(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 104283756 A
(43)申请公布日 2015.01.14

(21)申请号 201310277240.1

(22)申请日 2013.07.02

(71)申请人 杭州华三通信技术有限公司
地址 310053 浙江省杭州市高新技术产业开发区之江科技工业园六和路 310 号华为杭州生产基地

(72)发明人 王松波 林涛 张寅飞 任维春

(74)专利代理机构 北京德琦知识产权代理有限公司 11018
代理人 谢安昆 宋志强

(51)Int.Cl.
H04L 12/46(2006.01)
H04L 12/823(2013.01)

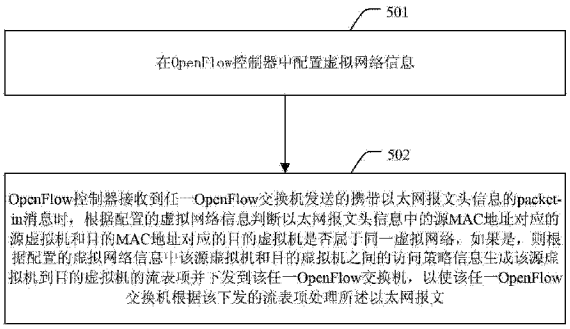
权利要求书4页 说明书10页 附图4页

(54)发明名称

一种实现分布式多租户虚拟网络的方法和装置

(57)摘要

本发明提供了一种实现分布式多租户虚拟网络的方法和装置，技术方案为：利用 OpenFlow 系统架构，通过 OpenFlow 协议部署数据中心的多租户分布式虚拟网络。系统由部署在一台服务器中的 OpenFlow 控制器和部署在其它各服务器上的 OpenFlow 交换机组成。在 OpenFlow 控制器中配置虚拟网络信息并根据配置的虚拟网络信息控制虚拟网络中虚拟网络实例之间的报文转发流程。本发明能够减少以太网报文负荷，并能够支持虚拟网络的无限扩展。



CN 104283756 A

1. 一种实现分布式多租户虚拟网络的方法，其特征在于，所述多租户虚拟网络中的一台服务器上部署有 OpenFlow 控制器，其它服务器中部署有虚拟机和 OpenFlow 交换机，该方法包括：

在 OpenFlow 控制器中配置虚拟网络的信息，所述虚拟网络的信息包括该虚拟网络中所有虚拟机信息以及虚拟机之间的访问策略信息，所述虚拟机信息包括该虚拟机的 MAC 地址；

OpenFlow 控制器接收到任一 OpenFlow 交换机发送的携带以太网报文的报文头信息的 packet-in 消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机是否属于同一虚拟网络，如果是，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该任一 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文；

其中，所述 packet-in 消息是该任一 OpenFlow 交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后发送的。

2. 根据权利要求 1 所述的方法，其特征在于，

所述虚拟机信息还包括虚拟机连接的 OpenFlow 交换机及连接端口信息，所述 OpenFlow 交换机为支持虚拟边缘端口汇聚 VEPA 转发模式的虚拟交换机，所述 OpenFlow 交换机通过一上行接口连接至外部的物理交换机；

所述生成该源虚拟机到目的虚拟机的流表项，包括：

在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第一流表项，该第一流表项包头域包括：源 MAC 地址为源虚拟机的 MAC 地址；目的 MAC 地址为目的虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至该源虚拟机的下行接口；该第一流表项的行动包括：通过该任一 OpenFlow 交换机上连接至外部物理交换机的上行接口转发所匹配的报文；

在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第二流表项，该第二流表项包头域与所述第一流表项相同，该第二流表项的行动包括：丢弃所匹配的报文。

3. 根据权利要求 2 所述的方法，其特征在于，

根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该 OpenFlow 交换机时，进一步根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该目的虚拟机到源虚拟机的流表项并下发到该 OpenFlow 交换机。

4. 根据权利要求 3 所述的方法，其特征在于，

所述生成该目的虚拟机到源虚拟机的流表项，包括：

在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第三流表项，该第三流表项包头域包括：源 MAC 地址为目的虚拟机的 MAC 地址；目的 MAC 地址为源虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至外部物理交换机的的上行接口；该第三流表项的行动包括：通过该任一 OpenFlow 交换机上连接至该源虚拟机的的下行接口转发所匹配的报文；

在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第四流表项，该第四流表项包头域与所述第三流表项相同，该第四流表项的行动包括：丢弃所匹配的报文。

5. 根据权利要求 4 所述的方法，其特征在于，该方法进一步包括：

OpenFlow 控制器接收到该任一 OpenFlow 交换机发送的虚拟机迁移事件通知时，更新发生迁移的虚拟机的虚拟机信息，并根据该虚拟机当前连接的 OpenFlow 交换机及连接端口信息，更新该虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项，并将生成的流表项下发到该任一 OpenFlow 交换机；其中，所述虚拟机迁移事件通知是该任一 OpenFlow 交换机检测到有虚拟机迁移到该任一 OpenFlow 交换机所在服务器时发送的。

6. 根据权利要求 1 所述的方法，其特征在于，

所述根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该 OpenFlow 交换机时，进一步向该 OpenFlow 交换机发送携带所述以太网报文的出接口信息的 packet-out 消息，以使该 OpenFlow 交换机在接收到该 Packet-out 消息的情况下根据该 packet-out 消息携带的出接口信息转发所述以太网报文，未接收到该 Packet-out 消息的情况下根据该下发的流表项处理所述以太网报文。

7. 一种在数据中心网络中实现分布式多租户虚拟网络的装置，其特征在于，所述数据中心网络中的一台服务器上部署有 OpenFlow 控制器，其它服务器中部署有虚拟机和 OpenFlow 交换机，该装置应用于所述 OpenFlow 控制器，包括：配置单元、接收单元、判断单元、控制单元、发送单元；

所述配置单元，用于在所述 OpenFlow 控制器中配置虚拟网络的信息，所述虚拟网络的信息包括该虚拟网络中所有虚拟机信息以及虚拟机之间的访问策略信息，所述虚拟机信息包括该虚拟机的 MAC 地址；

所述接收单元，用于接收任一 OpenFlow 交换机发送的携带以太网报文的报文头信息的 packet-in 消息；所述 packet-in 消息是该 OpenFlow 交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后发送的；

所述判断单元，用于接收单元接收到该 OpenFlow 交换机发送的携带以太网报文头信息的 packet-in 消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机是否属于同一虚拟网络；

所述控制单元，用于判断单元判定所述以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机属于同一虚拟网络时，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项；

所述发送单元，用于将控制单元生成的该源虚拟机到目的虚拟机的流表项下发到该任一 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文。

8. 根据权利要求 7 所述的装置，其特征在于，

所述虚拟机信息还包括虚拟机连接的 OpenFlow 交换机及连接端口信息，所述 OpenFlow 交换机为支持虚拟边缘端口汇聚 VEPA 转发模式的虚拟交换机，所述 OpenFlow 交

换机通过一上行接口连接至外部的物理交换机；

所述控制单元生成该源虚拟机到目的虚拟机的流表项，包括：

在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第一流表项，该第一流表项包头域包括：源 MAC 地址为源虚拟机的 MAC 地址；目的 MAC 地址为目的虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至该源虚拟机的下行接口；该第一流表项的行动包括：通过该任一 OpenFlow 交换机上连接至外部物理交换机的上行接口转发所匹配的报文；

在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第二流表项，该第二流表项包头域与所述第一流表项相同，该第二流表项的行动包括：丢弃所匹配的报文。

9. 根据权利要求 8 所述的装置，其特征在于，

所述控制单元在根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该 OpenFlow 交换机时，进一步根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该目的虚拟机到源虚拟机的流表项；

所述发送单元，用于将控制单元生成的该目的虚拟机到源虚拟机的流表项下发到该任一 OpenFlow 交换机。

10. 根据权利要求 9 所述的装置，其特征在于，

所述控制单元生成该目的虚拟机到源虚拟机的流表项，包括：

在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第三流表项，该第三流表项包头域包括：源 MAC 地址为目的虚拟机的 MAC 地址；目的 MAC 地址为源虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至外部物理交换机的的上行接口；该第三流表项的行动包括：通过该任一 OpenFlow 交换机上连接至该源虚拟机的的下行接口转发所匹配的报文；

在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第四流表项，该第四流表项包头域与所述第三流表项相同，该第四流表项的行动包括：丢弃所匹配的报文。

11. 根据权利要求 10 所述的装置，其特征在于，

所述接收单元，进一步用于接收该任一 OpenFlow 交换机发送的虚拟机迁移事件通知；

所述控制单元，进一步用于接收单元接收到该任一 OpenFlow 交换机发送的虚拟机迁移事件通知时，更新发生迁移的虚拟机的虚拟机信息，并根据该虚拟机当前连接的 OpenFlow 交换机及连接端口信息，更新该虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项；其中，所述虚拟机迁移事件通知是该任一 OpenFlow 交换机检测到有虚拟机迁移到该任一 OpenFlow 交换机所在服务器时发送的；

所述发送单元，用于将控制单元生成该发生迁移的虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项下发到该任一 OpenFlow 交换机。

12. 根据权利要求 11 所述的装置，其特征在于，

所述发送单元，用于将控制单元根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项下发到该任一 OpenFlow 交换

机时，向该任一 OpenFlow 交换机发送携带所述以太网报文的出接口信息的 packet-out 消息，以使该任一 OpenFlow 交换机在接收到该 Packet-out 消息的情况下根据该 packet-out 消息携带的出接口信息转发所述以太网报文，未接收到该 Packet-out 消息的情况下根据该下发的流表项处理所述以太网报文。

一种实现分布式多租户虚拟网络的方法和装置

技术领域

[0001] 本申请涉及通信技术领域，特别涉及一种实现分布式多租户虚拟网络的方法和装置。

背景技术

[0002] 数据中心的虚拟化技术主要包括 3 方面内容：网络虚拟化、存储虚拟化和服务器虚拟化，最主要的是服务器虚拟化。通过专用的虚拟化软件（如 VMware）管理，一台物理服务器能虚拟出多台虚拟机 VM，每个 VM 独立运行，互不影响，都有自己的操作系统和应用程序和虚拟的硬件环境，包括虚拟 CPU、内存、存储设备、IO 设备、虚拟交换机等。其中，虚拟交换机的应用日益广泛。

[0003] 目前，在虚拟交换机中开始应用边缘虚拟桥接技术 EVB。EVB 技术分为交换机 EVB 技术和服务器 EVB 技术。服务器 EVB 技术应用于数据中心服务器，在其上的虚拟交换机中实现，用于简化虚拟服务器的流量转发实现，对虚拟服务器的网络交换、流量管理和策略下发进行集中控制，并能在虚拟迁移时实现网络管理和策略的自动迁移。支持 EVB 的虚拟交换机分为 VEB（Virtual Ethernet Bridge，虚拟边缘交换机）和 VEPA（Virtual Edge Port Aggregator，虚拟边缘端口汇聚）。

[0004] VEPA 将虚拟机产生的网络流量全部交由与服务器相连的物理交换机进行处理，即使同一台服务器上的虚拟机间流量，也将在物理交换机上查表处理后，再回到目的虚拟机上。VEPA 方式不仅借助物理交换机解决了虚拟机间流量转发，同时还实现了对虚拟机流量的监管，并且将虚拟机接入层网络纳入到传统服务器接入网络管理体系中。

[0005] OpenFlow 技术最早由斯坦福大学提出，旨在基于现有 TCP/IP 技术条件，以创新的网络互联理念解决当前网络面对新业务产生的种种瓶颈。OpenFlow 的核心思想是将原本完全由交换机 / 路由器控制的数据包转发过程，转化为由 OpenFlow 交换机和 OpenFlow 控制器（Controller）来共同完成，从而实现了数据转发和路由控制的分离。OpenFlow 技术的提出给网络虚拟化的发展带来无限可能。

[0006] 参见图 1，图 1 是现有技术 OpenFlow 组网示意图，其中包括通过安全通道互联的 OpenFlow 交换机和 OpenFlow 控制器。OpenFlow 交换机在本地维护一个流表（Flow Table），流表由很多个流表项组成。每个流表项就是一个转发规则，由包头域（header field）、计数器（counters）和行动（action）组成；其中包头域可以包括进入接口，源 MAC 地址、目标 MAC 地址、类型，vlan id，vlan 优先级，源 IP 地址、目标 IP 地址、协议、IP ToS 位，TCP/UDP 目标端口、TCP/UDP 源端口等多个域，是流表项的标识；计数器用来计数流表项的统计数据；行动标明了与该流表项匹配的数据包应该执行的操作。数据包进入 OpenFlow 交换机之后，通过查询流表来获得转发的目的端口：如果在流表中有对应项，则直接进行快速转发；如果在流表中没有对应项，数据包就会被发送到控制服务器进行传输路径的确认，再根据下发结果进行转发。

[0007] 在现代化企业中，多个部门的信息数据共享数据中心网络，不同部门的信息数据

具有不同转发策略，构成多租户虚拟网络。如图 2 所示的数据中心网络示意图，其中包括站点网络 A 和站点网络 B，租户 1 和租户 2 共享站点网络 A 和站点网络 B 中的硬件设施(路由器，交换机)，使得数据中心网络成为多租户虚拟网络，在数据中心网络中，租户 A 和租户 B 之间的应用程序以及数据是相互隔离的。

[0008] 目前多采用 VLAN 实现多租户虚拟网络。IEEE802.Q 规范定义了 VLAN 网桥操作，允许在桥接局域网结构中实现定义、运行以及管理 VLAN 拓扑结构等操作，为标识带有 VLAN 信息的以太帧建立了一种标准方法。IEEE802.Q 使用 VLAN 将网络划分成多个逻辑单元，通过为每个租户分配一个或多个 VLAN，使得每个租户可以拥有独立于其它租户的数据中心网络。

[0009] 采用 VLAN 实现多租户虚拟网络具有以下缺点 ：以太网报文中需要携带 VLAN 标签，导致报文负荷比较大 ；IEEE802.1Q 规定仅支持 4K 个 VLAN，使得虚拟网络的扩展受限。

发明内容

[0010] 有鉴于此，本发明的目的在于提供一种实现分布式多租户虚拟网络的方法，该方法能够减少以太网报文负荷，并能够支持虚拟网络的无限扩展。

[0011] 为实现上述目的，本发明提供的技术方案为 ：

[0012] 一种实现分布式多租户虚拟网络的方法，所述多租户虚拟网络中的一台服务器上部署有 OpenFlow 控制器，其它服务器中部署有虚拟机和 OpenFlow 交换机，该方法包括 ：

[0013] 在 OpenFlow 控制器中配置虚拟网络的信息，所述虚拟网络的信息包括该虚拟网络中所有虚拟机信息以及虚拟机之间的访问策略信息，所述虚拟机信息包括该虚拟机的 MAC 地址 ；

[0014] OpenFlow 控制器接收到任一 OpenFlow 交换机发送的携带以太网报文的报文头信息的 packet-in 消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机是否属于同一虚拟网络，如果是，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该任一 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文 ；

[0015] 其中，所述 packet-in 消息是该任一 OpenFlow 交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后发送的。

[0016] 一种实现分布式多租户虚拟网络的装置，所述数据中心网络中的一台服务器上部署有 OpenFlow 控制器，其它服务器中部署有虚拟机和 OpenFlow 交换机，该装置应用于所述 OpenFlow 控制器，包括 ：配置单元、接收单元、判断单元、控制单元、发送单元 ；

[0017] 所述配置单元，用于在所述 OpenFlow 控制器中配置虚拟网络的信息，所述虚拟网络的信息包括该虚拟网络中所有虚拟机信息以及虚拟机之间的访问策略信息，所述虚拟机信息包括该虚拟机的 MAC 地址 ；

[0018] 所述接收单元，用于接收任一 OpenFlow 交换机发送的携带以太网报文的报文头信息的 packet-in 消息 ；所述 packet-in 消息是该 OpenFlow 交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后发送的 ；

[0019] 所述判断单元，用于接收单元接收到该 OpenFlow 交换机发送的携带以太网报文

头信息的 packet-in 消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机是否属于同一虚拟网络；

[0020] 所述控制单元，用于判断单元判定所述以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机属于同一虚拟网络时，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项；

[0021] 所述发送单元，用于将控制单元生成的该源虚拟机到目的虚拟机的流表项下发到该任一 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文。

[0022] 综上所述，本发明通过在数据中心网络中的一台服务器上配置 OpenFlow 控制器，在其它服务器中配置 OpenFlow 交换机，形成 OpenFlow 系统架构；通过在 OpenFlow 控制器中配置虚拟网络信息，并在 OpenFlow 交换机转发虚拟网络中的虚拟机之间的以太网报文且未查找到对应流表项时，由控制服务器根据配置的虚拟网络信息进行同一虚拟网络内的转发确认及相应流表项的生成和下发，使该 OpenFlow 交换机可以根据下发的流表项转发以太网报文。本发明不再使用传统虚拟网络基于 IEEE802.1Q 协议创建 VLAN 的管理方法，不需要在以太网报文中携带 VLAN 标签，因此可以减少以太网报文负荷；另外，本发明是通过在控制服务器中配置虚拟网络信息并根据配置的虚拟网络信息实现对虚拟网络的管理，不再受 IEEE802.1Q 中 VLAN 标签只支持 4096 个 VLAN 的方式，因此可以无限制的扩展虚拟网络。

附图说明

[0023] 图 1 是现有技术 OpenFlow 组网示意图；

[0024] 图 2 是现有技术数据中心网络示意图；

[0025] 图 3 是本发明实施例分布式多租户虚拟网络的系统架构示意图；

[0026] 图 4 是本发明实施例报文转发流程示意图；

[0027] 图 5 是本发明实施例实现分布式多租户虚拟网络的方法流程图；

[0028] 图 6 是本发明实施例实现分布式多租户虚拟网络的装置的结构示意图。

具体实施方式

[0029] 为使本发明的目的、技术方案及优点更加清楚明白，以下参照附图并举实施例，对本发明所述方案作进一步地详细说明。

[0030] 本发明实施例中，利用 OpenFlow 系统架构，通过 OpenFlow 协议部署数据中心的多租户分布式虚拟网络。系统由控制服务器和部署在各服务器上的 OpenFlow 交换机组成，在控制服务器中配置虚拟网络信息并由控制服务器管理虚拟网络。

[0031] 参见图 3，图 3 是本发明实施例分布式多租户虚拟网络的系统架构示意图，如图 3 所示，系统包括部署在服务器 0 上的 OpenFlow 控制器和部署在服务器 1 上 OpenFlow 交换机以及部署在服务器 2 上的 OpenFlow 交换机，为了实现多租户，在服务器 1 中还部署了多个虚拟机：VM-A、VM-B、VM-C、VM-D，在服务器 2 中部署了多个虚拟机：VM-A’、VM-B’、VM-C’、VM-D’。将 VM-A、VM-B、VM-A’、VM-B’ 划归到 Rose 虚拟网络，将 VM-C、VM-D、VM-C’、VM-D’ 划

归到 Bob 虚拟网络，从而形成由 Rose 和 Bob 两个虚拟网络。图 3 中，服务器 1 和服务器 2 上部署的 OpenFlow 交换机通过一上行接口（uplink 口）连接至外部的物理交换机，同时还通过一下行接口与自身所在服务器上部署的各个 VM 对应连接。

[0032] 在本发明实施例中，要实现图 3 所示的系统架构，首先需要在控制服务器中配置虚拟网络信息：Rose 虚拟网络信息和 Bob 虚拟网络信息。在实际应用中，可以根据实际需求在控制服务器中配置多个虚拟网络的虚拟网络信息。配置的虚拟网络信息中可以包括以下内容：虚拟网络标识、虚拟网络的网关 IP 地址及掩码、该虚拟网络中的所有虚拟机信息、以及虚拟机之间的访问策略信息等，其中，虚拟机信息包括虚拟机标识、虚拟机的 MAC 地址、虚拟机连接的 OpenFlow 交换机及连接端口等信息。这里，所述连接端口是指该 OpenFlow 交换机上与虚拟机连接的下行接口。

[0033] 在控制服务器中配置虚拟网络信息的例子如下（采用 Json 格式）：

[0034]

```
virtual Network ID: Rose
Gateway: 192.168.10.1
Gateway Mask: 255.255.255.0
Instance ID:    VM-A
        Mac address:00.00.00.00.00.01    /*VM-A 的 MAC 地址*/
Instance ID:  VM-B
        Mac address:00.00.00.00.00.02    /*VM-B 的 MAC 地址*/
Instance ID:  VM-A'
        Mac address:00.00.00.00.00.03    /*VM-A'的 MAC 地址*/
Instance ID:  VM-B'
        Mac Address:  00.00.00.00.04    /* VM-B'的 MAC 地址*/
Policy-ID:   Rose-Policy-test1
      VM-A deny VM-B;
      VM-A permit VM-A'
```

[0035] 其中，

[0036] virtual Network ID 表示虚拟网络标识，其值为 Rose，代表标识为 Rose 的虚拟网络；

[0037] Instance ID 表示虚拟机标识，在上述配置信息中，虚拟机标识有四个：VM-A、VM-B、VM-A'、VM-B'，代表四个虚拟机；紧跟在每个 Instance ID 下面一行的 MAC Address 表示该 Instance ID 代表的虚拟机的 MAC 地址，例如 VM-A 代表的虚拟机的 MAC 地址为 00.00.00.00.00.01；

[0038] Policy-ID 表示该虚拟网络的访问策略标识，其值为 Rose-Policy-test1，在 Rose-Policy-test1 代表的访问策略中，“VM-A deny VM-B”表示禁止 VM-A 和 VM-B 互访；“VM-A permit VM-A'”表示允许 VM-A 和 VM-A' 互访。在实际应用中，这些访问策略可以根

据实际需求定义，例如允许 VM-A 到 VM-B’的单向访问等。

[0039] 在 OpenFlow 控制器中配置了虚拟网络信息后，就可以根据配置的虚拟网络信息管理和控制虚拟网络中任意两个虚拟机之间的互访。下面以虚拟网络 Rose 中 VM-A 到 VM-A’的访问为例，对 OpenFlow 控制器管理虚拟网络的过程进行说明：

[0040] 参见图 4，图 4 是本发明实施例报文转发流程示意图，以图 3 中服务器 1 的 VM-A 访问服务器 2 的 VM-A’为例，包括以下步骤：

[0041] 步骤 401、VM-A 发送以太网报文，部署在服务器 1 中的 OpenFlow 交换机接收该以太网报文并查找该以太网报文对应的流表项。

[0042] 步骤 402，判断是否查找到该以太网报文对应的流表项，如果查找到，则转至步骤 408 执行，否则，继续执行步骤 403。

[0043] 步骤 403、部署在服务器 1 中的 OpenFlow 交换机向 OpenFlow 控制器发送携带以太网报文头信息的 packet-in 消息。

[0044] 这里，packet-in 消息中还可以携带以太网报文其它部分的信息，具体可以参照 OpenFlow 协议规定。

[0045] 步骤 404、OpenFlow 控制器接收到 packet-in 消息后，根据 packet-in 消息携带的以太网报文头信息确定以太网报文的源虚拟机 VM-A 以及目的虚拟机 VM-A’。

[0046] 根据以太网报文头中的源 MAC 地址确定以太网报文的源虚拟机，也即发送以太网报文的虚拟机，根据以太网报文中的目的 MAC 地址确定以太网报文的目的虚拟机，也即要接收该以太网报文的虚拟机。

[0047] 步骤 405、OpenFlow 控制器根据配置的虚拟网络信息判断 VM-A 和 VM-A’是否在同一虚拟网络，如果是，则继续执行步骤 406，否则，转至步骤 409 执行。

[0048] 步骤 406、控制服务器查找 VM-A 和 VM-A’所在虚拟网络的访问策略信息，确定 VM-A 和 VM-A’之间的访问策略，根据 VM-A 和 VM-A’之间的访问策略生成 VM-A 到 VM-A’的流表项，并将该流表项下发到部署在服务器 1 中的 OpenFlow 交换机。

[0049] 这里，在访问策略允许 VM-A 和 VM-A’互访时，根据 VM-A 和 VM-A’之间的访问策略生成 VM-A 到 VM-A’的流表项的包头域中，进入接口为 VM-A 连接的部署在服务器 1 上的 OpenFlow 交换机上的端口；源 MAC 地址为 VM-A 的 MAC 地址；目的 MAC 地址为 VM-A’的 MAC 地址；行动为通过服务器 1 的上行接口（uplink 接口）转发所匹配的报文，该上行接口为该服务器 1 上连接至外部物理交换机的接口，也是部署在服务器 1 上的 OpenFlow 交换机连接至外部物理交换机的接口，该接口预先被配置加入图 3 所示的分布式虚拟交换机系统。在访问策略禁止 VM-A 和 VM-A’互访时，可以生成用于丢弃报文的流表项，其中，该流表项的包头域与访问策略允许 VM-A 和 VM-A’互访时生成的 VM-A 到 VM-A’的流表项的包头域相同，行动包括：丢弃所匹配的报文。

[0050] 本步骤中，根据 VM-A 和 VM-A’之间的访问策略生成 VM-A 到 VM-A’的流表项时，还可以进一步根据 VM-A 和 VM-A’之间的访问策略生成 VM-A’到 VM-A 的流表项，并下发到部署在在服务器 1 中的 OpenFlow 交换机。其中，在访问策略允许 VM-A 和 VM-A’互访时，根据 VM-A 和 VM-A’之间的访问策略生成 VM-A’到 VM-A 的流表项的包头域中，进入接口为服务器 1 上的 OpenFlow 交换机上连接至外部物理交换机的 uplink 接口；源 MAC 地址为 VM-A’的 MAC 地址；目的 MAC 地址为 VM-A 的 MAC 地址；行动为通过服务器 1 上的 OpenFlow 交换机上

连接至 VM-A 的下行接口转发所匹配的报文。在访问策略禁止 VM-A 和 VM-A' 互访时，可以生成用于丢弃报文的流表项，其中，该流表项的包头域与访问策略允许 VM-A 和 VM-A' 互访时生成的 VM-A' 到 VM-A 的流表项的包头域相同，行动包括：丢弃所匹配的报文。

[0051] 本实施例中，OpenFlow 交换机可以是支持虚拟边缘端口汇聚（VEPA）转发模式的虚拟交换机。

[0052] 步骤 407、部署在服务器 1 中的 OpenFlow 交换机根据下发的流表项转发以太网报文，报文转发流程结束。

[0053] 此后，部署在服务器 1 中的 OpenFlow 交换机将会根据该下发的流表项转发后续的从 VM-A 发往 VM-A' 的以太网报文。

[0054] 步骤 408、根据查找到的流表项转发以太网报文，报文转发流程结束。

[0055] 步骤 409、丢弃以太网报文，报文转发流程结束。

[0056] 本步骤中，在丢弃以太网报文时，还可以生成用于丢弃该类以太网报文（这里将具有相同源 MAC 地址和目的 MAC 地址的以太网归为同一类以太网报文）的流表项并下发到部署在服务器 1 中的 OpenFlow 交换机，使得后续接收到该类以太网报文时根据该流表项执行报文丢弃操作。

[0057] 图 4 所示本发明实施例中，由于控制服务器在步骤 406 中生成 VM-A 到 VM-A' 的流表项时，已经确定了以太网报文在服务器 1 中的出接口（即上述的 uplink 口），因此，在下发 VM-A 到 VM-A' 的流表项时，还可以将以太网报文的出接口信息携带在 packet-out 消息发送到部署在服务器 1 中的 OpenFlow 交换机，从而使得部署在服务器 1 中的 OpenFlow 交换机直接根据 packet-out 消息中携带的出接口信息转发以太网报文，不再执行步骤 407 中根据下发的流表项转发以太网报文的操作。

[0058] 以上是 VM-A 访问 VM-A' 时，VM-A 所在服务器 1 对 VM-A 发往 VM-A' 的以太网报文的转发处理流程。当以太网报文到达 VM-A' 所在服务器 2 时，部署在服务器 2 中的 OpenFlow 交换机也需要对以太网报文执行转发处理，其报文转发处理流程与图 4 所示报文转发流程相同，不再赘述，不同之处在于，发送侧和接收侧的 OpenFlow 交换机查找到或生成的流表项中的具体内容不同。

[0059] 在实际应用中，可以允许虚拟机在其所在虚拟网络的多个服务器之间进行迁移，当发生虚拟机迁移时，虚拟机迁移前所在服务器中的 OpenFlow 交换机以及虚拟机迁移后所在服务器中的 OpenFlow 交换机都可以感知到该迁移事件，虚拟机迁移后所在服务器的 OpenFlow 交换机还会将该虚拟机迁移事件通知给 OpenFlow 控制器，OpenFlow 控制器接收到该虚拟机迁移事件通知后，需要更新发生迁移的虚拟机的虚拟机信息。例如，Rose 虚拟网络中的 VM-A 从服务器 1 迁移到服务器 2，服务器 2 中的 OpenFlow 交换机会向 OpenFlow 控制器发送虚拟机迁移事件通知，OpenFlow 控制器接收到该虚拟机迁移事件通知后，需要对 Rose 虚拟网络信息中 VM-A 对应的虚拟机信息进行修改。

[0060] 另外，由于虚拟机迁移后，相应的流表项也会发生变化，例如，VM-A 从服务器 1 迁移到服务器 2 后，由于迁移后 VM-A 所在的服务器及连接的 OpenFlow 交换机发生变化，从迁移后 VM-A 到 VM-B、VM-A'、以及 VM-B' 的流表项与迁移前 VM-A 到 VM-B、VM-A'、以及 VM-B' 的流表项均不相同，不能再按照原来的流表项进行以太网报文转发，从 VM-B、VM-A'、VM-B' 到迁移后 VM-A 的流表项与原来的从 VM-B、VM-A'、VM-B' 到迁移前 VM-A 的流表项也均不相

同，不能再按照原来的流表项进行以太网报文转发。

[0061] 为此，OpenFlow 控制器在接收到虚拟机迁移后所在服务器发送的虚拟机迁移事件通知后，还需要根据该虚拟机当前连接的 OpenFlow 交换机（也即该虚拟机迁移后所在服务器中的 OpenFlow 交换机）及连接端口信息，更新该虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项，并将生成的流表项下发到该虚拟机当前连接的 OpenFlow 交换机。

[0062] 基于以上的原理性说明，本发明提供了一种实现分布式多租户虚拟网络的方法，下面结合图 5 进行说明。

[0063] 图 5 是本发明实施例实现分布式多租户虚拟网络的方法流程图，其中，分布式多租户虚拟网络中的一台服务器上部署有 OpenFlow 控制器，其它服务器中部署有虚拟机和 OpenFlow 交换机，该方法主要包括以下步骤：

[0064] 步骤 501、在 OpenFlow 控制器中配置虚拟网络信息。

[0065] 这里，所述虚拟网络信息包括该虚拟网络中所有虚拟机信息和该虚拟网络中的虚拟机之间的访问策略信息，所述虚拟机信息包括虚拟机标识和虚拟机的 MAC 地址。

[0066] 步骤 502、OpenFlow 控制器接收到任一 OpenFlow 交换机发送的携带以太网报文头信息的 packet-in 消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源 MAC 地址对应的源虚拟机和目的 MAC 地址对应的目的虚拟机是否属于同一虚拟网络，如果是，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文。

[0067] 这里，所述 packet-in 消息是该任一 OpenFlow 交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后向 OpenFlow 控制器发送的。

[0068] 所述虚拟机信息还可以包括虚拟机连接的 OpenFlow 交换机及连接端口信息，所述 OpenFlow 交换机为支持虚拟边缘端口汇聚 VEPA 转发模式的虚拟交换机，所述 OpenFlow 交换机通过一上行接口连接至外部的物理交换机。

[0069] 所述生成该源虚拟机到目的虚拟机的流表项，包括：

[0070] 在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第一流表项，该第一流表项包头域包括：源 MAC 地址为源虚拟机的 MAC 地址；目的 MAC 地址为目的虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至该源虚拟机的下行接口；该第一流表项的行动包括：通过该任一 OpenFlow 交换机上连接至外部物理交换机的上行接口转发所匹配的报文；

[0071] 在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第二流表项，该第二流表项包头域与所述第一流表项相同，该第二流表项的行动包括：丢弃所匹配的报文。

[0072] 另外，根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该任一 OpenFlow 交换机时，还可以同时根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该目的虚拟机到源虚拟机的流表项并下发到该任一 OpenFlow 交换机。

[0073] 所述生成该目的虚拟机到源虚拟机的流表项，包括：

[0074] 在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机

互访时，生成第三流表项，该第三流表项包头域包括：源MAC地址为目的虚拟机的MAC地址；目的MAC地址为源虚拟机的MAC地址；进入接口为该任一OpenFlow交换机上连接至外部物理交换机的上行接口；该第三流表项的行动包括：通过该任一OpenFlow交换机上连接至该源虚拟机的下行接口转发所匹配的报文；

[0075] 在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第四流表项，该第四流表项包头域与所述第三流表项相同，该第四流表项的行动包括：丢弃所匹配的报文。

[0076] 图5所示本发明实施例中，当发生虚拟机迁移时，OpenFlow控制器还会接收到OpenFlow交换机发送的虚拟机迁移事件通知，这时，需将更新虚拟网络信息中该发生迁移的虚拟机的虚拟机信息，并根据该虚拟机当前连接的OpenFlow交换机及连接端口信息，更新该虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项，并将生成的流表项下发到该虚拟机当前连接的OpenFlow交换机；其中，所述虚拟机迁移事件通知是该虚拟机当前连接的OpenFlow交换机检测到有虚拟机迁移到该虚拟机当前连接的OpenFlow交换机所在服务器时发送的。

[0077] 图5所示本发明实施例中，OpenFlow控制器将生成的所述以太网报文的源虚拟机到目的虚拟机的流表项下发到OpenFlow交换机时，还可以向该OpenFlow交换机发送携带所述以太网报文的出接口信息的packet-out消息，以使该OpenFlow交换机在接收到该Packet-out消息的情况下根据该packet-out消息携带的出接口信息转发所述以太网报文，未接收到该Packet-out消息的情况下根据该下发的流表项处理所述以太网报文。

[0078] 本发明还提供了一种在数据中心网络中实现分布式多租户虚拟网络的装置，下面结合图6进行说明。

[0079] 参见图6，图6是本发明实施例实现分布式多租户虚拟网络的装置的结构示意图，所述分布式多租户虚拟网络中的一台服务器上部署有OpenFlow控制器，其它服务器中部署有虚拟机和OpenFlow交换机，该装置应用于所述OpenFlow控制器，包括：配置单元601、接收单元602、判断单元603、控制单元604、发送单元605；其中

[0080] 配置单元601，用于在所述OpenFlow控制器中配置虚拟网络的信息，所述虚拟网络的信息包括该虚拟网络中所有虚拟机信息以及虚拟机之间的访问策略信息，所述虚拟机信息包括该虚拟机的MAC地址；

[0081] 接收单元602，用于接收任一OpenFlow交换机发送的携带以太网报文的报文头信息的packet-in消息；所述packet-in消息是该OpenFlow交换机接收到所述以太网报文且确定不存在所述以太网报文对应的流表项之后发送的；

[0082] 判断单元603，用于接收单元602接收到该OpenFlow交换机发送的携带以太网报文头信息的packet-in消息时，根据配置单元601配置的虚拟网络信息判断以太网报文头信息中的源MAC地址对应的源虚拟机和目的MAC地址对应的目的虚拟机是否属于同一虚拟网络；

[0083] 控制单元604，用于判断单元603判定所述以太网报文头信息中的源MAC地址对应的源虚拟机和目的MAC地址对应的目的虚拟机属于同一虚拟网络时，则根据配置单元601配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项；

[0084] 发送单元 605，用于将控制单元 604 生成的该源虚拟机到目的虚拟机的流表项下发到该任一 OpenFlow 交换机，以使该任一 OpenFlow 交换机根据该下发的流表项处理所述以太网报文。

[0085] 上述装置中，

[0086] 所述虚拟机信息还包括虚拟机连接的 OpenFlow 交换机及连接端口信息，所述 OpenFlow 交换机为支持虚拟边缘端口汇聚 VEPA 转发模式的虚拟交换机，所述 OpenFlow 交换机通过一上行接口连接至外部的物理交换机；

[0087] 所述控制单元 604 生成该源虚拟机到目的虚拟机的流表项，包括：

[0088] 在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第一流表项，该第一流表项包头域包括：源 MAC 地址为源虚拟机的 MAC 地址；目的 MAC 地址为目的虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至该源虚拟机的下行接口；该第一流表项的行动包括：通过该任一 OpenFlow 交换机上连接至外部物理交换机的上行接口转发所匹配的报文；

[0089] 在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第二流表项，该第二流表项包头域与所述第一流表项相同，该第二流表项的行动包括：丢弃所匹配的报文。

[0090] 上述装置中，

[0091] 所述控制单元 604 在根据配置单元 601 配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项时，进一步根据配置单元 601 配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该目的虚拟机到源虚拟机的流表项；

[0092] 所述发送单元 605，用于将控制单元 604 生成的该目的虚拟机到源虚拟机的流表项下发到该任一 OpenFlow 交换机。

[0093] 上述装置中，

[0094] 所述控制单元 604 生成该目的虚拟机到源虚拟机的流表项，包括：

[0095] 在该源虚拟机和目的虚拟机之间的访问策略信息允许该源虚拟机和目的虚拟机互访时，生成第三流表项，该第三流表项包头域包括：源 MAC 地址为目的虚拟机的 MAC 地址；目的 MAC 地址为源虚拟机的 MAC 地址；进入接口为该任一 OpenFlow 交换机上连接至外部物理交换机的上行接口；该第三流表项的行动包括：通过该任一 OpenFlow 交换机上连接至该源虚拟机的下行接口转发所匹配的报文；

[0096] 在该源虚拟机和目的虚拟机之间的访问策略信息禁止该源虚拟机和目的虚拟机互访时，生成第四流表项，该第四流表项包头域与所述第三流表项相同，该第四流表项的行动包括：丢弃所匹配的报文。

[0097] 上述装置中，

[0098] 所述接收单元 602，进一步用于接收该任一 OpenFlow 交换机发送的虚拟机迁移事件通知；

[0099] 所述控制单元 604，进一步用于接收单元 602 接收到该任一 OpenFlow 交换机发送的虚拟机迁移事件通知时，更新发生迁移的虚拟机的虚拟机信息，并根据该虚拟机当前连接的 OpenFlow 交换机及连接端口信息，更新该虚拟机与其所在虚拟网络中其它各虚拟机

之间的流表项；其中，所述虚拟机迁移事件通知是该任一 OpenFlow 交换机检测到有虚拟机迁移到该任一 OpenFlow 交换机所在服务器时发送的；

[0100] 所述发送单元，用于将控制单元生成该发生迁移的虚拟机与其所在虚拟网络中其它各虚拟机之间的流表项下发到该任一 OpenFlow 交换机。。

[0101] 上述装置还包括发送单元 605；

[0102] 所述发送单元605，用于将控制单元604根据配置单元601配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项下发到该任一 OpenFlow 交换机时，进一步向该任一 OpenFlow 交换机发送携带所述以太网报文的出接口信息的 packet-out 消息，以使该任一 OpenFlow 交换机在接收到该 Packet-out 消息的情况下根据该 packet-out 消息携带的出接口信息转发所述以太网报文，未接收到该 Packet-out 消息的情况下根据该下发的流表项处理所述以太网报文。

[0103] 以上所述，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，均应包含在本发明的保护范围之内。

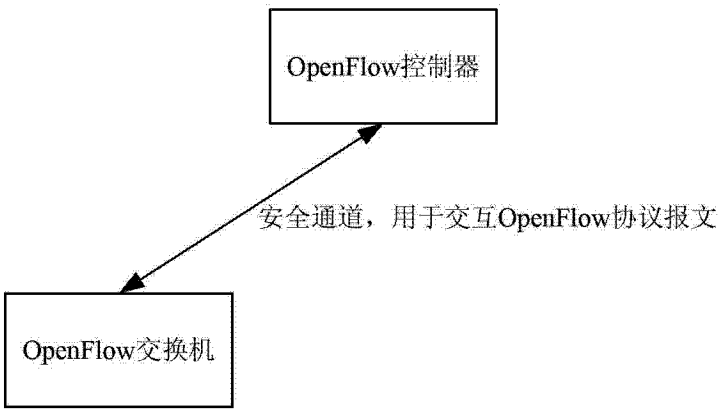

OpenFlow控制器
安全通道，用于交互OpenFlow协议报文
OpenFlow交换机

图 1

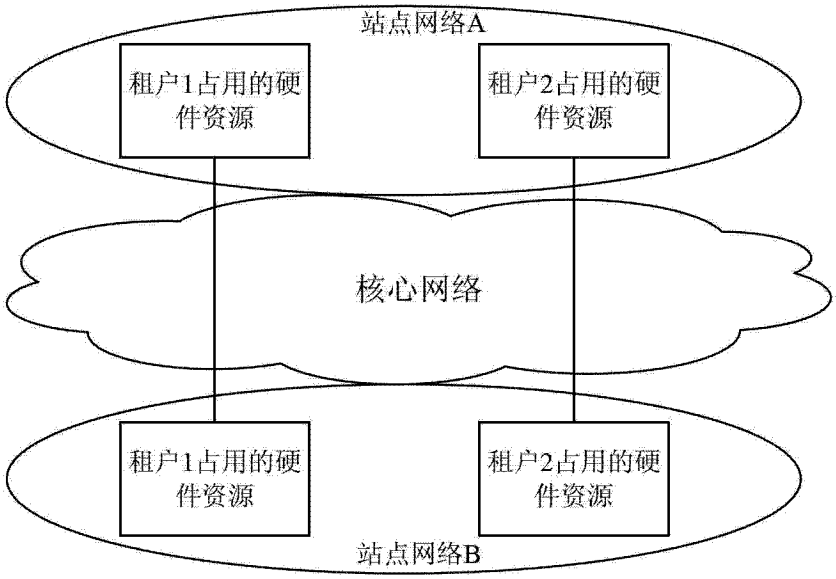

站点网络A
租户1占用的硬件资源
租户2占用的硬件资源
核心网络
租户1占用的硬件资源
租户2占用的硬件资源
站点网络B

图 2

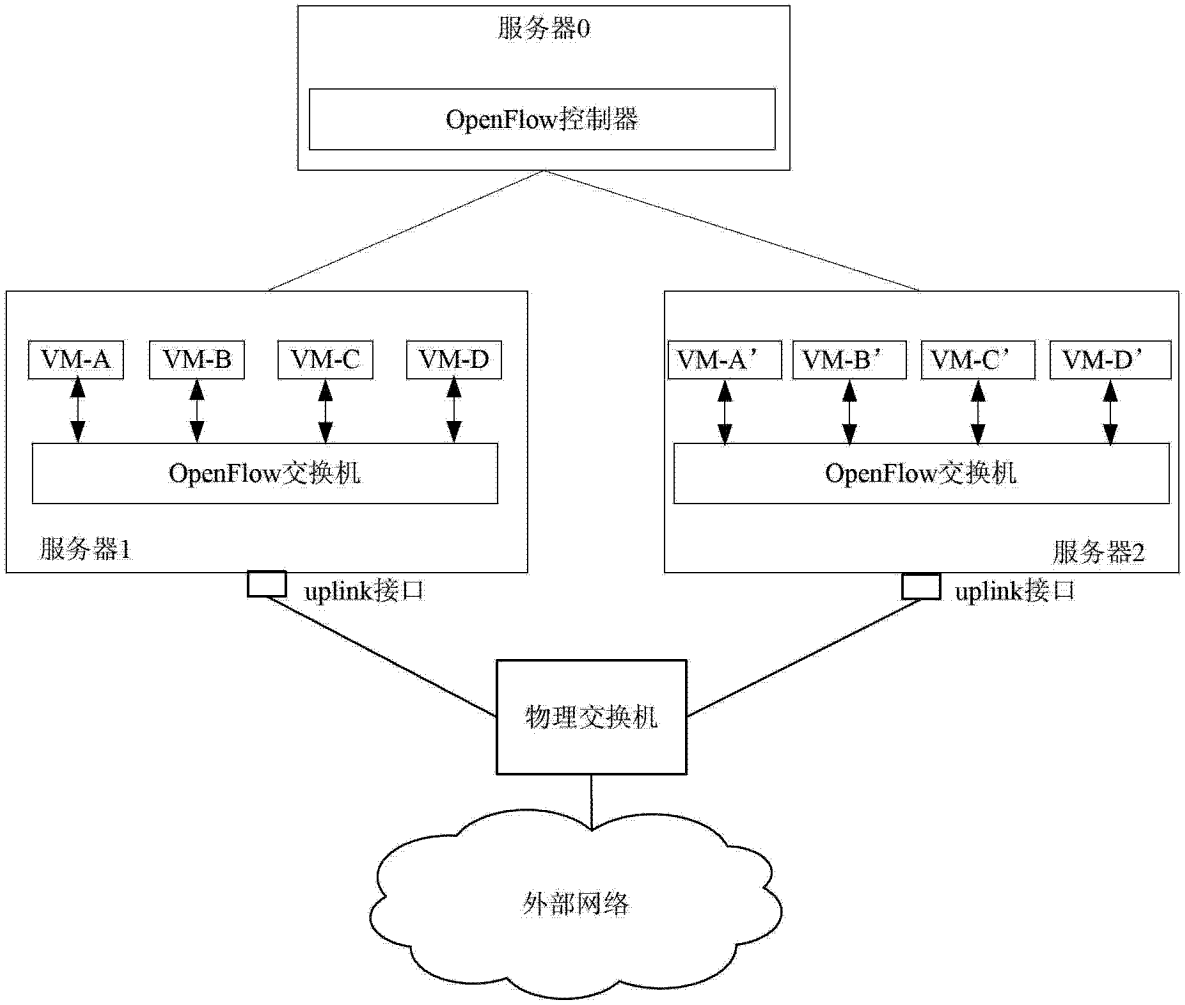

服务器0
OpenFlow控制器
VM-A
VM-B
VM-C
VM-D
OpenFlow交换机
服务器1
VM-A'
VM-B'
VM-C'
VM-D'
OpenFlow交换机
服务器2
uplink接口
uplink接口
物理交换机
外部网络

图 3

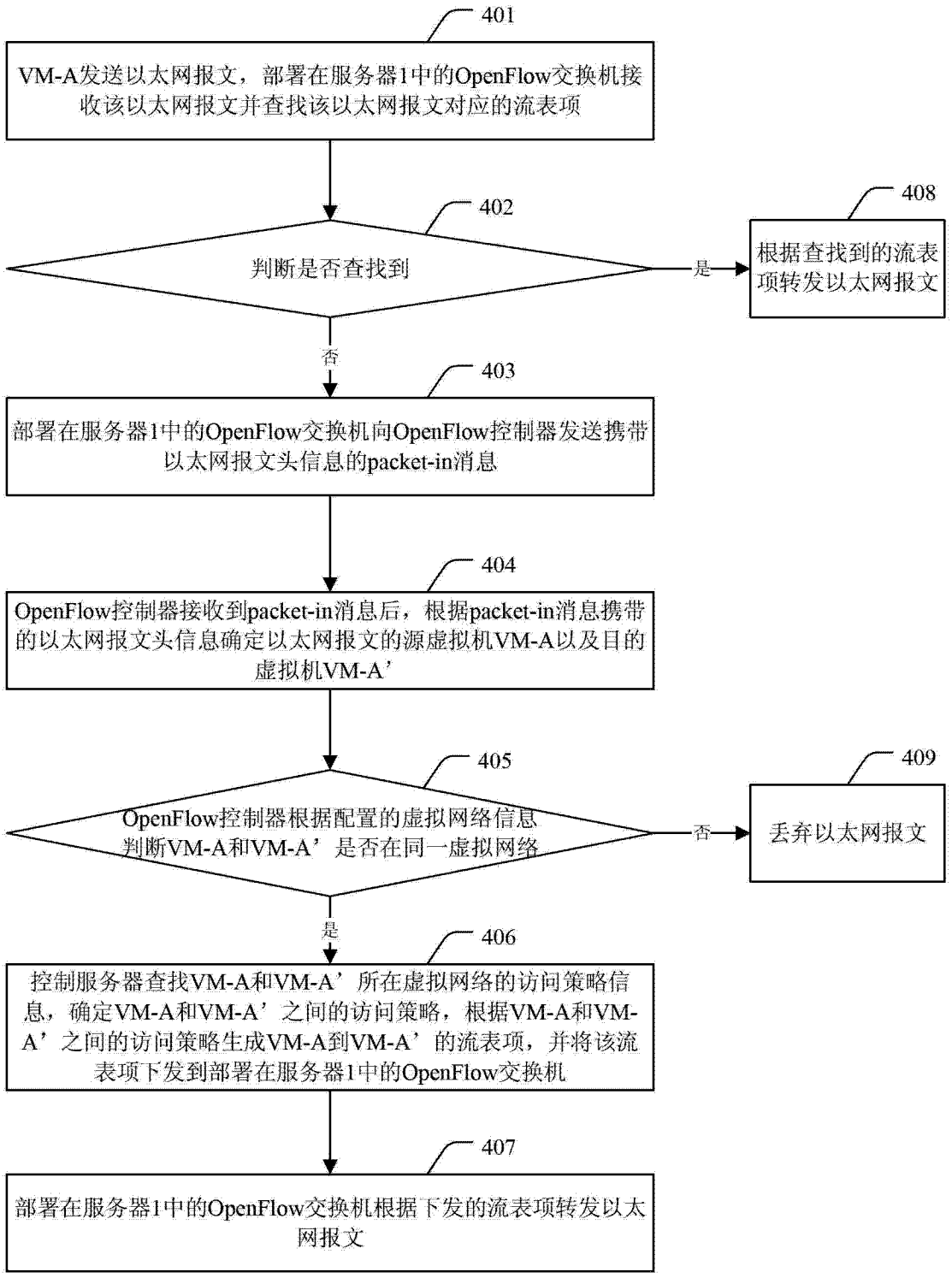

401
VM-A发送以太网报文，部署在服务器1中的OpenFlow交换机接收该以太网报文并查找该以太网报文对应的流表项
402
判断是否查找到
是
408
根据查找到的流表项转发以太网报文
否
403
部署在服务器1中的OpenFlow交换机向OpenFlow控制器发送携带以太网报文头信息的packet-in消息
404
OpenFlow控制器接收到packet-in消息后，根据packet-in消息携带的以太网报文头信息确定以太网报文的源虚拟机VM-A以及目的虚拟机VM-A'
405
OpenFlow控制器根据配置的虚拟网络信息判断VM-A和VM-A'是否在同一虚拟网络
否
409
丢弃以太网报文
是
406
控制服务器查找VM-A和VM-A'所在虚拟网络的访问策略信息，确定VM-A和VM-A'之间的访问策略，根据VM-A和VM-A'之间的访问策略生成VM-A到VM-A'的流表项，并将该流表项下发到部署在服务器1中的OpenFlow交换机
407
部署在服务器1中的OpenFlow交换机根据下发的流表项转发以太网报文

图 4

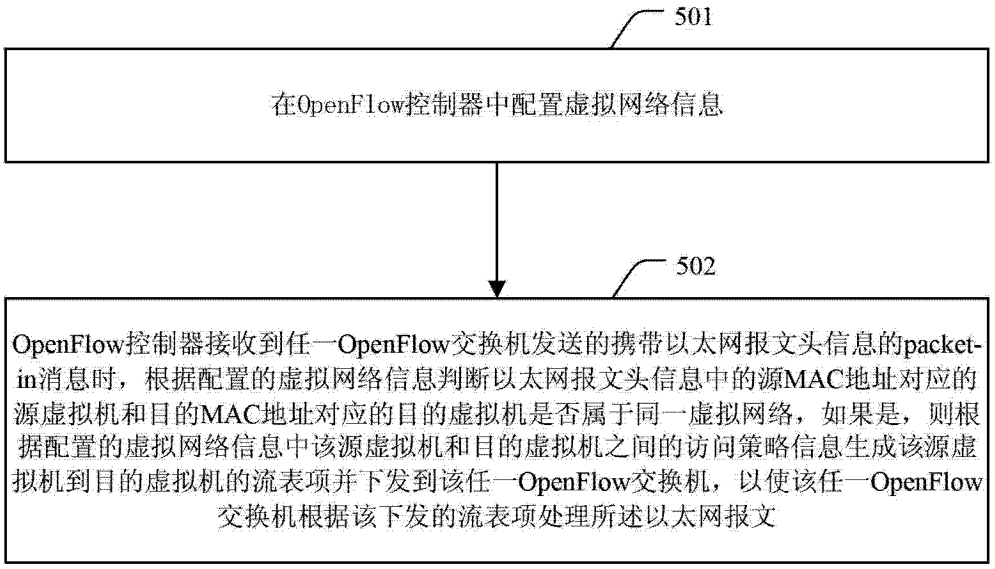

501
在OpenFlow控制器中配置虚拟网络信息
502
OpenFlow控制器接收到任一OpenFlow交换机发送的携带以太网报文头信息的packet-in消息时，根据配置的虚拟网络信息判断以太网报文头信息中的源MAC地址对应的源虚拟机和目的MAC地址对应的目的虚拟机是否属于同一虚拟网络，如果是，则根据配置的虚拟网络信息中该源虚拟机和目的虚拟机之间的访问策略信息生成该源虚拟机到目的虚拟机的流表项并下发到该任一OpenFlow交换机，以使该任一OpenFlow交换机根据该下发的流表项处理所述以太网报文

图 5

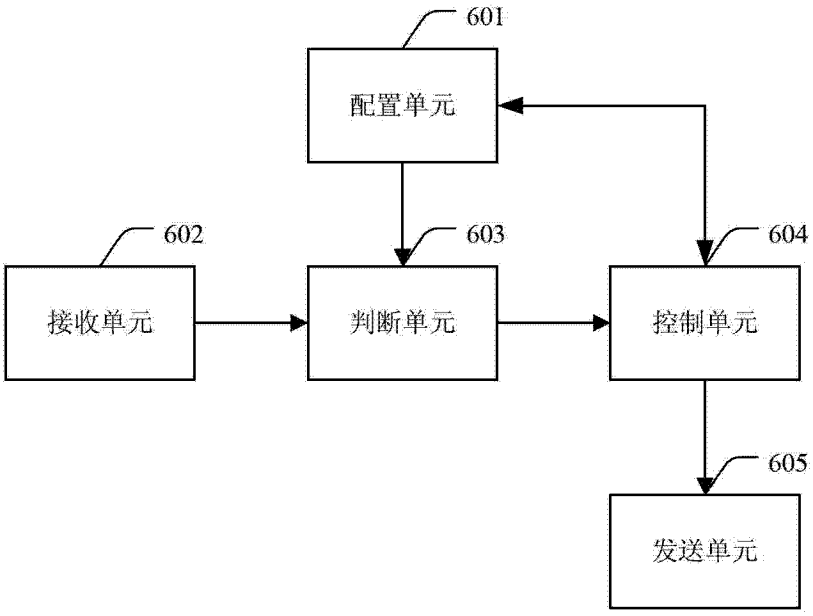

601
配置单元
602
603
604
接收单元
判断单元
控制单元
605
发送单元

图 6