CONFIGURATION SANITY CHECK FOR VXLAN

Soumosir Dutta, Nitish Garg, Nalini N

School of Computing Science and Engineering,
VIT University, Vellore, TN, India

---***---

Abstract: *With the increasing datacenter complexity there is need of various advancement in technology to support the demand by the production environment. Optimal utilization of resources in layer 2 or datalink layer requires vlan for separation of broadcast domain. Vmobility is required in modern datacenters to have the server utilized to use many VMs at a time depending upon the requirements. There has to be moves for VMs between data centers for efficient utilization of RAM, physical memory, power, bandwidth of the links hence a layer 2 tunneling technology has to be implemented for users in one VM to be logically in the same domain even if it's physical VM has been moved.Advancement upon the VXLAN are to be performed for better utilization of the network resources which is discussed further.*

Keywords: Multi-Layer switches, VLAN, VXLAN, VM, VMobility, VTEP

1. Introduction

Digitization had lead to increase in data to be processed and stored hence the importance of data center design has been of prime importance. Technologies are implemented to make a network more efficient and have optimal use of resources in data center. Vlan has been introduced for making multiple broadcast domain to function under the same switch. VMobility is moving a vm from one server to another and is done for utilization of resources like RAM, memory, power and for host mobility in a network. For mobility tunneling is required for host too have the perception of same network hence VXLAN is being introduced. VXLAN is the overlay technology which provides layer 2 communication over a layer 3 network. Network resources like the Vteps and floodlist in VXLAN had to be managed so as to have less outage and save time troubleshooting configurations in networking devices.A tool has to be implemented for configuration management and improving the management in production network scenarios.

2. Related work

2.1 VLAN

Computer networks which are local to a specific region is segregated by the local area network. They are mainly used to interconnect computers or servers within a limited reach like that of a college network or company inter network. They architecture of a LAN mainly revolve around connecting switches as networking device which perform both data link or layer 2 and network layer or layer 3 functionalities. The devices which performs L2 frames forwarding are primarily known as switches and those responsible for L3 packet forwarding are known as routers. In production network the L3 forwarding devices are named vendor specific. These networking devices has a combination of hardware forwarding agent known as ASIC and a software which runs on the CPU block for programming the logic which determines hardware forwarding. The ASIC contains the hardware chip which are responsible for data forwarding in bits if seen by modular point of view.

The OSI standard has separated the networking viewpoint into seven layers in which the layer 1 or physical layer has transmission units as bits and consists of cabling standards. Some examples are the coaxial cable which are used mostly in datacenters and also Shielded twisted pair and unshielded twisted pair which gives better reliability.

Data link layer or layer 2 does the forwarding of frames and the forwarding is done by the networking devices like hub or a switch. We will be only concentrating upon switch as hub is beyond our scope. The switch performs its operation by the mac address table in it. Any packet that hits the switch is decapsulated and the destination mac address is seen and matched with the mac address table in switch if present and then forwarded to the port or physical interface present in it or to the cpu as per the logic and rules defined. A broadcast message is however forwarded to all the ports except for the incoming port. A domain where the broadcast messages gets forwarded is known as broadcast message. As switch forwards the message to all the ports it cannot restrict the broadcast. Routers which is a layer 3 or network layer devices does the forwarding based upon the layer 3 address or the IP address in general. A router therefore has the capability to restrict a broadcast message. Hence routers can separate broadcast domains but even having that due to several use case and ever growing networks it's not feasible to have many routers in a network mind the consideration of cost in a production environment, hence a technology is being introduced to segregate the broadcast domain here itself in switches or layer 2 device. This is primarily known as vlans.

Vlans are virtual lans which are used to segregate broadcast domains in data link layer or layer 2 in osi standards. The separation is necessary for today's ever increasing networks and subnetworks with i a organization. When we take into consideration of organization like that of an educational institute like Engineering college there will be stream decision as well as division based upon the designation like students and faculties primarily. Other than that division we have sub parts in students like UGs and PGs and also distant education separately. There will be division again in hostels and main educational block.

Apart from that a college supports various streams like that of computer science, electrical engineering , mechanical engineering and many others including the division of branch based upon specialization.

Taking into account all the separate domains it will be suboptimal if individual network addresses are used for all of them. Hence there should be a way to separate the broadcast domain in data link layer itself for utilizing the whole of network. Thus Vlan comes into picture. Vlan helps us create separate broadcast domain resulting into meeting our requirements for data to be forwarded to only specific domains when as per requirements.

Vlan works by creating a vlan id with the vlan number for identifying a vlan. There can be 4096 vlans as per the standard bit set.There are two modes in it primarily Access and Trunk mode.After a vlan is being configured in a switch ports should be allocated to vlans to bring the vlan interface up. The allocation of a port is done by giving access to the vlan.

But by the access option only one vlan access can be done by the port. Here there is no tagging done in the packet to identify the vlan but the mechanism is handled by the switch allowing the packet coming from a certain vlan access port to only be forwarded into those remaining ports where the same vlan is allowed.

The second mode is the trunk mode. A port when given under trunk mode by default can access all the vlans. There is also a option to allow certain vlans in a port depending upon the requirements. The trunk port mode has a tagging mechanism where a 802.1q tag is added to the ethernet frame to identify the vlans allowed. Here a 4 byte header is being added to the ethernet frame to identify the vlan number. But due to this separation in broadcast domain ,for two separate vlans to communicate inter vlan routing comes to picture where switch virtual interface(SVI) is required.

SVIs are the logical interface where ip addresses are given and are generally gateways for the devices connected via access port allowed in vlan. These svi's required for inter-vlan routing as the device acts as a router where the packet is routed to directly connected logical interface i.e the SVIs.

Having vlan in one's network is advantageous as the networking parameters are used efficiently and provides optimal utilization.

2.2 VMobility

Everyone nowadays is digitizing. Everything and every business is taking the help of digital solutions like mobile and web applications, websites, digital media to expand their business and outreach. This results in generating large amounts of data which needs to be stored for further processing or analysis and future use. Data these days is increasing at a very fast rate due to rapid inclination of large population towards digital world.

This is creating a problem to store such large amounts of ever-increasing data. Where to store this data so that a person sitting at a different location in the world can access this data without much difficulty and time-delay (latency). Data centers are the places where this data is stored. These are the facilities where such data is stored in data servers and are provided to such applications and websites when queried for. To provide seamless and flawless delivery of this data, data centers are located at various locations of the world.

Care is also taken to take into account the chances of data loss due to some natural calamity or some other unavoidable mis-happenings. So data centers are set up taking into account redundancy. Multiple copies of data are kept at various locations to prevent data loss.

Virtualization is the concept where actual physical servers are replaced with virtual machines where these storage, networking and other infrastructure services run. These are cloud computing and virtualization technologies. These helps to make a better use of actual physical resources. As these physical resources cost a lot, virtualization helps to make an optimal use of them as in data centers the requirement or usage of these resources can be fluctuating a lot based on varying data traffic. To make optimal usage of the physical hardware, without incurring much cost and also simultaneously catering the high rates of data traffic, virtual machines (VM) are run on servers which in turn run various services. VMs are hosted on actual heavy duty physical servers which have ample physical resources to host multiple VMs. Virtualization in data centers is an umbrella term covering all the processes, tools, and technologies that run data centers and allow to run services on top of virtualization layers. This also helps to run multiple virtual data centers on existing physical data center infrastructure. These can be used for different services or applications or organizations. Cost of running data centers also include the electricity cost to run the network devices like switches, routers, firewalls and servers, cost to keep the data centers cool as running these devices at such high data rates for very large period of time (may span over years) continuously generate a lot of heat [2] .

With the increase in size of data centers and in their number, the effort to maintain these virtualization technologies and tools is also increasing. The data centers of same organization providing similar services may be located at different locations which may differ in geographical locations or in same physical data center.

To let the two servers communicate with each other as if they were present in same VLAN, VXLAN (Virtual Extensible Local Area Network) is used. VXLAN is tunneling technique which overlays Layer 2 over Layer 3 network. The servers in two different (actual or physical) data centers need to be configured in same way so that the data is routed to the external world exactly the same way from both the servers. This techniques also helps in movement of a VM from one data center to the other in a working state. If both VMs are configured exactly the same, this movement will not affect the data traffic of the user availing the services, and he won't even get to know of any network change. For two servers to have exactly same configuration, they should belong to same logical local network, but extending a local network over Layer 3 domain is not possible for which again VXLAN comes into picture. VXLAN lets to migrate the VMs from one data center to other flawlessly. This movement of VMs is called vmobility. For vmobility first the files of the VM is transferred to the new location. Data files and the state of the virtual machine are the two types of files need to be transferred for vmobility. State files store the state of the VM. It stores the current running state at which the VM is currently running. It is a kind of snapshot that has the change log for the virtual machine's disk files (VMDK). It helps to restore the VM to a particular point of time from where you want to restore. For vmobility, the state files are copied so that the VM can be restored and the data files of the VM which have the data on which the VM is working.

To have a successful vmobility, a lot of parameters need to be checked first. The parameters need to be checked for successful vmobility are:

1. CPU compatibility

2. Interface for vmobility

3. Central mass storage which is shared

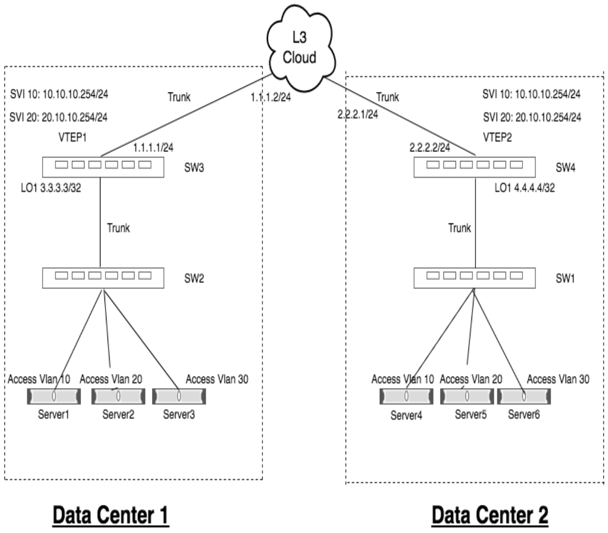
4. Naming for virtual port groups must be same

5. Target host must have sufficient resources like RAM, Hard Disk

2.3 VXLAN

VXLAN is a network virtualization technology that helps to have scalability in a large scale distributed data center network. It is the overlay of Layer 2 or data link layer over the layer 3 or network layer. The layer 3 traffic is encapsulated and an outer layer 2 headers are added to make it a layer 2 forwarding keeping the inner contents of the traffic unaffected by the layer 3 routing or the internet medium connecting two datacenters. After the layer 3 packet comes to a VXLAN enabled interface there is a 8 bytes VXLAN header added to the packet out of which 24bits are assigned for VNI or virtual network identifier. Apart from that a outer mac address are added which does the outer layer 2 forwarding and also an outer ip is added which does the outer routing. A destination udp is also added which is 4789 which is specifically assigned VXLAN communication i.e the layer 4 port number. The vlan to VNI mapping which is present in an VXLAN enabled interface is matched with the packet and then VNI mapping is added to the VXLAN header encapsulated in the packet [1].

For VXLAN to work in a networking device there has to be a vtep or otherwise known as logical interface necessary for VXLAN packet forwarding.Vlan to Vni mapping should be made available for the packet to get converted to VXLAN tagged frame. There are primarily two types of communication in VXLAN i.e the VXLAN bridging and VXLAN routing. For VXLAN bridging there should be communication between 2 devices under the same vlan and for VXLAN routing, there is communication between two separate vlans. The multilayer switch in which VXLAN interface is enabled floods the VXLAN encapsulated packet into the vteps present in its floodlist as for the devices to communicate the vteps of the devices should be present in the floodlist of each other. They also should be listening on the same udp port[5].



3. Proposed work:Topology

The above network design is a simple topology of a working VXLAN. In this there are 4 Multi-Layer switches which provide the basic working functionality of VXLAN tunneling. There are 4 virtual machines running, 2 in each data center. One virtual machine in data center 1 is in VLAN 10 and other in VLAN 20. Similarly in the second data center one is in VLAN 10 and other in VLAN 20. The lower 2 switches (SW2 and SW1) are basically present in a data center which provide connectivity of the data center servers to the external public IP Domain. The upper two switches are basically providing Layer 3 routing over the public IP domain. The links to virtual machines from lower switches is access link only allowing traffic of that VLAN. Connection between lower and upper switches is in trunk mode. There is redundant trunk connection between the two upper switches over Layer 3 network. Loopbacks are configured on the upper two switches. IP routing is enabled on the upper two switches to communicate over Layer 3 network. VTEPs (Virtual Tunneling Endpoints) are configured on the two loopback switches. Using VXLAN, the Layer 2 frame is encapsulated by outer UDP header, outer IP header, and outer MAC header, so that this can flow over Layer 3 network as Layer 2 link.[3]

The virtual machines run on the ESX servers which are represented by the system PCs in the above architecture. Server 1 in Data Center 1 can communicate with server 3 in Data Center 2 as if it was in same VLAN.

3.1 Mis-config

In growing datacenter, configuration changes are pushed to the networking devices on day-to-day basis by a network administrator. Frequent changes could lead to misconfiguration due to human errors and could lead to network outages. Troubleshooting during such incidents is time consuming, hence configuration management and sanity check for the devices are needed to avoid such issue. VXLAN config sanity check is designed to solve above problems. Various areas where mis-configurations can occur are:

1. The source IP of VTEP interface could be not configured or inactive.

2. Flood-list of VTEP is not properly configured

3. VLAN to VNI mapping is not correctly done

4. SVIs are not configured correctly

5. Error in configuring UDP destination port

4. Implementation

The topology has been configured properly and tests has been carried out on it. This tool will display all the configurations done on the VTEPs of the VXLAN network on the basis of the tech-support information from the switch fed into the tool. Firstly it will check if all the single VTEPs are configured correctly individually and then it will check if all the VTEPs are configured correctly in relation to each other in the VXLAN network setup[4].

Main benefits from the use of this tool are:

1. A lot of time will be saved while checking the configurations done on VTEP or checking for the possible mis-configurations.
2. It will greatly reduce human errors and efforts while configuring the network or troubleshooting it.
3. Reduce network downtime by a great factor as knowing the root cause of problem and fixing it would be easier and faster.
4. Pre-check before deploying the network on actual infrastructure can be done so that any possible errors that may come into picture are sorted out in the dry-run phase itself.
5. Scalability to check multiple VTEPs is a very important factor in checking large scale networks having large number of VTEPs as logging into each and every switch and then checking for their configurations is a very tedious job and again there are chances that the errors in configurations are skipped.

The main groups of users who will get benefit from this tool are:

1. TAC engineer

2. End users which include system engineers and the network administrators of data centers.

5. Conclusion

The above discussed test topology for VXLAN was successfully implemented first in virtual environments and then on physical devices. The VXLAN setup was up and running properly in both the environments. Then the tool was implemented and was tested against various test cases. The tool was tested for various configurations extracted from the switches both in virtual and real world. The results obtained from the tool were in accordance with what was expected.

6. References

1. RFC7348, M. Mahalingam, D. Dutt, K. Duda, P. Agarwal, L. Kreeger, T. Sridhar, M. Bursell, C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014.

2. Steve Herrod (August 30, 2011). "Towards Virtualized Networking for the Cloud". VMware. Retrieved 2013-02-25.

3. "Virtual Extensible LAN (VXLAN) Best Practices (White Paper)" (PDF). Cisco Systems. 2016-01-08. Retrieved 2016-07-04.

4. "Arista Expands Leaf Switch Product Portfolio" (Press release). Arista Networks. 22 October 2014. Retrieved 8 November 2014. Arista's updated VXLAN implementation eliminates the need for multicast in the underlay network by using Head End Replication for forwarding broadcast, multicast and unknown unicast traffic".

5. "Arista and VMware have coauthored a new standard in cloud networking: the Virtual eXtensible LAN (VXLAN)" (PDF). Retrieved 2013-12-01.

6. M. Mahalingam; et al. (February 22, 2013). "VXLAN: A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks". Retrieved 2013-02-25.