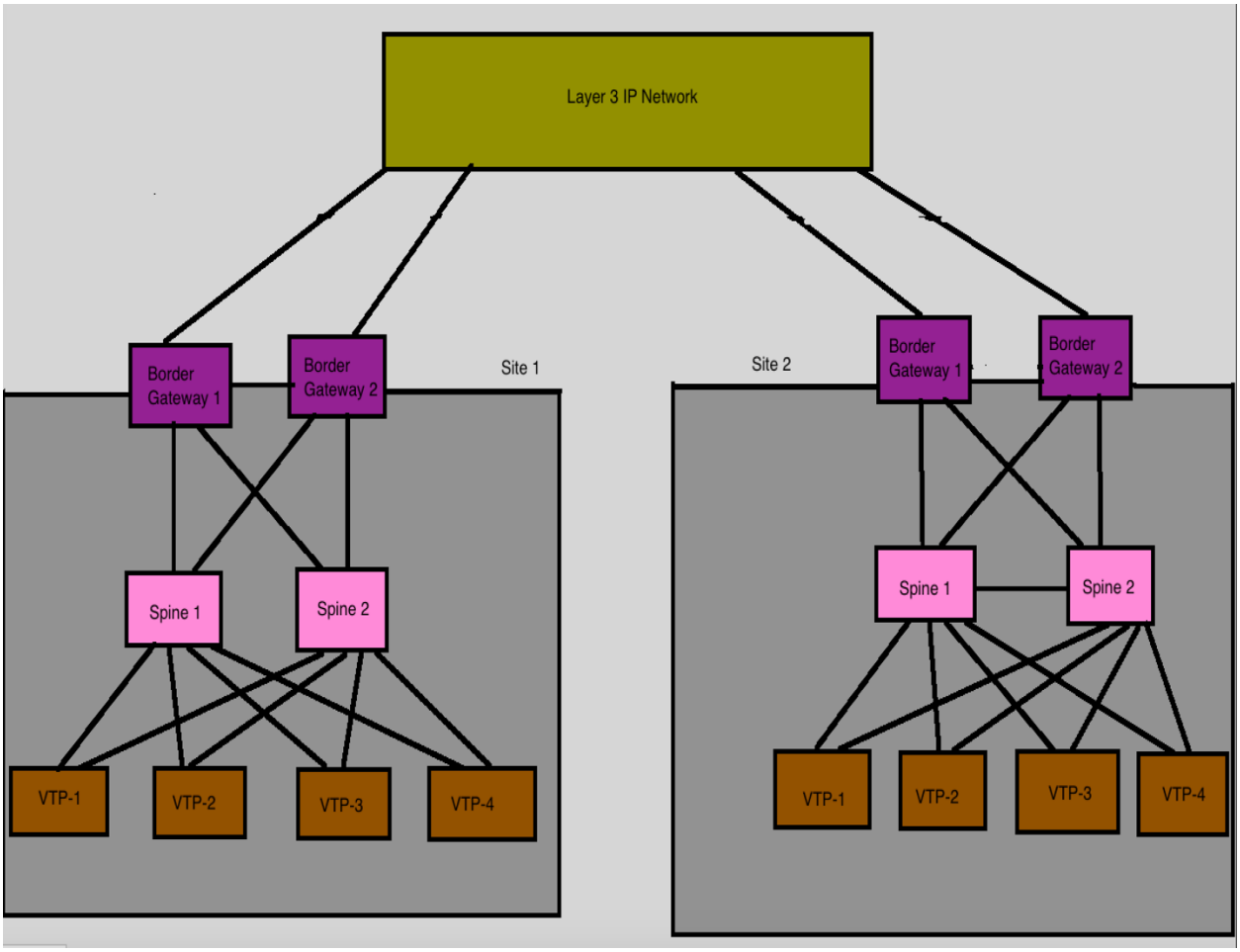
Figure 3: Multi-Pod VXLAN



Multi-Site VXLAN

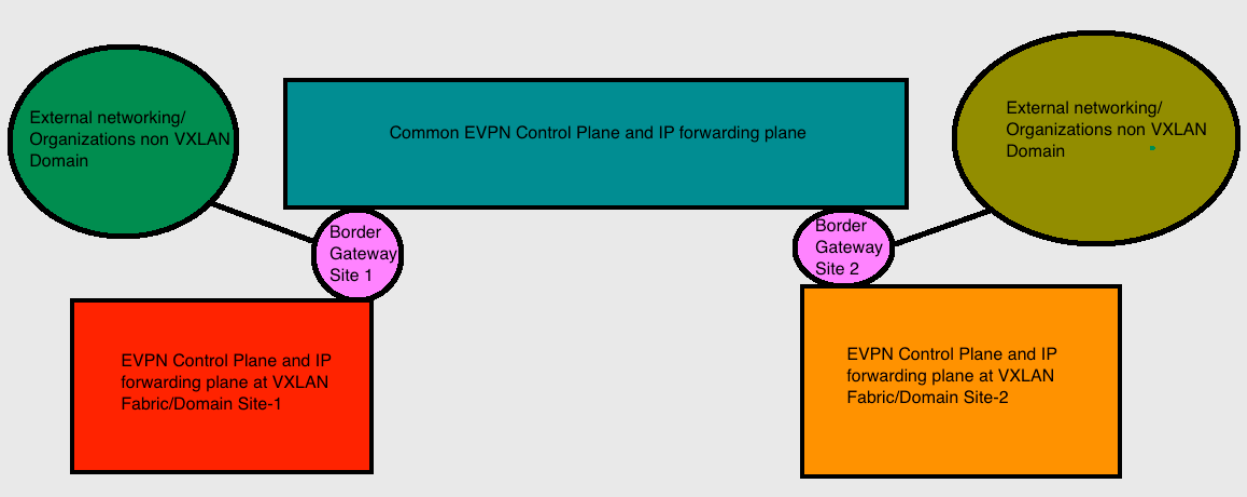
In a multi-site VXLAN, multiple VXLAN domains are connected together using Border Gateways. All the VTEPs in Domain 1 forms adjacency only with other VTEPs in the same domain and Border Gateway in the same domain. Each site will have one border gateway that will connect to other sites Border gateway. This research paper focuses on design, functionality, and implementation of Multi-Site Border Gateway in detail. Following figure 4 shows a typical Multi-Site VXLAN network look like.

Figure 4: Multi-Site VXLAN



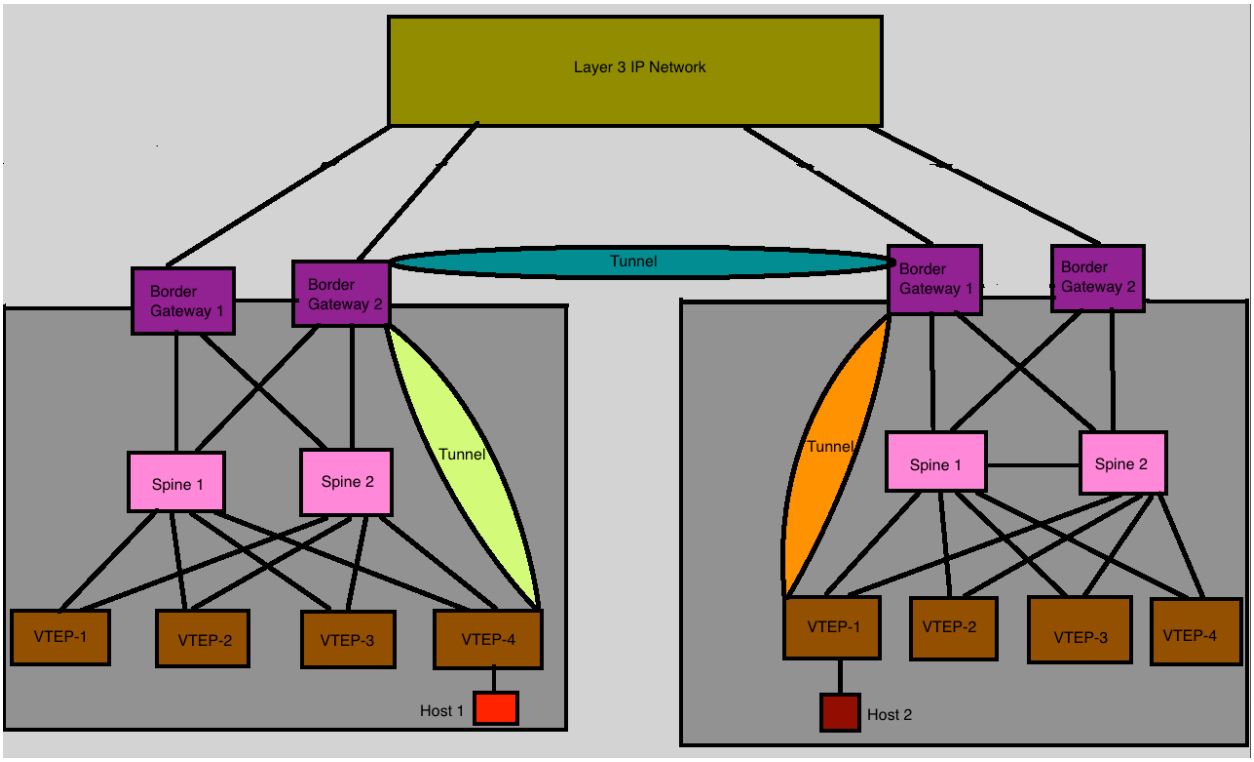
While implementing Border gateway, for redundancy purpose, Border gateway can be implemented in Virtual Port Channel (VPC) mode or anycast mode. For all devices in a VXLAN domain or fabric, both layer 2 and layer 3 external network is reachable via border gateway. Connecting multiple VXLAN domain or Fabrics together is similar to connecting multiple different local EVPN control planes and IP forwarding domains together via common EVPN control plane and IP forwarding domain as shown in the figure 5 below.

Figure 5: Control Plane and IP forwarding Plane



In multi-site VXLAN Fabric every node is assigned a unique site scope identifier. Border Gateway are part of common EVPN control Plane and IP forwarding Plane and local site's control plane and IP forwarding plane. When a device at site–1 wants to send traffic to device at site-2, the packet first comes to local VTEP at site 1. Site-1 VTEP will know this remote host via Border Gateway. The VTEP at site 1 initiate VXLAN Tunnel with Local Border Gateway at Site 1. When Border Gateway at Site-1 gets the packet, it does the lookup for destination host reachability and identifies that final destination is towards Border gateway 2. Border Gateway at Site 1 forms another VXLAN tunnel with Border Gateway at Site 2. When the Border Gateway at site-2 get the packet, it does another lookup and finds out the final destination is beyond local VTEP ate site 2, that time Border Gateway at site 2 forms another VXLAN tunnel with VTEP at site 2. When VTEP at site 2 get the packet, it does another lookup and identifies that the final destination is locally connected in traditional ethernet segment and it decapsulates the packet and send out towards classical ethernet domain. This packet flow is shown in figure 6 below.

Figure 6: Multi-site VXLAN Tunnel



Multi-site VXLAN Design Criteria

Deciding whether to use Single pod VXLAN deployment, or Multi-pod VXLAN deployment or multi-site VXLAN deployment is a very critical. Some of the criteria to consider are as follows.

High Availability: Multi-site VXLAN enables high availability within the data center (specific to local VXLAN domain) and across multiple data centers which are geographically separated (Cisco, 2017).

Distributed Application: It enables workload mobility and multi-tenancy. Application servers can be moved or migrated easily to another datacenter without changing IP address of the server and without impacting the applications and client services. Application servers can work in active-

passive state or active-active mode and provide disaster recovery capability and business continuity. In case of any natural disaster business can continue to run from other location without human intervention (Cisco, 2017).

Scalability: Multi-site VXLAN EVPN solution is more scalable overlay solution. Adding a new site or removing a site is just a matter of establishing connectivity with new Border Gateway.

Physical Connectivity: The availability of physical connectivity to connect the two different data center is first criteria that needs to be considered. Most of the time the physical connectivity is already in place. That time is required to check if the physical link can support VXLAN encapsulation and can it support required maximum transmission unit (MTU). VTEP should not fragment encapsulated VXLAN packet and if the path does not support required MTU size those packets will get dropped (Del Piccolo, Amamou, Haddadou, & Pujolle, 2016). Also, the design needs to take into consideration if different types of links that can be used. Sometimes only dark fiber links may be available, other places layer-3 direct links may be available (Cisco, 2017).

Business Requirement: If requirement of the business is such that it cannot tolerate any downtime then multi-site VXLAN helps achieve that design goal.

Fault Isolation: When multiple sites are connected together, problem in one site can easily affect other side. For example, if there is broadcast storm in one site, it can easily pass on to another site. In case of multi-site design broadcast domain is limited only up to border gateway (Cisco, 2017).

Cost saving: Multi-site VXLAN can be configured over existing layer-3 network to implement disaster recovery site without too much additional investments.

Multi-Site VXLAN EVPN Design Component

Multi-Site VXLAN has all the component same as Multi-pod VXLAN plus one additional component which is called Border Gateway which is used to connect two different VXLAN domains at two different sites. Following are various components of Multi-site VXLAN

VXLAN Tunnel Endpoint (VTEP): When VTEP receives packet from local LAN and final destination is beyond VXLAN fabric, VTEP looks for VLAN ID in incoming frame and assign associated VXLAN Network Identified (VNID). Then it encapsulates the frames and send it over VXLAN fabric. Remote VTEP decapsulates the frame and sends the Ethernet frame towards Local Lan (Dhodapkar, 2019).

Border Gateway: Border gateway is a device that interacts with devices within the same site and devices that at in another site. Border gateway function can be implemented on a leaf device, or spine device, or a special device can be implemented for Border gateway function. Border Gateways provide connectivity to external network (Non VXLAN) using VRF-Lite (Cisco, 2018).

Technologies related to VXLAN

MP-BGP

Multi-Protocol Border Gateway Protocol (MP-BGP) is an extension to Border Gateway Protocol (BGP). BGP only supports address family IPV4 Unicast. MP-BGP supports multiple address families like

1. IPv4 and IPv6 unicast and multicast
2. L2VPN – Layer-2 VPN. Example VPLS (Virtual Private Lan Services) or EVPN
3. VPN4 and VPN6 – Layer3 VPN. Example MPLS

To be able carry these multi-protocol information MP-BGP has following features and capabilities

1. Address Family Identifier (AFI): - This will specify which address family is being used. For Example, IPv4 unicast, IPv6-unicsast, l2VPN EVPN etc
2. Subsequent Address Family Identifier (SAFI) – This will specify more information about the address family. For Example, VPLS, BGP VPNs, SD-WAN Capabilities etc. (Nalawade, Cui 2019)
3. Multi-protocol reachable Network Layer Reachability Information (MP_UNREACH_NLRI) : Used to transfer unreachable network layer reachability information.
4. BGP Capabilities Advertisement: This is used to advertise the capabilities BGP or MP-MGP capabilities.

L2VPN

Layer 2 VPN provides end to end Layer 2 connection via Service providers MPLS or IP core network. Service provider forwards frames based on Layer 2 information and no need to route customer's packets based on layer 3 information. When L2VPN service is used customer has control over their layer 3 policies such as QOS, and routing policies. L2VPN consist of many services like Virtual Private Lan Services (VPLS), Virtual Private Wire Services (VPWS), Point to point layer 2 VPN, IP-Only L2 VPN, Ethernet VPN (EVPN) etc. (Bitar, Heron, Farren, Zinin, 2017).

EVPN

Ethernet Virtual Private Network is one of the ways to implement L2VPN. EVPN is a standard based technology that connects different layer-2 domains over Multiprotocol Label Switching (MPLS) network or IP core network. EVPN evolved from family of MPLS based Layer 2 Virtual Private Network (L2VPN) and L3VPN solution which provides much better support in ISP based environment supporting multiple customer (Makowski & Grosso, 2019).

MP-BGP L2VPN EVPN

L2VPN EVPN is an address family of MP-BGP that is used as control plane for VXLAN for layer 2 reachability information, exchanging IP to MAC information, end host reachability information, and VTEP discovery. MP-BGP EVPN is an industry standard control plane protocol for VXLAN. The prior version of VXLAN which is also called flood and learn used data plane-based method to discover VTEP, exchange host reachability information, and exchange IP-to-MAC information which was not scalable enough.

The benefits of using MP-BGP EVPN as control plane protocols are as follows

1. MP-BGP EVPN protocol is standard based and works for all vendors platforms.
2. Learning host reachability information and VTEP discovery is control plane based which makes MP-BGP EVPN based VXLAN more scalable and robust.
3. L2VPN EVPN address family carries layer2 and layer3 reachability information and so it can provide integrated bridging and routing using VXLAN.
4. ARP related information is exchanged using control plane and hence reducing network broadcast.

5. MP-BGP EVPN enables configuring same gateway IP address called anycast gateway IP address on multiple VTEPs across fabric.

MP-BGP Route Types

MP-BGP uses 5 types of route types for IP prefix and other advertisements (Dhodapkar, 2019).

1. Type 1 Ethernet Auto Discovery route
2. Type 2 Route update Contains Layer 2 VNI, MAC and IP, ARP resolution information
3. Type 3 Route type contains EVPN Ingress Replication information
4. Type 4 Route is for ethernet Segment
5. Type 5 Route type contains IP prefix route information.

VXLAN Multi-Site Design and Deployment

Multi-Site VXLAN in simple words means if any organization has multiple data center sites and each of that Data Center site has its own local VXLAN fabric. Now when connecting those two different VXLAN fabric site using EVPN Multi-site architecture, it is called VXLAN Multi-site. Different networking vendors may call this architecture by different names, but the idea remains same to connect multiple VXLAN fabrics together. Cisco call this architecture VXLAN multisite and Arista call this architecture Data Center Interconnect. The VXLAN multi-site study in this paper is based on Cisco's implementation of VXLAN to connects multiple VXLAN fabric. Multi-

site EVPN based VXLAN using Border Gateway is based on IETF draft-sharma-multi-site-evpn (Sharma, Banarjee, & Sivaramu, 2016).

Placement of Border Gateway

The Border gateway can be located in two different possible places. One possible location is at the leaf layer where dedicated border gateway can be placed. Other possible location to place border gateway is co-located with Spine layer. When a border gateway is co-located at spine layer it performs multiple functions like Route Reflector, external connectivity function, east to west traffic flow, Rendezvous Point (Cisco, 2018)

Anycast Border Gateway

Anycast Border Gateway shares same virtual IP between two or more border gateways. Adding multiple Border Gateways allows scaling out the capacity without a lot of interdependency between the Border Gateways. This Virtual IP address is referred as Anycast IP address. Anycast gateway IP address is used as source IP address in the outer header when communication is happening within the site or between the sites. When the anycast gateway IP, address is used traffic load balancing is possible using Equal Cost Multipath (ECMP) mechanism. This avoids polarization, supports load balancing of traffic and increases resiliency (Cisco, 2018).

Primary VTEP IP (PIP)

For the purpose of forwarding broadcast, multicast, and unknown unicast traffic (BUM) between different sites, Border Gateways uses their unique IP address called the primary VTEP IP (PIP) address. The PIP address is advertised within the site and between the sites as it needs to be used for broadcast, multicast and unknown unicast communication. Border Gateway uses the PIP addresses for BUM replication either in the multicast underlay or when advertising BUM traffic

using ingress replication method using BGP EVPN type 3 route. The PIP address is also used by Border Gateway to advertise externally learned network prefixes (Non VXLAN) into VXLAN fabric. Border Gateways are connected to external network using VRF-Lite and they redistribute those external Prefixes into BGP EVPN. Routes learned using OSPF, EIGRP, Static routing, or using any other dynamic routing protocol method can be redistributed into BGP EVPN. When the devices in the VXLAN fabric our to reach out to those external prefixes, the next hop for next is the PIP addresses which is located on Border gateway. Once Border Gateway receives those packets it can send them to outside world (Cisco, 2018).

Designated Forwarder

When multiple Border Gateways are implemented Broadcast, Multicast, and unknown unicast traffic replication function is distributed among the Border Gateways on round robin basis. One Border Gateway will be designated for half of the VNIs and another Border Gateway will act as designated forwarder for remaining half VNI. The designated border gateway will synchronize with another border gateway using BGP EVPN Route Type 4 (Cisco, 2018).

Border Gateway Failure Scenarios

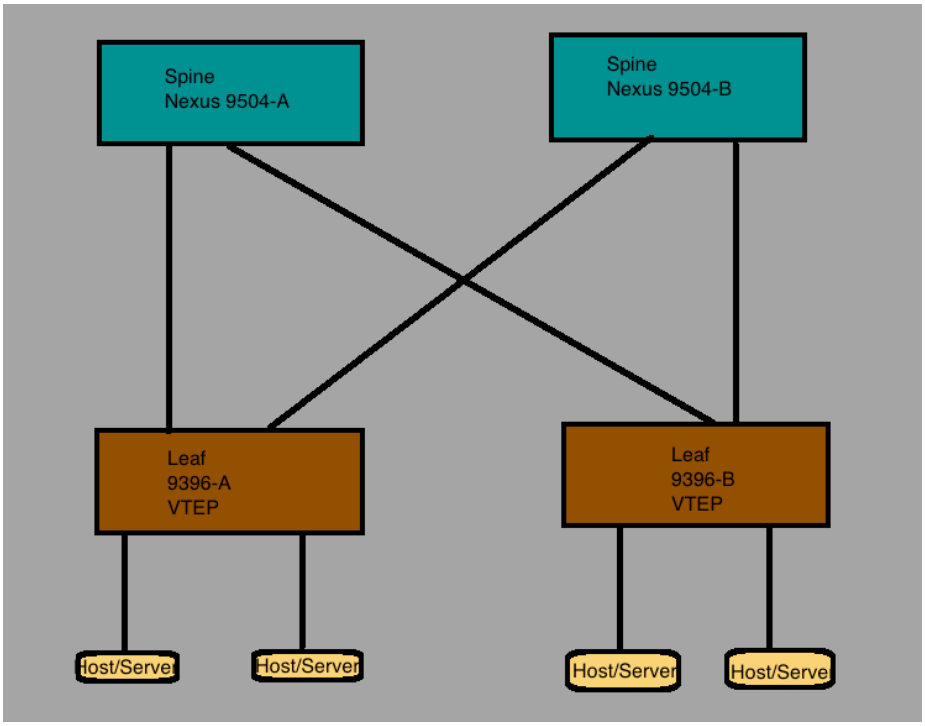
When the Border Gateway links going to the internal network such as Spines or other VTEPS are going down or links going to external network are going down, that Border Gateway will withdraw the routes advertised by it to others. It will stop advertising the its virtual IP external network, it will withdraw all the route type 2, route type 3. Route type 4 and route type 5) those were advertised by that border gateway. When all the routes advertised by the failed border gateway are advertised it gets isolated from the network and remaining Border gateways takes over the functionality (Cisco, 2018).

Brief Overview of VXLAN Configuration

In this section brief overview of VXLAN configuration is discussed. To explain BGP EVPN VLXAN Multi-site configuration, it is required to quick overview of BGP EVPN VXLAN single site configuration. BGP EVPN VXLAN Multi-site configuration is all about how to join the multiple sides. The VXLAN configuration steps discussed below are Cisco Nexus platform (Lalith, 2018)

Following network topology shown in figure 7 is used to discuss the configuration.

Figure 7. VXLAN Topology used for configuration of VXLAN single site



Sample VTEP configuration

Step 1 : In this step enable required features on Nexus. Specifically, OSPF for underlay network, PIM protocol for multicast core, interface Vlan feature for Switched Virtual Interface (SVI)

configuration, vn-segment-vlan-based for mapping VLAN to VNI, nv overlay for VXLAN

overlay

```
nv overlay evpn
feature ospf
feature bgp
feature pim
feature interface-vlan
feature vn-segment-vlan-based
feature lacp
feature vpc
feature nv overlay
```

Step 2 : In this step Map Layer 2 VLANs to VNID, define anycast gateway mac, Layer 3 SVI

(SVI 144 and SVI 145) and associate Tenant (SVI 144, 145 Customer' SVI) VRF to the SVI,

and define Layer 3 VNI for inter VNI communication.

```
#### Configure this command mac address on all VTEP for VM mobility
fabric forwarding anycast-gateway-mac 0005.0005.0005
!
ip pim rp-address 130.130.130.1 group-list 224.0.0.0/4
!
ip pim ssm range 232.0.0.0/8
!
vlan 1,10,30,40,100,200
!
########## Map Layer 3 100 VLANs to VNID 10000100
vlan 100
name L3-VNI-VLAN-10
vn-segment 10000100
!
########## Map Layer 2 VLANs to VNID 10000300
vlan 303
vn-segment 10000300
!
### Layer 3 VRF used for Inter-VNI traffic
vrf context EVPN-L3-VNI-100
vni 10000100
rd auto
address-family ipv4 unicast
route-target both auto
route-target both auto evpn
```

```
!
## This Interface is for Layer 3 VNI and does not need IP address
interface Vlan100
 no shutdown
 vrf member EVPN-L3-VNI-100
 ip forward
!
### Defining Host SVI and associating it to Layer 3 VNI
interface Vlan300
 no shutdown
 vrf member EVPN-L3-VNI-100

 ip address 72.6.30.1/24
 fabric forwarding mode anycast-gateway
```

Step 3 Configure underlay routing protocol and overlay routing protocol on VTEP. This example

uses OSPF as underlay and BGP L2VPN EVPN as overlay

```
router ospf UNDERLAY

!
router bgp 65000
 address-family ipv4 unicast
 address-family l2vpn evpn
 neighbor 192.168.9.9
 remote-as 65000
 update-source loopback2
 address-family ipv4 unicast
 address-family l2vpn evpn
 send-community extended
 vrf EVPN-L3-VNI-100
 address-family ipv4 unicast
 advertise l2vpn evpn
!
evpn
 vni 10000300 l2
 rd auto
 route-target import auto
 route-target export auto
```

Step 4 : Define NVE interface which a the logical interface used for VXLAN and it encapsulates

and decapsulates VXLAN Packets.

```
interface nve1
no shutdown
source-interface loopback2
#### Specify BGP control plane is used to exchange updates.
host-reachability protocol bgp
### associate-vrf is used for for layer3 vni.
member vni 10000100 associate-vrf
member vni 10000300
suppress-arp
mcast-group 239.1.1.10
!
interface Ethernet1/4
 description "Going to Spine"
 no switchport
 ip address 92.68.19.1/24
 ip router ospf UNDERLAY area 0.0.0.0
 ip pim sparse-mode
 no shutdown
!
interface Ethernet1/16
Description Interface to Host A.
 switchport mode trunk
!
interface loopback2
Description for BGP Peering.
ip address 92.68.11.11/32
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
```

Sample Spine Configuration

```
nv overlay evpn
feature ospf
feature bgp
feature pim
feature interface-vlan
feature vn-segment-vlan-based
feature lacp
feature vpc
feature nv overlay
```

```
!
ip pim rp-address 92.68.9.9 group-list 224.0.0.0/4
!
ip pim ssm range 232.0.0.0/8
!
interface Ethernet1/11
 ip address 92.68.19.9/24
 ip router ospf UNDERLAY area 0.0.0.0
 ip pim sparse-mode
 no shutdown
!
interface Ethernet1/6
Description Connection to VTEP
ip address 192.168.29.9/24
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
no shutdown
!
interface Ethernet1/9
Description Connection to VTEP
ip address 192.168.39.9/24
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
no shutdown
!
### Condfigure Spine as Rendezvous Point
interface loopback1
ip address 92.68.9.9/32
ip router ospf UNDERLAY area 0.0.0.0
ip pim sparse-mode
!
router ospf UNDERLAY
!
router bgp 65000
 log-neighbor-changes
 address-family ipv4 unicast
 address-family l2vpn evpn
 retain route-target all
 template peer VTEPPEERS
 remote-as 65000
 update-source loopback1
 address-family ipv4 unicast
 send-community both
 route-reflector-client
 address-family l2vpn evpn
 send-community both
```

```
route-reflector-client
neighbor 92.68.11.11
inherit peer VTEPPEERS
neighbor 92.68.22.22
inherit peer VTEPPEERS
```

VXLAN BGP EVPN Multi-site sample configuration

Step 1: Configure VXLAN site ID on the Border Gateway

This example shows how to configure site it on a Cisco Nexus Switch

Border-Gateway#config terminal

Bortder-Gateway(config)#evpn multi-site border-gateway 100

Step 2 : Create VXLAN Interface and define BGP as host reachability advertisement protocol

Bortder-Gateway(config)#Interface nve1

Border-Gateway(config-if)#host-reachability protocol bgp

Conclusion:

VXLAN multi-site EVPN deployment is fairly new technology which being adopted at very fast rate because of its many advantages. VXLAN multi-site fabric provides hierarchal topology and multiple overlay domains. Its benefits include scalability, resiliency, loop prevention, transport media independence, open standard, failure containment, and it is flexible and independent. As the use of this technology grow, it will gain more maturity at technology level and also at the

industry expertise level. Since it is open standard based, and interoperable with different vendors, different transport media and various routing protocol, it has very high chance of industry acceptance.

Reference:

Nalawade, G., & Cui, Y. (2019). Subsequent Address Family Identifiers (SAFI) Parameters. Retrieved from https://www.iana.org/assignments/safi-namespace/safi-namespace.xhtml

Naranjo, E. F., & Gustavo, D. S. C. (2017;2018;). Underlay and overlay networks: The approach to solve addressing and segmentation problems in the new networking era: VXLAN encapsulation with cisco and open source networks. Paper presented at the, 2017- 1-6. doi:10.1109/ETCM.2017.8247505

Singh, T., Jain, V., & Babu, G. S. (2017). VXLAN and EVPN for data center network transformation. Paper presented at the 1-6. doi:10.1109/ICCCNT.2017.8203947

B. Buresh, D. Eline, D. Jansen, J. Gmitter, J. Ostermiller, J. Moreno, K. Lei, L. Quan, L. Krattiger, M. Ardica, R. Parameswaran, R. Tappenden, S. Kondalam, A modern open and scalable fabric VXLAN EVPN, 2017

D. Jansen, L. Krattiger, Building Data Centers with VXLAN BGP EVPN: A Cisco NX-OS Perspective, San Jose, California:Ciscopress, April 2017.

M. Mahalingam, D. Dutt, Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlay Virtualized Layer 2 Networks over layer 3 Networks, vol. RFC7348, 2014

Bitar, N., Heron, G., Farren, A., & Zinin, A. (2017). Layer 2 Virtual Private Networks (l2vpn). Retrieved from https://datatracker.ietf.org/wg/l2vpn/about/

Dhodapkar, S. (2019). Troubleshooting BGP EVPN. Retrieved from https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2019/pdf/BRKDCN-3040.pdf

Sharma, R., Banarjee, A., & Sivaramu, R. (2016). Multi-site EVPN based VXLAN using Border Gateways. Multi-Site EVPN Based VXLAN Using Border Gateways. Retrieved from https://tools.ietf.org/html/draft-sharma-multi-site-evpn-00

Varma, L. (2018). Configuration and Verification VXLAN with MP-BGP EVPN Control Plane. Retrieved from https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/200952-Configuration-and-Verification-VXLAN-wit.html

Del Piccolo, V., Amamou, A., Haddadou, K., & Pujolle, G. (2016). A survey of network isolation solutions for multi-tenant data centers. IEEE Communications Surveys & Tutorials, 18(4), 2787-2821. doi:10.1109/COMST.2016.2556979

Dai, B., Xu, Y., Xu, G., Huang, B., & Qin, P. (2017). Enabling network innovation in data center networks with software defined networking: A survey. Journal of Network and Computer Applications, 94, 33-49. doi:10.1016/j.jnca.2017.07.004

Wang, Y., & Lin, Y. (2019). Circuit-based logical layer 2 bridging in software-defined data center networking. International Journal of Communication Systems, 32(16), e4128-n/a. doi:10.1002/dac.4128

Rodriguez-Natal, A., Paillisse, J., Coras, F., Lopez-Bresco, A., Jakab, L., Portoles-Comeras, M., . . . Cabellos-Aparicio, A. (2017). Programmable overlays via OpenOverlayRouter. IEEE Communications Magazine, 55(6), 32-38. doi:10.1109/MCOM.2017.1601056

Makowski, Ł., & Grosso, P. (2019). Evaluation of virtualization and traffic filtering methods for container networks. Future Generation Computer Systems, 93, 345-357. doi:10.1016/j.future.2018.08.012

Ricart-Sanchez, R., Malagon, P., Salva-Garcia, P., Perez, E. C., Wang, Q., & Alcaraz Calero, J. M. (2018). Towards an FPGA-accelerated programmable data path for edge-to-core communications in 5G networks. Journal of Network and Computer Applications, 124, 80-93. doi:10.1016/j.jnca.2018.09.012

Cisco, C. (2018). VXLAN EVPN Multi-Site Design and Deployment. *VXLAN EVPN Multi-Site Design and Deployment*. Retrieved from https://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/white-paper-c11-739942.pdf

Cisco, C (2017). Building Hierarchical Fabrics with VXLAN EVPN multi-site. Retried from www.cisco.com/c/dam/en/us/products/collateral/switches/nexus-9000-series-switches/at-a-glance-c45-739422.pdf