基于 VXLAN 的 EVPN 技术研究与实现

钟耿辉[1],唐加山[2]
(1. 南京邮电大学 通信与信息工程学院,江苏 南京 210003;
2. 南京邮电大学 理学院,江苏 南京 210003)

摘 要:VXLAN 作为 overlay 网络技术的代表,为解决云数据中心的组网问题提供了有效的技术支持,使网络更具可扩展性。但传统 VXLAN 技术是基于数据平面的,存在诸多局限性。早在 EVPN 之前,运营商网络 PE 设备间的 MAC 或 ARP 学习是基于传统网桥功能的,以泛洪学习模式运行。在这种模式下,终端虚机信息学习和 VTEP 的自动发现均基于数据平面,缺少控制平面来发布终端虚机的可达信息。为此,将 VXLAN 作为数据平面,基于 VXLAN 封装的 MP-BGP EVPN 作为控制平面,改变了这种泛洪学习模式,提供了远端 VTEP 下终端虚机基于控制平面的学习模式,并设计实现了在 VXLAN overlay网络中二层和三层转发的控制平面方案。该方案将 MAC/IP 路由作为控制平面消息,利用该路由携带的信息进行终端虚拟机信息的学习,并实现了虚拟机之间的互通,取代了泛洪学习模式,有效减少了网络流量。

关键词:EVPN;VXLAN;数据平面;控制平面
中图分类号:TP31 **文献标识码**:A **文章编号**:1673-629X(2017)05-0046-05
doi:10.3969/j.issn.1673-629X.2017.05.010

Research and Implementation of EVPN Technology with VXLAN

ZHONG Geng-hui[1], TANG Jia-shan[2]
(1. College of Telecommunication & Information Engineering, Nanjing University of Posts & Telecommunications, Nanjing 210003, China;
2. College of Science, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: As the representative of overlay network, VXLAN (Virtual eXtensible LAN) has provided effective support to conduct network of cloud data center and makes it more scalable. However, traditional VXLAN based on data platform is limited in some ways. Before EVPN, MAC or ARP learning in PE of operator network, based on bridge, is performed by flooding-learning pattern. Based on data platform, terminal VM learning and VTEP automatic finding has missed control platform to release reachability information of terminal VM. Therefore, with VXLAN as data platform and MP-BGP EVPN encapsulated by VXLAN as control platform, the flooding-learning pattern has been changed. It also provides learning pattern of terminal VM under remote VTEP based on control platform. A forwarding scheme between bridge and route in VXLAN overlay network has also been realized. In this scheme, MAC/IP routing is used as a control plane. Information attached by this route has been used to study knowledge of terminal virtual machines, and the communication between virtual machine has been implemented, which has replaced the flooding-learning pattern and reduced network traffic effectively.

Key words: EVPN; VXLAN; data platform; control platform

0 引 言

VXLAN(Virtual eXtensible LAN)[1-2]是一种实现网络虚拟化[3-4]的 overlay 技术,主要应用于数据中心组网,通过隧道封装在一个共享的三层 underlay 网络上实现二层扩展。由于在 overlay 网络中,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,即只需要 VTEP(VXLAN Tunnel End Point)[5]设备对 VXLAN 报文进行封装和解封装操作,接入交换机只需要学习 VTEP 设备的 MAC 或 ARP 信息。因此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。然而 VXLAN 作为数据平面技术[6],存在的局限性也很明显。Overlay 的 BUM(广播、未知单播、组播)流量需要封装到广播

收稿日期:2016-05-19 **修回日期**:2016-09-09 **网络出版时间**:2017-03-13
基金项目:教育部留学回国人员基金(BJ206004)
作者简介:钟耿辉(1991-),男,硕士研究生,研究方向为现代通信中的智能信号处理技术;唐加山,教授,硕士生导师,研究方向为现代通信中的智能信号处理技术、信道辨识与均衡、复杂网络。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20170313.1545.012.html

VXLAN 报文中,在 underlay 网络中广播转发到远端 VTEP。这样会带来一些问题:Underlay 网络需要使能广播,而有些租户不愿意在数据中心内使能广播;对 BUM 报文采用泛洪的方式导致网络扩展性有限;且现在的数据中心组网对网络提出了更多的新需求,如负载均衡、故障快速收敛等,都不能很好支持。

通过对传统网络下 VXLAN 技术缺陷的分析,引出 MP-BGP EVPN 控制平面出现的必要性,给出了具体研究控制平面的方案,并实现了控制平面方案下虚机之间的互通。

1　传统网络下的 VXLAN 技术

1.1　MAC/ARP 学习的缺陷

首先,交换机上需要学习 MAC/ARP 来指导转发。VXLAN 技术可以在一定程度上抑制 BUM 流量。为了避免 ARP 报文泛洪占用核心网络带宽,可以在本地学习并维护一个 ARP 泛洪抑制表项。以后当该 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,从而避免二次网络泛洪。ARP 泛洪抑制功能在一定程度上减少了 ARP 泛洪的次数,但在学习 MAC/ARP 时仍然需要泛洪。在虚拟机规模快速增长的云数据中心网络,这种泛洪学习模式会给数据中心网络带来巨大的压力,并不能很好地支持大规模数据中心组网。

1.2　对端自动发现及功能性限制

在数据中心组网中,一台 PE(Provider Edge router)设备可能需要发现对端的 PE 设备以相互形成 VTEP(VXALN Tunnel End-Point)对端。在传统的 VPLS(Virtual Private LAN Service,虚拟专用局域网业务)[7-8]实现中,尽管有依赖于数据平面的泛洪来进行 VTEP 远端自动发现功能,但在大规模数据中心组网中,租户更加希望能够通过其他特殊的方式来实现 VTEP 对端自动发现。而且还有一些功能如负载均衡、快速收敛、水平分割、部署简易性等都不能很好地满足要求。

2　MP-BGPEVPN 控制平面设计

2.1　EVPN 控制平面概述

在大规模数据中心组网下,为了改善传统网络下一系列缺陷,需要一种全新的技术来满足现在云计算虚拟化的网络需求。EVPN 的出现能够很好地改善上述问题。把 MP-BGP EVPN(多协议边界网关协议以太网虚拟专用网)作为 VXLAN 的控制平面,实现了网络控制平面与数据平面的分离。EVPN 定义了一种新的 BGP NLRI,称为 EVPN NLRI,并定义了 4 种路由类型:Ethernet Auto-Discovery (A-D)路由、MAC/IP Advertisement 路由、Inclusive Multicast Ethernet Tag 路由、Ethernet Segment 路由[9],还有将 IP Prefix 路由[10]也应用到 EVPN。通过这五种可达路由信息的发布、撤销和接收处理来实现控制平面。其中,第一种路由包含 A-D per ES 路由和 A-D per EVI 路由。通过该路由的通告可以发现同一个或多个 VPN 实例中具有相同 ESI 的 PE 成员,这是 EVPN 中非常重要的功能,用于实现多归属接入的负载分担、水平分割、快速收敛及减少 MAC 地址路由通告数目等各种应用场景[11]。当 PE 设备学习到本地虚机的 MAC 或 ARP 信息时,就可以通告第二种路由类型。PE 可以同时通告 MAC 路由和 MAC/IP 路由,因此可以分开管理 MAC 表和 ARP 表。如果 PE 收到 ESI 为有效的 MAC/IP 路由时,且该 PE 也连接到此 ESI,此时 PE 不更改转发状态,因为这种情况属于多归属组网,收到该 MAC/IP 路由的 PE 实际与该 MAC 是直连的,以此保证本地路由较远端路由优先。第三种路由类型用来发现数据中心[12-13]内的对端 PE,并在 PE 成员之间自动建立 Tunnel,形成 VTEP 对端。第四种路由类型用来发现连接到同一 ES 的 PE 成员,此路由只会被那些连接了同一 ES 的 PE 接收,其他 PE 不接收,可用于 DF 选举。第五种路由类型可用来在 PE 还未学到本地虚机的 ARP 信息时发布一个网段路由到远端 PE,远端 PE 收到该路由可以下发网段 FIB 表和 ARP 表,以此指导转发。MP-BGP EVPN 控制平面在解决传统 VXLAN 网络局限性的同时还带来了其他益处:

(1)MP-BGP EVPN 基于行业标准,允许多家设备供应商之间设备的互通。

(2)允许通过控制平面来学习二层和三层可达信息,可以构建更具鲁棒性和扩展性的 VXLAN overlay 网络[14]。

(3)MP-BGP VPN 技术已经经过实践考验,用来支持多租户的 VXLAN overlay 网络使其具有很好的可靠性。

(4)EVPN 协议族同时携带二层和三层可达路由信息,为 VXLAN overlay 网络提供了集成路由选择和桥接方案。

(5)在本地 VTEP 上,通过基于协议虚拟机 MAC/IP 路由的发布和配置 ARP 代答来最小化网络泛洪。

(6)为各种流量提供最佳转发路径。

(7)提供了一种机制来建立二层的多归属。

2.2　EVPN 控制平面下 MAC 和 ARP 学习的设计

MP-BGP EVPN 设计用来发布 NLRI(网络可达消息)到网络。EVPN NLRI 独一无二的一个特征是同时包含终端虚机二层和三层的可达消息。可以同时发布 EVPN VXALN overlay 网络的 MAC 和 IP 地址。这是支

持 VXALN IRB(集成路由选择和桥接)的基础。

VXLAN 是一个二层的扩展技术,因此二层 MAC 地址需要被发布。在 overlay 网络中,相同 VNI 中的终端虚机之间的流量需要被转发,这就意味着在给定的 VNI 中,VTEP 设备需要学习到其他终端虚机的 MAC 地址。通过 BGP EVPN 控制平面来发布 MAC 地址可以有效减少或者说限制 BUM 报文在 VXLAN 中的泛洪。本地 VTEP 收到路由可达消息时可以学到远端虚机的 MAC 地址。

三层的虚机 IP 地址也通过 MP-BGP EVPN 来通告,远端 PE 通过接收该路由来学习虚机的 ARP 信息,这样 VXLAN 流量就可以通过最佳路径被转发到目的终端虚机。当远端 VTEP 还没有通过 MP-BGP 学习到本地虚机 IP 路由时,本地 VTEP 可以通告 IP Prefix 路由到 VXLAN 网络中,通过学习到的网段路由转发流量到目的终端虚机。EVPN NLRI 由 BGP 携带。其中,RD 用来保证同一个路由中不同 VRF 实例的唯一性,RT 用来决定不同 VRF 实例中路由的通告和引入。而且,为了进一步简化配置,RD 和 RT 可以通过简单配置来自动生成。MP-BGP EVPN 虚机 NLRI 学习和通告的基本过程如图 1 所示。

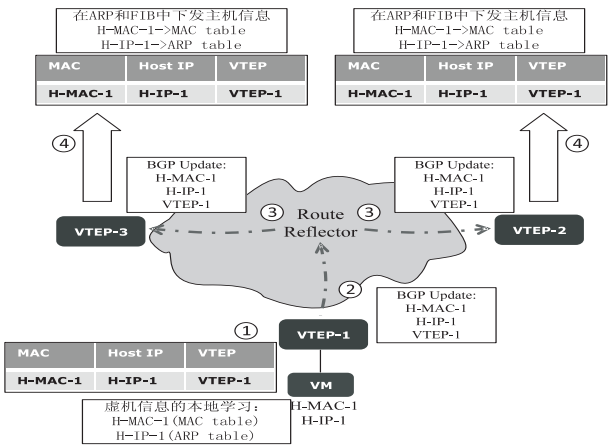

图 1 MP-BGP EVPN 可达消息的通告与学习图

(1)本地虚机学习。

VTEP-1 学习本地下挂的虚机的 MAC 地址和 IP 地址,MAC 使用标准的以太网 MAC 地址学习方式,IP 通过虚机发送的 GARP、RARP 或者虚机对网关的 ARP 请求来学习。VTEP-1 学习到本地的 MAC 和 ARP 信息,可以下发 ARP 表、MAC 表。也可以通过配置“mac-learning disable”来禁止本地 MAC 的下发,进一步减小 MAC 表规模,在大规模数据中心组网下本地虚机数量很大时效果明显。本地下发的 MAC 表、ARP 表上报给 L2VPN,再通告给 BGP,在 EVPN MAC 和 EVPN ARP 表中维护数据。

(2)EVPN 路由公告。

学习到本地虚机的 MAC 和 IP 地址后,VTEP-1 通过 MP-BGP EVPN 控制平面将虚机的信息通过 MAC/IP 路由通告给其他 VTEP,该路由可以携带下一跳、RD、ESI、ETAG、MAC、IP、VXLAN、L3VNI、R-MAC 和 RT 属性。将 NLRI 发送给 BGP,BGP 根据建立的 BGP 会话将此路由通告到所有 BGP 对端。其中,每个 VTEP 有一个唯一的系统 MAC 地址,这个 MAC 地址又称为 R-MAC 地址,用作 VXLAN 报文的内部 MAC 地址。在网络中每个租户的 VRF 实例映射到一个唯一的 L3VNI(Layer-3 VNI)。所有的 VXLAN 流量使用这个 L3VNI 封装在 VXLAN 头并提供接收 VTEP 的 VRF 信息。接收的 VTEP 使用这个 VNI 来决定这个 VRF 信息要往哪里转发,同时这个 VNI 提供数据平面的强制三层分离。

(3)远端虚机学习。

MP-BGP EVPN 用一个 EVPN 路由中 BGP 扩展团体属性来传递 export RT。当一个 EVPN VTEP 接收一个 EVPN 路由时,会用本地配置的 import RT 和接收路由的 export RT 相匹配决定是引入还是丢弃该路由。同时,通过配置“vxlan tunnel arp-learning disable”和“vxlan tunnel mac-learning disable”来避免通过数据平面学习 MAC 和 ARP 信息,只允许控制平面的 MAC 和 ARP 学习。

(4)表项下发。

VTEP-2 和 VTEP-3 收到 MAC/IP 路由,三层 RT 匹配引入该路由,就可以学到 VTEP-1 下虚机的信息,在 BGP 进程中的 ARP 和 MAC 模块维护远端虚机的 EVPN ARP 和 EVPN MAC 表项。再下发 ARP、代答 ARP 和 MAC 等表项,需要下发的表项根据配置而定。

3 EVPN 控制平面下的 VXLAN overlay 网络的研究、设计与实现

3.1 IRB(集成路由选择和桥接)方式的支持性分析

以上阐述了 MP-BGP EVPN 控制平面下的概要设计,接下来进一步研究该控制平面下数据中心组网的具体实现。VXLAN overlay 网络中的终端虚机通过同时通告二层和三层可达消息,MP-BGP EVPN 控制平面可以提供 IRB 方案。每台 VTEP 通过数据平面本地学习下挂虚机的 MAC 和 IP 地址,然后通过 MP-BGP EVPN 控制平面来通告这些信息。远端 VTEP 通过控制平面来学习远端信息。这种方法有效限制网络泛洪并为终端虚机之间的信息发布提供更好的控制能力。

3.2 IRB 方案的设计与研究

IETF EVPN 草案定义了两种 IRB 方式:对称 IRB 和非对称 IRB。重点研究实现对称 IRB 方式,这种方式具有更好的可扩展性。

通过 MP-BGP EVPN 控制平面可实现数据中心多

种组网,采用集中式网关方式时,不同 VXLAN 之间的流量以及 VXLAN 访问外界网络的流量全部由集中式 EVPN 网关处理,网关压力较大,并加剧了网络带宽资源的消耗。因此重点研究分布式网关组网,同时在网关上配置 ARP 代理功能。因此,分布式网关组网下,每台 VTEP 设备都可作为 EVPN 网关,很好地缓解了网关的压力。所谓 ARP 代理,是指虚机为请求远端虚机的 MAC 地址发送一个 ARP 请求报文时,若网关发现 ARP 请求报文的 IP 地址和网关 IP 地址在同一网段,即使网关未学到远端虚机准确的 ARP 信息,也给本地虚机返回一个 ARP 应答消息,MAC 地址为网关的 MAC 地址。然后网关再去请求虚机的 ARP 信息并学习。此时若本地虚机向远端虚机发送报文,网关发现目的 MAC 是自己,则网关处理报文后走三层转发。

如图 2 所示,当一个报文从 VM 1 发送到 VM 2 时,数据平面利用 VXLAN 封装报文,把内部的源、目的 MAC 地址分别改成本地 VTEP 和远端 egress VTEP 的 R-MAC 地址。内部的源、目的 IP 分别为本地和远端终端虚机的 IP 地址。然后把 L3VNI 或 VXLAN 值写进 VXLAN 头做封装。外部源目的 IP 地址分别为本地和远端的 VTEP 地址,外部源、目的 MAC 地址分别为本地和远端 VTEP 直连口 MAC 地址。这样 VXLAN 报文就封装好了,该报文的出接口为 Tunnel 口。egress VTEP 接收到这个 VXLAN 封装报文后,首先去掉 VXLAN 头来解封装,然后可以看到内部的报文头,因为内部目的 MAC 地址就是自己的 R-MAC 地址,接着将内层报文在 L3VNI 所对应的 VRF 中做三层转发。

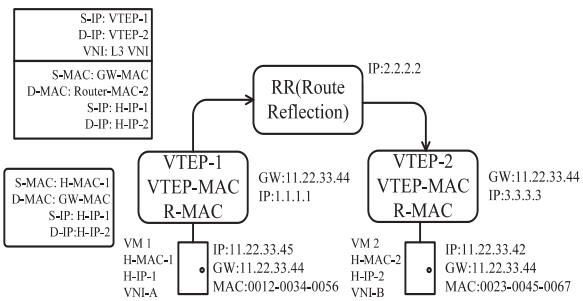

图 2　对称 IRB 中 VXLAN 报文转发图

3.3　IRB 方案的实现

图 2 完成了一个基本的分布式网关组网,并在网关上配置代理功能。简单的实现只需要两台分布式网关 VTEP1 和 VTEP2 做 leaf,两台虚拟机 VM 1 和 VM 2 以及一台 spine 用作 RR(Route Reflection),spine 只用于反射路由,因此对于 VTEP-1 和 VTEP-2 可以看作是直连的。

3.3.1　VTEP-1 配置

(1)配置 EVPN 租户的 VRF 实例,首先创建一个 VRF 实例,定义 VRF route distinguisher,并在 address-family ipv4 和 address-family evpn 下定义 VRF router target 引入和输出策略。

(2)对每个租户 VRF 实例绑定一个 Layer-3 VNI。

(3)创建一个 VSIIF 网关虚接口,关联 VSI 虚接口到用户 VRF 实例,同时配置每台 VTEP 的虚拟 MAC 地址和每个 VNI 的网关 IP 地址。研究分布式网关下 ARP 代理组网,需具有 ARP 代理和分布式网关功能。

(4)创建一个 EVI 实例,并关联 VSI 虚接口,同时设置 VXLAN-ID,并定义 VXLAN 封装报文下 EVI 实例的 route distinguisher 和 route target。另外,由于设计不下发本地虚机的 MAC 地址,需要配置去使能学习本地虚机 MAC。

(5)配置 MP-BGP 邻居,创建 BGP l2VPN EVPN 地址族,使能 EVPN,并使能 BGP 对端邻居,通常邻居是对端 VTEP 环回口地址。

(6)VTEP-2 的配置类似 VTEP-1,而且要求两个 EVPN 网关配置相同的网关虚拟 MAC 地址,一个 VNI 中配置相同的网关 IP 地址,作为默认的 IP 网关。

3.3.2　MAC/ARP 学习以及报文转发测试

(1)VM1 发送一个 GARP、RARP 或 ARP 请求报文,VTEP-1 通过数据平面学习到 VM1 的 MAC 和 IP 地址,下发 ARP 表项给驱动。因为设计分布式网关配置代理时,同网段也走三层转发,所以本地的 MAC 表项其实不用下发来指导转发,设计通过配置“mac-learning disable”命令来禁止本地 MAC 的下发,在大规模组网下减轻了 MAC 表项规模的限制。本地 ARP 表项上报给 BGP 进程,BGP 进程维护本地 EVPN ARP 表项信息,其中 ARP 信息如下所示:

viewarp

IP-address　MAC-address　VID　Interface/Link ID　Aging　Type

11.22.33.45　0012-0034-0056　0　0x0　10　Dynamic

(2)VTEP-1 学习到 VM1 的 ARP 信息后,通过 MP-BGP EVPN 控制平面将 MAC/IP 路由通告给所有 BGP 对端,通告的 MAC/IP 路由携带 Nexthop、RD、IP-address、MAC-address、ESI、Label1(VXLAN-ID)、Label2(Layer-3 VNI)、R-MAC、RT 等信息。

(3)VTEP-2 接收到 MAC/IP 路由,下一跳是 VTEP-1 地址。

(4)VTEP-2 收到 MAC/IP 路由后,通过 RIB 模块维护相关路由信息,并下发路由表,下一跳是 VTEP-1 地址。ARP 模块维护相关的远端 ARP 数据信息,并下发下一跳 ARP 表项。如下所示:

view routing-table vpn-instance 1

Destination/Mask　Proto　Pre　Cost　Next Hop　Interface

11.22.33.45/32　BGP　255　0　1.1.1.1　Vsi0

view arp

IP - address MAC - address VID Interface/Link ID Aging Type

1.1.1.1 7e1e-b58a-0100 1 Tunnel0 N/A Rule

(5)VTEP-2 将 ARP 请求在 VSI 中广播,VM2 会收到广播报文,这样 VM2 就可以学到网关和远端 VM1 的 ARP 信息并下发 ARP 表项,其中 MAC 地址为网关 MAC 地址。VM1 返回一个 ARP 应答报文,通过 ARP 应答报文,相同的原理 VTEP-2 上下发 VM2 的 ARP 表,VTEP-1 上可以下发路由表和下一跳 ARP 表。这样一个基于 MP-BGP EVPN 控制平面的分布式组网就可以按照最佳转发路径发送报文了。

(6)接下来做进一步的验证,从 VM1 上发送一条 ping 报文给 VM2,在 VM1 上可见流量转发正常,同时观察 VTEP-1 上流量转发流程正常,该方案切实可行。测试结果如图 3 和图 4 所示。

```
ping -c 1 11.22.33.42
Ping 11.22.33.42 (11.22.33.42): 56 data bytes, press CTRL_C to break
56 bytes from 11.22.33.42: icmp_seq=0 ttl=253 time=1.816 ms
--- Ping statistics for 11.22.33.42 ---
1 packet(s) transmitted, 1 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.816/1.816/1.816/0.000 ms
```

图 3 VM1 ping 报文信息图

```
*Mar 19 09:33:40:658 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN fsinput: Received a packet from interface GE2/0/1, Service Instance 0,
               Control-Word=0x0, PktLen=102.
*Mar 19 09:33:40:658 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN output: VSI gateway interface Vsi0 transmitted a packet, VSI=1,
              Link-ID=0x5000000, PktLen=98.
*Mar 19 09:33:40:658 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN output: Sent a packet on interface Tun0, KeyType=Vxlan,
              KeyID=7, PktLen=98.
*Mar 19 09:33:40:658 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN Output: Packet sent for L3VNI result: 0.
*Mar 19 09:33:40:658 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN fsinput: Packet delivered to the VSI gateway interface of VSI (0), Result=0.
*Mar 19 09:33:40:659 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN fsinput: Received a packet from interface Tun0, KeyType=Vxlan,
               KeyID=7, PktLen=98.
*Mar 19 09:33:40:659 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN output: VSI gateway interface Vsi1 transmitted a packet, VSI=0,
              Link-ID=0x0, PktLen=98.
*Mar 19 09:33:40:659 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L2VPN output: Sent a packet to interface GE2/0/1, PktLen=98.
*Mar 19 09:33:40:659 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
Packet sent result: 0.
*Mar 19 09:33:40:659 2016 L2VFW/7/PACKET: -MDC=1-Slot=2;
L3VNI fsinput: Packet delivered to the VSI interface(1413), Result=0.
```

图 4 VTEP-1 ping 报文转发信息图

(7)在 VTEP1 和 VTEP2 之间通过抓包软件可以看到 VXLAN 封装的报文。正如上述设计所描述的,VTEP1 内部的源、目的 MAC 地址分别为自己 VTEP1 的 R-MAC 地址和远端 VTEP2 的 R-MAC 地址,内部的源、目的 IP 分别为本地虚拟机 VM1 和远端虚拟机 VM2 的 IP 地址。再把 L3VNI 写进 VXLAN 头做封装,本次实验 L3VNI 值为 7。外部源、目的 IP 地址分别为本地和远端的 VTEP 地址,外部源、目的 MAC 地址分别为本地和远端 VTEP 直连口 MAC 地址,见图 5。

```
Frame 108: 148 bytes on wire (1184 bits), 148 bytes captured (1184 bits)
Ethernet II, Src: 7e:1e:b5:8a:01:18 (7e:1e:b5:8a:01:18),
              Dst: 7e:de:d8:7e:03:07 (7e:de:d8:7e:03:07)
Internet Protocol Version 4, Src: 1.1.1.1 (1.1.1.1),
                             Dst: 3.3.3.3 (3.3.3.3)
User Datagram Protocol, Src Port: 21501 (21501),
                        Dst Port: 4789 (4789)
Virtual eXtensible Local Area Network
    VXLAN Network Identifier (VNI): 7
Ethernet II, Src: 7e:1e:b5:8a:01:00 (7e:1e:b5:8a:01:00),
              Dst: 7e:de:d8:7e:03:00 (7e:de:d8:7e:03:00)
Internet Protocol Version 4, Src: 11.22.33.45 (11.22.33.45),
                             Dst: 11.22.33.42 (11.22.33.42)
Internet Control Message Protocol
```

图 5 VXLAN 封装下报文抓包信息图

4 结束语

大规模数据中心组网依赖于网络虚拟化技术,而 VXLAN 的出现解决了传统网络暴露的很多问题,在网络虚拟化技术中应用广泛。将 VXLAN 结合 EVPN 控制平面,进一步优化大规模数据中心网络,通过控制平面实现 MAC 和 ARP 的学习,抑制网络泛洪。而且分析控制平面实现传统 VXLAN 数据中心网络无法实现的功能。但是云计算还在不断发展,对大规模数据中心网络要求还会越来越高,这些技术依然需要不断进行完善和深入研究。

参考文献:

[1] 缪仕福. VXLAN 网络技术研究[J]. 科技资讯,2015(4):15-16.

[2] 刘付桂兰. 虚拟局域网新技术 VXLAN 研究[J]. 福建电脑,2014,30(11):88-89.

[3] 姚 青. 网络虚拟化的关键技术研究[D]. 南京:南京邮电大学,2013.

[4] 赵慧玲,解云鹏,史 凡. 网络虚拟化及网络功能虚拟化技术探讨[J]. 中兴通讯技术,2014,20(3):8-11.

[5] 孙铭浩. VXLAN 隧道的设计与实现[D]. 哈尔滨:哈尔滨工业大学,2014.

[6] 张届新,傅志仁,吴志明,等. VXLAN 在云数据中心组网的应用[J]. 电信科学,2015,31(9):163-169.

[7] 邓翠华,冯玉珉. 基于 MPLS 的 VPLS 技术分析[J]. 山西电子技术,2006(6):27-29.

[8] 李 山,白 桦. 构建基于 VPLS 城域网精品方案[J]. 现代电信科技,2008,38(3):54-58.

[9] Uttaro J,Drake J,Henderickx W. BGP MPLS-Based Ethernet VPN[EB/OL]. (2015-10-14)[2016-04-27]. https://datatracker. ietf. org/doc/rfc7432/.

[10] Rabadan J,Henderickx W,Palislamovic S,et al. IP prefix advertisement in EVPN[EB/OL]. (2015-09-14). [2016-04-27]. https://datatracker. ietf. org/doc/draft-ietf-bess-evpn-prefix-advertisement/.

[11] 何晓明,唐 宏,刘志华,等. 以太网 VPN 技术在云数据中心互联应用的研究[J]. 电信科学,2012,28(8):138-144.

[12] 朱明明,夏寅贲,徐小飞. 基于 SDN 的数据中心网络研究[J]. 邮电设计技术,2014(3):23-29.

[13] 宋文文,李 莉. 云数据中心大二层网络技术研究[J]. 中国教育网络,2013(12):34-36.

[14] 马文杰. Overlay Network 技术在云计算数据中心中的应用[J]. 河南科技,2014(11):6-7.