VXLAN概述及使用场景

H3C交换机技术专题篇

VXLAN

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和园区接入网络。

哦哟这么厉害的？那到底什么是VXLAN？

不急，我们这就来瞧一瞧：

1、为什么要用VXLAN？

2、VXLAN实现原理

3、VXLAN使用场景

Why VXLAN?

随着数据中心规模和容量的不断扩大，当前的传统网络主要遇到了如下几方面的问题：

一、有限的VLAN数量

当前的虚拟局域网（VLAN），仅能提供4096个VLAN标记，不足以提供必要的可扩展性。

二、二层网络防环路困难

随着二层网络的不断扩大，网络设备二层防环压力增大，更为复杂。

三、三层网络边界限制

虚拟机的无缝迁移要求IP地址不能改变，因此在传统网络中，虚拟机的迁移受到三层网络边界限制。

总之就是蛮愁的……但是不用怕！我们可以使用VXLAN！

VXLAN技术完美解决了上述问题，它是一种在常用的网络和虚拟基础架构的顶层创建“浮动”虚拟域的方法，可以在现有网络上创建大量虚拟域，并且虚拟域彼此之间以及与底层网络之间完全隔离，优势超大：

一、灵活性大

VXLAN在标准三层IP网络上运行，不再需要构建和管理庞大的二层基础传输层网络。通过支持跨交换机和单元边界的“扩展集群”，数据中心服务器与存储的利用率和灵活性可实现最大化。

二、利于投资保护

VXLAN在标准交换机硬件上运行，交换机上无需进行软件升级，也不必采用特殊的代码版本。

三、支持大量的租户

使用24位的标识符，最多可支持2的24次方即16777216个VXLAN，使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。

四、易于维护

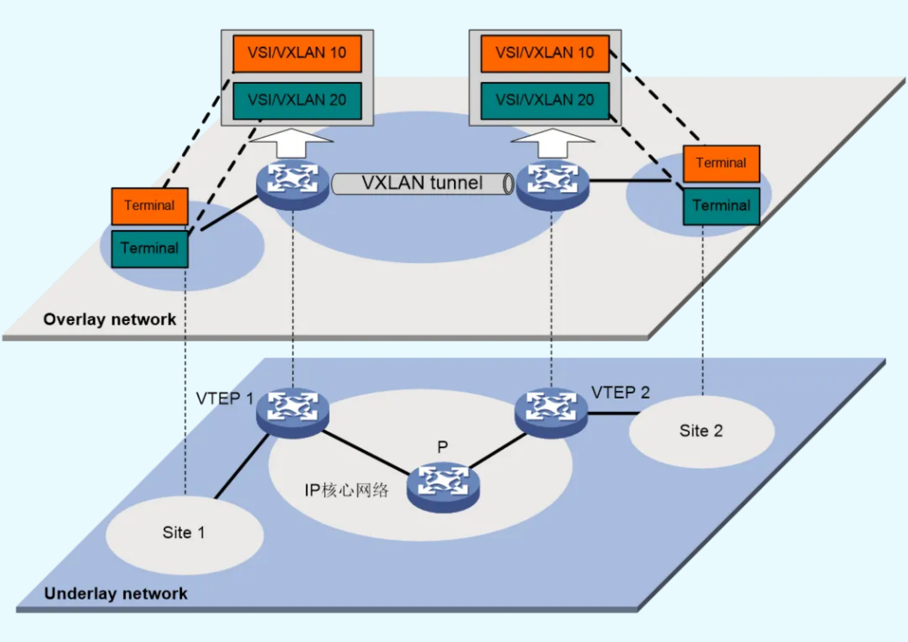
基于IP网络组建大二层网络，使得网络部署和维护更加容易，并且可以充分地利用现有的IP网络技术，例如利用等价路由进行负载分担等；只有IP核心网络的边缘设备需要进行VXLAN处理，网络中间设备只需根据IP头转发报文，降低了网络部署的难度和费用。

VXLAN实现原理

VXLAN网络模型

VXLAN技术将已有的三层物理网络作为Underlay网络，在其上构建出虚拟的二层网络，即Overlay网络。Overlay网络通过封装技术、利用Underlay网络提供的三层转发路径，实现租户二层报文跨越三层网络在不同站点间传递。对于租户来说，Underlay网络是透明的，同一租户的不同站点就像工作在一个局域网中。

VXLAN网络模型示意图：



用户终端设备可以是PC机、无线终端设备、服务器上创建的VM等，属于相同VXLAN的用户终端处于同一个逻辑二层网络，彼此之间二层互通；属于不同VXLAN的用户终端之间二层隔离。

VTEP是VXLAN的边缘设备，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VXLAN的相关处理都在VTEP上进行。

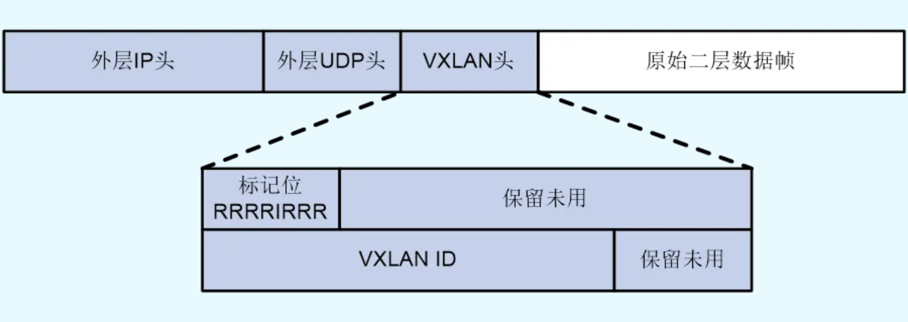
VXLAN隧道是两个VTEP之间的点到点逻辑隧道。VTEP为数据帧封装VXLAN头、UDP头和IP头后，通过VXLAN隧道将封装后的报文转发给远端VTEP，远端VTEP对其进行解封装。

VXLAN报文封装格式

VXLAN报文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：

1、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。

2、VXLAN ID：用来标识一个VXLAN网络，长度为24比特。



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。

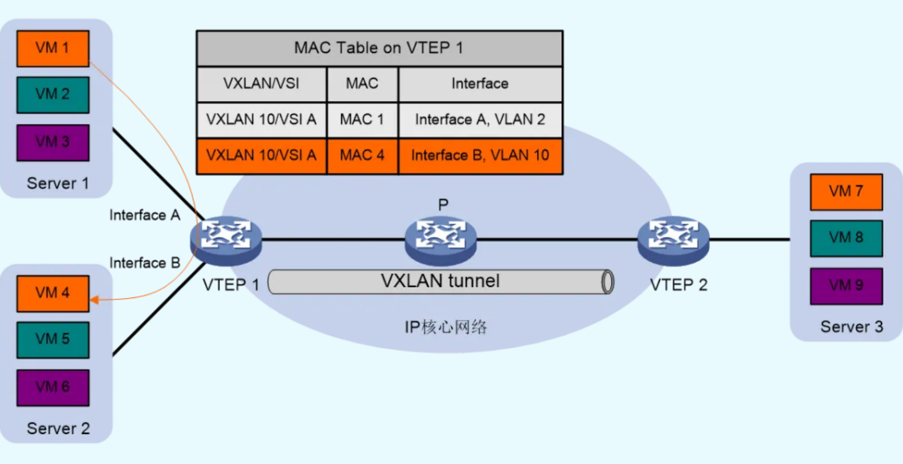
2、识别接收到的报文所属的VXLAN，以便将报文的源MAC地址学习到VXLAN对应的VSI，并在该VSI内转发该报文。

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。

VXLAN使用场景

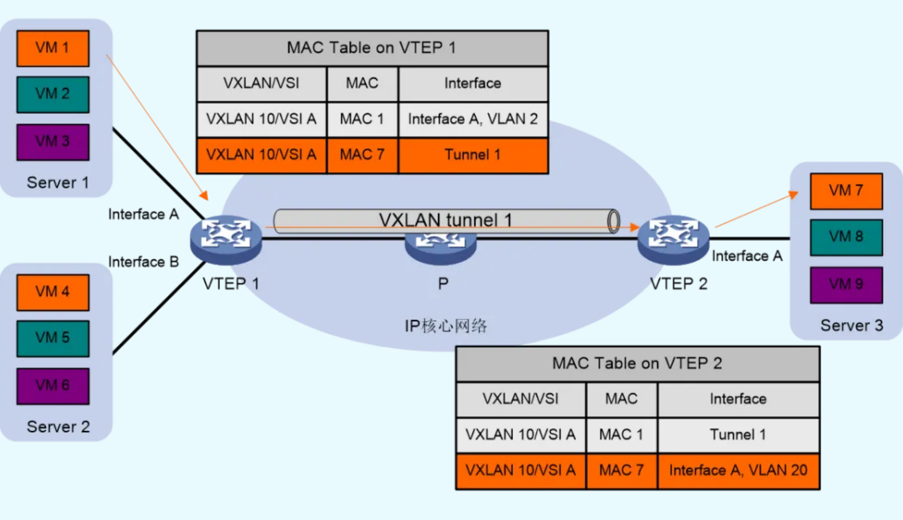
一、站点内单播流量的转发：



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。

如图所示，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Interface A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Interface B，所在VLAN为VLAN 10，则将以太网帧从接口Interface B的VLAN 10内发送给VM 4。

二、站点间单播流量的转发：



如图所示，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Interface A（所在VLAN为VLAN 2）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Interface A（所在VLAN为VLAN 20）。

6、VTEP 2从接口Interface A的VLAN 20内将数据帧发送给VM 7。

—— end ——