Integration of Data Center Network Technologies VxLAN, BGP, EVPN

Adriana-Elena Rădoi[1]
[1]Department of Communications and Military Electronic Systems, Military Technical Academy, Bucharest, Romania
Contact author e-mail: radoi.adriana@gmail.com

Cristian-Iulian Rîncu[1]
[1]Department of Communications and Military Electronic Systems, Military Technical Academy, Bucharest, Romania
Contact author e-mail: iulian.rincu@gmail.com

***Abstract* — The implementation of network overlays, such as Virtual eXtensible Local Area Network (VXLAN) as the data plane and Border Gateway Protocol/Ethernet VPN (BGP/EVPN) as the control plane for virtual network distribution[5], plays an important role in the integration of VXLAN BGP EVPN technologies in modern Data Centers, allowing a better flow control over the network. With this regard, the enhancement of dedicated networks, and, furthermore, of switching and routing capabilities has been achieved, thereby reducing the costs of device-by-device implementation and management towards a more centralized, integrated system. This paper presents the new network architecture, along with its related protocols, not only for control plane, but also for data plane, demonstrating the need to replace the traditional network architectures and protocols, such as Spanning Tree Protocol and vPC.**

Keywords — Virtual eXtensible Local Area Network, Border Gateway Protocol, Ethernet Virtual Private Network, Overlay, Underlay, Spanning Tree Protocol, Data Center, Spine-Leaf

I. INTRODUCTION

The technological progress and its accelerated spreading have imposed the interconnection of computerized systems all over the world, through the implementation of communication systems, in order to provide network connectivity between various research and development centers. Thus, there have been created specialized spaces in the interior of buildings, in order to house computer systems and their associated hardware components, bearing the name of "Data Centers". When talking about the amount of integrated technologies in Data Centers, one of the most important aspects to be taken into consideration is not only the solutions of delivering data traffic in a fast and efficient way, but also the management of large, scalable networks, using different network pieces of equipment, bought from different vendors.

Early implementation of data center network technologies leverage traditional loop prevention techniques such as Spanning Tree Protocol (STP), which not only imposes restrictions on network design and resiliency, but it also results in an inefficient use of available network links due to the blocking of redundant paths[6]. It likewise brings up a series of disadvantages such as the increased convergence time of network groups of pieces of equipment, or limitations in terms of scalability in the network topology. Taking these factors into consideration, the development of data centers led to the emergence of virtual Port Channel (vPC) or FabricPath technologies, which have proven not to meet the requirements for the expansion of data centers, not only in terms of the number of configured pieces of equipment which can be interconnected, but also of their various platforms, provided by different vendors. Therefore, the integration of VXLAN BGP Ethernet VPN technologies in a Spine-Leaf network hierarchical topology, organised on two physical layers, has represented a solution for the unification of data centers, the novelty element of this type of implementation being given by the combination of a physical underlay network with a virtual overlay one. The fundamental requirements of such an implementation are not only scalability, but also security, the optimization of waiting time in order to fulfill an application request and virtual machines (VMs) mobility through the network topology. Moreover, the configuration of modern data center solutions has imposed the use of automatized processes in order to set the necessary resources by nowadays needs.

The goal of this article is to analyze network performance and packets behaviour, following the process of encapsulation, respectively decapsulation, regarding the integration of Virtual eXtensible Local Area Network (VXLAN), Border Gateway Protocol (BGP) and Ethernet Virtual Private Network (EVPN) in a modern Data Center network topology, in order to emphasise the benefits of this solution and the improvements that it can bring to network approaches, from an administrative domain perspective.

This article describes, not only from a theoretical point of view, but also from a practical one, the integration of data center network technologies, through the implementation and analysis of a Spine-Leaf network topology, using GNS3 environment. The addresed scenarios are not only based on the deployment of three data centers through different configuration methods, combining specific protocols and technologies, such as OSPF, BGP, EVPN, VXLAN, MPLS, VRF, but they also rely on the implementation of data centers interconnection solutions, such as the configuration of Type 5 EVPN Routes. Their main objective is to highlight the advantages of implementing modern data centers, in comparison with the limitations of traditional network infrastructure. Various aspects, such as the implementations of specific solutions for building virtual networks, will be covered, taking into consideration the increased redundancy within such networks, which allows redirecting data traffic to other routes in a short period of time, by using other available pieces of equipment, in the event of damage of their embedded systems.

2022 14th International Conference on Communications (COMM) | 978-1-6654-9485-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/COMM54429.2022.9817218

II. VXLAN BGP WITH ETHERNET VPN

A "Spine-Leaf" network architecture allows the use of all transmission links between pieces of network equipment situated in different physical layers, thus providing redundancy and the possibility of implementing the routing strategy called Equal Cost Multipath (ECMP), characteristic to the configuration of a dynamic routing protocol within the network topology.

An "overlay" network is a virtual network built by means of a logical tunnel, implemented over the infrastructure of a physical one, the "underlay", defined by the configuration of a dynamic routing protocol, which executes the exchange of information between the layer three switches in the topology.

A. *VXLAN – Virtual eXtensible Local Area Network*

Within a VXLAN network topology, each device builds its own database, containing information corresponding to the locally connected VMs. Packets are transmitted through a virtual tunnel, called VXLAN segment, identified by a specific 24-bit field, called Virtual Network Identifier (VNI). Each VNI ensures the isolation of traffic associated to a VXLAN segment. The VXLAN header encapsulation process can define a Data Link layer tunneling scheme over a Network layer infrastructure, each end of the logical tunnel being generically called Virtual Tunnel Endpoint (VTEP).

B. *BGP – Border Gateway Protocol*

The solution of integrating the technologies VXLAN, BGP and EVPN is particularly based on the configuration of Multiprotocol BGP (MP-BGP) in order to transport specific reachability information for separate tenants. In accordance with this statement, External BGP is used within the Underlay network to allow devices, located in different autonomous systems (AS), to make their IP addresses on the loopback interface over which the MP-BGP session is established, known to each other. The Overlay networks works as a single AS, thus achieving accessibility information exchange in the network topology.

C. *EVPN – Ethernet Virtual Private Network*

The configuration of EVPN technology allows a classification of the routes between network devices, according to the parameters they posses. Therefore, there are EVPN type 1 routes, corresponding to an Ethernet segment, type 2 MAC/IP routes, associated with the learning of the MAC addresses of VMs, type 3 EVPN routes, characterising of learning the IP addresses of the VMs and, nevertheless, type 5 EVPN routes, corresponding to the implementation of Data Center Interconnection. EVPN eliminates the need of the traditional flood and learn mechanisms, replacing it with the implementation of „Address Resolution Protocol (ARP) Suppression", which gave the devices the possibility of storing the learned MAC addresses of other directly connected devices.

EVPN transports information between source and destination VTEPs within a VXLAN, using MP-BGP formatting to identify reachability of specific IP and MAC address destinations behind a VTEP.[1]

A control plane protocol defines how devices exchange topology information and make best path decisions. This, in turn, influences data plane behaviour, which determines how data packets are transported over a network path.[2] In the absence of a control plane protocol, topology information is exchanged according to the data plane protocol, similar to the "Flood and Learn" process, each VTEP performing its own remote MAC address learning while forwarding Broadcast, Unknown Unicast and Multicast (B.U.M) traffic. EVPN uses MP-BGP to let VTEPs advertise and learn each other's MAC addresses, stored in their MAC address table, behaving like a control plane technology.

III. NETWORK SPACE IMPLEMENTATION

In order to implement the three virtual Data Centers, integrating the presented technologies, we will use the network software emulator GNS3, as it offers the possibility of simulating the performances of physical pieces of equipment in a virtual environment, efficiently using the server resources on which the software is installed.

The implemented topology consists of three Data Centers, with a symmetrical architecture, consisting of two Spine switches and three Leaf switches, implemented with virtual QFX switches, running the Junos Operating System, consisting of a vQFX Routing Engine and a vQFX Packet Forwarding Engine, a virtualised version of the QFX10000, using the RAM Memory and the CPU of the GNS3 Server. The *first scenario* exemplifies a centrally routed EVPN VXLAN environment, using the dynamic routing protocol eBGP in order to establish the communication between network devices. The *second scenario* combines an edge routed bridging environment with the configuration of eBGP as a dynamic routing protocol. The Spine devices participate only in the dynamic routing process, while the Leaf devices perform the encapsulation and decapsulation process with the VXLAN header. The *third instance* presents the configuration of OSPF, used to interconnect network devices through their virtual Loopback interfaces in a centrally routed EVPN VXLAN environment.

To allow the exchange of EVPN routes between VTEP devices, we configure iBGP within the Overlay network, all of the equipments being part of the same autonomous system (for example: 65000). The source IP for all of the initiated iBGP sessions is the one on the virtual Loopback interface. The Underlay network is using eBGP as a dynamic routing protocol for the first and the second Data Center, while the third topology is using OSPF, the specific configuration being highlighted in Figure 2. The routing table resulting from the configuration of iBGP and eBGP in a single topology is also illustrated below in Figure 1.

```
Spine1.1# run show bgp summary
Threading mode: BGP I/O
Groups: 2 Peers: 6 Down peers: 0
Peer                     AS          InPkt        OutPkt       Last  Up/Dwn
State|#Active/Received/Accepted/Damped...
172.16.10.1              65201          69            70      28:24 2/3/3/0
0/0/0/0
172.16.10.3              65202          72            69      27:53 2/4/4/0
0/0/0/0
172.16.10.5              65203          70            69      28:41 2/3/3/0
0/0/0/0
192.168.100.11           65000          18            17       7:24 Establ
  bgp.evpn.0: 0/0/0/0
192.168.100.12           65000          11             9       3:46 Establ
  bgp.evpn.0: 0/0/0/0
192.168.100.13           65000           6             4       1:48 Establ
  bgp.evpn.0: 0/0/0/0
```

Figure 1. EBGP table corresponding to the Underlay and Overlay networks

```
{master:0}[edit protocols ospf]
Leaf3.1# show
area 0.0.0.0 {
    interface lo0.0;
    interface xe-0/0/1.0;
    interface xe-0/0/2.0;
    interface xe-0/0/0.0 {
        passive; } }
```

Figure 2. OSPF configuration in the Underlay Network

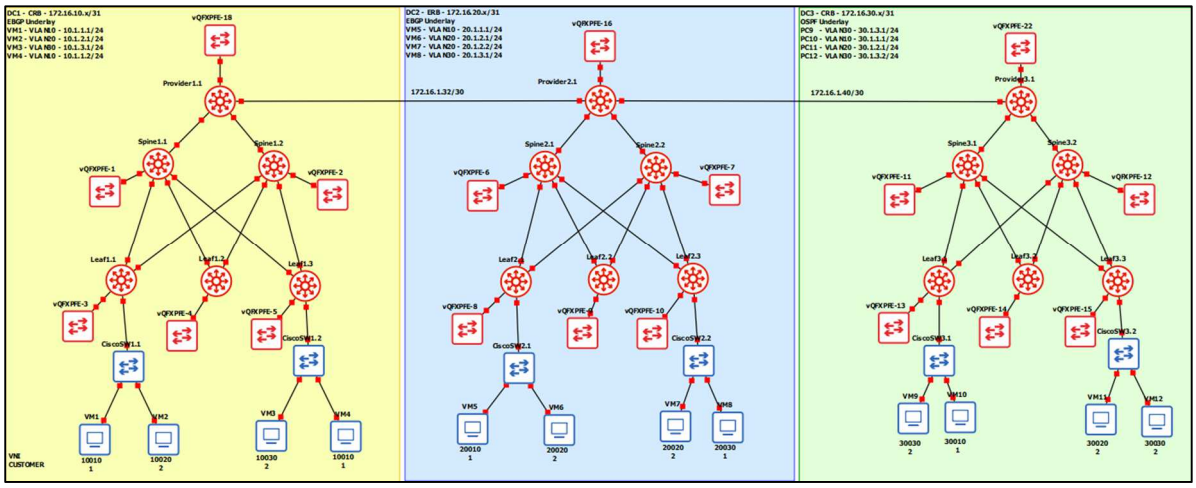

Figure 3. The topology used for simulation in GNS3

The topology implemented in GNS3, used to integrate the three Data Center technologies: VXLAN for the data plane, BGP and EVPN for the control plane management of the virtual network distribution and to analyze packet behaviour in a Spine-Leaf topology, based on the architecture selected for the routing process between different Virtual Network Identifiers, is illustrated above in Figure 3.

The configuration of VXLAN technology is associated with mapping each configured VLAN to a specific VNI, creating VXLAN segments, as illustrated in Figure 4.

```
{master:0}[edit vlans]
Leaf2.1# show
v10 {
    vlan-id 10;
    vxlan {
        vni 10010; } }
v20 {
    vlan-id 20;
    vxlan {
        vni 10020; } }
v30 {
    vlan-id 30;
    vxlan {
        vni 10030; } }
```

Figure 4. Mapping VLANs to the corresponding VNIs

Furthermore, the implementation of a Distributed IP Anycast Gateway will be accomplished using two methods, as follows: configuring network-level functionalities within Spine and Leaf devices in a Centrally Routed Bridging (CRB) environment or within Leaf devices in an Edge Routed Bridging (ERB) Overlay. When a host cannot resolve the IP-to-MAC mapping when trying to communicate with another endpoint situated in a different subnet, it initiates an ARP request for the IP address of its default gateway, being the IP anycast gateway configured on the most appropriate VTEP. The information from the ARP Snooping process executed on this VTEP is retrieved, thus populating the BGP EVPN control protocol. The configuration steps for Ethernet VPN are mainly shown below in Figure 5.

```
{master:0}[edit switch-options]
Spine1.1# show
vtep-source-interface lo0.0;
route-distinguisher 192.168.100.1:1;
vrf-target {
    target:65000:1;
    auto;}
{master:0}[edit routing-options]
Spine1.1# show
router-id 192.168.100.1;
autonomous-system 65000;
forwarding-table {
    export Load-Balance-Policy;}
```

Figure 5. Ethernet VPN configuration

The technique called Integrated Routing and Bridging (IRB) allows a protocol to perform both routing and switching functions on a virtual interface of a Layer 3 switch. We assign the configured VLANs to the appropriate IRB interface so that the packets can be forwarded to the default gateway and, then, to another VNI, without using a physical router. In order for the traffic to be routed between different VLANs, we create units with IP addresses associated with the VLANs. Thefore, the L3 switch automatically creates direct routes to the subsequent subnets and uses them to forward the traffic suitably. The configuration of the IRB interfaces are illustrated below in Figure 6.

```
{master:0}[edit interfaces irb]
Spine1.1# show
unit 10 {
    virtual-gateway-accept-data;
    family inet {
        address 10.1.1.101/24 {
            virtual-gateway-address 10.1.1.254;
        }
    }
    virtual-gateway-v4-mac 00:00:a3:d9:00:01; }
unit 20 {
    virtual-gateway-accept-data;
    family inet {
        address 10.1.2.101/24 {
            virtual-gateway-address 10.1.2.254;
        }
    }
    virtual-gateway-v4-mac 00:00:a3:d9:00:01; }
unit 30 {
    virtual-gateway-accept-data;
    family inet {
        address 10.1.3.101/24 {
            virtual-gateway-address 10.1.3.254;
        }
    }
    virtual-gateway-v4-mac 00:00:a3:d9:00:01; }
```

Figure 6. The configuration of IRB Interfaces

In the CRB Overlay we configure VLANs at the leaf devices, and IRB interfaces for routing at the spine devices, whereas in the ERB Overlay routing is performed at an edge location of the topology, namely at the Leaf devices.[3] [4]

The implementation of VRF technology, shown in Figure 7, strongly suggest a specific method of traffic separation between traffic coming from tenants VMs, thus contributing to the security of the topology we are currently analyzing, allowing the configuration of logically separated routing instances within a single physical equipment. This likewise allows the implementation of a routing table specific to each configured VRF.

```
{master:0}[edit routing-instances]
Spine1.1# show
customer1 {
    instance-type vrf;
    interface irb.10;
    interface lo0.1;
    route-distinguisher 192.168.100.20:1;
    vrf-target target:65000:1; }
customer2 {
    instance-type vrf;
    interface irb.20;
    interface irb.30;
    interface lo0.2;
    route-distinguisher 192.168.100.21:1;
    vrf-target target:65000:1; }
```

Figure 7. The configuration of VRF technology

IV. Network Performance and Analysis

The newly implemented technologies, such as VXLAN, Border Gateway Protocol and Ethernet VPN come in opposition to the old technologies, such as Spanning-Tree Protocol in order to replace the configuration of a switched access topology, using trunk links between the layers, with a routed access one, following the ECMP routing strategy. Thus, due to the routing strategy, configured in the underlay network, Spanning-Tree will not be allowed to block links, allowing the network infrastructure administrator to span VLANs end-to-end, improving high-availability and traffic distribution. Therefore, critical applications will function properly without interruptions when end users are likely to utilize them. Moreover, each of the customers service providers have will be allowed to manage their resources efficiently and maintain service continuity, due to traffic separation and high scalability offered by VXLAN, allowing the configuration of 16 million isolated networks.

In the process of data transmission, frames from one VM are transmitted to a Cisco L2 switch, where the encapsulation with the VLAN header DOT1Q is performed, resulting in a classic Ethernet framework. As Figure 8 illustrates, the destination MAC address inscribed in the frame coming from the source VM is the virtual gateway MAC of the IRB interface on the Leaf router, the first analysed network architecture being an ERB one. Moreover, the source MAC address is the one of the source VM. The L2 switch executes a broadcast ARP Request in order to map the IP address of the L3 Gateway to its own MAC address. Thus, the ARP Suppression process is being executed, in order to eliminate the need of flood and learn.

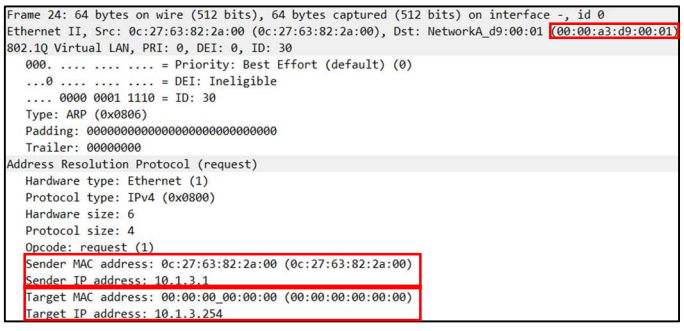

Figure 8. ARP Request on the flow from VM3 to VM2 in DC1

The CRB configuration performs routing at a central level of the Spine-Leaf network topology, namely at the Spine devices. The Leaf device will accomplish the encapsulation process with the VXLAN header and the external IP header, containing the IP addresses of the source Leaf and the next hop, namely the selected Spine device, letting the frames turn into packets and transit the topology network in order to reach the destination Leaf endpoint. This process is shown below in Figure 9. The packet also contains an UDP header, specific for VXLAN encapsulation process, where the destination port is 4789, assigned by IANA for the implementation of VXLAN. In contrast with CRB architecture, the ERB configuration performs routing at an edge level of the Spine-Leaf network topology, namely at the Leaf devices. Thus, the VXLAN virtual segments are no longer performed between Spine and Leaf devices, but between source and destination Leafs, their IP addresses corresponding to the ones in the external IP header.

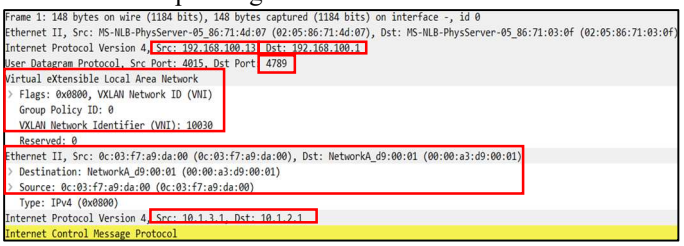

Figure 9. The encapsulation with VXLAN header process, performed by a Leaf device

As we can see in Figure 10, the Wireshark capture shows how MP-BGP sends keepalive messages in the Underlay network, in order to maintain the communication between network devices.

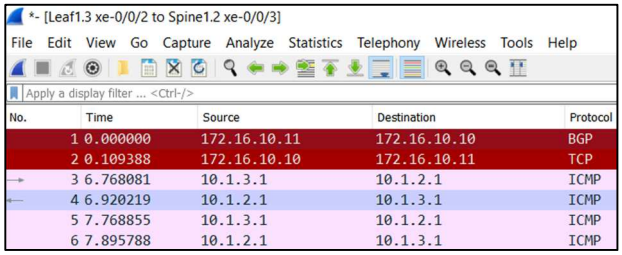

Figure 10. ICMP and keepalive messages between Leaf and Spine devices

The major difference between CRB and ERB architectures is that the first hop gateway capability is moved to the leaf level concerning the ERB architecture, using IRB interfaces with anycast addressing. The network architecture primarily depends on the amount of traffic flow travelling the entire Data Center, as in the case of network topologies where there is a lot of East-West traffic, specific to sending packets between VMs, the ERB architecture is particularly prefferrable. A common disadvantage of the ERB architecture consists of the performance of the physical pieces of equipment in a Data Center, as the Spine devices are only Route Reflectors in the network topology, actively participating in the dynamic routing process. When the Leaf devices don't meet the required performances, the CRB architecture is highly needed to be implemented.

V. Conclusions

The presented Spine-Leaf Data Center network topology, built on a standard-based architecture, allowed us to make a detailed analysis of the integration of specific protocols and technologies, including VXLAN and Ethernet VPN solutions, as well as MP-BGP and OSPF protocols. The use of a dynamic routing protocol permitted an automatic update of the routing tables corresponding to the implemented topologies, providing a short to zero downtime in case of the failure of a node in the

topology, in contrast to the use of Spanning Tree Protocol. Furthermore, it not only provides IP connectivity in the Underlay network, but also redundancy and the possibility of using all the connection elements between devices, following the ECMP routing strategy, contributing in this way to traffic optimization in the network topology, as routes are transported to other locations through BGP flows. The VXLAN technology configuration offered us the possibility of implementing a tunneling process, by encapsulating the frames coming from directly connected VMs with a VXLAN header, building a set of virtual segments over the infrastructure of a physical Underlay network, thereby offering scalability in terms of the number of Virtual Network Identifiers. The analyzed topology puts into light how customer traffic is separated due to the VRF technology configuration, in this way binding customer-specific instances to customer-owned virtual interfaces. This is offering not only an advantage in terms of security, isolating traffic traveling across the network, but also flexibility and scalability. The need for the integration of BGP and EVPN with VXLAN technology eliminates the need to transfer B.U.M. traffic using data plane protocol, in a flood and learn manner, introducing us to the control plane protocol, facilitating information advertising between VXLAN speakers. The presented Data Center Interconnection model is an efficient way to interconnect locations from different individual geographic areas of a small service provider. This is possible not only due to the MPLS technology, which introduced packet delivery using labels, but also due to the Type 5 EVPN routes carried from one Data Center to another.

We analyzed packets behaviour in a Spine-Leaf topology, following the processes of encapsulation, respectively decapsulation with the VXLAN header from the source VM to the destination end-host, using the Wireshark network protocol analyzer. Likewise, we made references in terms of differences between the different network architectures used for the process of routing the traffic between Virtual Network Identifiers. Moreover, in communication network which support critical applications, both in the National Defense System and in the governmental and civilian field, redundancy is one of the points of view that must be taken into account, so it is accomplished by configuring at least two Spine devices in each of the implemented Data Centers, the flow of the packets between Leaf devices being mediated by them.

REFERENCES

[1] Lukas Krattiger, Shyam Kapadia, David Jansen, ”Building Data Centers with VxLAN BGP, EVPN: A Cisco NX-OS Perspective”, Cisco Press, 2017

[2] Deepti Chandra, ”Data Center Deployment with EVPN/VxLAN”, Vervante, 2017

[3] https://www.juniper.net/documentation/en_US/release-independent/solutions/topics/task/configuration/centrally-routed-overlay-cloud-dc-configuring.html, accesed on February 17th, 2022

[4] https://www.juniper.net/documentation/en_US/release-independent/solutions/topics/task/configuration/edge-routed-overlay-cloud-dc-configuring.html, accesed on February 17th, 2022

[5] Richard Li, Kiran Makhijani, Lin Han, ”Cloudcasting: A New Architecture for Cloud Centric Networks”, American Research Center, Huawei Technologies, CTRQ 2016

[6] Brenden Buresh, Dan Eline, David Jensen, Jason Gmitter, Jeff Ostermiller, Jose Moreno, Kenny Lei, Lilian Quan, Lukas Krattiger, Max Ardica, Rahul Parameswaran, Rob Tappenden, Satish Kondalam, ”A Modern, Open and Scalable Fabric VXLAN EVPN”, Cisco