

CloudEngine 系列交换机

VXLAN 技术白皮书

文档版本 01

发布日期 2014-09-20



版权所有 © 华为技术有限公司 2014。 保留一切权利。

非经本公司书面许可,任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部,并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标,由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束,本文档中描述的全部或部分产品、服务或 特性可能不在您的购买或使用范围之内。除非合同另有约定,华为公司对本文档内容不做任何明示或默示的声 明或保证。

由于产品版本升级或其他原因,本文档内容会不定期进行更新。除非另有约定,本文档仅作为使用指导,本文 档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址: 深圳市龙岗区坂田华为总部办公楼 邮编: 518129

网址: http://enterprise.huawei.com

目 录

1 VXLAN 简介	1
2 原理描述	
2.1 基本概念	
2.2 报文格式	
2.3 隧道建立与维护	
2.4 数据报文转发	9
2.5 VXLAN QoS	15
3 应用场景	17
3.1 同网段终端用户通信的应用	18
3.2 不同网段终端用户通信的应用	19
3.3 在虚拟机迁移场景中的应用	20
4 基于 SDN 控制器的 VXLAN 配置示例	2 3
5 参考标准和协议	39

1 vxlan 简介

介绍VXLAN的定义、目的和受益。

定义

VXLAN(Virtual eXtensible Local Area Network)是VLAN扩展方案草案,采用MAC in UDP(User Datagram Protocol)封装方式,是NVo3(Network Virtualization over Layer 3)中的一种网络虚拟化技术。

目的

作为云计算的核心技术之一,服务器虚拟化凭借其大幅降低IT成本、提高业务部署灵活性、降低运维成本等优势已经得到越来越多的认可和部署。

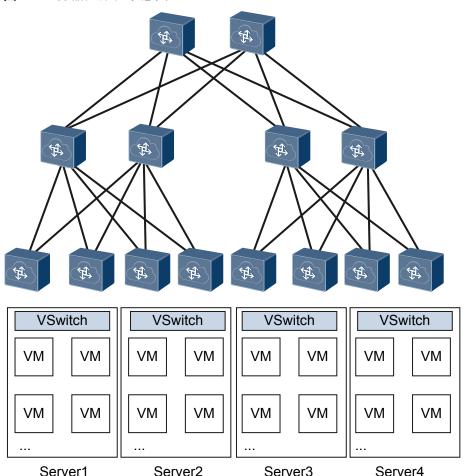


图 1-1 服务器虚拟化示意图

如**图1-1**所示,一台服务器可虚拟多台虚拟机,而一台虚拟机相当于一台主机。主机的数量发生了数量级的变化,这也为虚拟网络带来了如下问题:

● 虚拟机规模受网络规格限制

在大二层网络环境下,数据报文是通过查询MAC地址表进行二层转发,而MAC地址表的容量限制了虚拟机的数量。

● 网络隔离能力限制

当前主流的网络隔离技术是VLAN或VPN(Virtual Private Network),在大规模的虚拟化网络中部署存在如下限制:

- 由于IEEE 802.1Q中定义的VLAN Tag域只有12比特,仅能表示4096个VLAN,无 法满足大二层网络中标识大量用户群的需求。
- 传统二层网络中的VLAN/VPN无法满足网络动态调整的需求。

● 虚拟机迁移范围受网络架构限制

虚拟机启动后,可能由于服务器资源等问题(如CPU过高,内存不够等),需要将虚拟机迁移到新的服务器上。为了保证虚拟机迁移过程中业务不中断,则需要保证虚拟机的IP地址、MAC地址等参数保持不变,这就要求业务网络是一个二层网络,且要求网络本身具备多路径的冗余备份和可靠性。

针对大二层网络,VXLAN的提出很好地解决了上述问题。

针对虚拟机规模受网络规格限制

VXLAN将虚拟机发出的数据包封装在UDP中,并使用物理网络的IP/MAC地址作为外层头进行封装,对网络只表现为封装后的参数。因此,极大降低了大二层网络对MAC地址规格的需求。

● 针对网络隔离能力限制

VXLAN引入了类似VLAN ID的用户标识,称为VXLAN网络标识VNI(VXLAN Network ID),由24比特组成,支持多达16M(($2^{^24}$ -1)/ $1024^{^2}$)的VXLAN段,从而满足了大量的用户标识。

● 针对虚拟机迁移范围受网络架构限制 通过VXLAN构建大二层网络,保证了在虚拟迁移时虚拟机的IP地址、MAC地址等 参数保持不变。

受益

随着数据中心在物理网络基础设施上实施服务器虚拟化的快速发展,作为NVo3技术之一的VXLAN:

- 通过24比特的VNI可以支持多达16M的VXLAN段的网络隔离,对用户进行隔离和标识不再受到限制,可满足海量租户。
- 除VXLAN网络边缘设备,网络中的其他设备不需要识别虚拟机的MAC地址,减轻 了设备的MAC地址学习压力,提升了设备性能。
- 通过采用MAC in UDP封装来延伸二层网络,实现了物理网络和虚拟网络解耦,租户可以规划自己的虚拟网络,不需要考虑物理网络IP地址和广播域的限制,大大降低了网络管理的难度。

2 _{原理描述}

关于本章

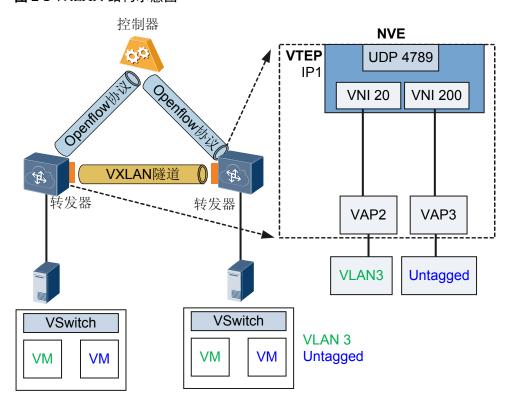
介绍VXLAN的实现原理。

- 2.1 基本概念
- 2.2 报文格式
- 2.3 隧道建立与维护
- 2.4 数据报文转发
- 2.5 VXLAN QoS

2.1 基本概念

VXLAN通过采用MAC in UDP封装来延伸二层网络,是大二层虚拟网络扩展的隧道封装技术。在大二层网络,为了方便控制与部署引入了SDN(Software Defined Network)控制器概念。控制器通过OpenFlow协议将信息下发给转发器,以实现控制器统一维护管理。

图 2-1 VXLAN 结构示意图



结合图2-1介绍控制器和VXLAN相关概念。

● 通过SDN控制器可简化网络操作维护,相关概念如表2-1所示。

表 2-1 控制器相关概念

概念	描述
控制器 (Controller)	控制器是OpenFlow协议的控制面服务器,所有的路径计算与管理都由独立的控制器完成。 通常,刀片服务器即可作为控制器。
转发器	OpenFlow协议的转发平面设备,只处理数据转发任务。
OpenFlow协议	OpenFlow协议是SDN中的重要协议,是控制器和转发器的通信通道。控制器通过OpenFlow协议将信息下发给转发器。

概念	描述
OpenFlow流表	OpenFlow流表用来指导转发器进行报文转发。转发器根据流表中的匹配项对报文进行匹配,匹配上则执行相应的动作。如,流表的匹配域是源MAC地址,动作是向某个端口转发,则转发器根据这个流表,对匹配这个MAC地址的报文进行向某个端口转发的动作。
	控制器通过下发流表直接控制转发器的转发行为,实现控制器 对转发器的控制。

● 在控制器上配置VXLAN,通过OpenFlow协议将VXLAN信息下发给转发器。 通过VXLAN,虚拟网络可接入大量租户,且租户可以规划自己的虚拟网络,不需 要考虑物理网络IP地址和广播域的限制,降低了网络管理的难度。相关概念如表2-2 所示。

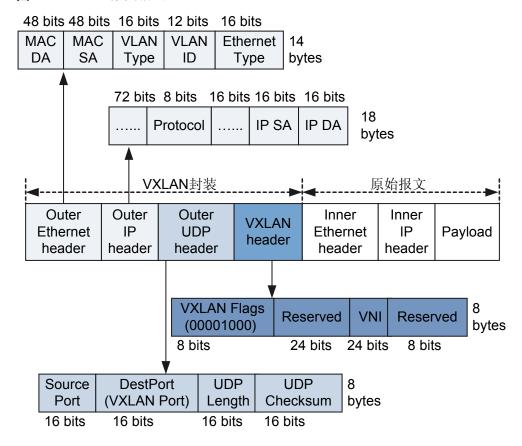
表 2-2 VXLAN 相关概念

概念	描述
NVE (Network Virtualization Edge)	网络虚拟边缘节点NVE,实现网络虚拟化功能的网络实体。
VAP (Virtual Access Point)	虚拟接入点VAP统一为二层子接口,用于接入数据报文。 为二层子接口配置不同的流封装,可实现不同的数据报文接入 不同的二层子接口。
VTEP (VXLAN Tunnel Endpoints)	VTEP是VXLAN隧道端点,封装在NVE中,用于VXLAN报文的封装和解封装。 VTEP与物理网络相连,分配有物理网络的IP地址,该地址与虚拟网络无关。 VXLAN报文中源IP地址为本节点的VTEP地址,VXLAN报文中目的IP地址为对端节点的VTEP地址,一对VTEP地址就对应着一个VXLAN隧道。
VNI (VXLAN Network Identifier)	VXLAN网络标识VNI类似VLAN ID,用于区分VXLAN段,不同VXLAN段的虚拟机不能直接二层相互通信。一个VNI表示一个租户,即使多个终端用户属于同一个VNI,也表示一个租户。VNI由24比特组成,支持多达16M((2^24-1)/1024^2)的租户。

2.2 报文格式

VXLAN是MAC in UDP的网络虚拟化技术,所以其报文封装是在原始以太报文之前添加了一个UDP封装及VXLAN头封装。具体报文格式如图2-2所示。

图 2-2 VXLAN 报文格式



● VXLAN头封装

- Flags: 8比特,取值为00001000。
- VNI: VXLAN网络标识,24比特,用于区分VXLAN段。
- Reserved: 24比特和8比特,必须设置为0。
- 外层UDP头封装

目的UDP端口号是4789。源端口号是内层以太报文头通过哈希算法计算后的值。

● 外层IP头封装

源IP地址为发送报文的虚拟机所属VTEP的IP地址;目的IP地址是目的虚拟机所属VTEP的IP地址。

- 外层Ethernet头封装
 - SA: 发送报文的虚拟机所属VTEP的MAC地址。
 - DA: 目的虚拟机所属VTEP上路由表中直连的下一跳MAC地址。
 - VLAN Type: 可选字段, 当报文中携带VLAN Tag时, 该字段取值为0x8100。
 - Ethernet Type: 以太报文类型, IP协议报文该字段取值为0x0800。

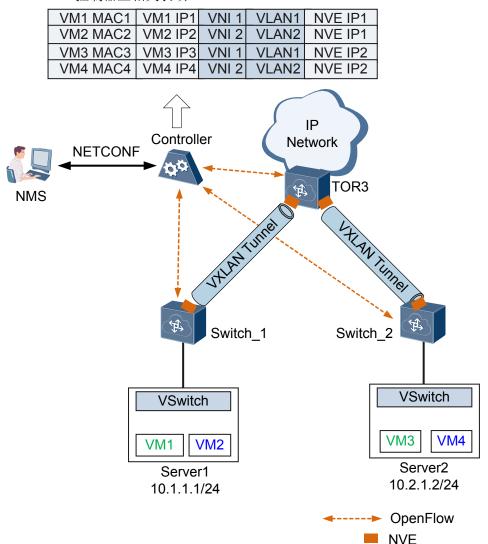
2.3 隧道建立与维护

控制器和转发器的邻居关系建立是通过OpenFlow协议完成,建立的通道称为OpenFlow通道。

控制器和转发器的邻居关系建立过程请见OpenFlow Agent配置。

图 2-3 控制器与转发器数据同步示意图

控制器上相关表项



用户可通过网管或CLI配置或维护控制器。图2-3所示,网管和控制器通过NETCONF协议建立连接后,用户通过网管配置控制器。在控制器上创建NVE,指定VTEP的源、目的IP地址,配置静态MAC地址或ARP表项,由此在控制器上生成静态MAC地址表项或静态ARP表项。

转发器和控制器之间OpenFlow通道建立成功后,转发器会主动将自身节点信息、接口信息上报给控制器。而控制器将静态MAC地址表项或静态ARP表项下发给转发器,控制器通过OpenFlow管理转发器上的静态MAC地址表项或静态ARP表项。

● 当控制器和转发器发生OpenFlow断连,转发器不会删除静态MAC地址表项和静态 ARP表项,以保证流量正常转发。 ● 当控制器和转发器重新建立连接后,转发器会主动将本地保存的静态MAC地址表项和静态ARP表项重新上报给控制器,以确保控制器和转发器上的静态MAC地址表项和静态ARP表项数据一致。

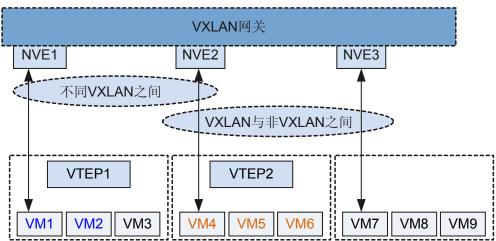
2.4 数据报文转发

在VXLAN网络中,业务接入点统一表现为二层子接口,通过在二层子接口上配置流封装实现不同的接口接入不同的数据报文。广播域统一表现为BD(Bridge-Domain),将二层子接口关联BD后,可实现数据报文通过BD转发。

由于VXLAN网络中通过VNI标识不同租户,将VNI以1:1方式映射到BD,可实现不同的租户通过不同的BD转发。

和VLAN类似,不同VNI之间的VXLAN,及VXLAN和非VXLAN之间不能直接进行二层通信。为了使VXLAN之间,以及VXLAN和非VXLAN之间能够进行通信,VXLAN引入了VXLAN网关,如图2-4所示。

图 2-4 VXLAN 网关示意图



VXLAN网关分为:

- 二层网关:用于同一网段的终端用户通信。 VXLAN二层网关收到用户报文,根据报文中包含的目的MAC地址类型,报文转发 流程分为:
 - MAC地址是BUM(Broadcast&Unknown-unicast&Multicast)地址,按照BUM报 文转发流程进行转发。
 - MAC地址是已知单播地址,按照**已知单播报文转发流程**进行转发。
- 三层网关: 用于非同一网段的终端用户通信或VXLAN和非VXLAN用户间的通信。

BUM 报文转发流程

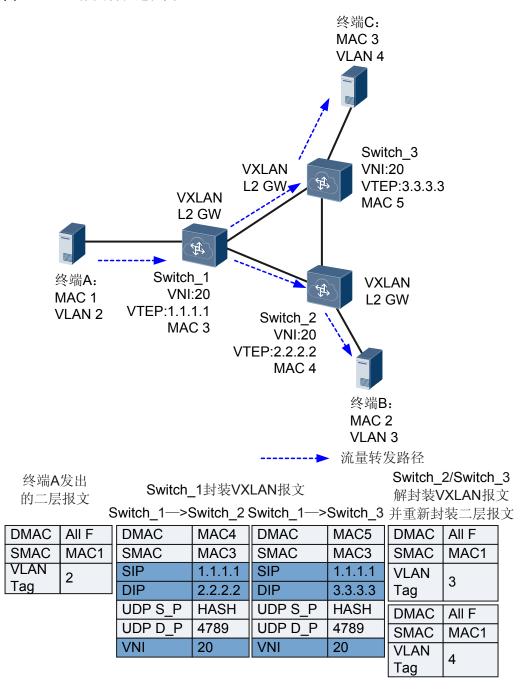
当BUM报文进入VXLAN隧道,接入端VTEP采用头端复制方式进行报文的VXLAN封装。BUM报文出VXLAN隧道,出口端VTEP对报文解封转。BUM报文具体转发流程如图2-5所示。

□ 说明

头端复制:接口收到BUM (Broadcast&Unknown-unicast&Multicast)报文,本地VTEP通过控制平面获取属于同一个VNI的VTEP列表,将收到的BUM报文根据VTEP列表进行复制并发送给属于同一个VNI的所有VTEP。

通过头端复制完成BUM报文的广播,不需要依赖组播路由协议。

图 2-5 BUM 报文转发过程图



1. Switch_1收到来自终端A的报文,根据报文中接入的端口和VLAN信息获取对应的二层广播域,并判断报文的目的MAC是否为BUM MAC。

- 是,在对应的二层广播域内广播,并跳转到2。
- 不是,通过已知单播报文转发流程。
- 2. Switch_1上VTEP根据对应的二层广播域获取对应VNI的头端复制隧道列表,依据获取的隧道列表进行报文复制,并进行VXLAN封装。基于每个出端口和VXLAN封装信息封装VXLAN头和外层IP信息,并从出端口转发。
- 3. Switch_2/Switch_3上VTEP收到VXLAN报文后,根据UDP目的端口号、源/目的IP地址、VNI判断VXLAN报文的合法有效性。依据VNI获取对应的二层广播域,然后进行VXLAN解封装,获取内层二层报文,判断报文的目的MAC是否为BUM MAC。
 - 是,在对应的二层广播域内非VXLAN侧进行广播处理。
 - 不是,再判断是否是本机MAC。
 - 是,上送主机处理。
 - 不是,在对应的二层广播域内查找出接口和封装信息,并跳转到4。
- 4. Switch_2/Switch_3根据查找到的出接口和封装信息,为报文添加VLAN Tag,转发给对应的终端B/C。

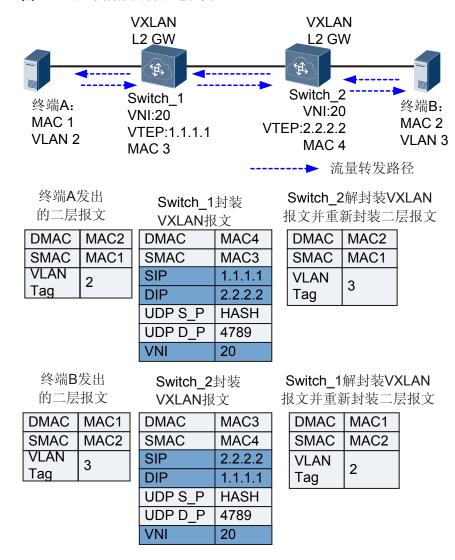
∭说明

终端B/C回应终端A的消息,按照**已知单播报文转发流程**进行转发。

已知单播报文转发流程

已知单播报文具体转发流程如图2-6所示。

图 2-6 已知单播报文转发过程图



- 1. Switch_1收到来自终端A的报文,根据报文中接入的端口和VLAN信息获取对应的二层广播域,并判断报文的目的MAC是否为已知单播MAC。
 - 是,再判断是否为本机MAC。
 - 是,上送主机处理。
 - 不是,在对应的二层广播域内查找出接口和封装信息,并跳转到2。
 - 不是,在对应的二层广播域内进行广播处理,并跳转到2。
- 2. Switch 1上VTEP根据查找到的出接口和封装信息进行VXLAN封装和报文转发。
- 3. Switch_2上VTEP收到VXLAN报文后,根据UDP目的端口号、源/目的IP地址、VNI 判断VXLAN报文的合法有效性。依据VNI获取对应的二层广播域,然后进行VXLAN 解封装,获取内层二层报文,判断报文的目的MAC是否为已知单播报文MAC。
 - 是,在对应的二层广播域内查找出接口和封装信息,并跳转到4。
 - 不是,再判断是否是本机MAC。
 - 是,上送主机处理。

- 不是,通过BUM报文转发流程。

4. Switch_2根据查找到的出接口和封装信息,为报文添加VLAN Tag,转发给对应的终端B。

三层网关

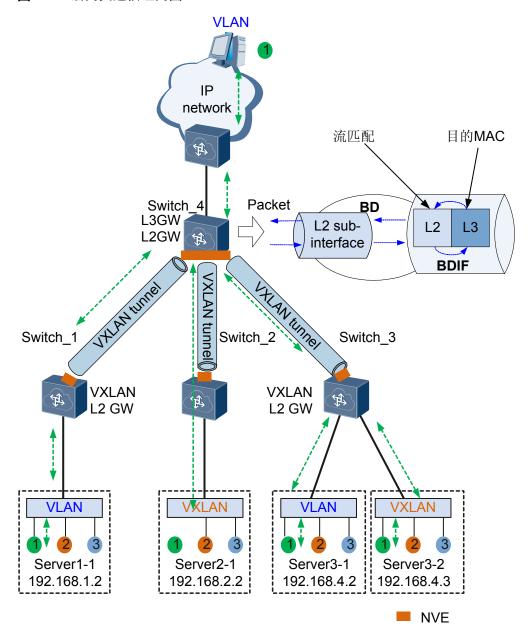
不同网段的VXLAN间通信,及VXLAN和非VXLAN的通信,需要通过IP路由实现。

在三层网关上创建BD,将VNI以1:1方式映射到BD,基于BD创建BDIF接口,通过BDIF接口配置IP地址实现不同网段的VXLAN间,及VXLAN和非VXLAN的通信。

∭说明

BDIF接口类似VLANIF接口。

图 2-7 三层网关通信组网图



如图2-7所示,三层网关通信具体实现过程如下:

- 1. 作为VXLAN二层网关的Switch_4收到VXLAN报文后进行解封装,确认内层报文中的DMAC是否是本网关接口的MAC地址。
 - 是,转给对应目的网段的三层网关处理,并跳转2。
 - 不是,在对应的二层广播域内查找出接口和封装信息。
- 2. 作为VXLAN三层网关的Switch_4剥除内层报文的以太封装,解析目的IP。根据目的IP查找ARP表项,确认DMAC、VXLAN隧道出接口及VIN等信息。
 - 没有VXLAN隧道出接口及VIN信息,进行三层转发。
 - 有VXLAN隧道出接口及VIN信息,跳转3。
- 3. 作为VXLAN二层网关的Switch_4重新封装VXLAN报文,其中内层报文以太头中的SMAC是网关接口的MAC地址。

□□说明

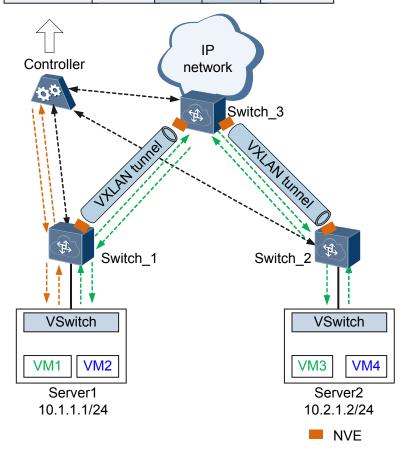
Switch 4与其他Switch之间的通信,请参见二层网关实现原理。

ARP 代答

如图2-8所示,在传统的终端用户互通过程中,如VM1初次访问VM3时,VM1会向VM3 发送ARP广播请求报文。ARP请求报文会在二层网络内广播,VM3收到ARP广播请求报文后进行ARP单播应答。

图 2-8 ARP 代答组网图

	VM1 MAC1	VM1 IP1	VNI 1	VLAN1	NVE IP1
ĺ	VM2 MAC2	VM2 IP2	VNI 2	VLAN2	NVE IP1
ĺ	VM3 MAC3	VM3 IP3	VNI 1	VLAN1	NVE IP2
	VM4 MAC4	VM4 IP4	VNI 2	VLAN2	NVE IP2



为了避免ARP广播请求报文给网络带来广播风暴,可在图2-8所示的控制器上使能ARP代答功能。VM1发送ARP请求报文,请求目的主机VM3的MAC地址具体实现过程如下:

- 1. VM1发送ARP请求报文(SMAC:MAC1, SIP:IP1, DMAC:FF-FF, DIP: IP3)。
- 2. NVE收到ARP请求报文后,通过Openflow通道上送控制器处理。
- 3. 控制器根据IP3查询用户信息库,获得VM3对应的MAC地址MAC3。
- 4. 控制器封装ARP应答报文,通过Openflow通道发送给NVE。
- 5. NVE根据控制器指定的出端口(ARP请求报文的入端口)将ARP应答报文发送给VM1。

2.5 VXLAN QoS

VXLAN QoS用来实现原始报文携带的QoS优先级、设备内部优先级(又称为本地优先级,是设备内部区分报文服务等级的优先级)与封装后报文优先级之间的转换,从而设备根据内部优先级提供有差别的QoS服务质量。

Network

Switch_1

VXLAN tunnel

Switch_2

VSwitch

VM1

VM2

NVE

如图2-9所示,VXLAN QoS实现的原始报文携带的QoS优先级、设备内部优先级与封装后报文优先级之间的转换过程如下。

- 1. 原始报文由二层子接口进入Switch_1设备,原始报文按照子接口上指定的VLAN上 绑定的DiffServ模板进行映射,将原始报文的802.1p优先级映射为设备内部优先级 (PHB行为和报文颜色),以此入队列。
- 2. 报文出Switch_1进入隧道,对原始报文进行加封装(外层依次添加VXLAN报文头、UDP报文头、IP报文头和以太报文头),加封装报文外层的802.1p优先级和DSCP优先级由原始报文内部优先级按照DiffServ域的缺省模板进行映射。报文按照映射后的优先级在隧道中进行传输。
- 3. 报文出隧道时,按照隧道接口上配置的信任类型802.1p或DSCP(以太网接口处于三层模式时只能信任DSCP),按照DiffServ域的缺省模板进行映射,映射为设备内部优先级,进入队列进行传输。
- 4. 最后,由设备内部优先级按照子接口上指定的VLAN上绑定的DiffServ域模板进行映射,映射到出接口报文的802.1p优先级,报文按照映射后的优先级进行传输。

3 应用场景

关于本章

介绍VXLAN的应用场景。

- 3.1 同网段终端用户通信的应用
- 3.2 不同网段终端用户通信的应用
- 3.3 在虚拟机迁移场景中的应用

3.1 同网段终端用户通信的应用

业务描述

数据中心用来在Internet网络基础设施上加速信息的传递,企业、运营商都在大力建设数据中心。目前,规模化、虚拟化、云计算已成为数据中心的发展方向,同时,数据中心为适应更大的业务量并降低维护成本,逐渐向大二层技术及虚拟化迁移。

随着数据中心在物理网络基础设施上实施服务器虚拟化的快速发展,作为NVo3技术之一的VXLAN技术具有很强的适应性,为数据中心提供了良好的解决方案。

组网描述

如图3-1所示,某企业在不同的数据中心中都拥有VM,且位于同一网段。现需要实现不同数据中心相同ID的VM的互通。

Controller Network VXLAN tunnel 事 红 Switch 2 Switch 1 VSwitch **VSwitch** VM1: VLAN10 VM2 VM1 VM2 VM1 VM2: VLAN20 Server2 Server1 10.1.1.1/24 10.1.1.2/24 NVE

图 3-1 同网段终端用户通信的应用组网图

特性部署

如图3-1所示,可将交换机设备作为VXLAN二层网关,交换机设备之间建立VXLAN隧道,通过VXLAN二层网关实现同一网段终端用户互通。

- 1. 控制器和转发器之间通过OpenFlow协议建立OpenFlow通道。
- 2. 通过NETCONF协议,网管人员在控制器上配置VXLAN。配置成功后,控制器通过 OpenFlow通道将VXLAN信息下发给转发器。
- 3. 转发器根据收到的信息建立VXLAN二层网关,转发器间建立VXLAN隧道。

4. 通过VXLAN隧道,依据静态MAC地址表项,实现同一网段终端用户(VM)互通。

3.2 不同网段终端用户通信的应用

业务描述

数据中心用来在Internet网络基础设施上加速信息的传递,企业、运营商都在大力建设数据中心。目前,规模化、虚拟化、云计算已成为数据中心的发展方向,同时,数据中心为适应更大的业务量并降低维护成本,逐渐向大二层技术及虚拟化迁移。

随着数据中心在物理网络基础设施上实施服务器虚拟化的快速发展,作为NVo3技术之一的VXLAN技术具有很强的适应性,为数据中心提供了良好的解决方案。

组网描述

如图3-2所示,某企业在不同的数据中心中都拥有VM,且位于不同网段。现需要实现不同数据中心同一企业的VM互通。

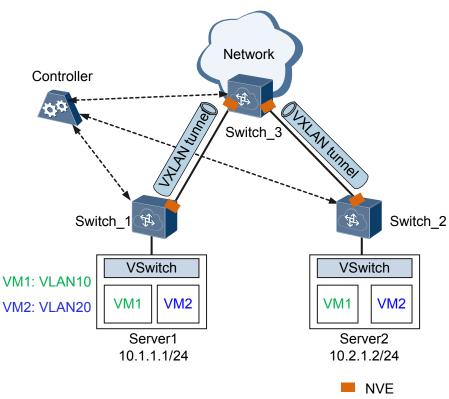


图 3-2 不同网段终端用户通信的应用组网图

特性部署

如图3-2所示,可将Switch_3设备作为VXLAN三层网关,其他交换机设备作为VXLAN二层网关,交换机设备之间建立VXLAN隧道,通过VXLAN三层网关实现不同网段终端用户互通。

- 1. 控制器和转发器之间通过OpenFlow协议建立OpenFlow通道。
- 2. 通过NETCONF协议,网管人员在控制器上配置VXLAN。配置成功后,控制器通过 OpenFlow通道将VXLAN信息下发给转发器。
- 3. 转发器Switch_1、Switch_2根据收到的信息建立VXLAN二层网关,Switch间建立 VXLAN隧道。转发器Switch_3根据收到的信息建立VXLAN三层网关。
- 4. 在控制器上,为VXLAN三层网关创建三层逻辑接口BDIF接口配置IP地址,并使能ARP代答功能。
- 5. 通过VXLAN三层网关和ARP代答功能,依据静态ARP地址表项,实现不同网段终端用户(VM)互通。

3.3 在虚拟机迁移场景中的应用

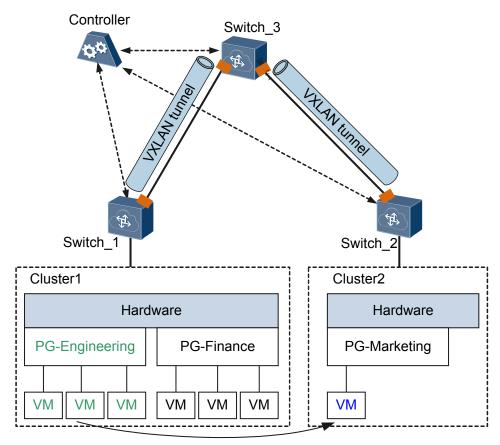
业务描述

当前数据中心网络中企业通过部署服务器虚拟化来达到整合IT资源、提升资源利用效率、降低开支的目的。随着虚拟化水平的不断提高,物理服务器上虚拟机的数量在不断增加,虚拟化环境下运行的应用数量也在不断上升,为虚拟网络带来了很大的挑战。

组网描述

如**图3-3**所示,某企业在数据中心中有两个群集Cluster,其中工程部门和财务部门都在Cluster1上,营销部门在Cluster2上。

图 3-3 企业分布组网图



Engineering (VLAN10):10.1.1.1/24 Finance (VLAN20):10.10.1.1/24 Marketing (VLAN30):10.100.1.1/24

Cluster1上显示计算空间不足,而Cluster2未充分利用。网络管理员需要将工程部门迁移到Cluster2上,而不影响业务。

特性部署

为了保证工程部门迁移过程中业务不中断,则需要保证工程部门的IP地址、MAC地址等参数保持不变,这就要求两个Cluster属于一个二层网络。如果使用传统方法解决此问题,这可能需要网络管理员购买新的物理设备以分离流量,并可能导致诸如VLAN散乱、网络成环以及系统和管理开销等问题。

为了成功将工程部门迁移到Cluster2,可通过VXLAN实现。VXLAN是MAC in UDP的网络虚拟化技术,只要物理网络支持IP转发,所有IP路由可达的终端用户即可建立一个大范围二层网络。

- 控制器和转发器之间通过OpenFlow协议建立OpenFlow通道。
- 通过NETCONF协议,网管人员在控制器上配置VXLAN。配置成功后,控制器通过 OpenFlow通道将VXLAN相关信息下发给转发器。
- 转发器根据收到的信息建立VXLAN隧道和VXLAN三层网关。

● 通过VXLAN隧道和VXLAN三层网关,工程部门在迁移过程中可保证网络无感知。

4 基于 SDN 控制器的 VXLAN 配置示例

该示例主要介绍如何在SDN控制器上配置VXLAN三层网关实现不同网段的租户通信。

组网需求

为了方便控制与部署,引入了控制器概念。控制器通过OpenFlow协议将信息下发给转发器,以实现控制器统一维护管理。

作为云计算的核心技术之一,服务器虚拟化凭借其大幅降低IT成本、提高业务部署灵活性、降低运维成本等优势已经得到越来越多的认可和部署。服务器虚拟化后,可支持多租户接入。因为一台服务器可虚拟多台虚拟机,而一台虚拟机相当于一台主机,主机数量发生了数量级的变化,这也为虚拟网络带来问题。例如:虚拟机规模受网络规格限制、网络隔离能力限制及虚拟机迁移范围受网络架构限制。

为了充分发挥服务器虚拟化的优势、解决服务器虚拟化后带来的问题,可以通过部署 VXLAN(Virtual eXtensible Local Area Network)实现服务器虚拟化后可满足16M租户接入,且租户可以规划自己的虚拟网络,不需要考虑物理网络IP地址和广播域的限制。

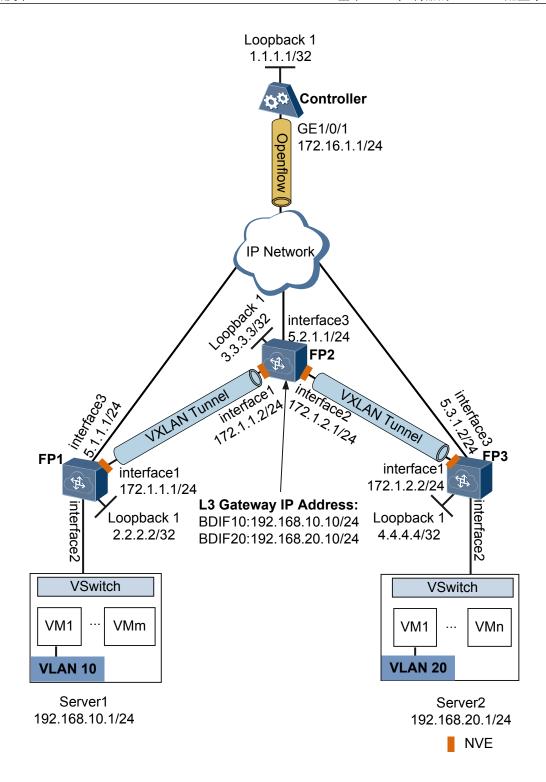
如图4-1所示,某企业在不同的数据中心中都拥有自己的VM,服务器1上的VM1属于VLAN10,服务器2上的VM1属于VLAN20,且两台服务器位于不同网段。现需要通过VXLAN实现不同数据中心相同VM的互通。

图 4-1 基于 SDN 控制器的 VXLAN 典型组网图

□ 说明

本例的配置主要在SNC Controller上进行。

本例中FP设备上的interface1、interface2、interface3分别代表10GE1/0/1、10GE1/0/2、10GE1/0/3。



注意事项

配置VXLAN前,必须保证控制器和转发器三层路由可达。

为了使VXLAN之间,以及VXLAN和非VXLAN之间能够进行通信,VXLAN引入了VXLAN网关。VXLAN网关分为:

- 二层网关:用于同一网段的租户通信。
- 三层网关:用于不同网段的租户通信或VXLAN和非VXLAN用户间的通信。 为了成功实现不同网段的VM互通,VM的缺省网关地址必须是对应三层网关上的BDIF接口的IP地址。

无论二层网关还是三层网关,VXLAN的配置相同,不同点在于三层网关上需要创建BDIF接口并配置IP地址。

配置思路

采用如下的思路配置基于SDN控制器的VXLAN:

- 1. 分别在转发器FP1、FP2、FP3上配置隧道模式。
- 2. 分别在控制器Controller、转发器FP1、FP2、FP3上配置路由协议,保证网络三层互通。

本示例以OSPF(Open Shortest Path First)为例。

- 3. 分别在控制器Controller、转发器FP1、FP2、FP3上配置Openflow,建立控制器和转发器之间的通信通道。
- 4. 在控制器上配置VXLAN三层网关,配置思路如下:
 - a. 分别为转发器FP1、FP3配置业务接入点,实现不同的业务报文通过不同的接口接入网络。
 - b. 分别为转发器FP1、FP2、FP3配置VXLAN,FP1与FP2、FP2与FP3之间建立 VXLAN隧道,通过VXLAN隧道实现数据转发。
 - c. 分别为转发器FP1、FP2、FP3配置静态MAC地址表,指导报文正确二层转发。
 - d. 为FP3配置三层逻辑接口BDIF接口,并配置IP地址,确定三层网关地址。
 - e. 为FP3配置静态ARP地址表,指导业务报文正确三层转发。
 - f. 分别为FP1、FP2使能ARP代答,避免ARP广播请求报文给网络带来广播风暴。

操作步骤

步骤1 配置隧道模式

#配置转发器FP1。

<Switch> system-view
[~Switch] sysname FP1
[*Switch] commit
[~FP1] ip tunnel mode vxlan

[*FP1] commit

∭说明

此命令功能需要保存配置并重启设备才能生效,您可以选择立即重启或完成所有配置后再重启。

在FP2、FP3与配置FP1类似,这里不再赘述。具体配置过程略,请参考本配置举例中的配置文件。

步骤2 配置路由协议

按**图4-1**分别配置控制器Controller、转发器FP1、FP2、FP3各接口IP地址。配置OSPF时,注意需要发布控制器、转发器的32位Loopback接口地址。

● 配置转发器FP1

表 4-1 FP1 配置

步骤	命令	说明
1	[-FP1] vlan batch 100 300 [*FP1] interface vlanif 100 [*FP1-Vlanif100] ip address 172.1.1.1 24 [*FP1-Vlanif100] quit [*FP1] interface vlanif 300 [*FP1-Vlanif300] ip address 5.1.1.1 24 [*FP1-Vlanif300] quit [*FP1] interface loopback 1 [*FP1-LoopBack1] ip address 2.2.2.2 32 [*FP1-LoopBack1] quit [*FP1] interface 10ge 1/0/3 [*FP1-10GE1/0/3] port link-type trunk [*FP1-10GE1/0/3] quit [*FP1-10GE1/0/3] quit [*FP1] interface 10ge 1/0/1 [*FP1-10GE1/0/1] port trunk allow-pass vlan 300 [*FP1-10GE1/0/1] port trunk allow-pass vlan 100 [*FP1-10GE1/0/1] quit	配置接口IP地址。
2	[*FP1] ospf [*FP1-ospf-1] area 0 [*FP1-ospf-1-area-0.0.0.0] network 2.2.2.2 0.0.0.0 [*FP1-ospf-1-area-0.0.0.0] network 5.1.1.0 0.0.0.255 [*FP1-ospf-1-area-0.0.0.0] network 172.1.1.0 0.0.0.255 [*FP1-ospf-1-area-0.0.0.0] quit [*FP1-ospf-1] quit [*FP1] commit	配置OSPF。

● 在Controller、FP2、FP3与配置FP1类似,这里不再赘述。具体配置过程略,请参考本配置举例中的配置文件。

OSPF成功配置后,控制器与转发器,转发器与转发器之间可通过OSPF协议发现对方的 Loopback接口的IP地址,并能互相ping通。

以Controller的显示为例。

```
[~Controller] ping 4.4.4.4

PING 4.4.4: 56 data bytes, press CTRL_C to break

Reply from 4.4.4: bytes=56 Sequence=1 tt1=255 time=4 ms

Reply from 4.4.4: bytes=56 Sequence=2 tt1=255 time=2 ms

Reply from 4.4.4: bytes=56 Sequence=3 tt1=255 time=2 ms

Reply from 4.4.4: bytes=56 Sequence=3 tt1=255 time=2 ms

Reply from 4.4.4: bytes=56 Sequence=4 tt1=255 time=2 ms

Reply from 4.4.4: bytes=56 Sequence=5 tt1=255 time=1 ms

--- 4.4.4.4 ping statistics ---
5 packet(s) transmitted
5 packet(s) received
0.00% packet loss

round-trip min/avg/max = 1/2/4 ms
```

步骤3 配置Openflow

● 在控制器Controller上配置与转发器FP1的链接

表 4-2 Controller 配置

步骤	命令	说明
1	[~Controller] sdn controller	将设备设置为SDN控制器,并进入SDN控制器视图。
2	[*Controller-sdn-controller] openflow listening-ip 1.1.1.1	配置SDN控制器侦听地 址。
3	[*Controller-sdn-controller] fp-id 10	配置转发器FP1的标识是 FP10,并进入FP视图。
4	[*Controller-sdn-controller-fp10] openflow controller	配置控制器与转发器之间 的通信通道采用 Openflow连接并进入 Openflow视图。
5	[*Controller-sdn-controller-fp10-openflow] peer-address 2.2.2.2 [*Controller-sdn-controller-fp10-openflow] quit [*Controller-sdn-controller-fp10] quit [*Controller-sdn-controller] quit [*Controller] commit	指定转发器FP1的 Loopback地址。

在Controller上配置与FP2、FP3的链接与配置FP1类似,这里不再赘述。具体配置过程略,请参考本配置举例中的配置文件。

● 配置转发器FP1

表 4-3 FP1 配置

步骤	命令	说明
1	[~FP1] sdn agent	将FP1设置为转发器,并 进入SDN Agent视图。
2	[*FP1-sdn-agent] controller-ip 1.1.1.1	指定控制器的Loopback 地址,并进入Controller 视图。
3	[*FP1-sdn-agent-ctrl-1.1.1.1] openflow agent	配置转发器与控制器之间 的通信通道采用 Openflow连接并进入 Openflow Agent视图。
4	[*FP1-sdn-agent-ctrl-1.1.1.1-openflow] transport-address 2.2.2.2 [*FP1-sdn-agent-ctrl-1.1.1.1] quit [*FP1-sdn-agent] quit [*FP1] commit	配置Openflow连接的本端地址。

转发器FP2、FP3配置与FP1类似,这里不再赘述。具体配置过程略,请参考本配置举例中的配置文件。

上述配置成功后,分别在转发器和控制器上执行**display sdn openflow session**命令可查看到Openflow连接的状态是"REGISTERED",说明Openflow通道建立成功。以控制器显示为例:

[~Controller] display sdn openflow session

FPID	AgentAddr	ListeningAddr	UpTime	State
	2. 2. 2. 2 3. 3. 3. 3 4. 4. 4. 4	1. 1. 1. 1 1. 1. 1. 1 1. 1. 1. 1	0d00h23m28s 0d00h55m17s 0d00h00m05s	REGISTERED

步骤4 在控制器上配置VXLAN三层网关

- 1. 配置业务接入点
 - 在控制器上配置FP1

表 4-4 FP1 配置

步骤	命令	说明
1	[~Controller] sdn fp service	使能转发器服务。
2	[*Controller-sdn-service] fp-id 10	创建并进入FP1转发器 视图。
3	[*Controller-sdn-service-fp10] bridge-domain 10 [*Controller-sdn-service-fp10-bd10] quit	创建广播域BD。
4	[*Controller-sdn-service-fp10] interface vserviceif 10:1	创建vServiceIf接口, 并进入vServiceIf接口 视图。
		在VXLAN网络中,业 务接入点统一表示为二 层vServiceIf子接口。
5	[*Controller-sdn-service-fp10-vServiceIf10:1] binding interface 10ge1/0/2 [*Controller-sdn-service-fp10-vServiceIf10:1] quit	为vServiceIf接口绑定 物理接口。
6	[*Controller-sdn-service-fp10] interface vserviceif 10:1.1 mode 12	创建二层vServiceIf子 接口,并进入子接口视 图。
7	[*Controller-sdn-service-fp10-vServiceIf10:1.1] encapsulation dot1q vid 10	配置二层vServiceIf子 接口的流封装类型,允 许接口接收携带一层 VLAN Tag是10的报文。
8	[*Controller-sdn-service-fp10-vServiceIf10:1.1] bridge-domain 10 [*Controller-sdn-service-fp10-vServiceIf10:1.1] quit [*Controller-sdn-service-fp10] quit [*Controller-sdn-service] commit	将二层vServiceIf子接口加入BD,允许报文通过广播域BD转发。

● 在控制器上配置FP3

表 4-5 FP3 配置

步骤	命令	说明
1	[~Controller-sdn-service] fp-id 30	创建并进入FP3转发器 视图。
2	[*Controller-sdn-service-fp30] bridge-domain 20 [*Controller-sdn-service-fp30-bd20] quit	创建广播域BD。
3	[*Controller-sdn-service-fp30] interface vserviceif 30:1	创建vServiceIf接口, 并进入vServiceIf接口 视图。
4	[*Controller-sdn-service-fp30-vServiceIf30:1] binding interface 10ge1/0/2 [*Controller-sdn-service-fp30-vServiceIf30:1] quit	为vServiceIf接口绑定 物理接口。
5	[*Controller-sdn-service-fp30] interface vserviceif 30:1.1 mode 12	创建二层vServiceIf子 接口,并进入子接口视 图。
6	[*Controller-sdn-service-fp30-vServiceIf30:1.1] encapsulation dot1q vid 20	配置二层vServiceIf子 接口的流封装类型,允 许接口接收携带一层 VLAN Tag是20的报文。
7	[*Controller-sdn-service-fp30-vServiceIf30:1.1] bridge-domain 20 [*Controller-sdn-service-fp30-vServiceIf30:1.1] quit [*Controller-sdn-service-fp30] quit [*Controller-sdn-service] commit	将二层vServiceIf子接口加入BD,允许报文通过广播域BD转发。

2. 配置VXLAN

● 在控制器上配置FP1

表 4-6 FP1 配置

步骤	命令	说明
1	[~Controller-sdn-service] fp-id 10	进入FP1转发器视图。
2	[~Controller-sdn-service-fp10] bridge-domain 10 [~Controller-sdn-service-fp10-bd10] vxlan vni 10 [*Controller-sdn-service-fp10-bd10] quit	创建VXLAN网络标识 VNI并关联广播域BD, 将VNI以1:1方式映射到 BD,通过BD转发流量。
3	[*Controller-sdn-service-fp10] interface nve 10:1	创建NVE接口,并进入 NVE接口视图。

步骤	命令	说明
4	[*Controller-sdn-service-fp10-Nve10:1] source 2.2.2.2	配置源端VTEP的IP地 址。
5	[*Controller-sdn-service-fp10-Nve10:1] overlay-encapsulation vxlan	配置NVE接口的封装类 型是VXLAN。
6	[*Controller-sdn-service-fp10-Nve10:1] vni 10 head-end peer-list 3.3.3.3 [*Controller-sdn-service-fp10-Nve10:1] quit [*Controller-sdn-service-fp10] quit [*Controller-sdn-service] commit	配置头端复制列表。通过头端复制列表,源端NVE接口将收到的BUM(Broadcast&Unknownunicast&Multicast)报文,根据VTEP列表进行复制并发送给属于同一个VNI的所有VTEP。

● 在控制器上配置FP2

表 4-7 FP2 配置

步骤	命令	说明
1	[~Controller-sdn-service] fp-id 20	进入FP2转发器视图。
2	[*Controller-sdn-service-fp20] bridge-domain 10 [*Controller-sdn-service-fp20-bd10] vxlan vni 10 [*Controller-sdn-service-fp20-bd10] quit [*Controller-sdn-service-fp20] bridge-domain 20 [*Controller-sdn-service-fp20-bd20] vxlan vni 20 [*Controller-sdn-service-fp20-bd20] quit	创建VXLAN网络标识 VNI并关联广播域BD, 将VNI以1:1方式映射到 BD,通过BD转发流量。
3	[*Controller-sdn-service-fp20] interface nve 20:1	创建NVE接口,并进入 NVE接口视图。
4	[*Controller-sdn-service-fp20-Nve20:1] source 3.3.3.3	配置源端VTEP的IP地 址。
5	[*Controller-sdn-service-fp20-Nve20:1] overlay- encapsulation vxlan	配置NVE接口的封装类 型是VXLAN。
6	[*Controller-sdn-service-fp20-Nve20:1] vni 10 head- end peer-list 2.2.2.2 [*Controller-sdn-service-fp20-Nve20:1] vni 20 head- end peer-list 4.4.4.4 [*Controller-sdn-service-fp20-Nve20:1] quit [*Controller-sdn-service-fp20] quit [*Controller-sdn-service] commit	配置头端复制列表。

● 在控制器上配置FP3

表 4-8 FP3 配置

步骤	命令	说明
1	[~Controller-sdn-service] fp-id 30	进入FP3转发器视图。
2	[~Controller-sdn-service-fp30] bridge-domain 20 [~Controller-sdn-service-fp30-bd20] vxlan vni 20 [*Controller-sdn-service-fp30-bd20] quit	创建VXLAN网络标识 VNI并关联广播域BD, 将VNI以1:1方式映射到 BD,通过BD转发流量。
3	[*Controller-sdn-service-fp30] interface nve 30:1	创建NVE接口,并进入 NVE接口视图。
4	[*Controller-sdn-service-fp30-Nve30:1] source 4.4.4.4	配置源端VTEP的IP地 址。
5	[*Controller-sdn-service-fp30-Nve30:1] overlay- encapsulation vxlan	配置NVE接口的封装类 型是VXLAN。
6	[*Controller-sdn-service-fp30-Nve30:1] vni 20 head- end peer-list 3.3.3.3 [*Controller-sdn-service-fp30-Nve30:1] quit [*Controller-sdn-service-fp30] quit [*Controller-sdn-service] commit	配置头端复制列表。

上述配置成功后,在控制器上指定转发器视图下执行**display vxlan vni**命令可查看到 VNI的状态是Up; 执行**display vxlan tunnel**命令可查看到VXLAN隧道的信息。以 FP2转发器显示为例:

[~Controller-sdn-service] fp-id 20 [~Controller-sdn-service-fp20] display vxlan vni

Number of vxlan vni: 2

VNI	BD-ID	State
10	10	UP
20	20	UP

 $[\hbox{-}Controller-sdn-service-fp20}] \ \, \textbf{display vxlan tunnel}$

Number of Vxlan tunnel : 2

Tunnel ID	Source	Destination	State	Туре
33686018	3. 3. 3. 3	2. 2. 2. 2	up	static
67372036	3. 3. 3. 3	4. 4. 4. 4	up	static

- 3. 在控制器上分别为转发器FP1、FP2、FP3配置静态MAC地址表
 - 在控制器上配置FP1

表 4-9 FP1 配置

步骤	命令	说明
1	[~Controller-sdn-service-fp20] quit [~Controller-sdn-service] fp-id 10 [~Controller-sdn-service-fp10] mac-address static 38eb-d921-0301 bridge-domain 10 nve 10:1 peer 3.3.3.3 vni 10 [*Controller-sdn-service-fp10] quit	配置静态MAC地址表 项。

● 在控制器上配置FP2

表 4-10 FP2 配置

步骤	命令	说明
1	[*Controller-sdn-service] fp-id 20 [*Controller-sdn-service-fp20] mac-address static 38eb-d911-0301 bridge-domain 10 nve 20:1 peer 2. 2. 2. 2 vni 10 [*Controller-sdn-service-fp20] mac-address static 38eb-d931-0301 bridge-domain 20 nve 20:1 peer 4. 4. 4. 4 vni 20 [*Controller-sdn-service-fp20] quit	配置静态MAC地址表项。

● 在控制器上配置FP3

表 4-11 FP3 配置

步骤	命令	说明
1	[*Controller-sdn-service] fp-id 30 [*Controller-sdn-service-fp30] mac-address static 38eb-d921-0302 bridge-domain 20 nve 30:1 peer 3.3.3.3 vni 20 [*Controller-sdn-service-fp30] quit [*Controller-sdn-service] commit	配置静态MAC地址表项。

上述配置成功后,在控制器上转发器视图下执行**display mac-address vxlan**命令可查看到VXLAN的静态MAC地址表项信息。

在控制器上以转发器FP2显示为例。

[~Controller-sdn-service] ${\bf fp\hbox{--}id}~{\bf 20}$

[~Controller-sdn-service-fp20] display mac-address vxlan

MAC Address	BD/VLAN	PORT	PEER-IP	VNI	Туре
38eb-d911-0301		Nve20:1	2. 2. 2. 2	10	static
38eb-d931-0301		Nve20:1	4. 4. 4. 4	20	static

Total items: 2

4. 在控制器上为FP2转发器创建BDIF接口并配置IP地址

表 4-12 FP2 配置

步骤	命令	说明
1	[~Controller-sdn-service-fp20] interface vbdif 20:10	创建BDIF10接口,并进 入BDIF10接口视图。
2	[*Controller-sdn-service-fp20-Vbdif20:10] ip address 192.168.10.10 24 [*Controller-sdn-service-fp20-Vbdif20:10] quit	配置BDIF10接口IP地址。
3	[*Controller-sdn-service-fp20] interface vbdif 20:20	创建BDIF20接口,并进入BDIF20接口视图。
4	[*Controller-sdn-service-fp20-Vbdif20:20] ip address 192.168.20.10 24 [*Controller-sdn-service-fp20-Vbdif20:20] quit [*Controller-sdn-service-fp20] commit	配置BDIF20接口IP地 址。

5. 在控制器上为FP2转发器配置静态ARP表

表 4-13 FP2 配置

步骤	命令	说明
1	[~Controller-sdn-service-fp20] arp static 192.168.10.1 38eb-d911-0301 vni 10 source-ip 3.3.3.3 peer-ip 2.2.2.2 [*Controller-sdn-service-fp20] arp static 192.168.20.1 38eb-d931-0301 vni 20 source-ip 3.3.3.3 peer-ip 4.4.4.4 [*Controller-sdn-service-fp20] commit	配置静态ARP表项。

上述配置成功后,在控制器上FP2转发器视图下执行display arp all命令可查看到所有的ARP表项。

[~Controller-sdn-service-fp20] display arp all						
IP ADDRESS	MAC ADDRESS	EXPIRE(M)	TYPE	INTERFACE	VPN-INSTANCE	
			VLAN/CEVLAN	PVC		
192. 168. 10. 1	38eb-d911-0301		S	VxLAN-Tunne	1	
192. 168. 20. 1	38eb-d931-0301		S	VxLAN-Tunne	1	
192. 168. 10. 10	38ba-8fab-6903		I -	Vbdif20:10		
192. 168. 20. 10	38ba-8fab-6903		I -	Vbdif20:20		
Total:4	Dynamic:0	Static:2	Interface	:2		

6. 在控制器上为转发器FP1、FP3使能ARP代答功能

表 4-14 FP2 配置

步骤	命令	说明
1	[~Controller-sdn-service-fp20] quit [~Controller-sdn-service] fp-id 10 [~Controller-sdn-service-fp10] bridge-domain 10 [~Controller-sdn-service-fp10-bd10] arp 12-proxy enable [*Controller-sdn-service-fp10-bd10] quit [*Controller-sdn-service-fp10] quit [*Controller-sdn-service] fp-id 20 [*Controller-sdn-service-fp20] bridge-domain 20 [*Controller-sdn-service-fp20-bd20] arp 12-proxy enable [*Controller-sdn-service-fp20-bd20] quit [*Controller-sdn-service-fp20] quit [*Controller-sdn-service] commit	使能ARP代答功能。

步骤5 检查配置结果

在VLAN10中的VM1上配置缺省网关为BDIF10接口的IP地址192.168.10.10/24。在VLAN20中的VM1上配置缺省网关为BDIF20接口的IP地址192.168.20.10/24。配置完成后,不同网段的VLAN10和VLAN20中的VM能够相互ping通。

----结束

配置文件

● 控制器的配置文件

```
sysname Controller
sdn controller
openflow listening-ip 1.1.1.1
fp-id 10
 openflow controller
  peer-address 2.2.2.2
fp-id 20
 openflow controller
  peer-address 3.3.3.3
fp-id 30
 openflow\ controller
  peer-address 4.4.4.4
sdn fp service
fp-id 10
fp-id 20
fp-id 30
interface\ Gigabitethernet 1/0/1
undo portswitch
ip address 172.16.1.1 255.255.255.0
interface LoopBack1
ip address 1.1.1.1 255.255.255.255
#
interface NULLO
#
ospf 1
area 0.0.0.0
 network 1.1.1.1 0.0.0.0
 network 172.16.1.0 0.0.0.255
```

```
sdn controller
openflow listening-ip 1.\,1.\,1.\,1
fp-id 10
 openflow controller
  peer-address 2.2.2.2
 openflow controller
  peer-address 3.3.3.3
 openflow controller
  peer-address 4.4.4.4
sdn fp service
fp-id 10
 #
 bridge-domain 10
  vxlan vni 10
  arp 12-proxy enable
 interface Nvel0:1
  source 2.2.2.2
  vni 10 head-end peer-list 3.3.3.3
 interface vServiceIf10:1
  binding interface 10GE1/0/2
 interface vServiceIf10:1.1 mode 12
  encapsulation dot1q vid 10
  bridge-domain 10
 mac-address static 38eb-d921-0301 bridge-domain 10 Nve10:1 peer 3.3.3.3 vni 10
sdn controller
openflow listening-ip 1.1.1.1
fp-id 20
 openflow controller
  peer-address 3.3.3.3
 openflow controller
  peer-address 2.2.2.2
 openflow controller
  peer-address 4.4.4.4
sdn fp service
fp-id 20
 #
 bridge-domain 10
  vxlan vni 10
 bridge-domain 20
  vxlan vni 20
  arp 12-proxy enable
 interface Nve20:1
  source 3.3.3.3
  vni 10 head-end peer-list 2.2.2.2
  vni 20 head-end peer-list 4.4.4.4
 interface Vbdif20:10
  ip address 192.168.10.10 255.255.255.0
 interface Vbdif20:20
  ip address 192.168.20.10 255.255.255.0
 arp static 192.168.10.1 38eb-d911-0301 vni 10 source-ip 3.3.3.3 peer-ip 2.2.2.2
 arp static 192.168.20.1 38eb-d931-0301 vni 20 source-ip 3.3.3.3 peer-ip 4.4.4.4
 mac-address static 38eb-d911-0301 bridge-domain 10 Nve20:1 peer 2.2.2.2 vni 10
 \verb|mac-address| static 38eb-d931-0301| bridge-domain 20 \ Nve20:1| peer 4.4.4.4 \ vni \ 20
sdn controller
```

```
openflow listening-ip 1.1.1.1
 fp-id 30
 openflow controller
  peer-address 4.4.4.4
  openflow controller
  peer-address 2.2.2.2
  openflow controller
  peer-address 3.3.3.3
sdn fp service
fp-id 30
  bridge-domain 20
  vxlan vni 20
  interface Nve30:1
  source 4.4.4.4
   vni 20 head-end peer-list 3.3.3.3
  interface vServiceIf30:1
  binding interface 10GE1/0/2
  interface vServiceIf30:1.1 mode 12
   encapsulation dot1q vid 20
  {\tt bridge-domain}\ 20
 \verb|mac-address| static 38eb-d921-0302| bridge-domain 20 Nve30:1 peer 3.3.3.3 vni 20
#
return
```

● FP1的配置文件

```
#
sysname FP1
vlan batch 10 100 300
sdn agent
controller-ip 1.1.1.1
  openflow agent
  transport-address 2.2.2.2
interface Vlanif100
ip address 172.1.1.1 255.255.255.0
interface Vlanif300
ip address 5.1.1.2 255.255.255.0
interface 10GE\ 1/0/1
port link-type trunk
port trunk allow-pass vlan 100
interface 10GE \ 1/0/2
port link-type trunk
port trunk allow-pass vlan 10
interface 10GE\ 1/0/3
port link-type trunk
port trunk allow-pass vlan 300
interface LoopBack1
ip address 2.2.2.2 255.255.255.255
ospf 1
area 0.0.0.0
 network 2.2.2.2 0.0.0.0
 network 5.1.1.0 0.0.0.255
  network 172.1.1.0 0.0.0.255
```

return

● FP2的配置文件

```
sysname FP2
#
vlan batch 100 200 300
sdn agent
controller-ip 1.1.1.1
 openflow agent
  transport-address 3.3.3.3
interface Vlanif100
ip address 172.1.1.2\ 255.255.255.0
interface Vlanif200
ip address 172.1.2.1 255.255.255.0
interface Vlanif300
ip address 5.2.1.2 255.255.255.0
interface 10GE 1/0/1
port link-type trunk
port trunk allow-pass vlan 100
interface 10GE \ 1/0/2
port link-type trunk
port trunk allow-pass vlan 200
interface 10GE 1/0/3
port link-type trunk
port trunk allow-pass vlan 300
interface LoopBack1
ip address 3.3.3.3 255.255.255.255
#
ospf 1
area 0.0.0.0
 network 3.3.3.3 0.0.0.0
 network 5.2.1.0 0.0.0.255
 network 172.1.1.0 0.0.0.255
 network 172.1.2.0 0.0.0.255
return
```

● FP3的配置文件

```
#
sysname FP3
#
vlan batch 20 200 300
#
sdn agent
  controller-ip 1.1.1.1
  openflow agent
    transport-address 4.4.4.4
#
interface Vlanif200
  ip address 172.1.2.2 255.255.255.0
#
interface Vlanif300
  ip address 5.3.1.2 255.255.255.0
#
interface 10GE 1/0/1
  port link-type trunk
  port trunk allow-pass vlan 200
```

```
#
interface 10GE 1/0/2
port link-type trunk
port trunk allow-pass vlan 20
#
interface 10GE 1/0/3
port link-type trunk
port trunk allow-pass vlan 300
#
interface LoopBackl
ip address 4. 4. 4. 4 255. 255. 255. 255
#
ospf 1
area 0.0.0.0
network 4. 4. 4. 4 0. 0. 0. 0
network 5. 3. 1. 0 0. 0. 0. 255
network 172. 1. 2. 0 0. 0. 0. 255
#
return
```

5 参考标准和协议

介绍VXLAN的参考标准和协议。

与VXLAN特性相关的参考标准及协议如下:

文档	描述	备注
draft-ietf-NVo3- framework-04	Framework for DC Network Virtualization	-
draft-ietf-NVo3- dataplane- requirements-02	NVo3 Data Plane Requirements	-
draft- mahalingam- dutt-dcops- vxlan-06	A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks	-