(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2015 年 7 月 9 日 (09.07.2015)

WIPO | PCT

(10) 国际公布号
WO 2015/100656 A1

(51) 国际专利分类号:
H04L 12/701 (2013.01)

(21) 国际申请号: PCT/CN2013/091202

(22) 国际申请日: 2013 年 12 月 31 日 (31.12.2013)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: **吴天议 (WU, Tianyi)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

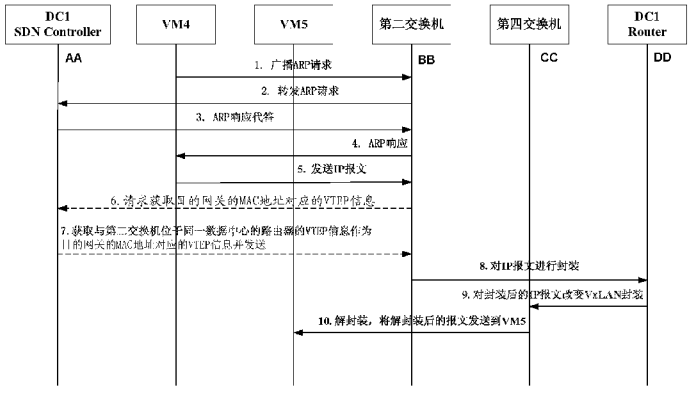
本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

WO 2015/100656 A1

(54) **Title**: METHOD AND DEVICE FOR IMPLEMENTING VIRTUAL MACHINE COMMUNICATION

(54) **发明名称** : 一种实现虚拟机通信的方法和装置



1 ARP REQUEST IS BROADCAST
2 ARP REQUEST IS FORWARDED
3 ARP RESPONSE IS PICKED UP
4 ARP RESPONSE
5 IP PACKET IS SENT
6 REQUEST TO OBTAIN VTEP INFORMATION CORRESPONDING TO MAC ADDRESS OF DESTINATION GATEWAY
7 VTEP INFORMATION OF ROUTER LOCATED IN SAME DATA CENTRE AS SECOND SWITCH IS OBTAINED TO BE VTEP INFORMATION CORRESPONDING TO MAC ADDRESS OF DESTINATION GATEWAY, AND IS SENT
8 ENCAPSULATION IS PERFORMED ON IP PACKET
9 ALTERATION VXLAN ENCAPSULATION IS PERFORMED ON ENCAPSULATED IP PACKET
10 DE-ENCAPSULATION IS PERFORMED, AND DE-ENCAPSULATED PACKET IS SENT TO VM5
AA DC1 SDN CONTROLLER
BB SECOND SWITCH
CC FOURTH SWITCH
DD DC1 ROUTER

图 5 / **Fig. 5**

(57) **Abstract**: Provided are a method and device. A first switch receives an ARP response from an SDN controller, the ARP response carrying a MAC address of a destination gateway. According to the MAC address of the destination gateway, the first switch obtains VTEP information corresponding to the MAC address of the destination gateway, a router corresponding to the VTEP information being located in a first data center. According to the VTEP information, the first switch sends an IP packet to the router corresponding to the VTEP information, so that the router corresponding to the VTEP information sends the IP packet to a second virtual machine via a tunnel between the router and a second switch. Therefore the SDN controller picks up, thereby reducing the transmission bandwidth occupancy of a broadcast packet. The packet only passes through the router of the first data center, thereby reducing a roundabout route of the packet between data centers.

(57) **摘要**: 本发明实施例提供的方法和装置，第一交换机从

[见续页]

SDN 控制器接收 ARP 响应，所述 ARP 响应携带目的网关的 MAC 地址，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心；所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。因此 SDN 控制器代答，减少了广播报文对传输带宽的占用，并且只经过了第一数据中心的路由器，减少了报文在数据中心间的迂回。

一种实现虚拟机通信的方法和装置

技术领域

本发明实施例涉及网络通信领域，特别涉及一种实现虚拟机通信的方法和装置。

背景技术

现有技术应用虚拟可扩展局域网（VxLAN， Virtual extensible LAN）技术扩展网络虚拟化，用以得到足够数量的虚拟网络，来满足用户使用。

每一个数据中心可能运行多个虚拟网络的虚拟机(VM， Virtual Machine)。当一个数据中心的 VM 与另一数据中心的 VM 通信时，它会向第一跳路由器发出广播地址解析协议(ARP, Address Resolution Protocol)请求，以广播获取目的网关的 MAC 地址，然后 ARP 请求被广播到各 VM。获取到目的网关的 MAC 地址后，通过目的网关的 MAC 地址，将报文发送到各个数据中心的路由器，然后到达目的 VM，从而实现 VM 之间的通信。但是这种 ARP 广播到各 VM 的方式，会占用大量传输带宽；而且报文会通过各数据中心的路由器才能到达目的 VM，在数据中心间迂回比较严重。

发明内容

本发明实施例提出了一种实现虚拟机通信的方法和装置，以减少广播报文对传输带宽的占用，和减少报文在数据中心间的迂回。

第一方面，本发明实施例提出了一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，该方法包括：所述第一交换机从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址；

所述第一交换机向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求；

所述第一交换机从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带

所述目的网关的媒体接入控制 MAC 地址;

所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心;

所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

结合第一方面，在第一种可能的实现方式中，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括：所述第一交换机向所述 SDN 控制器发送请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息；接收所述 SDN 控制器发送的所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第一方面，在第二种可能的实现方式中，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括：所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第一方面的第二种可能的实现方式中，在第三种可能的实现方式中，所述第一交换机查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第一方面的第三种可能的实现方式中，在第四种可能的实现方式中，当所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述第一交换机根据以下至少之一：路由器的负载和调度策略，选择所述目

的网关的 MAC 地址对应的 VTEP 信息。

结合第一方面或者第一方面的第一种至第四种任意一种可能的实现方式，在第五种可能的实现方式中，所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机具体包括:

所述第一交换机对从所述第一虚拟机接收的 IP 报文进行虚拟可扩展局域网 VxLAN 封装，其中所述 VxLAN 封装的目的 IP 地址为获取的所述 VTEP 信息，虚拟网络标识 VNI 为所述第一子网的标识，所述 IP 报文的目的 MAC 地址为所述目的网关的 MAC 地址;

所述 VTEP 信息对应的路由器，将目的 IP 地址改为所述第二交换机的 IP 地址，VNI 改为所述第二子网的标识，目的 MAC 地址改为所述第二虚拟机的 MAC 地址，然后将改变 VxLAN 封装的 IP 报文，发送给所述第二交换机;

所述第二交换机对接收的改变 VxLAN 封装的 IP 报文，进行 VxLAN 解封装后，发送给所述第二虚拟机。

第二方面，本发明实施例提出了一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，该方法包括: 所述第一数据中心的软件定义网络 SDN 控制器从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

所述 SDN 控制器根据所述目的网关的 IP 地址，获取所述目的网关的媒体接入控制 MAC 地址，并向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，以使所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心，使所述

VTEP 信息对应的路由器将 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

结合第二方面，在第一种可能的实现方式中，所述 SDN 控制器接收所述第一交换机发送的请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息;

所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第二方面的第一种可能的实现方式中，在第二种可能的实现方式中，所述 SDN 控制器查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第二方面的第二种可能的实现方式中，在第三种可能的实现方式中，当所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述 SDN 控制器根据以下至少之一: 路由器的负载和调度策略，选择所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第二方面或者第二方面的第一种至第三种任意一种可能的实现方式，在第四种可能的实现方式中，若所述第二虚拟机位于第二数据中心，则还包括:

所述第一数据中心的 SDN 控制器从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、所述第二子网的虚拟网络标识 VNI，和所述第二交换机的 VTEP 信息，以在所述第一数据中心的 SDN 控制器的控制下，使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所

述第二交换机的隧道发送给所述第二虚拟机。

第三方面，本发明实施例提出了一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与所述第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，所述第一交换机包括：接收模块，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址；

发送模块，用于向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求；

所述接收模块还用于从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的网关的媒体接入控制 MAC 地址；

获取模块，用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心；

所述发送模块，还用于根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

结合第三方面，在第一种可能的实现方式中，所述获取模块用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为：

所述获取模块用于向所述 SDN 控制器发送请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第三方面，在第二种可能的实现方式中，所述获取模块用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为：

所述获取模块用于根据所述目的网关的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第三方面的第二种可能的实现方式中，在第三种可能的实现方式中，所述获取模块用于查询所述第一交换机保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为：

所述获取模块用于根据所述目的网关的 MAC 地址，查询所述第一交换机保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

第四方面，本发明实施例提出了一种实现虚拟机通信软件定义网络 SDN 控制器，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于所述第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，所述 SDN 控制器包括：

接收模块，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址；

处理模块，用于根据所述目的网关的 IP 地址，获取所述目的网关的媒体接入控制 MAC 地址；

发送模块，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，以使所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心，使所述 VTEP 信息对应的路由器将 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

结合第四方面，在第一种可能的实现方式中，所述接收模块还用于接收所述第一交换机发送的请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息；

所述处理模块还用于根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第四方面的在第一种可能的实现方式中，在第二种可能的实现方式中，所述处理模块用于查询所述 SDN 控制器保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为：

所述处理模块用于根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第四方面的第二种可能的实现方式中，在第三种可能的实现方式中，当所述处理模块用于根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述处理模块用于根据以下至少之一：路由器的负载和调度策略，选择所述目的网关的 MAC 地址对应的 VTEP 信息。

结合第四方面或者第四方面的第一种至第三种任意一种可能的实现方式，在第四种可能的实现方式中，若所述第二虚拟机位于第二数据中心，则所述处理模块还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、第二子网的虚拟网络标识 VNI，和所述第二交换机的 VTEP 信息。

第五方面，本发明实施例提出了一种实现虚拟机通信的交换机，包括：

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行如第一方面或第一方面的第一种至第五种任意一种可能的实现方式。

第六方面，本发明实施例提出了一种实现虚拟机通信软件定义网络 SDN

控制器，包括处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行第二方面或第二方面的第一种至第四种任意一种可能的实现方式。

本发明实施例，第一交换机从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心；所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。因此 SDN 控制器代答，减少了广播报文对传输带宽的占用，并且只经过了第一数据中心的路由器，而不像现有技术可能同时经过第一数据中心的路由器和第二数据中心的路由器，减少了报文在数据中心间的迂回。

第七方面，本发明实施例提出了一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，该方法包括：所述第一交换机从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

所述第一交换机根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址；

根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址；

所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

结合第七方面，在第一种可能的实现方式中，所述第一交换机根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址具体包括:

所述第一交换机向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求，所述 ARP 请求携带所述目的虚拟机的 IP 地址;

所述第一交换机从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。

结合第七方面或第七方面的第一种可能的实现方式中，在第二种可能的实现方式中，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机向所述 SDN 控制器发送请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息;

接收所述 SDN 控制器发送的所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

结合第七方面或第七方面的第一种可能的实现方式中，在第三种可能的实现方式中，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机根据所述目的虚拟机的 MAC 地址，查询本地保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

根据第七方面的第三种可能的实现方式中，在第四种可能的实现方式中，所述第一交换机从所述 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、和所述第二交换机的 VTEP 信息，以通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

第八方面，本发明实施例提出了一种实现虚拟机通信的方法，应用于

至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，该方法包括：所述第一数据中心的软件定义网络 SDN 控制器从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

所述 SDN 控制器根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址，并向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址，以使所述第一交换机根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址，使所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

结合第八方面，在第一种可能的实现方式中，所述 SDN 控制器接收所述第一交换机发送的请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息；

所述 SDN 控制器根据所述目的虚拟机的 MAC 地址，查询本地保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

结合第八方面或第八方面的第一种可能的实现方式中，在第二种可能的实现方式中，所述第一数据中心的 SDN 控制器从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息，并将所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息下发给所述第一交换机。

第九方面，本发明实施例提出了一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与所述第一交换机通信，所述第二虚拟机位于第二数据中

心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，所述交换机包括：

接收模块，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

处理模块，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址；

发送模块，用于通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

结合第九方面，在第一种可能的实现方式中，所述处理模块用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的 MAC 地址具体为：

所述处理模块用于向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求，所述 ARP 请求携带所述目的虚拟机的 IP 地址，从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。

结合第九方面或第九方面的第一种可能的实现方式中，在第二种可能的实现方式中，所述处理模块用于根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体为：

所述处理模块用于向所述 SDN 控制器发送请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

结合第九方面或第九方面的第一种可能的实现方式中，在第三种可能的实现方式中，所述处理模块用于根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体为：

所述处理模块用于根据所述目的虚拟机的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

结合第九方面的第三种可能的实现方式中，在第四种可能的实现方式中，所述处理模块还用于从所述第一数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、和所述第二交换机的 VTEP 信息。

第十方面，本发明实施例提出了一种实现虚拟机通信软件定义网络 SDN 控制器，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于所述第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，所述 SDN 控制器包括:

接收模块，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址;

获取模块，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址;

发送模块，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址，以使所述第一交换机根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址，使所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

结合第十方面，在第一种可能的实现方式中，所述接收模块还用于接收所述第一交换机发送的请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息;

所述获取模块还用于根据所述目的虚拟机的 MAC 地址，查询所述 SDN 控制器保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

根据第十方面或第十方面的第一种可能的实现方式中，在第二种可能

的实现方式中，所述获取模块还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息；

所述发送模块还用于将所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息下发给所述第一交换机。

第十一方面，本发明实施例提出了一种实现虚拟机通信的交换机，包括：

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行如第七方面或第七方面的第一种至第四种任意一种可能的实现方式。

第十二方面，本发明实施例提出了一种实现虚拟机通信软件定义网络 SDN 控制器，包括处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行第八方面或第八方面的第一种至第二种任意一种可能的实现方式。

本发明实施例，所述第一交换机根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的 MAC 地址；根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址；所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。因此 SDN 控制器代答，减少了广播报文对传输带宽的占用，即不用经过第一数据中心的路由器，也不用经过第二数据中心的路由器，减少了报文在两个数据中心间的迂回。

附图说明

为了更清楚地说明本发明实施例的技术方案，下面将对现有技术或实施

例中所需要使用的图作简单地介绍，显而易见地，下面描述中的图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些图获得其他的图。

图 1a 是根据本发明实施例提供的一种实现虚拟机通信的系统架构图；

图 1b 是根据本发明实施例提供的一种实现虚拟机通信的系统架构图；

图 2 是根据本发明实施例提供的一种配置方法示意图；

图 3 是根据本发明实施例提供的一种配置方法流程图；

图 4a 是根据本发明实施例提供的 VM 增加时的流程图；

图 4b 是根据本发明实施例提供的 VM 删除时的流程图；

图 5 是根据本发明实施例提供的一种实现 VM 通信的流程图；

图 6 是根据本发明实施例提供的一种实现 VM 通信的流程图；

图 7 是根据本发明实施例提供的一种实现 VM 通信的流程图；

图 8 是根据本发明实施例提供的一种交换机结构示意图；

图 9 是根据本发明实施例提供的一种控制器结构示意图；

图 10 是根据本发明实施例提供的一种交换机结构示意图；

图 11 是根据本发明实施例提供的一种控制器结构示意图；

图 12 是根据本发明实施例提供的一种硬件结构示意图。

具体实施方式

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述。

本发明实施例提出了一种实现虚拟机通信的方法、和系统，能够减少广播报文对传输带宽的占用，和减少报文在数据中心间的迂回。

本发明实施例的系统架构

本发明实施例提供的实现虚拟机通信的方法可以在如图 1a 和图 1b 所示的系统架构中实现。图 1a 和图 1b 仅是示例性说明，并不限定具体的组网方式。

虚拟可扩展局域网（VxLAN，Virtual extensible LAN）是在3层网络上借助封装来实现网络虚拟化的。具体地，VxLAN封装借助VxLAN隧道端点（VTEP，VxLAN Tunnel End Point）来完成。该端点可以通过一台hypervisor交换机或物理接入交换机来提供的。VTEP之间建立隧道(Tunnel)，通过该隧道传输虚拟网络的数据。

如图1a所示，该系统包括第一数据中心（DC1），第二数据中心（DC2）。DC1内包括虚拟机(VM，Virtual Machine) VM1、VM2、VM3和VM4。DC2内包括虚拟机VM5、VM6和VM7。

网络管理器（NetworkManager），部署于任意数据中心中，可实现与各数据中心的软件定义网络控制器（SDN Controller，Software Defined Network Controller）通信。网络管理器的功能，具体参见图2和图3的描述。

每个数据中心可以各部署一个SDN控制器，各个数据中心的SDN控制器互相通信。具体的，SDN控制器集群之间可以是Full-Mesh全互联，也可以是部分相连。也可以多个数据中心，部署一个SDN控制器，本发明对此不作限定。

位于DC1内的VM1属于第一子网（也可以表示为子网1），该子网的虚拟网络标识（VNI，Virtual Network Identifier)以11111为例。第一子网在DC1的网关的IP地址以192.168.1.1为例，媒体接入控制(MAC，Media Access Control）地址为MAC1。第一子网在DC2的网关的IP地址以192.168.1.1为例，MAC地址为MAC1。第一子网在DC1的网关的IP地址和第一子网在DC2的网关的IP地址可以相同或者不同。第一子网在DC1的网关的MAC地址和第一子网在DC2的网关的MAC地址可以相同或者不同。在IP地址和MAC地址相同的情况下（也可以称为双活），本方案的效果更明显。

VM1通过第一交换机（Switch）与第一子网外的VM通信。这里的交换机可以是虚拟的，也可以是物理的。本发明对此不作限定。第一交换机的IP

地址以 10.10.10.2 为例。第一交换机在隧道中以 VTEP1 表示。VM1 的 IP 地址以 192.168.1.2 为例，MAC 地址表示为 MAC-VM1。

位于 DC1 内的 VM2、VM3、VM4 和位于 DC2 内的 VM7 属于同一子网，称为第二子网（也可以表示为子网 2）。该子网的 VNI 以 22222 为例。第二子网在 DC1 的网关的 IP 地址以 192.168.2.1 为例，MAC 地址为 MAC2。第二子网在 DC2 的网关的 IP 地址以 192.168.2.1 为例，MAC 地址为 MAC2。第二子网在 DC1 的网关的 IP 地址和第二子网在 DC2 的网关的 IP 地址可以相同或者不同。第二子网在 DC1 的网关的 MAC 地址和第二子网在 DC2 的网关的 MAC 地址可以相同或者不同。

VM2、VM3 和 VM4 通过第二交换机与第一子网外的 VM 通信。这里的交换机可以是虚拟的，也可以是物理的。第二交换机的 IP 地址以 10.10.10.3 为例,在隧道中以 VTEP2 表示。VM7 通过第三交换机与第二子网外的 VM 通信。第三交换机的 IP 地址以 11.11.11.3 为例，在隧道中以 VTEP3 表示。VM2 的 IP 地址以 192.168.2.2 为例，MAC 地址表示为 MAC-VM2。VM3 的 IP 地址以 192.168.2.3 为例，MAC 地址表示为 MAC-VM3。VM4 的 IP 地址以 192.168.2.4 为例，MAC 地址表示为 MAC-VM4。VM7 的 IP 地址以 192.168.2.7 为例，MAC 地址表示为 MAC-VM7。

位于 DC2 内的 VM5 和 VM6 属于第三子网（也可以表示为子网 3），该子网的 VNI 以 33333 为例。第三子网在 DC1 的网关的 IP 地址以 192.168.3.1 为例，MAC 地址为 MAC3。第三子网在 DC2 的网关的 IP 地址以 192.168.3.1 为例，MAC 地址为 MAC3。第三子网在 DC1 的网关的 IP 地址和第三子网在 DC2 的网关的 IP 地址可以相同或者不同。第三子网在 DC1 的网关的 MAC 地址和第三子网在 DC2 的网关的 MAC 地址可以相同或者不同。

VM5 和 VM6 通过第四交换机与第三子网外的 VM 通信。这里的交换机可以是虚拟的，也可以是物理的。第四交换机的 IP 地址以 11.11.11.4 为例，在隧道中以 VTEP4 表示。VM5 的 IP 地址以 192.168.3.5 为例，MAC 地址表示为

MAC-VM5。VM6 的 IP 地址以 192.168.3.6 为例，MAC 地址表示为 MAC-VM6。

在 DC1 内，部署一台或多台路由器（router），以部署路由器 1 为例说明。路由器 1 在隧道中以 VTEP5 表示，IP 地址以 10.10.10.5 为例。类似地，在 DC2 内，部署一台或多台路由器，以部署路由器 2 为例说明。路由器 1 在隧道中以 VTEP6 表示，IP 地址以 11.11.11.6 为例。

部署多个路由器的情况，可以参照图 1b。比如路由器 1 和路由器 3 可以组成集群，从而可以消除单路由器的瓶颈，实现路由器间的弹性伸缩。本发明对部署多少个路由器不做限定，当然也可以大于 2 个。其他信息与图 1a 类似，不再详述。

参见图 2，描述网络管理器对任一 SDN 控制器的配置过程，包括但不限于以下三条命令：AddRouterInfo、AddSwitchInfo、和 ADDPeerSDNCtrl。

其中 AddRouterInfo，用来配置路由器的 IP 地址和 DC 标识信息，以表 1 为例说明：

表 1

项目	类型	描述
RouterIPAddress	String	路由器的 IP 地址
DCName	String	数据中心的标识

AddSwitchInfo，用来配置交换机的 VTEP 信息，以表 2 为例说明：

表 2

项目	类型	描述
VTEPIPAddress	String	交换机作为 VTEP 的 IP 地址
DCName	String	数据中心的标识

ADDPeerSDNCtrl，用来配置对端 SDN 控制器的 IP 地址，对端 DC 的示识，以表 3 为例说明：

表 3

项目	类型	描述
PeerSDNCtrlIPAddress	String	对端 SDN 控制器的 IP 地址
PeerDCName	String	对端数据中心的标识

可选地，SDN 控制器把表 1-表 3 的各项信息，以流表的形式保存。

然后，结合图 3，描述网络管理器通过 SDN 控制器对各子网进行配置的过程。

A1)、NetworkManager 向 DC1 内的 SDN 控制器添加子网与路由器的关联关系，具体的添加命令以 AddRouterInterface 为例。以添加子网 1 为例，子网 1 的 VNI 为 11111，该子网在 DC1 的网关的 IP 地址为 192.168.1.1/24，24 表示子网掩码为 255.255.255.0，MAC 地址为 MAC1。

可选地，SDN 控制器把添加的信息以流表的形式保存，下面类似地，SDN 控制器会把其他添加的信息，也以流表的形式保存。

A2)、DC1 内的 SDN 控制器向 DC1 内的路由器 1 添加子网与路由器的关联关系，具体的添加命令以 AddRouterGwIP 为例。

A3)、在子网 1 中创建了 VM 的情况下，SDN 控制器向本数据中心的 Switch 添加网关的 MAC 地址对应的 VTEP 信息，具体的添加命令以 AddGwMacVTEP 为例。以图 1a 和图 1b 中信息为例，如果子网 1 中 VM1 发送报文，若该报文对应的网关的 MAC 地址为 MAC1，则将该报文发送给 VTEP 为 10.10.10.5 的路由器，其中 VNI 确定路由器上的处理接口。本数据中心的 Switch 把网关的 MAC 地址对应的 VTEP 信息以流表的形式保存。

B1)、NetworkManager 向 DC2 内的 SDN 控制器添加子网与路由器的关联关系，具体的添加命令以 AddRouterInterface 为例。以添加子网 1 为例，子网 1 的 VNI 为 11111，该子网在 DC2 的网关的 IP 地址为 192.168.1.1/24，24 表示子网掩码为 255.255.255.0，MAC 地址为 MAC1。

B2)、DC2 内的 SDN 控制器向 DC2 内的路由器 2 添加子网与路由器的关联关系，具体的添加命令以 AddRouterGwIP 为例。

B3)、在子网 1 中创建了 VM 的情况下，SDN 控制器向本数据中心的 Switch 添加网关的 MAC 地址对应的 VTEP 信息，具体的添加命令以 AddGwMacVTEP 为例。以图 1a 和图 1b 中信息为例，如果子网 1 中 VM1 发送报文，若该报文对应的网关的 MAC 地址为 MAC1，则将该报文发送给 VTEP 为 11.11.11.6 的路由器，其中 VNI 确定路由器上的处理接口。

步骤 B1）、B2）、B3）和步骤 A1）、A2）、A3）之间没有先后顺序之分。类似地，网络控制器和 SDN 控制器可以添加其他子网与路由器的关联关系。

在图 3 所述的配置过程中，子网 1 在 DC1 的网关的 IP 地址和在 DC2 内的网关的 IP 地址相同，子网 1 在 DC1 的网关的 MAC 地址和在 DC2 内的网关的 IP 地址相同，但是路由器的 VTEP 信息（即指路由器的 VTEP 的 IP 地址）不同。

下面结合图 4a，描述如何创建 VM 以及实现信息同步的过程。以创建 VM2 为例说明。

A1)、向 DC1 的第二交换机发起创建虚拟机 VM2 的命令，具体的创建命令以 Create VM on Switch 为例。

A2)、DC1 的第二交换机创建了 VM2，然后将 VM2 虚拟端口与对应的 VTEP 信息通知到 DC1 的 SDN Controller，具体的通知命令以 VM vPort VTEP notify 为例。对 VM2，其 MAC 地址为 MAC-VM2，IP 地址为 192.168.2.2。属于第二子网，VNI 为 22222。VM2 对应的 VTEP 信息为 10.10.10.3。

SDN 控制器可以根据端口信息，获得该端口的 VNI。

A3)、DC1 的 SDN Controller 将 VM2 虚拟端口与对应的 VTEP 信息通知给 DC2 的 SDNController，具体的通知命令以 VM vPort VTEP notify 为例。

A4.1)、DC1 的 SDN Controller 将 VM2 的信息添加到 DC1 的路由器中，具体的添加命令以 AddVMMacVTEP 为例，包括 VM2 的 IP=192.168.2.2，

MAC=MAC-VM2，VNI=222222，VTEP=10.10.10.3。

A4.2)、DC2 的 SDN Controller 将 VM2 的信息添加到 DC2 的路由器中，具体的添加命令以 AddVMMacVTEP 为例，包括 VM2 的 IP=192.168.2.2，MAC=MAC-VM2，VNI=222222，VTEP=10.10.10.3。

A3）和 A4.1）之间没有先后顺序之分，A4.1）与 A4.2）之间也没有先后顺序之分。

A5)、DC1 的 SDN 控制器向 DC1 的 Switch 添加网关的 MAC 地址对应的 VTEP 信息。具体的添加命令以 AddGwMacVTEP 为例。以 VM2 为例，如果子网 2 中 VM2 发送报文，该报文对应的网关的 MAC 地址为 MAC2，则将该报文发送给 VTEP 为 10.10.10.5 的路由器。VNI=22222，用于确定路由器上的处理接口。

如图 3 的 B3)和 A3)所示，若在子网中创建了 VM 的情况下，SDN 控制器向本数据中心的 Switch 添加网关的 MAC 地址对应的 VTEP 信息。若在图 3 创建子网的时候，VM 仍未建立。那么可以通过图 4a 的 A5)向本数据中心的 Switch 添加网关的 MAC 地址对应的 VTEP 信息。

A5)与 A3）-A4.2）之间，无先后顺序之分。

下面结合图 4b，描述如何删除 VM 以及实现信息同步的过程。以删除 VM2 为例说明。

B1)、向 DC1 的第二交换机发起删除虚拟机 VM2 的命令，具体的删除命令以 delete VM on Switch 为例。

B2)、DC1 的第二交换机删除了 VM2，然后通知 DC1 的 SDN Controller 将 VM2 的虚拟端口与对应的 VTEP 信息删除，具体的删除命令以 delete VM vPort notify 为例。

对应图 4a 的 A5)，网关的 MAC 地址对应的 VTEP 信息也可以在步骤 B2)删除

B3)、DC1 的 SDN Controller 将删除 VM2 的信息通知给 DC2 的 SDNController，具体的通知命令以 delete VM vPort notify 为例。

B4.1)、DC1 的 SDN Controller 将 DC1 的路由器中的 VM2 的 MAC 地址及 VTEP 信息删除，具体的删除命令以 DeleteRouterMacVTEP 为例，包括 VM2 的 IP=192.168.2.2，MAC=MAC-VM2，VNI=222222。

B4.2）、DC2 的 SDN Controller 将 DC2 的路由器中的 VM2 的 MAC 地址及 VTEP 信息删除，具体的删除命令以 DeleteRouterMacVTEP 为例，包括 VM2 的 IP=192.168.2.2，MAC=MAC-VM2，VNI=222222。

B3）和 B4.1）之间没有先后顺序之分，B4.1）与 B4.2）之间也没有先后顺序之分。

虚拟机迁移可以拆解为 VM 创建和 VM 删除两个操作，具体可以参见图 4a 和图 4b 描述的流程，不再赘述。

经过图 2-图 4 的配置过程，DC1 的 SDN 控制器获得了 DC1 中路由器 1 的信息、第一交换机的信息、第二交换机的信息， DC1 中各子网与路由器的关联关系、以及 DC1 中各虚拟机的 MAC 地址和 VTEP 信息，同时 DC1 的 SDN 控制器也获得了 DC2 的 SDN 控制器的信息，以及 DC2 的 SDN 控制器控制下的虚拟机的 MAC 地址和 VTEP 的信息，可以是 DC2 的 SDN 控制器控制下的部分虚拟机的 MAC 地址和 VTEP 的信息，也可以是 DC2 的 SDN 控制器控制下的全部虚拟机的 MAC 地址和 VTEP 的信息。类似地，DC2 的 SDN 控制器获得了 DC2 中路由器 2 的信息、第三交换机的信息、第四交换机的信息， DC2 中各子网与路由器的关联关系、以及 DC2 中各虚拟机的 MAC 地址和 VTEP 信息，同时 DC2 的 SDN 控制器也获得了 DC1 的 SDN 控制器的信息，以及 DC1 的 SDN 控制器控制下的虚拟机的 MAC 地址和 VTEP 的信息，可以是 DC1 的 SDN 控制器控制下的部分虚拟机的 MAC 地址和 VTEP 的信息，也可以是 DC1 的 SDN 控制器控制下的全部虚拟机的 MAC 地址和 VTEP 的信息。

在图 1a 和图 1b 所述的系统架构上，如何实现虚拟机通信，具体参见下面实施例的描述。

实现虚拟机通信的方法

下面介绍本发明实施例提供的实现虚拟机通信的方法，该方法可以在前述实施例的系统架构上实现。

如图 5 所示的实现虚拟机通信的方法实施例，用于实现位于数据中心 DC1 的虚拟机 VM4 与位于数据中心 DC2 的虚拟机 VM5 通信，VM4 位于第二子网，VM5 位于第三子网。VM4 通过第二交换机与第二子网外的 VM 通信，VM5 通过第四交换机与第三子网外的 VM 通信。该方法包括以下步骤：

1）VM4 需要与 VM5 通信。VM4 查询路由表，发现 VM4 与 VM5 位于不同子网，则 VM4 使用第二子网的默认网关 192.168.2.1 作为目的网关。VM4 向第二交换机发送广播地址解析协议(ARP, Address Resolution Protocol)请求，以广播获取目的网关的 MAC 地址，该 ARP 请求携带所述目的网关的 IP 地址。

2）第二交换机接收所述 ARP 请求，并将所述 ARP 请求转发给第一数据中心的 SDN 控制器。

3）第一数据中心的 SDN 控制器接收所述 ARP 请求，根据所述 ARP 请求携带的目的网关的 IP 地址，查询本地保存的流表，获取目的网关的 MAC 地址为 MAC2。SDN 控制器向所述第二交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址。并且，此 ARP 响应相当于对广播 ARP 请求进行了代答。

可选地，该 ARP 请求还可以携带接收 ARP 响应的第二交换机的端口信息， SDN 控制器根据所述端口信息，查询流表，获得所述端口信息对应的 VNI。获取所述目的网关的 MAC 地址具体包括：所述 SDN 控制器根据所述目的网关的 IP 地址和端口信息对应的 VNI，获取所述目的网关的 MAC 地址。

可选地，该 ARP 请求还可以携带接收 ARP 响应的第二交换机的端口的 VNI。获取所述目的网关的 MAC 地址具体包括：所述 SDN 控制器根据所述目的网关的 IP 地址和第二交换机的端口的 VNI，获取所述目的网关的 MAC 地址。

4）第二交换机接收所述 ARP 响应，并将所述 ARP 响应转发给 VM4。

5）VM4 向第二交换机发送 IP 报文，该 IP 报文的目的 MAC(DestMAC)为 MAC2，目的 IP（DestIP）为 VM5 的 IP 地址：192.168.3.5，源 MAC 为 VM4 的 MAC（以 MAC-VM4 表示），源 IP 为 VM4 的 IP 地址：192.168.2.4。

经过图 3、图 4a 和图 4b 的配置过程，第二交换机可能缓存了 MAC2 对应的 VTEP 信息。如果第二交换机发现 MAC2 对应的 VTEP 信息存在，则直接执行步骤 8）。如果第二交换机发现 MAC2 对应的 VTEP 信息不存在，则执行步骤 6）和步骤 7）。

6）第二交换机向本数据中心的 SDN 控制器发送请求，以请求获取目的网关的 MAC 地址对应的 VTEP 信息。

7）SDN 控制器查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第二交换机发送所述目的网关的 MAC 地址对应的 VTEP 信息。

当 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上的 VTEP 信息，则以与所述第二交换机位于同一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

在本实施例中，根据 MAC2，如果获取了 VTEP5（IP 为 10.10.10.5）和 VTEP6（IP 为 11.11.11.6）两条 VTEP 信息，则选择与第二交换机位于同一数据中心的 VTEP5 作为目的网关的 MAC 地址对应的 VTEP。

当第一数据中心内，与目的网关的 MAC 地址对应的 VTEP 信息有两条或以上时，如图 1b 所示的多路由器的方案，SDN 控制器根据以下至少之一：路由器的负载和调度策略，为 VM4 选择合适的 VTEP，把选择的合适的 VTEP 的 IP 地址发送给所述第二交换机。路由器的负载包括但不限于路由器的流量、和路由器的 CPU 负荷。调度策略包括但不限于 RondRobin、最小负载优先、和最大负载优先。

8）第二交换机获取目的网关的 MAC 地址对应的 VTEP 信息，然后将从 VM4

接收的 IP 报文 VxLAN 封装后发送给该 VTEP 信息对应的路由器。其中 VxLAN 封装的 IP 报文的源 IP 为第二交换机的 IP 地址 10.10.10.3，目的 IP 为获取的 VTEP 信息（10.10.10.5），VNI 为 22222。

如果第二交换机查询本地的流表，发现 MAC2 对应的 VTEP 信息不存在，则通过上述的步骤 6）和 7），获取目的网关的 MAC 地址对应的 VTEP 信息；或者，

第二交换机查询本地的流表，获取目的网关的 MAC 地址对应的 VTEP 信息。具体的，获取了两条或以上的 VTEP 信息时，则以与所述第二交换机位于同一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。同一数据中心有多个路由器的时候，采取步骤 7）类似的路由器的负载和调度策略进行选择，不再详细描述。

9）所述第一数据中心的路由器，对接收的报文改变 VxLAN 封装，通过第一数据中心的路由器与第四交换机之间的隧道，发送给第四交换机。

经过图 3、图 4a 和图 4b 的配置过程，第一数据中心的路由器缓存了第二数据中心的虚拟机的 MAC 地址和 VTEP 信息，因此可以直接在第一数据中心的路由器与第四交换机之间建立 VxLAN 隧道，并通过第一数据中心的路由器与第四交换机之间隧道，直接将改变封装后的报文发送给第四交换机，而不用通过第二数据中心的路由器。

改变 VxLAN 封装具体为：将步骤 8）的 VxLAN 封装的源 IP 改为 10.10.10.5，目的 IP 改为第四交换机的 IP 地址 11.11.11.4，VNI 改为 33333；然后将步骤 5）的 IP 报文的目的 MAC 改为 MAC-VM5，源 MAC 改为 MAC2。经过对 VxLAN 封装的改变，第四交换机收到封装后的 IP 报文，直接进行处理，不会再触发广播 ARP 请求的流程。

10）第四交换机进行 VxLAN 解封装后，将 IP 报文发送给 VM5。

VM5 向 VM4 发送 IP 报文的过程，与此类似，不再赘述。

经过上述流程，实现了跨数据中心跨子网的通信，而且 VM 发送的广播

ARP 请求，可由 SDN 控制器代答，减少了广播报文对传输带宽的占用。并且 VM 通信时，只经过了第一数据中心的路由器，而不用同时经过第一数据中心的路由器和第二数据中心的路由器，减少了报文在数据中心间的迂回。

如图 6 所示的实现虚拟机通信的方法实施例，用于实现位于数据中心 DC1 的虚拟机 VM3 与位于数据中心 DC2 的虚拟机 VM7 通信，VM3 和 VM7 均位于第二子网。VM3 通过第二交换机与第二子网外的 VM 通信，VM7 通过第三交换机与第二子网外的 VM 通信。该方法包括以下步骤:

1）VM3 需要与 VM7 通信。VM3 查询路由表，发现 VM3 与 VM7 位于同一子网。VM3 向第二交换机发送广播地址解析协议(ARP, Address Resolution Protocol)请求，以广播获取目的虚拟机的 MAC 地址。该 ARP 请求携带 VM7 的 IP 地址。

第二交换机接收所述 ARP 请求。经过图 4a 和图 4b 的配置过程，第二交换机可能缓存了 VM7 的 MAC 地址。如果第二交换机发现 VM7 的 MAC 地址存在，则直接执行步骤 4）。如果第二交换机发现 VM7 的 MAC 地址不存在，则执行步骤 2）和步骤 3）。

2）第二交换机将所述 ARP 请求转发给第一数据中心的 SDN 控制器。

3）第一数据中心的 SDN 控制器接收所述 ARP 请求，根据所述 ARP 请求携带的目的虚拟机的 IP 地址，查询本地保存的流表，获取目的虚拟机的 MAC 地址为 MAC-VM7。SDN 控制器向所述第二交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。并且，此 ARP 响应相当于对广播 ARP 请求进行了代答。

4）第二交换机接收所述 ARP 响应，并将所述 ARP 响应转发给 VM3。

5）VM3 向第二交换机发送 IP 报文，该 IP 报文的目的 MAC(DestMAC)为 MAC-VM7，目的 IP（DestIP）为 VM7 的 IP 地址：192.168.2.7，源 MAC 为 VM3 的 MAC（以 MAC-VM3 表示），源 IP 为 VM3 的 IP 地址：192.168.2.3。

经过图 4a 和图 4b 的配置过程，第二交换机可能缓存了 MAC-VM7 对应

的 VTEP 信息。如果第二交换机发现 MAC-VM7 对应的 VTEP 信息存在，则直接执行步骤 8）。如果第二交换机发现 MAC-VM7 对应的 VTEP 信息不存在，则执行步骤 6）和步骤 7）。

6）第二交换机向本数据中心的 SDN 控制器发送请求，以请求获取目的虚拟机的 MAC 地址对应的 VTEP 信息。

7）SDN 控制器查询本地保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第二交换机发送所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

在本实施例中，根据 MAC-VM7，获取的 VTEP 为 VTEP3，VTEP 信息为 11.11.11.3。

8）第二交换机获取目的虚拟机的 MAC 地址对应的 VTEP 信息，然后将从 VM3 接收的 IP 报文 VxLAN 封装后发送给该 VTEP 信息对应的交换机。其中 VxLAN 封装的 IP 报文的源 IP 为第二交换机的 IP 地址 10.10.10.3，目的 IP 为获取的 VTEP 信息（11.11.11.3）。

如果第二交换机查询本地的流表，发现 MAC-VM7 对应的 VTEP 信息不存在，则通过上述的步骤 6）和 7），获取目的虚拟机的 MAC 地址对应的 VTEP 信息；或者，

第二交换机查询本地的流表，获取目的虚拟机的 MAC 地址对应的 VTEP 信息。

经过图 2、图 3、图 4a 和图 4b 的配置方法，所述第一数据中心的 SDN 控制器从所述第二数据中心的 SDN 控制器获取了 VM7 的 IP 地址、MAC 地址和第三交换机的 VTEP 信息，并将 VM7 的 IP 地址、MAC 地址和第三交换机的 VTEP 信息下发给了第二交换机。第二交换机从而能够通过第二交换机与第三交换机的隧道将 IP 报文发送给 VM7，即不用经过数据中心 1 的路由器，也不用经过数据中心 2 的路由器。

9）第三交换机进行 VxLAN 解封装后，将 IP 报文发送给 VM7。

VM7 向 VM3 发送 IP 报文的过程，与此类似，不再赘述。

经过上述流程，实现了跨数据中心同一子网的通信，而且 VM 发送的广播 ARP 请求，可由 SDN 控制器代答，减少了广播报文对传输带宽的占用。并且 VM 通信时，即不用经过第一数据中心的路由器，也不用经过第二数据中心的路由器，减少了报文在两个数据中心间的迂回。

如图 7 所示的实现虚拟机通信的方法实施例，用于实现位于数据中心 DC1 的虚拟机 VM1 与位于同一数据中心 DC1 的另一虚拟机 VM2 通信，VM1 位于第一子网，VM2 位于第二子网，第一子网与第二子网不同。VM1 通过第一交换机与第一子网外的 VM 通信，VM2 通过第二交换机与第二子网外的 VM 通信。该方法包括以下步骤:

1）VM1 需要与 VM2 通信。VM1 查询路由表，发现 VM2 与 VM1 位于不同子网，则 VM1 使用第一子网的默认网关 192.168.1.1 作为目的网关。VM1 向第一交换机发送广播地址解析协议(ARP, Address Resolution Protocol)请求，以广播获取目的网关的 MAC 地址，该 ARP 请求携带所述目的网关的 IP 地址。

2）第一交换机接收所述 ARP 请求，并将所述 ARP 请求转发给第一数据中心的 SDN 控制器。

3）第一数据中心的 SDN 控制器接收所述 ARP 请求，根据所述 ARP 请求携带的目的网关的 IP 地址，查询本地保存的流表，获取目的网关的 MAC 地址为 MAC1。SDN 控制器向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址。并且，此 ARP 响应相当于对广播 ARP 请求进行了代答。

可选地，该 ARP 请求还可以携带接收 ARP 响应的第一交换机的端口信息， SDN 控制器根据所述端口信息，查询流表，获得所述端口信息对应的 VNI。获取所述目的网关的 MAC 地址具体包括: 所述 SDN 控制器根据所述目的网关的 IP 地址和端口信息对应的 VNI，获取所述目的网关的 MAC 地址。

可选地，该 ARP 请求还可以携带接收 ARP 响应的第一交换机的端口的 VNI。获取所述目的网关的 MAC 地址具体包括：所述 SDN 控制器根据所述目的网关的 IP 地址和第一交换机的端口的 VNI，获取所述目的网关的 MAC 地址。

4）第一交换机接收所述 ARP 响应，并将所述 ARP 响应转发给 VM1。

5）VM1 向第一交换机发送 IP 报文，该 IP 报文的目的 MAC(DestMAC)为 MAC1，目的 IP(DestIP)为 VM2 的 IP 地址：192.168.2.2，源 MAC 为 VM1 的 MAC（以 MAC-VM1 表示），源 IP 为 VM1 的 IP 地址：192.168.2.1。

经过图 3、图 4a 和图 4b 的配置过程，第一交换机可能缓存了 MAC1 对应的 VTEP 信息。如果第一交换机发现 MAC1 对应的 VTEP 信息存在，则直接执行步骤 8）。如果第一交换机发现 MAC1 对应的 VTEP 信息不存在，则执行步骤 6）和步骤 7）。

6）第一交换机向本数据中心的 SDN 控制器发送请求，以请求获取目的网关的 MAC 地址对应的 VTEP 信息。

7）SDN 控制器查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机发送所述目的网关的 MAC 地址对应的 VTEP 信息。

当 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上的 VTEP 信息，则以与所述第一交换机位于同一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

在本实施例中，根据 MAC1，如果获取了 VTEP5（IP 为 10.10.10.5）和 VTEP6（IP 为 11.11.11.6）两条 VTEP 信息，则选择与第一交换机位于同一数据中心的 VTEP5 作为目的网关的 MAC 地址对应的 VTEP。

当第一数据中心内，与目的网关的 MAC 地址对应的 VTEP 信息有两条或以上时，如图 1b 所示的多路由器的方案，SDN 控制器根据以下至少之一：路由器的负载和调度策略，为 VM4 选择合适的 VTEP，把选择的合适的 VTEP

的 IP 地址发送给所述第二交换机。路由器的负载包括但不限于路由器的流量、和路由器的 CPU 负荷。调度策略包括但不限于 RondRobin、最小负载优先、和最大负载优先。

8) 第一交换机获取目的网关的 MAC 地址对应的 VTEP 信息，然后将从 VM1 接收的 IP 报文 VxLAN 封装后发送给该 VTEP 信息对应的路由器。其中 VxLAN 封装的 IP 报文的源 IP 为第一交换机的 IP 地址 10.10.10.2，目的 IP 为获取的 VTEP 信息（10.10.10.5），VNI 为 11111。

如果第一交换机查询本地的流表，发现 MAC1 对应的 VTEP 信息不存在，则通过上述的步骤 6）和 7），获取目的网关的 MAC 地址对应的 VTEP 信息；或者

第一交换机查询本地的流表，获取目的网关的 MAC 地址对应的 VTEP 信息。具体的，获取了两条或以上的 VTEP 信息时，则以与所述第一交换机位于同一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。同一数据中心有多个路由器的时候，采取步骤 7）类似的路由器的负载和调度策略进行选择，不再详细描述。

9）所述第一数据中心的路由器，对接收的报文改变 VxLAN 封装，通过第一数据中心的路由器与第二交换机之间的隧道，发送给第二交换机。

改变 VxLAN 封装具体为：将步骤 8）的 VxLAN 封装的源 IP 改为 10.10.10.5，目的 IP 改为第二交换机的 IP 地址 10.10.10.3，VNI 改为 22222；然后将步骤 5）的 IP 报文的目的 MAC 改为 MAC-VM2，源 MAC 改为 MAC1。经过对 VxLAN 封装的改变，第二交换机收到封装后的 IP 报文，直接进行处理，不会再触发广播 ARP 请求的流程。

10）第二交换机进行 VxLAN 解封装后，将 IP 报文发送给 VM2。

经过上述流程，实现了同一数据中心跨子网的通信，而且 VM 发送的广播 ARP 请求，可由 SDN 控制器代答，减少了广播报文对传输带宽的占用。并且 VM 通信时，只经过了数据中心 1 的路由器，而不像现有技术可能同时

经过数据中心 1 的路由器和数据中心 2 的路由器，减少了报文在数据中心间的迂回。对多个路由器的方案，还能够实现路由器间的弹性伸缩。

同一数据中心同一子网的通信，不需要经过路由器。以 VM2 与 VM3 通信为例，VM2 发送报文到第二交换机，然后第二交换机将报文发送给 VM3。

上述流程描述了本发明实施例在各种场景下的应用，当然并不限于这些场景。

本发明实施例的装置

如图 8 所示，本发明实施例提供一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与所述第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信。所述第一交换机包括:

接收模块 801，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

发送模块 803，用于向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求;

所述接收模块 801 还用于从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的网关的媒体接入控制 MAC 地址;

获取模块 802，用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心;

所述发送模块 803，还用于根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

获取模块 802 用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为:

获取模块 802 用于向所述 SDN 控制器发送请求，以请求获取所述目的

网关的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的网关的 MAC 地址对应的 VTEP 信息。或者

获取模块 802 用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为：

获取模块 802 用于根据所述目的网关的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息。当获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

如图 9 所示，本发明实施例提供一种实现虚拟机通信的软件定义网络 SDN 控制器，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于所述第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，所述 SDN 控制器包括：

接收模块 901，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址；

处理模块 902，用于根据所述目的网关的 IP 地址，获取所述目的网关的媒体接入控制 MAC 地址；

发送模块 903，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，以使所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心，使所述 VTEP 信息对应的路由器将 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

接收模块 901 还用于接收所述第一交换机发送的请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息；

处理模块 902 还用于根据所述目的网关的 MAC 地址，查询所述 SDN 控

制器保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的网关的 MAC 地址对应的 VTEP 信息。当查询所述 SDN 控制器保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。并且当处理模块 902 根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，处理模块 902 根据以下至少之一：路由器的负载和调度策略，选择所述目的网关的 MAC 地址对应的 VTEP 信息。

若所述第二虚拟机位于第二数据中心，则处理模块 902 还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、第二子网的 VNI，和所述第二交换机的 VTEP 信息。

本发明实施例图8提供的交换机和图9提供的SDN控制器，可以用来执行图2、图3、图4a和图4b的配置过程，以及用来执行图5和图7所述的方法，具体的过程，参见系统实施例和方法实施例的描述，不再赘述。

本发明实施例图 8 提供的交换机和图 9 提供的 SDN 控制器，可以实现跨子网的通信，而且 VM 发送的广播 ARP 请求，可由 SDN 控制器代答，减少了广播报文对传输带宽的占用。并且 VM 通信时，只经过了数据中心 1 的路由器，而不用同时经过数据中心 1 的路由器和数据中心 2 的路由器，减少了报文在数据中心间的迂回。

如图 10 所示，本发明实施例提供一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与所述第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，所述交换机包括：

接收模块 1001，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

处理模块 1002，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址；

发送模块 1003，用于通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

处理模块 1002 用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址具体为：向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求，所述 ARP 请求携带所述目的虚拟机的 IP 地址，从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。

处理模块 1002 用于根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体为：处理模块 1002 用于向所述 SDN 控制器发送请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的虚拟机的 MAC 地址对应的 VTEP 信息；或者，

处理模块 1002 用于根据所述目的虚拟机的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

处理模块 1002 还用于从所述第一数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、和所述第二交换机的 VTEP 信息。

如图 11 所示，本发明实施例提供一种实现虚拟机通信的软件定义网络 SDN 控制器，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于所述第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，所述 SDN 控制器包括：

接收模块 1101，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

获取模块 1102，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址；

发送模块 1103，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址，以使所述第一交换机根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址，使所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

接收模块 1101 还用于所述第一交换机发送的请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息；获取模块 1102 还用于根据所述目的虚拟机的 MAC 地址，查询所述 SDN 控制器保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

获取模块 1102 还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息；发送模块 1103 还用于将所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息下发给所述第一交换机。

本发明实施例图10提供的交换机和图11提供的SDN控制器，可以用来执行图2、图3、图4a和图4b的配置过程，以及用来执行图6所述的方法，具体的过程，参见系统实施例和方法实施例的描述，不再赘述。

本发明实施例图 10 提供的交换机和图 11 提供的 SDN 控制器，可以实现跨数据中心同一子网的通信，而且 VM 发送的广播 ARP 请求，可由 SDN 控制器代答，减少了广播报文对传输带宽的占用。并且 VM 通信时，即不用经过数据中心 1 的路由器，也不用经过数据中心 2 的路由器，减少了报文在

两个数据中心间的迁回。

为本发明实施例提供的另一实现虚拟机通信的装置1200，包括:

处理器 101、存储器 102、系统总线（简称总线）105。处理器 101 和存储器 102 之间通过系统总线 105 连接并完成相互间的通信。

处理器 101 可能为单核或多核中央处理单元，或者为特定集成电路，或者为被配置成实施本发明实施例的一个或多个集成电路。

存储器 102 可以为高速 RAM 存储器，也可以为非易失性存储器（non-volatile memory），例如至少一个硬盘存储器。

存储器 102 用于存储计算机执行指令 1021。具体的，计算机执行指令 1021 中可以包括程序代码。

当计算机运行时，处理器 101 运行计算机执行指令 1021，可以执行图 2-图 7 任意之一的流程。

本发明实施例还提供一种虚拟机通信的计算机程序产品，包括存储了程序代码的计算机可读存储介质，所述程序代码包括的指令用于执行图 2-图 7 任意之一的流程。

本领域普通技术人员将会理解，本发明的各个方面、或各个方面的可能实现方式可以被具体实施为系统、方法或者计算机程序产品。因此，本发明的各方面、或各个方面的可能实现方式可以采用完全硬件实施例、完全软件实施例（包括固件、驻留软件等等），或者组合软件和硬件方面的实施例的形式，在这里都统称为“电路”、“模块”或者“系统”。此外，本发明的各方面、或各个方面的可能实现方式可以采用计算机程序产品的形式，计算机程序产品是指存储在计算机可读介质中的计算机可读程序代码。

计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质包含但不限于电子、磁性、光学、电磁、红外或半导体系统、设备或者装置，或者前述的任意适当组合，如随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器(EPROM 或者快闪

存储器)、光纤、便携式只读存储器(CD-ROM)。

计算机中的处理器读取存储在计算机可读介质中的计算机可读程序代码，使得处理器能够执行在流程图中每个步骤、或各步骤的组合中规定的功能动作；生成实施在框图的每一块、或各块的组合中规定的功能动作的装置。

计算机可读程序代码可以完全在用户的计算机上执行、部分在用户的计算机上执行、作为单独的软件包、部分在用户的计算机上并且部分在远程计算机上，或者完全在远程计算机或者服务器上执行。也应该注意，在某些替代实施方案中，在流程图中各步骤、或框图中各块所注明的功能可能不按图中注明的顺序发生。例如，依赖于所涉及的功能，接连示出的两个步骤、或两个块实际上可能被大致同时执行，或者这些块有时候可能被以相反顺序执行。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

权 利 要 求

1、一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，其特征在于:

所述第一交换机从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

所述第一交换机向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求;

所述第一交换机从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的网关的媒体接入控制 MAC 地址;

所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心;

所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

2、如权利要求 1 所述的方法，其特征在于，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机向所述 SDN 控制器发送请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息;

接收所述 SDN 控制器发送的所述目的网关的 MAC 地址对应的 VTEP 信息。

3、如权利要求 1 所述的方法，其特征在于，所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息。

4、如权利要求 3 所述的方法，其特征在于，所述第一交换机查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

5、如权利要求 4 所述的方法，其特征在于，当所述第一交换机根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述第一交换机根据以下至少之一: 路由器的负载和调度策略，选择所述目的网关的 MAC 地址对应的 VTEP 信息。

6、如权利要求 1 至 5 任一所述的方法，其特征在于，所述第一交换机根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机具体包括:

所述第一交换机对从所述第一虚拟机接收的 IP 报文进行虚拟可扩展局域网 VxLAN 封装，其中所述 VxLAN 封装的目的 IP 地址为获取的所述 VTEP 信息，虚拟网络标识 VNI 为所述第一子网的标识，所述 IP 报文的目的 MAC 地址为所述目的网关的 MAC 地址;

所述 VTEP 信息对应的路由器，将目的 IP 地址改为所述第二交换机的 IP 地址，VNI 改为所述第二子网的标识，目的 MAC 地址改为所述第二虚拟机的 MAC 地址，然后将改变 VxLAN 封装的 IP 报文，发送给所述第二交换机;

所述第二交换机对接收的改变 VxLAN 封装的 IP 报文，进行 VxLAN 解封装后，发送给所述第二虚拟机。

7、一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚

拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，其特征在于:

所述第一数据中心的软件定义网络 SDN 控制器从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

所述 SDN 控制器根据所述目的网关的 IP 地址，获取所述目的网关的媒体接入控制 MAC 地址，并向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，以使所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心，使所述 VTEP 信息对应的路由器将 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

8、如权利要求 7 所述的方法，其特征在于，还包括:

所述 SDN 控制器接收所述第一交换机发送的请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息;

所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的网关的 MAC 地址对应的 VTEP 信息。

9、如权利要求 8 所述的方法，其特征在于，所述 SDN 控制器查询本地保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体包括:

所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

10、如权利要求 9 所述的方法，其特征在于，当所述 SDN 控制器根据所述目的网关的 MAC 地址，查询本地保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述 SDN 控制器根据以下至少

之一:路由器的负载和调度策略,选择所述目的网关的 MAC 地址对应的 VTEP 信息。

11、如权利要求 7 至 10 任一所述的方法，其特征在于，若所述第二虚拟机位于第二数据中心，则还包括:

所述第一数据中心的 SDN 控制器从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、所述第二子网的虚拟网络标识 VNI，和所述第二交换机的 VTEP 信息，以在所述第一数据中心的 SDN 控制器的控制下，使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

12、一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，其特征在于:

所述第一交换机从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址;

所述第一交换机根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址;

根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址;

所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

13、如权利要求 12 所述的方法，其特征在于，所述第一交换机根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址具体包括:

所述第一交换机向所述第一数据中心的软件定义网络 SDN 控制器转发

所述 ARP 请求，所述 ARP 请求携带所述目的虚拟机的 IP 地址;

所述第一交换机从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。

14、如权利要求 12 或 13 所述的方法，其特征在于，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机向所述 SDN 控制器发送请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息;

接收所述 SDN 控制器发送的所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

15、如权利要求 12 或 13 所述的方法，其特征在于，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体包括:

所述第一交换机根据所述目的虚拟机的 MAC 地址，查询本地保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

16、如权利要求 15 所述的方法，其特征在于，还包括:

所述第一交换机从所述 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、和所述第二交换机的 VTEP 信息，以通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

17、一种实现虚拟机通信的方法，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，其特征在于:

所述第一数据中心的软件定义网络 SDN 控制器从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址;

所述 SDN 控制器根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机

的媒体接入控制 MAC 地址，并向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址，以使所述第一交换机根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址，使所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

18、如权利要求 17 所述的方法，其特征在于，还包括:

所述 SDN 控制器接收所述第一交换机发送的请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息;

所述 SDN 控制器根据所述目的虚拟机的 MAC 地址，查询本地保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

19、如权利要求 17 或 18 所述的方法，其特征在于，还包括:

所述第一数据中心的 SDN 控制器从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息，并将所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息下发给所述第一交换机。

20、一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心的第一子网并且与所述第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，其特征在于，所述第一交换机包括:

接收模块，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

发送模块，用于向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求;

所述接收模块还用于从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应

携带所述目的网关的媒体接入控制 MAC 地址;

获取模块，用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心;

所述发送模块，还用于根据所述 VTEP 信息，发送 IP 报文给所述 VTEP 信息对应的路由器，以使所述 VTEP 信息对应的路由器将所述 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

21、如权利要求 20 所述的交换机，其特征在于，所述获取模块用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为:

所述获取模块用于向所述 SDN 控制器发送请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的网关的 MAC 地址对应的 VTEP 信息。

22、如权利要求 20 所述的交换机，其特征在于，所述获取模块用于根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为:

所述获取模块用于根据所述目的网关的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息。

23、如权利要求 22 所述的交换机，其特征在于，所述获取模块用于查询所述第一交换机保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为:

所述获取模块用于根据所述目的网关的 MAC 地址，查询所述第一交换机保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

24、一种实现虚拟机通信的软件定义网络 SDN 控制器，其特征在于，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第

二虚拟机的系统，所述第一虚拟机位于所述第一数据中心的第一子网并且与第一交换机通信，所述第二虚拟机位于第二子网并且与第二交换机通信，所述 SDN 控制器包括:

接收模块，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的网关的 IP 地址;

处理模块，用于根据所述目的网关的 IP 地址，获取所述目的网关的媒体接入控制 MAC 地址;

发送模块，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的网关的 MAC 地址，以使所述第一交换机根据所述目的网关的 MAC 地址，获取所述目的网关的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息对应的路由器位于所述第一数据中心，使所述 VTEP 信息对应的路由器将 IP 报文通过所述路由器与所述第二交换机的隧道发送给所述第二虚拟机。

25、如权利要求 24 所述的控制器，其特征在于，

所述接收模块还用于接收所述第一交换机发送的请求，以请求获取所述目的网关的 MAC 地址对应的 VTEP 信息;

所述处理模块还用于根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的网关的 MAC 地址对应的 VTEP 信息。

26、如权利要求 25 所述的控制器，其特征在于，所述处理模块用于查询所述 SDN 控制器保存的流表，获取所述目的网关的 MAC 地址对应的 VTEP 信息具体为:

所述处理模块用于根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取两条或以上 VTEP 信息时，以位于所述第一数据中心的路由器的 VTEP 信息作为所述目的网关的 MAC 地址对应的 VTEP 信息。

27、如权利要求 26 所述的控制器，其特征在于，当所述处理模块用于

根据所述目的网关的 MAC 地址，查询所述 SDN 控制器保存的流表，获取了两条或以上位于所述第一数据中心的路由器的 VTEP 信息时，所述处理模块用于根据以下至少之一：路由器的负载和调度策略，选择所述目的网关的 MAC 地址对应的 VTEP 信息。

28、如权利要求 24 至 27 任一所述的控制器，若所述第二虚拟机位于第二数据中心，则所述处理模块还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、第二子网的虚拟网络标识 VNI，和所述第二交换机的 VTEP 信息。

29、一种实现虚拟机通信的第一交换机，应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于第一数据中心并且与所述第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，其特征在于，所述交换机包括：

接收模块，用于从所述第一虚拟机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

处理模块，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址，根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址；

发送模块，用于通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

30、如权利要求 29 所述的交换机，其特征在于，所述处理模块用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的 MAC 地址具体为：

所述处理模块用于向所述第一数据中心的软件定义网络 SDN 控制器转发所述 ARP 请求，所述 ARP 请求携带所述目的虚拟机的 IP 地址，从所述 SDN 控制器接收 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址。

31、如权利要求 29 或 30 所述的交换机，其特征在于，所述处理模块用于根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体为：

所述处理模块用于向所述 SDN 控制器发送请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，接收所述 SDN 控制器发送的所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

32、如权利要求 29 或 30 所述的交换机，其特征在于，所述处理模块用于根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息具体为：

所述处理模块用于根据所述目的虚拟机的 MAC 地址，查询所述第一交换机保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

33、如权利要求 32 所述的交换机，其特征在于，所述处理模块还用于从所述第一数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址、和所述第二交换机的 VTEP 信息。

34、一种实现虚拟机通信的软件定义网络 SDN 控制器，其特征在于，所述 SDN 控制器位于第一数据中心，并且应用于至少包括第一虚拟机与第二虚拟机的系统，所述第一虚拟机位于所述第一数据中心并且与第一交换机通信，所述第二虚拟机位于第二数据中心并且与第二交换机通信，所述第一虚拟机和所述第二虚拟机位于同一子网，所述 SDN 控制器包括：

接收模块，用于从所述第一交换机接收地址解析协议 ARP 请求，所述 ARP 请求携带目的虚拟机的 IP 地址；

获取模块，用于根据所述目的虚拟机的 IP 地址，获取所述目的虚拟机的媒体接入控制 MAC 地址；

发送模块，用于向所述第一交换机发送 ARP 响应，所述 ARP 响应携带所述目的虚拟机的 MAC 地址，以使所述第一交换机根据所述目的虚拟机的 MAC 地址，获取所述目的虚拟机的 MAC 地址对应的虚拟可扩展局域网隧道端

点 VTEP 信息，所述 VTEP 信息为所述第二交换机的 IP 地址，使所述第一交换机通过所述第一交换机与所述第二交换机的隧道将 IP 报文发送给所述第二虚拟机。

35、如权利要求 34 所述的控制器，其特征在于，所述接收模块还用于接收所述第一交换机发送的请求，以请求获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息；

所述获取模块还用于根据所述目的虚拟机的 MAC 地址，查询所述 SDN 控制器保存的流表，获取所述目的虚拟机的 MAC 地址对应的 VTEP 信息，并向所述第一交换机返回所述目的虚拟机的 MAC 地址对应的 VTEP 信息。

36、如权利要求 34 或 35 所述的控制器，其特征在于，所述获取模块还用于从所述第二数据中心的 SDN 控制器获取所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息；

所述发送模块还用于将所述第二虚拟机的 IP 地址、MAC 地址和所述第二交换机的 VTEP 信息下发给所述第一交换机。

37、一种实现虚拟机通信的交换机，其特征在于，包括：

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行如权利要求 1 至 6 任一所述的方法。

38、一种实现虚拟机通信的软件定义网络 SDN 控制器，其特征在于，包括：

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信；

所述存储器，用于存储计算机执行指令；

所述处理器，用于运行所述计算机执行指令，执行如权利要求 7 至 11

任一所述的方法。

39、一种实现虚拟机通信的交换机，其特征在于，包括:

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信;

所述存储器，用于存储计算机执行指令;

所述处理器，用于运行所述计算机执行指令，执行如权利要求 12 至 16 任一所述的方法。

40、一种实现虚拟机通信的软件定义网络 SDN 控制器，其特征在于，包括:

处理器、存储器和系统总线，所述处理器和所述存储器之间通过所述系统总线连接并完成相互间的通信;

所述存储器，用于存储计算机执行指令;

所述处理器，用于运行所述计算机执行指令，执行如权利要求 17 至 19 任一所述的方法。

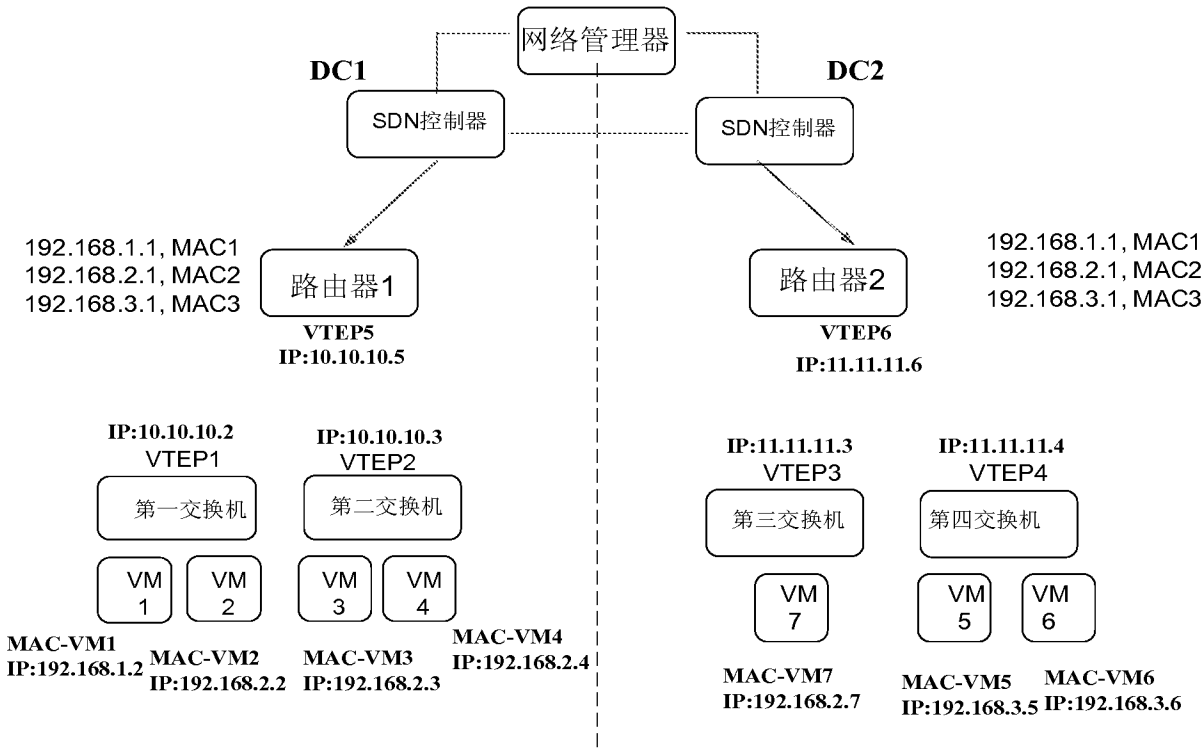

网络管理器
DC1
DC2
SDN控制器
SDN控制器
192.168.1.1, MAC1
192.168.2.1, MAC2
192.168.3.1, MAC3
路由器1
VTEP5
IP:10.10.10.5
路由器2
VTEP6
IP:11.11.11.6
192.168.1.1, MAC1
192.168.2.1, MAC2
192.168.3.1, MAC3
IP:10.10.10.2
VTEP1
IP:10.10.10.3
VTEP2
IP:11.11.11.3
VTEP3
IP:11.11.11.4
VTEP4
第一交换机
第二交换机
第三交换机
第四交换机
VM 1
VM 2
VM 3
VM 4
VM 7
VM 5
VM 6
MAC-VM1
IP:192.168.1.2
MAC-VM2
IP:192.168.2.2
MAC-VM3
IP:192.168.2.3
MAC-VM4
IP:192.168.2.4
MAC-VM7
IP:192.168.2.7
MAC-VM5
IP:192.168.3.5
MAC-VM6
IP:192.168.3.6

图 1a

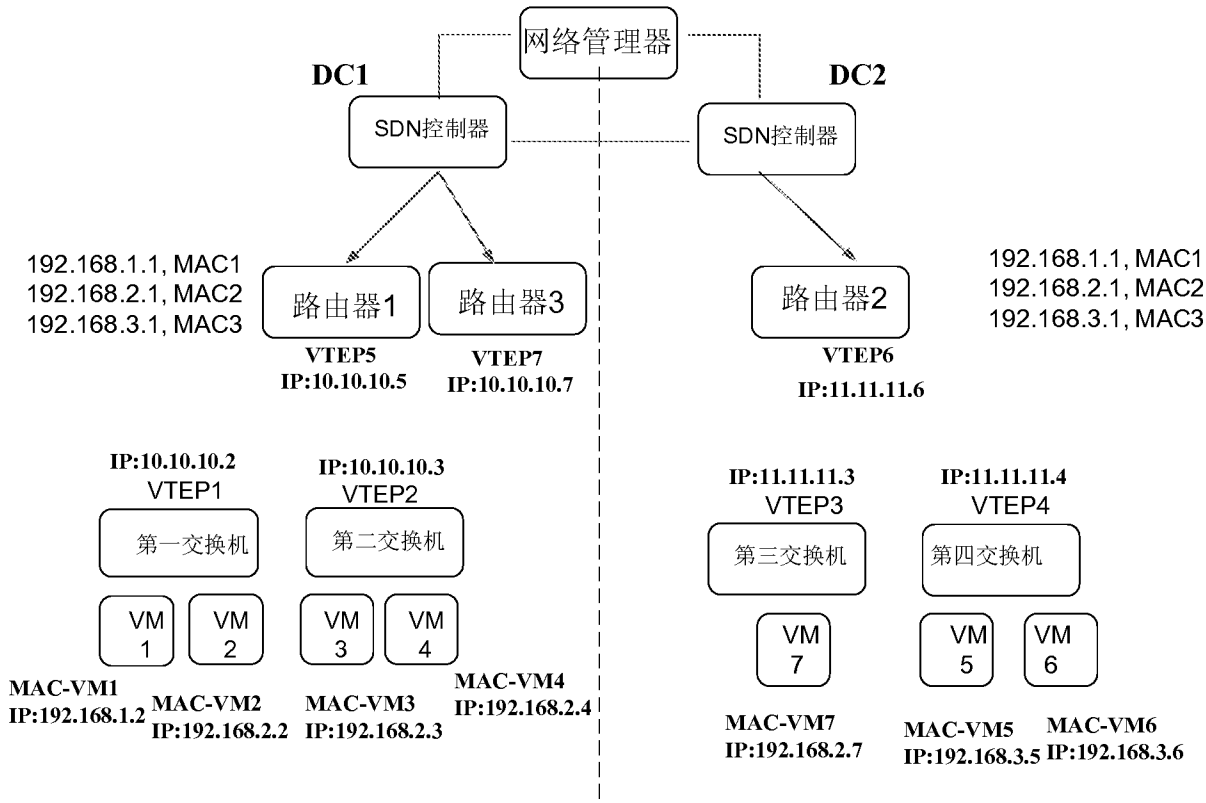

网络管理器
DC1
DC2
SDN控制器
SDN控制器
192.168.1.1, MAC1
192.168.2.1, MAC2
192.168.3.1, MAC3
路由器1
路由器3
路由器2
192.168.1.1, MAC1
192.168.2.1, MAC2
192.168.3.1, MAC3
VTEP5
IP:10.10.10.5
VTEP7
IP:10.10.10.7
VTEP6
IP:11.11.11.6
IP:10.10.10.2
VTEP1
IP:10.10.10.3
VTEP2
IP:11.11.11.3
VTEP3
IP:11.11.11.4
VTEP4
第一交换机
第二交换机
第三交换机
第四交换机
VM 1
VM 2
VM 3
VM 4
VM 7
VM 5
VM 6
MAC-VM1
IP:192.168.1.2
MAC-VM2
IP:192.168.2.2
MAC-VM3
IP:192.168.2.3
MAC-VM4
IP:192.168.2.4
MAC-VM7
IP:192.168.2.7
MAC-VM5
IP:192.168.3.5
MAC-VM6
IP:192.168.3.6

图 1b

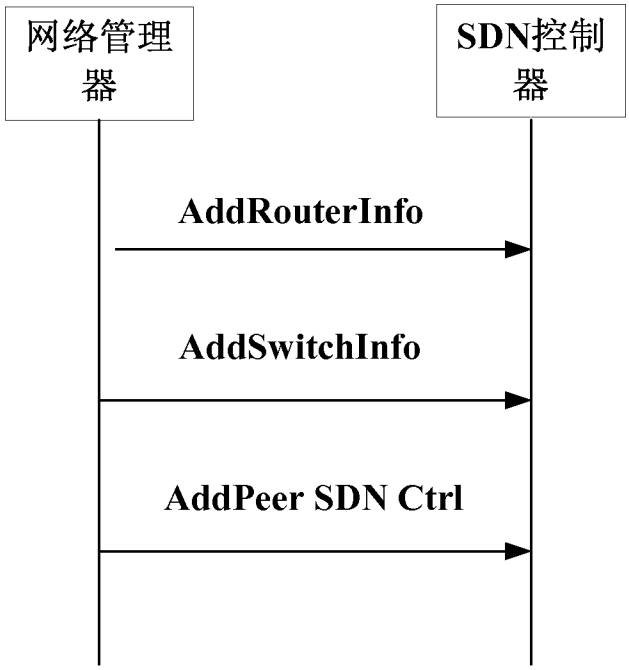

网络管理器
SDN控制器
AddRouterInfo
AddSwitchInfo
AddPeer SDN Ctrl

图 2

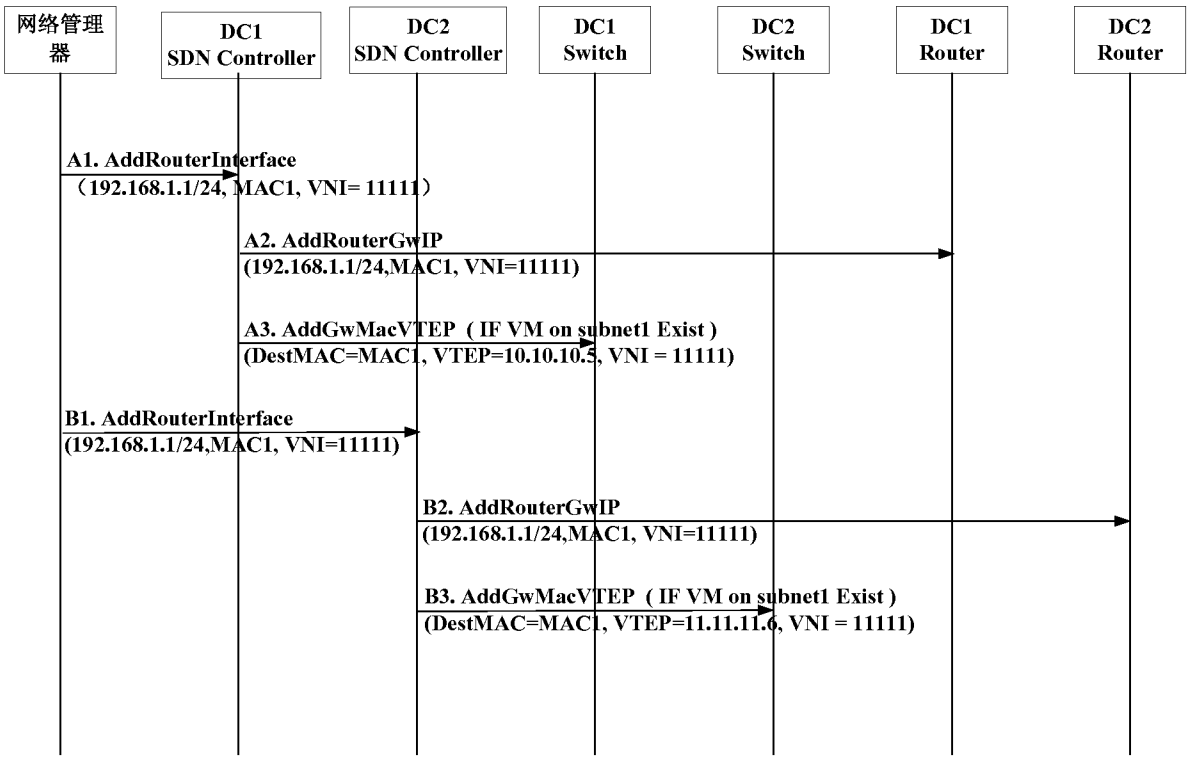

网络管理器
DC1 SDN Controller
DC2 SDN Controller
DC1 Switch
DC2 Switch
DC1 Router
DC2 Router
A1. AddRouterInterface
（192.168.1.1/24, MAC1, VNI= 11111）
A2. AddRouterGwIP
(192.168.1.1/24,MAC1, VNI=11111)
A3. AddGwMacVTEP （IF VM on subnet1 Exist）
(DestMAC=MAC1, VTEP=10.10.10.5, VNI = 11111)
B1. AddRouterInterface
(192.168.1.1/24,MAC1, VNI=11111)
B2. AddRouterGwIP
(192.168.1.1/24,MAC1, VNI=11111)
B3. AddGwMacVTEP （IF VM on subnet1 Exist）
(DestMAC=MAC1, VTEP=11.11.11.6, VNI = 11111)

图 3

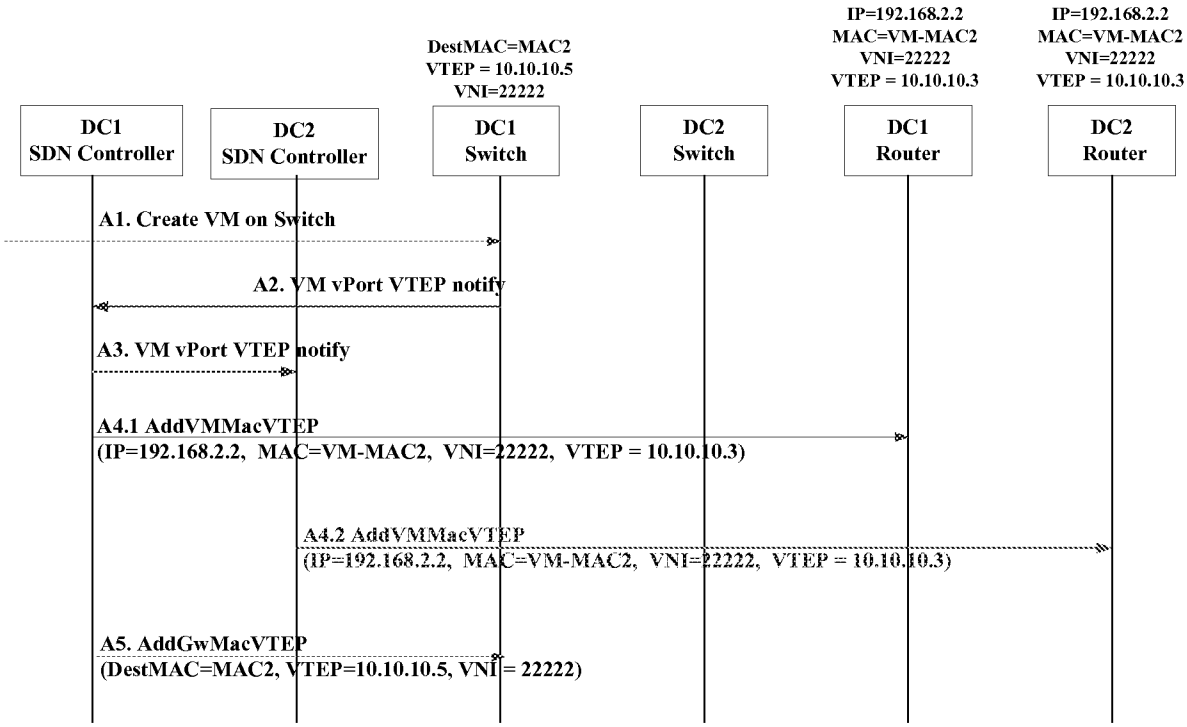

DestMAC=MAC2
VTEP = 10.10.10.5
VNI=22222
IP=192.168.2.2
MAC=VM-MAC2
VNI=22222
VTEP = 10.10.10.3
IP=192.168.2.2
MAC=VM-MAC2
VNI=22222
VTEP = 10.10.10.3
DC1 SDN Controller
DC2 SDN Controller
DC1 Switch
DC2 Switch
DC1 Router
DC2 Router
A1. Create VM on Switch
A2. VM vPort VTEP notify
A3. VM vPort VTEP notify
A4.1 AddVMMacVTEP
(IP=192.168.2.2, MAC=VM-MAC2, VNI=22222, VTEP = 10.10.10.3)
A4.2 AddVMMacVTEP
(IP=192.168.2.2, MAC=VM-MAC2, VNI=22222, VTEP = 10.10.10.3)
A5. AddGwMacVTEP
(DestMAC=MAC2, VTEP=10.10.10.5, VNI = 22222)

图 4a

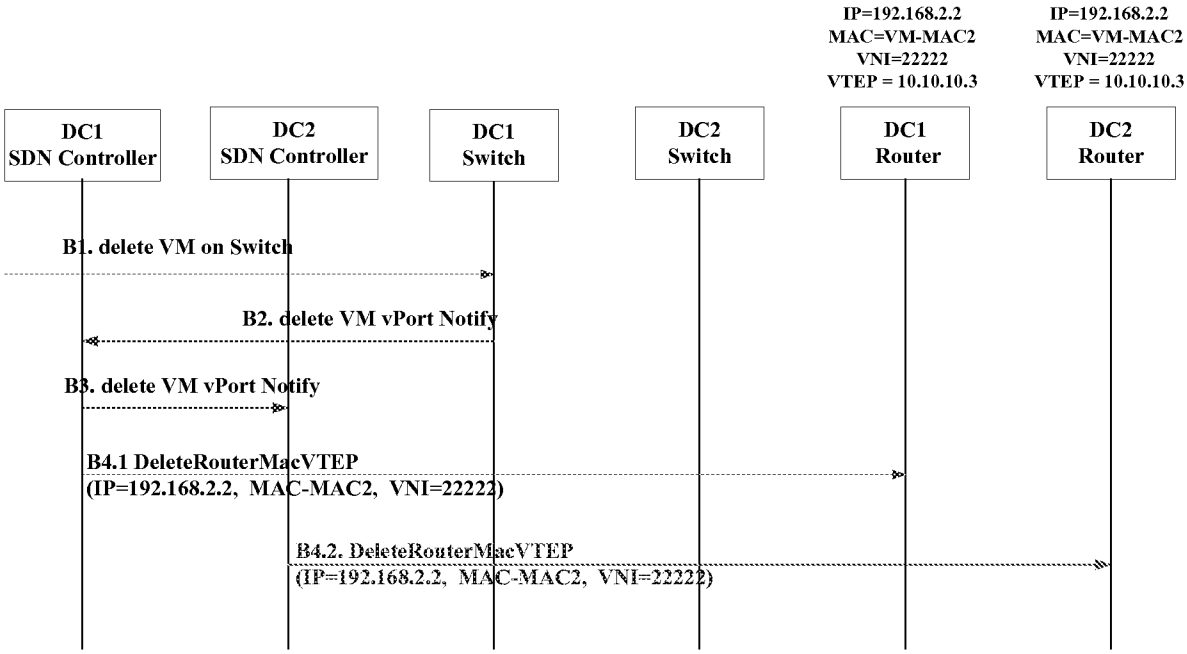

IP=192.168.2.2
MAC=VM-MAC2
VNI=22222
VTEP = 10.10.10.3
IP=192.168.2.2
MAC=VM-MAC2
VNI=22222
VTEP = 10.10.10.3
DC1 SDN Controller
DC2 SDN Controller
DC1 Switch
DC2 Switch
DC1 Router
DC2 Router
B1. delete VM on Switch
B2. delete VM vPort Notify
B3. delete VM vPort Notify
B4.1 DeleteRouterMacVTEP
(IP=192.168.2.2, MAC-MAC2, VNI=22222)
B4.2. DeleteRouterMacVTEP
(IP=192.168.2.2, MAC-MAC2, VNI=22222)

图 4b

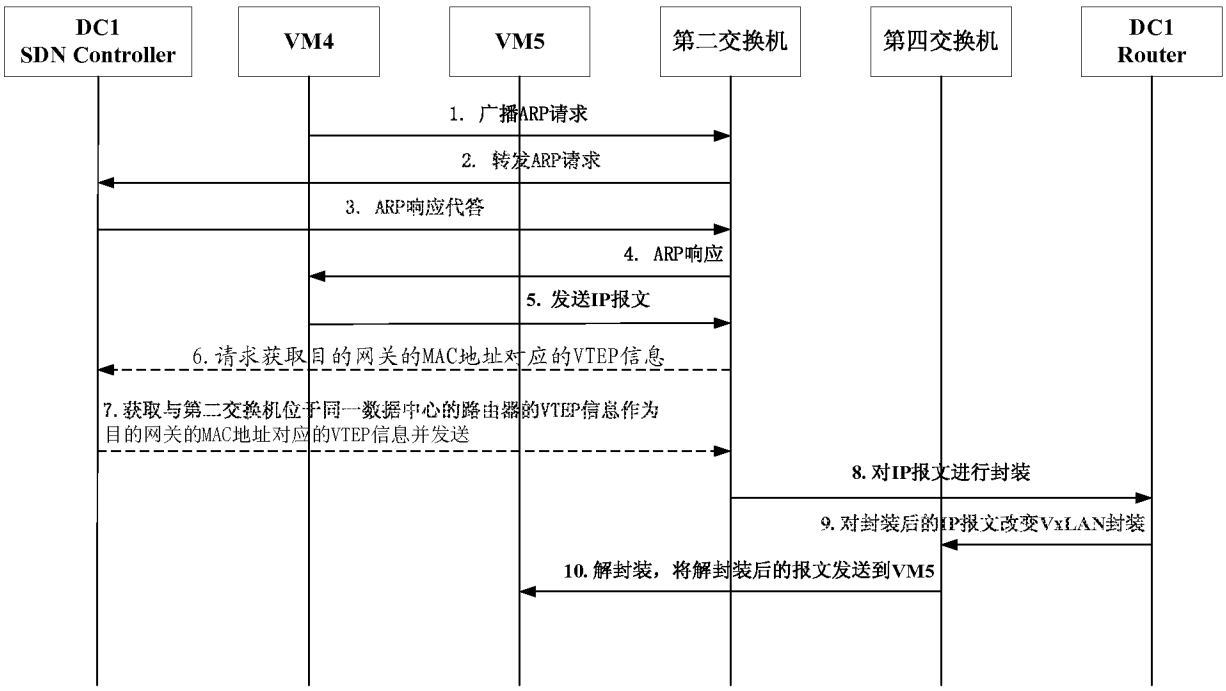

DC1
SDN Controller
VM4
VM5
第二交换机
第四交换机
DC1
Router
1. 广播ARP请求
2. 转发ARP请求
3. ARP响应代答
4. ARP响应
5. 发送IP报文
6.请求获取目的网关的MAC地址对应的VTEP信息
7.获取与第二交换机位于同一数据中心的路由器的VTEP信息作为
目的网关的MAC地址对应的VTEP信息并发送
8.对IP报文进行封装
9.对封装后的IP报文改变VxLAN封装
10.解封装，将解封装后的报文发送到VM5

图 5

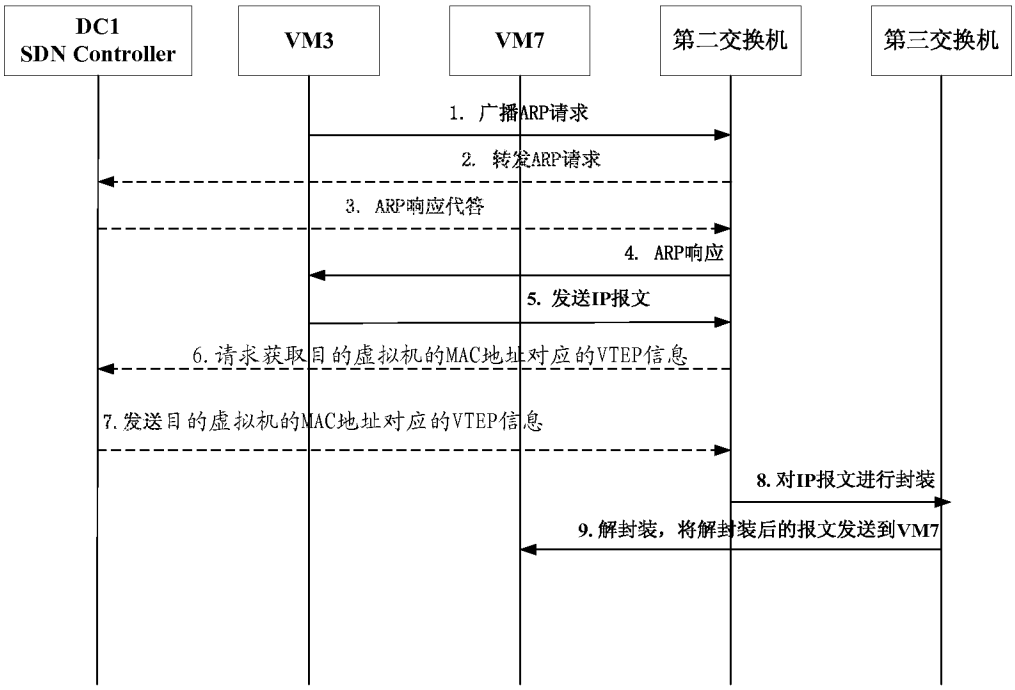

DC1
SDN Controller
VM3
VM7
第二交换机
第三交换机
1. 广播ARP请求
2. 转发ARP请求
3. ARP响应代答
4. ARP响应
5. 发送IP报文
6.请求获取目的虚拟机的MAC地址对应的VTEP信息
7.发送目的虚拟机的MAC地址对应的VTEP信息
8.对IP报文进行封装
9.解封装，将解封装后的报文发送到VM7

图 6

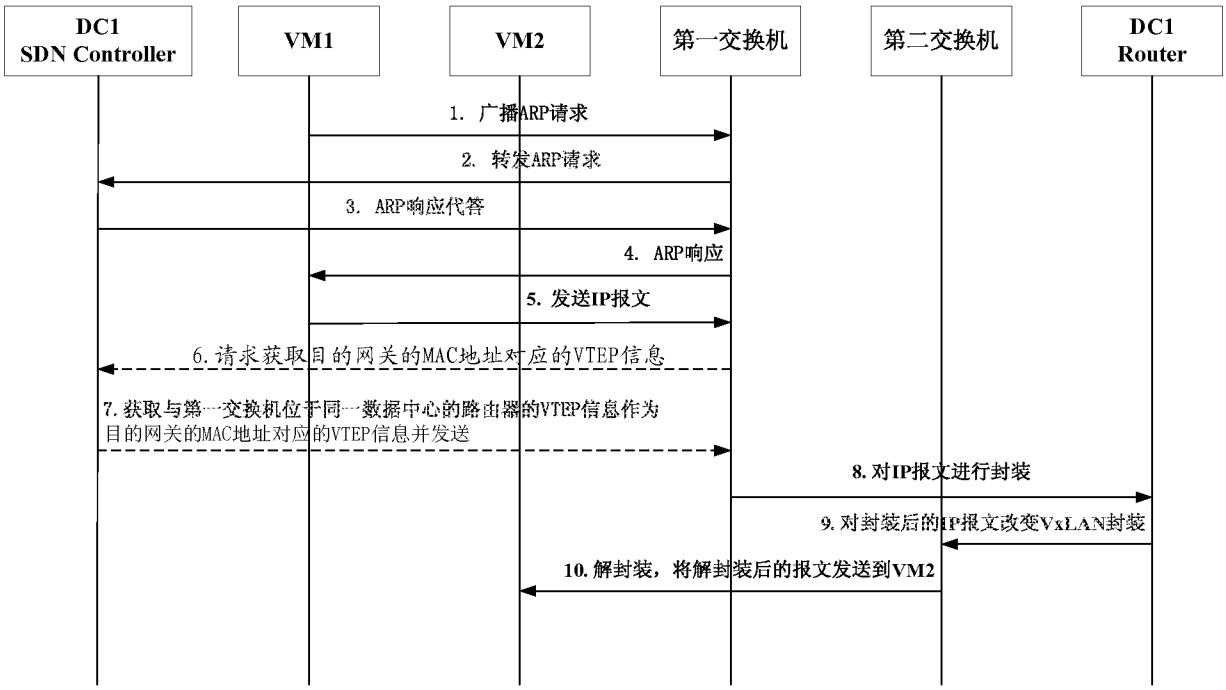

DC1
SDN Controller
VM1
VM2
第一交换机
第二交换机
DC1
Router
1. 广播ARP请求
2. 转发ARP请求
3. ARP响应代答
4. ARP响应
5. 发送IP报文
6.请求获取目的网关的MAC地址对应的VTEP信息
7.获取与第一交换机位于同一数据中心的路由器的VTEP信息作为目的网关的MAC地址对应的VTEP信息并发送
8.对IP报文进行封装
9.对封装后的IP报文改变VxLAN封装
10.解封装，将解封装后的报文发送到VM2

图 7

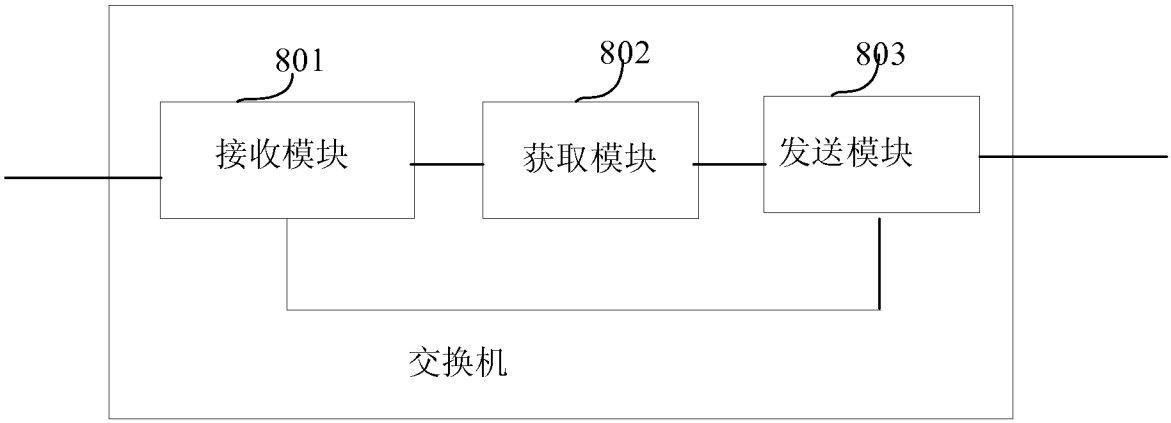

801
802
803
接收模块
获取模块
发送模块
交换机

图 8

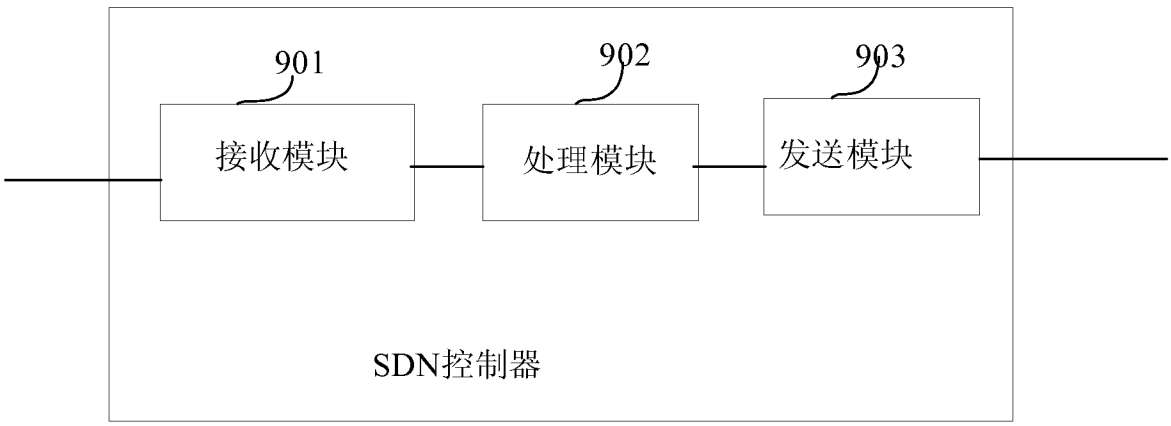

901
902
903
接收模块
处理模块
发送模块
SDN控制器

图 9

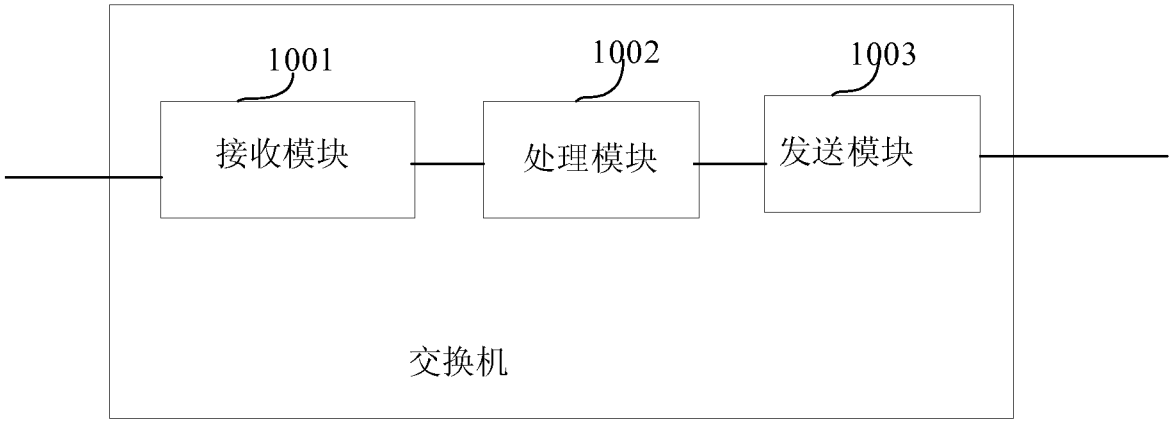

1001
1002
1003
接收模块
处理模块
发送模块
交换机

图 10

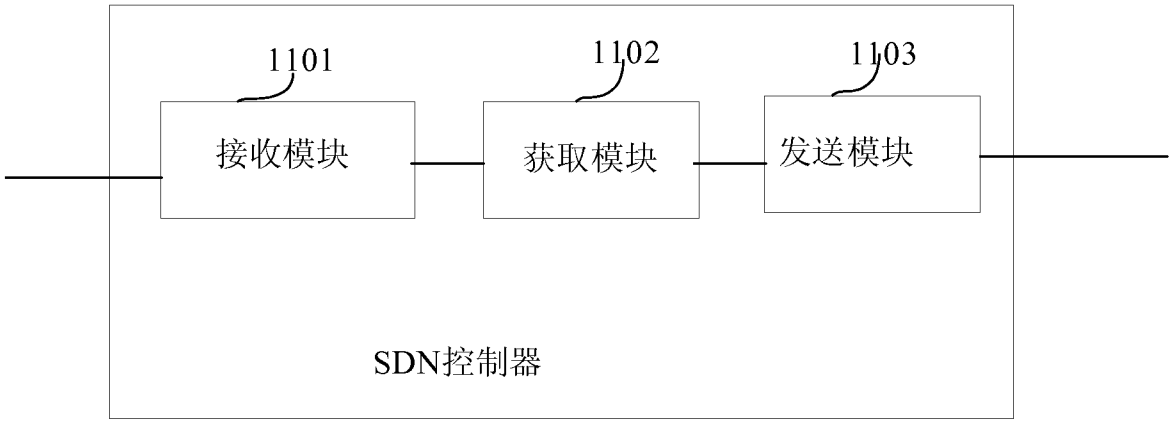

1101
1102
1103
接收模块
获取模块
发送模块
SDN控制器

图 11

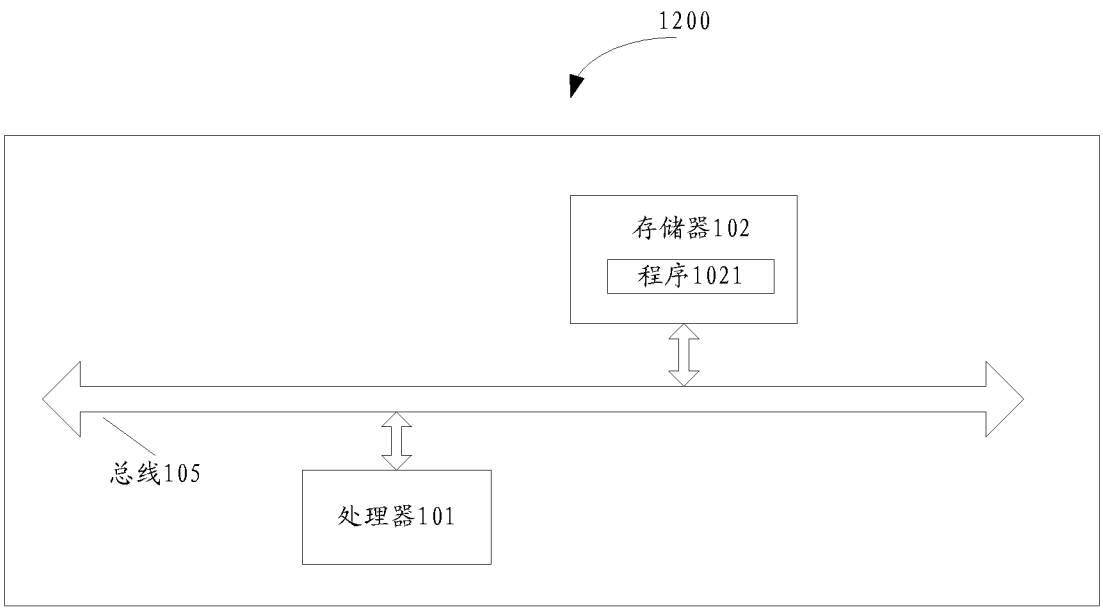

1200
存储器102
程序1021
总线105
处理器101

图 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/091202

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/701 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, VEN, CNKI: vtep, mac, ip, sdn, vm, virtual machine, arp, address resolution protocol, vxlan tunnel end point

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 102857416 A (ZTE CORP.) 02 January 2013 (02.01.2013) the whole document	1-40
A	US 20130266019 A1 (FUTUREWEI TECHNOLOGIES INC.) 10 October 2013 (10.10.2013) the whole document	1-40
A	US 8259726 B2 (SUBRAMANIAN KRISHNAMURTHY) 04 September 2012 (04.09.2012) the whole document	1-40

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 June 2014	26 June 2014
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer YIN,Yue Telephone No. (86-10) 62089467

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/091202

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102857416 A	02 January 2013	WO 2013185715 A1	19 December 2013
US 20130266019 A1	10 October 2013	None	
US 8259726 B2	04 September 2012	US 2011292939 A1	01 December 2011

国际检索报告

国际申请号
PCT/CN2013/091202

A. 主题的分类

H04L 12/701 (2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNABS;CNTXT;VEN;CNKI:vtep，虚拟机，虚机，地址解析协议，mac，ip，sdn，软件定义网络，vm，virtual w machine，arp，address w resolution w protocol，vxlan w tunnel w end w point

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 102857416A (中兴通讯股份有限公司) 2013年 1月 02日 (2013-01-02) 全文	1-40
A	US 20130266019A1 (FUTUREWEI TECHNOLOGIES INC) 2013年 10月 10日 (2013-10-10) 全文	1-40
A	US 8259726B2 (SUBRAMANIAN KRISHNAMURTHY等) 2012年 9月 04日 (2012-09-04) 全文	1-40

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2014年 6月 10日	2014年 6月 26日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国	殷跃
传真号 (86-10)62019451	电话号码 (86-10)62089467

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2013/091202

检索报告引用的专利文件	公布日（年/月/日）	同族专利	公布日（年/月/日）
CN 102857416A	2013年 1月 02日	WO 2013185715A1	2013年 12月 19日
US 20130266019A1	2013年 10月 10日	无	
US 8259726B2	2012年 9月 04日	US 2011292939A1	2011年 12月 01日