



(12) 发明专利

(10) 授权公告号 CN 102263704 B

(45) 授权公告日 2014. 03. 26

(21) 申请号 201110257140. 3

审查员 连立杰

(22) 申请日 2011. 09. 01

(73) 专利权人 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 周万 朱国平

(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 12/70 (2013. 01)

(56) 对比文件

CN 101557336 A, 2009. 10. 14,

US 2008/0225853 A1, 2008. 09. 18,

CN 101730170 A, 2010. 06. 09,

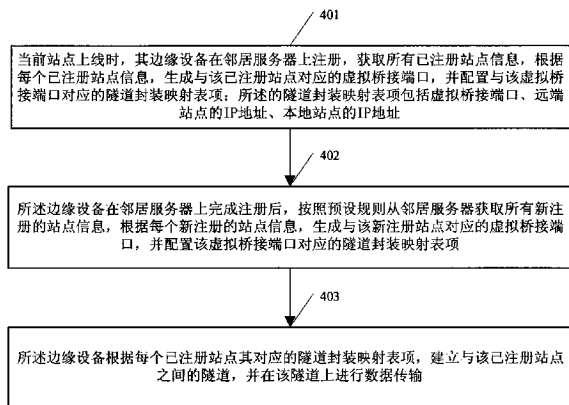
权利要求书3页 说明书9页 附图3页

(54) 发明名称

一种支持数据中心二层互联的拓扑构建方法和装置

(57) 摘要

本发明提供了一种支持数据中心二层互联的拓扑构建方法和装置,该方法包括:当前站点上线时,向邻居服务器注册,获取所有已注册站点信息,生成与每个已注册站点对应的虚拟桥接端口;当前站点在邻居服务器注册后,按照预设规则获取所有新注册的站点信息,并生成与每个新注册的站点对应的虚拟桥接端口;通过与每个已注册站点对应的虚拟桥接端口和该已注册站点建立隧道,进行数据传输。本发明能够自动发现站点。



1. 一种支持数据中心二层互联的拓扑构建方法,其特征在于,该方法包括:

当前站点上线时,其边缘设备在邻居服务器上注册,获取所有已注册站点信息,根据每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;所述的隧道封装映射表项包括虚拟桥接端口、远端站点的 IP 地址、本地站点的 IP 地址;

当前站点上线后,其边缘设备按照预设规则从邻居服务器获取所有新注册的站点信息,根据每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口以及配置与该虚拟桥接端口对应的隧道封装映射表项;

所述边缘设备根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

2. 如权利要求 1 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,所述边缘设备在邻居服务器上注册,获取所有已注册站点信息的方法为:边缘设备向邻居服务器发送携带自身所在站点信息的注册请求报文,接收所述邻居服务器在接收到该注册请求报文后返回的注册应答报文,并获取接收到的注册应答报文中携带的所有已注册站点信息。

3. 如权利要求 2 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,预先配置注册重试时间间隔列表;

如果在当前注册重试时间间隔超时的时候仍未接收到邻居服务器返回的注册应答报文,则再次向邻居服务器发送注册请求报文,并以注册重试时间间隔列表中的下一个时间间隔值作为注册重试时间间隔,直到接收到邻居服务器返回的注册应答报文;

如果当前注册重试时间间隔是列表中的最后一个时间间隔值作为注册重试时间间隔,且等待超时后仍未接收到邻居服务器回应的注册应答报文,则重新向邻居服务器发送注册请求报文并以注册重试时间间隔列表中的第一个时间间隔值作为注册重试时间间隔进行新一轮的注册重试。

4. 如权利要求 3 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,所述按照预设规则从邻居服务器获取所有新注册的站点信息的方法为:边缘设备每隔第一预设时间向邻居服务器发送注册请求报文,并在接收到邻居服务器回应的注册应答报文之后,获取该注册应答报文中携带的所有已注册站点信息,根据每个已注册站点信息查找对应的隧道封装映射表项,如果未查找到,则确定该已注册站点为新注册的站点,并获取该新注册的站点信息。

5. 如权利要求 4 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,所述站点信息包括该站点的桥 MAC 地址;

所述根据每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口之前,进一步包括:判断该已注册站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该已注册站点对应的虚拟桥接端口,否则,生成与该已注册站点对应的虚拟桥接端口;

所述根据每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口之前,进一步包括:判断该新注册的站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该新注册的站点对应的虚拟桥接端口,否则,生成与该新注册的站点对应的虚拟桥接端口。

6. 如权利要求 5 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,所述生成与该已注册站点对应的虚拟桥接端口之后,进一步包括:向该已注册站点发送解析请求报文,并在接收到该已注册站点回应的解析应答报文后,将该虚拟桥接端口的状态设置为启动 UP。

7. 如权利要求 6 所述的支持数据中心二层互联的拓扑构建方法,其特征在于,所述生成与该已注册站点对应的虚拟桥接端口时,进一步将该虚拟桥接端口的状态设置为故障 DOWN;

所述生成与该已注册站点对应的虚拟桥接端口之后,每隔第二预设时间向该已注册站点发送解析请求报文,如果连续发送预设次数的解析请求报文均未收到相应的解析应答报文,则将该已注册站点对应的虚拟桥接端口的状态设置为 DOWN,设置该虚拟桥接端口对应的老化定时器,并在该老化定时器超时后删除该虚拟桥接端口。

8. 如权利要求 1-7 中任一权项所述的支持数据中心二层互联的拓扑构建方法,其特征在于,该方法进一步包括:

当前站点下线时,其边缘设备向邻居服务器发送携带当前站点的 IP 地址的注销请求报文,并由邻居服务器将该注销请求报文发送到所有其它已注册站点;

边缘设备接收到邻居服务器发来的注销请求报文时,根据注销请求报文中携带的下线站点的 IP 地址,查找对应的隧道封装映射表项,并将查找到的隧道封装映射表项对应的虚拟桥接端口删除。

9. 一种边缘设备,应用于数据中心站点的二层互联和拓扑构建,其特征在于,该边缘设备包括:注册单元、控制单元、传输单元;

所述注册单元,用于该边缘设备所在的站点上线时,在邻居服务器上注册,获取所有已注册站点信息;用于该边缘设备所在的站点上线后,按照预设规则从邻居服务器获取所有新注册的站点信息;

所述控制单元,根据注册单元获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;用于根据注册单元获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口以及配置与该虚拟桥接端口对应的隧道封装映射表项;

所述传输单元,用于根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

10. 如权利要求 9 所述的边缘设备,其特征在于,所述注册单元在邻居服务器上注册,获取所有已注册站点信息时,用于:向邻居服务器发送携带该边缘设备所在站点信息的注册请求报文,接收所述邻居服务器在接收到该注册请求报文后返回的注册应答报文,并获取接收到的注册应答报文中携带的所有已注册站点信息。

11. 如权利要求 10 所述的边缘设备,其特征在于,所述注册单元,用于存储预先配置的注册重试时间间隔列表;

所述注册单元向邻居服务器发送注册请求报文之后,如果在当前注册重试时间间隔超时的时侯仍未接收到邻居服务器返回的注册应答报文,则再次向邻居服务器发送注册请求报文,并以注册重试时间间隔列表中的下一个时间间隔值作为注册重试时间间隔,直到接收到邻居服务器返回的注册应答报文;如果当前注册重试时间间隔是列表中的最后一个时

间间隔值,且等待超时后仍未接收到邻居服务器回应的注册应答报文,则重新向邻居服务器发送注册请求报文并以注册重试时间间隔列表中的第一个时间间隔值作为注册重试时间间隔进行新一轮的注册重试。

12. 如权利要求 11 所述的边缘设备,其特征在于,所述注册单元按照预设规则从邻居服务器获取所有新注册的站点信息时,用于:每隔第一预设时间向邻居服务器发送注册请求报文,并在接收到邻居服务器回应的注册应答报文之后,获取该注册应答报文中携带的所有已注册站点信息,根据每个已注册站点信息在存储单元中查找对应的隧道封装映射表项,如果未查找到,则确定该已注册站点为新注册的站点,并获取该新注册的站点信息。

13. 如权利要求 12 所述的边缘设备,其特征在于,所述注册单元获取的站点信息包括该站点的桥 MAC 地址;

所述控制单元在根据注册单元获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口之前,进一步用于:判断该已注册站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该已注册站点对应的虚拟桥接端口,否则,生成与该已注册站点对应的虚拟桥接端口;

所述控制单元在根据注册单元获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口之前,进一步用于:判断该新注册的站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该新注册的站点对应的虚拟桥接端口,否则,生成与该新注册的站点对应的虚拟桥接端口。

14. 如权利要求 13 所述的边缘设备,其特征在于,所述控制单元生成与该已注册站点对应的虚拟桥接端口之后,进一步用于:向该已注册站点发送解析请求报文,并在接收到该已注册站点回应的解析应答报文后,将该虚拟桥接端口的状态设置为启动 UP。

15. 如权利要求 14 所述的边缘设备,其特征在于,所述控制单元生成与该已注册站点对应的虚拟桥接端口时,进一步将该虚拟桥接端口的状态设置为故障 DOWN;

所述控制单元在生成与该已注册站点对应的虚拟桥接端口之后,每隔第二预设时间向该已注册站点发送解析请求报文,如果连续发送预设次数的解析请求报文均未收到相应的解析应答报文后,将该已注册站点对应的虚拟桥接端口的状态设置为 DOWN,设置该虚拟桥接端口对应的老化定时器,并在该老化定时器超时后删除该虚拟桥接端口。

16. 如权利要求 9-15 中任一权项所述的边缘设备,其特征在于,所述注册单元,进一步用于:当前站点下线时,向邻居服务器发送携带当前站点的 IP 地址的注销请求报文,并由邻居服务器将该注销请求报文发送到所有其它已注册站点;用于接收到邻居服务器发来的注销请求报文时,根据注销请求报文中携带的下线站点的 IP 地址,在存储单元中查找对应的隧道封装映射表项,并将查找到的隧道封装映射表项对应的虚拟桥接端口删除。

一种支持数据中心二层互联的拓扑构建方法和装置

技术领域

[0001] 本发明涉及网络通信技术领域,特别涉及一种支持数据中心二层互联的拓扑构建方法和装置。

背景技术

[0002] 数据中心,是各种业务的提供中心,是数据处理、数据存储和数据交换的中心。为了更好的服务,提供商通常在异地部署多个数据中心,实现负载分担和高可靠性,并实现虚拟机在数据中心之间的自由迁移。由于虚拟机迁移过程对用户透明,不能改变 IP 地址,所以必须在分布于异地的数据中心之间实现二层网络互联。

[0003] 数据中心互联,需要满足以下基本要求:各站点互相独立,站点内的拓扑和故障互不影响;对站点间传输数据时使用的技术与站点位置、提供商的网络无特殊要求,具有传输无关性;使用多归宿提供冗余接入,并在站点间避免流量环路,保证高可靠性;站点内拓扑结构实现灵活;运营维护简单,可快速新增和减少站点。

[0004] 目前,在公网只支持 IP 技术时,只有思科的覆盖传输虚拟化 (OTV) 技术能够实现数据中心互联。在各数据中心的边缘设备的连接到核心网的物理端口上配置 IP 地址,并关联一个或多个 OTV 虚拟接口,每个 OTV 虚拟接口可以作为二层转发表的出接口。下面结合附图对 OTV 技术进行说明。

[0005] 参见图 1,图 1 是现有技术应用 OTV 技术实现数据中心二层互联的组网示意图。其中网络 10 为核心网,其上运行 pUMRP 协议,网络 12、16 分别是数据中心站点,其上运行 cUMRP 协议,网络 14 是构建在网络 10 上的虚拟网络,其上运行 oUMRP 协议。在网络 12 中,包含边缘设备 X11,交换机 S11,路由器 R11、R12,主机 H3 通过交换机 S11 接入,主机 H12 通过路由器 R11 接入。在网络 16 中,包含边缘设备 X12,交换机 S12、S13,路由器 R14、R15,主机 H1 通过交换机 S13 接入。

[0006] 参见图 2,图 2 是图 1 中边缘设备 X11 的数据平面内部实现图。其中,E1 和 E2 分别是连接主机 H2 和连接主机 H3 的内部接口,E3 是连接核心网的外部接口,OTV 虚拟接口 01:IPA 关联到 E3。在 X11 中存储 MAC 地址映射表(如图 2 中所示的 MAC 地址映射表),其中的 H2/E1、H3/E2 表项是通过 MAC 地址学习得到的,H1/01:IPA 表项则是从远端 OTV 设备通告的 IGP 协议报文中得到的(如图 2 中的线条 21 所示,表示将远端站点通告的 IGP 协议报文中携带的 MAC 地址写入 MAC 地址映射表),在图 1 中,远端 OTV 设备即为边缘设备 X12。例如,当 X11 从 E1 接收 H2 发来的报文时,进行 MAC 地址学习即可得到 H2/E1 表项并写入 MAC 地址映射表(如图 2 中的线条 22 所示,表示对在内部以太网端口接收到的报文进行 MAC 地址学习),同时还将学习到的 MAC 地址映射表项通过 IGP 协议报文通告给远端 OTV 设备;同样的道理,X11 也会从远端 OTV 设备通告的 IGP 协议报文中学习得到 H1/01:IPA 表项。根据 MAC 地址映射表,X11 可以将来自于 H2 或 H3 且目的地址为 H1 的报文进行 OTV 封装,并从 01:IPA 关联的物理接口 E3 发送出去;将来自于 H1 且目的地址为 H2 或 H3 的报文解封装后从相应的 E2 或 E3 接口发送出去。

[0007] 在图 1 所示的组网中,如果需要支持广播,则需要预先在 X11 和 X12 上配置由核心网分配的任意源组播 (ASM/Bidir) 类型的 IP 组播组。当 X11 在内部接口接收到广播报文时,将广播报文进行 OTV 封装后发送到核心网的组播树中,由核心网组播转发到远端 OTV 设备 X12, X12 将接收到的广播报文解封装后,在其所在的数据中心网络内广播。如果需要支持组播,则需要在网络 12 和 16 上运行 IGMP/MLD 侦听,在 X11 和 X12 上记录组播信息到组播转发表,在核心网中构建组播组的特定源组播 (SSM) 树,并根据该组播树组播转发数据中心之间的组播报文。

[0008] 可见,如果在应用 OTV 技术实现数据中心二层互联的组网中,核心网必须支持组播,否则,将无法实现广播转发和组播转发需求,进而也无法实现数据中心二层互联。如果不借助于核心网的组播技术实现数据中心二层互联,则需要各数据中心站点之间能够直接进行通信。这就需要各数据中心能够自动发现其它数据中心站点,并与自动发现的数据中心站点建立二层隧道,进行数据传输。然而,目前还不存在自动发现数据中心站点的方法。

发明内容

[0009] 有鉴于此,本发明的目的在于提供一种支持数据中心二层互联的拓扑构建方法,该方法能够自动发现数据中心站点。

[0010] 为了达到上述目的,本发明提供了一种实现数据中心二层互联的方法,该方法包括:

[0011] 当前站点上线时,其边缘设备在邻居服务器上注册,获取所有已注册站点信息,根据每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;所述的隧道封装映射表项包括虚拟桥接端口、远端站点的 IP 地址、本地站点的 IP 地址;

[0012] 当前站点上线后,其边缘设备按照预设规则从邻居服务器获取所有新注册的站点信息,根据每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口以及配置与该虚拟桥接端口对应的隧道封装映射表项;

[0013] 所述边缘设备根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

[0014] 本发明还提供了一种边缘设备,应用与数据中心的二层互联和拓扑构建,该边缘设备包括:注册单元、控制单元、传输单元;

[0015] 所述注册单元,用于该边缘设备所在的站点上线时,在邻居服务器上注册,获取所有已注册站点信息;用于该边缘设备所在的站点上线后,按照预设规则从邻居服务器获取所有新注册的站点信息;

[0016] 所述控制单元,根据注册单元获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;用于根据注册单元获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口以及配置与该虚拟桥接端口对应的隧道封装映射表项;

[0017] 所述传输单元,用于根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

[0018] 由上面的技术方案可知,本发明中,当前站点上线时,向邻居服务器注册,获取所

有已注册站点信息,针对每个已注册站点生成虚拟桥接端口;当前站点在邻居服务器注册后,周期性获取所有新注册站点信息,并针对每个新注册站点生成虚拟桥接端口;通过针对每个已注册站点生成的虚拟桥接端口和该已注册站点建立隧道,进行数据传输。应用本发明,可以通过虚拟桥接端口在数据中心站点之间建立隧道,实现数据中心二层互连;还可以根据各数据中心站点在邻居服务器中的注册情况,周期性从邻居服务器中获取所有已注册站点信息,从而使各数据中心站点能够动态发现其它数据中心站点。

附图说明

[0019] 图 1 是现有技术应用 OTV 技术实现数据中心二层互联的组网示意图;

[0020] 图 2 是图 1 中边缘设备 X11 的数据平面内部实现图;

[0021] 图 3 是本发明实施例支持数据中心站点二层互联的网络拓扑图;

[0022] 图 4 是本发明实施例支持数据中心二层互联的拓扑构建方法流程图;

[0023] 图 5 是本发明实施例边缘设备的结构示意图。

具体实施方式

[0024] 为了使本发明的目的、技术方案及优点更加清楚明白,下面结合附图并举实施例,对本发明的技术方案进行详细说明。

[0025] 参见图 3,图 3 是本发明实施例支持数据中心站点二层互联的网络拓扑图,只画出了各站点的边缘设备,站点内部拓扑不是本发明的重点,因此未在图中画出。图 3 中,站点 A、B、C、D、E 分别通过各自的边缘设备 1、2、3、4、5 接入核心网;另外,还在核心网中部署了邻居服务器,用于记录当前在线的各数据中心站点信息。这里,邻居服务器也可以部署在一台边缘设备上。需要说明的是,本申请文件中所述的站点,也即是数据中心站点。

[0026] 参见图 4,图 4 是本发明实施例支持数据中心二层互联的拓扑构建方法流程图,包括以下步骤:

[0027] 步骤 401、当前站点上线时,其边缘设备在邻居服务器上注册,获取所有已注册站点信息,根据获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;所述的隧道封装映射表项包括虚拟桥接端口、远端站点的 IP 地址、本地站点的 IP 地址。

[0028] 这里,各站点上线时,需要先向邻居服务器注册,并在完成注册后上线。邻居服务器记录所有已注册站点信息,并在每个站点向其注册时,将当前所有已注册站点信息通知该站点,该站点获得所有已注册站点信息后,可以针对每个已注册站点生成虚拟桥接端口,并配置相应的隧道封装映射表项。

[0029] 所述生成的虚拟桥接端口是虚拟以太网端口,没有物理实体,但具有与物理以太网端口相同的功能,关于虚拟桥接端口,将在后面进行详细说明。

[0030] 步骤 402、所述边缘设备在邻居服务器上完成注册后,按照预设规则从邻居服务器获取所有新注册的站点信息,根据获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口,并配置该虚拟桥接端口对应的隧道封装映射表项。

[0031] 本步骤中,当前站点在邻居服务器上注册完毕后,如果又有新的站点在邻居服务器完成注册后上线,则当前站点需要获取该新注册的站点信息,并根据获取的新注册的站

点信息建立与该新注册的站点之间的隧道,实现与该新注册的站点的二层互联。因此,站点上线后,其边缘设备仍需要定期获取新注册的站点信息,并生成与该新注册站点对应的虚拟桥接端口。

[0032] 需要说明的是,这里的新注册站点只是相对于当前站点来说的,也就是说,只要当前站点尚未生成与该已注册站点对应的虚拟桥接端口,则该已注册站点对当前站点而言,就是新注册的站点,需要获取该新注册的站点信息,并根据获取的新注册的站点信息建立与该新注册站点之间的隧道,实现与该新的已注册站点的二层互联。

[0033] 步骤 403、所述边缘设备根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

[0034] 这里,当前站点的边缘设备在针对每个已注册站点生成虚拟桥接端口后,可以通过该虚拟桥接端口建立与该已注册站点之间的隧道,当需要和该已注册站点通信时,在该隧道上进行数据通信,实现与该已注册站点的二层互联。

[0035] 图 4 所示的本发明实施例中,预先定义了注册请求报文和注册应答报文。其中,注册请求报文是各数据中心的边缘设备向邻居服务器发送的报文,携带发送该注册请求报文的边缘设备的桥 MAC 地址和 IP 地址信息;注册应答报文是邻居服务器在接收到边缘设备的注册请求报文之后回应的报文,携带了当前已经在邻居服务器中注册的站点信息,所述站点信息包括该站点的桥 MAC 地址和 IP 地址。这里,由于数据中心站点是通过边缘设备接入核心网的,数据中心站点的桥 MAC 地址也即该数据中心站点的边缘设备的桥 MAC 地址;数据中心的 IP 地址也即该数据中心站点的边缘设备的 IP 地址。

[0036] 另外,每个站点的边缘设备上需要预先配置邻居服务器的 IP 地址,用于根据该 IP 地址向邻居服务器发送注册请求报文。如果边缘设备向邻居服务器发送注册请求报文后,能够成功接收到邻居服务器回应的注册应答报文,则说明站点注册成功。

[0037] 步骤 401 中所述当前站点上线时,其边缘设备在邻居服务器上注册,获取所有已注册站点信息的具体过程为:边缘设备向邻居服务器发送携带了自身桥 MAC 地址和 IP 地址的注册请求报文;邻居服务器接收到该边缘设备发来的注册请求报文之后,将存储的所有已注册站点信息携带在注册应答报文中发送给该边缘设备,并将当前站点记为已注册站点,将接收到的注册请求报文中携带的桥 MAC 地址和 IP 地址存储为当前站点信息;该边缘设备接收到邻居服务器的注册应答报文后,获取该注册应答报文中携带的所有已注册站点信息。至此,边缘设备在邻居服务器上完成注册。

[0038] 在正常情况下,站点上线时,其边缘设备向邻居服务器发送注册请求报文,邻居服务器收到注册请求报文后,会发送注册应答报文到该边缘设备。但是,如果网络故障,或邻居服务器发生故障,则边缘设备和邻居服务器之间的通信则可能中断,这时,边缘设备可以进行多次注册尝试,例如周期性发送注册请求报文,直到接收到邻居服务器的注册应答报文,或者按照预先设置的注册规则进行注册。

[0039] 本发明实施例中,预先设置注册重试时间间隔列表,其中存储多个时间间隔值。以所述多个时间间隔值为发送注册请求报文的时间间隔,进行一轮注册重试,如果该轮注册重试中未收到相应的注册应答报文,则以所述多个时间间隔值为发送注册请求报文的时间间隔开始新一轮的注册重试。因此,所述边缘设备向邻居服务器发送注册请求报文之后,进一步包括:如果当前注册重试时间间隔超时的时候仍未接收到邻居服务器返回的注册应答

报文,则再次向邻居服务器发送注册请求报文,并以注册重试时间间隔列表中的下一个时间间隔值作为注册重试时间间隔,直到接收到邻居服务器返回的注册应答报文;如果当前注册重试时间间隔是注册重试时间间隔列表中的最后一个时间间隔值,且等待超时后仍未接收到邻居服务器回应的注册应答报文,则重新向邻居服务器发送注册请求报文,并以注册重试时间间隔列表中的第一个时间间隔值作为注册重试时间间隔进行新一轮的注册重试。

[0040] 假设注册重试时间间隔列表中的时间间隔值分别为 1, 2, 4, 以秒为时间单位,则站点在邻居服务器上的注册过程为:

[0041] 边缘设备向邻居服务发送注册请求报文,并等待 1 秒,如果超过 1 秒未收到邻居服务器的注册应答报文,则再次向邻居服务器发送注册请求报文,并等待 2 秒,如果超过 2 秒未收到邻居服务器的注册应答报文,则等待 4 秒,如果超过 4 秒仍未收到邻居服务器的注册应答报文,则重新按照注册重试时间间隔列表中的时间间隔值进行新一轮的重试注册。

[0042] 图 4 所示的本发明实施例中,当前站点在邻居服务器注册并上线之后,如果有新的站点在邻居服务器注册并上线,则当前站点也需要与该新注册站点之间实现二层互联,因此,需要获取该新注册站点信息,并针对该新注册站点生成虚拟桥接端口。这里,可以按照预设规则获取新注册站点信息,例如,可以周期性向邻居服务器请求获取新注册站点信息,也可以由邻居服务器在接受新站点的注册后,通知所有其它已注册站点该新注册站点信息。本实施例中,只对当前站点周期性向邻居服务器请求获取新注册站点信息的方法进一步详细说明。

[0043] 当前站点在邻居服务器注册并上线后,周期性向邻居服务器请求获取新注册站点信息的方法为:当前站点的边缘设备每隔第一预设时间向邻居服务器发送注册请求报文,并在接收到邻居服务器返回的注册应答报文之后,获取该注册应答报文中携带的所有已注册站点信息,根据每个已注册站点信息查找对应的隧道封装映射表项,如果未查找到,则确定该已注册站点为新注册的站点,并获取该新注册的站点信息。

[0044] 这里,站点信息中包含该站点的 IP 地址,因此,可以将该已注册站点信息中的 IP 地址与每个隧道封装映射表项中的远端站点的 IP 地址进行比较,如果相同,则找到对应的隧道封装映射表项,如果没有找到相同的,则说明该已注册站点是新的已注册站点。

[0045] 在实际应用中,边缘设备可能配置有多个 IP 地址,因此,可以根据 IP 地址将该边缘设备虚拟为多个虚拟边缘设备,其他站点可以针对所述多个虚拟边缘设备,生成多个虚拟桥接端口,建立与该边缘设备所在站点之间的多条隧道。这种情况下,两个虚拟边缘设备之间,则不允许针对对方生成虚拟桥接端口,否则,会造成环路。例如,假设边缘设备上配置有 IP1 和 IP2 两个 IP 地址,分别作为虚拟边缘设备 1 和虚拟边缘设备 2 的 IP 地址,如果虚拟边缘设备 1 首先向邻居服务器注册,邻居服务器会将虚拟边缘设备 1 所在站点记录为已注册站点;当虚拟边缘设备 2 向邻居服务器注册时,则邻居服务器会将虚拟边缘设备 1 所在的站点信息发送到虚拟边缘设备 2,同时将虚拟边缘设备 2 所在站点记录为已注册站点;这样,虚拟边缘设备 1 和虚拟边缘设备 2 之间会根据针对彼此生成的虚拟桥接端口建立二层隧道。当在两个虚拟边缘设备之间进行数据传输时,实际上是把边缘设备发送到核心网的报文经核心网传输后重新发送回该边缘设备,显然,这种情况是不能允许的。

[0046] 本发明实施例中,由于每个边缘设备的桥 MAC 地址唯一,可以通过边缘设备的桥

MAC 地址唯一区分一个站点。因此,边缘设备针对每个已注册站点生成虚拟桥接端口之前,进一步包括:判断该已注册站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该已注册站点对应的虚拟桥接端口,否则,生成与该已注册站点对应的虚拟桥接端口;边缘设备针对每个新注册的站点生成虚拟桥接端口之前,进一步包括:判断该新注册站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该新注册站点对应的虚拟桥接端口,否则,生成与该新注册站点对应的虚拟桥接端口。

[0047] 图 4 所示的本发明实施例中,当前站点上线并生成与每个已注册站点对应的虚拟桥接端口之后,在正常的情况下,可以与该已注册站点进行正常的通信。但是,如果出现网络故障,或者其中一个站点掉线,则两个站点之间不能进行正常通信。因此,当前站点的边缘设备生成与已注册站点生成虚拟桥接端口之后,需要先检测是否能够与该已注册站点进行正常通信,如果能够正常通信,则可以通过该已注册站点对应的虚拟桥接端口建立与该已注册站点间的隧道,并在该隧道上进行数据传输,否则,无法进行数据传输。因此,在生成虚拟桥接端口时,可以先将虚拟桥接端口的状态设置为不能正常工作,例如,故障(DOWN)状态。而且,由于网络故障或站点掉线可以发生在任何时刻,因此,两个站点之间需要定期检测是否可以和对端站点进行正常通信。

[0048] 为此,本实施例中,还预先定义了解析请求报文和解析应答报文,其中,解析请求报文是当前站点的边缘设备生成与每个已注册站点对应的虚拟桥接端口之后,发送到该已注册站点的报文,用于检测是否可以和该已注册站点进行正常通信;解析应答报文是该已注册站点接收到当前站点的边缘设备发送的解析请求报文后回应的解析应答报文。

[0049] 当前站点的边缘设备向该已注册站点发送解析请求报文,并接收到该已注册站点回应的解析应答报文之后,将该虚拟桥接端口的状态设置为启动(UP),至此,两个站点之间可以通过隧道进行数据传输。此后,边缘设备可以定期向该已注册站点发送注册请求报文,检测是否可以和该已注册站点进行正常通信,如果能够接收到该已注册站点回应的解析应答报文,则说明可以和该已注册站点进行正常通信,需要将针对该已注册站点生成的虚拟桥接端口的状态设置为 UP,或者在该虚拟桥接端口的状态为 UP 时,保持其 UP 状态不变。

[0050] 本发明实施例中,站点之间可以周期性向每个已注册站点发送解析请求报文和接收解析应答报文来判断是否可以和对端站点进行正常的通信,如果边缘设备向该已注册站点连续发送预设次数的解析请求报文均未收到相应的解析应答报文,则认为该已注册站点下线,可以将该已注册站点对应的虚拟桥接端口的状态设置为 DOWN,设置该虚拟桥接端口对应的老化定时器,并在该老化定时器超时后删除该虚拟桥接端口。

[0051] 图 4 所示的本发明实施例中,如果当前站点下线,则可以通知邻居服务器删除当前站点信息,并由邻居服务器通知其它已注册站点,使得各站点能够及时获知其它站点的下线情况。为了实现这个目的,可以预先定义注销请求报文,该报文中携带下线站点的桥 MAC 地址和 IP 地址信息。当前站点下线时,其边缘设备向邻居服务器发送注销请求报文;邻居服务器接收到该注销请求报文之后,根据该注销请求报文中携带的桥 MAC 地址和 IP 地址确定下线站点为当前站点,则删除当前站点信息,并向其它所有已注册站点发送该注销请求报文;其它各注册站点收到该注销请求报文后,根据该注销请求报文中携带的 IP 地址查找对应的隧道封装映射表项,进而确定针对当前站点生成的虚拟桥接端口,并将针对当前站点生成的虚拟桥接端口删除。

[0052] 实际上,当前站点下线时,也可以直接通过每个已注册站点生成的虚拟桥接端口,向该已注册站点发送注销请求报文,使得该已注册站点能够删除针对当前站点生成的虚拟桥接端口;但是,这样可能会有部分已注册站点不能及时获知当前站点下线。例如,站点 X 上线后获取了当前站点的信息,生成了与当前站点对应当虚拟桥接端口,此时,若当前站点要下线,且尚未获知站点 X 已上线,则将无法通知站点 X,使得站点 X 不能及时获知当前站点下线。

[0053] 另外,邻居服务器也会对注册的站点进行老化。当邻居服务器接收到站点的注册请求报文时,刷新该站点的状态,并启动老化定时器,如果在老化定时器超时之前,再次接到该站点的注册请求报文,则重新刷新该站点状态;如果老化定时器超时,未收到该站点的注册请求报文,则认为该站点掉线,生成注销请求报文,并分别发送给所有其它已注册站点,使得所有其它已注册站点删除针对该站点生成的虚拟桥接端口。

[0054] 此外,由于邻居服务器发送注销请求报文时,有可能因网络原因造成报文丢失。因此边缘设备在周期性向邻居服务器发送注册请求报文,并接收邻居服务器回应的注册应答报文时,针对注册应答报文中携带的所有已注册站点信息,除了判断确定新注册站点外,还可以判断确定当前已经在邻居服务器中注销的站点,并删除该注销站点对应的虚拟桥接端口。针对注册应答报文中携带的所有已注册站点信息,判断确定当前已经在邻居服务器中注销的站点的方法为:将边缘设备中存储的每个虚拟桥接端口对应的隧道封装映射表项中的远端站点的 IP 地址,与注册应答报文中携带的所有已注册站点信息中的站点 IP 地址进行比较,如果与注册应答报文中携带的所有已注册站点信息中的站点 IP 地址均不相同,则将该隧道封装映射表项对应的虚拟桥接端口删除。

[0055] 图 4 所示的本发明实施例中,当前站点生成与已注册站点对应的虚拟桥接端口,并配置相应的隧道封装映射表项之后,可以与该已注册站点进行二层互联。步骤 403 中所述边缘设备根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输可以采用以下方式实现:

[0056] 边缘设备在内部以太网端口接收到需要发送到该已注册站点的以太网报文时,根据针对该已注册站点生成的虚拟桥接端口对应的隧道封装映射表项对该以太网报文进行隧道封装后发送出去;该经隧道封装后的报文将在核心网中通过隧道发送到该已注册站点;这里,对以太网报文进行隧道封装,可以是将隧道封装映射表项中的远端站点的 IP 地址、本地站点的 IP 地址分别作为该报文的目的 IP 地址和源 IP 地址进行封装。

[0057] 边缘设备从核心网接收到经隧道封装,且目的 IP 地址是本机 IP 地址的 IP 报文时,将该 IP 报文解隧道封装得到源 IP 地址、目的 IP 地址、以及原始的以太网报文,再根据该原始以太网报文的目的 MAC 转发该原始以太网报文到相应的内部以太网端口。这里,根据解隧道封装得到的源 IP 地址和目的 IP 地址可以查找到对应的隧道封装映射表项,进而确定对应的虚拟桥接端口,该虚拟桥接端口也即接收该原始以太网报文的端口。

[0058] 另外,当边缘设备在内部以太网端口收到以太网报文时,可以对该以太网报文进行 MAC 地址学习,并且,可以将学习到的新的 MAC 地址通过隧道通告给其它已注册站点;边缘设备也可以将该已注册站点通告的 MAC 地址、以及接收该通告的虚拟桥接端口记录到 MAC 地址映射表中,用于在内部以太网端口接收到需要发送到该已注册站点的以太网报文时,根据报文的目的 MAC 查找 MAC 地址映射表,确定对应的虚拟桥接端口,进而根据该虚拟

桥接端口对应的隧道封装映射表项对该以太网报文进行隧道封装后发送出去。边缘设备在虚拟桥接端口接收到以太网报文时,也可以进行 MAC 地址学习,学习方法与对在内部以太网端口收到以太网报文的 MAC 地址学习方法相同。

[0059] 此外,当前站点的边缘设备生成每个已注册站点生成虚拟桥接端口之后,还可以通过和该已注册站点间的隧道进行 VLAN 协商,协商该虚拟桥接端口上可以支持的 VLAN。协商 VLAN 的方法是:边缘设备接收该已注册站点发来的携带了该已注册站点能够支持的 VLAN 的报文,并将该已注册站点能够支持的 VLAN 与当前站点的边缘设备能够支持的 VLAN 进行交集运算,所得结果即为该虚拟桥接端口能够支持的 VLAN。

[0060] 需要说明的是,本发明中,站点之间通过隧道进行数据传输时,如果传输的以太网报文为广播或组播报文,则在站点内部传输时,可以按照现有技术进行相应的广播或组播方式进行报文转发。但是,当通过核心网传输到其它已注册站点时,在发送到核心网之前,已经根据虚拟桥接端口对应的隧道封装映射表项对该以太网报文进行了隧道封装,因此,在核心网中是以单播方式传输的。例如:边缘设备在内部以太网端口接收到的以太网报文是广播报文时,将该以太网报文泛洪到包括虚拟桥接端口在内的所有端口,当通过虚拟桥接端口发送报文时,对该以太网报文进行相应的隧道封装后发送到核心网中,并在核心网中以单播方式发送。

[0061] 本发明还提供了一种边缘设备,该边缘设备为数据中心的边缘设备,应用于数据中心站点的二层互联和拓扑构建。

[0062] 参见图 5,图 5 是本发明实施例边缘设备的结构示意图,该边缘设备包括:注册单元 501、控制单元 502、传输单元 503;其中,

[0063] 注册单元 501,用于该边缘设备所在的站点上线时,在邻居服务器上注册,获取所有已注册站点信息;用于该边缘设备所在的站点上线后,按照预设规则从邻居服务器获取所有新注册的站点信息;

[0064] 控制单元 502,根据注册单元 501 获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口,并配置与该虚拟桥接端口对应的隧道封装映射表项;用于根据注册单元 501 获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口以及配置与该虚拟桥接端口对应的隧道封装映射表项;

[0065] 传输单元 503,用于根据每个已注册站点对应的隧道封装映射表项,建立与该已注册站点之间的隧道,并在该隧道上进行数据传输。

[0066] 所述注册单元 501 在邻居服务器上注册,获取所有已注册站点信息时,用于:向邻居服务器发送携带该边缘设备所在站点信息的注册请求报文,接收所述邻居服务器在接收到该注册请求报文后返回的注册应答报文,并获取接收到的注册应答报文中携带的所有已注册站点信息。

[0067] 所述注册单元 501,用于存储预先配置的注册重试时间间隔列表;

[0068] 所述注册单元 501 向邻居服务器发送注册请求报文之后,如果在当前注册重试时间间隔超时的时候仍未接收到邻居服务器返回的注册应答报文,则再次向邻居服务器发送注册请求报文,并以注册重试时间间隔列表中的下一个时间间隔值作为注册重试时间间隔,直到接收到邻居服务器返回的注册应答报文;如果当前注册重试时间间隔是列表中的最后一个时间间隔值,且等待超时后仍未接收到邻居服务器回应的注册应答报文,则重新

向邻居服务器发送注册请求报文并以注册重试时间间隔列表中的第一个时间间隔值作为注册重试时间间隔进行新一轮的注册重试。

[0069] 所述注册单元 501 按照预设规则从邻居服务器获取所有新注册的站点信息时,用于:每隔第一预设时间向邻居服务器发送注册请求报文,并在接收到邻居服务器回应的注册应答报文之后,获取该注册应答报文中携带的所有已注册站点信息,根据每个已注册站点信息在存储单元中查找对应的隧道封装映射表项,如果未查找到,则确定该已注册站点为新注册的站点,并获取该新注册的站点信息。

[0070] 所述注册单元 501 获取的站点信息包括该站点的桥 MAC 地址;

[0071] 所述控制单元 502 在根据注册单元 501 获取的每个已注册站点信息,生成与该已注册站点对应的虚拟桥接端口之前,进一步用于:判断该已注册站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该已注册站点对应的虚拟桥接端口,否则,生成与该已注册站点对应的虚拟桥接端口;

[0072] 所述控制单元 502 在根据注册单元 501 获取的每个新注册的站点信息,生成与该新注册的站点对应的虚拟桥接端口之前,进一步用于:判断该新注册的站点的桥 MAC 地址是否与当前站点的桥 MAC 地址相同,如果相同,则不生成与该新注册的站点对应的虚拟桥接端口,否则,生成与该新注册的站点对应的虚拟桥接端口。

[0073] 所述控制单元 502 生成与该已注册站点对应的虚拟桥接端口之后,进一步用于:向该已注册站点发送解析请求报文,并在接收到该已注册站点回应的解析应答报文后,将该虚拟桥接端口的状态设置为启动 UP。

[0074] 所述控制单元 502 生成与该已注册站点对应的虚拟桥接端口时,进一步将该虚拟桥接端口的状态设置为故障 DOWN;

[0075] 所述控制单元 502 在生成与该已注册站点对应的虚拟桥接端口之后,每隔第二预设时间向该已注册站点发送解析请求报文,如果连续发送预设次数的解析请求报文均未收到相应的解析应答报文后,将该已注册站点对应的虚拟桥接端口的状态设置为 DOWN,设置该虚拟桥接端口对应的老化定时器,并在该老化定时器超时后删除该虚拟桥接端口。

[0076] 所述注册单元 501,进一步用于:当前站点下线时,向邻居服务器发送携带当前站点的 IP 地址的注销请求报文,并由邻居服务器将该注销请求报文发送到所有其它已注册站点;用于接收到邻居服务器发来的注销请求报文时,根据注销请求报文中携带的下线站点的 IP 地址,在存储单元中查找对应的隧道封装映射表项,并将查找到的隧道封装映射表项对应的虚拟桥接端口删除。

[0077] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

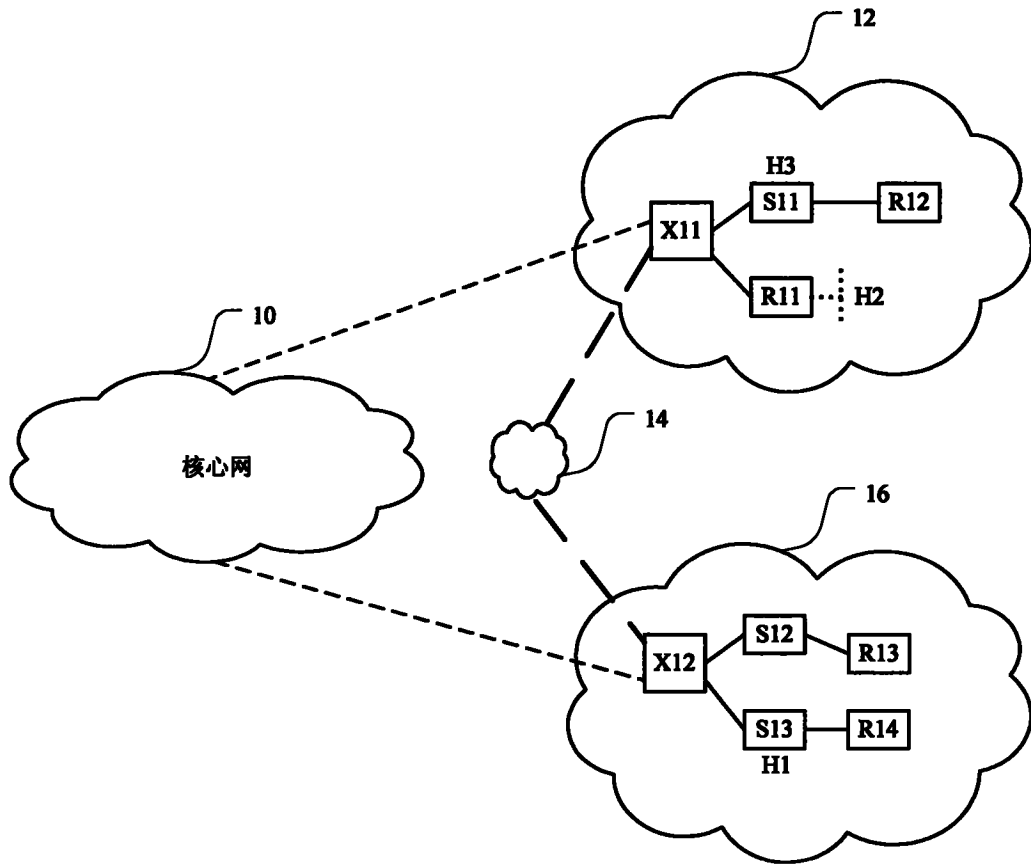


图 1

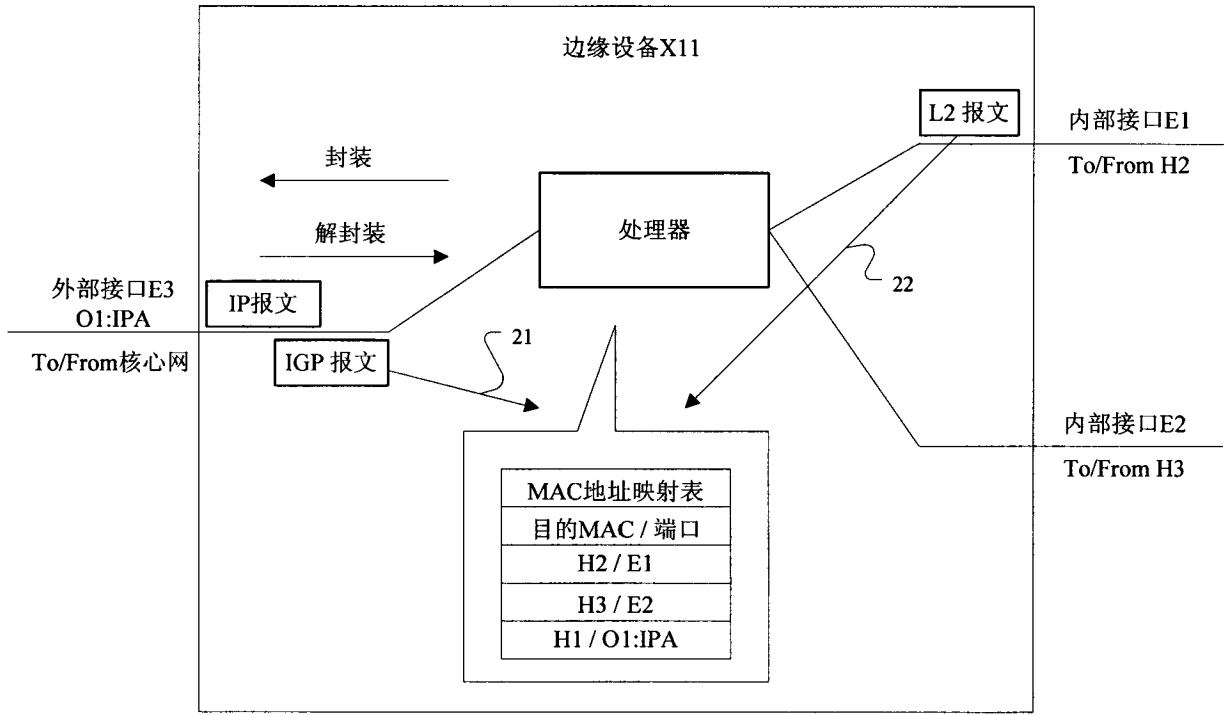


图 2

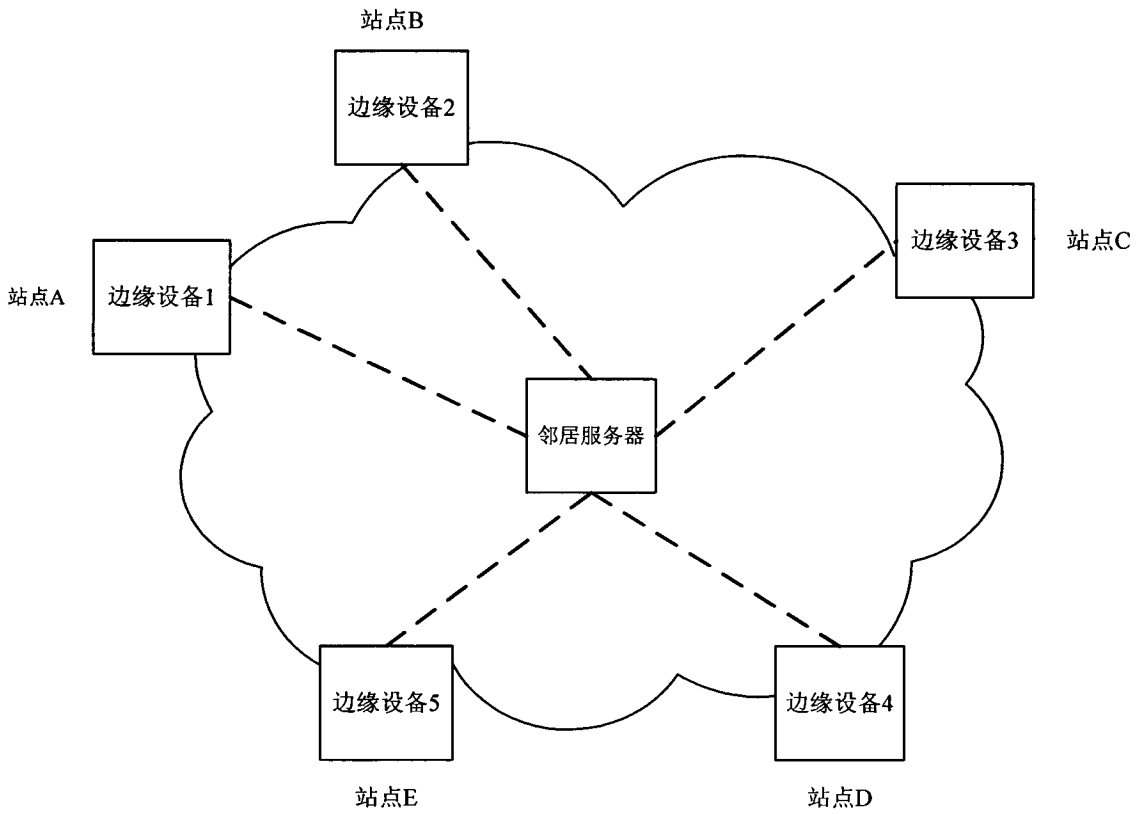


图 3

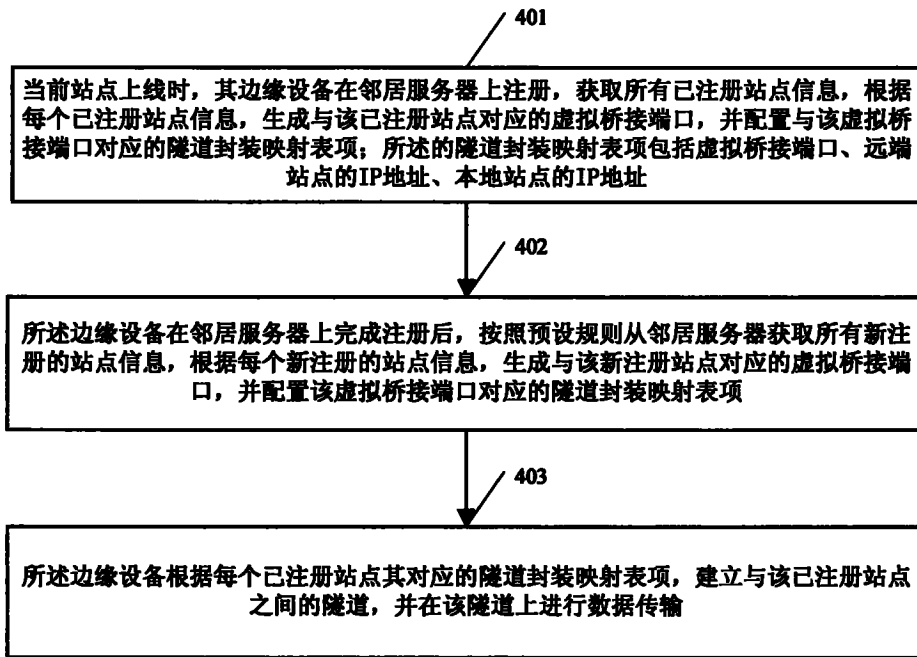


图 4

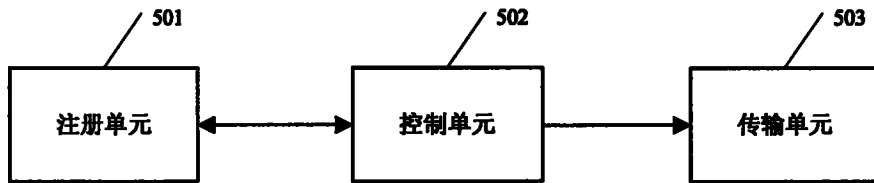


图 5