



(12)发明专利

(10)授权公告号 CN 107995123 B

(45)授权公告日 2020.05.01

(21)申请号 201610948570.2

(22)申请日 2016.10.27

(65)同一申请的已公布的文献号
申请公布号 CN 107995123 A

(43)申请公布日 2018.05.04

(73)专利权人 贵州白山云科技股份有限公司
地址 550000 贵州省贵阳市南明区富源北路22号富源医药物流园二期41

(72)发明人 苗辉 庄吴敏 李逸骏

(74)专利代理机构 厦门市精诚新创知识产权代理有限公司 35218

代理人 方惠春

(51)Int.Cl.

H04L 12/803(2013.01)

(56)对比文件

CN 104301417 A,2015.01.21,
CN 104301417 A,2015.01.21,
CN 104144120 A,2014.11.12,
CN 103078804 A,2013.05.01,
US 2015/0304219 A1,2015.10.22,
Jaeyoung Kim等.Next-Hop Selection Algorithm over ECMP.《2006 Asia-Pacific Conference on Communication》.2006,

审查员 徐苏宁

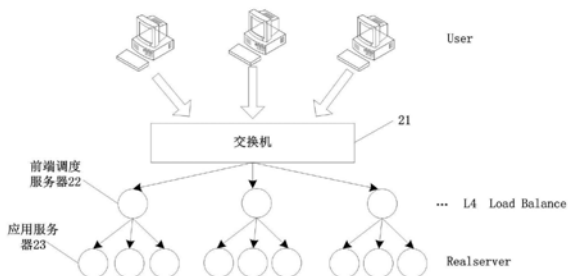
权利要求书2页 说明书11页 附图3页

(54)发明名称

一种基于交换机的负载均衡系统和方法

(57)摘要

本发明公开了一种基于交换机的负载均衡系统和方法,该系统采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中前端调度服务器与应用服务器的个数均大于1。在本发明中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题。



1. 一种基于交换机的负载均衡系统,其特征在于,包括:交换机、N台前端调度服务器、M台应用服务器,其中,N和M均为大于1的整数;

所述交换机,用于接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器,由该前端调度服务器将接收到的所述请求数据包发送至相应的所述应用服务器,经由所述应用服务器将所述请求数据包的响应结果返回给所述用户端;

所述前端调度服务器,用于将接收到的所述请求数据包发送至相应的所述应用服务器;

所述应用服务器,用于响应所述前端调度服务器发送的所述请求数据包,并将相应的响应结果返回给所述用户端;

所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同;

其中,所述交换机接收用户端发送的请求数据包,这一操作,还用于抵抗SYN Flood攻击;所述交换机还用于在接收用户端发送的请求数据包之前,获取与各所述前端调度服务器之间的通信链路的链路状态信息,并根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息,以及向各所述前端调度服务器广播更新后的链路状态信息;

所述前端调度服务器还用于发送和接收所述链路状态信息;

以及所述交换机还用于将所述ECMP路由表项中各路径条目进行取模计算,并根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化,若确定任一条与所述前端调度服务器之间的通信链路断开后,则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变,并将断开的通信链路的流量重新进行哈希计算,分配到其他通信链路上;若确定所述ECMP路由表项中通信链路对应的路径条目增多,则将活跃的通信链路上的部分流量分配至新增通信链路上。

2. 一种基于如权利要求1所述的负载均衡系统的负载均衡方法,其特征在于,该方法包括:

交换机接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址;

根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目;

将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器,由该前端调度服务器将接收到的所述请求数据包发送至相应的所述应用服务器,经由所述应用服务器将所述请求数据包的响应结果返回给所述用户端;

所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同;

其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击;在接收用户端发送的请求数据包之前,获取与各所述前端调度服务器之间的通信链路的链路状态信息;

根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息,并向各所述前端调度服务器广播更新后的链路状态信息;

将所述ECMP路由表项中各路径条目进行取模计算;

根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化,若确定任一条与所述前端调度服务器之间的通信链路断开后,则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变,并将断开的通信链路的流量重新进行哈希计算,分配到其他的通信链路上;若确定所述ECMP路由表项中通信链路对应的路径条目增多,则将活跃的通信链路上的部分流量分配至新增通信链路上。

3. 一种基于如权利要求1所述的负载均衡系统的负载均衡方法,其特征在于,该方法包括:

前端调度服务器接收所述交换机发送的来自用户端的请求数据包;其中,所述请求数据包是由所述交换机根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器;

将所述请求数据包发送至相应的所述应用服务器,以由所述应用服务器将所述请求数据包的响应结果返回给所述用户端;

所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同;

其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击;所述前端调度服务器发送和接收与所述交换机之间的链路状态信息。

4. 一种基于如权利要求1所述的负载均衡系统的负载均衡方法,其特征在于,该方法包括:

应用服务器响应所述前端调度服务器发送的来自所述用户端的所述请求数据包,并将相应的响应结果返回给所述用户端;

其中,所述请求数据包是由所述交换机将接收到的来自用户端的所述请求数据包根据预设均衡条件确定分配的相同的下一跳地址,并根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目后,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器所得到的;

所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同;

其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击。

一种基于交换机的负载均衡系统及方法

技术领域

[0001] 本发明涉及网络通信技术领域,尤其涉及一种基于交换机的负载均衡系统及方法。

背景技术

[0002] LVS是Linux Virtual Server的简写,是一个虚拟的服务器集群系统,采用IP负载均衡技术和基于内容请求分发技术实现的负载均衡。

[0003] IP负载均衡技术是在负载调度器的实现技术中效率最高的。在已有的IP负载均衡技术中,主要有通过网络地址转换(NAT,Network Address Translation)将一组服务器构成一个高性能的、高可用的虚拟服务器,称之为VS/NAT技术(Virtual Server via Network Address Translation)。在分析VS/NAT的缺点和网络服务的非对称性的基础上,还提出了通过直接路由实现虚拟服务器的方法VS/DR(Virtual Server via Direct Routing)和通过IP隧道实现虚拟服务器的方法VS/TUN(Virtual Server via IP Tunneling),它们可以极大地提高系统的伸缩性。VS/NAT、VS/DR和VS/TUN技术是LVS集群中实现的三种IP负载均衡技术。除此之外,还有一种名为“FullNat”的负载均衡技术。

[0004] 如图1所示,采用IP负载均衡技术的LVS集群服务通过分层方式来提供服务。

[0005] 1、负载调度层:它位于整个集群系统的最前端,有两台(NAT、DR或TUN方式)或者两台以上(FullNat方式)的负载调度器(即前端调度服务器,或者简称调度器)组成。

[0006] 2、后端应用服务层:实际是由一组运行应用服务的机器组成。后端应用服务层的机器(简称Realserver)的硬件性能并不需要完全统一,可以各异,因为前端调度服务器可以人为地定义调度机制去调度后端应用服务器。

[0007] 其工作原理如下:当有大量的数据需要访问某种应用服务(WWW服务、DNS服务)时,数据首先经过负载调度器。负载调度器通过各种调度算法,比如,轮询调度算法、加权调度算法、ECP算法等,将数据有目标地发送到多台后端应用服务器,使得后端应用服务器比较有效地、均衡地提供服务。当后端应用服务器宕机,或者不能提供服务的时候,负载调度器也能通过检测机制(如Keep alive)发现,然后将不能提供服务的后端应用服务器剔除。

[0008] 现有的LVS集群服务可以有效地提供稳定、可靠的服务,但同时也带来了如下问题:

[0009] 1、LVS集群服务系统中前端的调度服务器都是主备架构,也就是说,只有一台服务器能提供服务,剩余的一台或者多台调度服务器都是备用的,并且主备架构的联系机制也是比较复杂的。当主调度服务器出现异常不能提供服务,主备服务器需要有各种机制去发现主调度器宕机以后,备用调度服务器接管主调度服务器的角色,从而使得主备调度服务器的角色不能充分利用资源,备用调度服务器是一种长时间等待的角色,不能主动提供服务,导致资源浪费。

[0010] 2、当出现大规模的SYN Flood攻击的时候,前端调度服务器提供服务的只有一台主调度服务器,不能很好的抵抗恶意的大量访问,会出现主调度服务器负载过大的问题,甚

至会造成主调度服务器不能对外提供服务的情况发生。

[0011] 3、主备架构或者一主多备架构的负载调度服务器扩展性比较差,不能同时提供两个或者两个上的调度服务器提供服务。

[0012] 也就是说,现有LVS集群服务系统存在资源浪费、抗干扰能力差以及扩展性差等问题。

发明内容

[0013] 本发明实施例提供了一种基于交换机的负载均衡系统及方法,用以解决现有LVS集群服务系统存在资源浪费、抗干扰能力差以及扩展性差的问题。

[0014] 本发明实施例提供了一种基于交换机的负载均衡系统,包括:交换机、N台前端调度服务器、M台应用服务器,其中,N和M均为大于1的整数;

[0015] 所述交换机,用于接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器,以由该前端调度服务器将接收到的所述请求数据包发送至相应的所述应用服务器,经由所述应用服务器将所述请求数据包的响应结果返回给所述用户端;

[0016] 所述前端调度服务器,用于将接收到的所述请求数据包发送至相应的所述应用服务器;

[0017] 所述应用服务器,用于响应所述前端调度服务器发送的所述请求数据包,并将相应的响应结果返回给所述用户端;

[0018] 所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同;

[0019] 其中,所述交换机接收用户端发送的请求数据包,这一操作,还用于抵抗SYN Flood攻击。

[0020] 进一步地,所述交换机,还用于在接收用户端发送的请求数据包之前,获取与各所述前端调度服务器之间的通信链路的链路状态信息,并根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息,以及向各所述前端调度服务器广播更新后的链路状态信息;

[0021] 所述前端调度服务器,还用于发送和接收所述链路状态信息。

[0022] 进一步地,所述交换机,还用于将所述ECMP路由表项中各路径条目进行取模计算,并根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化,若确定任一条与所述前端调度服务器之间的通信链路断开后,则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变,并将断开的通信链路的流量重新进行哈希计算,分配到其他的通信链路上;若确定所述ECMP路由表项中通信链路对应的路径条目增多,则将活跃的通信链路上的部分流量分配至新增通信链路上。

[0023] 进一步地,本发明实施例还提供了基于上述负载均衡系统的负载均衡方法,该方法包括:

[0024] 交换机接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包

分配的相同的下一跳地址；

[0025] 根据确定的所述下一跳地址，从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目；

[0026] 将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器，以由该前端调度服务器将接收到的所述请求数据包发送至相应的所述应用服务器，经由所述应用服务器将所述请求数据包的响应结果返回给所述用户端；

[0027] 所述预设均衡条件：芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同；

[0028] 其中，所述交换机接收用户端发送的请求数据包，这一步骤，还用于抵抗SYN Flood攻击。

[0029] 进一步地，所述在接收用户端发送的请求数据包之前，所述方法还包括：

[0030] 获取与各所述前端调度服务器之间的通信链路的链路状态信息；

[0031] 根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息，并向各所述前端调度服务器广播更新后的链路状态信息。

[0032] 进一步地，所述方法还包括：

[0033] 将所述ECMP路由表项中各路径条目进行取模计算；

[0034] 根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化，若确定任一条与所述前端调度服务器之间的通信链路断开后，则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变，并将断开的通信链路的流量重新进行哈希计算，分配到其他的通信链路上；若确定所述ECMP路由表项中通信链路对应的路径条目增多，则将活跃的通信链路上的部分流量分配至新增通信链路上。

[0035] 进一步地，本发明实施例还提供了另一基于上述负载均衡系统的负载均衡方法，该方法包括：

[0036] 前端调度服务器接收所述交换机发送的来自用户端的请求数据包；其中，所述请求数据包是由所述交换机根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址，根据确定的所述下一跳地址，从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目，并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器；

[0037] 将所述请求数据包发送至相应的所述应用服务器，以由所述应用服务器将所述请求数据包的响应结果返回给所述用户端；

[0038] 所述预设均衡条件：芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同；

[0039] 其中，所述交换机接收用户端发送的请求数据包，这一步骤，还用于抵抗SYN Flood攻击。

[0040] 进一步地，所述方法还包括：

[0041] 所述前端调度服务器发送和接收与所述交换机之间的链路状态信息。

[0042] 进一步地，本发明实施例还提供了另一基于上述负载均衡系统的负载均衡方法，该方法包括：

[0043] 应用服务器响应所述前端调度服务器发送的来自所述用户端的所述请求数据包，

并将相应的响应结果返回给所述用户端；

[0044] 其中,所述请求数据包是由所述交换机将接收到的来自用户端的所述请求数据包根据预设均衡条件确定分配的相同的下一跳地址,并根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目后,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器所得到的；

[0045] 所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同；

[0046] 其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击。

[0047] 本发明有益效果如下:

[0048] 本发明实施例采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中,前端调度服务器与应用服务器的个数均大于1。在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

附图说明

[0049] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,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,对于本领域的普通技术人员来讲,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。

[0050] 图1所示为现有采用LVS集群服务的负载均衡系统；

[0051] 图2所示为本发明实施例一所述的基于交换机的负载均衡系统的结构示意图；

[0052] 图3所示为本发明实施例一所述的负载均衡系统的具体工作流程图；

[0053] 图4所示为本发明实施例一中具体的Hash计算过程；

[0054] 图5所示为本发明实施例二所述的负载均衡方法的流程示意图；

[0055] 图6所示为本发明实施例三所述的负载均衡方法的流程示意图。

具体实施方式

[0056] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,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它实施例,都属于本发明保护的范围。

[0057] 实施例一:

[0058] 本发明实施例一提供了一种基于交换机的负载均衡系统,如图2所示,其为本发明实施例一所述的基于交换机的负载均衡系统的结构示意图,所述负载均衡系统包括:交换

机21、N台前端调度服务器22、M台应用服务器23,其中,N和M均为大于1的整数;

[0059] 所述交换机21,用于接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器22,以由该前端调度服务器22将接收到的所述请求数据包发送至相应的所述应用服务器,经由所述应用服务器23将所述请求数据包的响应结果返回给所述用户端,其中,所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同。

[0060] 所述前端调度服务器22,可用于将接收到的所述请求数据包发送至相应的所述应用服务器23;

[0061] 所述应用服务器23,用于响应所述前端调度服务器22发送的所述请求数据包,并将相应的响应结果返回给所述用户端。

[0062] 需要说明的是,所述预设均衡条件,具体的应用场景为:对于任意的网络四元组对,定义hash函数 $f(ip, port) = y$ 总是属于集合ECMP路由表项,即任意一个四元组对都能够hash计算为ECMP路由表项中的某个槽位。在ECMP路由表项中,每个槽位都唯一定义一个下一条地址。因此任意一个地址都能够通过hash计算唯一确定一个ECMPECMP路由表项中的槽位,进而确定相同的下一条地址。

[0063] 另外,需要说明的是,所述交换机接收用户端发送的请求数据包,这一操作,还用于抵抗SYN Flood攻击。

[0064] 再有,需要说明的是,ECMP是一种IP路由等价多路径,是基于OSI七层模型网络层路由层面的计算结果。等价多路径的意义在于:在存在多条不同链路到达同一目的地址的网络环境中,如果使用传统的路由技术,发往该目的地址的数据包只能利用其中的一条链路,其他链路处于备份状态或无效状态,并且在动态路由环境下路径的相互切换需要一定的时间;而ECMP协议可以在该网络环境下同时使用多条链路,不仅增加了传输带宽,而且可以无时延无丢包地备份失效链路的数据传输。ECMP最大的特点就是实现了等价情况下,多路径负载均衡和链路备份的目的。在实际应用中,可以根据前端调度服务器22集群中前端调度服务器22数量的多少确定配置ECMP的成员个数。

[0065] 也就是说,本发明实施例所述的负载均衡系统不局限于2台主备前端调度服务器的架构,可以很好地横向扩展2台以上的主主架构前端调度服务器,避免了主备前端调度服务器在故障切换时各种复杂的联系机制,以及主备前端调度服务器中备用前端调度服务器长时间等待的角色,采用不仅限于2台的主主架构前端调度服务器提高了资源的利用效率,并且当出现恶意攻击时,各主主架构中的前端调度服务器能同时工作,同时抵抗攻击流量,使得容灾能力增强,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0066] 下面结合图2说明本发明实施例的所述负载均衡系统的整体架构。

[0067] User层为普通用户访问,分布在不同地理位置的成千上万的访问用户,是带有大量访问应用请求的用户数据包,通过互联网访问连接到达基于交换机的服务器集群。

[0068] 本发明实施例所述基于交换机的负载均衡系统的系统架构分为三层架构:交换机

层、前端服务器层和Realserver层(应用服务器)。(1)交换机层是互联网数据中心的网络设备,在交换机层做L4负载均衡,控制4层一致性Hash,将流量均摊至前端服务器层的子节点;需要说明的是,交换机的ECMP(Equal Cost MultiPath)功能可以基于数据流的五元组做hash计算,并结合预设均衡条件,确定下一跳地址,进而在各链路之间均衡流量。(2)前端服务器层会运行等价路由协议,为各应用服务器配置一样的loopback地址,配置到达该地址的静态路由,并将下一跳指向不同应用服务器的物理口地址。且该前端服务器层至少包含2个前端调度服务器,至少两个前端调度服务器具备相同的地址;每个前端服务器的IP地址都是唯一的;通过ECMP将多个个体调度服务器串型,形成了新的架构的集群方式,便于调度服务器集群的横向扩展,解决了主备架构的前端调度服务器无法应对负载过大、浪费资源、扩展性差等问题。(3)RealServer层即应用服务器层为通常的网页服务器、日志服务器、应用服务器等,可以统称为应用服务器集群RS。

[0069] 进一步地,所述交换机,还可用于在接收用户端发送的请求数据包之前,获取与各所述前端调度服务器之间的通信链路的链路状态信息,并根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息,以及向各所述前端调度服务器广播更新后的链路状态信息;

[0070] 所述前端调度服务器,还可用于发送和接收所述链路状态信息。

[0071] 具体地,在交换机上运行路由协议,接收或广播与前端调度服务器的链路状态信息。并指定该交换机可以接收和广播链路状态信息。该交换机可以启用一个或多个虚拟接口,通过虚拟接口与前端调度服务器交互链路状态信息。通常情况下,一个前端调度服务器只能与指定的交换机进行链路状态信息的更新和交互。当指定的交换机更新链路状态信息之后,会再向其他的前端调度服务器发布更新后的链路状态信息,其他的前端调度服务器接收更新后的链路状态信息,以达到和交换机的路由信息保持一致的目的。

[0072] 进一步地,根据上述负载均衡系统的架构信息,并结合图3说明本发明实施例所述的负载均衡系统的具体工作流程。

[0073] (1)、当用户来访问调度服务器时,用户端的请求数据包经过互联网到达数据中心的机房交换机;

[0074] (2)、交换机检查数据包的目的地址,并查看ECMP路由表项,根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,并根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,同时获取该路径条目对应的下一跳出口,该下一跳出口即为所述前端调度服务器的网络端口,该网络端口具备发送和接收功能;然后将所述请求数据包发往所述选定的等价路径对应的前端调度服务器;当在ECMP路由表项中未找到有效路径时,将所述请求数据包直接丢弃。

[0075] 具体地,举一例子说明相同的下一跳地址的计算过程:例如,根据预设均衡条件确定请求数据包分配的相同的下一跳地址的计算过程可如图4所示。其中,通信链路L3的路由表项分别有10.10.10.0/24和20.20.20.0/24两个NH,都被hash计算为槽位3,并且槽位3被唯一指定下一条地址为1.1.1.1,因此,加上Hash偏移计算后,确定这两个地址都有相同的下一条地址NH:1.1.1.1为下一跳出口地址。(3)、当前端调度服务器收到请求数据包后,其根据预先设定的均衡方式将相应的请求数据包发往后端的应用服务器集群(RS)。

[0076] 其中,预先设定的均衡方式,可理解为实际上是取决于对ECMP槽位的定义,例如64

个槽位被平均分为8组,则每一组被分配到流量的概率相等。如果64个槽位被平均分为7组,则必定至少有一组分配流量的概率与别组不同,本发明实施例对此不作赘述。

[0077] 具体地,前端调度服务器层可以分为L4/L7出口,以IP/端口来辨识,与Realserver层节点相连。实际上是以二元组对(源ip和目的ip)参与hash计算,而L7则增加了源端口和目的端口参与hash计算,进而选取相应出口,本发明实施例对此不作赘述。(4)、应用服务器集群响应请求,将响应数据包发送给发出请求的用户端。

[0078] 当另外一个用户来访问调度服务器时,同理步骤(1)到(4)完成用户的访问响应。

[0079] 在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0080] 进一步地,所述交换机,还可用于将所述ECMP路由表项中各路径条目进行取模计算,并根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化,若确定任一条与所述前端调度服务器之间的通信链路断开后,则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变,并将断开的通信链路的流量重新进行哈希计算,分配到其他的通信链路上;若确定所述ECMP路由表项中通信链路对应的路径条目增多,则将活跃的通信链路上的部分流量分配至新增通信链路上。

[0081] 也就是说,本发明实施例所述的负载均衡系统可实现前端调度服务器的增减,并在前端调度服务器有增减变换的情况下,可实现其他通信链路不中断的连贯的负载均衡。且由交换机负责一致性Hash的问题,灵活性强,对整个架构而言包转发能力增强,而且在遍历方面具有更大的优势。

[0082] 在传统的数据中心实现负载均衡时,通常是通过ECMP和TOR(top of rack)互联,在TOR设备上通过ECMP将数据流均衡分发到负载均衡集群的各个成员机(即本实施例所述的前端调度服务器)。而现有动态路由协议在TOR和负载均衡集群之间生成ECMP路由表项,当ECMP路由表项中某条链路因故障失效后,动态路由协议会重新收敛。从TOR设备到负载均衡集群的流量会重新均衡,这样会打乱负载均衡集群成员机上原先维护的会话状态,整个集群需要重建会话,导致部分会话中断。

[0083] 而本发明所述的负载均衡系统,可在交换机层做一致性hash,解决一台服务器宕机后,会话会全部打乱的问题。具体地,一致性hash是基于已有的ECMP条目数进行取模计算的,当其中一条通信链路断开后,其他通信链路的取模不变,而将断开的那条通信链路的流量,重新hash到其他通信链路上,这样就能够维持其他通信链路的TCP连接保持不变。也就是说,本发明实施例所述的负载均衡系统中的一致性hash设计与通常的ECMP只能用于UDP链路不同,本发明实施例提出的负载均衡系统的架构还能适用于TCP链路。

[0084] 本发明实施例采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中,前端调度服务器与应用服务器的个数均大于1。在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能

更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题。且由交换机负责一致性Hash的问题,灵活性强,对整个架构而言包转发能力增强,而且在遍历方面具有更大的优势,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0085] 实施例二

[0086] 基于与本发明实施例一相同的发明构思,本发明实施例二提供了一种基于负载均衡系统的负载均衡方法,该负载均衡方法的具体实施可参见上述实施例一中的相关描述,重复之处不再赘述,其执行主体为交换机,流程示意图如图5所示,该方法包括:

[0087] 步骤501:交换机接收用户端发送的请求数据包,并根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址。

[0088] 其中,所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同。

[0089] 其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击。

[0090] 步骤502:根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目。

[0091] 步骤503:将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器,由该前端调度服务器将接收到的所述请求数据包发送至相应的所述应用服务器,经由所述应用服务器将所述请求数据包的响应结果返回给所述用户端。

[0092] 进一步地,所述在接收用户端发送的请求数据包之前,所述方法还可包括步骤A1-A2:

[0093] 步骤A1:获取与各所述前端调度服务器之间的通信链路的链路状态信息。

[0094] 步骤A2:根据所述链路状态信息更新所述ECMP路由表项中的各路径条目对应的链路信息,并向各所述前端调度服务器广播更新后的链路状态信息。

[0095] 进一步地,所述方法还可包括步骤B1-B2:

[0096] 步骤B1:将所述ECMP路由表项中各路径条目进行取模计算。

[0097] 步骤B2:根据更新后的所述链路状态信息确定是否有与所述前端调度服务器之间的通信链路的增减变化,若确定任一条与所述前端调度服务器之间的通信链路断开后,则保持所述ECMP路由表项中其他通信链路对应的所述路径条目的取模不变,并将断开的通信链路的流量重新进行哈希计算,分配到其他的通信链路上;若确定所述ECMP路由表项中通信链路对应的路径条目增多,则将活跃的通信链路上的部分流量分配至新增通信链路上。

[0098] 本发明实施例采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中,前端调度服务器与应用服务器的个数均大于1。在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,

同时还解决了传统主备LVS架构服务集群的负载过大的问题。且由交换机负责一致性Hash的问题,灵活性强,对整个架构而言包转发能力增强,而且在遍历方面具有更大的优势,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0099] 实施例三

[0100] 基于与本发明实施例一相同的发明构思,本发明实施例三提供了另一种基于负载均衡系统的负载均衡方法,该负载均衡方法的具体实施可参见上述实施例一中的相关描述,重复之处不再赘述,其执行主体为前端调度服务器,流程示意图如图6所示,该方法包括:

[0101] 步骤601:前端调度服务器接收所述交换机发送的来自用户端的请求数据包;其中,所述请求数据包是由所述交换机根据预设均衡条件确定所述请求数据包分配的相同的下一跳地址,根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器。

[0102] 其中,所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同。

[0103] 其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击。

[0104] 步骤602:将所述请求数据包发送至相应的所述应用服务器,以由所述应用服务器将所述请求数据包的响应结果返回给所述用户端。

[0105] 进一步地,所述方法还可包括:

[0106] 前端调度服务器发送和接收与所述交换机之间的链路状态信息。

[0107] 本发明实施例采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中,前端调度服务器与应用服务器的个数均大于1。在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题。且由交换机负责一致性Hash的问题,灵活性强,对整个架构而言包转发能力增强,而且在遍历方面具有更大的优势,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0108] 实施例四

[0109] 基于与本发明实施例一相同的发明构思,本发明实施例四提供了另一种基于负载均衡系统的负载均衡方法,该负载均衡方法的具体实施可参见上述实施例一中的相关描述,重复之处不再赘述,其执行主体为应用服务器,该方法包括:

[0110] 应用服务器响应所述前端调度服务器发送的来自所述用户端的所述请求数据包,并将相应的响应结果返回给所述用户端。

[0111] 其中,所述请求数据包是由所述交换机将接收到的来自用户端的所述请求数据包

根据预设均衡条件确定数据包分配的相同的下一跳地址,并根据确定的所述下一跳地址,从等价路由ECMP路由表项中选取与所述下一跳地址相对应的路径条目后,并将所述请求数据包按照所选的所述路径条目对应的通信链路发送至相应的所述前端调度服务器所得到的。

[0112] 所述预设均衡条件:芯片计算的哈希值相同、ECMP路由表项中各路径条目的下一跳个数的总数相同、相同偏移的下一跳存储的出口信息相同。

[0113] 其中,所述交换机接收用户端发送的请求数据包,这一步骤,还用于抵抗SYN Flood攻击。

[0114] 本发明实施例采用基于交换机的负载均衡架构,根据预设均衡条件确定来自用户端的请求数据包分配的相同的下一跳地址,并将交换机接收到的请求数据包根据确定的相同的下一跳地址发送至前端调度服务器,再转发至应用服务器,从而实现了负载均衡,其中,前端调度服务器与应用服务器的个数均大于1。在本发明实施例所述的负载均衡系统中,由于交换机转发为纯硬件转发,端口均为线速,所以交换机做负载均衡相较于LVS性能更好,并且避免了故障切换时各种复杂的主备联系机制,提高了资源的利用率,扩展性强,同时还解决了传统主备LVS架构服务集群的负载过大的问题。且由交换机负责一致性Hash的问题,灵活性强,对整个架构而言包转发能力增强,而且在遍历方面具有更大的优势,另外,由于本系统是由交换机接收用户端的访问请求,因此还具有能够抵御大规模的SYN Flood攻击的效果。

[0115] 本领域技术人员应明白,本发明的实施例可提供为方法、装置(设备)、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0116] 本发明是参照根据本发明实施例的方法、装置(设备)和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0117] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0118] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0119] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优

选实施例以及落入本发明范围的所有变更和修改。

[0120] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

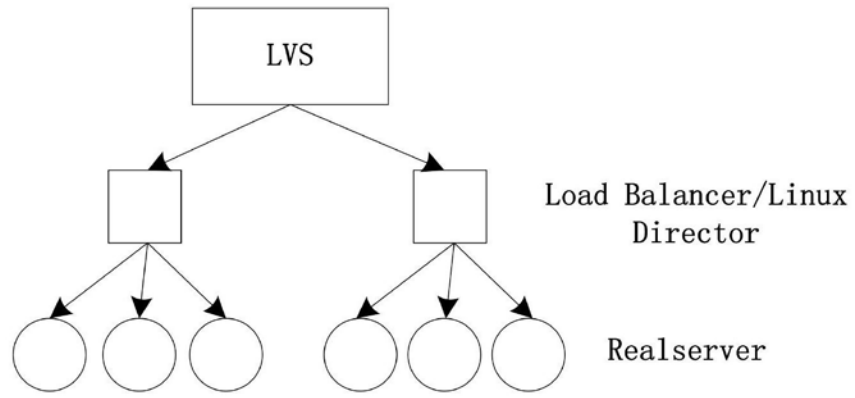


图1

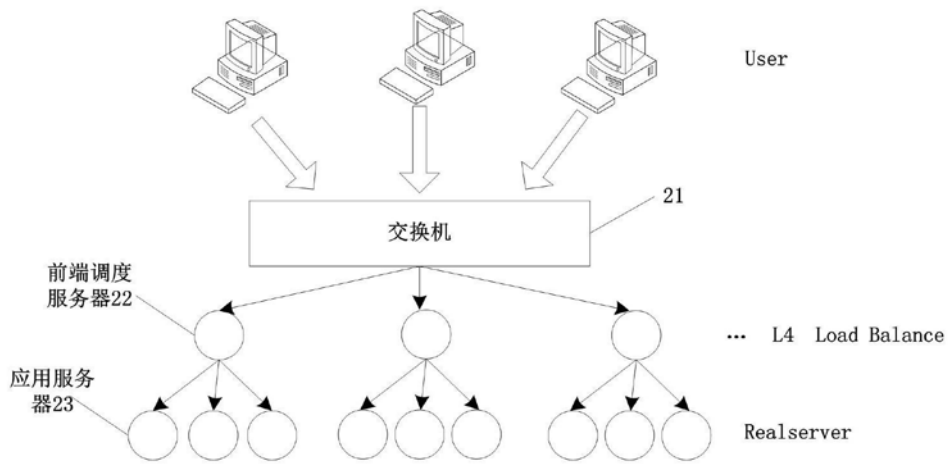


图2

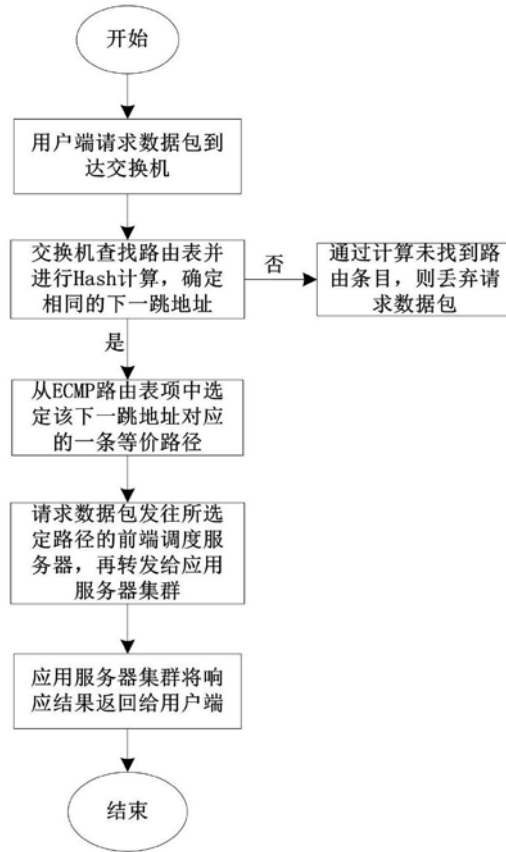


图3

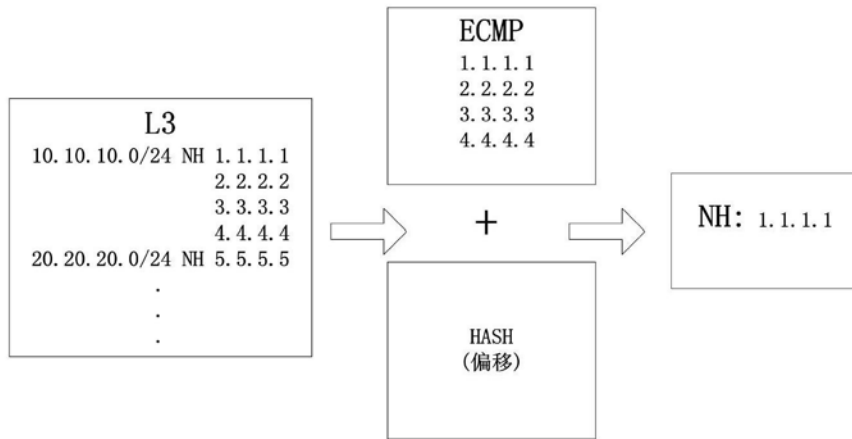


图4

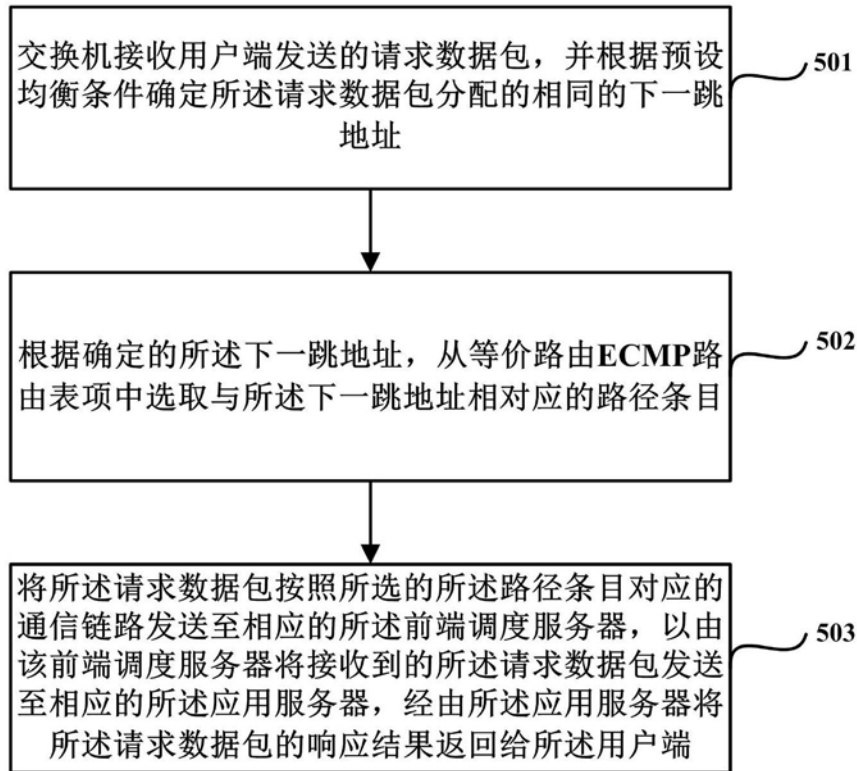


图5

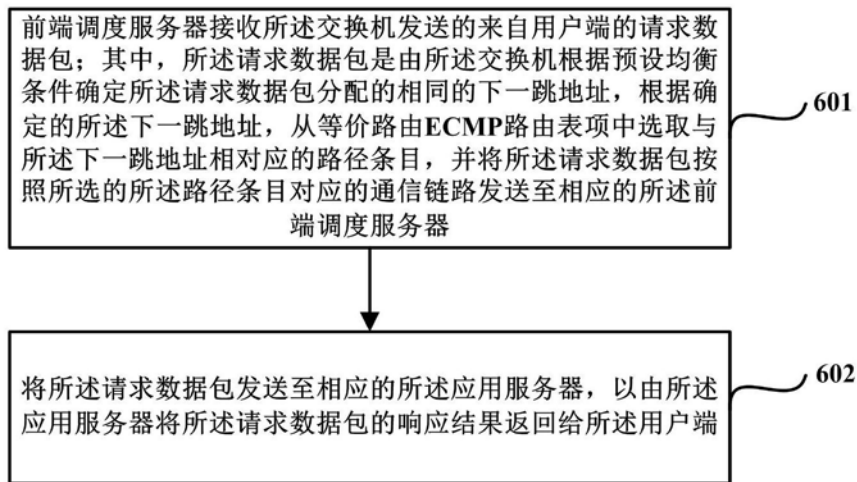


图6