[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04L 12/56 (2006.01)
H04L 12/28 (2006.01)

[12] 发明专利说明书

专利号 ZL 200310114612.5

[45] 授权公告日 2007 年 4 月 25 日

[11] 授权公告号 CN 1312889C

[22] 申请日 2003.12.17
[21] 申请号 200310114612.5
[73] 专利权人 浪潮电子信息产业股份有限公司
地址 250014 山东省济南市山大路 224 号
[72] 发明人 侯宗浩 董小社 张 露 刘爱华 胡雷钧
[56] 参考文献
WO03041355A1 2003.5.15
US6266335B1 2001.7.24
负载均衡技术浅析 刘爱洁,电信工程技术与标准化,第 6 期 2002
一种动态网络负载平衡集群的实践方法 林凡，杨晨晖,厦门大学学报（自然科学版），第 42 卷第 4 期 2003
一种具有高可用性的通用负载均衡技术 骆宗阳，董玮文，杨宇航，王澄,计算机工程，第 29 卷第 2 期 2003
审查员 罗 玮
[74] 专利代理机构 济南信达专利事务所有限公司
代理人 姜 明

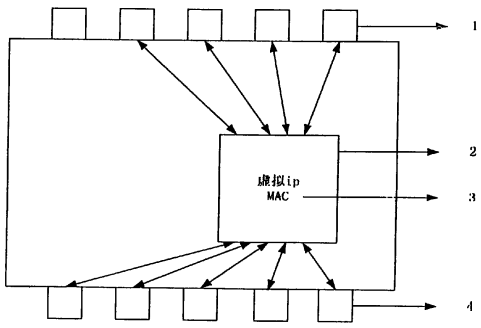
权利要求书 2 页 说明书 6 页 附图 5 页

[54] 发明名称

集群网络的单一地址流量分发器

[57] 摘要

本发明提供一个集群网络单一地址流量分发器，该分发器通过将多个物理端口绑定成一个虚拟物理端口来扩展上行的流量，并根据客户端的 IP 地址将网络流量分发到不同的处理节点来提高或扩展处理能力。 它是将后台的多台服务节点映射为单一的 IP 和 MAC 地址。 请求该 IP 的数据包将被分配到装置中不同的物理端口，分流过程以 ASIC 技术来实现，流量分类的依据是对客户端 IP 地址求 Hash 值，由于同一用户的数据被分配在同一物理端口，可为与物理端口相连的节点提供持续服务。 本发明的装置可应用于要求对上行流量进行分流的场合：第三层负载均衡器，集群路由器和集群防火墙的流量分配，以及大流量集群的多个调度器的上行流量的分流。



1、集群网络单一地址流量分发器，其特征在于实现步骤如下：

a. 该分发器是通过在单一地址流量分发器上将多个物理端口绑定成一个虚拟物理端口来扩展上行的数据流量，并根据客户端的IP 地址将网络流量分配到不同的处理节点上来实现第三层负载均衡；

b. 该分发器是通过将前后关联的数据包分发到同一个物理端口上以提高处理能力；

c. 该分发器还通过功能的硬件化以提高数据的转发速度；

d. 该分发器对外提供一个单一虚拟 IP 地址和单一的虚拟物理地址 MAC;

e. 该分发器可绑定物理端口的数目取决于分发器所拥有的物理端口数目，当物理端口数目为 N 时，那么分发器可绑定的物理端口数为 2 ～ (N-1);

f. 第三层负载均衡是以位于网络第三层客户端的IP 地址为依据将数据报文分发到不同的绑定物理端口，每个绑定物理端口连接的节点是并行处理VIP 请求的一部分，连接绑定物理端口的节点是真实服务器，第四层负载均衡器是防火墙、路由器或是一个二级转发器或多级转发器；

g. 前后关联的数据包，是来自同一用户的相互之间具有一定关系的数据包，需要同一台服务器节点进行处理，同一连接的第一个包或后续的包以及ftp 建立的一次控制连接和多个数据连接都具有相同安全套接字层的 http 请求；

h. 保证前后关联的数据包分发到相同的物理端口是同一源IP 的数据包被接收后，通过Hash 算法所得到的Hash 值是相同的，相同的Hash 值对应同一物理端口，这些数据包都从这个物理端口被分发出去；

i. 功能的硬件化是将Hash 算法程序固化在ASIC 芯片中作为分发器的核心来提高数据转发速度， Hash 算法采用以下公式表示：

Value ＝ Mod (Source_IP * 2654435761,n)

2654435761 是介于32 位IP 地址之间的一个质数，n 是模，是绑定的物理端口数或是绑定物理端口的数目的整数倍；

j. 以客户端的IP 地址为依据，计算客户端的IP 地址的Hash 值，查询Hash 转发表可得到该IP 地址所发数据包应经过的物理端口，若Hash 转发表没有该 Hash 值的纪录，该分发器按照加权最小连接方法或者加权轮巡方法为新的 Hash 值分配物理端口，并记录到 Hash 转发表中，Hash 转发表定期记录每个

Hash 值所对应网络流量，物理端口号由该装置内部指定，Hash 值和物理端口号采用一对一或多对一的对应关系；当Hash 值的数量定义为绑定物理端口数量的2 的N 次方倍，N 为非负整数，则根据其流量和参照物理端口信息登记表，IP 地址可计算出的Hash 值的数量大于绑定物理端口的数量，并选择恰当的方式，记录每个Hash 值所对应的网络流量，当物理端口流量不均衡时，重新组合物理端口号与 Hash 值的关系以均衡各物理端口分配的网络流量。

2、根据权利1 所述的分发器，其特征在于高扩展性是通过扩展连接绑定物理端口上的节点来实现，其扩展的最大数目由该单一地址流量分发器的可绑定物理端口数目决定，高可用性是当与某个物理端口相连的节点出现问题时，单一地址流量分发器根据实现协商的备份策略将失效节点所对应物理端口的数据包发送到备份节点所对应的物理端口上，物理端口信息登记表记录每个物理端口的网络数据流量的上下限，以及连接该物理端口和节点的地址信息，物理端口流量达到上限U_line 时，说明与该物理端口所连接的节点的处理能力已经达到 上限，该上限由用户根据相连物理端口的处理能力指定，转发报文到绑定物理端口，把报文的目的IP 地址和MAC 地址从单一地址流量分发器所设定的虚拟IP 地址和虚拟MAC 地址修改为该绑定物理端口所对应节点的IP 地址和虚拟 MAC 地址。

集群网络的单一地址流量分发器

一、技术领域

本发明涉及计算机领域，是一种用于构建高可用性、高可扩展性集群网络的单一地址流量分发器。

二、技术背景

快速增长的网上活动使网络服务器的访问量大大增加，导致网络的许多关键部位成为性能瓶颈或者带宽瓶颈。目前，普遍采用第四层负载均衡器或者交换机物理端口的Trunking（链路聚集）的方法解决这个问题。第四层负载均衡器可以解决性能瓶颈的问题，但是它只能提供一个上行的入口，无法增加上行的网络带宽；Trunking 技术，可以解决网络带宽问题，但是Trunking 的物理端口必须连接同一台服务节点，无法提高处理性能。

三、发明内容

本发明的目的是设计单一地址流量分发器，通过将多个物理端口绑定成一个虚拟物理端口来扩展上行的流量，并根据客户端的IP 地址将网络流量分到不同的处理节点实现提高扩展处理能力。而且该分发器保证了前后关联的数据包分配到同一个物理端口做后续的处理。

实现步骤如下：

a. 该分发器是通过在单一地址流量分发器上将多个物理端口绑定成一个虚拟物理端口来扩展上行的数据流量，并根据客户端的IP 地址将网络流量分配到不同的处理节点上来实现第三层负载均衡；

b. 该分发器是通过将前后关联的数据包分发到同一个物理端口上以提高处理能力；

c. 该分发器还通过功能的硬件化以提高数据的转发速度；

d. 该分发器对外提供一个单一虚拟 IP 地址和单一的虚拟物理地址 MAC;

e. 该分发器可绑定物理端口的数目取决于分发器所拥有的物理端口数目，当物理端口数目为 N 时，那么分发器可绑定的物理端口数为 2 ～ (N-1);

f. 第三层负载均衡是以位于网络第三层客户端的IP 地址为依据将数据报文分发到不同的绑定物理端口，每个绑定物理端口连接的节点是并行处理 VIP

请求的一部分,连接绑定物理端口的节点，可以是真实服务器，第四层负载均衡器是防火墙、路由器或是一个二级转发器或多级转发器；

g. 前后关联的数据包，是来自同一用户的相互之间具有一定关系的数据包，需要同一台服务器节点进行处理，同一连接的第一个包或后续的包，ftp 建立的一次控制连接和多个数据连接具有相同安全套接字层（ssl）的 http 请求；

h. 保证前后关联的数据包分发到相同的物理端口是同一源IP 的数据包被接收后，通过Hash 算法所得到的Hash 值是相同的，相同的Hash 值对应同一物理端口，这些数据包都从这个物理端口被分发出去；

i. 功能的硬件化是将Hash 算法程序固化在ASIC 芯片中作为分发器的核心来提高数据转发速度， Hash 算法采用以下公式表示：

Value = Mod (Source_IP * 2654435761,n)

2654435761 是介于32 位IP 地址之间的一个质数，n 是模,是绑定的物理端口数或是绑定物理端口的数目的整数倍；

k. 以客户端的IP 地址为依据,计算客户端的IP 地址的Hash 值，查询Hash 转发表可得到该IP 地址所发数据包应经过的物理端口，若Hash 转发表没有该Hash 值的纪录，该分发器按照加权最小连接方法或者加权轮巡方法为新的Hash 值分配物理端口，并记录到 Hash 转发表中，Hash 转发表定期记录每个Hash 值所对应网络流量，物理端口号由该装置内部指定，Hash 值和物理端口号可以采用一对一或多对一的对应关系；当定义为绑定物理端口数量的2 的N 次方倍，N 为非负整数，则根据其流量和参照物理端口信息登记表，IP 地址可计算出的Hash 值的数量大于绑定物理端口的数量，并选择恰当的方式，记录每个Hash 值所对应的网络流量，当物理端口流量不均衡时，重新组合物理端口号与 Hash 值的关系以均衡各物理端口分配的网络流量。

四、附图说明

附图 1 为单一地址流量分发器结构示意图；

1 为绑定得物理端口；2 为内部交换芯片；3 为虚拟的IP 地址和MAC 地址；4 为客户访问物理端口。

附图 2 为接收报文流程图；

附图 3 为转发器所用表格；

1、转发表：描述物理端口和 Hash 值得对应关系。

2、物理端口信息登记表：纪录每个物理端口的信息，包括流量信息，和连

接在该物理端口的节点地址信息。

3、物理端口备份关系表：纪录物理端口的备份信息。

附图 4 为防火墙应用方案；

采用两个单一地址流量分发器，中间连接多个防火墙服务器。

附图 5 为单一地址流量分发器级联应用方案。

五、 实施方式：

参照附图说明对本发明的分发器作以下详细的说明。

1，绑定多个物理端口对外提供单一虚拟 IP 和虚拟 MAC。

（如附图1 所示）本发明的分发器在内部提供一个虚拟的IP 地址和MAC 地址，为客户端提供一个单一的访问地址，当接收到ARP 包时，直接将虚拟的 MAC 返回给用户。该 IP 地址和 MAC 地址可以通过专用控制接口或网络登陆到单一地址流量分发器来设定。绑定物理端口的数目默认为该装置物理端口总数的一半，目的是保持进出该装置的网络流量的平衡，当然也可以根据用户的需要，调节绑定物理端口数。

2、（如附图3 所示）对客户端IP 地址按照Hash 算法进行计算，得 出 Hash 值，并根据Hash 转发表，将不同Hash 值的数据包转发到不同的绑定物理端口，实现可靠的第三层负载均衡功能。

（如附图2 流程图所示）当单一地址流量分发器接受一个数据包的时候，转发的方法是通过修改报文的目的IP 地址和MAC 地址。从客户端发来的数据包，其目的IP 地址和目的MAC 地址是单一地址流量分发器的虚拟IP 和虚拟 MAC，在经过该装置时，这两个地址被Hash 值所指定物理端口后端真实的节点的IP 地址和MAC 地址所代替，使数据包可以转发到该真实节点。同样返回的数据包，其目的IP 地址和MAC 地址是客户端的地址，源地址是真实节点的 IP 和MAC 地址，在通过单一地址流量分发器的时候，需要把源地址修改为虚拟地址。

3、采用恰当的 Hash 值与物理端口的分配方案可以使网络流量更均衡。

访问的数据包中有一些是通过代理，有一些是直接访问，这就造成了不同源IP 之间访问流量的不均衡。假设该装置有4 个可以绑定的物理端口，目前有 8 个源 IP，它们的流量分布如下：

IP1	IP2	IP3	IP4	IP5	IP6	IP7	IP8
800	700	500	200	300	200	100	100

如果Hash 算法中的n 取4， IP1 和IP2 得到同一个Hash 值，其数据包被分发到同一个物理端口，同样，IP3 和IP4，IP5 和IP6，IP7 和IP8 分别分发到相同的物理端口，那么各物理端口分配的流量如下：

物理端口号	1	2	3	4
相同 Hash 值 IP	IP1,IP2	IP3,IP4	IP5,IP6	IP7,IP8
流量	1500	700	500	200

如果Hash 算法中的n 取8， 则每个源IP 都得到一个Hash 值，本装置可以根据Hash 转发表中的每个Hash 值对应的流量，按照恰当的算法，合理的分配各 Hash 值对应的物理端口号，可以得到下面的分配方案：

物理端口号	1	2	3	4
相同 Hash 值 IP	IP1	IP2	IP3,IP6	IP4,IP5, IP7,IP8
流量	800	700	700	700

如上表所示可见网络流量更加均衡。

4、硬件实现功能

把 Hash 算法程序烧录在 ASIC(Application Specific Integrated Circuit) 芯片中作为该装置的核心，以提高转发的速度。

5、高可扩展性

该装置可以根据可绑定的物理端口数扩展连接到各物理端口的节点数（可绑定的物理端口数可以调整），其最大的扩展数是N-1，N 是该装置的物理端口数。高扩展性是通过扩展连接绑定物理端口上的节点来实现，其扩展的最大数目由该单一分发器的可绑定物理端口数目决定，高可用性是当与某个物理端口相连的节点出现问题时，单一地址流量分发器可以根据实现协商的备份策略将失效节点所对应物理端口的数据包发送到备份节点所对应的物理端口上，物理端口信息登记表，记录每个物理端口的网络数据流量的上下限，以及连接该物理端口和节点的地址信息，物理端口流量达到U_line 时，说明与该物理端口所连接的节点的处理能力已经达到 上限，该上限由用户根据相连物理端口的处理能力指定，转发报文到绑定物理端口，把报文的目的IP 地址和MAC 地址从单一地址流量分发器所设定的虚拟IP 地址和虚拟MAC 地址修改为该绑定物理端口所对应节点的 IP 地址和虚拟 MAC 地址。

6、高可用性

当与某个物理端口相连的服务器出现问题时，可以采取下面两种保证所提

供的服务正常运行：

a) 如果后续节点之间无备份，则调整hash 算法参数，自动根据流量调整的算法来改变结果和物理端口的对应关系。

b) 虚拟物理端口之间或后续节点之间有备份关系，则将失效物理端口的所有包发向具有备份功能的节点。虚拟物理端口之间若有一种对应关系，这种对应关系反映了各物理端口之间的备份关系，这些关系在该设备中可以通过一张表来描述，（如附图3 表3 所示），当某个物理端口对应的链路不通时（处理节点宕机或链路问题），发往各物理端口的包将发到备份物理端口。

7、可以采用的应用方式

a) 直接连接真实服务器 (充当小型集群的负载均衡器)

按照Hash 方法分配用户，同一用户的所有数据包将被分发到相同的真实服务器上，由于同一连接的源IP 是相同的，因此，所有数据都会被发向同一个物理端口，即被同一个真实服务器接收。

b) 连接多个集群调度器

如果集群上行数据量比较大，可以使用多个负载均衡器，该设备可以为多个负载均衡器提供唯一的虚拟ip 地址，由于负载均衡器既要均衡负载，也要保证将属于同一连接的数据包发送到后台的同一台真实服务器去处理，所以，必须将属于前后相关的数据包分配给同一个负载均衡器，该设备以IP 地址来区分不同的上行数据包，只要源IP 相同，数据包将被送到设备的同一物理端口，从而为同一调度器来调度。

c) 连接集群防火墙

防火墙容易成为网络的瓶颈，因此，构建集群防火墙是防火墙的一个发展趋势，集群防火墙也需要对网络的数据包进行分流。可用单一地址流量分发器完成此工作，如附图 4 所示。

d) 实现多级转发器

（如附图3、附图5 和表2 所示）该设备可以级联，级联的基础是一个物理端口对应不同的多个hash 值，如设备1 给port1 分配的hash 值为1，2，3，4，5；给port5 分配的21、22、23、24、25。则设备2 可进一步给设备2 的1、2、3、4、5 物理端口分配hash 值1、2、3、4、5，设备3 可进一步给设备3 的物理端口1、2、3、4、5 分配hash 值21、22、22、23、24、25，。此时，其物理端口信息登记表中的 U_line 值可以视二级转发器绑定的数目设置为连接

二级转发器的整数倍。

e) 连接机群路由器

集群路由器可用交换机去构造，此设备可代替交换机，构造方法同集群防火墙。

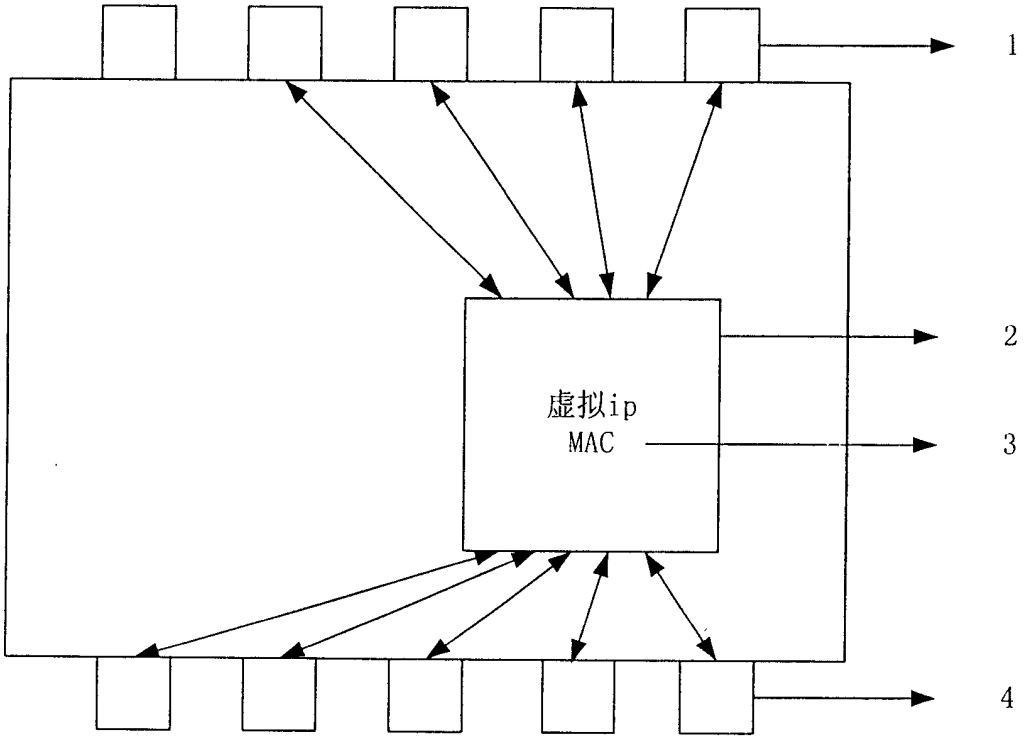


图 1

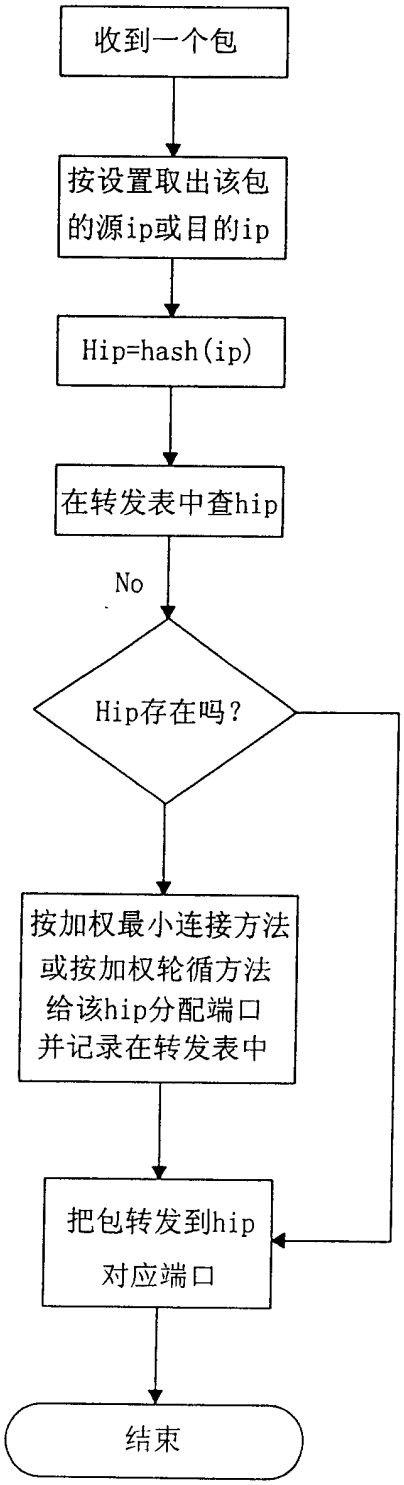

收到一个包
按设置取出该包的源ip或目的ip
Hip=hash(ip)
在转发表中查hip
No
Hip存在吗?
按加权最小连接方法或按加权轮循方法给该hip分配端口并记录在转发表中
把包转发到hip对应端口
结束

图2

表 1：转发表

port 号	Hash 值	流量

表 2：端口信息登记表

Port 号	下线	上限	前 1s 流量	前 5 分钟流量	节点 IP	节点 MAC	权值
	D_line	U_line	U_line	U_line			

表 3：端口备份关系

port 号	备份 port 号

图 3

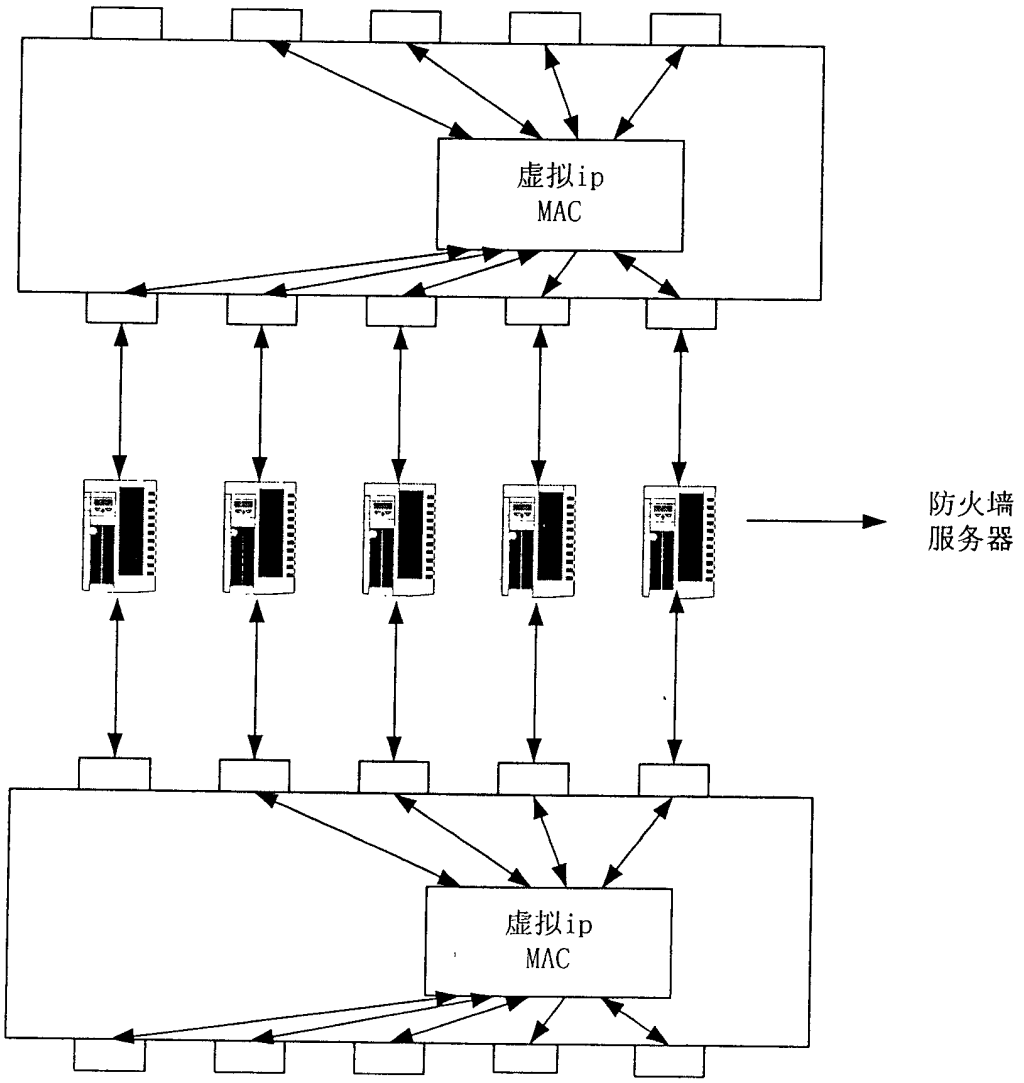


图 4

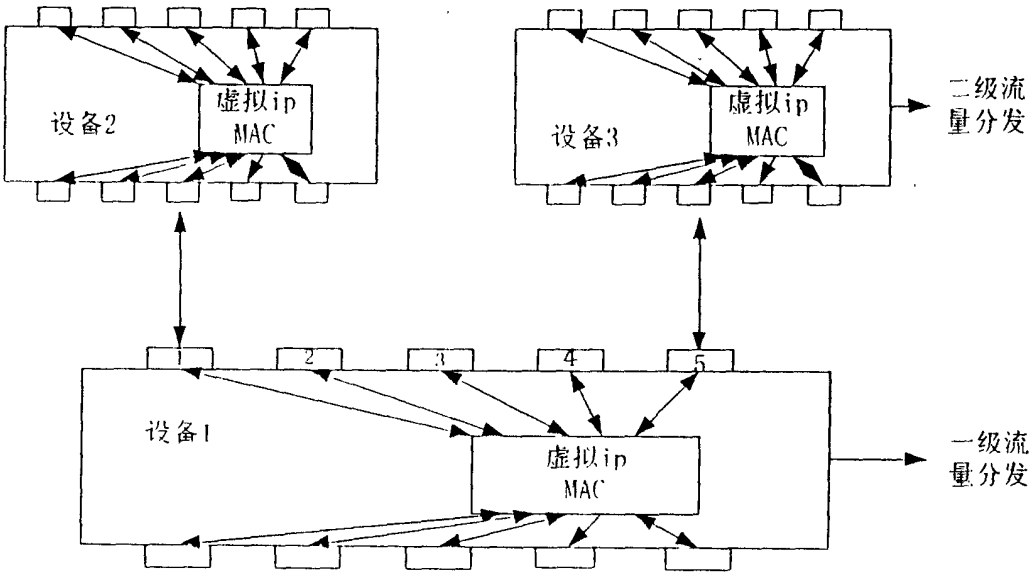


图 5