UNIS S10600 系列交换机

VXLAN 配置指导

北京紫光恒越网络科技有限公司
http://www.unis-hy.com

资料版本：6W100-20160311
产品版本：S10600-CMW710-R7178

环境保护

本产品符合关于环境保护方面的设计要求，产品的存放、使用和弃置应遵照相关国家法律、法规要求进行。

前 言

本配置指导主要介绍 VXLAN（Virtual eXtensible LAN，可扩展虚拟局域网络）工作原理及相关配置。VXLAN 可以基于已有的服务提供商或企业 IP 网络，为分散的物理站点提供二层互联，并能够为不同的租户提供业务隔离。

前言部分包含如下内容：

读者对象

本手册主要适用于如下工程师：

- 网络规划人员
- 现场技术支持与维护人员
- 负责网络配置和维护的网络管理员

本书约定

1. 命令行格式约定

格　式	意　义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从多个选项中仅选取一个。
[x | y | ...]	表示从多个选项中选取一个或者不选。
{ x | y | ... }*	表示从多个选项中至少选取一个。
[x | y | ...]*	表示从多个选项中选取一个、多个或者不选。
&<1-n>	表示符号&前面的参数可以重复输入1～n次。
#	由“#”号开始的行表示为注释行。

2. 图形界面格式约定

格　式	意　义
< >	带尖括号“< >”表示按钮名，如“单击<确定>按钮”。

格　式	意　义
[ ]	带方括号“[ ]”表示窗口名、菜单名和数据表，如“弹出[新建用户]窗口”。
/	多级菜单用“/”隔开。如[文件/新建/文件夹]多级菜单表示[文件]菜单下的[新建]子菜单下的[文件夹]菜单项。

3. 各类标志

本书还采用各种醒目标志来表示在操作过程中应该特别注意的地方，这些标志的意义如下：

 警告	该标志后的注释需给予格外关注，不当的操作可能会对人身造成伤害。
 注意	提醒操作中应注意的事项，不当的操作可能会导致数据丢失或者设备损坏。
 提示	为确保设备配置成功或者正常工作而需要特别关注的操作或信息。
 说明	对操作内容的描述进行必要的补充和说明。
 窍门	配置、操作、或使用设备的技巧、小窍门。

4. 图标约定

本书使用的图标及其含义如下：

	该图标及其相关描述文字代表一般网络设备，如路由器、交换机、防火墙等。
	该图标及其相关描述文字代表一般意义下的路由器，以及其他运行了路由协议的设备。
	该图标及其相关描述文字代表二、三层以太网交换机，以及运行了二层协议的设备。
	该图标及其相关描述文字代表无线控制器、无线控制器业务板和有线无线一体化交换机的无线控制引擎设备。
	该图标及其相关描述文字代表无线接入点设备。
	该图标及其相关描述文字代表无线终结单元。
	该图标及其相关描述文字代表无线终结者。
	该图标及其相关描述文字代表无线Mesh设备。
	该图标代表发散的无线射频信号。
	该图标代表点到点的无线射频信号。

	该图标及其相关描述文字代表防火墙、UTM、多业务安全网关、负载均衡等安全设备。
	该图标及其相关描述文字代表防火墙插卡、负载均衡插卡、NetStream插卡、SSL VPN插卡、IPS插卡、ACG插卡等安全插卡。

5. 端口编号示例约定

本手册中出现的端口编号仅作示例，并不代表设备上实际具有此编号的端口，实际使用中请以设备上存在的端口编号为准。

产品配套资料

紫光恒越 S10600 交换机的配套资料包括如下部分：

大类	资料名称	内容介绍
硬件描述与安装	安全兼容性手册	列出产品的兼容性声明，并对兼容性和安全的细节进行说明
	快速入门	指导您对设备进行初始安装、配置，通常针对最常用的情况，减少您的检索时间
	安装指导	帮助您详细了解设备硬件规格和安装方法，指导您对设备进行安装
	单板手册	帮助您详细了解单板的硬件规格
业务配置	配置指导	帮助您掌握设备软件功能的配置方法及配置步骤
	命令参考	详细介绍设备的命令，相当于命令字典，方便您查阅各个命令的功能
运行维护	版本说明书	帮助您了解产品版本的相关信息（包括：版本配套说明、兼容性说明、特性变更说明、技术支持信息）及软件升级方法

技术支持

用户支持邮箱：zgsm_service@thunis.com

技术支持热线电话：400-910-9998（手机、固话均可拨打）

网址：http://www.unis-hy.com

资料意见反馈

如果您在使用过程中发现产品资料的任何问题，可以通过以下方式反馈：

E-mail：zgsm_info@thunis.com

感谢您的反馈，让我们做得更好！

目 录

1 VXLAN简介

VXLAN（Virtual eXtensible LAN，可扩展虚拟局域网络）是基于 IP 网络、采用“MAC in UDP”封装形式的二层 VPN 技术。VXLAN 可以基于已有的服务提供商或企业 IP 网络，为分散的物理站点提供二层互联，并能够为不同的租户提供业务隔离。VXLAN 主要应用于数据中心网络。

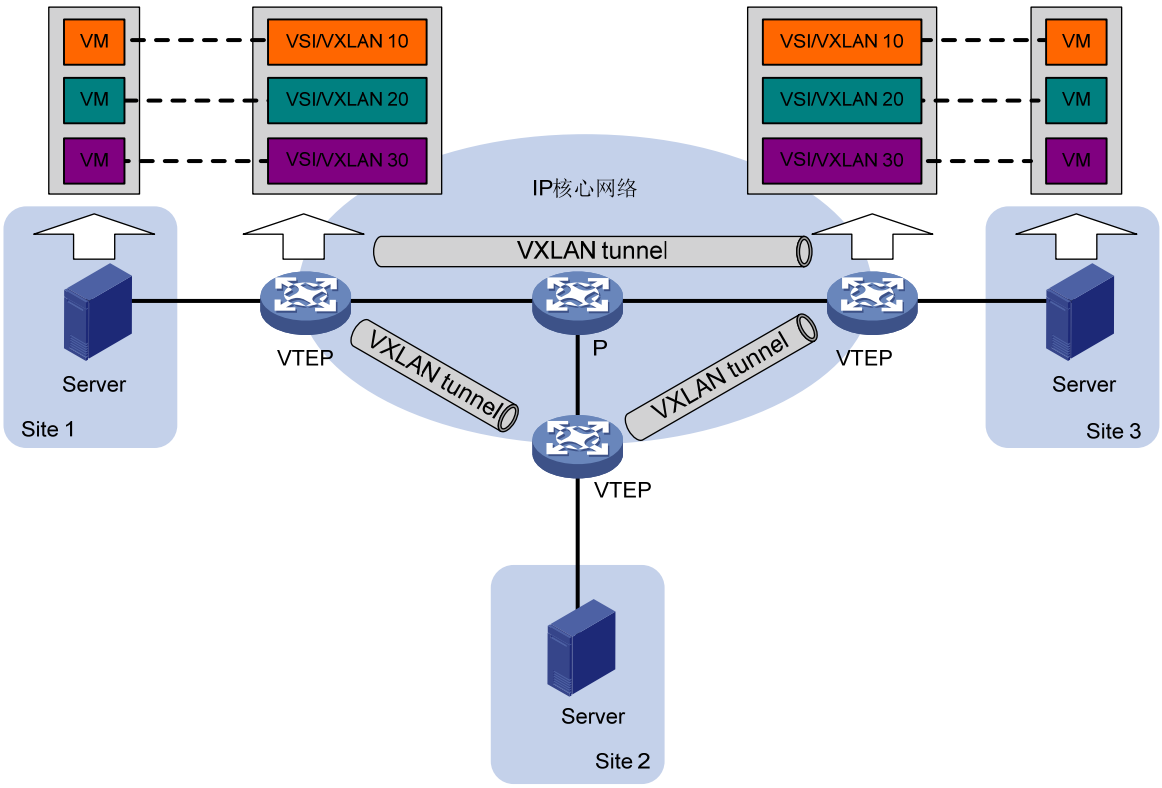
VXLAN 具有如下特点：

- 支持大量的租户：使用 24 位的标识符，最多可支持 2 的 24 次方（16777216）个 VXLAN，支持的租户数目大规模增加，解决了传统二层网络 VLAN 资源不足的问题。
- 易于维护：基于 IP 网络组建大二层网络，使得网络部署和维护更加容易，并且可以充分地利用现有的 IP 网络技术，例如利用等价路由进行负载分担等；只有 IP 核心网络的边缘设备需要进行 VXLAN 处理，网络中间设备只需根据 IP 头转发报文，降低了网络部署的难度和费用。

目前，设备只支持基于 IPv4 网络的 VXLAN 技术，不支持基于 IPv6 网络的 VXLAN 技术。

1.1 VXLAN网络模型

图1-1 VXLAN 网络模型示意图



如 图 1-1 所示，VXLAN的典型网络模型中包括如下几部分：

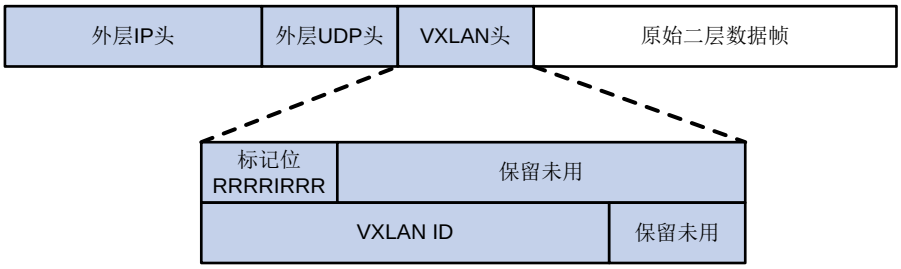
- VM（Virtual Machine，虚拟机）：在一台服务器上可以创建多台虚拟机，不同的虚拟机可以属于不同的 VXLAN。属于相同 VXLAN 的虚拟机处于同一个逻辑二层网络，彼此之间二层互

通；属于不同 VXLAN 的虚拟机之间二层隔离。VXLAN 通过 VXLAN ID 来标识，VXLAN ID 又称 VNI（VXLAN Network Identifier，VXLAN 网络标识符），其长度为 24 比特。

- VTEP（VXLAN Tunnel End Point，VXLAN 隧道端点）：VXLAN 的边缘设备。VXLAN 的相关处理都在 VTEP 上进行，例如识别以太网数据帧所属的 VXLAN、基于 VXLAN 对数据帧进行二层转发、封装/解封装报文等。VTEP 可以是一台独立的物理设备，也可以是虚拟机所在的服务器。
- VXLAN 隧道：两个 VTEP 之间的点到点逻辑隧道。VTEP 为数据帧封装 VXLAN 头、UDP 头、IP 头后，通过 VXLAN 隧道将封装后的报文转发给远端 VTEP，远端 VTEP 对其进行解封装。
- 核心设备：IP核心网络中的设备（如 图 1-1 中的P设备）。核心设备不参与VXLAN处理，仅需要根据封装后报文的目的IP地址对报文进行三层转发。
- VSI（Virtual Switching Instance，虚拟交换实例）：VTEP 上为一个 VXLAN 提供二层交换服务的虚拟交换实例。VSI 可以看作是 VTEP 上的一台基于 VXLAN 进行二层转发的虚拟交换机，它具有传统以太网交换机的所有功能，包括源 MAC 地址学习、MAC 地址老化、泛洪等。VSI 与 VXLAN 一一对应。

1.2 VXLAN报文封装格式

图1-2 VXLAN 报文封装示意图



如 图 1-2 所示，VXLAN报文的封装格式为：在原始二层数据帧外添加 8 字节VXLAN头、8 字节UDP头和 20 字节IP头。其中，UDP头的目的端口号为VXLAN UDP端口号（缺省为 4789）。VXLAN头主要包括两部分：

- 标记位：“I”位为 1 时，表示 VXLAN 头中的 VXLAN ID 有效；为 0，表示 VXLAN ID 无效。其他位保留未用，设置为 0。
- VXLAN ID：用来标识一个 VXLAN 网络，长度为 24 比特。

1.3 VXLAN运行机制

VXLAN 运行机制可以概括为：

(1) 发现远端 VTEP，在 VTEP 之间建立 VXLAN 隧道，并将 VXLAN 隧道与 VXLAN 关联。

(2) 识别接收到的报文所属的 VXLAN，以便将报文的源 MAC 地址学习到 VXLAN 对应的 VSI，并在该 VSI 内转发该报文。

(3) 学习虚拟机的 MAC 地址。

(4) 根据学习到的 MAC 地址表项转发报文。

1.3.1 建立VXLAN隧道，并将VXLAN隧道与VXLAN关联

为了将 VXLAN 报文传递到远端 VTEP，需要创建 VXLAN 隧道，并将 VXLAN 隧道与 VXLAN 关联。

1. 创建VXLAN隧道

VXLAN 隧道的建立方式为：手工配置 Tunnel 接口，并指定隧道的源和目的 IP 地址分别为本端和远端 VTEP 的 IP 地址。

2. 关联VXLAN隧道与VXLAN

VXLAN 隧道与 VXLAN 关联的方式为：手工将 VXLAN 隧道与 VXLAN 关联。

1.3.2 识别报文所属的VXLAN

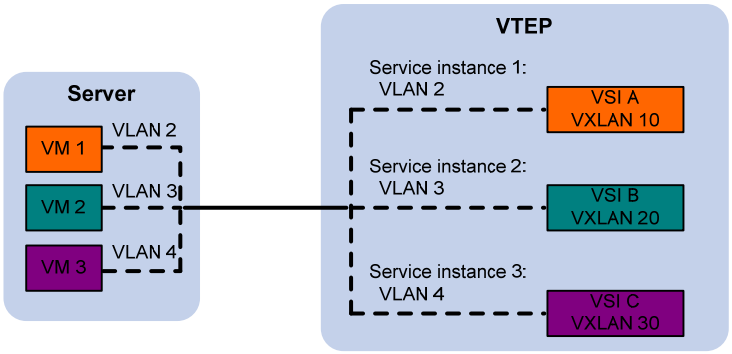
1. 本地站点内接收到数据帧的识别

VTEP 将连接本地站点的以太网服务实例（Service Instance）与 VSI 关联。VTEP 从以太网服务实例接收到数据帧后，查找与其关联的 VSI，VSI 内创建的 VXLAN 即为该数据帧所属的 VXLAN。

在 VXLAN 中，与 VSI 关联的以太网服务实例也称为 AC（Attachment Circuit，接入电路）。以太网服务实例在二层以太网接口上创建，它定义了一系列匹配规则，用来匹配从该二层以太网接口上接收到的数据帧。

如 图 1-3 所示，VM 1 属于VLAN 2，在VTEP上配置以太网服务实例 1 匹配VLAN 2 的报文，将以太网服务实例 1 与VSI A绑定，并在VSI A内创建VXLAN 10，则VTEP接收到VM 1 发送的数据帧后，可以判定该数据帧属于VXLAN 10。

图1-3 二层数据帧所属 VXLAN 识别



2. VXLAN隧道上接收报文的识别

对于从 VXLAN 隧道上接收到的 VXLAN 报文，VTEP 根据报文中携带的 VXLAN ID 判断该报文所属的 VXLAN。

1.3.3 学习MAC地址

MAC 地址学习分为本地 MAC 地址学习和远端 MAC 地址学习两部分。

1. 本地MAC地址学习

是指 VTEP 对本地站点内虚拟机 MAC 地址的学习。VTEP 接收到本地虚拟机发送的数据帧后，判断该数据帧所属的 VSI，并将数据帧中的源 MAC 地址（本地虚拟机的 MAC 地址）添加到该 VSI 的 MAC 地址表中，该 MAC 地址对应的接口为接收到数据帧的接口。

VXLAN 不支持静态配置本地 MAC 地址。

2. 远端MAC地址学习

是指 VTEP 对远端站点内虚拟机 MAC 地址的学习。远端 MAC 地址的学习方式有如下几种：

- 静态配置：手工指定远端 MAC 地址所属的 VSI（VXLAN），及其对应的 VXLAN 隧道接口。
- 通过报文中的源MAC地址动态学习：VTEP从VXLAN隧道上接收到远端VTEP发送的VXLAN报文后，根据 VXLAN ID 判断报文所属的 VXLAN，对报文进行解封装，还原二层数据帧，并将数据帧中的源 MAC 地址（远端虚拟机的 MAC 地址）添加到所属 VXLAN 对应 VSI 的 MAC 地址表中，该 MAC 地址对应的接口为 VXLAN 隧道接口。

静态配置的远端 MAC 地址表项优先级高于源 MAC 地址动态学习的表项。

1.3.4 接入模式

接入模式分为以下两种：

- VLAN 接入模式：从本地站点接收到的、发送给本地站点的以太网帧必须带有 VLAN tag。VTEP 从本地站点接收到以太网帧后，删除该帧的所有 VLAN tag，再转发该数据帧；VTEP 发送以太网帧到本地站点时，为其添加 VLAN tag。采用该模式时，VTEP 不会传递 VLAN tag 信息，不同站点可以独立地规划自己的 VLAN，不同站点的不同 VLAN 之间可以互通。
- Ethernet 接入模式：从本地站点接收到的、发送给本地站点的以太网帧可以携带 VLAN tag，也可以不携带 VLAN tag。VTEP 从本地站点接收到以太网帧后，保持该帧的 VLAN tag 信息不变，转发该数据帧；VTEP 发送以太网帧到本地站点时，不会为其添加 VLAN tag。采用该模式时，VTEP 会在不同站点间传递 VLAN tag 信息，不同站点的 VLAN 需要统一规划，否则无法互通。

缺省情况下，接入模式为 VLAN 模式。下文对于流量转发过程的介绍均以 VLAN 模式为例。

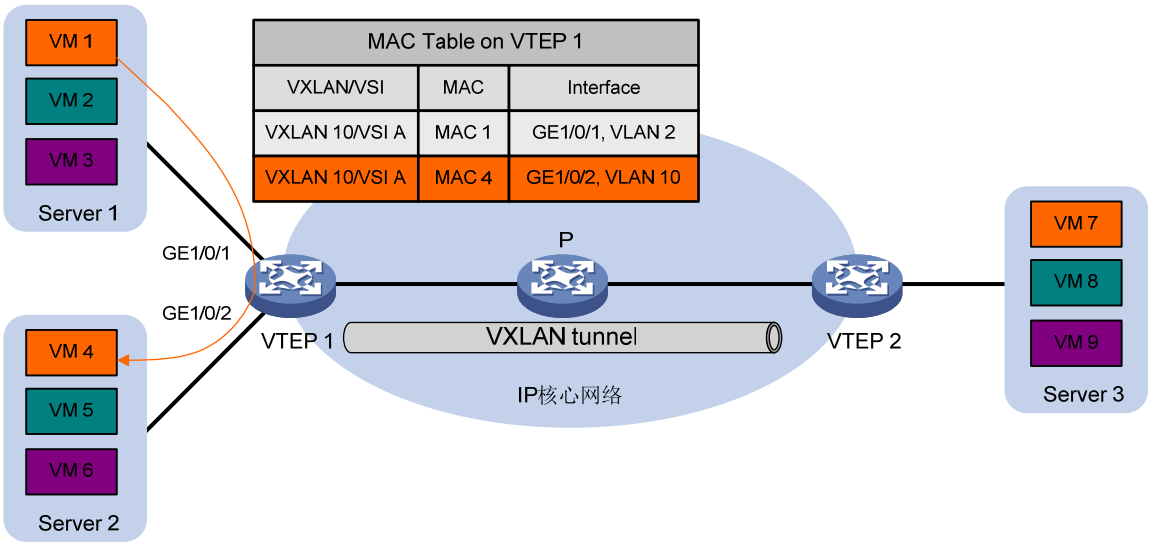
1.3.5 转发单播流量

完成本地和远端 MAC 地址学习后，VTEP 在 VXLAN 内转发单播流量的过程如下所述。

1. 站点内流量

对于站点内流量，VTEP 判断出报文所属的 VSI 后，根据目的 MAC 地址查找该 VSI 的 MAC 地址表，从相应的本地接口转发给目的 VM。

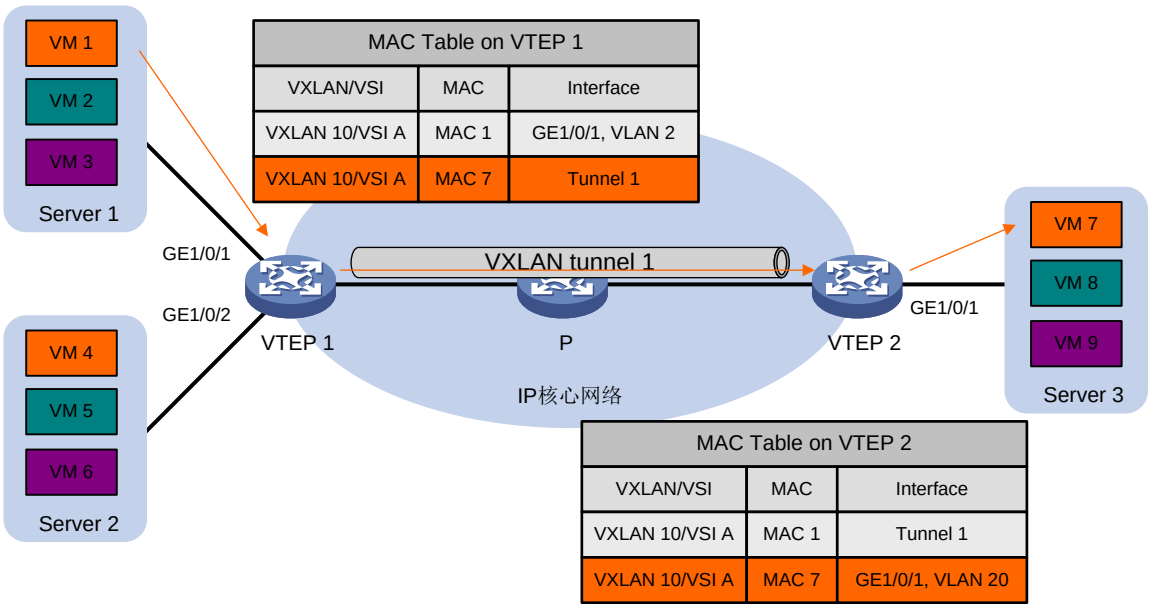
图1-4 站点内单播流量转发



如 图 1-4 所示，VM 1（MAC地址为MAC 1）发送以太网帧到VM 4（MAC地址为MAC 4）时，VTEP 1 从接口GigabitEthernet1/0/1 收到该以太网帧后，判断该数据帧属于VSI A（VXLAN 10），查找VSI A的MAC地址表，得到MAC 4 的出接口为GigabitEthernet1/0/2，所在VLAN为VLAN 10，则将以太网帧从接口GigabitEthernet1/0/2 的VLAN 10 内发送给VM 4。

2. 站点间流量

图1-5 站点间单播流量转发



如 图 1-5 所示，以VM 1（MAC地址为MAC 1）发送以太网帧给VM 7（MAC地址为MAC 7）为例，站点间单播流量的转发过程为：

(1) VM 1 发送以太网数据帧给 VM 7，数据帧的源 MAC 地址为 MAC 1，目的 MAC 为 MAC 7，VLAN tag 为 2。

(2) VTEP 1 从接口 GigabitEthernet1/0/1 收到该数据帧后，判断该数据帧属于 VSI A（VXLAN 10），查找 VSI A 的 MAC 地址表，得到 MAC 7 的出端口为 Tunnel1。

(3) VTEP 1 为数据帧封装 VXLAN 头、UDP 头和 IP 头后，将封装好的报文通过 VXLAN 隧道 Tunnel1、经由 P 设备发送给 VTEP 2。

(4) VTEP 2 接收到报文后，根据报文中的 VXLAN ID 判断该报文属于 VXLAN 10，并剥离 VXLAN 头、UDP 头和 IP 头，还原出原始的数据帧。

(5) VTEP 2 查找与 VXLAN 10 对应的 VSI A 的 MAC 地址表，得到 MAC 7 的出端口为 GigabitEthernet1/0/1，所在 VLAN 为 VLAN 20。

(6) VTEP 2 从接口 GigabitEthernet1/0/1 的 VLAN 20 内将数据帧发送给 VM 7。

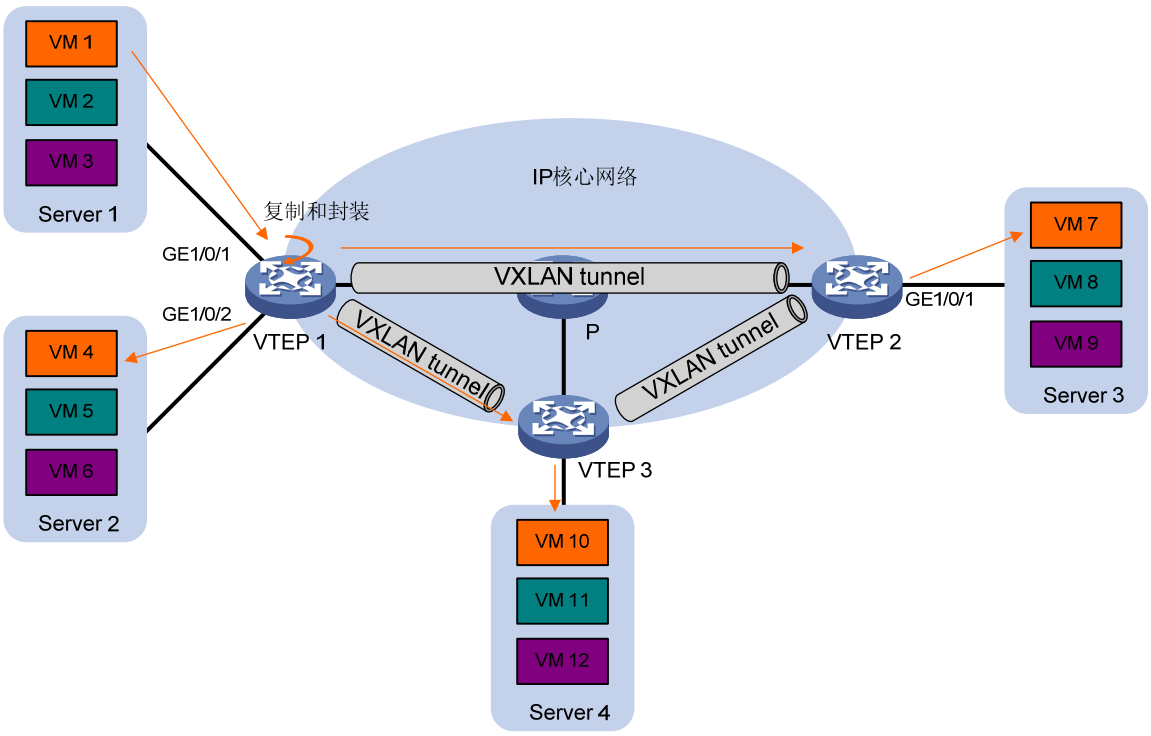
1.3.6 转发泛洪流量

泛洪流量包括组播、广播和未知单播流量。根据复制方式的不同，流量泛洪方式分为单播路由方式（头端复制）和组播路由方式（核心复制）两种。

1. 单播路由方式（头端复制）

在单播路由方式下，VTEP 负责复制报文，采用单播方式将复制后的报文通过本地接口发送给本地站点，并通过 VXLAN 隧道发送给 VXLAN 内的所有远端 VTEP。

图1-6 单播路由方式转发示意图



如 图 1-6 所示，单播路由方式的泛洪流量转发过程为：

(1) VTEP 1 接收到本地虚拟机发送的组播、广播和未知单播数据帧后，判断数据帧所属的 VXLAN，通过该 VXLAN 内除接收接口外的所有本地接口和 VXLAN 隧道转发该数据帧。通过 VXLAN

隧道转发数据帧时，需要为其封装 VXLAN 头、UDP 头和 IP 头，将泛洪流量封装在多个单播报文中，发送到 VXLAN 内的所有远端 VTEP。

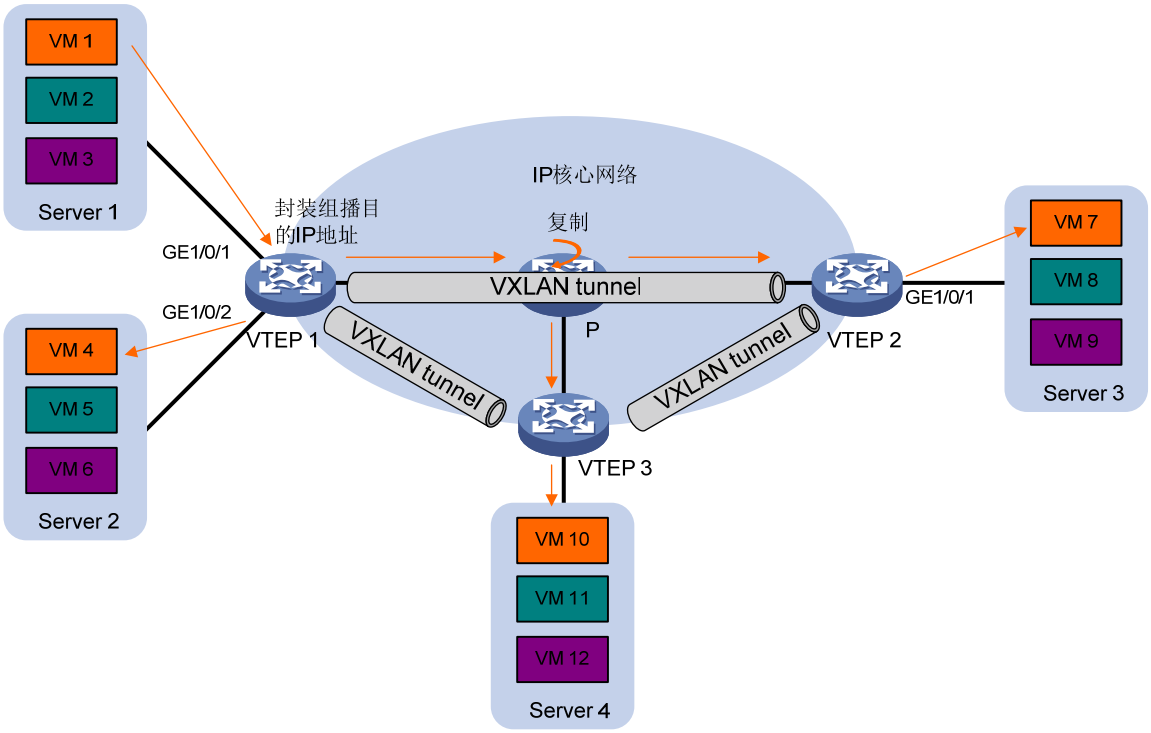
(2) 远端 VTEP（VTEP 2 和 VTEP 3）接收到 VXLAN 报文后，解封装报文，将原始的数据帧在本地站点的指定 VXLAN 内泛洪。为了避免环路，远端 VTEP 从 VXLAN 隧道上接收到报文后，不会再将其泛洪到其他的 VXLAN 隧道。

2. 组播路由方式（核心复制）

数据中心网络中需要通过 IP 核心网络进行二层互联的站点较多时，采用组播路由方式可以节省泛洪流量对核心网络带宽资源的占用。

在组播路由方式下，同一个 VXLAN 内的所有 VTEP 都加入同一个组播组，利用组播路由协议（如 PIM）在 IP 核心网上为该组播组建立组播转发表项。VTEP 接收到泛洪流量后，不仅在本地站点内泛洪，还会为其封装组播目的IP地址，封装后的报文根据已建立的组播转发表项转发到远端VTEP。

图1-7 组播路由方式转发示意图



如 图 1-7 所示，组播路由方式的泛洪流量转发过程为：

(1) VTEP 1 接收到本地虚拟机发送的组播、广播和未知单播数据帧后，判断数据帧所属的 VXLAN，不仅通过该 VXLAN 内除接收接口外的所有本地接口将数据帧转发到本地站点，还会为其封装 VXLAN 头、UDP 头和 IP 头（目的 IP 地址为组播地址）通过组播转发表项将其发送到远端 VTEP。

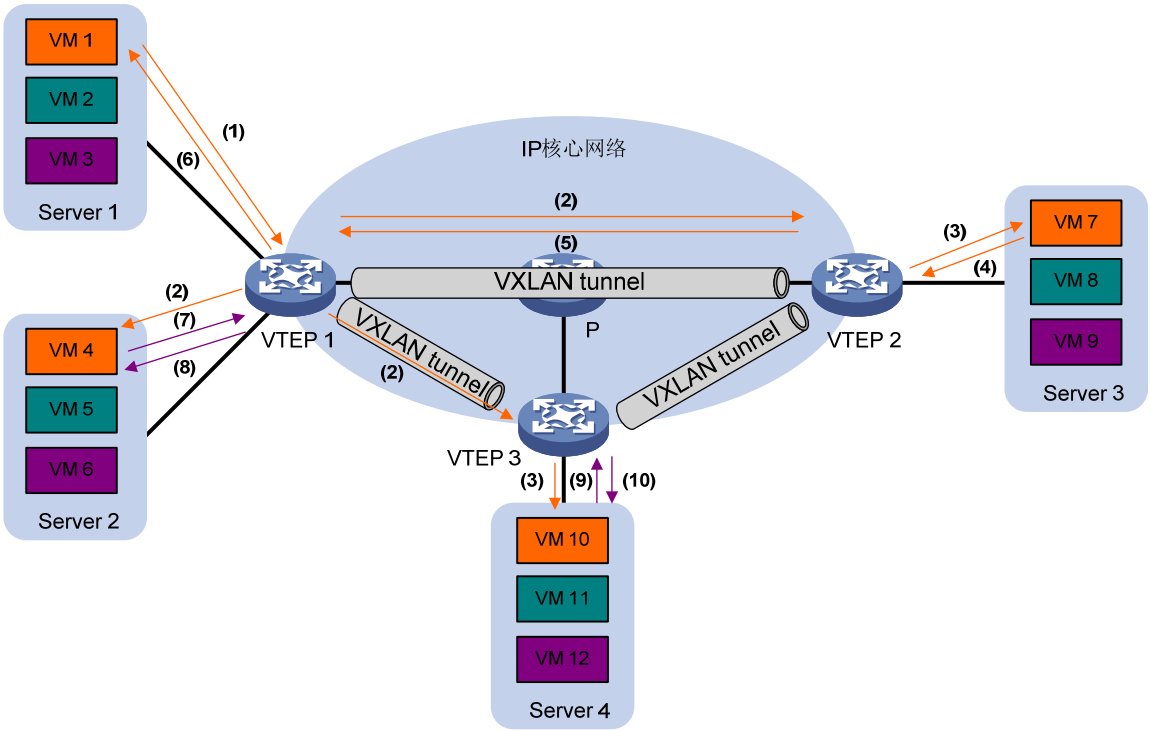
(2) 在 IP 核心网内，P 设备根据已经建立的组播转发表项复制并转发该组播报文。

(3) 远端 VTEP（VTEP 2 和 VTEP 3）接收到 VXLAN 报文后，解封装报文，将原始的数据帧在本地站点的指定 VXLAN 内泛洪。为了避免环路，远端 VTEP 从 VXLAN 隧道上接收到报文后，不会再将其泛洪到其他的 VXLAN 隧道。

1.4 ARP泛洪抑制

为了避免广播发送的ARP请求报文占用核心网络带宽，VTEP从本地站点、VXLAN隧道接收到ARP请求和ARP应答报文后，根据该报文在本地建立ARP泛洪抑制表项。后续当VTEP收到本站点内虚拟机请求其他虚拟机MAC地址的ARP请求时，优先根据ARP泛洪抑制表项进行代答。如果没有对应的表项，则将ARP请求泛洪到核心网。ARP泛洪抑制功能可以大大减少ARP泛洪的次数。

图1-8 ARP泛洪抑制示意图



如 图 1-8 所示，ARP泛洪抑制的处理过程如下：

(1) 虚拟机 VM 1 发送 ARP 请求，获取 VM 7 的 MAC 地址。

(2) VTEP 1 根据接收到的ARP请求，建立VM 1 的ARP泛洪抑制表项，并在VXLAN内泛洪该ARP请求（图 1-8 以单播路由泛洪方式为例）。

(3) 远端 VTEP（VTEP 2 和 VTEP 3）解封装 VXLAN 报文，获取原始的 ARP 请求报文后，建立 VM 1 的 ARP 泛洪抑制表项，并在本地站点的指定 VXLAN 内泛洪该 ARP 请求。

(4) VM 7 接收到 ARP 请求后，回复 ARP 应答报文。

(5) VTEP 2 接收到 ARP 应答后，建立 VM 7 的 ARP 泛洪抑制表项，并通过 VXLAN 隧道将 ARP 应答发送给 VTEP 1。

(6) VTEP 1 解封装 VXLAN 报文，获取原始的 ARP 应答，并根据该应答建立 VM 7 的 ARP 泛洪抑制表项，之后将 ARP 应答报文发送给 VM 1。

(7) 在 VTEP 1 上建立 ARP 泛洪抑制表项后，虚拟机 VM 4 发送 ARP 请求，获取 VM 1 或 VM 7 的 MAC 地址。

(8) VTEP 1 接收到 ARP 请求后，建立 VM 4 的 ARP 泛洪抑制表项，并查找本地 ARP 泛洪抑制表项，根据已有的表项回复 ARP 应答报文，不会对 ARP 请求进行泛洪。

(9) 在 VTEP 3 上建立 ARP 泛洪抑制表项后，虚拟机 VM 10 发送 ARP 请求，获取 VM 1 的 MAC 地址。

(10) VTEP 3 接收到 ARP 请求后，建立 VM 10 的 ARP 泛洪抑制表项，并查找本地 ARP 泛洪抑制表项，根据已有的表项回复 ARP 应答报文，不会对 ARP 请求进行泛洪。

1.5 协议规范

与 VXLAN 相关的协议规范有：

- IETF 草案：draft-mahalingam-dutt-dcops-vxlan-04

2 VXLAN配置限制和指导

2.1 硬件限制

- VXLAN 公网侧端口必须位于 SG 系列接口板上。
- VXLAN 用户侧端口必须位于下列 SE 系列接口板上：
  - LSUM2GP24TSSE0
  - LSUM2GP44TSSE0
  - LSUM2GT24PTSSE0
  - LSUM2GT48SE0
  - SG 系列接口板
- 当设备处于 IRF 模式时，IRF 物理端口必须位于 SG 系列接口板上，VXLAN 才能正常运行。有关 IRF 模式、IRF 物理端口的介绍，请参见“虚拟化技术配置指导”中的“IRF”。

2.2 软件限制

- 多个 VXLAN 隧道共用一个公网侧端口时，这些隧道必须使用设备的同一个 VLAN 接口来转发报文。
- 请不要在 VXLAN 网络中配置 MPLS、EVI 相关功能。有关 MPLS 和 EVI 的详细介绍，请分别参见“MPLS 配置指导”和“EVI 配置指导”。
- 端口使能了 EVB 功能后，不支持配置 VXLAN 功能，否则二者均将无法正常工作。有关 EVB 的介绍，请参见“EVB 配置指导”。
- 在 VXLAN 组网中，IP 核心网络中的设备只需要配置路由协议，确保 VTEP 之间路由可达。VXLAN 相关配置都在 VTEP 上进行。
- VTEP 上配置了 AC（与 VSI 关联的以太网服务实例）的端口不支持三层组播功能。
- 为保证 VTEP 与本地站点的正常对接，应在配置了 AC 的端口上关闭生成树协议（**undo stp enable**）。
- 在组播路由方式下，不允许在核心设备上创建 VXLAN 或 VXLAN 隧道，否则会导致 VXLAN 报文转发不通。

3 配置VXLAN

3.1 VXLAN配置任务简介

表3-1 VXLAN 配置任务简介

配置任务	说明	详细配置
创建VSI和VXLAN	必选	3.2

配置任务	说明	详细配置
配置VXLAN隧道	必选	3.3
关联VXLAN与VXLAN隧道	必选	3.4
配置AC与VSI关联	必选	3.5
管理本地和远端MAC地址	可选	3.6
配置VXLAN组播路由泛洪方式	可选	3.7
配置VSI泛洪抑制	可选	3.8
配置VXLAN报文的目的UDP端口号	可选	3.9
配置VXLAN报文检查功能	可选	3.10
配置ARP泛洪抑制	可选	3.11

3.2 创建VSI和VXLAN

表3-2 创建 VSI 和 VXLAN

操作	命令	说明
进入系统视图	**system-view**	-
使能L2VPN功能	**l2vpn enable**	缺省情况下，L2VPN功能处于关闭状态
创建VSI，并进入VSI视图	**vsi** *vsi-name*	缺省情况下，设备上不存在任何VSI
（可选）设置VSI的描述信息	**description** *text*	缺省情况下，未配置VSI的描述信息
开启当前的VSI	**undo shutdown**	缺省情况下，VSI处于开启状态
（可选）配置VSI的MTU值	**mtu** *mtu*	缺省情况下，VSI的MTU值为1500字节
（可选）开启VSI的MAC地址学习功能	**mac-learning enable**	缺省情况下，VSI的MAC地址学习功能处于开启状态 对于VXLAN，设备不支持关闭VSI的MAC地址学习功能，执行**undo mac-learning enable**命令后不生效
创建VXLAN，并进入VXLAN视图	**vxlan** *vxlan-id*	缺省情况下，设备上不存在任何VXLAN 在一个VSI下只能创建一个VXLAN 不同VSI下创建的VXLAN，其VXLAN ID不能相同

3.3 配置VXLAN隧道

手工创建 VXLAN 隧道时，隧道的源端地址和目的端地址需要分别手工指定为本地和远端 VTEP 的接口地址。在同一台设备上，VXLAN 隧道模式的不同 Tunnel 接口建议不要同时配置完全相同的源端地址和目的端地址。

关于隧道的详细介绍及 Tunnel 接口下的更多配置命令，请参见“三层技术-IP 业务配置指导”中的“隧道”。关于 **interface tunnel**、**source** 和 **destination** 命令的详细介绍，请参见“三层技术-IP 业务命令参考”中的“隧道”。

表3-3 手工创建 VXLAN 隧道

操作	命令	说明
进入系统视图	**system-view**	-
配置VXLAN隧道的全局源地址	**tunnel global source-address** *ip-address*	缺省情况下，未配置VXLAN隧道的全局源地址 如果隧道下未配置源地址或源接口，则隧道会使用全局源地址作为隧道的源地址
（可选）配置保留VXLAN	**reserved vxlan** *vxlan-id*	缺省情况下，未指定保留VXLAN 当需要开启隧道的BFD检测功能时，必须配置保留VXLAN
创建模式为VXLAN隧道的Tunnel接口，并进入Tunnel接口视图	**interface tunnel** *tunnel-number* **mode vxlan**	缺省情况下，设备上不存在任何Tunnel接口 在隧道的两端应配置相同的隧道模式，否则会造成报文传输失败
配置隧道的源端地址或源接口	**source** { *ipv4-address* | *interface-type interface-number* }	缺省情况下，没有设置VXLAN隧道的源端地址和源接口 如果设置的是隧道的源端地址，则该地址将作为封装后VXLAN报文的源IP地址；如果设置的是隧道的源接口，则该接口的主IP地址将作为封装后VXLAN报文的源IP地址 需要注意的是，采用VXLAN组播路由泛洪方式时，VXLAN隧道的源接口不能是Loopback接口，源端地址不能是Loopback接口的地址
配置隧道的目的端地址	**destination** *ipv4-address*	缺省情况下，未指定隧道的目的端地址 隧道的目的端地址是对端设备上接口的IP地址，该地址将作为封装后VXLAN报文的目的地址
（可选）开启隧道的BFD检测功能，并设置BFD报文的目的MAC地址	**tunnel bfd enable destination-mac** *mac-address*	缺省情况下，隧道的BFD检测功能处于关闭状态 开启隧道的BFD检测功能后，VTEP将自动建立单跳控制报文方式的BFD会话对VXLAN隧道的状态进行检测。检测方式为：隧道两端的VTEP设备均周期性地向配置的目的MAC地址发送BFD控制报文，并对报文进行VXLAN隧道封装，如果在5秒内没有接收到对端发送的BFD控制报文，则将隧道状态置为Defect，隧道接口状态仍为Up。当VXLAN隧道恢复正常后，隧道状态可自动恢复Up 执行本命令的同时，需要在系统视图下执行**reserved vxlan**命令配置保留VXLAN。否则，BFD会话无法up

3.4 关联VXLAN与VXLAN隧道

一个 VXLAN 可以关联多条 VXLAN 隧道。一条 VXLAN 隧道可以关联多个 VXLAN，这些 VXLAN 共用该VXLAN隧道，VTEP 根据VXLAN报文中的VXLAN ID来识别隧道传递的报文所属的VXLAN。

VTEP 接收到某个 VXLAN 的泛洪流量后，如果采用单播路由泛洪方式，则 VTEP 将在与该 VXLAN 关联的所有 VXLAN 隧道上发送该流量，以便将流量转发给所有的远端 VTEP。

表3-4 手工关联 VXLAN 与 VXLAN 隧道

操作	命令	说明
进入系统视图	**system-view**	-
进入VSI视图	**vsi** *vsi-name*	-
进入VXLAN视图	**vxlan** *vxlan-id*	-
配置VXLAN与VXLAN隧道关联	**tunnel** { *tunnel-number* | **all** }	缺省情况下，VXLAN没有与任何VXLAN隧道关联 VTEP必须与相同VXLAN内的其它VTEP建立VXLAN隧道，并将该隧道与VXLAN关联

3.5 配置AC与VSI关联

3.5.1 配置以太网服务实例与VSI关联

将以太网服务实例与 VSI 关联后，从该接口接收到的、符合以太网服务实例报文匹配规则的报文，将通过查找关联 VSI 的 MAC 地址表进行转发。以太网服务实例提供了多种报文匹配规则（包括接口接收到的所有报文、所有携带 VLAN Tag 的报文和所有不携带 VLAN Tag 的报文等），为报文关联 VSI 提供了更加灵活的方式。

配置以太网服务实例与 VSI 关联时，需要注意：

- 为确保转发正常，端口上以太网服务实例的报文匹配规则需要与该端口上允许通过的 VLAN、VLAN 报文是否带 Tag 配置保持一致。
- 如果以太网服务实例采用缺省的报文匹配规则（**encapsulation default**）或匹配携带 VLAN Tag 的报文（**encapsulation tagged**），为确保转发正常，请指定接入模式为 Ethernet。
- 当接入模式为 VLAN 时，如果端口接收到的报文不带 Tag，需要配置报文匹配规则为 **encapsulation untagged**。

表3-5 配置以太网服务实例与 VSI 关联

操作		命令	说明
进入系统视图		**system-view**	-
进入二层接口视图	进入二层以太网接口视图	**interface** *interface-type interface-number*	二者选其一
	进入二层聚合接口视图	**interface bridge-aggregation** *interface-number*	
将二层接口加入本地站点VLAN	配置端口的链路类型	**port link-type** { **access** | **trunk** | **hybrid** }	缺省情况下，所有端口的链路类型均为Access类型
	将当前端口加入本地站点VLAN	**port access vlan** *vlan-id*	三者选其一 本地站点VLAN必须是设备上已创建的VLAN
		port trunk permit vlan { *vlan-id-list* | **all** }	
		port hybrid vlan *vlan-id-list* { **tagged** |	

操作	命令	说明
	untagged }	
创建以太网服务实例，并进入以太网服务实例视图	**service-instance** *instance-id*	缺省情况下，不存在任何以太网服务实例
配置以太网服务实例的报文匹配规则	**encapsulation s-vid** *vlan-id* [**only-tagged**]	三者选其一 缺省情况下，未配置任何报文匹配规则
	encapsulation s-vid *vlan-id* **c-vid** *vlan-id*	
	encapsulation { **default** | **tagged** | **untagged** }	
将以太网服务实例与VSI关联，（可选）并指定接入模式	**xconnect vsi** *vsi-name* [**access-mode** { **ethernet** | **vlan** }]	缺省情况下，以太网服务实例没有与VSI关联 配置本功能时，如果不指定**access-mode**参数，将采用缺省接入模式VLAN

3.6 管理本地和远端MAC地址

本地 MAC 地址只能动态学习，不能静态配置。在动态添加、删除本地 MAC 地址时，可以记录日志信息。

远端 MAC 地址表项可以静态添加，也可以根据接收到的 VXLAN 报文内封装的源 MAC 地址自动学习。

3.6.1 配置本端MAC地址添加/删除的日志功能

执行本配置后，VXLAN 添加、删除本地 MAC 地址时，将产生日志信息。生成的日志信息将被发送到设备的信息中心，通过设置信息中心的参数，决定日志信息的输出规则（即是否允许输出以及输出方向）。

表3-6 配置本端 MAC 地址添加/删除的日志功能

操作	命令	说明
进入系统视图	**system-view**	-
开启VXLAN本地MAC地址添加/删除的日志功能	**vxlan local-mac report**	缺省情况下，VXLAN添加/删除本地MAC地址时不会记录日志信息

3.6.2 添加静态远端MAC地址

表3-7 添加静态远端 MAC 地址

操作	命令	说明
进入系统视图	**system-view**	-

操作	命令	说明
添加静态远端MAC地址表项	**mac-address static** *mac-address* **interface tunnel** *tunnel-number* **vsi** *vsi-name*	缺省情况下，设备上不存在任何静态的远端MAC地址表项 **interface tunnel** *tunnel-number*参数指定的隧道接口必须与**vsi** *vsi-name*参数指定的VSI对应的VXLAN关联，且该VXLAN必须已经创建，否则配置将失败

3.6.3 开启远端MAC地址自动学习功能

缺省情况下，设备可以自动学习远端 MAC 地址。如果网络中存在攻击，为了避免学习到错误的远端 MAC 地址，也可以手工关闭远端 MAC 地址自动学习功能。

表3-8 开启远端 MAC 地址自动学习功能

操作	命令	说明
进入系统视图	**system-view**	-
开启远端MAC地址自动学习功能	**undo vxlan tunnel mac-learning disable**	缺省情况下，远端MAC地址自动学习功能处于开启状态

3.7 配置VXLAN组播路由泛洪方式

配置 VXLAN 组播路由泛洪方式时，需要完成以下配置任务：

- 在 VTEP 和核心设备上使能 IP 组播路由功能。
- 在核心设备上配置 IGMP 和组播路由协议。由于 VTEP 同时作为组播源和组播接收者，因此推荐使用双向 PIM 作为组播路由协议。

表3-9 配置 VXLAN 组播路由泛洪方式

操作	命令	说明
进入系统视图	**system-view**	-
进入VSI视图	**vsi** *vsi-name*	-
进入VXLAN视图	**vxlan** *vxlan-id*	-
配置VXLAN泛洪的组播地址和组播报文的源IP地址	**group** *group-address* **source** *source-address*	缺省情况下，未指定VXLAN泛洪的组播地址和组播报文的源IP地址，VXLAN采用单播路由方式泛洪 执行本命令后，VTEP将加入指定的组播组。同一VXLAN的所有VTEP都要加入相同的组播组 为确保报文转发正常，VXLAN组播报文的源IP地址（*source-address*）应指定为当前设备配置的处于up状态的VXLAN隧道的源端地址；如果当前设备配置了多个VXLAN隧道，要求这些隧道的源端地址配置相同

操作	命令	说明
进入组播报文源IP地址所在接口的接口视图	**interface** *interface-type interface-number*	组播报文源IP地址是指通过**group**命令中的**source**参数指定的地址
在接口上使能IGMP协议的主机功能	**igmp host enable**	缺省情况下，接口上IGMP协议的主机功能处于关闭状态 执行本命令后，当前接口将作为IGMP主机，即从该接口收到IGMP查询报文后，通过该接口发送组播组的报告报文，以便接收该组播组的报文 只有通过**multicast routing**命令使能IP组播路由后，本命令才会生效

3.8 配置VSI泛洪抑制

缺省情况下，VTEP 从本地站点内接收到目的 MAC 地址未知的单播数据帧后，会在该 VXLAN 内除接收接口外的所有本地接口和 VXLAN 隧道上泛洪该数据帧，将该数据帧发送给 VXLAN 内的所有站点。如果用户希望把该类数据帧限制在本地站点内，不通过 VXLAN 隧道将其转发到远端站点，则可以通过本命令手工禁止 VXLAN 对应 VSI 的泛洪功能。

禁止泛洪功能后，为了将某些 MAC 地址的数据帧泛洪到远端站点以保证某些业务的流量在站点间互通，可以配置选择性泛洪的 MAC 地址，当数据帧的目的 MAC 地址匹配该 MAC 地址时，该数据帧可以泛洪到远端站点。

表3-10 配置 VSI 泛洪抑制

操作	命令	说明
进入系统视图	**system-view**	-
进入VSI视图	**vsi** *vsi-name*	-
关闭VSI的泛洪功能	**flooding disable**	缺省情况下，VSI泛洪功能处于开启状态
（可选）配置VSI选择性泛洪的MAC地址	**selective-flooding mac-address** *mac-address*	缺省情况下，设备上不存在任何VSI选择性泛洪MAC地址 如果用户只希望某些目的MAC地址的报文可以泛洪到其他站点，可以先通过**flooding disable**命令关闭泛洪功能，再通过本命令配置选择性泛洪的MAC地址

3.9 配置VXLAN报文的目的UDP端口号

属于同一个 VXLAN 的 VTEP 设备上需要配置相同的 UDP 端口号。

表3-11 配置 VXLAN 报文的目的 UDP 端口号

操作	命令	说明
进入系统视图	**system-view**	-
配置VXLAN报文的目的UDP端口号	**vxlan udp-port** *port-number*	缺省情况下，VXLAN报文的目的

操作	命令	说明
		UDP端口号为4789

3.10 配置VXLAN报文检查功能

通过本配置可以实现对接收到的 VXLAN 报文内层封装的以太网数据帧是否携带 VLAN tag 进行检查：VTEP 接收到 VXLAN 报文并对其解封装后，若内层以太网数据帧带有 VLAN tag，则丢弃该 VXLAN 报文。

需要注意的是：远端 VTEP 上通过 **xconnect vsi** 命令的 **access-mode** 参数配置接入模式为 **ethernet** 时，VXLAN 报文可能携带 VLAN tag。这种情况下建议不要在本端 VTEP 上执行 **vxlan invalid-vlan-tag discard** 命令，以免错误地丢弃报文。

表3-12 配置 VXLAN 报文检查功能

操作	命令	说明
进入系统视图	**system-view**	-
配置丢弃内层数据帧含有VLAN tag的VXLAN报文	**vxlan invalid-vlan-tag discard**	缺省情况下，不会检查VXLAN报文内层封装的以太网数据帧是否携带VLAN tag

3.11 配置ARP泛洪抑制

配置 ARP 泛洪抑制时需要注意：

- 当同时执行 **flooding disable** 命令关闭了 VSI 的泛洪功能时：
  - 如果要与远端站点互通，则两端 VTEP 都需要为对端站点添加静态远端 MAC 地址表项（相关命令为 **mac-address static**）。
  - 建议通过 **mac-address timer** 命令配置动态 MAC 地址的老化时间大于 25 分钟（ARP 泛洪抑制表项的老化时间），以免 MAC 地址在 ARP 泛洪抑制表项老化之前老化，产生黑洞 MAC 地址。
- 当 VXLAN 网络采用组播路由（核心复制）方式转发泛洪流量时：
  - 若需要使用 ARP 泛洪抑制功能，必须保证所有 VTEP 设备均开启 ARP 泛洪抑制功能；
  - 如果需要和其他厂商的 VTEP 设备互通，则不能使用 ARP 泛洪抑制功能。

表3-13 配置 ARP 泛洪抑制

操作	命令	说明
进入系统视图	**system-view**	-
进入VSI视图	**vsi** *vsi-name*	-
开启ARP泛洪抑制功能	**arp suppression enable**	缺省情况下，ARP泛洪抑制功能处于关闭状态

3.12 VXLAN显示和维护

在完成上述配置后，在任意视图下执行 **display** 命令可以显示配置后 VXLAN 的运行情况，通过查看显示信息验证配置的效果。

在用户视图下，用户可以执行 **reset** 命令来清除 VXLAN 的相关信息。

表3-14 VXLAN 显示和维护

操作	命令
显示VSI的ARP泛洪抑制表项信息（独立运行模式）	**display arp suppression vsi** [**name** *vsi-name*] [**slot** *slot-number*] [**count**]
显示VSI的ARP泛洪抑制表项信息（IRF模式）	**display arp suppression vsi** [**name** *vsi-name*] [**chassis** *chassis-number* **slot** *slot-number*] [**count**]
显示VSI的MAC地址表信息	**display l2vpn mac-address** [**vsi** *vsi-name*] [**dynamic**] [**count**]
显示以太网服务实例的信息	**display l2vpn service-instance** [**interface** *interface-type interface-number* [**service-instance** *instance-id*]] [**verbose**]
显示VSI的信息	**display l2vpn vsi** [**name** *vsi-name*] [**verbose**]
显示IGMP执行主机行为的所有组播组信息	**display igmp host group** [*group-address* | **interface** *interface-type interface-number*] [**verbose**]
显示Tunnel接口信息	**display interface** [**tunnel** [*number*]] [**brief** [**description** | **down**]]
显示VXLAN关联的VXLAN隧道信息	**display vxlan tunnel** [**vxlan-id** *vxlan-id*]
清除VSI的ARP泛洪抑制表项	**reset arp suppression vsi** [**name** *vsi-name*]
清除VSI动态学习的MAC地址表项	**reset l2vpn mac-address** [**vsi** *vsi-name*]

说明

display interface tunnel 命令的详细介绍，请参见“三层技术-IP 业务命令参考”中的“隧道”。

3.13 VXLAN典型配置举例

3.13.1 VXLAN头端复制配置举例

1. 组网需求

Switch A、Switch B、Switch C 为与服务器连接的 VTEP 设备。虚拟机 VM 1、VM 2 和 VM 3 同属于 VXLAN 10。通过 VXLAN 实现不同站点间的二层互联，确保虚拟机在站点之间进行迁移时用户的访问流量不会中断。

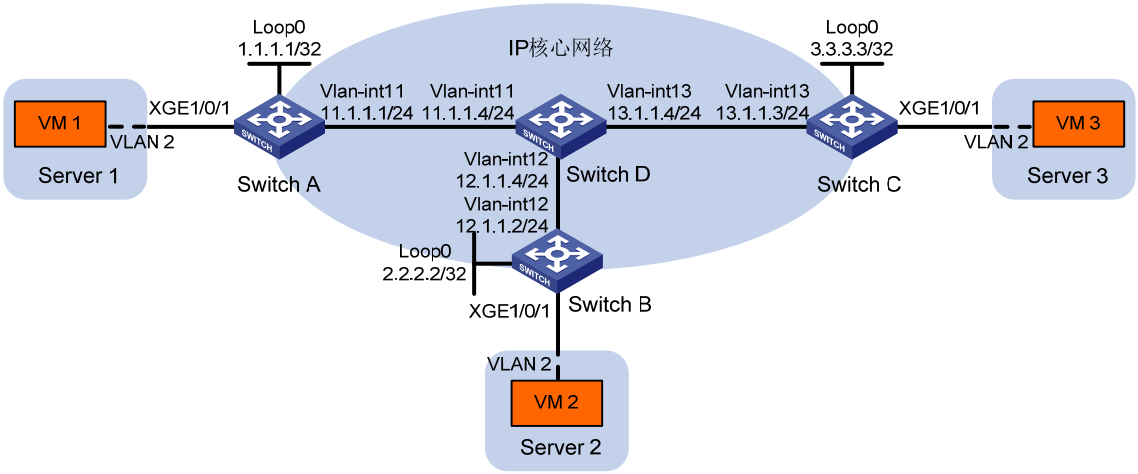
具体需求为：

- 不同 VTEP 之间手工建立 VXLAN 隧道。

- 手工关联 VXLAN 和 VXLAN 隧道。
- 通过源 MAC 地址动态学习远端 MAC 地址表项。
- 站点之间的泛洪流量采用头端复制的方式转发。

2. 组网图

图3-1 VXLAN 头端复制组网图



3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 3-1 配置各接口的IP地址和子网掩码，并在IP核心网络内配置OSPF协议，具体配置过程略。

(2) 配置 Switch A

开启 L2VPN 能力。

```
<SwitchA> system-view
[SwitchA] l2vpn enable
```

创建 VSI 实例 vpna 和 VXLAN 10。

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan10] quit
[SwitchA-vsi-vpna] quit
```

配置接口 Loopback0 的 IP 地址，作为隧道的源端地址。

```
[SwitchA] interface loopback0
[SwitchA-Loopback0] ip address 1.1.1.1 255.255.255.255
[SwitchA-Loopback0] quit
```

在 Switch A 和 Switch B 之间建立 VXLAN 隧道：

- 创建模式为 VXLAN 的隧道接口 Tunnel1
- 指定隧道的源端地址为本地接口 Loopback0 的地址 1.1.1.1
- 指定隧道的目的端地址为 Switch B 上接口 Loopback0 的地址 2.2.2.2。

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 1.1.1.1
[SwitchA-Tunnel1] destination 2.2.2.2
```

```
[SwitchA-Tunnel1] quit
```
在 Switch A 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 1.1.1.1
[SwitchA-Tunnel2] destination 3.3.3.3
[SwitchA-Tunnel2] quit
```
配置 Tunnel1 和 Tunnel2 与 VXLAN 10 关联。
```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan10] tunnel 1
[SwitchA-vsi-vpna-vxlan10] tunnel 2
[SwitchA-vsi-vpna-vxlan10] quit
[SwitchA-vsi-vpna] quit
```
创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。
```
[SwitchA] vlan 2
[SwitchA-vlan2] quit
[SwitchA] interface ten-gigabitethernet 1/0/1
[SwitchA-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchA-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```
在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。
```
[SwitchA-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```
配置以太网服务实例 1000 与 VSI 实例 vpna 关联。
```
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-Ten-GigabitEthernet1/0/1] quit
```
(3) 配置 Switch B

开启 L2VPN 能力。
```
<SwitchB> system-view
[SwitchB] l2vpn enable
```
创建 VSI 实例 vpna 和 VXLAN 10。
```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan10] quit
[SwitchB-vsi-vpna] quit
```
配置接口 Loopback0 的 IP 地址，作为隧道的源端地址。
```
[SwitchB] interface loopback0
[SwitchB-Loopback0] ip address 2.2.2.2 255.255.255.255
[SwitchB-Loopback0] quit
```
在 Switch A 和 Switch B 之间建立 VXLAN 隧道。
```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 2.2.2.2
[SwitchB-Tunnel2] destination 1.1.1.1
[SwitchB-Tunnel2] quit
```
在 Switch B 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 2.2.2.2
```

```
[SwitchB-Tunnel3] destination 3.3.3.3
[SwitchB-Tunnel3] quit
```
配置 Tunnel2 和 Tunnel3 与 VXLAN10 关联。
```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan10] tunnel 2
[SwitchB-vsi-vpna-vxlan10] tunnel 3
[SwitchB-vsi-vpna-vxlan10] quit
[SwitchB-vsi-vpna] quit
```
创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。
```
[SwitchB] vlan 2
[SwitchB-vlan2] quit
[SwitchB] interface ten-gigabitethernet 1/0/1
[SwitchB-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchB-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```
在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。
```
[SwitchB-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```
配置以太网服务实例 1000 与 VSI 实例 vpna 关联。
```
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-Ten-GigabitEthernet1/0/1] quit
```
(4)　配置 Switch C

开启 L2VPN 能力。
```
<SwitchC> system-view
[SwitchC] l2vpn enable
```
创建 VSI 实例 vpna 和 VXLAN 10。
```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan10] quit
[SwitchC-vsi-vpna] quit
```
配置接口 Loopback0 的 IP 地址，作为隧道的源端地址。
```
[SwitchC] interface loopback0
[SwitchC-Loopback0] ip address 3.3.3.3 255.255.255.255
[SwitchC-Loopback0] quit
```
在 Switch A 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 3.3.3.3
[SwitchC-Tunnel1] destination 1.1.1.1
[SwitchC-Tunnel1] quit
```
在 Switch B 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 3.3.3.3
[SwitchC-Tunnel3] destination 2.2.2.2
[SwitchC-Tunnel3] quit
```
配置 Tunnel1 和 Tunnel3 与 VXLAN 10 关联。
```
[SwitchC] vsi vpna
```

```
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan10] tunnel 1
[SwitchC-vsi-vpna-vxlan10] tunnel 3
[SwitchC-vsi-vpna-vxlan10] quit
[SwitchC-vsi-vpna] quit
```

创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。

```
[SwitchC] vlan 2
[SwitchC-vlan2] quit
[SwitchC] interface ten-gigabitethernet 1/0/1
[SwitchC-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchC-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```

在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。

```
[SwitchC-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

配置以太网服务实例 1000 与 VSI 实例 vpna 关联。

```
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] quit
[SwitchC-Ten-GigabitEthernet1/0/1] quit
```

4. 验证配置

(1) 验证 VTEP 设备（下文以 Switch A 为例，其他设备验证方法与此类似）

查看 Switch A 上的 Tunnel 接口信息，可以看到 VXLAN 模式的 Tunnel 接口处于 up 状态。

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64kbps
Maximum Transmit Unit: 64000
Internet protocol processing: disabled
Last clearing of counters: Never
Tunnel source 1.1.1.1, destination 2.2.2.2
Tunnel protocol/transport UDP_VXLAN/IP
```

查看 Switch A 上的 VSI 信息，可以看到 VSI 内创建的 VXLAN、与 VXLAN 关联的 VXLAN 隧道、与 VSI 关联的以太网服务实例等信息。

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  Drop Unknown            : -
```

```
  Flooding                : Enabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name           Link ID    State  Type
    Tunnel1               0x5000001  Up     Manual
    Tunnel2               0x5000002  Up     Manual
  ACs:
    AC                              Link ID    State
    XGE1/0/1 srv1000                0          Up
```

查看 Switch A 上 VSI 的 MAC 地址表项信息，可以看到已学习到的 MAC 地址信息。

```
<SwitchA> display l2vpn mac-address
MAC Address      State    VSI Name                        Link ID/Name  Aging
cc3e-5f9c-6cdb   Dynamic  vpna                            Tunnel1       Aging
cc3e-5f9c-23dc   Dynamic  vpna                            Tunnel2       Aging
--- 2 mac address(es) found  ---
```

(2) 验证主机

虚拟机 VM 1、VM 2、VM 3 之间可以互访。

3.13.2 VXLAN核心复制配置举例

1. 组网需求

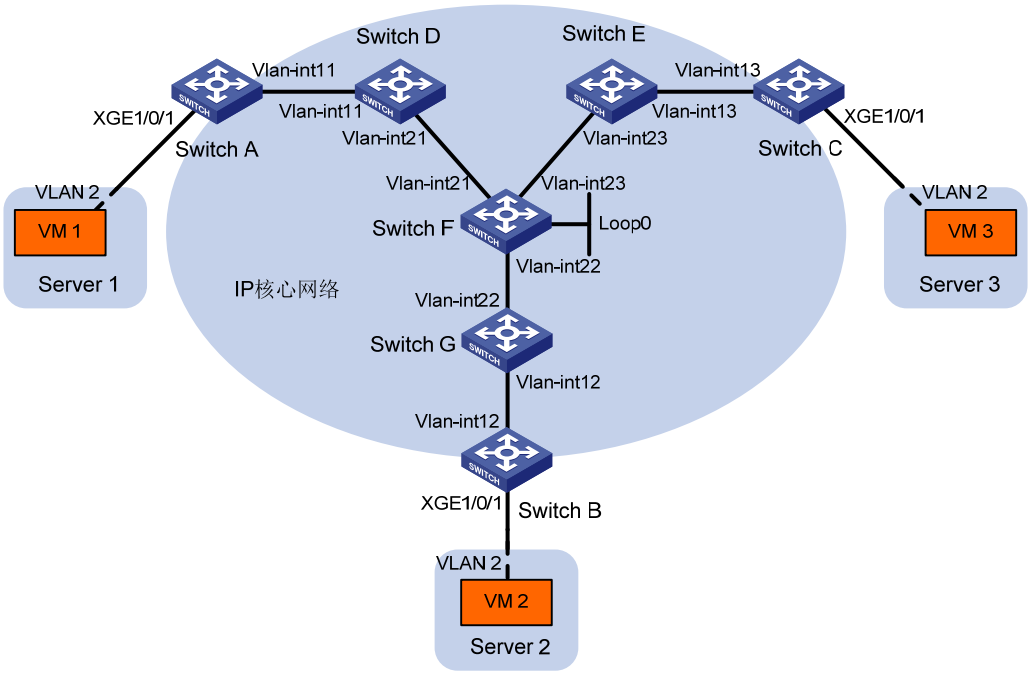
Switch A、Switch B、Switch C 为与服务器连接的 VTEP 设备。虚拟机 VM 1、VM 2 和 VM 3 同属于 VXLAN 10。通过 VXLAN 实现不同站点间的二层互联，确保虚拟机在站点之间进行迁移时用户的访问流量不会中断。

具体需求为：

- 不同 VTEP 之间手工建立 VXLAN 隧道。
- 手工关联 VXLAN 和 VXLAN 隧道。
- 通过源 MAC 地址动态学习远端 MAC 地址表项。
- 站点之间的泛洪流量采用核心复制的方式转发。

2. 组网图

图3-2 VXLAN 核心复制组网图



设备	接口	IP地址	设备	接口	IP地址
Switch A	Vlan-int11	11.1.1.1/24	Switch C	Vlan-int13	13.1.1.3/24
Switch D	Vlan-int11	11.1.1.4/24	Switch E	Vlan-int13	13.1.1.5/24
	Vlan-int21	21.1.1.4/24		Vlan-int23	23.1.1.5/24
Switch F	Vlan-int21	21.1.1.6/24	Switch G	Vlan-int12	12.1.1.7/24
	Vlan-int22	22.1.1.6/24		Vlan-int22	22.1.1.7/24
	Vlan-int23	23.1.1.6/24	Switch B	Vlan-int12	12.1.1.2/24
	Loop0	6.6.6.6/32			

3. 配置步骤

(1) 配置 IP 地址和单播路由协议

请按照 图 3-2 配置各接口的IP地址和子网掩码，并在IP核心网络内配置OSPF协议，具体配置过程略。

(2) 配置 Switch A

开启 L2VPN 能力。

```
<SwitchA> system-view
[SwitchA] l2vpn enable
```

使能 IP 组播路由。

```
[SwitchA] multicast routing
[SwitchA-mrib] quit
```

创建 VSI 实例 vpna 和 VXLAN 10。

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan10] quit
[SwitchA-vsi-vpna] quit
```

配置接口 Vlan-interface11 的 IP 地址，并在该接口上使能 IGMP 协议的主机功能。

```
[SwitchA] interface vlan-interface 11
[SwitchA-Vlan-interface11] ip address 11.1.1.1 24
[SwitchA-Vlan-interface11] igmp host enable
[SwitchA-Vlan-interface11] quit
```

在 Switch A 和 Switch B 之间建立 VXLAN 隧道：

- 创建模式为 VXLAN 的隧道接口 Tunnel1
- 指定隧道的源端地址为本地接口 Vlan-interface11 的地址 11.1.1.1
- 指定隧道的目的端地址为 Switch B 上接口 Vlan-interface12 的地址 12.1.1.2

```
[SwitchA] interface tunnel 1 mode vxlan
[SwitchA-Tunnel1] source 11.1.1.1
[SwitchA-Tunnel1] destination 12.1.1.2
[SwitchA-Tunnel1] quit
```

在 Switch A 和 Switch C 之间建立 VXLAN 隧道。

```
[SwitchA] interface tunnel 2 mode vxlan
[SwitchA-Tunnel2] source 11.1.1.1
[SwitchA-Tunnel2] destination 13.1.1.3
[SwitchA-Tunnel2] quit
```

配置 Tunnel1 和 Tunnel2 与 VXLAN 10 关联。

```
[SwitchA] vsi vpna
[SwitchA-vsi-vpna] vxlan 10
[SwitchA-vsi-vpna-vxlan10] tunnel 1
[SwitchA-vsi-vpna-vxlan10] tunnel 2
```

配置 VXLAN 泛洪的组播地址为 225.1.1.1，组播报文的源 IP 地址为 11.1.1.1。

```
[SwitchA-vsi-vpna-vxlan10] group 225.1.1.1 source 11.1.1.1
[SwitchA-vsi-vpna-vxlan10] quit
[SwitchA-vsi-vpna] quit
```

创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。

```
[SwitchA] vlan 2
[SwitchA-vlan2] quit
[SwitchA] interface ten-gigabitethernet 1/0/1
[SwitchA-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchA-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```

在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。

```
[SwitchA-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```

配置以太网服务实例 1000 与 VSI 实例 vpna 关联。

```
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchA-Ten-GigabitEthernet1/0/1-srv1000] quit
[SwitchA-Ten-GigabitEthernet1/0/1] quit
```

(3) 配置 Switch B

开启 L2VPN 能力。

```
<SwitchB> system-view
[SwitchB] l2vpn enable
```

使能 IP 组播路由。

```
[SwitchB] multicast routing
```

```
[SwitchB-mrib] quit
```
创建 VSI 实例 vpna 和 VXLAN 10。
```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan10] quit
[SwitchB-vsi-vpna] quit
```
配置接口 Vlan-interface12 的 IP 地址，并在该接口上使能 IGMP 协议的主机功能。
```
[SwitchB] interface vlan-interface 12
[SwitchB-Vlan-interface12] ip address 12.1.1.2 24
[SwitchB-Vlan-interface12] igmp host enable
[SwitchB-Vlan-interface12] quit
```
在 Switch A 和 Switch B 之间建立 VXLAN 隧道。
```
[SwitchB] interface tunnel 2 mode vxlan
[SwitchB-Tunnel2] source 12.1.1.2
[SwitchB-Tunnel2] destination 11.1.1.1
[SwitchB-Tunnel2] quit
```
在 Switch B 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchB] interface tunnel 3 mode vxlan
[SwitchB-Tunnel3] source 12.1.1.2
[SwitchB-Tunnel3] destination 13.1.1.3
[SwitchB-Tunnel3] quit
```
配置 Tunnel2 和 Tunnel3 与 VXLAN10 关联。
```
[SwitchB] vsi vpna
[SwitchB-vsi-vpna] vxlan 10
[SwitchB-vsi-vpna-vxlan10] tunnel 2
[SwitchB-vsi-vpna-vxlan10] tunnel 3
```
配置 VXLAN 泛洪的组播地址为 225.1.1.1，组播报文的源 IP 地址为 12.1.1.2。
```
[SwitchB-vsi-vpna-vxlan10] group 225.1.1.1 source 12.1.1.2
[SwitchB-vsi-vpna-vxlan10] quit
[SwitchB-vsi-vpna] quit
```
创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。
```
[SwitchB] vlan 2
[SwitchB-vlan2] quit
[SwitchB] interface ten-gigabitethernet 1/0/1
[SwitchB-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchB-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```
在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。
```
[SwitchB-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```
配置以太网服务实例 1000 与 VSI 实例 vpna 关联。
```
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchB-Ten-GigabitEthernet1/0/1-srv1000] quit
[SwitchB-Ten-GigabitEthernet1/0/1] quit
```
(4) 配置 Switch C

开启 L2VPN 能力。

```
<SwitchC> system-view
[SwitchC] l2vpn enable
```
使能 IP 组播路由。
```
[SwitchC] multicast routing
[SwitchC-mrib] quit
```
创建 VSI 实例 vpna 和 VXLAN 10。
```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan10] quit
[SwitchC-vsi-vpna] quit
```
配置接口 Vlan-interface13 的 IP 地址，并在该接口上使能 IGMP 协议的主机功能。
```
[SwitchC] interface vlan-interface 13
[SwitchC-Vlan-interface13] ip address 13.1.1.3 24
[SwitchC-Vlan-interface13] igmp host enable
[SwitchC-Vlan-interface13] quit
```
在 Switch A 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchC] interface tunnel 1 mode vxlan
[SwitchC-Tunnel1] source 13.1.1.3
[SwitchC-Tunnel1] destination 11.1.1.1
[SwitchC-Tunnel1] quit
```
在 Switch B 和 Switch C 之间建立 VXLAN 隧道。
```
[SwitchC] interface tunnel 3 mode vxlan
[SwitchC-Tunnel3] source 13.1.1.3
[SwitchC-Tunnel3] destination 12.1.1.2
[SwitchC-Tunnel3] quit
```
配置 Tunnel1 和 Tunnel3 与 VXLAN 10 关联。
```
[SwitchC] vsi vpna
[SwitchC-vsi-vpna] vxlan 10
[SwitchC-vsi-vpna-vxlan10] tunnel 1
[SwitchC-vsi-vpna-vxlan10] tunnel 3
```
配置 VXLAN 泛洪的组播地址为 225.1.1.1，组播报文的源 IP 地址为 13.1.1.3。
```
[SwitchC-vsi-vpna-vxlan10] group 225.1.1.1 source 13.1.1.3
[SwitchC-vsi-vpna-vxlan10] quit
[SwitchC-vsi-vpna] quit
```
创建 VLAN 2，并配置接入服务器的接口允许该 VLAN 通过。
```
[SwitchC] vlan 2
[SwitchC-vlan2] quit
[SwitchC] interface ten-gigabitethernet 1/0/1
[SwitchC-Ten-GigabitEthernet1/0/1] port link-type trunk
[SwitchC-Ten-GigabitEthernet1/0/1] port trunk permit vlan 2
```
在接入服务器的接口上创建以太网服务实例 1000，该实例用来匹配 VLAN 2 的数据帧。
```
[SwitchC-Ten-GigabitEthernet1/0/1] service-instance 1000
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] encapsulation s-vid 2
```
配置以太网服务实例 1000 与 VSI 实例 vpna 关联。
```
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] xconnect vsi vpna
[SwitchC-Ten-GigabitEthernet1/0/1-srv1000] quit
```

```
[SwitchC-Ten-GigabitEthernet1/0/1] quit
```

(5) 配置 Switch D

使能 IP 组播路由。

```
<SwitchD> system-view
[SwitchD] multicast routing
[SwitchD-mrib] quit
```

在接口 Vlan-interface11 上使能 IGMP 和 PIM-SM。

```
[SwitchD] interface vlan-interface 11
[SwitchD-Vlan-interface11] igmp enable
[SwitchD-Vlan-interface11] pim sm
[SwitchD-Vlan-interface11] quit
```

在接口 Vlan-interface21 上使能 PIM-SM。

```
[SwitchD] interface vlan-interface 21
[SwitchD-Vlan-interface21] pim sm
[SwitchD-Vlan-interface21] quit
```

使能双向 PIM。

```
[SwitchD] pim
[SwitchD-pim] bidir-pim enable
[SwitchD-pim] quit
```

(6) 配置 Switch E

使能 IP 组播路由。

```
<SwitchE> system-view
[SwitchE] multicast routing
[SwitchE-mrib] quit
```

在接口 Vlan-interface13 上使能 IGMP 和 PIM-SM。

```
[SwitchE] interface vlan-interface 13
[SwitchE-Vlan-interface13] igmp enable
[SwitchD-Vlan-interface13] pim sm
[SwitchE-Vlan-interface13] quit
```

在接口 Vlan-interface23 上使能 PIM-SM。

```
[SwitchE] interface vlan-interface 23
[SwitchE-Vlan-interface23] pim sm
[SwitchE-Vlan-interface23] quit
```

使能双向 PIM。

```
[SwitchE] pim
[SwitchE-pim] bidir-pim enable
[SwitchE-pim] quit
```

(7) 配置 Switch F

使能 IP 组播路由。

```
<SwitchF> system-view
[SwitchF] multicast routing
[SwitchF-mrib] quit
```

在各接口上使能 PIM-SM。

```
[SwitchF] interface vlan-interface 21
[SwitchF-Vlan-interface21] pim sm
```

```
[SwitchF-Vlan-interface21] quit
[SwitchF] interface vlan-interface 22
[SwitchF-Vlan-interface22] pim sm
[SwitchF-Vlan-interface22] quit
[SwitchF] interface vlan-interface 23
[SwitchF-Vlan-interface23] pim sm
[SwitchF-Vlan-interface23] quit
```

使能双向 PIM。

```
[SwitchF] pim
[SwitchF-pim] bidir-pim enable
```

将接口 Vlan-interface22 配置为 C-BSR，并将接口 Loopback0 配置为服务于双向 PIM 的 C-RP。

```
[SwitchF-pim] c-bsr 22.1.1.6
[SwitchF-pim] c-rp 6.6.6.6 bidir
[SwitchF-pim] quit
```

(8) 配置 Switch G

使能 IP 组播路由。

```
<SwitchG> system-view
[SwitchG] multicast routing
[SwitchG-mrib] quit
```

在接口 Vlan-interface12 上使能 IGMP 和 PIM-SM。

```
[SwitchG] interface vlan-interface 12
[SwitchG-Vlan-interface12] igmp enable
[SwitchD-Vlan-interface12] pim sm
[SwitchG-Vlan-interface12] quit
```

在接口 Vlan-interface22 上使能 PIM-SM。

```
[SwitchG] interface vlan-interface 22
[SwitchG-Vlan-interface22] pim sm
[SwitchG-Vlan-interface22] quit
```

使能双向 PIM。

```
[SwitchG] pim
[SwitchG-pim] bidir-pim enable
[SwitchG-pim] quit
```

4. 验证配置

(1) 验证 VTEP 设备（下文以 Switch A 为例，其他设备验证方法与此类似）

查看 Switch A 上的 Tunnel 接口信息，可以看到 VXLAN 模式的 Tunnel 接口处于 up 状态。

```
[SwitchA] display interface tunnel 1
Tunnel1
Current state: UP
Line protocol state: UP
Description: Tunnel1 Interface
Bandwidth: 64kbps
Maximum Transmit Unit: 64000
Internet protocol processing: disabled
Last clearing of counters: Never
Tunnel source 11.1.1.1, destination 12.1.1.2
```

```
Tunnel protocol/transport UDP_VXLAN/IP
```

查看 Switch A 上的 VSI 信息，可以看到 VSI 内创建的 VXLAN、与 VXLAN 关联的 VXLAN 隧道、与 VSI 关联的以太网服务实例等信息。

```
[SwitchA] display l2vpn vsi verbose
VSI Name: vpna
  VSI Index               : 0
  VSI State               : Up
  MTU                     : 1500
  Bandwidth               : -
  Broadcast Restrain      : -
  Multicast Restrain      : -
  Unknown Unicast Restrain: -
  MAC Learning            : Enabled
  MAC Table Limit         : -
  Drop Unknown            : -
  Flooding                : Enabled
  VXLAN ID                : 10
  Tunnels:
    Tunnel Name          Link ID    State  Type
    Tunnel1              0x5000001  Up     Manual
    Tunnel2              0x5000002  Up     Manual
    MTunnel1             0x6000000  Up     Auto
  ACs:
    AC                              Link ID    State
    XGE1/0/1 srv1000                0          Up
```

查看 Switch A 上 VSI 的 MAC 地址表项信息，可以看到已学习到的 MAC 地址信息。

```
<SwitchA> display l2vpn mac-address
MAC Address      State    VSI Name                        Link ID/Name  Aging
cc3e-5f9c-6cdb   Dynamic  vpna                            Tunnel1   Aging
cc3e-5f9c-23dc   Dynamic  vpna                            Tunnel2   Aging
--- 2 mac address(es) found  ---
```

查看 Switch A 上 IGMP 执行主机行为的所有组播组信息，可以看到接口 Vlan-interface11 下存在组播组 225.1.1.1 的信息。

```
<SwitchA> display igmp host group
IGMP host groups in total: 1
 Vlan-interface11(11.1.1.1):
  IGMP host groups in total: 1
   Group address      Member state      Expires
   225.1.1.1          Idle              Off
```

(2) 验证主机

虚拟机 VM 1、VM 2、VM 3 之间可以互访。