Overview

This chapter contains the following sections:

Information About VXLANs

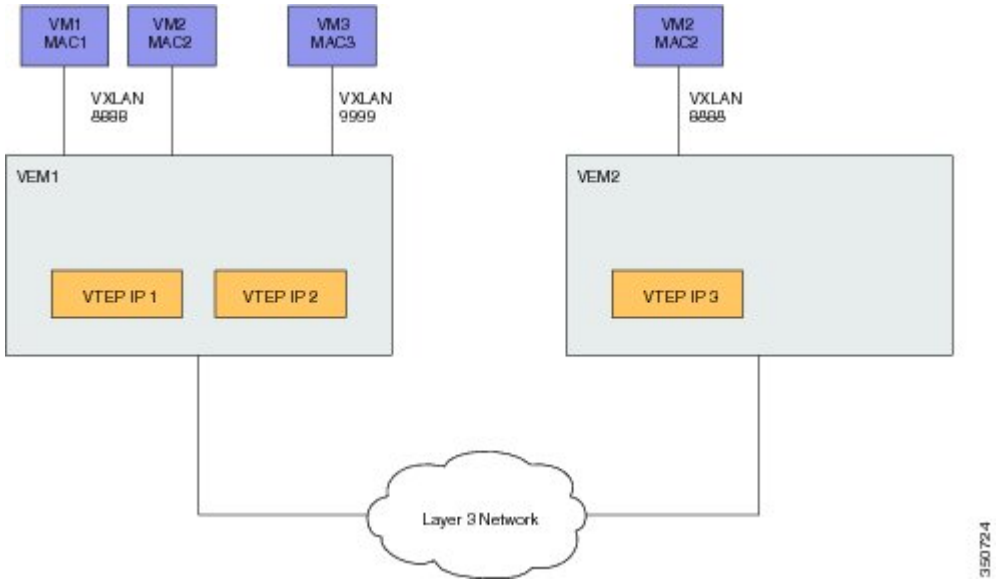
Overview of VXLANs

The Virtual Extensible LAN (VXLAN) technology enables you to create virtual domains by running a Layer 2 overlay network on top of Layer 3 with MAC-in-UDP encapsulation and a 24-bit VXLAN ID. The original Layer 2 frame from a Virtual Machine (VM) is encapsulated from within the Virtual Ethernet Module (VEM). Each VEM is assigned at least one IP address that is used as the source IP address when the encapsulated MAC frames are sent to other VEMs over the network.

The IP addresses, which are known as VXLAN Tunnel End Point (VTEP) IP addresses, are assigned to selected vmknics on the corresponding VEM. The encapsulation carries the VXLAN ID to scope the MAC

address of the payload frame. The VM's VXLAN ID is indicated within the port profile configuration of the vNIC and is applied when the VM connects to the network.

Figure 1: VXLAN Overview



A VXLAN supports two different modes for flood traffic:

- Multicast mode—A VXLAN uses an IP multicast network to send broadcast, multicast, and unknown unicast flood frames. Each multicast mode VXLAN has an assigned multicast group IP address. When a new VM joins a host in a multicast mode VXLAN, a VEM joins the assigned multicast group IP address by sending IGMP join messages. Flood traffic, broadcast, multicast and unknown unicast from the VM is encapsulated and is sent using the assigned multicast group IP address as the destination IP address. Packets sent to known unicast MAC addresses are encapsulated and sent directly to the destination server VTEP IP addresses.
- Unicast-only mode—A VXLAN uses each VEM's single unicast IP address as the destination IP address to send broadcast, multicast, and unknown unicast flood frames of the designated VTEP on each VEM that has at least one VM in the corresponding VXLAN. When a new VM joins the host in a unicast-mode VXLAN, a designated VTEP is selected for receiving flood traffic on that host. This designated VTEP is communicated to all other hosts through the Virtual Supervisor Module (VSM). Flood traffic (broadcast, multicast, and unknown unicast) is replicated on each VEM's designated VTEP in that VXLAN by encapsulating it with a VXLAN header. Packets are sent only to VEMs with a VM in that VXLAN. Packets that have a unicast MAC address are encapsulated and sent directly to the destination server's VTEP IP address.
    - MAC distribution mode (supported only in unicast mode)—In this mode, unknown unicast flooding in the network is eliminated. The VSM learns all the MAC addresses from the VEMs in all the VXLANs and distributes those MAC addresses with VTEP IP mappings to other VEMs. Therefore, there is no unknown unicast MAC address in the network when the VMs on the VEMs are communicating and controlled by the same VM.

Note MAC distribution works only for static MAC addresses. If dynamic MAC addresses are found on ports that use VXLANs that operate in MAC distribution mode, syslogs are generated to indicate that MAC distribution does not work with dynamic MAC addresses.

Note You can configure the above modes globally and override them for each bridge domain.

By default, if you upgrade the VSM from an earlier version of the Cisco Nexus 1000V to the current version with the segmentation feature enabled, all the VXLANs continue to operate in multicast mode. If you enable the feature when the VSM is running the current version of the Cisco Nexus 1000V, by default, the bridge domains change to unicast-only mode. You must explicitly enable MAC distribution mode because it is disabled by default.

After completing the software upgrade, you need to explicitly configure the segment mode to multicast mode.

Note During an upgrade, you cannot enable unicast-only mode unless you upgrade all VEMs and the VEM level.

VXLAN Tunnel Endpoints

Each VEM requires at least one IP/MAC address pair to terminate VXLAN packets. This IP/MAC address pair is known as the VXLAN Tunnel End Point (VTEP) IP/MAC addresses. The VEM supports IPv4 addressing for this purpose. The IP/MAC address that the VTEP uses is configured when you enter the **capability vxlan** command. You can have a maximum of four VTEPs in a single VEM.

One VTEP per VXLAN segment is designated to receive all broadcast, multicast, and unknown unicast flood traffic for the VEM.

When encapsulated traffic is destined to a VEM that is connected to a different subnet, the VEM does not use the VMware host routing table. Instead, the VTEPs initiate the Address Resolution Protocol (ARP) for remote VEM IP addresses. If the VTEPs in the different VEMs are in different subnets, you must configure the upstream router to respond by using the Proxy ARP.

VXLAN Gateway

VXLAN termination (encapsulation and decapsulation) is supported on virtual switches. As a result, the only endpoints that can connect into VXLANs are VMs that are connected to a virtual switch. Physical servers cannot be in VXLANs and routers or services that have traditional VLAN interfaces cannot be used by VXLAN networks. The only way that VXLANs can currently interconnect with traditional VLANs is through VM-based software routers.

The supported gateways are as follows:

- VMware vShield Edge

- Cisco VXLAN gateway
- Cisco ASA1000V

The configuration for such VLAN-VXLAN translation/mappings for the VXLAN gateway must be configured from the VSM and must always be a 1:1 mapping for each Layer 2 domain. Each VXLAN gateway can support multiple VLAN-VXLAN mappings.

VXLAN Trunks

A VXLAN trunk allows you to trunk multiple VXLANs on a single virtual Ethernet interface. In order to achieve this configuration, you must encapsulate a VLAN-VXLAN mapping on the virtual Ethernet interface.

VLAN-VXLAN mappings are applied on a virtual Ethernet interface using a port profile. A single port profile can support multiple VLAN-VXLAN mappings.

Multi-MAC Capability

You must use the multi-MAC capability feature to mark a virtual Ethernet interface as capable of sourcing packets from multiple MAC addresses. For example, you can use this feature if you have a virtual Ethernet port and you have enabled VXLAN trunking on it and the VM that is connected to the port bridges packets that are sourced from multiple MAC addresses.

By using this feature, you can easily identify such multi-MAC capable ports and handle live migration scenarios correctly for those ports.

Fragmentation

The VXLAN encapsulation overhead is 50 bytes. In order to prevent performance degradation due to fragmentation, the entire interconnection infrastructure between all VEMs that exchange VXLAN packets must be configured to carry 50 bytes more than what the VM VNICs are configured to send. For example, if the default VNIC configuration is 1500 bytes, the VEM uplink port profile, upstream physical switch port, and interswitch links, and any routers if present, must be configured to carry a maximum transmission unit (MTU) of at least 1550 bytes. If that is not possible, we recommend that you configure the MTU within the guest VMs to be smaller by 50 bytes.

If you do not configure a smaller MTU, the VEM attempts to notify the VM if it performs Path MTU (PMTU) Discovery. If the VM does not send packets with a smaller MTU, the VM fragments the IP packets. Fragmentation occurs only at the IP layer. If the VM sends a frame that is too large, the frame will be dropped after VXLAN encapsulation and if the frame does not contain an IP packet.

Scalability

Maximum Number of VXLANs

The Cisco Nexus 1000V supports up to 2048 VLANs and 2048 VXLANs with a combined maximum of 4096.

Supported Features

Jumbo Frames

Jumbo frames are supported by the Cisco Nexus 1000V if there is space on the frame to accommodate the VXLAN encapsulation overhead of at least 50 bytes, and the physical switch/router infrastructure has the capability to transport these jumbo-sized IP packets.

VXLAN Feature Disabled

As a safety precaution, do not use the **no feature segmentation** command if there are any ports associated with a VXLAN port profile. You must remove all associations before you can disable this feature. You can use the **no feature segmentation** command to remove all the VXLAN bridge domain configurations on the Cisco Nexus 1000V.

VXLAN Offload

The Cisco Nexus 1000V supports offloading VXLAN checksum and TSO computations of inner packets for VXLAN encapsulated packets. The VXLAN offload feature is supported only if an adapter supports the offload feature and VMware supports the offload feature on that adapter. To verify if the Cisco Nexus 1000V supports the VXLAN offload feature on an adapter, use the **vemcmd show pd-port** command on the host. The V flag in the Flags column indicates that the VXLAN offload feature is supported. The TSO computations are automatically offloaded when the VXLAN offload feature is supported.