(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106878168 A
(43)申请公布日 2017.06.20

(21)申请号 201710166638.6

(22)申请日 2017.03.20

(71)申请人 新华三技术有限公司
地址 310052 浙江省杭州市滨江区长河路466号

(72)发明人 宋小恒

(74)专利代理机构 北京柏杉松知识产权代理事务所(普通合伙) 11413
代理人 项京 马敬

(51)Int.Cl.
H04L 12/715(2013.01)
H04L 12/751(2013.01)

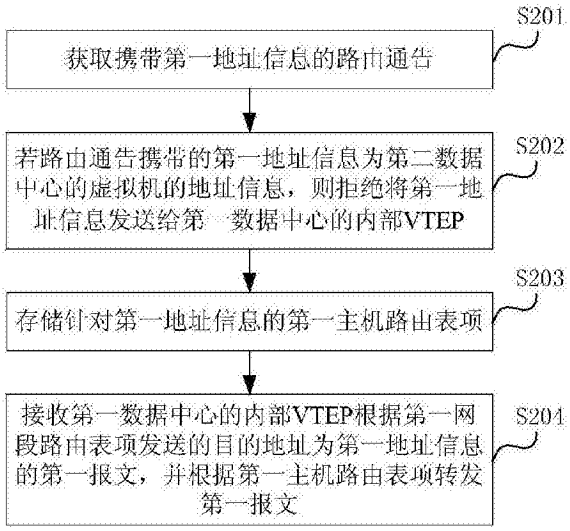
权利要求书3页 说明书12页 附图2页

(54)发明名称

一种报文转发方法及装置

(57)摘要

本发明实施例提供了一种报文转发方法及装置，应用于第一数据中心的第一边界VTEP，包括：获取携带第一地址信息的路由通告；若第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；存储针对第一地址信息的第一主机路由表项；接收第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为第一地址信息的第一报文，并根据第一主机路由表项转发第一报文；第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一网段的报文转发给第一边界VTEP，第一网段为所述第一地址信息所属的网段。应用本发明实施例，提高了分布式数据中心网络的稳定性。



CN 106878168 A

1.一种报文转发方法，其特征在于，应用于第一数据中心的第一边界可扩展虚拟局域网络隧道端点VTEP，所述方法包括：

获取携带第一地址信息的路由通告；

若所述第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP；

存储针对所述第一地址信息的第一主机路由表项；所述第一主机路由表项用于指示将目的地址与所述第一地址信息匹配的报文转发给所述第二数据中心的第二边界VTEP；

接收所述第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为所述第一地址信息的第一报文，并根据所述第一主机路由表项转发所述第一报文；所述第一网段路由表项用于指示将匹配上所述第一网段路由表项包括的第一网段的报文转发给所述第一边界VTEP，所述第一网段为所述第一地址信息所属的网段。

2.根据权利要求1所述的方法，其特征在于，所述若所述第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP的步骤，包括：

判断接收所述路由通告的接口是否为所述第一边界VTEP与第二数据中心的第二边界VTEP间的隧道；

若是，则确定所述第一地址信息为所述第二数据中心的虚拟机的地址信息；

拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP。

3.根据权利要求1或2项所述的方法，其特征在于，在所述获取携带第一地址信息的路由通告的步骤之后，所述方法还包括：

将所述第一地址信息发送给SDN控制器，以使所述SDN控制器生成针对所述第一网段的所述第一网段路由表项和第二网段路由表项，并将所述第一网段路由表项下发给所述第一数据中心的内部VTEP，将所述第二网段路由表项下发给所述第一边界VTEP，其中，所述第二网段路由表项用于指示将匹配上所述第二网段路由表项包括的所述第一网段的报文丢弃；

接收并存储所述SDN控制器下发的所述第二网段路由表项。

4.根据权利要求3所述的方法，其特征在于，所述存储针对所述第一地址信息的第一主机路由表项的步骤，包括：

接收并存储所述SDN控制器下发的所述SDN控制器生成针对所述第一地址信息的第一主机路由表项。

5.一种报文转发方法，其特征在于，应用于第一数据中心的第一内部可扩展虚拟局域网络隧道端点VTEP，所述第一内部VTEP中存储有所述第一数据中心的虚拟机的地址信息和对应的主机路由表项，未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项，所述方法包括：

接收所述第一数据中心的虚拟机发送的第二报文；

若未查找到与所述第二报文的目的地址匹配的主机路由表项，则将所述第二报文发送给所述第一数据中心的第一边界VTEP，以使所述第一边界VTEP将所述第二报文转发给目的虚拟机；所述第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。

6.根据权利要求5所述的方法，其特征在于，所述第一内部VTEP中存储有针对所有数据

中心的虚拟机的地址信息所属网段的网段路由表项;所述网段路由表项用于指示将匹配上所述网段路由表项包括的网段的报文转发给所述第一边界VTEP;

所述若未查找到与所述第二报文的目的地址匹配的主机路由表项,则将所述第二报文发送给所述第一数据中心的第一边界VTEP的步骤,包括:

若未查找到与所述第二报文的目的地址匹配的主机路由表项,则查找与所述第二报文的目的地址所属网段匹配的网段路由表项;

若查找到与所述第二报文的目的地址所属网段匹配的网段路由表项,则根据查找到的网段路由表项,将所述第二报文发送给所述第一数据中心的第一边界VTEP;

若未查找到与所述第二报文的目的地址所属网段匹配的网段路由表项,则丢弃所述第二报文。

7.一种报文转发装置,其特征在于,应用于第一数据中心的第一边界可扩展虚拟局域网络隧道端点VTEP,所述装置包括:

获取单元,用于获取携带第一地址信息的路由通告;

拒绝单元,用于若所述第一地址信息为第二数据中心的虚拟机的地址信息,则拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP;

存储单元,用于存储针对所述第一地址信息的第一主机路由表项;所述第一主机路由表项用于指示将目的地址与所述第一地址信息匹配的报文转发给所述第二数据中心的第二边界VTEP;

转发单元,用于接收所述第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为所述第一地址信息的第一报文,并根据所述第一主机路由表项转发所述第一报文;所述第一网段路由表项用于指示将匹配上所述第一网段路由表项包括的第一网段的报文转发给所述第一边界VTEP,所述第一网段为所述第一地址信息所属的网段。

8.根据权利要求7所述的装置,其特征在于,所述拒绝单元,具体用于:

判断接收所述路由通告的接口是否为所述第一边界VTEP与第二数据中心的第二边界VTEP间的隧道;若是,则确定所述第一地址信息为所述第二数据中心的虚拟机的地址信息;拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP。

9.根据权利要求7或8所述的装置,其特征在于,所述装置还包括:

发送单元,用于在获取携带第一地址信息的路由通告之后,将所述第一地址信息发送给SDN控制器,以使所述SDN控制器生成针对所述第一网段的所述第一网段路由表项和第二网段路由表项,并将所述第一网段路由表项下发给所述第一数据中心的内部VTEP,将所述第二网段路由表项下发给所述第一边界VTEP,其中,所述第二网段路由表项用于指示将匹配上所述第二网段路由表项包括的所述第一网段的报文丢弃;

所述存储单元,还用于接收并存储所述SDN控制器下发的所述第二网段路由表项。

10.根据权利要求9所述的装置,其特征在于,所述存储单元,具体用于:

接收并存储所述SDN控制器下发的所述SDN控制器生成针对所述第一地址信息的第一主机路由表项。

11.一种报文转发装置,其特征在于,应用于第一数据中心的第一内部可扩展虚拟局域网络隧道端点VTEP,所述第一内部VTEP中存储有所述第一数据中心的虚拟机的地址信息和对应的主机路由表项,未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项,

所述装置包括：

接收单元，用于接收所述第一数据中心的虚拟机发送的第二报文；

发送单元，用于若未查找到与所述第二报文的目的地址匹配的主机路由表项，则将所述第二报文发送给所述第一数据中心的第一边界VTEP，以使所述第一边界VTEP将所述第二报文转发给目的虚拟机；所述第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。

12.根据权利要求11所述的装置，其特征在于，所述第一内部VTEP中存储有针对所有数据中心的虚拟机的地址信息所属网段的网段路由表项；所述网段路由表项用于指示将匹配上所述网段路由表项包括的网段的报文转发给所述第一边界VTEP；

所述发送单元，具体用于：

若未查找到与所述第二报文的目的地址匹配的主机路由表项，则查找与所述第二报文的目的地址所属网段匹配的网段路由表项；

若查找到与所述第二报文的目的地址所属网段匹配的网段路由表项，则根据查找到的网段路由表项，将所述第二报文发送给所述第一数据中心的第一边界VTEP；

若未查找到与所述第二报文的目的地址所属网段匹配的网段路由表项，则丢弃所述第二报文。

一种报文转发方法及装置

技术领域

[0001] 本发明涉及通信技术领域，特别是涉及一种报文转发方法及装置。

背景技术

[0002] 如图1所示，在分布式数据中心网络中包括多个数据中心，每个数据中心包括：SDN(Software Defined Network，软件定义网络)控制器(如SDN控制器400、SDN控制器410)、内部VTEP(Virtual eXtensible Local Area Network Tunnel End Point，可扩展虚拟局域网络隧道端点)(如内部VTEP110、内部VTEP120、内部VTEP210、内部VTEP220)、边界VTEP(如边界VTEP100、边界VTEP200)、路由反射器(如路由反射器300、路由反射器310)和虚拟机(如虚拟机510、虚拟机520、虚拟机610、虚拟机620)；其中，一个数据中心内的内部VTEP间、内部VTEP与边界VTEP间运行IBGP(Internal Border Gateway Protocol，内部边界网关协议)用于本数据中心内的虚拟机的地址学习和发布；不同数据中心间采用DCI(Data Center Interconnection，数据中心二层互联)的方式连接，也就是不同数据中心间的边界VTEP间运行EBGP(External Border Gateway Protocol，外部边界网关协议)用于数据中心之间虚拟机的地址学习和发布。

[0003] 现有技术中，若虚拟机间需要进行通信，内部VTEP和边界VTEP需要学习虚拟机的地址(包括IP(Internet Protocol，网络协议)地址和MAC(Media Access Control，媒体访问控制)地址)。具体地，当边界VTEP学习到其他数据中心内的虚拟机的地址后，运行IBGP，向本数据中心内的内部VTEP发送路由通告，将学习到的其他数据中心内的地址告知本数据中心内的所有内部VTEP。此时，本数据中心内的所有内部VTEP和边界VTEP都存储有其他数据中心内的虚拟机的地址；另外，在内部VTEP和边界VTEP获取到针对虚拟机的地址的转发表项后，就可以根据转发表项进行报文转发了。

[0004] 基于上述情况，当分布式数据中心网络的规模较大时，该分布式数据中心网络常常会有虚拟机上线、下线或迁移，为了保证虚拟机间能够进行通信，分布式数据中心网络的各个数据中心中常常会有路由通告，以进行地址学习，这将给数据中心内的内部VTEP的CPU(Central Processing Unit，中央处理器)带来很大的冲击，进而影响分布式数据中心网络的稳定性。

发明内容

[0005] 本发明实施例的目的在于提供一种报文转发方法及装置，以提高分布式数据中心网络的稳定性。具体技术方案如下：

[0006] 一方面，本发明实施例公开了一种报文转发方法，应用于第一数据中心的第一边界VTEP，所述方法包括：

[0007] 获取携带第一地址信息的路由通告；

[0008] 若所述第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP；

[0009] 存储针对所述第一地址信息的第一主机路由表项，其中，所述第一主机路由表项用于指示将目的地址与所述第一地址信息匹配的报文转发给所述第二数据中心的第二边界VTEP；

[0010] 接收所述第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为所述第一地址信息的第一报文，并根据所述第一主机路由表项转发所述第一报文；所述第一网段路由表项用于指示将匹配上所述第一网段路由表项包括的第一网段的报文转发给所述第一边界VTEP，所述第一网段为所述第一地址信息所属的网段。

[0011] 二方面，本发明实施例公开了一种报文转发方法，应用于第一数据中心的第一内部VTEP，所述第一数据中心的内部VTEP中存储有所述第一数据中心的虚拟机的地址信息和对应的主机路由表项，未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项，所述方法包括：

[0012] 接收所述第一数据中心的虚拟机发送的第二报文；

[0013] 若未查找到与所述第二报文的目的地址匹配的主机路由表项，则将所述第二报文发送给所述第一数据中心的第一边界VTEP，以使所述第一边界VTEP将所述第二报文转发给目的虚拟机；所述第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。

[0014] 三方面，本发明实施例公开了一种报文转发装置，应用于第一数据中心的第一边界VTEP，所述装置包括：

[0015] 获取单元，用于获取携带第一地址信息的路由通告；

[0016] 拒绝单元，用于若所述第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将所述第一地址信息发送给所述第一数据中心的内部VTEP；

[0017] 存储单元，用于存储针对所述第一地址信息的第一主机路由表项；所述第一主机路由表项用于指示将目的地址与所述第一地址信息匹配的报文转发给所述第二数据中心的第二边界VTEP；

[0018] 转发单元，用于接收所述第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为所述第一地址信息的第一报文，并根据所述第一主机路由表项转发所述第一报文；所述第一网段路由表项用于指示将匹配上所述第一网段路由表项包括的第一网段的报文转发给所述第一边界VTEP，所述第一网段为所述第一地址信息所属的网段。

[0019] 四方面，本发明实施例公开了一种报文转发装置，应用于第一数据中心的第一内部VTEP，所述第一内部VTEP中存储有所述第一数据中心的虚拟机的地址信息和对应的主机路由表项，未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项，所述装置包括：

[0020] 接收单元，用于接收所述第一数据中心的虚拟机发送的第二报文；

[0021] 发送单元，用于若未查找到与所述第二报文的目的地址匹配的主机路由表项，则将所述第二报文发送给所述第一数据中心的第一边界VTEP，以使所述第一边界VTEP将所述第二报文转发给目的虚拟机；所述第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。

[0022] 本发明实施例中，第一数据中心的第一边界VTEP获得路由通告后，若确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送

给第一数据中心的内部VTEP;此时,第一数据中心的内部VTEP根据第一网段路由表项,将目的地址为第一地址信息的第一报文发送给第一边界VTEP,该第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一地址信息所属第一网段的报文转发给第一边界VTEP;第一边界VTEP根据存储的针对第一地址信息的第一主机路由表项转发第一报文。这样,其他数据中心中虚拟机的地址信息不会在本地数据中心进行通告,降低了数据中心中常常有路由通告的可能性,减小了因过多的路由通告给数据中心内的内部VTEP的CPU带来的冲击,提高了分布式数据中心网络的稳定性。当然,实施本发明的任一产品或方法并不一定需要同时达到以上所述的所有优点。

附图说明

[0023] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附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,还可以根据这些附图获得其他的附图。

[0024] 图1为一种分布式数据中心网络的示意图;

[0025] 图2为本发明实施例提供的一种报文转发方法的流程示意图;

[0026] 图3为本发明实施例提供的另一种报文转发方法的流程示意图;

[0027] 图4为本发明实施例提供的一种报文转发装置的结构示意图;

[0028] 图5为本发明实施例提供的另一种报文转发装置的结构示意图。

具体实施方式

[0029] 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,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0030] 下面通过具体实施例,对本发明进行详细说明。

[0031] 参考图2,图2为本发明实施例提供的一种报文转发方法的流程示意图,应用第一数据中心的第一边界VTEP。这里,第一数据中心可以为分布式数据中心网络中的任一数据中心,如图1所示的数据中心1和数据中心2。具体的,该方法包括:

[0032] S201:获取携带第一地址信息的路由通告;

[0033] 这里,路由通告可以用于告知内部VTEP和/或边界VTEP删除、添加或修改虚拟机的地址信息等,其中,地址信息可以包括:IP地址和MAC地址等。在本发明的一个实施例中,路由通告可以包括ARP(Address Resolution Protocol,地址解析协议)信息。

[0034] S202:若路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息,则拒绝将第一地址信息发送给第一数据中心的内部VTEP;

[0035] 这里,第二数据中心可以为分布式数据中心网络中除第一数据中心外的任一数据中心。

[0036] 实际应用中,第一数据中心的第一边界VTEP获取到的路由通告可以为第一数据中心的内部VTEP通过IBGP发送来的路由通告,也可以为第一边界VTEP根据自身连接的虚拟机

的地址信息生成的路由通告;此时,第一边界VTEP获取到的路由通告中携带的第一地址信息为第一数据中心中内的虚拟机的地址信息;

[0037] 第一数据中心的第一边界VTEP获取到的路由通告还可以为第二数据中心的第二边界VTEP通过EBGP发送来的路由通告;此时,第一边界VTEP获取到的路由通告中携带的第一地址信息为第二数据中心内的虚拟机的地址信息;

[0038] 基于上述情况,在本发明的其他实施例中,边界VTEP可以通过接收路由通告的接口确定路由通告携带的第一地址信息是否为第二数据中心的虚拟机的地址信息。具体的,若第一数据中心的第一边界VTEP接收路由通告的接口为第一边界VTEP与第二数据中心的第二边界VTEP间的隧道,则可以确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址;否则,可以确定路由通告携带的第一地址信息为第一数据中心的虚拟机的地址。

[0039] 另外,为了避免一个数据中心中常常有路由通告,在本发明的一个实施例中,一个数据中心的内部VTEP可以只学习本数据中心内的地址信息,即,第一边界VTEP若确定路由通告中携带的第一地址信息为第二数据中心的虚拟机的地址信息,则不将这个第一地址信息发送给第一数据中心的内部VTEP;若确定路由通告中携带的第一地址信息为第一数据中心的虚拟机的地址信息,则可以通过IBGP在第一数据中心内部发送路由通告,将这个第一地址信息发送给第一数据中心的内部VTEP。

[0040] 这样,只有在有虚拟机在第一数据中心中上线(或下线或迁移)的情况下,第一数据中心中才会存在路由通告,以进行地址学习,而在虚拟机在第二数据中心中上线(或下线或迁移)的情况下,第一数据中心中不会存在路由通告,这有效地避免了一个数据中心中常常有路由通告的问题,提高了分布式数据中心网络的稳定性。

[0041] S203:存储针对第一地址信息的第一主机路由表项;

[0042] 其中,第一主机路由表项可以为一条32位地址的转发表项,其存储在主机路由表中,这里,主机路由表为转发表的一部分;第一主机路由表项用于指示将目的地址与第一地址信息匹配的报文转发给第二数据中心的第二边界VTEP,此时,第一主机路由表项的出接口为第一边界VTEP与第二边界VTEP间的隧道。

[0043] 需要说明的是,主机路由表中不仅可以存储上述第一主机路由表项,还可以存储其他32位地址的主机路由表项,如:指示将报文转发给第一数据中心的内部VTEP的主机路由表项,指示将报文转发给本地端口对应的虚拟机的主机路由表项等。

[0044] 在本发明的一个实施例中,主机路由表项可以由VTEP(包括边界VTEP和内部VTEP)生成。具体地,第一数据中心的第一边界VTEP在获取到携带第一地址信息的路由通告之后,解析路由通告获得第一地址信息,根据获得的第一地址信息及接收该路由通告的接口,生成并存储针对第一地址信息的第一主机路由表项;

[0045] 在本发明的另一个实施例中,主机路由表项可以由SDN控制器生成。具体地,在第一数据中心的第一边界VTEP获取到携带第一地址信息的路由通告后,解析路由通告获得第一地址信息,将获得的第一地址信息发送给SDN控制器;

[0046] SDN控制器若确定接收到的第一地址信息为第二数据中心的虚拟机的地址信息,则生成针对第一地址信息的第一主机路由表项,将第一主机路由表项下发给第一边界VTEP;

[0047] 第一边界VTEP存储接收到的第一主机路由表项。

[0048] 值得一提的是，S203可以在S202之前执行，也可以与S202同时执行，还可以在S202之后执行，本发明实施例对此不进行限定。

[0049] S204：接收第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为第一地址信息的第一报文，并根据第一主机路由表项转发第一报文。

[0050] 其中，第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一网段的报文转发给第一边界VTEP，第一网段为第一地址信息所属的网段，此时，第一网段路由表项的出接口为：第一数据中心的内部VTEP和第一边界VTEP间的隧道。

[0051] 在本发明实施例中，第一数据中心的内部VTEP不会接收到携带第二数据中心的虚拟机的地址信息的路由通告，也就是，第一数据中心的内部VTEP不会学习到第二数据中心的虚拟机的地址信息(包括MAC地址和IP地址等)，无法获得针对第二数据中心的虚拟机的地址信息的转发表项，也就无法向第二数据中心的虚拟机转发报文。

[0052] 这种情况下，为了保证第一数据中心的虚拟机能够正常访问第二数据中心的虚拟机，可以在第一数据中心的内部VTEP上配置针对第二数据中心的虚拟机的地址信息所属网段的网段路由表项，这个网段路由表项为一条转发表项，用于指示将匹配上网段路由表项包括的网段的报文转发给第一数据中心的第一边界VTEP。

[0053] 这样，第一数据中心的内部VTEP就可以基于三层转发报文，也就是，当第一数据中心的内部VTEP接收到针对第二数据中心的虚拟机的报文时，可以根据目的地址(IP地址)匹配的网段路由表项，将接收到的报文转发给第一数据中心的第一边界VTEP；

[0054] 第一边界VTEP根据接收到的报文的目的地址匹配的主机路由表项，将这个接收到的报文转发至第二数据中心的第二边界VTEP，进而将这个接收到的报文发送给目的虚拟机。

[0055] 值得一提的是，在本发明实施例中，为了保证虚拟机间的正常通信，可以设置主机路由表项的优先级高于网段路由表项的优先级，也就是，内部VTEP和边界VTEP在接收到报文后，先匹配主机路由表项，在匹配不到主机路由表项时，再匹配网段路由表项。

[0056] 在本发明的其他实施例中，第一数据中心的第一边界VTEP可能还会接收到目的地址无法匹配到主机路由表项的报文，而第一边界VTEP中存储有所有数据中心的虚拟机的地址，这种情况下，就可以确定这个接收到的报文为无效的报文，需要丢弃这个接收到的报文。

[0057] 为了实现丢弃无效的报文的目的，在第一数据中心的第一边界VTEP中可以预先配置针对所有数据中心的虚拟机的地址信息所属网段的网段路由表项，这个网段路由表项为一条转发表项，用于指示将匹配上网段路由表项包括的这个网段的报文丢弃，该网段路由表项的出接口为丢弃端口。

[0058] 这样，第一数据中心的第一边界VTEP在基于三层转发报文时，若第一边界VTEP接收到目的地址(IP地址)无法匹配到主机路由表项的报文，根据目的地址匹配的网段路由表项，将这个报文丢弃；若第一边界VTEP接收到目的地址匹配到主机路由表项的报文，根据目的地址匹配的主机路由表项，转发这个报文；若第一边界VTEP接收到目的地址无法匹配到主机路由表项且目的地址无法匹配的网段路由表项的报文，则同样丢弃该报文。

[0059] 在本发明的其他实施例中，网段路由表项可以由用户手动配置到第一数据中心的内部VTEP和第一边界VTEP中，也可以由SDN控制器生成并下发给第一数据中心的内部VTEP

和第一边界VTEP。

[0060] 具体的SDN控制器生成网段路由表项的过程包括：

[0061] 在第一数据中心的第一边界VTEP获取到携带第一地址信息的路由通告之后，解析路由通告获得第一地址信息，将获得的第一地址信息发送给SDN控制器；

[0062] SDN控制器根据接收到的第一地址信息，生成针对这个第一地址信息所属第一网段的第一网段路由表项和第二网段路由表项，将第一网段路由表项下发给第一数据中心的内部VTEP，将第二网段路由表项下发给第一边界VTEP；其中，第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一网段的报文转发给第一边界VTEP，第二网段路由表项用于指示将匹配上第二网段路由表项包括的第一网段的报文丢弃；

[0063] 第一边界VTEP接收并存储SDN控制器下发的第二网段路由表项，第一数据中心的内部VTEP接收并存储SDN控制器下发的第一网段路由表项。

[0064] 应用上述实施例，第一数据中心的第一边界VTEP获得路由通告后，若确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；此时，第一数据中心的内部VTEP根据第一网段路由表项，将目的地址为第一地址信息的第一报文发送给第一边界VTEP，该第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一地址信息所属第一网段的报文转发给第一边界VTEP；第一边界VTEP根据存储的针对第一地址信息的第一主机路由表项转发第一报文。这样，其他数据中心中虚拟机的地址信息不会在本地数据中心进行通告，降低了数据中心中常常有路由通告的可能性，减小了因过多的路由通告给数据中心内的内部VTEP的CPU带来的冲击，提高了分布式数据中心网络的稳定性。

[0065] 基于上述报文转发方法，本发明实施例还提供了另一种报文转发方法。参考图3，图3为本发明实施例提供的另一种报文转发方法，该方法应用于第一数据中心的第一内部VTEP，该报文转发方法应用于第一数据中心的内部VTEP，第一数据中心的内部VTEP中存储有第一数据中心的虚拟机的地址信息和对应的主机路由表项，但未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项。具体地，该方法包括：

[0066] S301：接收第一数据中心的虚拟机发送的第二报文；

[0067] S302：若未查找到与第二报文的目的地址匹配的主机路由表项，则将第二报文发送给第一数据中心的第一边界VTEP。

[0068] 第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。这种情况下，第一边界VTEP就可以查询本地的主机路由表项，将第二报文转发给目的虚拟机了。

[0069] 若查找到与第二报文的目的地址匹配的主机路由表项，则将第二报文转发给查找到的主机路由表项中包括的第一数据中心的第二内部VTEP，进而由第二内部VTEP将第二报文转发给目的虚拟机。

[0070] 在本发明的一个实施例中，为了保证第一数据中心内的虚拟机能够正常的访问第二数据中心的虚拟机，第一数据中心的内部VTEP中存储有针对所有数据中心的虚拟机的地址信息所属网段的网段路由表项；这个网段路由表项用于指示将匹配上网段路由表项包括的网段的报文转发给第一数据中心的第一边界VTEP，也就是，这个网段路由表项的出接口为第一数据中心的内部VTEP与第一边界VTEP间的隧道；

[0071] 这种情况下，上述S302可以包括：

[0072] 若未查找到与第二报文的目的地址匹配的主机路由表项，则查找与第二报文的目的地址所属网段匹配的网段路由表项；

[0073] 若查找到与第二报文的目的地址所属网段匹配的网段路由表项，则根据查找到的网段路由表项，将第二报文发送给第一数据中心的第一边界VTEP；

[0074] 若未查找到与第二报文的目的地址所属网段匹配的网段路由表项，则丢弃第二报文。

[0075] 下面结合图1所示分布式数据中心网络，详细说明对本发明实施例提供的转发报文转发的过程；其中，数据中心1中，虚拟机510的IP地址为172.15.1.1、虚拟机520的IP地址为172.16.1.1；数据中心2中，虚拟机610的IP地址为172.15.1.2、虚拟机620的IP地址为172.16.1.2；此时，分布式数据中心网络中包括两个网段，分别为172.15.1.0/24和172.16.1.0/24；主机路由表项、网段路由表项由SDN控制器生成：

[0076] 1、虚拟机510上线后，内部VTEP110获取到虚拟机510的地址信息(包括虚拟机510的IP地址：172.15.1.1)，生成携带172.15.1.1的路由通告，将172.15.1.1上报给SDN控制器400，并通过内部VTEP110和边界VTEP100间的隧道，将携带172.15.1.1的路由通告发送给边界VTEP100，通过内部VTEP110和内部VTEP120间的隧道，将携带172.15.1.1的路由通告发送给内部VTEP120；

[0077] 2、SDN控制器400接收到172.15.1.1后，生成边界VTEP100对应的、针对172.15.1.1的主机路由表项A01，生成内部VTEP110对应的、针对172.15.1.1的主机路由表项A02，生成内部VTEP120对应的、针对172.15.1.1的主机路由表项A03，并将主机路由表项A01下发给边界VTEP100，将主机路由表项A02下发给内部VTEP110，将主机路由表项A03下发给内部VTEP120；

[0078] 其中，主机路由表项A01的出接口为内部VTEP110与边界VTEP100间的隧道，主机路由表项A02的出接口为本地端口，主机路由表项A03的出接口为内部VTEP110与内部VTEP120间的隧道；

[0079] 另外，SDN控制器400生成针对172.15.1.1所属网段172.15.1.0/24的网段路由表项A11、网段路由表项A12和网段路由表项A13，将网段路由表项A11下发给边界VTEP100，将网段路由表项A12下发给内部VTEP110，将网段路由表项A13下发给内部VTEP120；

[0080] 其中，网段路由表项A11的出接口为丢弃端口，网段路由表项A12的出接口为边界VTEP100与内部VTEP110间的隧道，网段路由表项A13的出接口为边界VTEP100与内部VTEP120间的隧道；

[0081] 3、边界VTEP100接收到携带172.15.1.1的路由通告后，存储172.15.1.1，并通过边界VTEP100和边界VTEP200间的隧道，将携带172.15.1.1的路由通告转发给边界VTEP200；

[0082] 4、边界VTEP200接收携带172.15.1.1的路由通告的接口为边界VTEP100和边界VTEP200间的隧道，确定172.15.1.1不是数据中心2的虚拟机的地址信息，在存储172.15.1.1后，拒绝将携带172.15.1.1的路由通告转发给数据中心2的内部VTEP；另外，边界VTEP200将172.15.1.1上报给SDN控制器410；

[0083] 这样，避免了数据中心2中常常有路由通告的问题，减小了对数据中心2内的内部VTEP的CPU的冲击，提高了分布式数据中心网络的稳定性。

[0084] 5、SDN控制器410确定172.15.1.1不是数据中心2的虚拟机的地址信息，生成边界VTEP200对应的、针对172.15.1.1的主机路由表项B01，并将主机路由表项B01下发给边界VTEP200；另外，生成172.15.1.1所属网段172.15.1.0/24的网段路由表项B11、网段路由表项B12和网段路由表项B13，将网段路由表项B11下发给边界VTEP200，将网段路由表项B12下发给内部VTEP210，将网段路由表项B13下发给内部VTEP220；

[0085] 其中，主机路由表项B01的出接口为边界VTEP200与边界VTEP100间的隧道，网段路由表项B11的出接口为丢弃端口，网段路由表项B12的出接口为边界VTEP200与内部VTEP210间的隧道，网段路由表项B13的出接口为边界VTEP200与内部VTEP220间的隧道；

[0086] 6、同理，在虚拟机520、虚拟机610和虚拟机620上线后，各个内部VTEP和边界VTEP间相互学习到的地址信息，这种情况下，各个内部VTEP和边界VTEP学习到的地址信息以及获得的转发表有：

[0087] 61、对于数据中心1，边界VTEP100学习到的地址信息有：172.15.1.1、172.16.1.1、172.15.1.2、172.16.1.2；获得的转发表有：

[0088] 针对172.15.1.1的主机路由表项A01；

[0089] 针对172.16.1.1的主机路由表项A04；

[0090] 针对172.15.1.2的主机路由表项A05；

[0091] 针对172.16.1.2的主机路由表项A06；

[0092] 针对网段172.15.1.0/24的网段路由表项A11；

[0093] 针对网段172.16.1.0/24的网段路由表项A14；

[0094] 其中，主机路由表项A04的出接口为边界VTEP100与内部VTEP120间的隧道，主机路由表项A05的出接口为边界VTEP100与边界VTEP200间的隧道，主机路由表项A06的出接口为边界VTEP100与边界VTEP200间的隧道，网段路由表项A14的出接口为丢弃端口；

[0095] 内部VTEP110学习到的地址信息有：172.15.1.1和172.16.1.1；获得的转发表项有：

[0096] 针对172.15.1.1的主机路由表项A02；

[0097] 针对172.16.1.1的主机路由表项A07；

[0098] 针对网段172.15.1.0/24的网段路由表项A12；

[0099] 针对网段172.16.1.0/24的网段路由表项A15；

[0100] 其中，主机路由表项A07的出接口为内部VTEP110与内部VTEP120间的隧道，网段路由表项A15的出接口为边界VTEP100与内部VTEP110间的隧道；

[0101] 内部VTEP120学习到的地址信息有：172.15.1.1和172.16.1.1；获得的转发表项有：

[0102] 针对172.15.1.1的主机路由表项A03；

[0103] 针对172.16.1.1的主机路由表项A08；

[0104] 针对网段172.15.1.0/24的网段路由表项A13；

[0105] 针对网段172.16.1.0/24的网段路由表项A16；

[0106] 其中，主机路由表项A08的出接口为本地端口，网段路由表项A16的出接口为边界VTEP100与内部VTEP120间的隧道；

[0107] 62、对于数据中心2，边界VTEP200学习到的地址信息有：172.15.1.1、172.16.1.1、

172.15.1.2、172.16.1.2;获得的转发表项有:

[0108] 针对172.15.1.1的主机路由表项B01;

[0109] 针对172.16.1.1的主机路由表项B02;

[0110] 针对172.15.1.2的主机路由表项B03;

[0111] 针对172.16.1.2的主机路由表项B04;

[0112] 针对网段172.15.1.0/24的网段路由表项B11;

[0113] 针对网段172.16.1.0/24的网段路由表项B14

[0114] 其中,主机路由表项B02的出接口为边界VTEP100与边界VTEP200间的隧道,主机路由表项B03的出接口为边界VTEP200与内部VTEP210间的隧道,主机路由表项B04的出接口为边界VTEP200与内部VTEP220间的隧道,网段路由表项B12的出接口为丢弃端口;

[0115] 内部VTEP210学习到的地址信息有:172.15.1.2和172.16.1.2;获得的转发表项有:

[0116] 针对172.15.1.2的主机路由表项B05;

[0117] 针对172.16.1.2的主机路由表项B06;

[0118] 针对网段172.15.1.0/24的网段路由表项B12

[0119] 针对网段172.16.1.0/24的网段路由表项B15

[0120] 其中,主机路由表项B05出接口为本地端口,主机路由表项B06出接口为内部VTEP210与内部VTEP220间的隧道,网段路由表项B14的出接口为边界VTEP200与内部VTEP210间的隧道;

[0121] 内部VTEP220学习到的地址信息有:172.15.1.2和172.16.1.2;获得的转发表项有:

[0122] 针对172.15.1.2的主机路由表项B07;

[0123] 针对172.16.1.2的主机路由表项B08;

[0124] 针对网段172.15.1.0/24的网段路由表项B13;

[0125] 针对网段172.16.1.0/24的网段路由表项B16;

[0126] 其中,主机路由表项B07出接口为内部VTEP210与内部VTEP220间的隧道,主机路由表项B08出接口为本地端口,网段路由表项B16出接口为边界VTEP200与内部VTEP220间的隧道;

[0127] 7、当虚拟机510访问虚拟机610时,虚拟机510获取到虚拟机610的IP地址172.15.1.2后,根据172.15.1.2构建目的IP地址为172.15.1.2的报文1;将报文1发送给内部VTEP110;

[0128] 内部VTEP110接收到报文1后,无法获取到与172.15.1.2匹配的主机路由表项,根据报文1的目的IP地址172.15.1.2,匹配网段路由表项,获得网段路由表项A12,网段路由表项A12的出接口为边界VTEP100与内部VTEP110间的隧道,通过边界VTEP100与内部VTEP110间的隧道将报文1发送给边界VTEP100;

[0129] 边界VTEP100根据报文1的目的IP地址172.15.1.2,匹配到主机路由表项A05,主机路由表项A05的出接口为边界VTEP100与边界VTEP200间的隧道,通过边界VTEP100与边界VTEP200间的隧道将报文1发送给边界VTEP200;

[0130] 边界VTEP200接收到报文1后,根据报文1的目的IP地址172.15.1.2,匹配到主机路

由表项B03，主机路由表项B03的出接口为边界VTEP200与内部VTEP210间的隧道，通过边界VTEP200与内部VTEP210间的隧道将报文1发送给内部VTEP210；

[0131] 内部VTEP210根据报文1的目的IP地址172.15.1.2，匹配到主机路由表项B05，主机路由表项B05的出接口为本地端口，通过本地端口，将报文1发送给虚拟机610。

[0132] 8、当虚拟机510访问虚拟机520时，虚拟机510获取到虚拟机610的IP地址172.16.1.1后，根据172.16.1.1构建目的IP地址为172.16.1.1的报文2；将报文2发送给内部VTEP110；

[0133] 内部VTEP110接收到报文2后，获取到与172.16.1.1匹配的主机路由表项A07，主机路由表项A07的出接口为内部VTEP110与内部VTEP120间的隧道，通过内部VTEP110与内部VTEP120间的隧道将报文2发送给内部VTEP120；

[0134] 内部VTEP120接收到报文2后，根据报文2的目的IP地址172.16.1.1，匹配到主机路由表项A08，主机路由表项A08的出接口为本地端口，通过本地端口，将报文2发送给本地用户侧的虚拟机520。

[0135] 值得一提的是，为了防止报文在数据中心中形成环路，可以在内部VTEP上配置以下规则：对于接收的边界VTEP发送的报文仅在本地的用户侧进行转发，不再进入隧道。

[0136] 应用上述实施例，第一数据中心的第一边界VTEP获得路由通告后，若确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；此时，第一数据中心的内部VTEP根据第一网段路由表项，将目的地址为第一地址信息的第一报文发送给第一边界VTEP，该第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一地址信息所属第一网段的报文转发给第一边界VTEP；第一边界VTEP根据存储的针对第一地址信息的第一主机路由表项转发第一报文。这样，其他数据中心中虚拟机的地址信息不会在本地数据中心进行通告，降低了数据中心中常常有路由通告的可能性，减小了因过多的路由通告给数据中心内的内部VTEP的CPU带来的冲击，提高了分布式数据中心网络的稳定性。

[0137] 参考图4，图4为本发明实施例提供的一种报文转发装置的结构示意图，应用于第一数据中心的第一边界VTEP，该装置包括：

[0138] 获取单元401，用于获取携带第一地址信息的路由通告；

[0139] 拒绝单元402，用于若第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；

[0140] 存储单元403，用于存储针对第一地址信息的第一主机路由表项；第一主机路由表项用于指示将目的地址与第一地址信息匹配的报文转发给第二数据中心的第二边界VTEP；

[0141] 转发单元404，用于接收第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为第一地址信息的第一报文，并根据第一主机路由表项转发第一报文；第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一网段的报文转发给第一边界VTEP，第一网段为第一地址信息所属的网段。

[0142] 在本发明的其他实施例中，拒绝单元402，具体可以用于：

[0143] 判断接收路由通告的接口是否为第一边界VTEP与第二数据中心的第二边界VTEP间的隧道；若是，则确定第一地址信息为第二数据中心的虚拟机的地址信息；拒绝将第一地址信息发送给第一数据中心的内部VTEP。

[0144] 在本发明的其他实施例中，上述报文转发装置还可以包括：

[0145] 发送单元(图4中未示出)，用于在获取携带第一地址信息的路由通告之后，将第一地址信息发送给SDN控制器，以使SDN控制器生成针对第一网段的第一网段路由表项和第二网段路由表项，并将第一网段路由表项下发给第一数据中心的内部VTEP，将第二网段路由表项下发给第一边界VTEP，其中，第二网段路由表项用于指示将匹配上第二网段路由表项包括的第一网段的报文丢弃；

[0146] 这种情况下，存储单元403，还可以用于接收并存储SDN控制器下发的第二网段路由表项。

[0147] 在本发明的其他实施例中，存储单元403，具体可以用于：

[0148] 接收并存储SDN控制器下发的SDN控制器生成针对第一地址信息的第一主机路由表项。

[0149] 应用上述实施例，第一数据中心的第一边界VTEP获得路由通告后，若确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；此时，第一数据中心的内部VTEP根据第一网段路由表项，将目的地址为第一地址信息的第一报文发送给第一边界VTEP，该第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一地址信息所属第一网段的报文转发给第一边界VTEP；第一边界VTEP根据存储的针对第一地址信息的第一主机路由表项转发第一报文。这样，其他数据中心中虚拟机的地址信息不会在本地数据中心进行通告，降低了数据中心中常常有路由通告的可能性，减小了因过多的路由通告给数据中心内的内部VTEP的CPU带来的冲击，提高了分布式数据中心网络的稳定性。

[0150] 参考图5，图5为本发明实施例提供的另一种报文转发装置的结构示意图，应用于第一数据中心的第一内部VTEP，第一内部VTEP中存储有第一数据中心的虚拟机的地址信息和对应的主机路由表项，未存储其它数据中心的虚拟机的地址信息和对应的主机路由表项，该装置包括：

[0151] 接收单元501，用于接收第一数据中心的虚拟机发送的第二报文；

[0152] 发送单元502，用于若未查找到与第二报文的目的地址匹配的主机路由表项，则将第二报文发送给第一数据中心的第一边界VTEP，以使第一边界VTEP将第二报文转发给目的虚拟机；其中，第一边界VTEP中存储有所有数据中心的虚拟机的地址信息和对应的主机路由表项。

[0153] 在本发明的其他实施例中，第一内部VTEP中存储有针对所有数据中心的虚拟机的地址信息所属网段的网段路由表项；网段路由表项用于指示将匹配上网段路由表项包括的网段的报文转发给第一边界VTEP；

[0154] 这种情况下，发送单元502，具体可以用于：

[0155] 若未查找到与第二报文的目的地址匹配的主机路由表项，则查找与第二报文的目的地址所属网段匹配的网段路由表项；

[0156] 若查找到与第二报文的目的地址所属网段匹配的网段路由表项，则根据查找到的网段路由表项，将第二报文发送给第一数据中心的第一边界VTEP；

[0157] 若未查找到与第二报文的目的地址所属网段匹配的网段路由表项，则丢弃第二报文。

[0158] 应用上述实施例，第一数据中心的第一边界VTEP获得路由通告后，若确定路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP；此时，第一数据中心的内部VTEP根据第一网段路由表项，将目的地址为第一地址信息的第一报文发送给第一边界VTEP，该第一网段路由表项用于指示将匹配上第一网段路由表项包括的第一地址信息所属第一网段的报文转发给第一边界VTEP；第一边界VTEP根据存储的针对第一地址信息的第一主机路由表项转发第一报文。这样，其他数据中心中虚拟机的地址信息不会在本地数据中心进行通告，降低了数据中心中常常有路由通告的可能性，减小了因过多的路由通告给数据中心内的内部VTEP的CPU带来的冲击，提高了分布式数据中心网络的稳定性。

[0159] 需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0160] 本说明书中的各个实施例均采用相关的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0161] 以上仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内所作的任何修改、等同替换、改进等，均包含在本发明的保护范围内。

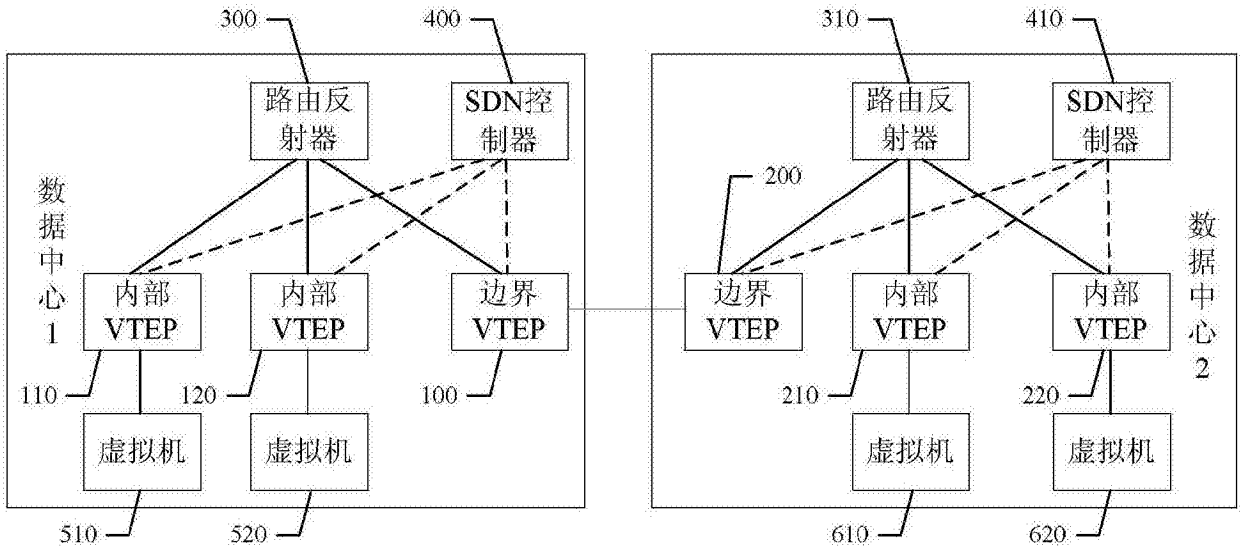

300
400
310
410
路由反射器
SDN控制器
路由反射器
SDN控制器
200
数据中心1
内部VTEP
内部VTEP
边界VTEP
边界VTEP
内部VTEP
内部VTEP
数据中心2
110
120
100
210
220
虚拟机
虚拟机
虚拟机
虚拟机
510
520
610
620

图1

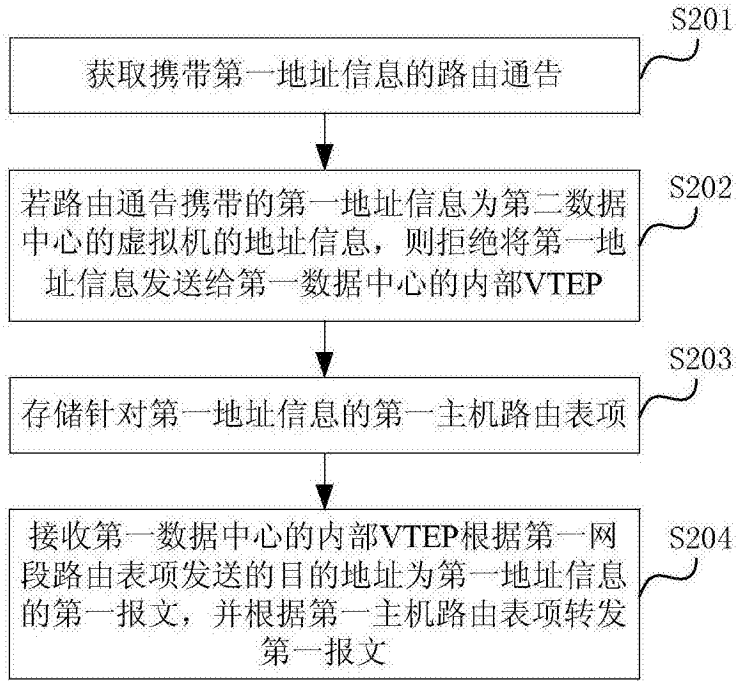

S201
获取携带第一地址信息的路由通告
S202
若路由通告携带的第一地址信息为第二数据中心的虚拟机的地址信息，则拒绝将第一地址信息发送给第一数据中心的内部VTEP
S203
存储针对第一地址信息的第一主机路由表项
S204
接收第一数据中心的内部VTEP根据第一网段路由表项发送的目的地址为第一地址信息的第一报文，并根据第一主机路由表项转发第一报文

图2

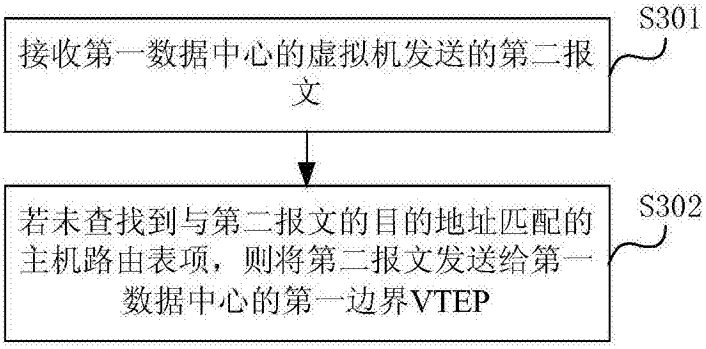

S301
接收第一数据中心的虚拟机发送的第二报文
S302
若未查找到与第二报文的目的地址匹配的主机路由表项，则将第二报文发送给第一数据中心的第一边界VTEP

图3

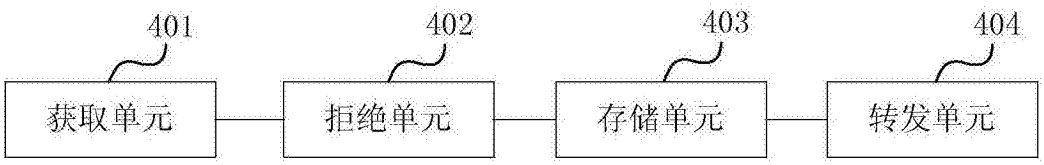

401
获取单元
402
拒绝单元
403
存储单元
404
转发单元

图4

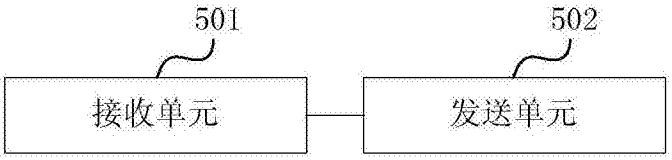

501
接收单元
502
发送单元

图5