Virtual Link Trunking (VLT)

Reference Architecture

Revision 2.0
Issued by: CTO Office/Systems Engineering
Author: P.Narayana Swamy

Table of Contents

1 Overview

With the rapid growth in the emerging IT market space, data centers, enterprise networks, campus networks are having a paradigm shift in the underlying network to cater the expanding business needs. Virtualization at all layers demand specific fabric architectures, while the traffic pattern in the next generation network is predominantly east-west, the network needs a highly resilient and agile fabric to meet the challenges for its seamless operations.

As networks scale, extending and managing the layer-2 fabric in a large network gets more complex. Extending multiple links for adding bandwidth may not be of much help with Spanning Tree Protocol running across the network. Any link failure with such STP deployed networks is fraught with network convergence resulting in sub optimal performance and not matching the underlying virtualized agile workload needs.

Virtual Link Trunking (VLT) overcomes the perilous layer-2 loop and blocked links by providing access to the underlying Top of Rack (ToR) switches in cascade and offer high resiliency and availability, effectively utilizing the multipath. However, diligent design eliminates the over-subscription of the traffic from the access to the aggregation. Based on the unique and specific requirements, the customer can deploy VLT/mVLT in their networks.

Migration from the existing STP based network to VLT involves meticulous planning, such that the configuration need to be planned before, and the VLT peers to converge, '*Peer-routing*' feature introduced in FTOS 9.2(0.0) release is highly recommended for active-active load sharing across the multiple VLAN's within VLT peers, ensuring high-availability. FTOS 9.2(0.0) adds supporting the routed (Peer-routing) VLT with unique features eliminating VRRP. With IPv6 and IPv4 dual stack, VRRP could be deployed for effective resiliency.

This document explains the dual node VLT deployment strategies with its associated network reference architecture. Various VLT deployment topologies are explained in this document with emphasis on best practices and recommendations for some of the network scenarios. This document also covers the configuration and troubleshooting of VLT using relevant show commands and different outputs.

1.1 Introduction to VLT

VLT in general term, ensembles two physical switch to represent as a single logical switch. With physical links as a port-channel, connecting two individual switches configured with VLT would logically group it as single entity only for the Access switches which connect to the VLT domain. Both the VLT peers have their own configuration, control and data planes. VLT allows creating port-channels across two switches, eliminating the need for spanning tree blocking states across its connectivity and effectively utilizing its entire VLT links connected to the access/ToR switches. The access device could be a switch, switch stack, dual NIC server or any other device supporting LACP port-channels. High-availability and redundancy in the network is achieved by its duality in physical connectivity.

1.2 VLT Implementation

Periodic hello messages are sent through the VLT Interconnect (VLTi) and the VLT control messages are sent in TLV format through the VLTi links for synchronizing the L2/L3 control planes across the two VLT peers. MAC, ARP tables, IGMP States are synchronized between the VLT peers ensuring traffic flow across the links and seamless failover in case of VLT link or node failure. The VLT feature ensures the local traffic on a VLT Switch takes the shortest path to the destination through the VLT links and not through the VLTi links. However VLTi carries the traffic during the link failure states. (Figure 1.0)

A backup link is established between the VLT peers normally through the management interface to exchange the periodic backup heartbeat messages through the out of band network. The backup link could be of any other physical interface, however as a best practice the management interface is used to exchange the heartbeat messages and its significance (Section 3.1) lies when all the members of the VLTi port-channel fails. With the priority election in VLT domain, the primary VLT peer takes control of handling LACP and RSTP control states and exchanging messages to the secondary VLT peer.

For all the VLANs configured on the VLT port-channels, the VLANs are dynamically mapped on the VLTi Links. When VRRP is implemented within the VLT domain, the VLT peers offer active-active load sharing. Traffic would be locally switched to its destination without redirecting to VRRP master. Similarly 'peer-routing' enables the active-active load sharing for the L3 VLAN traffic without enabling VRRP. With peer-routing the VLANs in the VLT fabric could scale beyond the limitation of VRRP groups.

Any broadcast traffic being flooded across the VLTi links prevents duplicate copies at each of the peer in the VLT domain. This is ensured by installing a port block, so that the packets coming in through the VLTi are blocked from egressing on the VLT ports. This block is installed only if both the peers have active ports on that VLT. If all the ports on one peer for that VLT are down, then this state is indicated to the other peer, which removes the block, so that the remote node continues to receive the data.

Significant features in the FTOS 9.2(0.0) release

- 'VLT Min-loss' feature to reduce the traffic loss and improve convergence time during failovers.
- VLT supports IPv6 with VRRPv3. Dual stack of IPv4 and IPv6 are supported which would certainly benefit for the customers migrating to IPv6.
- Multicast features with PIM-SM are supported.
- MXL Switches could form VLT within the chassis improving the resiliency of servers deployed within the M1000e chassis.
- IOA Switch forms auto VLT with a single CLI.

NOTE: Latest VLT features are supported on S4810, S4820T, MXL, IOM/IOA and Z9000 switches in FTOS 9.2(0.0) release

1.3 VLT Terminology

Virtual link trunk (VLT)	The combined port channel between an attached device (ToR switch) and the VLT peer switches.
VLT backup link	Monitors the vitality of a VLT peer switches. The backup link sends configurable, periodic keep alive messages between VLT peer switches
VLT interconnect (VLTi) or Inter Chassis Link (ICL)	Used to synchronize states between the VLT peer switches. Both ends must be on 10G or 40G interfaces.
VLT domain	Includes both VLT peer devices, the VLT interconnect, and all the port-channels in the VLT connected to the attached devices. It is also associated to the configuration mode that must be used to assign VLT global parameters.
Non-VLT (Orphan) Ports	Any ports not connected to VLT port-channel in the VLT node
VLT nodes	One pair of switches that are connected with the port channel known as the VLT inter-connect (VLTi).

1.4 Significant Advantages of VLT

- Loop free connectivity in layer-2 domain
- Faster Network convergence
- High-availability and redundancy
- Effective utilization of all the links
- Link level resiliency
- Active-Active Load sharing with VRRP.
- Active-Active load sharing with Peer-routing for Layer-3 VLAN
- Graceful failover of LACP during reload
- Agility in VM Migration under VLT domain.
- Unified access for virtualization, Web applications and Cloud computing
- High performance for Big Data networks
- Easier design and manageability of fabric with AFM.

1.5 Best Practices for Implementing VLT

1. FTOS version should be the same in both the VLT peers.
2. Backup link should pass through the OOB management network
3. Enable RSTP in both the VLT peers and tweak the timers for minimum values.
4. Keep the Primary VLT node as Root-Bridge, secondary node as backup root-bridge.
5. Use Identical System MAC address on both the VLT peers to avoid minimum traffic loss.
6. Configure unique Unit-id on VLT peers.
7. Adjust Delay-restore timer from the default 90 seconds based on the servers.
8. Configure LLDP to assist during troubleshooting link failures.
9. Use VLTi Links configured with static port-channel of more than one member.
10. Configure VLT links with LACP Port-channels.
11. In a hierarchical mVLT domain with OSPF, configure the Spine/Core primary node as DR.
12. Enter description fields of the VLT and VLTi links configuration for easier identification.
13. Configure RSTP, "LACP ungroup" CLI on VLT nodes and ToR for easier BMP boot process.
14. VLTi can be statically configured as mrouter port, when PIM is deployed on VLT VLAN.
15. In a PIM-SM environment, deploy a non-VLT core node as Rendezvous Point (RP)

1.6 Typical VLT Topology

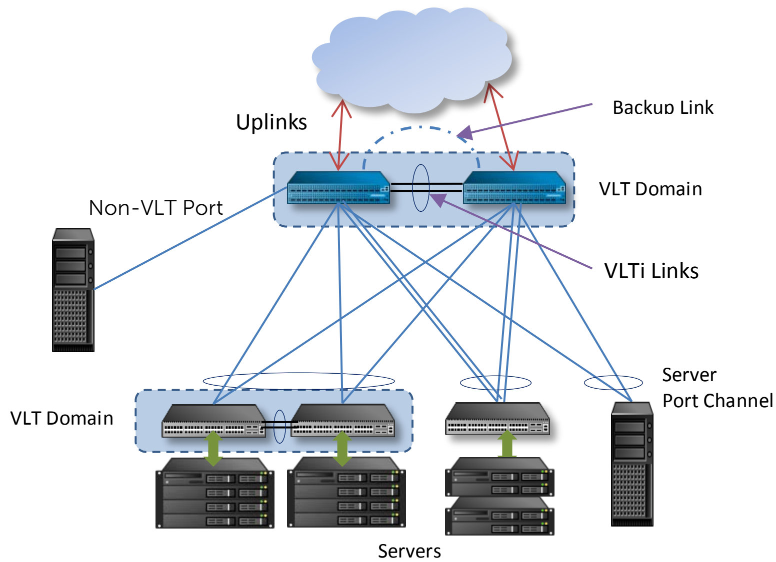

Figure 1.0 Virtual Link Trunking Topology

The VLT domain has VLTi links connecting between VLT Peers and VLT port-channels connecting to Single Access Switch, to a Switch Stack, server supporting LACP on its dual NIC, or it could connect to another VLT domain as shown in figure 1.0. The backup-link get connected through the OOB Management Network. Some hosts could directly connect through the Non-VLT ports.

2 Packet flow in VLT

Due to the inherent layer-2 multi-path hashing in the VLT port-channel links, the traffic might pass through any of the VLT peer before reaching its destination. The various colors of traffic stream flow are shown in the figure 2.0.

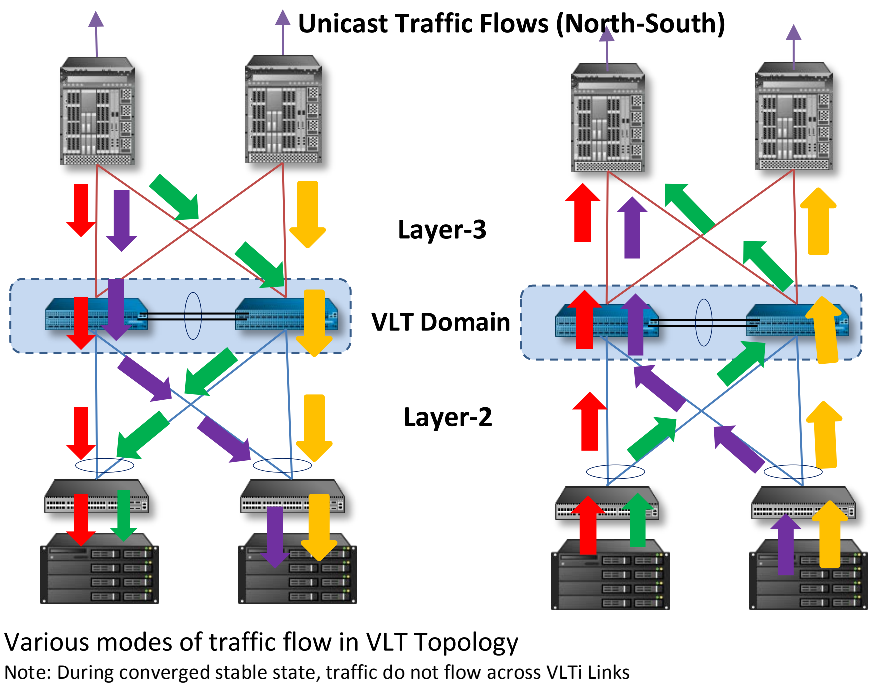

Figure 2.0 Typical traffic flow in VLT topology

During the converged stable state, with MAC and ARP sync in both VLT peers, the traffic flow (North-South) from the Core layer (Layer-3) to the VLT peer with ECMP path reaches its destination through the shortest path. Similarly the traffic flow (East-West) between the servers in the ToR switches, within the layer-2 domain would always take the optimal path as shown in figure 2.1

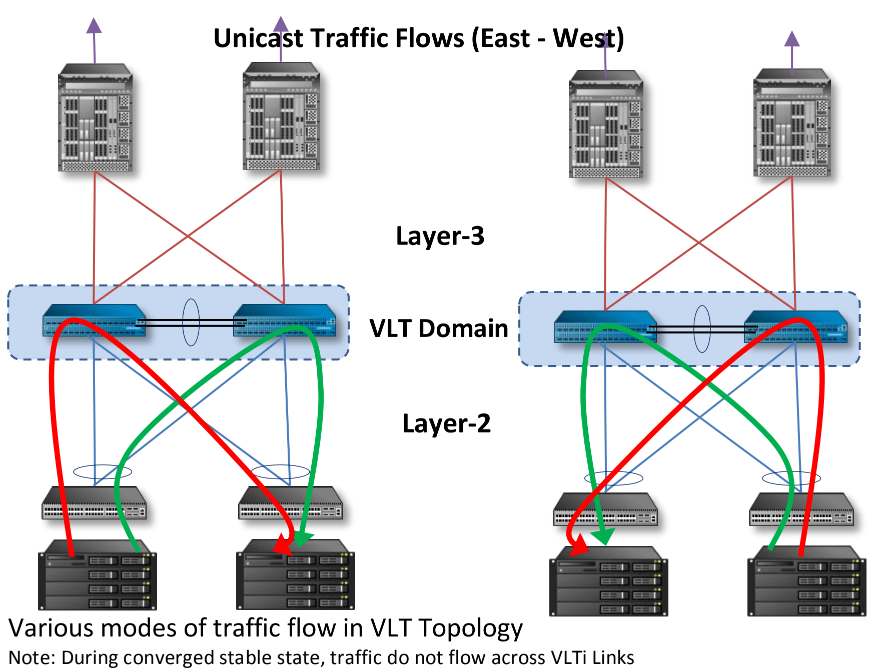

Figure 2.1 East-West traffic flow in VLT topology

3 Link Failover Scenarios

Any failure in the upstream layer-3 links would force the traffic towards the alternate ECMP path to the VLT domain and local switched to reach its destination. The link failure in the VLT port-channel might have the traffic passed through the VLTi as shown in the figure 3.0, since the MAC address learnt on the failed VLT would be now mapped to the VLTi port.

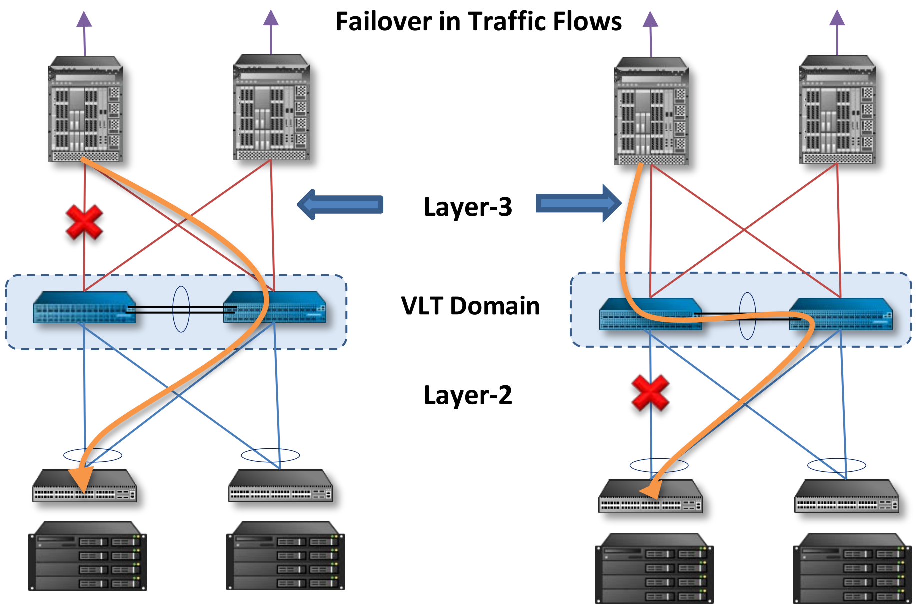

Figure 3.0 Traffic flow during link failure

For the north bound traffic with default gateway configured for specific Layer-3 VLAN in a VLT peer, where the traffic hashed and reaches the other VLT node, it would still pass through without diverting to the appropriate gateway node

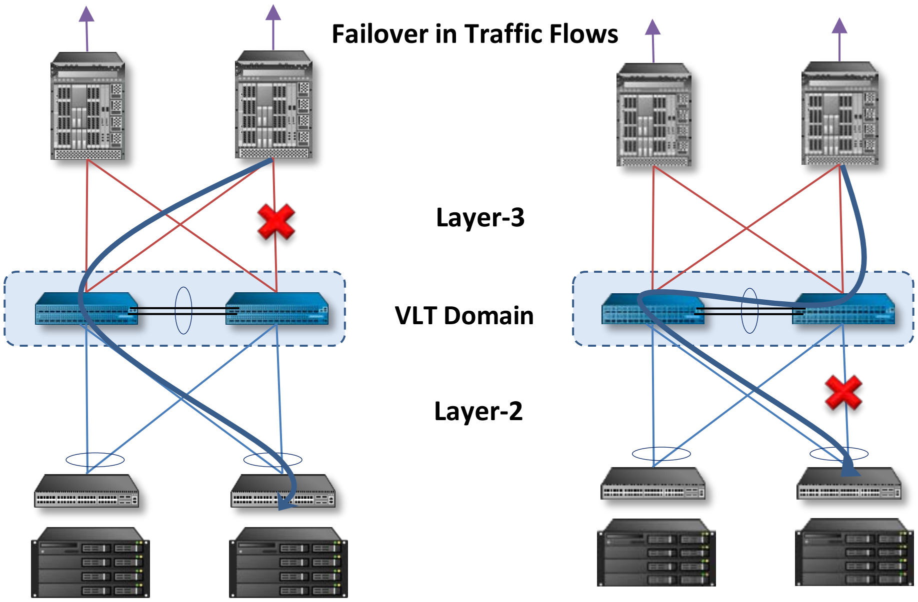

Figure 3.1 Traffic flow during link failure

This link failover feature with local switching mechanism is adopted in the latest releases of FTOS 9.2(0.0)

3.1 Split-Brain in VLT

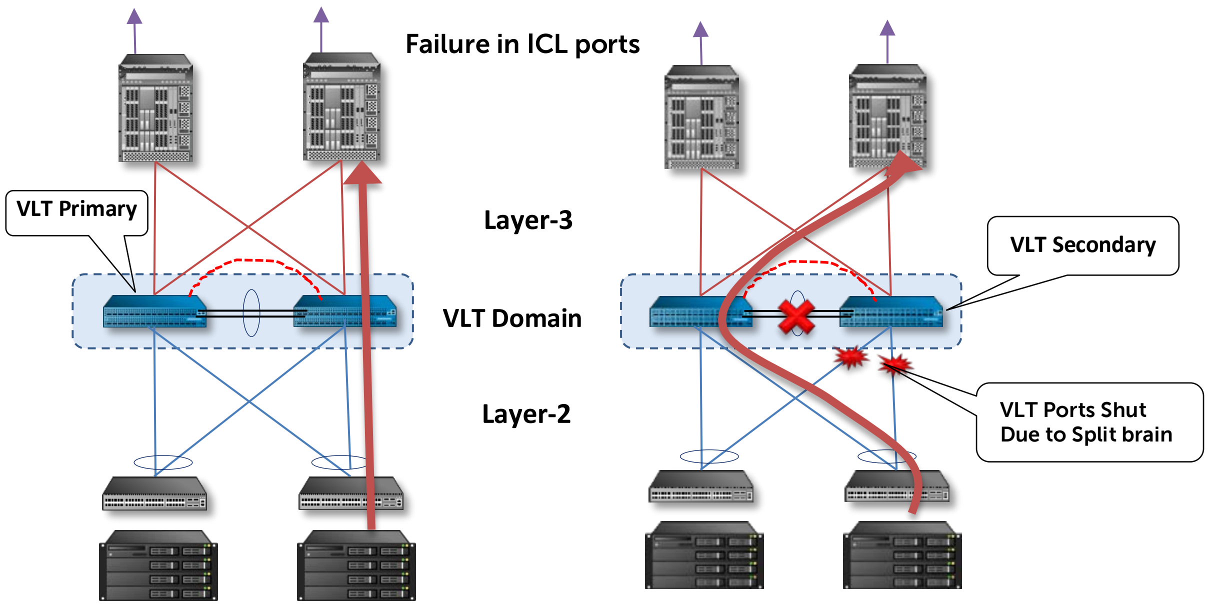

Figure 3.2 Traffic flow during VLTi failure

The backup heartbeat messages are exchanged between the VLT peers through the backup links of the OOB Management network. When the VLTi link (port-channel) fails, the MAC/ARP entries cannot be synchronized between the VLT peers through the failed VLTI link, hence the Secondary VLT Peer shuts the VLT port-channel forcing the traffic from the ToR switches to flow only through the primary VLT peer to avoid traffic black-hole. Similarly the return traffic on layer-3 also reaches the primary VLT node. This is Split-brain scenario and when the VLTI link is restored, the secondary VLT peer waits for the pre-configured time (*delay-restore*) for the MAC/ARP tables to synchronize before passing the traffic. In case of both VLTi and backup link failure, both the VLT nodes take primary role and continue to pass the traffic if the system mac is configured on both the VLT peers. However there would not be MAC/ARP synchronization.

3.2 VLTi Traffic

The VLTi links normally carry the following traffic

1. VLT Hello (domain-id, system-mac, unit-id, priority, version), election and other related messages in TLV format
2. MAC, ARP table, IGMP state, iSCSI Synchronization and other such type of messages
3. Traffic towards the non-VLT ports from one VLT peer to another VLT peer
4. Multicast and Broadcast traffic in the domain
5. Diverted traffic due to VLT link failure
6. Layer-3 protocol traffic forming adjacency over VLTi
7. VRRP hello messages.

3.3 Peer-routing

With FTOS 9.2(0.0) release VLT supports "*peer-routing*" feature. This feature enables the VLT node to act as a proxy gateway for the other VLT peer. As shown in the figure, the packets could be sent to either of the VLT port-channel members. Due to the hashing algorithm in the port-channel, if the packet is sent to the Peer-2 which is not the destined gateway for the hosts under the ToR Switch, the packet gets switched to the destined peer in the earlier release.

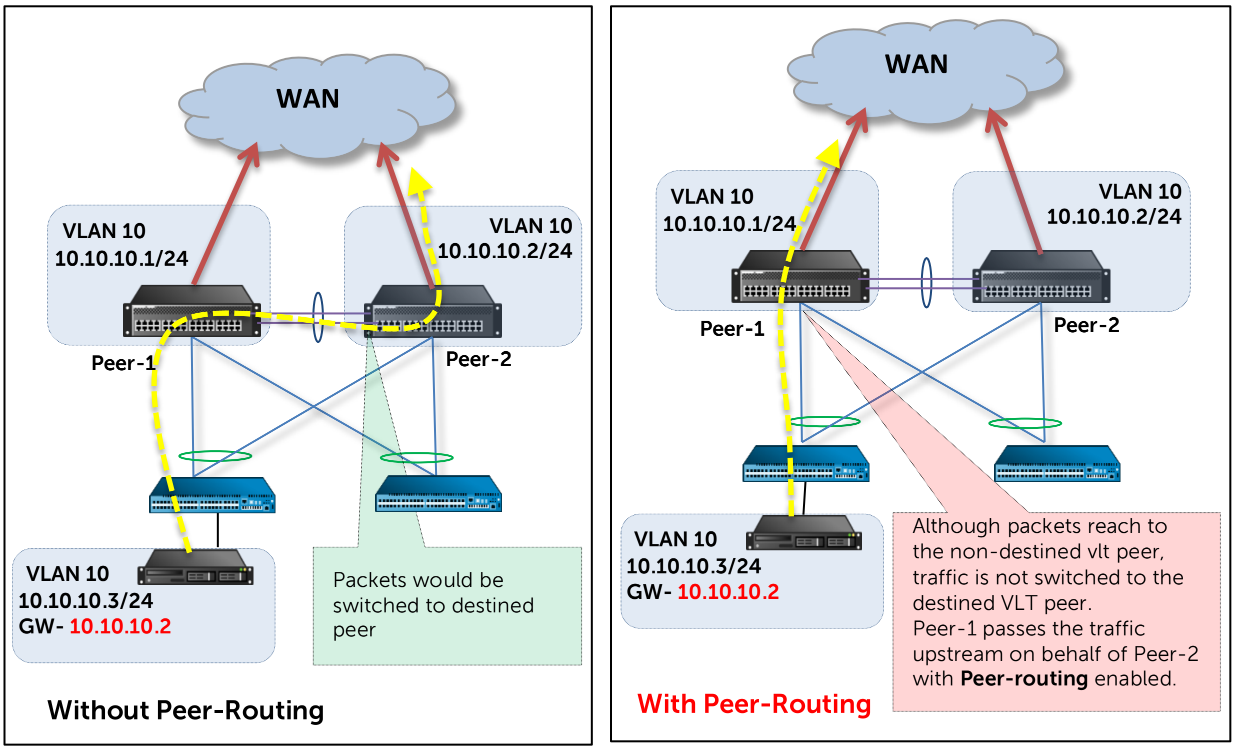

Figure 3.3 Traffic flow with & without Peer-Routing

However with Peer-routing feature in the FTOS 9.2(0.0) release, the VLT node acts as a proxy gateway only for its connected VLT peer. The packet received by the non-destined VLT node would act as a gateway and pass it upstream without tunneling to the destined peer with the following advantages:

1. Avoiding sub optimal routing
2. Latency is considerably reduced by avoiding another hop in the traffic path.
3. VLTi Port-Channel members could be reduced based on the specific design.

With peer-routing, VRRP need not be configured for those participating VLANs. Since both VLT nodes act as a gateway for its peer, irrespective of the gateway IP address, the traffic would tend to flow upstream without any latency. Also there is no limitation in the scaling of VLAN's. (With VRRP we have limitation of 255 VLANs).

3.3.1 Secondary IP address support for VLANs

Also the VLANs could be configured with different subnets as Secondary IP address, ensuring the IP connectivity across different subnets for the underlying Server nodes. However, these VLANs should be symmetrically configured on both the VLT peers in the same mode. Asymmetric configuration (layer-2 VLAN in node-1 and layer-3 VLAN in node-2) of same VLANs are not permitted in the VLT Peers.

Both the VLT nodes have to be configured for "*peer-routing*" under VLT configuration.

3.3.2 Peer-routing-timeout

By default the VLT nodes continuously act as proxy for its peer until the ARP times out. With the '*peer-routing-timeout*' enabled, if the VLT peer fails (or rebooted during maintenance), the VLT node acts as gateway only for the configured duration (maximum of 65535 seconds). However if the '*peer-routing-timeout*' is not configured, then the active node continues to pass the traffic on behalf of its failed peer, until the connected hosts ARP times out.

As a best practice, do not disable "peer-routing" when the VLT nodes are active and passing traffic through multiple VLAN's. Also '*peer-routing*' should be enabled before configuring VLANs.

3.3.3 VLAN Scalability

This 'peer-routing' feature also covers the vlan wildcarding internal architecture approach which helps in scaling the number of L3 vlans supported with the current FTOS. In the current release, with '*peer-routing*' enabled, per Vlan based local entry in the table is removed and VLAN field will be wild carded .A separate VLAN profile table will be used to set the status of Routing and Switching capability of each vlan. Activating VLAN wildcarding in FTOS scales more number of L3 vlans with VLT.

3.3.4 Routing Protocols

Routing protocols (RIP, OSPF, IS-IS, BGP) are supported with Routed VLT. As all the VLANs in the VLT are part of broadcast network, 'point-to-point' network configurations are not supported in OSPF and IS-IS. These routing protocols by default enables only broadcast network, hence no specific configuration is required on the routed VLAN interfaces. Also as a best practice, the Core/Spine VLT domain shall be configured as DR/DIS for effective convergence.

3.4 Graceful restart for LACP

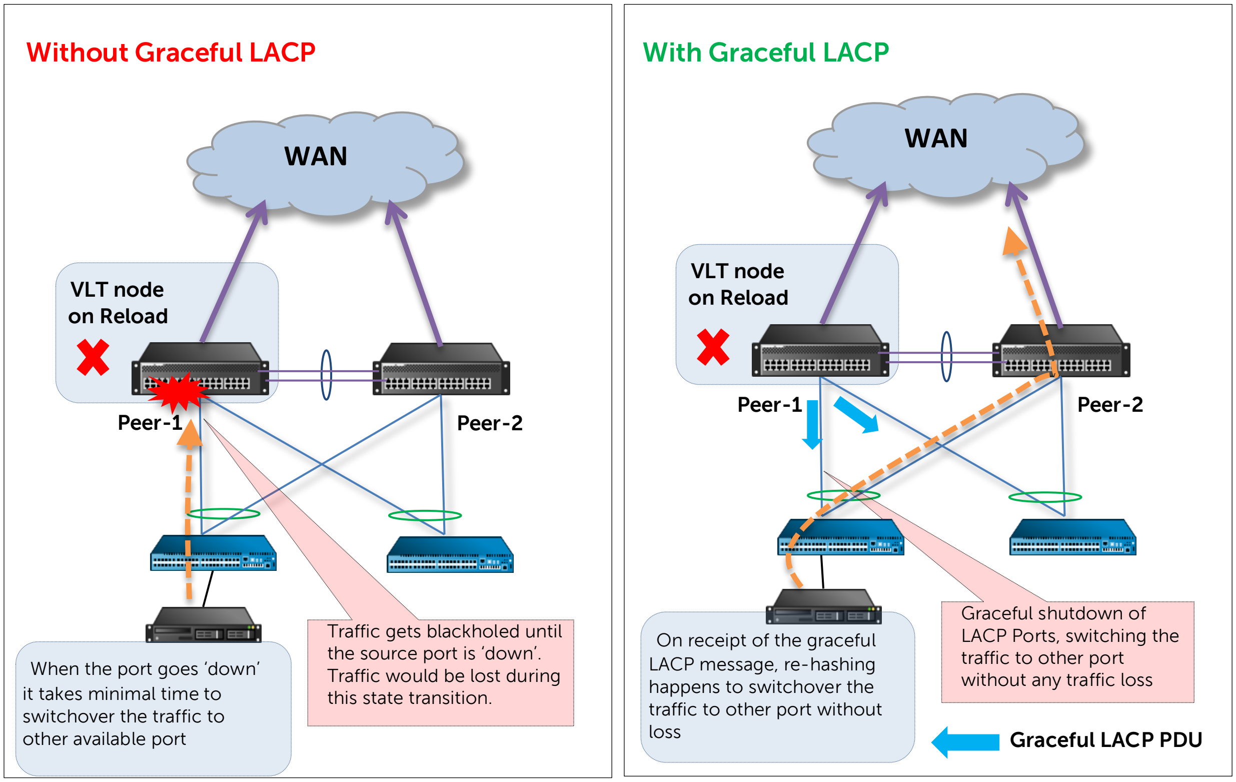

Figure 3.4 Traffic flow with & without Graceful LACP

With FTOS 9.2(0.0) release, 'graceful restart of LACP' feature is introduced.

In earlier FTOS release, if the VLT node gets reloaded, there is a definite time delay in switching the traffic towards the other active peer. The Interface status change, detection of link status at hardware with the subsequent trigger to switchover leads to transient traffic loss. However with FTOS 9.2(0.0) release, whenever the VLT nodes are reloaded, the graceful restart LACP sends a special PDU on all the ports of the VLT port-channel. On receiving the graceful LACP PDU, the ToR Switch sends an acknowledgement to the VLT node and detaches the ports from the VLT Port-Channel. Simultaneously the Port-channel in the ToR seamlessly switches the traffic on the other member of the VLT port-channel forcing the deterministic traffic towards the other VLT peer. This process ensures there is no traffic loss or blackholing due to the VLT node getting reloaded.

When the node comes up after reload, the VLT port-channels take a definite "restore-delay" time to pass the traffic. This is to ensure all the MAC/ARP synchronization happens between the peers. After the delay restore time the VLT port-channel from the reloaded node comes UP and continues to pass the traffic. With the above features, seamless traffic flow is ensured for the end-stations, any firmware upgrade could be done without major downtime which is essential for high availability and high SLA networks.

```
VLT-10-PEER-1#reload
Proceed with reload [confirm yes/no]: yes
All VLT LAG's gracefully shut down...!!!
LACP-5-PORT-UNGROUPED: PortChannel-033-Ungrouped: Interface Te 0/12 exited port-channel 33
```

3.5 Layer-3 Routing in VLT

Spanning the same VLANs across the racks, pods, and clusters is a major challenge in a data center. Deploying a flat layer-2 network has its own limitations with sub optimal utilization of the links, often fraught with perils of loop, restricted scalability etc.., With FTOS 9.2(0.0) release, the same VLAN extension across racks is made possible in a unique way of configuring layer-3 VLANs across the VLT nodes and the ToR Switches. Spanning the VLANs in mVLT architecture could interconnect and aggregate multiple racks with same VLAN. Moreover with interspersing of layer-2 and layer-3 VLANs in the mVLT the ARP table could scale considerably extending the scalability of the domain.

OSPF, IS-IS, BGP could be configured as the routing protocol for this extended VLANs. This would ensure routes learnt by specific VLT node would be advertised to all other participating nodes in the same VLAN. Since the IGP is configured on the VLAN interfaces, only the default broadcast mode is supported. Point-to-point mode of OSPF/IS-IS is not supported. BGP form adjacencies with the VLAN IP address between the VLT nodes and advertise the routes across different peers.

For a highly virtualized Data Centre, the host gateway could be interspersed across the multiple VLT domains, distributing the ARP tables across for optimal scalability requirements.

In case of static routes, the same static routes with the next-hop should be configured on both the VLT peers for optimal routing. However, if the next-hop reachability for the uplink is different, the appropriate NH has to be configured for each VLT peer.

Layer-3 VLT supports both IPv4 and IPv6 dual stack. However '*peer-routing*' is currently supported only for IPv4. In case of IPv4 and IPv6 dual stack requirement, VRRP could be configured for resiliency in operation for the end-stations.

FTOS 9.2(0.0) release scales more number of VLANs (layer-3 and layer-2). With L3-VLT, ARP/MAC addresses learnt by one of the VLT peer would be synchronized in both the VLT nodes for the corresponding VLANs.

The following sections highlight the IGP/BGP configuration in VLT domains:

- OSPF Configured VLT Domain
- IS-IS configured VLT domain
- BGP configured VLT domain

3.6 OSPF Configured VLT Domain

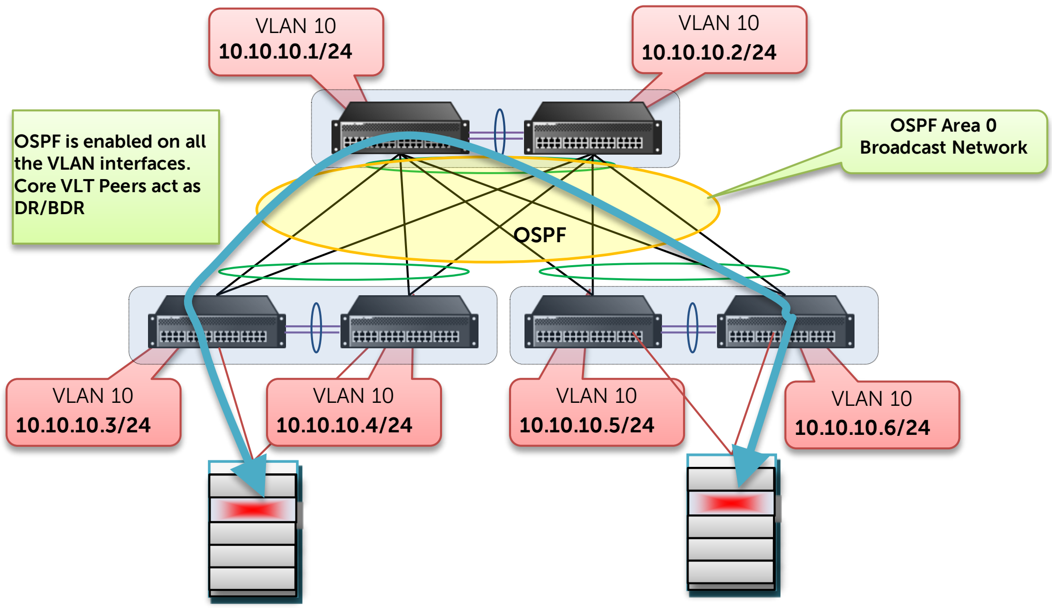

Figure 3.6 OSPF in VLT

OSPF is configured on VLAN interfaces as broadcast network (Default OSPF network). No other OSPF network (point-to-point) type is supported. Since all other VLT peers form adjacency with the Core VLT peers, as a best practice configure the primary node of the core vlt domain as DR and secondary node as BDR for optimal routing adjacencies. By configuring all the VLAN interfaces in the primary VLT peer of the aggregation layer, the priority could be incremented to act as OSPF DR. The number of OSPF adjacencies from each peer depends upon the VLAN subnets.

With OSPF converged, all the nodes have the routes and next-hop detail to reach other nodes. Albeit the VLAN interfaces are configured with OSPF, as a same broadcast domain, and underlying layer-2 dependency is achieved. For instance VM migration within the rack and across VLT domains could be orchestrated.

The OSPF configurations in VLT domain are as follows:

router ospf 1 router-id 10.10.10.1 network 192.168.0.3/16 area 0 network 192.169.0.3/16 area 0 **auto-cost reference-bandwidth 40000**	interface Vlan 10 ip address 192.168.0.3/16 tagged Port-channel 33 **ip ospf priority 100** no shutdown ! interface Vlan 20 ip address 192.169.0.3/16 tagged Port-channel 44 **ip ospf priority 100** no shutdown

3.7 IS-IS Configuration in VLT

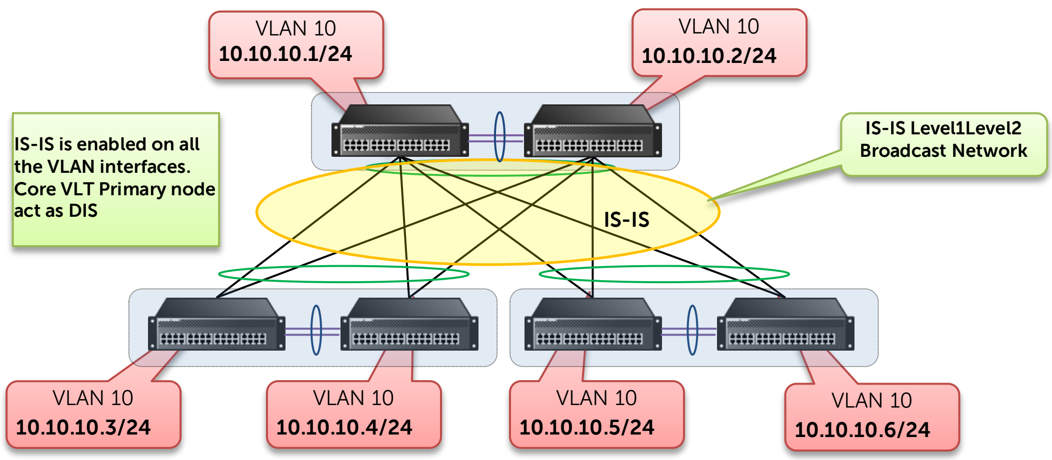

Figure 3.7 IS-IS in VLT

With multi-topology IS-IS, both IPv4 and IPv6 can be configured with single IS-IS process. Also the primary node of the core VLT domain could be configured as DIS, with priority of 100 across all VLAN interfaces, similarly priority 90 for all the interfaces in secondary node. However there is no concept of Backup DIS in IS-IS. This DIS mechanism reduces the LSP flooding in the LAN.

As a best practice set the '*metric-type*' as '*wide*' under IS-IS configuration for all the VLT nodes. Also configure the level as L1 or L2 or L1L2 (default). Configuring both level-1 and level-2 in the same VLT domain have reachability issues, unless route-leaking is configured.

The IS-IS configuration in VLT domain is as follows:

interface Vlan 10 ip address 100.1.1.1/24 ip router isis 1 **isis priority 100 level-1** **isis priority 100 level-2** no shutdown	router isis 1 description ISIS_VLT_Peer1 log-adjacency-changes **metric-style wide level-1** **metric-style wide level-2** net 49.0001.1111.2222.3333.4444.00

For the leaf VLT domains, the ISIS priority need not be configured. All other leaf VLT domains form the ISIS Level1/2 adjacencies with the Spine/Core VLT domain.

3.8 BGP in VLT

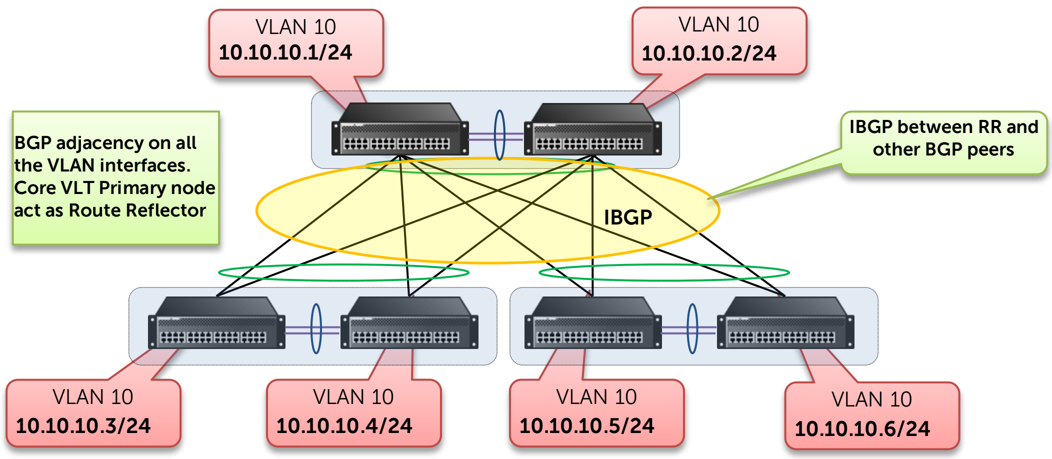

Figure 3.8 BGP in VLT

IBGP adjacencies are formed between the Core VLT domain nodes and all the other VLT nodes. As a best practice configure both the core VLT peers as Route-reflector. With core VLT domain acting as RR, BGP adjacencies are formed only between the RR and the Clients (other VLT domains).

Tweak the timers for faster convergence with keep-alive and hold time at a minimum.

BGP Configuration for Spine/Core VLT domain acting as Route-reflector(RR)

```
router bgp 65000
network 100.100.100.0/24
neighbor 192.169.0.1 remote-as 65000
neighbor 192.169.0.1 route-reflector-client
neighbor 192.169.0.1 no shutdown
neighbor 192.169.0.2 remote-as 65000
neighbor 192.169.0.2 route-reflector-client
neighbor 192.169.0.2 no shutdown
neighbor 192.169.0.4 remote-as 65000
neighbor 192.169.0.4 route-reflector-client
neighbor 192.169.0.4 no shutdown
neighbor 100:100:100:100::2 remote-as 65000
neighbor 100:100:100:100::2 route-reflector-client
neighbor 100:100:100:100::2 no shutdown
neighbor 100:100:100:100::3 remote-as 65500
neighbor 100:100:100:100::3 no shutdown
!
address-family ipv6 unicast
neighbor 100:100:100:100::2 activate
neighbor 100:100:100:100::2 route-reflector-client
neighbor 100:100:100:100::3 activate
exit-address-family
```

3.9 Layer-2 traffic in VLT Domain

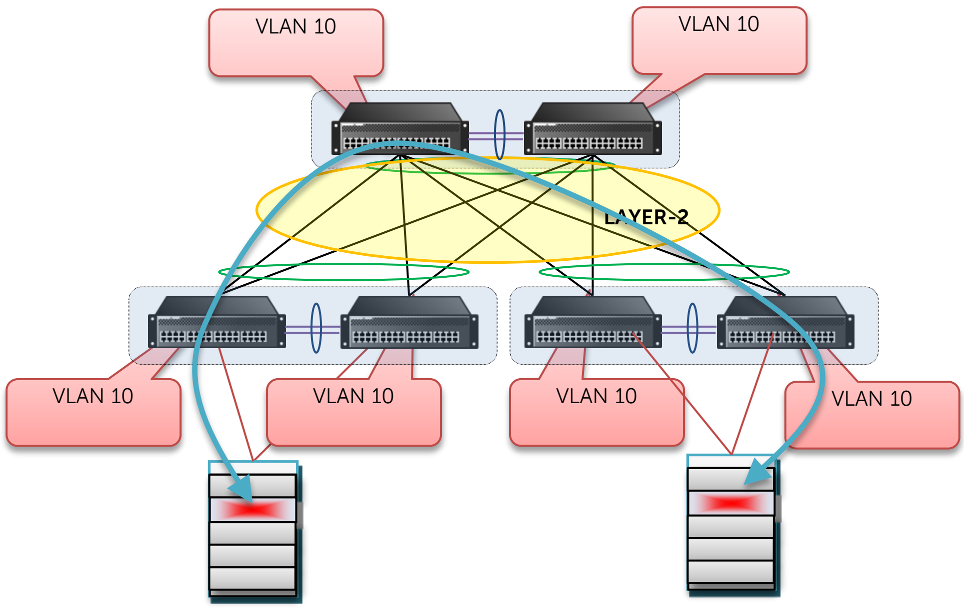

Figure 3.9 Layer-2 Traffic in VLT domain

Most of the Next gen Data centers envisages a continuous growth to meet their virtualization business needs scaling the layer-2 domain in progressive phases. In this illustration, VLAN 10, 30 and 50 configured as layer-2 spanning all the VLT domains. Any host in either of the VLT node gets synchronized with its VLT peer. MAC addresses of all the VLANs would be synchronized between the VLT peers.

```
VLT-10-PEER-1#show mac-address-table count
MAC Entries for all vlans :
Dynamic Address Count :                 1007
Static Address (User-defined) Count :   1
Sticky Address Count  :                 0
Total Synced Mac from Peer(N):          503
Total MAC Addresses in Use:             1008
```

```
VLT-10-PEER-1#show vlt counter mac
Total MAC VLT counters
----------------------
L2 Total MAC-Address Count:            1007
```

```
VLT-10-PEER-1#show mac-address-table

Codes: *N - VLT Peer Synced MAC
VlanId     Mac Address           Type          Interface        State
10      00:00:4c:54:8b:f6       Dynamic        Po 11            Active
10      00:01:e8:95:ec:97       Dynamic        Po 33            Active
10      00:01:e8:b3:ba:47       Dynamic        Po 33            Active
30      a0:00:a1:00:00:01       Dynamic        Po 11            Active
30      a0:00:a1:00:00:02       Dynamic        Po 11            Active
```

```
30      a0:00:a1:00:00:03        Dynamic         Po 11            Active
30      a0:00:a1:00:00:04        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:05        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:06        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:07        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:08        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:09        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:0a        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:0b        Dynamic (N)     Po 11            Active
30      a0:00:a1:00:00:0c        Dynamic         Po 11            Active
```

```
VLT-10-PEER-2#show vlt  statistics mac
VLT MAC Statistics
--------------------
L2 Info Pkts sent:0,    L2 Mac-sync Pkts Sent:7
L2 Info Pkts Rcvd:0,    L2 Mac-sync Pkts Rcvd:9
L2 Reg Request sent:0
L2 Reg Request rcvd:0

L2 Reg Response sent:0
L2 Reg Response rcvd:0
```

Any MAC address learnt from the connected host on the VLT node is instantaneously synchronized to its VLT peer. The VLT Peer synchronized MAC addresses are indicated with a (N) Flag to easily identify the MAC's learnt from its VLT peer. This MAC synchronization handles the traffic flow even if it is hashed and forwarded through the other member of the port-channel.

Network orchestration which requires a seamless migration of Virtual Machines could be achieved with this hierarchical multi VLT (mVLT) architecture which scales to the user needs, while effectively utilizing all the multipath links.

3.10 Layer-3 traffic in VLT domain

The following figure illustrates the interspersing of the Layer-3 VLANs and effectively scaling the ARP requirements within the designed VLT network fabric.

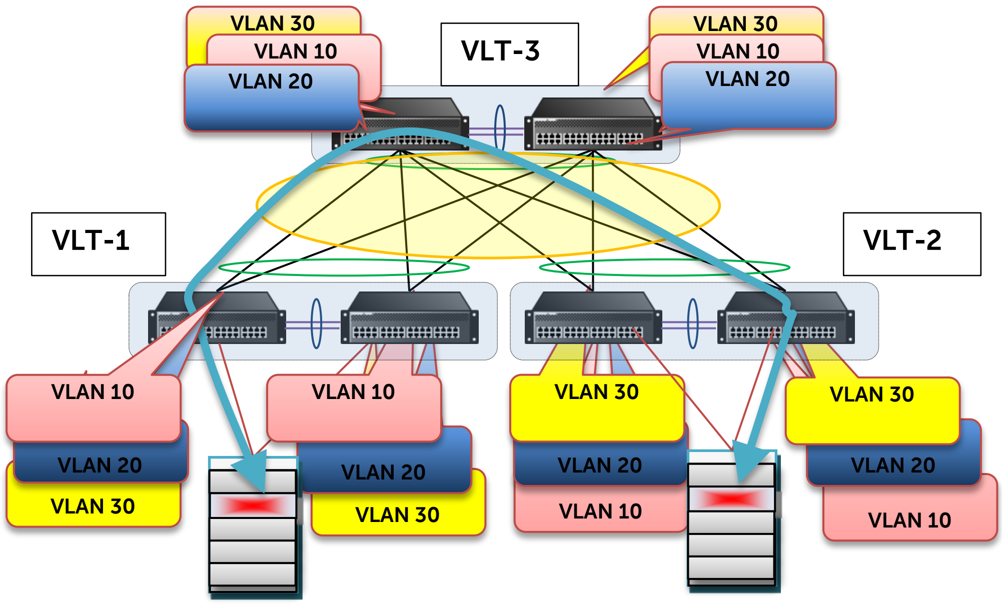

Figure 3.10 Layer-3 Traffic in VLT domain

From the above figure we have three VLT domains viz., VLT-1, VLT-2 and VLT-3 configured with three VLANs (VLAN10, 20 & 30). With routed-VLT, a VLAN could be configured as layer-3 in a VLT domain and layer-2 VLAN in all other VLT domains.

In this example, VLAN-10 is configured as layer-3 in VLT-1 and as layer-2 in other VLT domains. Similarly VLAN-30 configured as layer-3 in VLT-2 domain and as layer-2 in other VLT domains. The benefit of interspersing layer-3 IP addressing for these VLAN's distributes the ARP tables for the respective VLANs in the relevant VLT domains. Assuming 48 x 10G ports from each VLT node in a VLT domain connecting to 48 physical servers of dual NIC with approximate 50 VM's for each of the physical server necessitates 2400 ARP entries. By spreading the layer-3 VLANs across VLT domains, the ARP table requirement in each of the VLT domain could be optimally utilized and could scale beyond the limitation of single switch. ARP entries for VLAN 10 would be confined to VLT-1, similarly ARP entries for VLAN-30 in VLT-2 domain only.

At the core/aggregation layer VLT domain, common Layer-3 VLAN's are configured for inter vlan routing within the VLT domain. ARP scaling is well achieved by this design, while extending the VLAN connectivity across all the nodes.

The spine/core VLT domain forms adjacencies with the border routers advertising the end-station routes. The north bound traffic flows through the VLT node uplinks with the route table information.

The following table illustrates the ARP synchronization between the VLT nodes

```
VLT-10-PEER-1#show vlt  detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
11            11           UP            UP            10, 30, 50
33            33           UP            UP            10, 30, 50

VLT-10-PEER-1#show arp summary

Total Entries    Static Entries   Dynamic Entries   CPU
-----------------------------------------------------------------------------
----
2007             0                2007              CP
VLT-10-PEER-1#show vlt counter arp interface port-channel ?
<1-128>                Port channel identifier
VLT-10-PEER-1#show vlt counter arp interface port-channel 11
VLT Port-ID: 11 ARP Counter
-----------------------
Total Arp Entries Learnt :              1000
Total Arp Entries Synced :              1000
Total Non-VLT Arp entries Learnt:       0
Total Non-VLT Arp Entries Synced        0

VLT-10-PEER-1#show vlt statistics arp
VLT ARP Statistics
--------------------
ARP Tunnel Pkts sent:6
ARP Tunnel Pkts Rcvd:0
ARP Tunnel Pkts sent Non Vlt:0
ARP Tunnel Pkts Rcvd Non Vlt:0
ARP-sync Pkts Sent:83478
ARP-sync Pkts Rcvd:12442
ARP Reg Request sent:4
ARP Reg Request rcvd:3

VLT-10-PEER-2#show vlt brief
VLT Domain Brief
------------------
Domain ID:                      10
Role:                           Secondary
Role Priority:                  1000
ICL Link Status:                Up
HeartBeat Status:               Up
VLT Peer Status:                Up
Local Unit Id:                  1
Version:                        6(1)
Local System MAC address:       00:01:e8:8b:24:62
Remote System MAC address:      00:01:e8:8b:24:2c
Configured System MAC address:  a0:10:10:aa:aa:aa
Remote system version:          6(1)
Delay-Restore timer:            10 seconds
Peer-Routing :                  Enabled
Peer-Routing-timeout timer:     0 seconds
```

```
Multicast peer-routing timeout: 150 seconds


VLT-10-PEER-2#show vlt  statistics arp
VLT ARP Statistics
--------------------
ARP Tunnel Pkts sent:0
ARP Tunnel Pkts Rcvd:6
ARP Tunnel Pkts sent Non Vlt:0
ARP Tunnel Pkts Rcvd Non Vlt:0
ARP-sync Pkts Sent:17566
ARP-sync Pkts Rcvd:88581
ARP Reg Request sent:3
ARP Reg Request rcvd:4

VLT-10-PEER-2#show vlt counter
Total VLT Counters
-------------------
L2 Total MAC-Address Count:             2008
Total Arp Entries Learnt :              1000
Total Arp Entries Synced :              1002
Total Non-VLT Arp entries Learnt:       0
Total Non-VLT Arp Entries Synced        1
IGMP MRouter Vlans count :              1
IGMP Mcast Groups count  :              0
Total VLT Ndp Entries Learnt :          0
Total VLT Ndp Entries Synced :          0
Total Non-VLT Ndp Entries Learnt :      0
Total Non-VLT Ndp Entries Synced :      0
VLT-10-PEER-2#show vlt counter arp ?
interface               Interface statistics
|                       Pipe through a command
<cr>
VLT-10-PEER-2#show vlt counter arp
Total ARP VLT counters
----------------------
Total Arp Entries Learnt :              1000
Total Arp Entries Synced :              1002
Total Non-VLT Arp entries Learnt:       0
Total Non-VLT Arp Entries Synced        1

VLT-10-PEER-2#show vlt counter arp interface port-channel 11
VLT Port-ID: 11 ARP Counter
-----------------------
Total Arp Entries Learnt :              1000
Total Arp Entries Synced :              1000
Total Non-VLT Arp entries Learnt:       0
Total Non-VLT Arp Entries Synced        0
VLT-10-PEER-2#
```

3.11 IPv6 addressing in VLT

IPv6 addressing is supported in VLT domains from FTOS 9.2(0.0) release. The VLT domains could be configured with IPv4 or IPv6 address or with dual stack of IPv4 and IPv6 together.

However "*peer-routing*" is currently supported only for IPv4. In case of dual stack requirement, it is recommended to configure VRRP for resiliency between the VLT peers for both IPv4 and IPv6 hosts. With IPv6 in VLT, NDP messages are synchronized between the VLT peers, "show vlt statistics" indicates the NDP message statistics.

```
VLT-100-PEER-1#show running-config interface vlan 50
!
interface Vlan 50
ip address 50.50.50.1/24
ipv6 address 50:50:50:50::1/64
tagged Port-channel 33
ip ospf priority 100
!
vrrp-ipv6-group 50
  priority 200
  virtual-address 50:50:50:50::5
  virtual-address fe80::50:50:50:5
no shutdown
VLT-100-PEER-1#show vrrp ipv6 brief

Interface Group    Pri Pre State  Master addr     Virtual
addr(s)                    Description
-----------------------------------------------------------------------------
---------------------
Vl 50     IPv6 50  200 Y   Master fe80::201:e8f...50:50:50:50::5
fe80::50:50:50:5
```

```
VLT-100-PEER-2#show running-config interface vlan 50
!
interface Vlan 50
ip address 50.50.50.2/24
ipv6 address 50:50:50:50::2/64
tagged Port-channel 33
ip ospf priority 90
!
vrrp-ipv6-group 50
  virtual-address 50:50:50:50::5
  virtual-address fe80::50:50:50:5
no shutdown
VLT-100-PEER-2#show vrrp ipv6 brief

Interface Group    Pri Pre State  Master addr     Virtual
addr(s)                    Description
-----------------------------------------------------------------------------
---------------------
Vl 50     IPv6 50  100 Y   Backup fe80::201:e8f...50:50:50:50::5
fe80::50:50:50:5
```

3.12 mVLT in MXL Switches

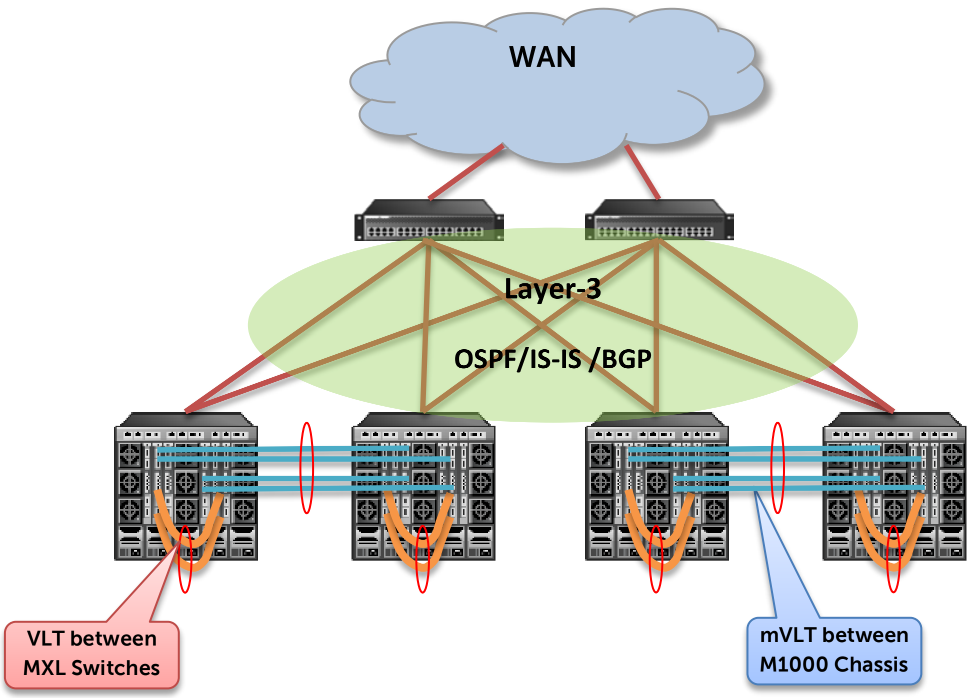

Figure 3.12 mVLT in MXL Switch

mVLT between M1000 Chassis - With FTOS 9.2(0.0) release, VLT is supported between MXL Switches within a Chassis (*Intra Chassis*) and extending as mVLT between M1000 Chassis (*Inter Chassis*). With this mVLT architecture, the VM's in the blade servers within the chassis could migrate to other chassis blade servers. This flexible and scalable architecture predominantly serves East-West traffic within the data centers. The upstream traffic extends through the uplink to the ToR/Core Switch with layer-3 connectivity.

This architecture focus on handling multiple VM's in the Blade servers within same VLAN's across multiple M1000 Chassis. VM migration within the rack and across the racks within the data center could be deployed. High availability is incorporated at all layers from the active servers to the network connectivity.

Active System Infrastructure with appropriate profiles could be deployed to aggregate the access traffic matching the customer workloads.

3.13 Auto VLT in IOA

Auto VLT deployment in IOM is supported from FTOS 9.2(0.0) release. The VLT domain with dual IOM in M1000 Chassis could be enabled with a single CLI:

stack-unit 0 iom-mode vlt

The above CLI enables VLT mode within the IOM blades.

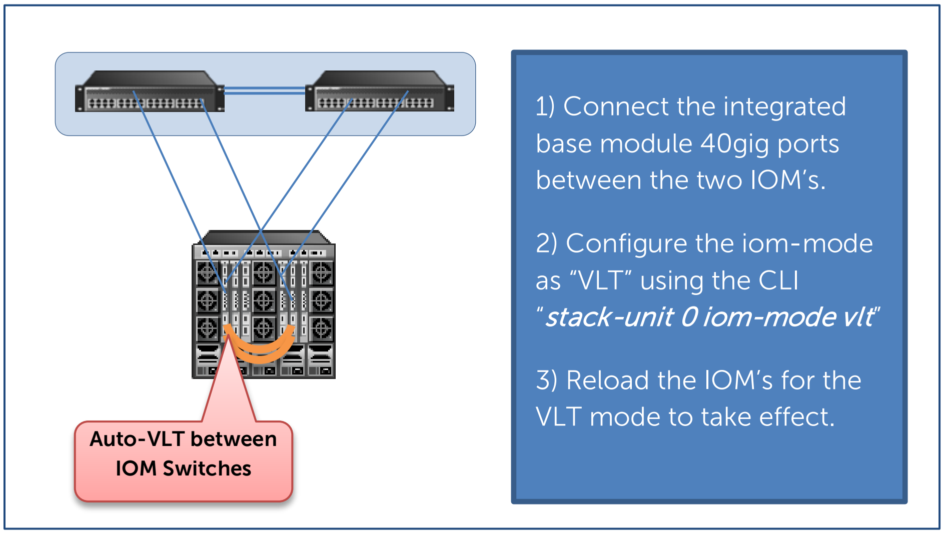

Figure 3.13 Auto VLT in IOA

Prerequisite:

1) Connect the integrated base module 40gig ports between the two IOM's.
2) Configure the iom-mode as "VLT" using the CLI "*stack-unit 0 iom-mode vlt*"
3) Reload the IOM's for the VLT mode to take effect.

Important points in IOA VLT

- With VLT in IOA, VLTi (ICL) port-channel will be always configured as Po 127 (static port-channel).
- Uplink VLT port-channel will always be Po 128 dynamic (with LACP)
- Server-side VLT port-channel will be always auto-assigned (LACP) between the port-channel ranges (1-126). ICL will be established between base module forty GIG ports (Fo 33 and 37)
- Uplink VLT port-channel will be established between ports 41 – 56 (optional module ports)
- Server-side VLT LAG will be established between ports 1 – 32.
- Back-up link will be established between the IOM's IP address (169.254.31.X)

```
FTOS(conf)#
FTOS(conf)#stack-unit 0 iom-mode ?
programmable-mux        Programmable Mux mode
stack                   Stack mode
standalone              Standalone mode
vlt                     Vlt mode
FTOS(conf)#stack-unit 0 iom-mode vlt
% You are about to configure Auto VLT to your IOA module, please reload the
IOA and then plug in the ICL cable for the changes to take effect.
```

FTOS#show run int po 127 ! interface Port-channel 127 mtu 12000 channel-member fortyGigE 0/33,37 no shutdown FTOS#	FTOS#show run int po 128 ! interface Port-channel 128 mtu 12000 portmode hybrid switchport fip-snooping port-mode fcf vlt-peer-lag port-channel 128 no shutdown FTOS#

```
FTOS#show vlt brief
 VLT Domain Brief
------------------
 Domain ID:                      1
 Role:                           Primary
 Role Priority:                  32768
 ICL Link Status:                Up
 HeartBeat Status:               Up
 VLT Peer Status:                Up
 Local Unit Id:                  0
 Version:                        6(1)
 Local System MAC address:       00:01:e8:43:00:08
 Remote System MAC address:      00:1e:c9:f1:02:0a
 Configured System MAC address:  00:01:00:03:00:02
 Remote system version:          6(1)
 Delay-Restore timer:            90 seconds
 Peer-Routing :                  Disabled
 Peer-Routing-timeout timer:     0 seconds
 Multicast peer-routing timeout: 150 seconds
FTOS#

FTOS#show vlt backup-link
 VLT Backup Link
-----------------
 Destination:                    169.254.31.25
 Peer HeartBeat status:          Up
 HeartBeat Timer Interval:       1
 HeartBeat Timeout:              3
 UDP Port:                       34998
 HeartBeat Messages Sent:        449505
 HeartBeat Messages Received:    449496
FTOS#
```

```
FTOS#show vlt detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
1             1            UP            UP            1
128           128          UP            UP            1
FTOS#


FTOS#show int po brief
Codes: L - LACP Port-channel

    LAG  Mode  Status       Uptime      Ports
L   1    L2    up           5d6h5m      Te 0/6     (Up)
    127  L2    up           5d6h5m      Fo 0/33    (Up)
                                        Fo 0/37    (Up)
L   128  L2    up           5d6h5m      Te 0/41    (Up)
                                        Te 0/42    (Up)
FTOS#

FTOS#show run int po 1 (auto config)
!
interface Port-channel 1
 mtu 12000
 portmode hybrid
 switchport
 vlt-peer-lag port-channel 1
 no shutdown
FTOS#
FTOS#show vlan
Codes: * - Default VLAN, G - GVRP VLANs, R - Remote Port Mirroring
VLANs, P - Primary, C - Community, I - Isolated
       O - Openflow
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   o - OpenFlow untagged, O - OpenFlow tagged
   G - GVRP tagged, M - Vlan-stack, H - VSN tagged
   i - Internal untagged, I - Internal tagged, v - VLT untagged, V -
VLT tagged

    NUM    Status    Description                     Q Ports
*   1      Active                                    U Po1(Te 0/6)
                                                     U Po127(Fo
0/33,37)
                                                     U Po128(Te 0/41-
42,44)
                                                     U Te 0/1-5,7-32
FTOS(conf)#int ten 0/6
FTOS(conf-if-te-0/6)#vlan tagged ?
VLAN-RANGE              Comma/Hyphen separated VLAN ID set
FTOS(conf-if-te-0/6)#vlan tagged 4 >>>>>>> Configure VLAN
FTOS(conf-if-te-0/6)#do show vlan

Codes: * - Default VLAN, G - GVRP VLANs, R - Remote Port Mirroring
VLANs, P - Primary, C - Community, I - Isolated
```

```
       O - Openflow
Q: U - Untagged, T - Tagged
   x - Dot1x untagged, X - Dot1x tagged
   o - OpenFlow untagged, O - OpenFlow tagged
   G - GVRP tagged, M - Vlan-stack, H - VSN tagged
   i - Internal untagged, I - Internal tagged, v - VLT untagged, V -
VLT tagged

    NUM    Status    Description                     Q Ports
*   1      Active                                    U Po1(Te 0/6)
                                                     U Po127(Fo

0/33,37)
                                                     U Po128(Te 0/41-

42,44)
                                                     U Te 0/1-5,7-32
    4      Active                                    T Po1(Te 0/6)
                                                     T Po128(Te 0/41-

42,44)
                                                     V Po127(Fo

0/33,37)
FTOS(conf-if-te-0/6)#
```

3.14 LAG features

16-member Link Aggregation Group

FTOS 9.2(0.0) release supports 16-member LAG. This would enhance the bundling capacity to the VLT nodes handling the traffic. Similarly 16 member port-channel LAG from the ToR, with 8 links connecting to VLT peer-1 and another 8 links connecting to VLT Peer-2 would handle less oversubscribed fabric requirement for specific customers. The following figure represents a 16 port 10G links from S4810 switch connected to 40G breakout of Z9000 Switch. This LAG support would generally assist for the large East-West traffic pattern based on the customer's unique workloads.

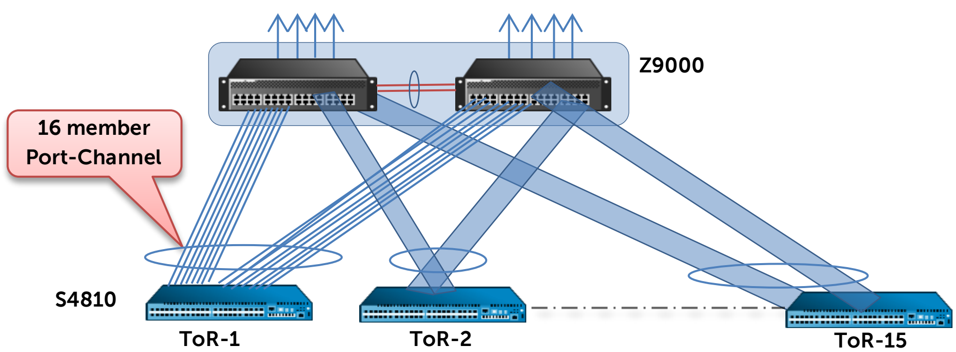

Figure 3.14 16-Members LAG in VLT

VLT LAG Scalability

128 VLT LAG Port-Channels per node could be supported in FTOS 9.2. With Z9000/S6000 Switches (32x40G or 128x10G ports), each port could be formed as VLT LAG connecting to its peer for highly over-subscribed environments such as Campus Networks, Small/medium Enterprise Networks etc.,

Depending upon the specific requirement of fabric subscription, the number of user ports could scale from 720x10G ports to 3600x10G and beyond.

4 VLT Peers connected with 10G VLTI

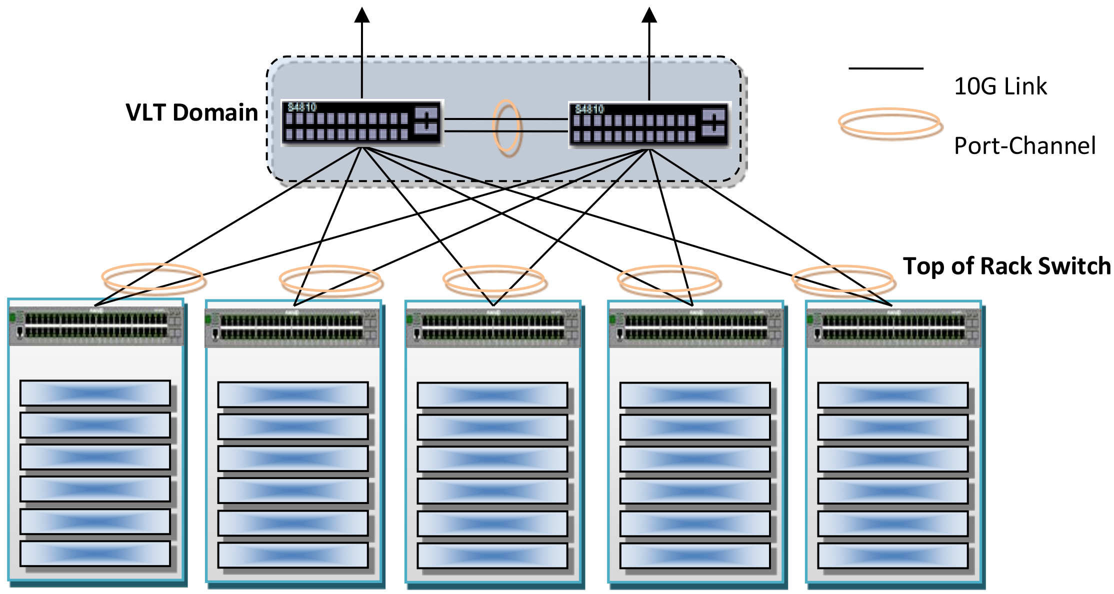

Figure 4.0 VLT with 10G VLTi links

VLT Peers connected through 10G VLTi links.

This topology represents the oversubscribed ToR to VLT aggregation switch. Any failure in the VLT links would force the traffic through the diverse path of the other VLT peer and may go via VLTI depending upon the topology. This setup could be deployed in small and medium enterprise networks, data centers, campus networks connecting server and storage clusters.

Generally, north to the VLT domain, aggregator would be connected to the core routers for north bound external traffic. Being layer 3 at the core the IGP could be extended up to the VLTI link with 'peer-routing'. IGP (OSPF/IS-IS) could be configured in the VLAN interfaces, which forms adjacency between the VLT peers,

With the MAC, ARP tables, IGMP State synchronization within the VLT peers, the VLT switches represent a single logical switch to the underlying Layer-2 ToR switches ensuring high availability and resiliency in operation.

5 VLT Peers connected with 40G VLTI

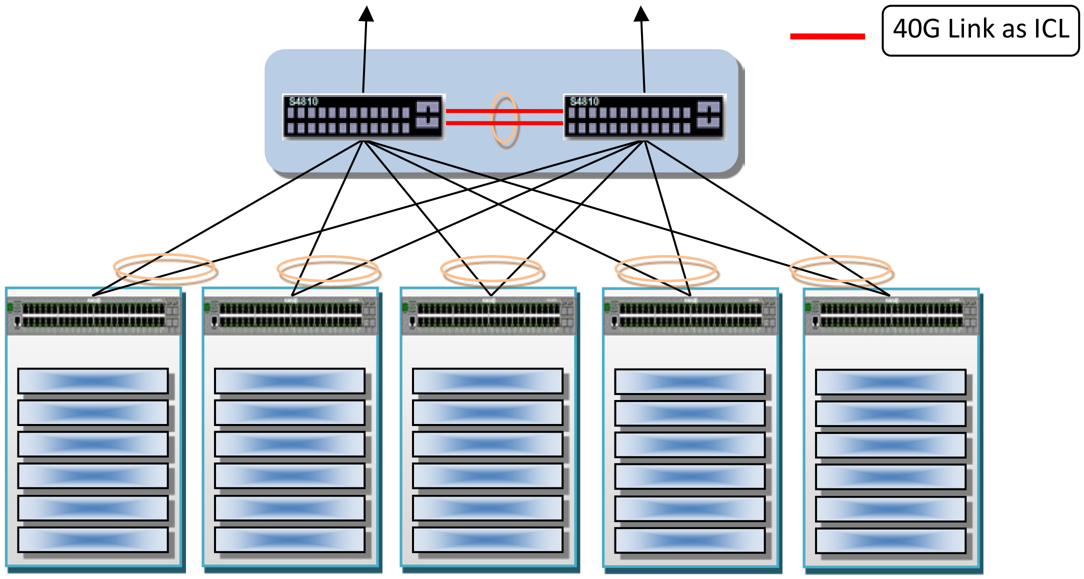

Figure 5.0 VLT with 40G VLTi links

VLT Peers connected through 40G VLTI links.

This topology could be used for non-blocking and oversubscribed ToR to VLT aggregation. Any failure in the VLT links would force the traffic through the diverse path of the other VLT peer and may go via VLTi depending upon the topology. This VLTi with two or more 40G can handle major traffic flow during failure scenarios. This setup could be deployed in Enterprise networks with Multicast traffic streams and High performance computing (HPC) Data centers connecting to server and storage Clusters.

Seamless VM migration within the rack and between the racks could be ensured with this topology. As all the port-channel links are effectively utilized, latency and congestion in traffic is considerably less comparing to the conventional dual or collapsed core network model.

6 VLT Peers connected to standalone Switches

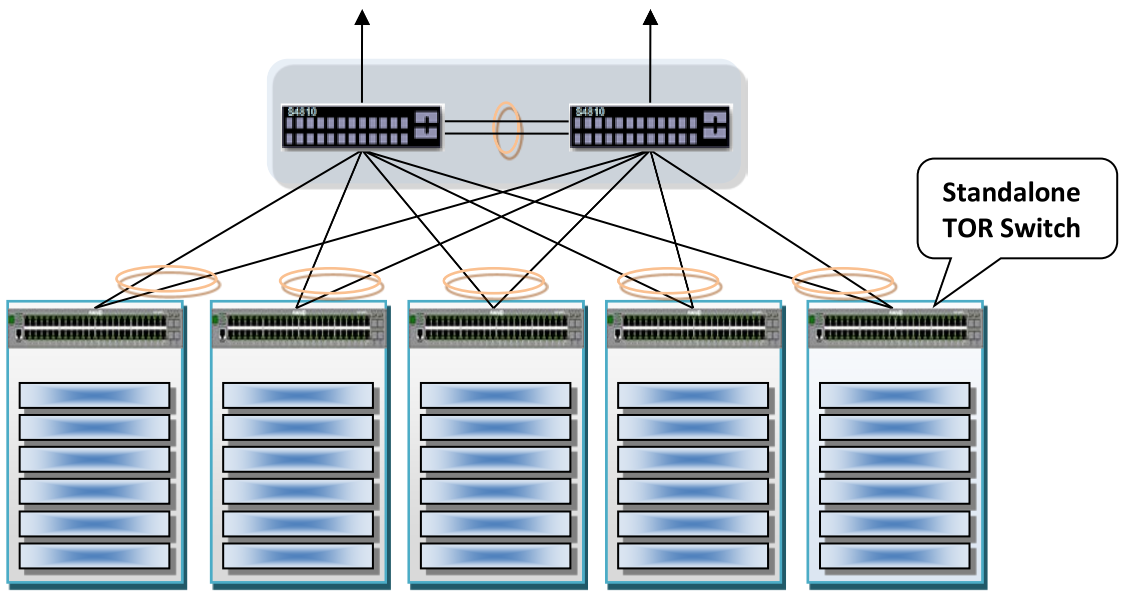

Figure 6.0 VLT domain connected to Standalone Switches

This simple topology represents the oversubscribed/non-blocking Top-of-Rack (ToR) to VLT aggregation. All the ToR switches could be standalone rack mounted switch with port-channels extending to VLT domains. This setup could be deployed in Campus, Enterprise networks connecting to server and storage Clusters. Albeit this is a simple and cost-effective setup, the redundancy at the ToR layer switch has to be matched with NIC teaming from the servers connecting to the adjoining ToR switches in the other racks. The physical implementation of this topology mandates the adjoining racks, ToR switches and servers having same VLAN. With dual NICs in the servers, the Twinax DAC Copper cables can be used to connect both the ToR switches.

7 VLT Peers connected to Stacked Switches

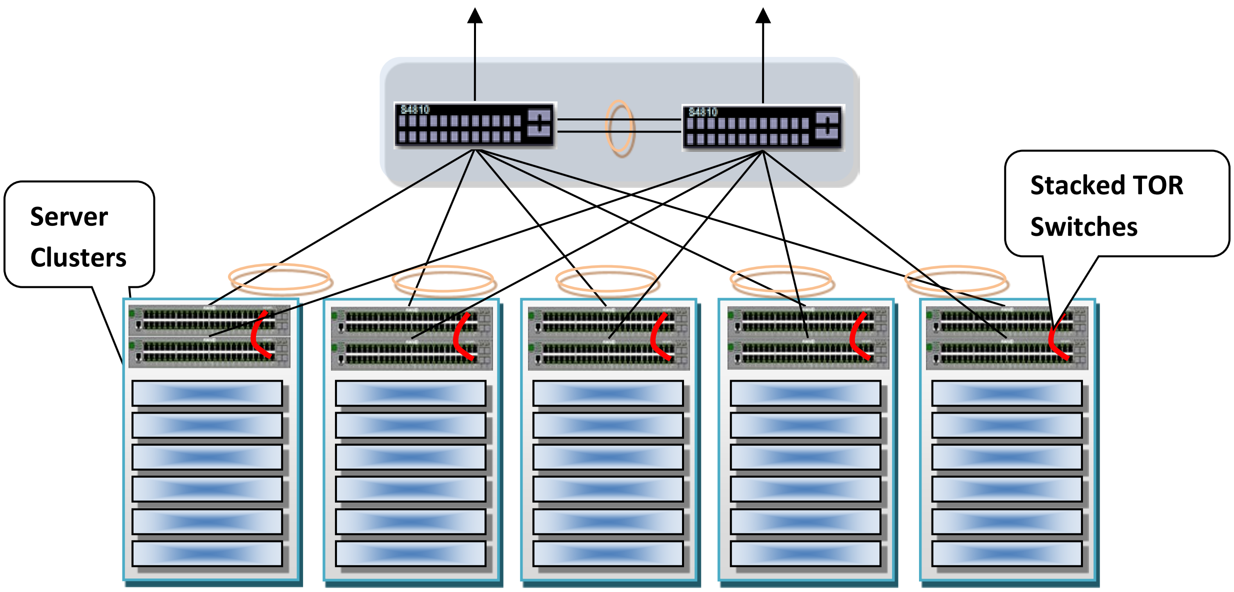

Figure 7.0 VLT peers connected to Stacked Switches

This simple topology represents the oversubscribed ToR to VLT aggregation. All the ToR switches are stacked switches with port-channels from stacked members extending to VLT domains. This setup could be deployed in Campus, Enterprise networks connecting to Server and Storage Clusters.

Stacked ToR switches offer redundancy at multiple levels for the Server clusters. Port-channel interface need to be connected from the stacked members to the VLT peers. Stacked ToR, Dual NICs in server, forms the extended access redundancy offering high availability for the network besides the VLT domain.

This topology offers over-subscribed ToR switch connecting the VLT domains. Based on the individual customer needs, and for the high port density requirement, stacked ToR's could be deployed connecting to VLT domains. However, the VLT nodes cannot be stacked.

8 Single VLAN across two ToR

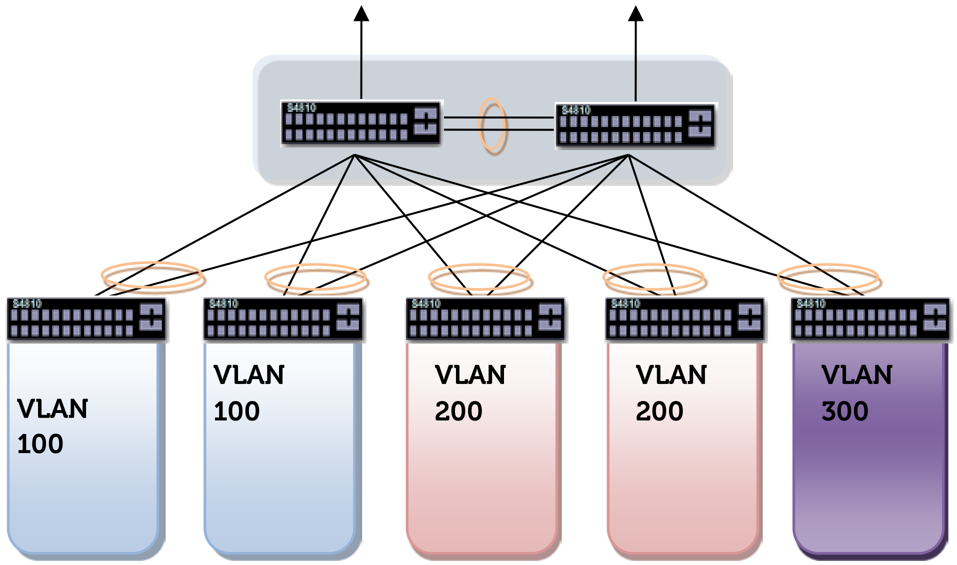

Figure 8.0 Single VLAN across ToR Switch

Single VLAN across two ToR

This is an idealistic topology, where single VLAN is configured for each ToR switch. Not many customers deploy this topology except few having multiple replication server/storage for specific application, workload aware VM migration, web-services with multi proxy web servers etc.

9 Multi VLANs across multi ToR

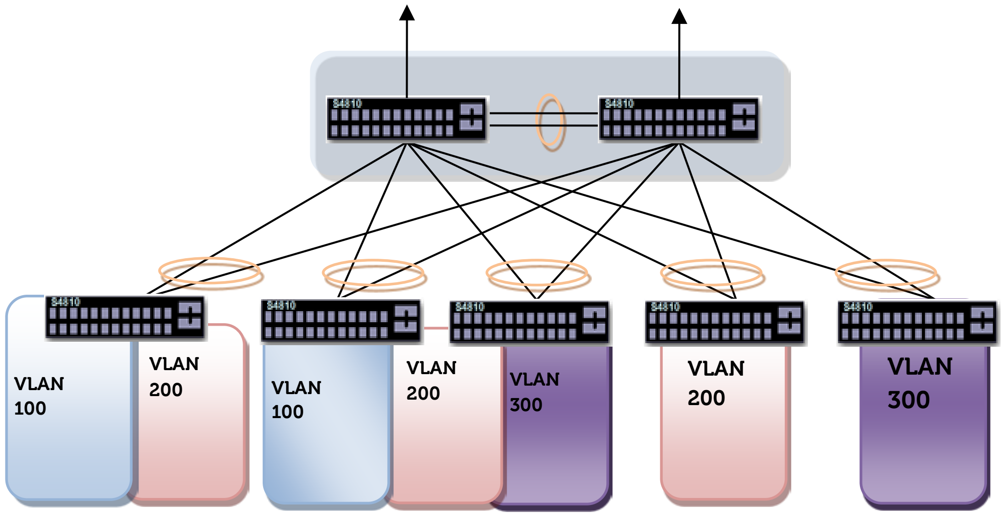

Figure 9.0 Multi VLAN across ToR

Multi VLANs across multi ToR

This topology is widely deployed with multi VLANs spanning multiple ToR switches connecting to the VLT domain. Redundancy at Server level is achieved by deploying multi services at multiple ToR switches. Server Load balancing could be deployed in these scenarios.
For all the VLANs created at the ToR switch, the VLT links need to be configured as tagged port for all those VLANs terminating at the VLT peers. The VLANs at the VLT peer would be configured as L3-VLAN with the corresponding IP addresses configured on both the VLT peers.

10 Inter-VLAN routing between ToR's

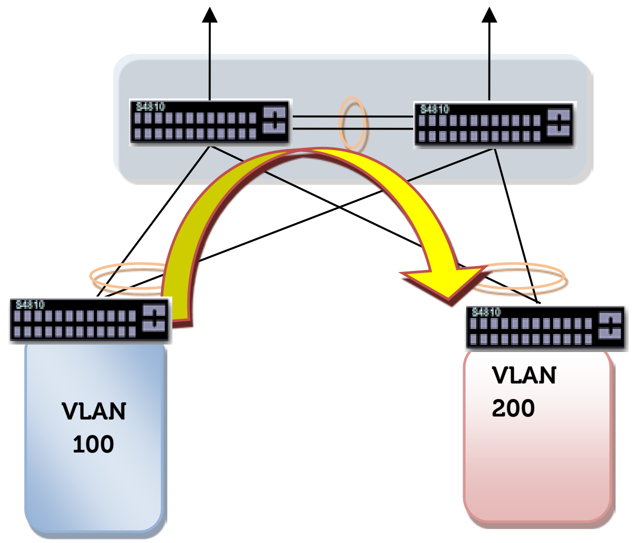

Figure 10.0 Inter VLAN routing in VLT

With the VLT links connected to the VLT domain, inter-vlan routing between multiple Layer-3 VLANs in the VLT domain happens at the receiving VLT peer and pass through the VLT port-channel of the destination VLAN. Due to the ARP and MAC Synchronization, the egress port details are synchronized in both the VLT peers. Any failure in the node/link is seamlessly handled.

With '*peer-routing*' feature, OSPF/IS-IS/BGP is supported for routing across VLAN subnets.

11 VLT in distributed core

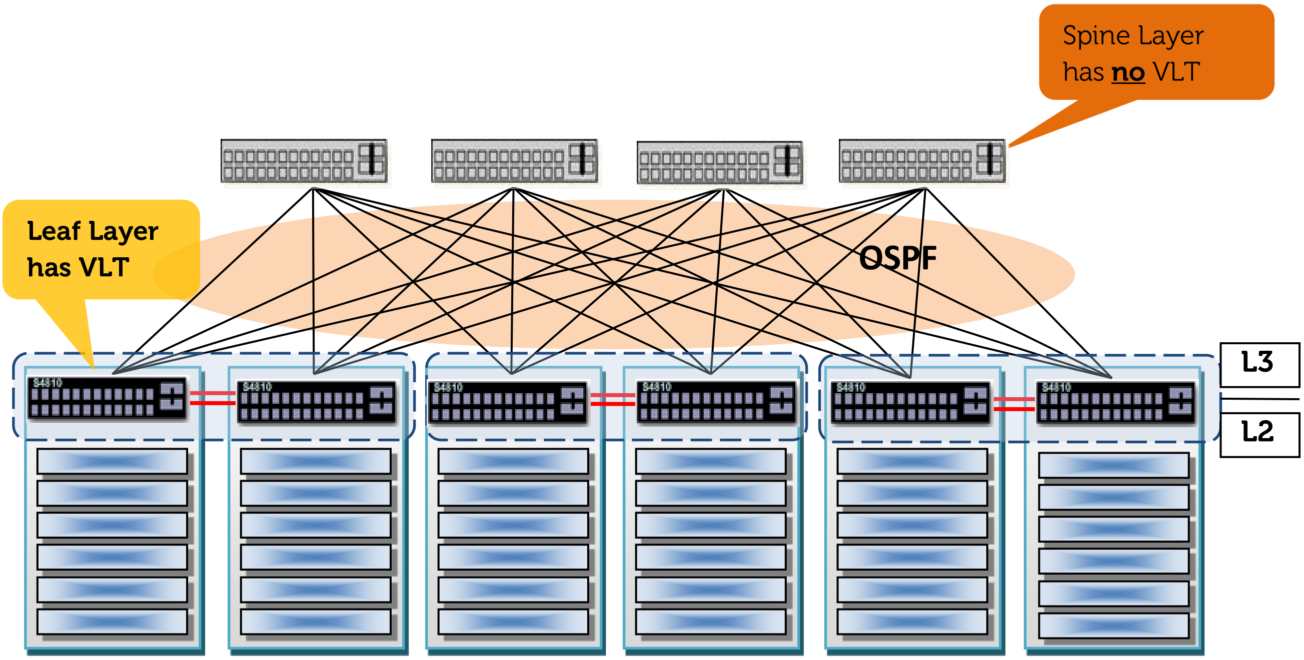

Figure 11.0 VLT in distributed Core

In a distributed Core with multi Spine and Leaf nodes, VLT has to be implemented in Leaf Nodes. VLT in *Spine layer* may not scale well in a distributed core network (CLOS) architecture. If the multiple nodes in the Spine layer has VLT configured, with leaf node VLT connecting to Spine node VLT (as in mVLT) without any redundant links, this could be a normal multi VLT topology. however if the redundant links, as in a distributed core network, connects to the other leaf VLT in domain-2, spanning the same VLANs, then there is a possibility of loop in the fabric for a layer-2 topology. Deploying STP to avoid loops could block such multi-links in distributed core. Further, implementing VLT in the Spine layer could deprive the available capacity of the fabric, as few ports would be used for VLTi.

With layer-3 in the fabric, there is not much advantage of VLT implemented in the spine layer. However, with VLT in the leaf nodes, this could be a major advantage, as NIC teaming or Server load balancing, redundancy in the ToR level could be achieved, as the access would be of layer-2.

The **Active Fabric Manager** (AFM 1.5) has the VLT features incorporated for auto configuring the VLT peers and the ToR switches based on the user's requirement, besides managing the distributed core. With auto configuring VLT domains, VLTI ports and the VLT port-channels, it would be much easier for the customer to manage the VLT domain through AFM.

11.1 Simplified illustration of VLT in Leaf layer

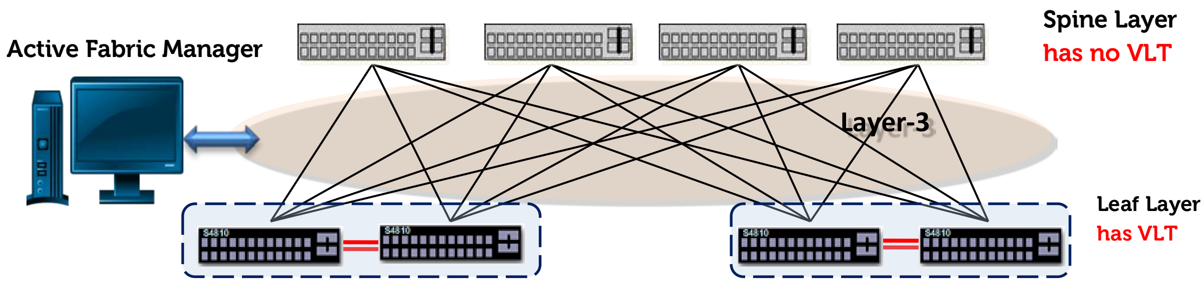

Figure 11.1 VLT in Leaf layer

VLT in Spine Layer

Distributed Core with VLT nodes in Spine layer and multiple VLT domains in the leaf layer could be deployed by AFM 1.5. (Figure 11.2) The uplinks would be terminated at the Spine layer unlike the leaf layer in the conventional distributed core. The dual Spine nodes are of Z9000 and leaf nodes are S4810. 40G links are interlinked between Z9000 and S4810 in this architecture.

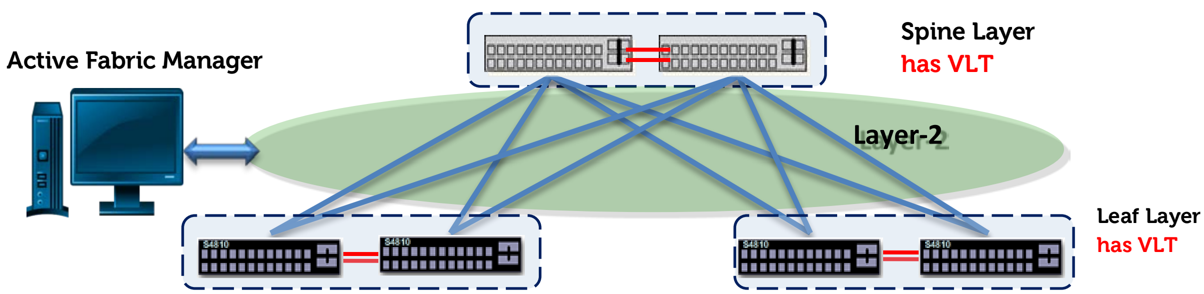

Figure 11.2 VLT in Spine and leaf layer

The summary view of AFM1.5 to design and deploy VLT domains as in figure 11.3

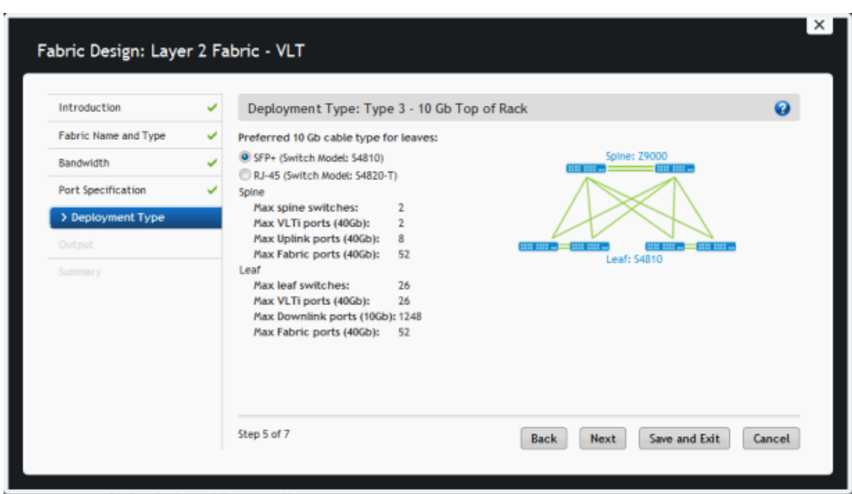

Figure 11.3 AFM fabric design summary view

12 STP flavors across ToR

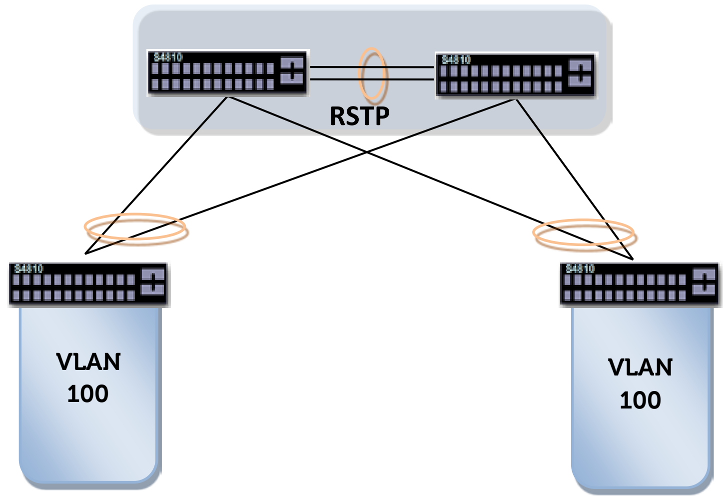

Figure 12.0 STP in VLT

Current VLT feature supports only RSTP in the VLT domain. No other STP flavors are supported in the VLT domain. Deploying RSTP in the VLT domain improves the convergence during the VLT and ToR nodes coming up. Moreover it eliminates any loop due to misconfigured ports. RSTP states are synchronized in both VLT peers; any RSTP BPDU received on the secondary VLT peer would be tunneled to the primary VLT peer for process. Only the primary VLT peer generates and forwards the BPDU on the VLT Port.

RSTP in this VLT domain would keep all the ports in the **FWD** state, as there is no loop in the topology. Primarily RSTP is required to isolate network loop in case of incorrect cabling in VLT domain. As a best practice, always keep the primary VLT member as the Root Bridge, VLT Secondary node with bridge priority higher than primary and lower than ToR, and do not change the bridge-priorities in the ToR/Access switches. Root-guard can be enabled on VLT port-channels. Also configure the host connected ports as edge-ports in the ToR to avoid TCN when servers are booted.

F10S4810AGGR2#**show spanning-tree rstp interface port-channel 40**

Port-channel 40 is designated Forwarding
Port is a Virtual Link Trunk port

Edge port:no (default) port guard :none (default)
Link type: point-to-point (auto) bpdu filter:disable (default)
Bpdu guard :disable bpduguard shutdown-on-violation : disable RootGuard: disable LoopGuard disable
Bpdus sent 29, received 0

```
Interface                                   Designated
 Name     PortID  Prio Cost  Sts          Cost    Bridge ID        PortID
--------- -------- ---- ------- ----------- ------- -------------------- --------
Po 40     128.41   128  1800   FWD(vlt)   600    32768 0001.e88b.1c9a 128.41
```

```
F10S4810AGGR2#


F10S4810AGGR2#show spanning-tree rstp interface port-channel 90

Port-channel 90 is root forwarding
Port is a Virtual Link Trunk Interconnect port

Edge port:no (default) port guard :none (default)
Link type: point-to-point (auto) bpdu filter:disable (default)
Bpdu guard :disable bpduguard shutdown-on-violation : disable RootGuard: disable LoopGuard
disable
Bpdus sent 1, received 31

Interface                                Designated
 Name     PortID  Prio Cost   Sts       Cost      Bridge ID        PortID
--------- -------- ---- ------- ----------- ------- -------------------- --------
Po 90     128.91   128  600     FWD(vltI) 600     4096  0001.e88b.1d3f 128.91
F10S4810AGGR2#
F10S4810AGGR2#
```

The VLTi link would be always in the STP forward state. The Interface status would be indicated as "*Forward*" with "*VLT*" and "*VLTi* "keyword as indicated in the above CLI snapshot.

RSTP Configuration

```
protocol spanning-tree rstp
 no disable
 bridge-priority 0     <<<<<<< Apply "0" for Primary VLT node and " 4096" for secondary VLT node
forward-delay 4
hello-time 1
```

13 RSTP in VLT Peers and MSTP in the ToR

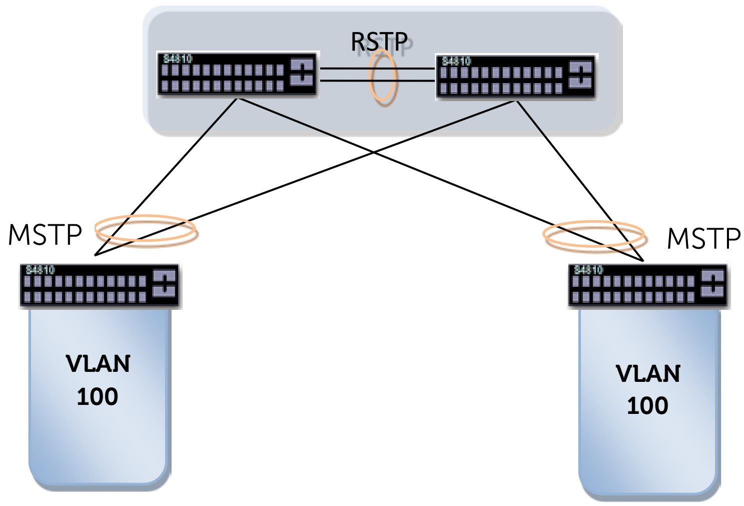

Figure 13.0 RSTP and MSTP in VLT

With MSTP configured in the ToR switches, they fall back to RSTP, once they receive the RSTP BPDU from the VLT Primary switch and converge with RSTP. As a best practice ensure the VLT Primary node as Root Bridge and secondary node as Secondary Root Bridge.

```
F10S4810AGGR1(conf-rstp)#do show spanning-tree rstp brief | except DIS
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 4096, Address 0001.e88b.1d3f
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 4096, Address 0001.e88b.1d3f
We are the root
Configured hello time 2, max age 20, forward delay 15

Interface                              Designated
 Name      PortID  Prio Cost   Sts       Cost      Bridge ID       PortID
---------- -------- ---- ------- ----------- ------- -------------------- --------
Po 40     128.41  128  1800   FWD(vlt)   0      4096  0001.e88b.1d3f 128.41
Po 90     128.91  128  600    FWD(vltI)  0      4096  0001.e88b.1d3f 128.91

Interface
 Name      Role   PortID   Prio Cost   Sts       Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 40     Desg   128.41   128  1800    FWD       0      (vlt) P2P No
Po 90     Desg   128.91   128  600     FWD       0      (vltI)P2P No
F10S4810AGGR1(conf-rstp)#
```

14 RSTP in VLT peer and PVST+ within the ToR's

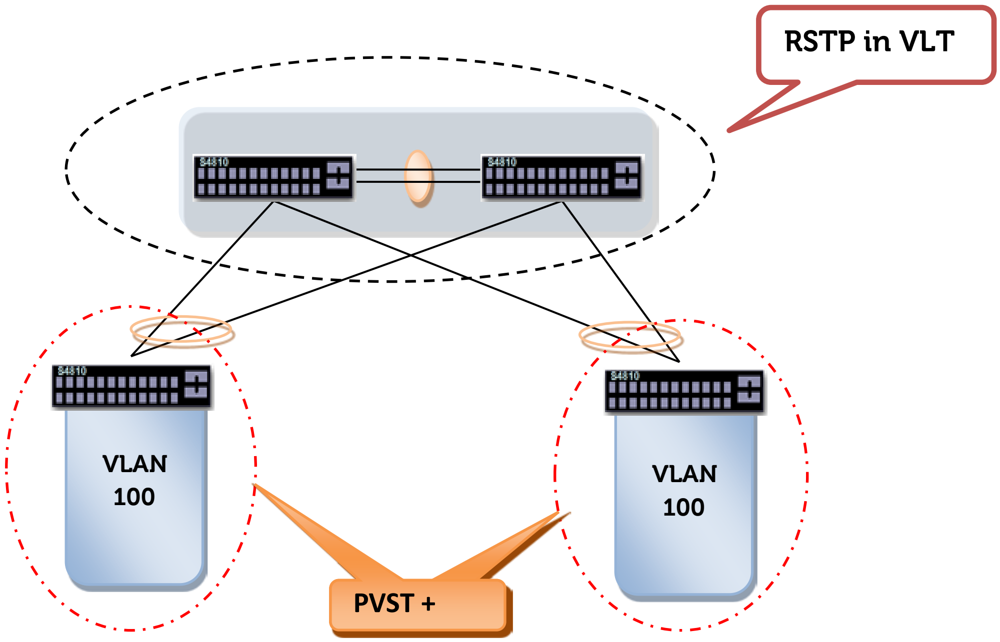

Figure 14.0 RSTP and PVST+ in VLT

When the VLT domain with RSTP is connected to ToR's configured with PVST+, the PVST + BPDUs received on the VLT peer get tunneled into the VLT domain for the non-default vlans. Few customer scenarios with multi-vendor device might have PVST+ configured in their ToR switches. As a best practice for this scenario, tag the VLT ports for the respective non-default VLANs, for the PVST BPDU's to pass through the VLT domain for convergence.

Configuration example:

ToR1 with RPVST+ enabled

spanning-tree ***mode rapid-pvst***
spanning-tree etherchannel guard misconfig
spanning-tree extend system-id

```
!
VLAN0002
  Spanning tree enabled protocol rstp
  Root ID    Priority    32770
             Address     001d.a23e.0d00
             This bridge is the root          <<<<<<<<<<<<<< (TOR 1) is root for vlan 2
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    32770  (priority 32768 sys-id-ext 2)
             Address     001d.a23e.0d00
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
             Aging Time  300 sec
Interface           Role Sts Cost      Prio.Nbr Type
------------------- ---- --- --------- -------- --------------------------------
Po10                Desg FWD 4         128.560  P2p
```

```
VLAN0003
  Spanning tree enabled protocol rstp
  Root ID    Priority    32768
             Address     0001.e8a4.fc7e    <<<<<<<<<<<< bridge id of TOR2 (Z9000)
             Cost        4
             Port        560 (Port-channel10)
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
  Bridge ID  Priority    32771  (priority 32768 sys-id-ext 3)
             Address     001d.a23e.0d00
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
             Aging Time  300 sec
Interface          Role Sts Cost      Prio.Nbr Type
------------------- ---- --- --------- -------- --------------------------------
Po10               Root FWD 4         128.560  P2p

VLAN0004
  Spanning tree enabled protocol rstp
  Root ID    Priority    32768
             Address     0001.e8a4.fc7e    <<<<<<<<<<< TOR 2 is root for vlan 4
             Cost        4
             Port        560 (Port-channel10)
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
  Bridge ID  Priority    32772  (priority 32768 sys-id-ext 4)
             Address     001d.a23e.0d00
             Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
             Aging Time  300 sec
Interface          Role Sts Cost      Prio.Nbr Type
------------------- ---- --- --------- -------- --------------------------------
Po10               Root FWD 4         128.560  P2p
```

Z9000 (ToR2) Output :

```
FTOS(conf-pvst)#do show span pvst vlan 2 brief
VLAN 2
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 32770, Address 001d.a23e.0d00
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 61440, Address 0001.e8a4.fc7e
Configured hello time 2, max age 20, forward delay 15

Interface                                  Designated
 Name      PortID   Prio Cost   Sts        Cost      Bridge ID          PortID
---------- -------- ---- ------  ----------- ------- -------------------- --------
Po 100     128.101  128  1800    FWD       1800    32770 001d.a23e.0d00  128.560
Po 128     128.129  128  200000  DIS       1800    32768 0001.e8a4.fc7e  128.129

Interface
 Name      Role   PortID   Prio Cost    Sts        Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 100     Root   128.101  128  1800    FWD        1800    P2P       No
Po 128     Dis    128.129  128  200000  DIS        1800    P2P       No
FTOS(conf-pvst)#
```

```
FTOS(conf-pvst)#do show span pvst vlan 3 brief
VLAN 3
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 32768, Address 0001.e8a4.fc7e
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 32768, Address 0001.e8a4.fc7e
We are the root of Vlan 3
Configured hello time 2, max age 20, forward delay 15

Interface                                    Designated
 Name      PortID   Prio Cost    Sts        Cost       Bridge ID       PortID
---------- -------- ---- ------  ----------- ------- -------------------- --------
Po 100     128.101  128  1800    FWD       0      32768 0001.e8a4.fc7e  128.101
Po 128     128.129  128  200000  DIS       0      32768 0001.e8a4.fc7e  128.129

Interface
 Name      Role   PortID   Prio Cost    Sts       Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 100     Desg   128.101  128  1800    FWD       0      P2P      No
Po 128     Dis    128.129  128  200000  DIS       0      P2P      No
FTOS(conf-pvst)#

FTOS(conf-pvst)#do show span pvst vlan 4 br
VLAN 4
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 32768, Address 0001.e8a4.fc7e
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 32768, Address 0001.e8a4.fc7e
We are the root of Vlan 4
Configured hello time 2, max age 20, forward delay 15
Interface                                    Designated
 Name      PortID   Prio Cost    Sts        Cost       Bridge ID       PortID
---------- -------- ---- ------  ----------- ------- -------------------- --------
Po 100     128.101  128  1800    FWD       0      32768 0001.e8a4.fc7e  128.101
Po 128     128.129  128  200000  DIS       0      32768 0001.e8a4.fc7e  128.129

Interface
 Name      Role   PortID   Prio Cost    Sts       Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 100     Desg   128.101  128  1800    FWD       0      P2P      No
Po 128     Dis    128.129  128  200000  DIS       0      P2P      No
```

15 BGP on the upstream links from VLT Peers

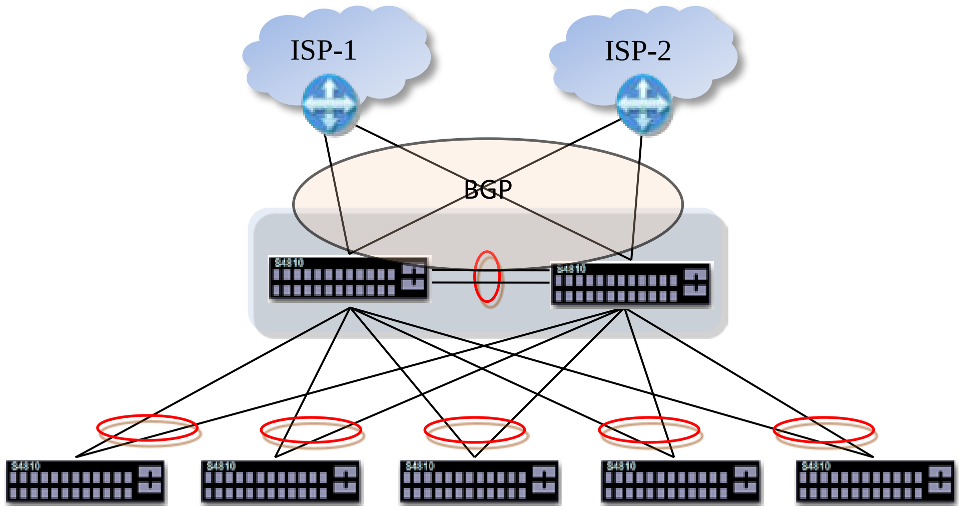

Figure 15.0 BGP on VLT uplinks

Depending on the number of BGP peers and the routes the VLT nodes can peer directly to the upstream ISP's, or through Core routers peering to ISP's, having IBGP to the VLT nodes. With BGP policies on the ingress route-maps, preference for specific NLRI can be implemented. If the Core Router needs to have complete Internet routes, then the route-maps with policies for more specific routes along with default-route could be advertised to the VLT peers. As a best practice, configure the same ASN in both the VLT peers. IBGP peering has to be established between VLT peers.

In case of hierarchical VLT domains, the Spine/Core layer has to be configured as Route-reflectors, if IBGP is configured across VLT domains with '*peer-routing*' enabled.

16 Scenarios with non-VLT ports

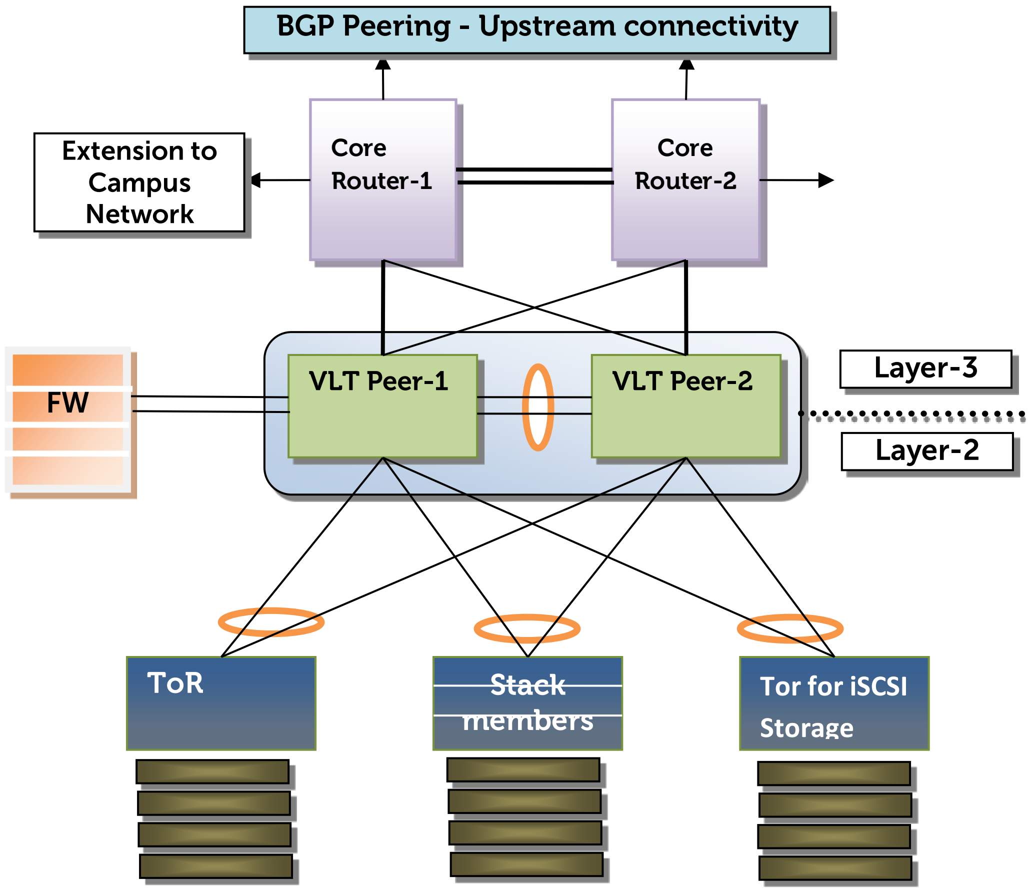

Figure 16.0 VLAN's with non-VLT port

Enterprise networks- Provisioning of Firewall and Load-balancers

Though many of the scenarios explained in the earlier VLT setup, firewall and load balancer connecting to non-VLT ports might be required in certain networks, where Policy based routing directs matching traffic to Firewall for deep packet inspection. For the hosts which are connected in the non-vlt ports, the traffic might pass through the VLTi links. Hence the bandwidth required has to be defined for such scenarios. With FTOS 9.2(0.0) release, the MAC and ARP tables for the orphan ports are synchronized between the VLT peers.

With geographically located at different regions, extending the network could be achieved in multiple ways. With Service providers offering the backbone, MPLS or BGP with core routers connecting to its peers can extend the VPN services. Alternately, if the dark fiber is laid across all these locations, layer-2 or layer-3 services can be extended as explained in the layer-2 domain extension section.

17 VRRP in VLT

In a VLT domain with MAC and ARP synchronized, the VRRP hello messages are sent to the other peer and offers active-active load balancing unlike the normal Master-Standby operation. Any of the upstream packets hitting the VRRP Standby node in the VLT domain need not send it across to the VRRP Master node. The received packet would be routed to its destination from the VLT peer as shown in the figure 17.0.

As a best practice configure identical VRRP configurations on both VLT peers in a domain. Track interface configuration has no significance in the VLT domain.

With FTOS 9.2(0.0), VRRP for IPv4 and IPv6 Dual stack is supported

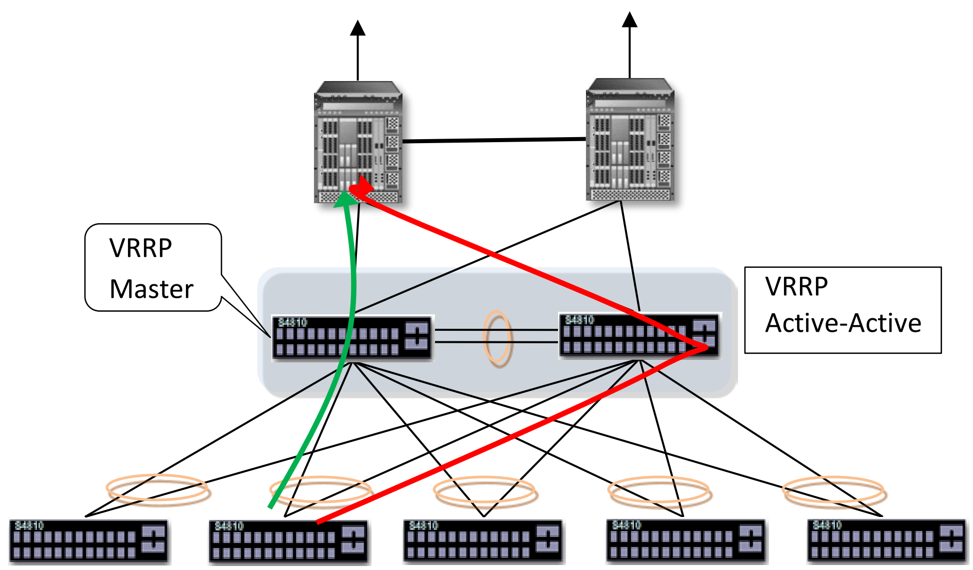

Figure 17.0 VRRP in VLT

VRRP Configuration

VLT Peer-1	**VLT Peer-2**
interface Vlan 200	interface Vlan 200
ip address 10.0.0.252/24	**ip address 10.0.0.253/24**
tagged Port-channel 1-2	tagged Port-channel 1-2
!	!
vrrp-group 1	**vrrp-group 1**
no preempt	no preempt
virtual-address 10.0.0.254	**virtual-address 10.0.0.254**
no shutdown	no shutdown
!	!
interface Vlan 201	interface Vlan 201
ip address 11.0.0.252/24	ip address 11.0.0.253/24
tagged Port-channel 1-2	tagged Port-channel 1-2
!	!
vrrp-group 1	vrrp-group 1
no preempt	no preempt
virtual-address 11.0.0.254	virtual-address 11.0.0.254
no shutdown	no shutdown

18 Multicast in VLT

VLT nodes have IGMP messages synchronized across its peer through the VLTi messages. Any of the IGMP Joins received at the VLT nodes are synchronized with its peer, so the IGMP tables are identical in the VLT domain.

VLT nodes learn the mrouter ports from the upstream IGMP Querier messages. If layer-3 VLANs are enabled with PIM-SM, then VLTi ports can be statically configured as mrouter ports. However, it is not required if the VLT domain has only IGMP-Snooping enabled.

Layer-3 VLANs in the VLT nodes can be enabled with PIM-Sparse mode which offers Active-Active PIM-SM support. Either of the VLT peer act as DR or BDR. However, as a best practice configure the non-VLT Core router to act as RP for the multicast group.

Any multicast source traffic arriving at the VLT peer would be sent to the receiver and also to the VLTi link. However with the port block at the VLT link of the peer node the multicast traffic is dropped and not duplicated (Figure 18.0). However, if the receiver is connected in a non-VLT port on that peer, the multicast traffic would reach the receiver through the VLTi.

VLTi link would be added as mrouter port if the IGMP Query is learnt through VLTi. If the multicast groups are more with multiple receivers and Source, then the VLTi link bandwidth has to be increased to handle the multicast flooding on the VLTi links.

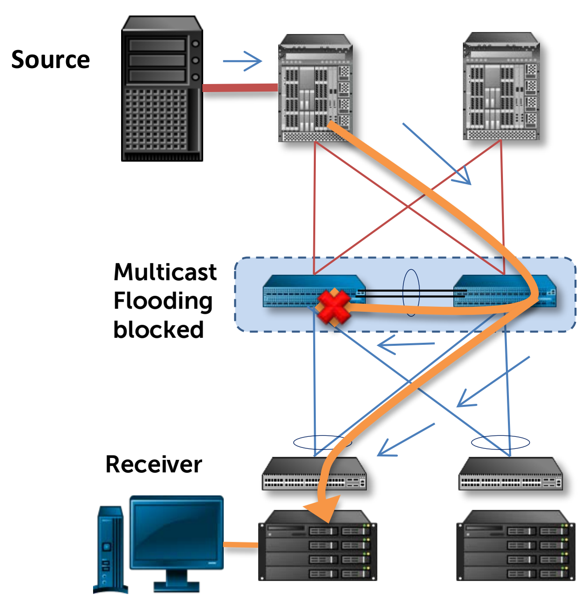

Figure 18.0 Multicast in VLT

With FTOS 9.2(0.0) release the following features are implemented for multicast

- PIM routes synchronization between VLT nodes
- Outgoing Interface (OIF) Sync for PIM routes
- PIM First hop (FHR) and Last hop (LHR)
- PIM-SM Intermediate Router support
- Show ip mroute on VLT context
- Multicast peer-routing timeout
- VLT mismatch for multicast

18.1 Multicast in hierarchical VLT topology

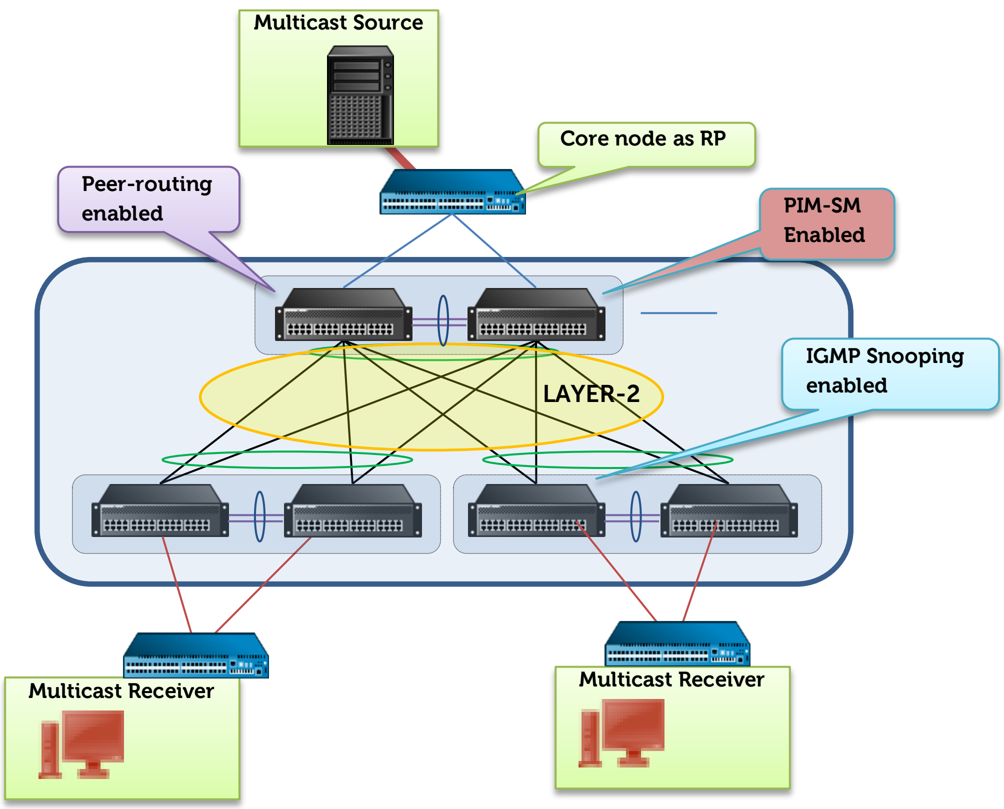

Figure 18.1 Multicast in hierarchical Layer-2 VLT domain

The Leaf VLT domains configured with layer-2 VLAN's and has IGMP-snooping enabled. With IGMP-Snooping on both the VLT nodes, the multicast forwarding states are synchronized. PIM-Sparse mode has to be enabled on the layer-3 VLANs participating in the multicast traffic for the Spine VLT domain.

The topology shown in the figure 18.1 meets the scalability needs for large multicast groups. The following points are to be ensured while deploying multicast service through VLT domain

- With VLT in Spine and leaf nodes, the RP (Rendezvous Point) for the multicast group shall be the non-VLT node preferably the core node in upstream.
- Currently VLT node cannot be configured as RP. As a workaround, RP shall be configured with single multicast enabled Layer-3 VLAN, connected to Spine VLT domains.
- Disable IGMP-snooping on the upstream links.

For various other VLT topologies with layer-3 between VLT domains, only a single multicast enabled Layer-3 VLAN has to be configured for the RP reachability. However other L3 VLANs not participating in multicast could be spanned between the Spine and Leaf VLT domains.

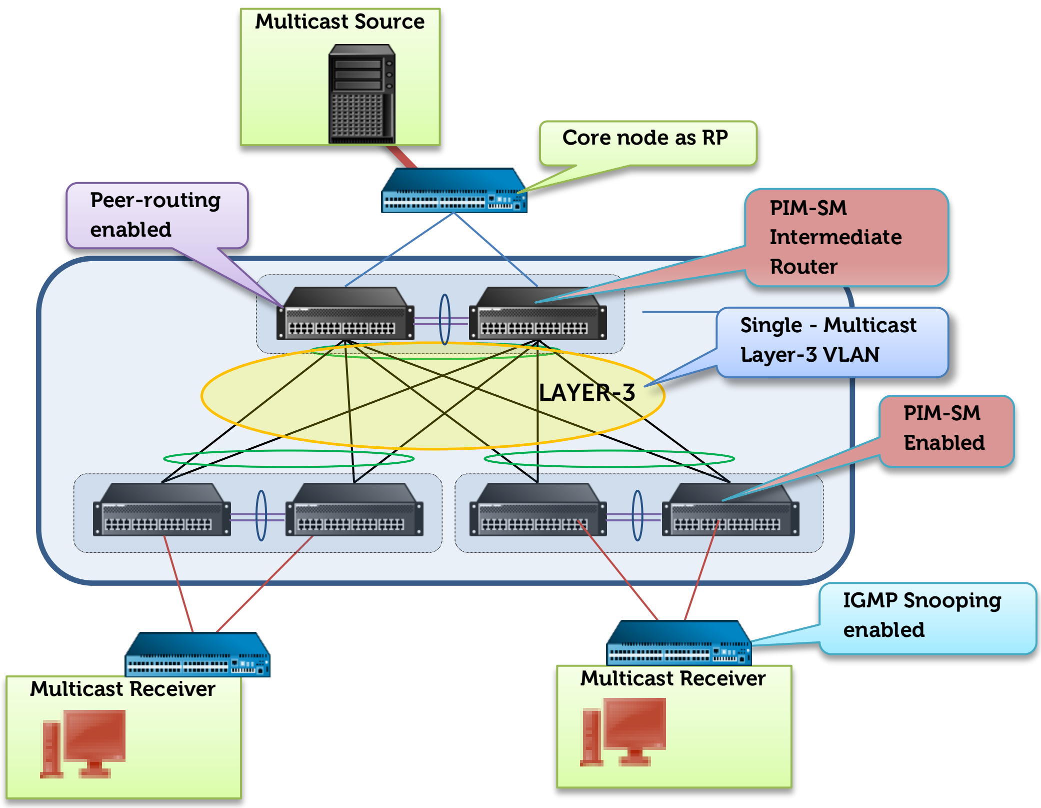

Figure 18.2 Multicast in hierarchical VLT with Layer-3 VLT domain

In a hierarchical VLT domain with Spine and Leaf VLT domains configured for peer-routing, optimal flow for multicast traffic is ensured by implementing a single multicast enabled layer-3 VLAN between the VLT domains for the reachability of the RP or Multicast Source. In this topology (Figure 18.2), the Spine VLT domain acts as Intermediate PIM-SM Router.

As a workaround for this limitation, the Leaf node VLT domains have to be configured with Layer-2 VLANs, enabled with IGMP-Snooping, and the Spine VLT domain configured with PIM-SM as shown in figure 18.1

Multicast peer-routing timeout

By default the multicast peer-routing timeout is configured for 150 seconds. In case of VLT peer node failure, the alternate node forwards the multicast traffic on behalf of its VLT peer for the configured timeout period in seconds. As a best practice, the duration shall be configured more than 105 seconds (matching PIM hold timer value).

VLT Mismatch for multicast

%STKUNIT0-M:CP %PIM-3-PIM_VLT_MRT_IIF_MISMATCH: VLT peers failed to choose the same interface as Incoming Interface for the Route (176.16.1.10, 225.0.0.2)

```
VLT-100-SPINE1-FTOS#show vlt mismatch

Global IP Multicast Status
  ---------------------------
     IP Version          Local Status      Peer Status
     ----------          -------------     ------------
     IPV4                 Disabled          Enabled
VLT-100-SPINE1-FTOS#
```

```
VLT-100-SPINE2-FTOS# show vlt inconsistency ip mroute

Spanned Multicast Routing IIF Inconsistency

 Multicast Route                LocalIIF      PeerIIF
 ---------------                ----------    ---------
 (0.0.0.0 , 225.0.0.0)             Vl 18      Vl 4001
VLT-100-SPINE2-FTOS#
```

```
VLT-100-SPINE1-FTOS#show vlt statistics multicast
 VLT Multicast Statistics
-------------------------------
 Info Pkts Sent:                  13
 Info Pkts Rcvd:                  6
 Reg Request Sent:                2
 Reg Request Rcvd:                4
 Reg Response Sent:               2
 Reg Response Rcvd:               2
 Route updates sent to Peer:      4
 Route updates rcvd from Peer:   10
 Route update pkts sent to Peer:  3
 Route update pkts rcvd from Peer: 7
```

```
VLT-100-SPINE1-FTOS#show ip mroute vlt count

IP Multicast Statistics
1 routes using 648 bytes of memory
```

```
1 groups, 1.00 average sources per group

Group: 225.0.0.0, Source count: 1
  Source: 90.114.112.11

VLT-100-SPINE1-FTOS#
```

Syslog for multicast

For any misconfigured static routes, if the Incoming interface (VLAN) for (*,G) or (S,G) route differs in both peers, the following syslog would be displayed.

```
Aug 6 09:56:00: %STKUNIT0-M:CP %PIM-3-PIM_VLT_MRT_IIF_MISMATCH: Route
(0.0.0.0, 225.0.0.0) - Detected IIF mismatch between VLT peers. Local
IIF is Vl 18, Remote IIF is Vl 4001

Aug 6 09:57:03: %STKUNIT0-M:CP %PIM-3-PIM_VLT_MRT_CLEAR_IIF_MISMATCH:
Route (0.0.0.0, 225.0.0.0) - Cleared IIF mismatch between VLT Peers.
IIF chosen is Vl 4001

Aug 6 10:01:27: %STKUNIT0-M:CP %PIM-3-PIM_VLT_MRT_IIF_MISMATCH: Route
(0.0.0.0, 225.0.0.0) - Detected IIF mismatch between VLT peers. Local
IIF is Vl 18, Remote IIF is Vl 4001
```

18.2 Multicast configuration and outputs from Leaf and Spine VLT nodes

```
VLT-10-LEAF-1#show vlt statistics igmp-snoop
VLT Igmp-Snooping Statistics
-------------------------------
IGMP Info Pkts sent:     5
IGMP Info Pkts Rcvd:     7
IGMP Reg Request sent:   2
IGMP Reg Request rcvd:   1
IGMP Reg Response sent:  1
IGMP Reg Response rcvd:  1
IGMP PDU Tunnel Pkt sent:83
IGMP PDU Tunnel Pkt rcvd:58
IGMP Tunnel PDUs sent:   257
IGMP Tunnel PDUs rcvd:   320

VLT-10-LEAF-1#show running-config igmp
!
ip igmp snooping enable

VLT-10-LEAF-1#show vlt statistics igmp-snoop
VLT Igmp-Snooping Statistics
-------------------------------
IGMP Info Pkts sent:     5
IGMP Info Pkts Rcvd:     7
IGMP Reg Request sent:   2
IGMP Reg Request rcvd:   1
IGMP Reg Response sent:  1
IGMP Reg Response rcvd:  1
IGMP PDU Tunnel Pkt sent:83
IGMP PDU Tunnel Pkt rcvd:58
IGMP Tunnel PDUs sent:   257
IGMP Tunnel PDUs rcvd:   320

VLT-10-LEAF-1#show ip igmp groups
Total Number of Groups: 5
IGMP Connected Group Membership
Group
Address     Interface                   Mode          Uptime    Expires
Last Reporter
225.1.1.1         Vlan
10                  IGMPv2        00:05:59  00:01:21  192.168.20.20
  Member Ports: Po 11
225.1.1.2         Vlan
10                  IGMPv2        00:05:59  00:01:21  192.168.20.20
  Member Ports: Po 11
225.1.1.3         Vlan
10                  IGMPv2        00:05:59  00:01:21  192.168.20.20
  Member Ports: Po 11
225.1.1.4         Vlan
10                  IGMPv2        00:05:59  00:01:21  192.168.20.20
```

```
  Member Ports: Po 11
225.1.1.5        Vlan
10                  IGMPv2        00:05:59  00:01:21  192.168.20.20
  Member Ports: Po 11

VLT-10-LEAF-2#show running-config igmp
!
ip igmp snooping enable

VLT-10-LEAF-2#show vlt statistics igmp-snoop
VLT Igmp-Snooping Statistics
-------------------------------
IGMP Info Pkts sent:     7
IGMP Info Pkts Rcvd:     5
IGMP Reg Request sent:   1
IGMP Reg Request rcvd:   2
IGMP Reg Response sent:  1
IGMP Reg Response rcvd:  1
IGMP PDU Tunnel Pkt sent:58
IGMP PDU Tunnel Pkt rcvd:87
IGMP Tunnel PDUs sent:   320
IGMP Tunnel PDUs rcvd:   264

VLT-10-LEAF-2#show ip igmp groups
Total Number of Groups: 5
IGMP Connected Group Membership
Group
Address     Interface                    Mode          Uptime    Expires
Last Reporter
225.1.1.1        Vlan
10                  IGMPv2        00:07:09  00:02:06  192.168.20.20
  Member Ports: Po 11
225.1.1.2        Vlan
10                  IGMPv2        00:07:09  00:02:06  192.168.20.20
  Member Ports: Po 11
225.1.1.3        Vlan
10                  IGMPv2        00:07:09  00:02:06  192.168.20.20
  Member Ports: Po 11
225.1.1.4        Vlan
10                  IGMPv2        00:07:09  00:02:06  192.168.20.20
  Member Ports: Po 11
225.1.1.5        Vlan
10                  IGMPv2        00:07:09  00:02:06  192.168.20.20
  Member Ports: Po 11
```

```
VLT-100-SPINE-1#show running-config pim
!
ip pim rp-address 90.114.112.1 group-address 224.0.0.0/4

VLT-100-SPINE-1(conf-if-range-vl-10,vl-20,vl-100)#show config
!
interface Vlan 10
ip address 192.168.0.3/16
tagged Port-channel 33
ip pim sparse-mode
ip ospf priority 100
no shutdown
!
interface Vlan 20
ip address 192.169.0.3/16
tagged Port-channel 44
ip pim sparse-mode
ip ospf priority 100
no shutdown
!
interface Vlan 100
ip address 100.100.100.1/24
untagged Port-channel 55
ip pim sparse-mode
ip ospf priority 100
no shutdown

VLT-100-SPINE-1#show running-config | grep multi
ip multicast-routing

VLT-100-SPINE-1#show ip pim  summary
PIM TIB version 110
Uptime 01:37:07
Entries in PIM-TIB/MFC : 5/0

Active Modes :
        PIM-SM

Interface summary:
        3 active PIM interfaces
        0 passive PIM interfaces
        4 active PIM neighbors

TIB summary:
        5/0 (*,G) entries in PIM-TIB/MFC
        0/0 (S,G) entries in PIM-TIB/MFC
        0/0 (S,G,Rpt) entries in PIM-TIB/MFC

        2 PIM nexthops
        1 RPs
        0 sources
```

```
        0 Register states

Message summary:
        20/885 Joins sent/received
        20/110 Prunes sent/received
        0/0 Candidate-RP advertisements sent/received
        0/0 BSR messages sent/received
        0/0 State-Refresh messages sent/received
        0/0 MSDP updates sent/received
        0/0 Null Register messages sent/received
        0/0 Register-stop messages sent/received

Data path event summary:
        10 no-cache messages received
        24 last-hop switchover messages received
        0/0 pim-assert messages sent/received
        0/0 register messages sent/received

Memory usage:
        TIB             : 2440 bytes
        Nexthop cache   : 232 bytes
        Interface table : 5000 bytes
        Neighbor table  : 672 bytes
        RP Mapping      : 180 bytes


VLT-100-SPINE-1#show ip pim  tib

PIM Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned,
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join
SPT,
       M - MSDP created entry, A - Candidate for MSDP Advertisement
       K - Ack-Pending State
Timers: Uptime/Expires
Interface state: Interface, next-Hop, State/Mode

(*, 225.1.1.1), uptime 00:09:38, expires 00:00:00, RP 90.114.112.1,
flags: SCPJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:

(*, 225.1.1.2), uptime 00:09:38, expires 00:00:00, RP 90.114.112.1,
flags: SCPJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:

(*, 225.1.1.3), uptime 00:09:38, expires 00:00:00, RP 90.114.112.1,
flags: SCPJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
```

```
(*, 225.1.1.4), uptime 00:09:38, expires 00:00:00, RP 90.114.112.1,
flags: SCPJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:

(*, 225.1.1.5), uptime 00:09:38, expires 00:00:00, RP 90.114.112.1,
flags: SCPJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:

VLT-100-SPINE-1#show ip mroute vlt

IP Multicast Routing Table
Flags: S - Synced

(*, 225.1.1.1), uptime 00:09:51  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(90.114.112.2, 225.1.1.1), uptime 00:09:06  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(*, 225.1.1.2), uptime 00:09:51  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(90.114.112.2, 225.1.1.2), uptime 00:09:06  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(*, 225.1.1.3), uptime 00:09:53  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(90.114.112.2, 225.1.1.3), uptime 00:09:08  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)
```

```
(*, 225.1.1.4), uptime 00:09:53  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(90.114.112.2, 225.1.1.4), uptime 00:09:08  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(*, 225.1.1.5), uptime 00:09:55  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

(90.114.112.2, 225.1.1.5), uptime 00:09:09  flags: S
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10 (S)
    Vlan 20 (S)

VLT-100-SPINE-1#show ip pim  neighbor
Neighbor                  Interface      Uptime/Expires     Ver  DR
Prio/Mode GR
Address
192.168.0.4               Vl 10          00:51:00/00:01:21  v2   1  /
DR S
192.169.0.4               Vl 20          00:14:06/00:01:40  v2   1  /
DR S
100.100.100.2             Vl 100         01:34:44/00:01:42  v2   1  /
S
100.100.100.3             Vl 100         01:35:00/00:01:25  v2   1  /
DR S

VLT-100-SPINE-2#show ip pim  neighbor
Neighbor                  Interface      Uptime/Expires     Ver  DR
Prio/Mode GR
Address
192.168.0.3               Vl 10          00:51:37/00:01:25  v2   1  /
S
192.169.0.3               Vl 20          00:14:44/00:01:34  v2   1  /
S
100.100.100.1             Vl 100         01:35:21/00:01:18  v2   1  /
S
100.100.100.3             Vl 100         01:35:22/00:01:18  v2   1  /
DR S
```

```
VLT-100-SPINE-2#show ip mroute vlt

IP Multicast Routing Table
Flags: S - Synced

(*, 225.1.1.1), uptime 00:12:27
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(90.114.112.2, 225.1.1.1), uptime 00:11:41
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(*, 225.1.1.2), uptime 00:12:27
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(90.114.112.2, 225.1.1.2), uptime 00:11:41
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(*, 225.1.1.3), uptime 00:12:28
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(90.114.112.2, 225.1.1.3), uptime 00:11:42
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(*, 225.1.1.4), uptime 00:12:28
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(90.114.112.2, 225.1.1.4), uptime 00:11:43
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
```

```
    Vlan 20

(*, 225.1.1.5), uptime 00:12:29
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

(90.114.112.2, 225.1.1.5), uptime 00:11:44
  Incoming interface: Vlan 100
  Spanned outgoing interface list:
    Vlan 10
    Vlan 20

VLT-100-SPINE-2#  show ip pim tib

PIM Multicast Routing Table
Flags: D - Dense, S - Sparse, C - Connected, L - Local, P - Pruned,
       R - RP-bit set, F - Register flag, T - SPT-bit set, J - Join
SPT,
       M - MSDP created entry, A - Candidate for MSDP Advertisement
       K - Ack-Pending State
Timers: Uptime/Expires
Interface state: Interface, next-Hop, State/Mode

(*, 225.1.1.1), uptime 00:12:42, expires 00:00:00, RP 90.114.112.1,
flags: SCJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(90.114.112.2, 225.1.1.1), uptime 00:11:56, expires 00:01:56, flags:
CT
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(*, 225.1.1.2), uptime 00:12:42, expires 00:00:00, RP 90.114.112.1,
flags: SCJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(90.114.112.2, 225.1.1.2), uptime 00:11:56, expires 00:01:56, flags:
CT
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
```

```
    Vlan 20  Forward/Sparse   00:12:38/Never

(*, 225.1.1.3), uptime 00:12:42, expires 00:00:00, RP 90.114.112.1,
flags: SCJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(90.114.112.2, 225.1.1.3), uptime 00:11:56, expires 00:01:56, flags:
CT
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(*, 225.1.1.4), uptime 00:12:42, expires 00:00:00, RP 90.114.112.1,
flags: SCJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(90.114.112.2, 225.1.1.4), uptime 00:11:56, expires 00:01:56, flags:
CT
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(*, 225.1.1.5), uptime 00:12:42, expires 00:00:00, RP 90.114.112.1,
flags: SCJ
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never

(90.114.112.2, 225.1.1.5), uptime 00:11:56, expires 00:01:56, flags:
CT
  Incoming interface: Vlan 100, RPF neighbor 100.100.100.3
  Outgoing interface list:
    Vlan 10  Forward/Sparse   00:12:42/Never
    Vlan 20  Forward/Sparse   00:12:38/Never
```

19 S5000 Switches in VLT

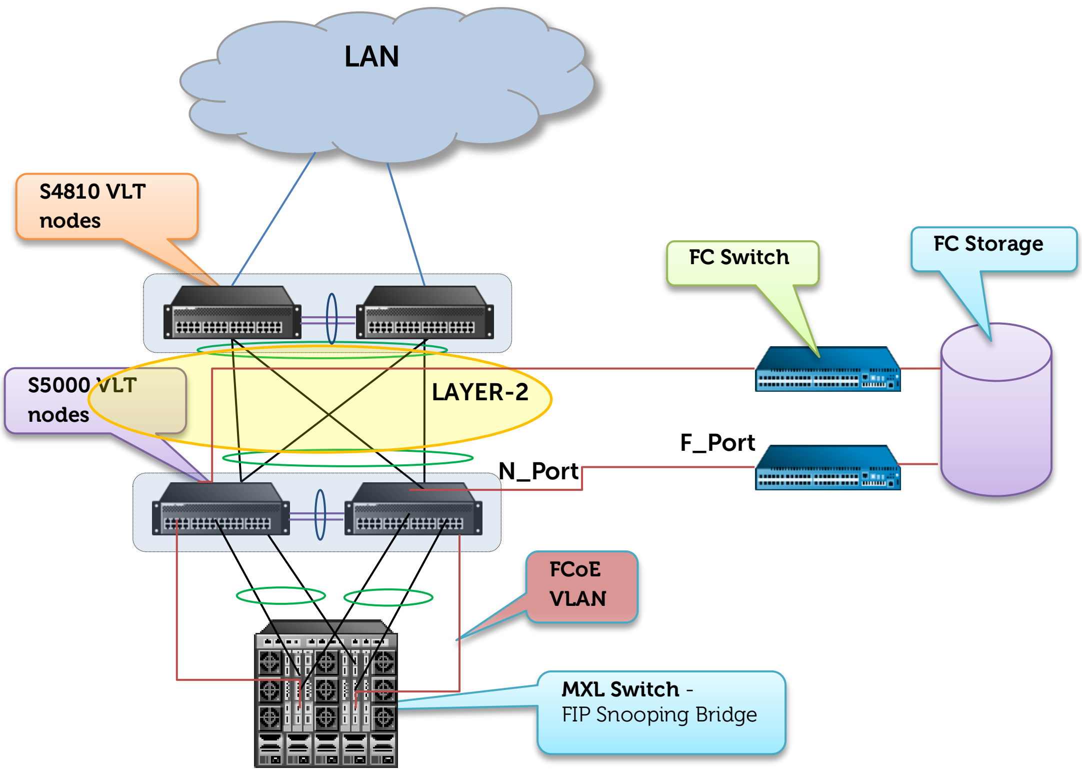

Figure 19.0 S5000 Switches in VLT for Converged Infrastructure

Currently with S5000 Switches in a VLT domain, FCoE is not supported over VLT. However, with unique FCoE VLAN mapped to specific ports, FCoE traffic is separately taken from MXL Switchport to S5000 VLT nodes. MXL act as FSB (FIP Snooping Bridge). The N_Port from the S5000 Switches are connected directly to FC Switch. In this topology the VLT redundancy benefits for the LAN traffic only.

For further deployment models with S5000 in a converged infrastructure please go through the following link

Dell Networking S5000: The Building Blocks of Unified Fabric and LAN/SAN Convergence

Dell Networking S5000: Deployment of a Converged Infrastructure with FCoE

However, the S5000 Switches with modular ports could benefit the Small/Medium Enterprise or Campus Networks which could populate with Ethernet ports and offer VLT features for LAN traffic.

20 Data Centre interconnect thro' VLTi with DWDM optics

Extending Layer 2 Data centers through DWDM rings

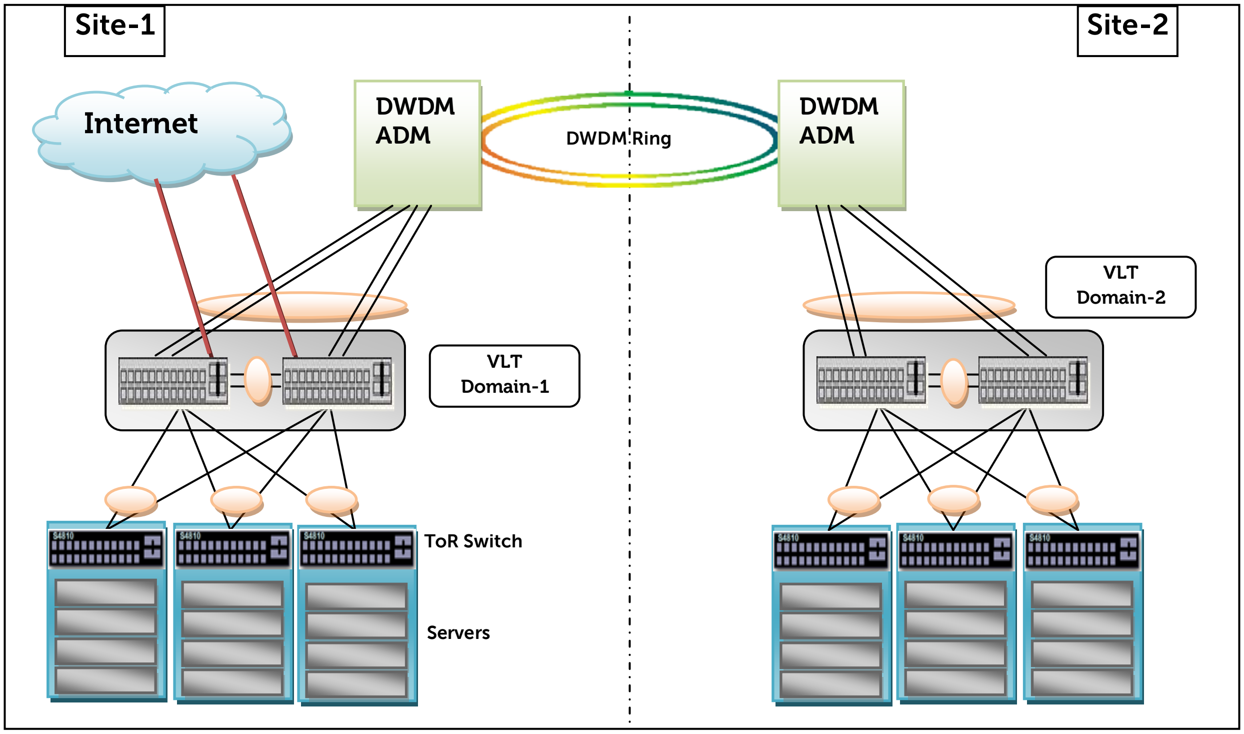

Figure 19.0 DCI with VLT

This topology is to extend the layer-2 domain over different regions. For an enterprise network having offices collocated at geographically different regions, extending the layer-2 domain of their network could be achieved through the DWDM connectivity depending on their extent.

The VLT domain-1 in the Hub site (Site-1) connects to VLT domain-2 in Remote Site (Site-2). The VLT Port-Channels between the multi VLT domains could be configured with long-timeout for LACP. The VLT links pass through the DWDM Add-drop multiplexer. The Non-VLT ports connect to the upstream connectivity.

VLAN's are tagged on the VLT links connecting to DWDM links at both locations. Depending upon the VLAN's configured on the respective ToR switch, the VLT links connecting to the ToR could be tagged or untagged. If multiple VLAN's are deployed on the same ToR switch, then tagging is necessary.

Virtual servers deployed in Site-1 could be live-migrated to Site-2 with this topology. Application agility with varying workloads need for VM live motion across different sites could be deployed with this architecture.

20.1 DWDM Interconnect multiple VLT domains aggregated with hierarchical VLT

Aggregated VLT domains extending layer-2 Circuits through DWDM/CWDM optics

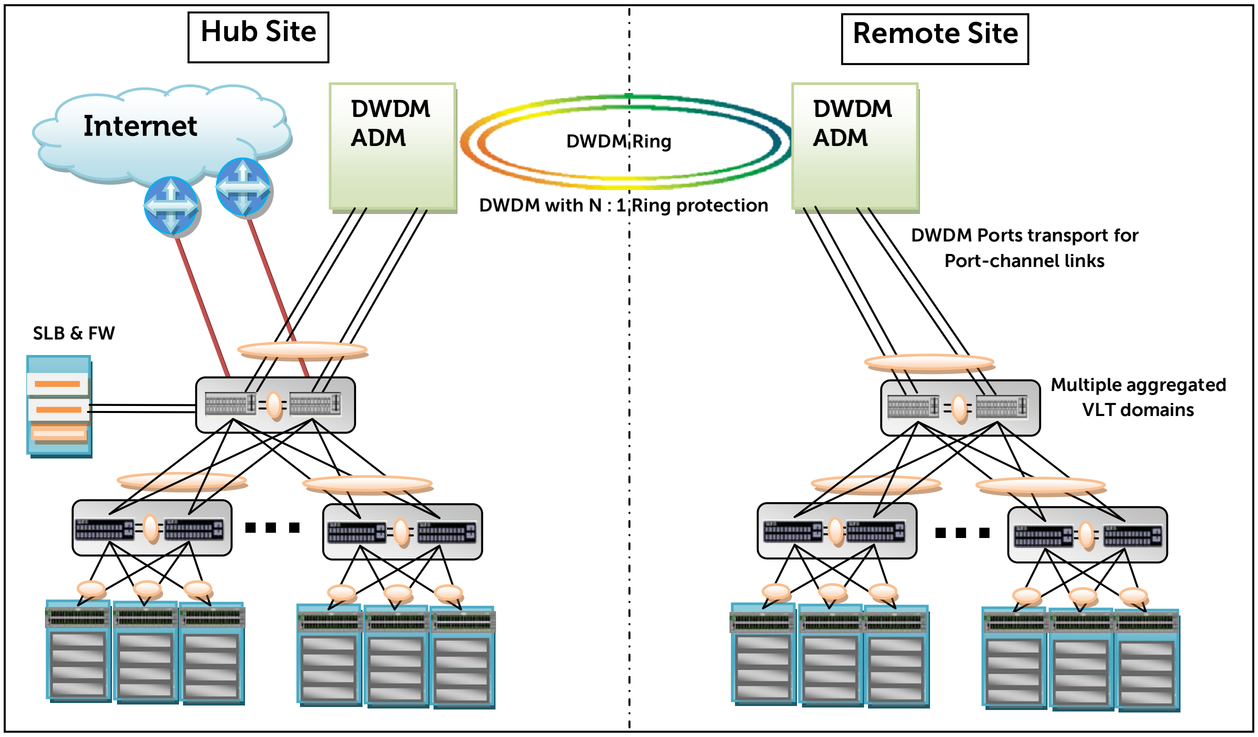

Figure 19.1 Hierarchical VLT interconnect with DWDM/CWDM

This topology depicts extending multi VLT thro DWDM Optical transport. Also the extended mVLT's could have their "local" mVLT's as well, extending the layer-2 links within the data center. Based on the over subscription rate and the server density, the mVLT could be extended within the limitation. The above diagram depicts the hierarchical setup of VLT domains. Multiple Remote sites connect to the central Hub site. With DWDM add drop multiplexer, tuned lambda carries the remote-site traffic to the respective hub sites.

CAPEX for this network architecture is less comparing to other contemporary layer-2 extension feature (Fabric path, l2vpn, VPLS etc.,) High availability and resiliency across all layers are built in this topology, covering from N:1 protection from the DWDM ring to the optical ports linking to the VLT port-channels.

Albeit the North bound traffic is located at Central site, depending on the unique customers requirement, another core connectivity could be made at Remote site as well.

Alternately, if the distance is within 10KM, then 40G –LR4 CWDM optics could be used to directly connect 40G QSFP+ ports interconnecting VLT domains without any DWDM mux.

For the appropriate optics, please refer
http://salesedge.dell.com/doc?id=0901bc82806791dc&ll=sr

21 VLT Scalability

VLT Switch port capacity

S4810 Switch
48x 10G ports and 4 x40G ports
For any of the VLT setup, with ***non-blocking*** capacity
30 x10G for downlink (VLT) and 30x10G for upstream
4x10G for VLTi links
This non-blocking is for the VLT peers only.

Z9000 Switch
32x 40G ports
For any of the VLT setup, with ***non-blocking*** capacity
15x 40G for downlink (VLT) and 15x40G for upstream
2x 40G for VLTi links
This non-blocking is for the VLT peers only, however with multi VLT, the Over-subscription ratio increments with the other VLT domain as well.

For the ToR switches connecting to S4810 VLT peers with port-channel of 16 members (8 in each VLT peer) only 7 ToR nodes could be connected for each VLT domain. This scenario suits well for predominant north-south traffic in the network. This could be changed if the user requires over subscription at different levels/layers. The ToR could be S55/S60 switches and the number of port-channel members could be less than 8 - 6/4/2 members. If the significance is for East-West traffic in the network, then over-subscription in the VLT links could be implemented.

21.1 Use cases

1. In the data center environment where multiple VM's hosted in Single Server scaling beyond this range, meticulous planning has to be made for positioning servers under each VLT domain with the scalability limits in mind. Not many ToR nodes aggregate to the same VLT domain. In the scalable data center network, multiple VLT domains connect to group of ToR nodes spanning multiple VLAN's, while the upstream of VLT domains connect to the Core router.

2. For enterprise networks with server-to-server and servers to upstream traffic present in the networks, over-subscription could be the ideal solution. With predominant east-west traffic pattern, multiple VLT domains to multiple ToR nodes could match server clusters to storage, with users in various VLAN groups.

3. For HPC networks, depending on the baseline, the over-subscription could be derived. The major advantage over the existing topology would be of eliminating STP in the access links. Faster convergence and reduced latency offer maximum advantage to the High performance computing networks.

22 Example Configuration

VLT Peer-1

```
F10S4810AGGR1#show run vlt
!
vlt domain 10
 peer-link port-channel 90
 back-up destination 10.16.151.102
 primary-priority 1
 system-mac mac-address
00:01:e8:8b:11:11
 unit-id 0
 delay-restore 10
 peer-routing
--------------------------------------
--------
F10S4810AGGR1#show vlt brief
 VLT Domain Brief
------------------
 Domain ID:                    10
 Role:                         Primary
 Role Priority:                1
 ICL Link Status:              Up
 HeartBeat Status:             Up
 VLT Peer Status:              Up
 Local Unit Id:                0
 Version:                      6(1)
 Local System MAC address:
00:01:e8:8b:1d:3f
 Remote System MAC address:
00:01:e8:8b:1c:9a
 Configured System MAC
address:00:01:e8:8b:11:11
 Remote system version:        6(1)
 Delay-Restore timer:          10
seconds
Peer-Routing :
Enabled
Peer-Routing-timeout timer:     0
seconds
Multicast peer-routing timeout: 150
seconds

F10S4810AGGR1#
--------------------------------------
--------
F10S4810AGGR1#show vlt detail
Local LAG Id  Peer LAG Id  Local
Status  Peer Status  Active VLANs
------------  -----------  -----------
-  -----------  -------------
30            30           DOWN
DOWN          10, 20, 30, 60, 203
40            40           UP
UP            10, 110, 120, 201-300
50            50           DOWN
```

VLT Peer-2

```
F10S4810AGGR2#show run vlt
!
vlt domain 10
 peer-link port-channel 90
 back-up destination 10.16.151.101
 primary-priority 65535
 system-mac mac-address
00:01:e8:8b:11:11
 unit-id 1
 delay-restore 10
 peer-routing
--------------------------------------
--------
F10S4810AGGR2#show vlt brief
 VLT Domain Brief
------------------
 Domain ID:                    10
 Role:
Secondary
 Role Priority:                65535
 ICL Link Status:              Up
 HeartBeat Status:             Up
 VLT Peer Status:              Up
 Local Unit Id:                1
 Version:                      6(1)
 Local System MAC address:
00:01:e8:8b:1c:9a
 Remote System MAC address:
00:01:e8:8b:1d:3f
 Configured System MAC
address:00:01:e8:8b:11:11
 Remote system version:        6(1)
 Delay-Restore timer:          10
seconds
Peer-Routing :
Enabled
Peer-Routing-timeout timer:     0
seconds
Multicast peer-routing timeout: 150
seconds

F10S4810AGGR2#
--------------------------------------
--------
F10S4810AGGR2#show vlt detail
Local LAG Id  Peer LAG Id  Local
Status  Peer Status  Active VLANs
------------  -----------  -----------
-  -----------  -------------
30            30           DOWN
DOWN          10, 20, 30, 60
40            40           UP
UP            110, 120, 201-300
```

DOWN 10, 110, 120, 201-300 60 60 DOWN DOWN 80 80 80 DOWN DOWN 90 --- F10S4810AGGR1#**show vlt role** VLT Role ---------- VLT Role: Primary System MAC address: 00:01:e8:8b:11:11 Primary Role Priority: 1 Local System MAC address: 00:01:e8:8b:1d:3f Local System Role Priority: 1 Local Unit Id: 0 --- F10S4810AGGR1#**show run int po 40** ! interface Port-channel 40 description F10S4810ToR1 no ip address switchport vlt-peer-lag port-channel 40 no shutdown --- F10S4810AGGR1#**show run int po 90** ! interface Port-channel 90 description VLT_CHANNEL no ip address no shutdown F10S4810AGGR1#	50 50 DOWN DOWN 10, 110, 120, 201-300 60 60 DOWN DOWN 80 80 80 DOWN DOWN 90 --- F10S4810AGGR2#**show vlt role** VLT Role ---------- VLT Role: Secondary System MAC address: 00:01:e8:8b:11:11 Primary Role Priority: 1 Local System MAC address: 00:01:e8:8b:1c:9a Local System Role Priority: 65535 Local Unit Id: 1 --- F10S4810AGGR2#**show vlt statistics** VLT Domain Statistics HeartBeat Messages Sent: 452309 HeartBeat Messages Received: 452306 ICL Hello's Sent: 150769 ICL Hello's Received: 150768 Domain Mismatch Errors: 0 Version Mismatch Errors: 0 Config Mismatch Errors: 19 VLT MAC Statistics ---------------- L2 Info Pkts sent:36, L2 Mac-sync Pkts Sent:325 L2 Info Pkts Rcvd:35, L2 Mac-sync Pkts Rcvd:461 L2 Reg Request sent:1 L2 Reg Request rcvd:2 L2 Reg Response sent:1 L2 Reg Response rcvd:1 VLT Igmp-Snooping Not Enabled VLT ARP Statistics ---------------- ARP Tunnel Pkts sent:1 ARP Tunnel Pkts Rcvd:5 ARP-sync Pkts Sent:4 ARP-sync Pkts Rcvd:7 ARP Reg Request sent:2 ARP Reg Request rcvd:2 F10S4810AGGR2#

```
POD2-AGG1#show running-config | n
Current Configuration ...
! Version 1-0(0-298)
! Last configuration change at Fri May 10 13:16:28 2013 by default
! Startup-config last updated at Mon May 13 13:43:13 2013 by default
!
boot system stack-unit 0 primary system: B:
boot system stack-unit 0 secondary system: B:
boot system stack-unit 0 default system: A:
boot system gateway 10.11.200.254
!
redundancy auto-synchronize full
redundancy disable-auto-reboot stack-unit
!
redundancy disable-auto-reboot stack-unit 0
redundancy disable-auto-reboot stack-unit 1
redundancy disable-auto-reboot stack-unit 2
redundancy disable-auto-reboot stack-unit 3
redundancy disable-auto-reboot stack-unit 4
redundancy disable-auto-reboot stack-unit 5
redundancy disable-auto-reboot stack-unit 6
redundancy disable-auto-reboot stack-unit 7
!
hardware watchdog
!
logging console errors
logging coredump stack-unit all
!
hostname POD2-AGG1
!
enable password 7 b125455cf679b208e79b910e85789edf
!
username admin password 7 1d28e9f33f99cf5c privilege 15
!
default vlan-id 4094
!
vlt domain 22
 peer-link port-channel 127
 back-up destination 13.1.1.15
 primary-priority 1
 system-mac mac-address 00:00:22:22:22:22
 unit-id 0
 peer-routing
 peer-routing-timeout 300
!
stack-unit 0 provision Z9000
!
interface fortyGigE 0/0
 description CORE1
 no ip address
!
 port-channel-protocol LACP
  port-channel 102 mode active
 no shutdown
!
interface fortyGigE 0/4
 no ip address
 shutdown
!
interface fortyGigE 0/8
 no ip address
 shutdown
!
```

```
interface fortyGigE 0/12
 no ip address
 shutdown
!
interface fortyGigE 0/16
 no ip address
 shutdown
!
interface fortyGigE 0/20
 no ip address
 shutdown
!
interface fortyGigE 0/24
 no ip address
 shutdown
!
interface fortyGigE 0/28
 no ip address
 shutdown
!
interface fortyGigE 0/32
 description CORE2
 no ip address
!
 port-channel-protocol LACP
  port-channel 102 mode active
 no shutdown
!
interface fortyGigE 0/36
 no ip address
 shutdown
!
interface fortyGigE 0/40
 no ip address
 shutdown
!
interface fortyGigE 0/44
 no ip address
 shutdown
!
interface fortyGigE 0/48
 no ip address
 shutdown
!
interface fortyGigE 0/52
 no ip address
 shutdown
!
interface fortyGigE 0/56
 no ip address
 shutdown
!
interface fortyGigE 0/60
 no ip address
 shutdown
!
interface fortyGigE 0/64
 description TOR1
 no ip address
!
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
```

```
!
interface fortyGigE 0/68
 description TOR2
 no ip address
!
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
!
interface fortyGigE 0/72
 no ip address
 no shutdown
!
interface fortyGigE 0/76
 no ip address
 no shutdown
!
interface fortyGigE 0/80
 no ip address
 shutdown
!
interface fortyGigE 0/84
 no ip address
 shutdown
!
interface fortyGigE 0/88
 no ip address
 shutdown
!
interface fortyGigE 0/92
 no ip address
 shutdown
!
interface fortyGigE 0/96
 no ip address
 shutdown
!
interface fortyGigE 0/100
 no ip address
 shutdown
!
interface fortyGigE 0/104
 no ip address
 shutdown
!
interface fortyGigE 0/108
 no ip address
 shutdown
!
interface fortyGigE 0/112
 no ip address
 shutdown
!
interface fortyGigE 0/116
 no ip address
 shutdown
!
interface fortyGigE 0/120
 description VLTi
 no ip address
 no shutdown
!
interface fortyGigE 0/124
```

```
 description VLTi
 no ip address
 no shutdown
!
interface ManagementEthernet 0/0
 ip address 13.1.1.14/24
 no shutdown
!
interface ManagementEthernet 1/0
 no shutdown
!
interface ManagementEthernet 2/0
 no shutdown
!
interface ManagementEthernet 3/0
 no shutdown
!
interface ManagementEthernet 4/0
 no shutdown
!
interface ManagementEthernet 5/0
 no shutdown
!
interface ManagementEthernet 6/0
 no shutdown
!
interface ManagementEthernet 7/0
 no shutdown
!
interface Loopback 0
 ip address 11.1.1.5/32
 no shutdown
!
interface Port-channel 1
 description TOR
 no ip address
 switchport
 rate-interval 30
 vlt-peer-lag port-channel 1
 no shutdown
!
interface Port-channel 102
 no ip address
 switchport
 rate-interval 30
 vlt-peer-lag port-channel 102
 no shutdown
!
interface Port-channel 127
 description VLTi
 no ip address
 channel-member fortyGigE 0/120,124
 rate-interval 30
 no shutdown
!
interface Vlan 1
 ip address 10.2.1.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.1.254
 no shutdown
```

```
!
interface Vlan 2
 ip address 10.2.2.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.2.254
 no shutdown
!
interface Vlan 3
 ip address 10.2.3.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.3.254
 no shutdown
!
interface Vlan 4
 ip address 10.2.4.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.4.254
 no shutdown
!
interface Vlan 5
 ip address 10.2.5.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.5.254
 no shutdown
!
interface Vlan 6
 ip address 10.2.6.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.6.254
 no shutdown
!
interface Vlan 7
 ip address 10.2.7.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.7.254
 no shutdown
!
interface Vlan 8
 ip address 10.2.8.1/24
 tagged Port-channel 1
!
 vrrp-group 1
  priority 110
  virtual-address 10.2.8.254
 no shutdown
```

```
!
interface Vlan 100
 ip address 10.100.1.5/16
 tagged Port-channel 1,102
!
 vrrp-group 1
  authentication-type simple 7 00d8d7903b20fd98
  priority 110
  virtual-address 10.100.100.100
 no shutdown
!
interface Vlan 2001
 ip address 10.20.1.3/24
 tagged Port-channel 102
 no shutdown
!
interface Vlan 4094
!untagged Port-channel 127
!
router ospf 1
 log-adjacency-changes
 graceful-restart mode planned-only
 network 10.20.0.0/16 area 0
 network 10.2.0.0/16 area 0
 auto-cost reference-bandwidth 40000
 maximum-paths 64
 default-information originate
 fast-converge 1
!
management route 0.0.0.0/0 13.1.1.1
!
logging 10.11.140.111
!
snmp-server community private rw
snmp-server community public ro
snmp-server enable traps bgp
snmp-server enable traps snmp authentication coldstart linkdown linkup
snmp-server enable traps vrrp
snmp-server enable traps lacp
snmp-server enable traps stp
snmp-server enable traps ecfm
snmp-server enable traps vlt
snmp-server enable traps xstp
snmp-server enable traps isis
snmp-server enable traps config
snmp-server enable traps envmon cam-utilization fan supply temperature
snmp-server enable traps eoam
snmp-server enable traps ecmp
snmp-server host 10.11.140.111 traps version 2c public udp-port 162
!
ntp server 10.11.140.111
!
clock timezone UTC 00 00
!
protocol lldp
 advertise management-tlv system-name
 no disable
!
script username admin /f10/flash/scripts/dfmsetup.sh
script username admin /usr/pkg/home/dfmcron.py
!
line console 0
 exec-timeout 0 0
```

```
line vty 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
line vty 7
line vty 8
line vty 9
!
 ecmp-group 1
  interface fortyGigE 0/0
  interface fortyGigE 0/4
  interface fortyGigE 0/8
  interface fortyGigE 0/12
  interface fortyGigE 0/16
  interface fortyGigE 0/20
  interface fortyGigE 0/24
  interface fortyGigE 0/28
  interface fortyGigE 0/32
  interface fortyGigE 0/36
  interface fortyGigE 0/40
  interface fortyGigE 0/44
  interface fortyGigE 0/48
  interface fortyGigE 0/52
  interface fortyGigE 0/56
  interface fortyGigE 0/60
  link-bundle-monitor enable
!
reload-type normal-reload
!
end
POD2-AGG1#
```

23 VLT Troubleshooting

There are specific "*show vlt*" commands, which display the detailed output of VLT parameters and its health status. The following CLI commands are useful to identify the basic VLT issues.

```
FTOS#show vlt?
backup-link      Backup-link information
brief            Brief information on the VLT
counter          Counter information on the VLT
detail           Detail information on the VLT
inconsistency    Inconsistent routes
mismatch         Vlt config mismatch information on the VLT
role             Role of VLT peer
statistics       Statistics of the VLT
```

```
FTOS#show vlt brief
 VLT Domain Brief
-------------------
Domain ID:                       10
Role:                            Primary
Role Priority:                   100
ICL Link Status:                 Up
HeartBeat Status:                Up
VLT Peer Status:                 Up
Local Unit Id:                   0
Version:                         6(1)
Local System MAC address:        00:01:e8:8b:24:2c
Remote System MAC address:       00:01:e8:8b:24:62
Configured System MAC address:   a0:10:10:aa:aa:aa
Remote system version:           6(1)
Delay-Restore timer:             30 seconds
Peer-Routing :                   Enabled
Peer-Routing-timeout timer:      0 seconds
Multicast peer-routing timeout: 150 seconds
```

```
FTOS#show vlt detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
10            10           UP            UP            2-4, 1000
100           100          UP            UP            2-4, 1000
```

```
FTOS#show vlt role
 VLT Role
----------
 VLT Role:                      Primary
 System MAC address:            a0:10:10:aa:aa:aa
 Primary Role Priority:         100
 Local System MAC address:      00:01:e8:8b:24:2c
 Local System Role Priority:    100
```

```
FTOS#show vlt backup-link
 VLT Backup Link
-----------------
 Destination:                    10.16.130.118
 Peer HeartBeat status:          Up
 HeartBeat Timer Interval:       1
 HeartBeat Timeout:              3
 UDP Port:                       34998
 HeartBeat Messages Sent:        14882
 HeartBeat Messages Received:    14825
```

VLT mismatch

```
VLT-10-PEER-1#show vlt mismatch
Domain
-----------
Parameters         Local               Peer
------------      -----------        ---------
   Domain-ID      11                   10
   Unit-ID        1                    1
   System-Mac     00:a1:bb:bb:bb:bb    a0:10:10:aa:aa:aa
```

For the VLT peering to be successful, the VLT '*domain-id*' should be configured with the same value in both VLT nodes. Similarly '*system-mac*' should be configured alike in both the VLT nodes. However '*unit-id*' should be unique in the VLT nodes. Any configuration error on these parameter leads to VLT mismatch and do not form VLT peering. Also FTOS logs with relevant VLT error messages:

```
Jun 25 09:24:45: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_DOMAINID_MISMATCH: VLT peer link will not be formed: Local VLT Domain-Id 11, Remote VLT Domain-Id 10 .
Jun 25 09:24:14: %STKUNIT0-M:CP %VLTMGR-6-VLT_HBEAT_DOWN: Heart beat link is down (down).
Jun 25 09:24:14: %STKUNIT0-M:CP %VLTMGR-6-VLT_PEER_STATUS: Peer chassis is down.
Jun 25 09:24:13: %STKUNIT0-M:CP %VLTMGR-6-VLT_ELECTION_ROLE: Chassis is transitioning to standalone role.
Jun 25 09:19:40: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_SYSTEM_MAC_MISMATCH: VLT peer link will not be formed: Local VLT System-Mac 00:a1:bb:bb:bb:bb, Remote VLT System-Mac a0:10:10:aa:aa:aa
Jun 25 09:17:45: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_UNITID_IDENTICAL: VLT peer link will not be formed: Local VLT UNIT-Id 1, Remote VLT UNIT-Id 1 .
Jun 25 06:41:43: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_CONFIG_MISMATCH_CLEARED: VLT peer link will be formed: No mismatches found
Jun 25 06:16:18: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_UNITID_MISMATCH: Local VLT UNIT-Id 0, Remote VLT UNIT-Id <UNKNOWN> .
Jun 25 06:14:02: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_CONFIG_MISMATCH_CLEARED: No mismatches found
Jun 25 06:13:58: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_UNITID_MISMATCH: Local VLT UNIT-Id <UNKNOWN>, Remote VLT UNIT-Id 1 .
Jun 25 06:13:34: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_SYSTEM_MAC_MISMATCH: Local VLT System-Mac 00:00:00:00:00:00, Remote VLT System-Mac a0:aa:aa:aa:aa:aa
Jun 25 06:13:04: %STKUNIT0-M:CP %VLTMGR-6-VLT_HBEAT_UP: Heart beat link is up.
Jun 25 06:12:08: %STKUNIT0-M:CP %VLTMGR-6-VLT_PEER_STATUS: Peer chassis is up.
Jun 25 06:12:08: %STKUNIT0-M:CP %VLTMGR-6-VLT_ICL_UP: InterChassis Link is up.
Jun 25 06:12:08: %STKUNIT0-M:CP %VLTMGR-6-VLT_ELECTION_ROLE: Chassis is transitioning to primary role.
```

'VLT statistics' indicates the probable error conditions for not forming VLT peering between the VLT nodes. To find the delta values (incrementing or decrementing) from initial state, clear the VLT statistics by '*clear vlt statistics*'. "*vlt brief*" indicates the status of VLT peers, ICL Link and heartbeat between the VLT nodes along with the node roles and priority.

```
VLT-100-PEER-1#show vlt statistics
VLT Domain Statistics
-----------------------
HeartBeat Messages Sent:       96882
HeartBeat Messages Received:   96885
ICL Hello's Sent:              32275
ICL Hello's Received:          32279
Domain Mismatch Errors:        0
Version Mismatch Errors:       0
Config Mismatch Errors:        0
```

```
VLT-10-PEER-1#show vlt brief
VLT Domain Brief
------------------
Domain ID:                      10
Role:                           Primary
Role Priority:                  100
ICL Link Status:                Up
HeartBeat Status:               Up
VLT Peer Status:                Up
Local Unit Id:                  0
Version:                        6(1)
Local System MAC address:       00:01:e8:8b:24:2c
Remote System MAC address:      00:01:e8:8b:24:62
Configured System MAC address:  a0:10:10:aa:aa:aa
Remote system version:          6(1)
Delay-Restore timer:            30 seconds
Peer-Routing :                  Enabled
Peer-Routing-timeout timer:     0 seconds
Multicast peer-routing timeout: 150 seconds
```

23.1 Debug outputs

VLT-PEER1#debug vlt ?	
backup	VLT Backup debugs
election	VLT Election debugs
errors	VLT Errors debugs
events	VLT interested IFM events
hello	VLT Hello debugs

VLT-PEER1#**debug vlt backup**
VLT backup debugging is ON
04:43:08 : mlagMgr: mlagMgrSendHeartBeatPkt
04:43:08 : mlagMgr: **mlagMgrSendHeartBeatPacketOut**
04:43:08 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:43:08 : mlagMgr: **mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1**
04:43:09 : mlagMgr: mlagMgrSendHeartBeatPkt
04:43:09 : mlagMgr: mlagMgrSendHeartBeatPacketOut
04:43:09 : mlagMgr: mlagMgrValidateHeartBeatPkt
<snip>

VLT-PEER1#**debug vlt events**
VLT IFM events debugging is ON
04:45:08 : mlagMgr: mlagMgrSendHeartBeatPkt
04:45:08 : mlagMgr: mlagMgrSendHeartBeatPacketOut
04:45:08 : mlagMgr: **mlagElecHandleHelloInPrimState - Entered remote-role:2**
04:45:08 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:45:08 : mlagMgr: mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1
04:45:09 : mlagMgr: mlagMgrSendHeartBeatPkt
04:45:09 : mlagMgr: mlagMgrSendHeartBeatPacketOut
04:45:09 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:45:09 : mlagMgr: mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1
<snip>

VLT-PEER1#**debug vlt hello**
VLT Hello debugging is ON
04:45:52 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:45:52 : mlagMgr: mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1
04:45:53 : mlagMgr: mlagMgrSendHeartBeatPkt
04:45:53 : mlagMgr: mlagMgrSendHeartBeatPacketOut
04:45:53 : mlagMgr**: Received Hello PDU**
04:45:53 : mlagMgr: mlagElecHandleHelloInPrimState - Entered remote-role:2
04:45:54 : mlagMgr: **Received Hello processed successfully**
04:45:54 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:45:54 : mlagMgr: mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1
04:45:54 : mlagMgr: mlagMgrSendHeartBeatPkt
04:45:54 : mlagMgr: mlagMgrSendHeartBeatPacketOut
04:45:54 : mlagMgr: mlagMgrValidateHeartBeatPkt
04:45:54 : mlagMgr: mlagMgrValidateHeartBeatPkt:Valid heartbeat message. ack:1
04:45:55 : mlagMgr: mlagMgrIclSend 263 pduType:1
04:45:55 : mlagMgr: **Hello packet sent!**
04:45:55 : mlagMgr: **Hello Timer started 3 secs**
04:45:55 : mlagMgr: mlagMgrSendHeartBeatPkt
04:45:56 : mlagMgr: mlagMgrSendHeartBeatPacketOut
<snip>

23.2 Syslog messages

If the VLT nodes are configured with mismatched domain-id's

Nov 27 11:04:48: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_DOMAINID_MISMATCH: VLT peer link will not be formed: Local VLT Domain-Id 1, Remote VLT Domain-Id 1000 .

If the VLT unit-id is configured only in one of the peer then we get the following syslog message:

Nov 27 11:00:43: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_UNITID_MISMATCH: Local VLT UNIT-Id <UNKNOWN>, Remote VLT UNIT-Id 0 .

If the mismatch is due to system MAC address

Nov 27 11:02:29: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_SYSTEM_MAC_MISMATCH: Local VLT System-Mac 00:00:00:00:00:00, Remote VLT System-Mac 00:00:00:00:00:11

When all the mismatches in the VLT domain are cleared, FTOS displays the following syslog:

Nov 27 11:02:54: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_CONFIG_MISMATCH_CLEARED: No mismatches found

If the backup link heartbeat is disrupted due to backup link failure or mismatch in backup link hello time

Nov 27 11:03:56: %STKUNIT0-M:CP %VLTMGR-6-VLT_HBEAT_DOWN: Heart beat link is down (down).

When the ICL (VLTi) link connecting VLT peers is down:

Nov 27 11:04:08: %STKUNIT0-M:CP %VLTMGR-6-VLT_ICL_DOWN: InterChassis Link is down (link down).
Nov 27 11:04:08: %STKUNIT0-M:CP %VLTMGR-6-VLT_PEER_STATUS: Peer chassis is down.

When the entire mismatches are cleared and all the links are UP, FTOS display the following:

Nov 27 11:05:32: %STKUNIT0-M:CP %VLTMGR-6-VLT_HBEAT_UP: Heart beat link is up.
Nov 27 11:05:34: %STKUNIT0-M:CP %VLTMGR-3-VLT_PEER_CONFIG_MISMATCH_CLEARED: VLT peer link will be formed: No mismatches found
Nov 27 11:05:34: %STKUNIT0-M:CP %VLTMGR-6-VLT_ICL_UP: InterChassis Link is up.
Nov 27 11:05:34: %STKUNIT0-M:CP %VLTMGR-6-VLT_PEER_STATUS: Peer chassis is up.

23.3 SNMP Traps

FTOS generates SNMP Trap for the various VLT mismatch objects. Also trap is generated when the VLTi link bandwidth exceeds 80% of its maximum capacity. With FTOS 9.2(0.0) release, multiple VLT MIB objects have been incorporated into FTOS for easier manageability and troubleshooting

FTOS(conf)#*snmp-server enable traps vlt*

Active Fabric Manager (AFM) polling for the VLT MIB and SNMP Traps from the VLT domain alerts the users for necessary corrective action.

24 FTOS Upgrade in VLT topology

FTOS upgrading procedure is almost the same as in other topology, except that certain preliminary steps to be taken care before upgrading the VLT nodes.

Step1:

- Upgrade Boot-flash partition A or B with the new image using upgrade command on both VLT peers.
- On both VLT peers, set Primary boot parameter to boot system from upgraded bootflash partition [A or B]

Step 2:

- Reload or Power-cycle one of the VLT Peer [Preferably Secondary VLT peer]. With FTOS 9.2(0.0) release, 'graceful lacp' feature ensures traffic flow through the other port-channel member.

Step3:

- Wait for Peer 2 to come up; VLT adjacency will be established.
  (Peer 2 – new image and Peer 1 - old image)

Step4:

- Peer 2 will start bringing up all the VLT LAG ports after a minute. Wait for all VLT LAG ports to come up.
- Use the command '*show vlt detail*' to confirm that all VLT ports in the local chassis are up.

Step5:

- After upgrade, save the running-configuration to memory using the "*write memory*" CLI command.

Step6:

- Ensure both Nodes are forwarding traffic.

Step 7:

- As soon as all the VLT ports are Up on Peer 2, repeat Step2 to Step5 for Peer1.

25 Conclusion

Virtual Link Trunking (VLT) feature offers efficient and optimal connectivity, cost effective and powerful alternative for the conventional layer-2 domain. As illustrated in various topologies, implementing VLT for the specific customer and migration of the existing legacy layer-2 networks to the VLT could be done without any complexity. VLT would certainly improve the high availability for any network and seamless traffic flow during failures at any layer. Convergence is the key and with VLT it elevates the network agility.

Dell Networking offers modular and stable FTOS for its high performance Ethernet Switches combined with low power consumption and low latency for the traffic, redefining the way the Data center, Enterprise, Campus networks are built, enabling greener solution and maximize customer satisfaction.