

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.



[12] 发明专利申请公布说明书

H04L 12/56 (2006.01)

H04L 12/28 (2006.01)

H04L 29/06 (2006.01)

G06F 15/16 (2006.01)

[21] 申请号 200810004223.X

[43] 公开日 2008年7月9日

[11] 公开号 CN 101217483A

[22] 申请日 2008.1.21

[21] 申请号 200810004223.X

[71] 申请人 中兴通讯股份有限公司

地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦

[72] 发明人 米丽娅 戴孟晋 方 马

[74] 专利代理机构 北京康信知识产权代理有限责任公司

代理人 尚志峰 吴孟秋

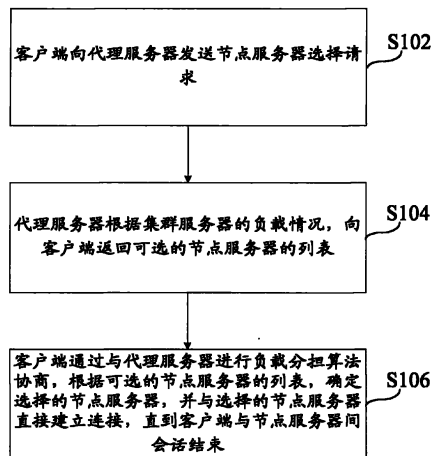
权利要求书 2 页 说明书 6 页 附图 3 页

[54] 发明名称

用于实现集群服务器内负载分担代理的方法

[57] 摘要

本发明公开了用于实现集群服务器内负载分担代理的方法，包括以下步骤：客户端向代理服务器发送节点服务器选择请求；代理服务器根据集群服务器的负载情况，向客户端返回可选的节点服务器的列表；以及客户端通过与代理服务器进行负载分担算法协商，根据可选的节点服务器的列表，确定选择的节点服务器，并与选择的节点服务器直接建立连接，直到客户端与节点服务器间会话结束。通过本发明，由于采取了最优服务器节点和次优服务器节点均提供给客户端的负载分担代理方法，增强了集群服务器的对外稳定性，且代理服务器不转发客户端和服务器间的应用数据，代理服务器不容易出现网络瓶颈。



1. 一种用于实现集群服务器内负载分担代理的方法，其特征在于，包括以下步骤：

客户端向代理服务器发送节点服务器选择请求；

所述代理服务器根据所述集群服务器的负载情况，向所述客户端返回可选的节点服务器的列表；以及

所述客户端通过与所述代理服务器进行负载分担算法协商，根据所述可选的节点服务器的列表，确定选择的节点服务器，并与所述选择的节点服务器直接建立连接，直到所述客户端与所述节点服务器间会话结束。

2. 根据权利要求1所述的方法，其特征在于，所述可选的节点服务器的列表中的节点服务器分成与所述客户端相对应的一个最优节点服务器和一个或多个次优节点服务器。
3. 根据权利要求2所述的方法，其特征在于，所述代理服务器根据所述客户端的所述请求和流量进行计算，以分析各个所述节点服务器的负载情况，找到所述最优节点服务器。
4. 根据权利要求2所述的方法，其特征在于，所述客户端首先与所述最优节点服务器连接，如果与所述最优节点服务器的连接出现异常，则所述客户端与所述一个或多个次优节点服务器中的一个连接。

-
5. 根据权利要求4所述的方法,其特征在于,所述客户端自主选择所述一个或多个次优节点服务器中的一个进行连接。
 6. 根据权利要求1所述的方法,其特征在于,所述客户端根据所述客户端的情况决定所述代理服务器使用的所述负载分担算法。

用于实现集群服务器内负载分担代理的方法

技术领域

本发明涉及计算机应用以及通讯领域，更具体地，涉及用于实现集群服务器内负载分担代理的方法。

背景技术

Internet 的快速增长使各种网络服务器，通讯服务器，面对的访问者数量快速增加，服务器变得不胜负荷，如果无法及时处理大量的用户服务请求，那么服务器将出现中断的情况。以往在解决这些问题的方法是采用更强计算能力的服务器来替换原来的服务器，即使这样，单台服务器的负载能力也是有限的，不可能无限扩展，同时，高档服务器的价格是随着服务器的性能呈现指数型上升，因此，采用多台廉价服务器组成负载分担的系统模型日渐成为主流，从而扩展出了负载分担技术。负载分担系统主要是将集中在一台服务器上的用户服务请求分发到多台服务器上。在负载分担方式出现的初期，有不少网络的设计采用域名轮转的方式，即是一个域名对应多台服务器，作为一种廉价的方案，域名轮转的方式可以解决一些服务器的负载问题，但是，由于这种负载分担的方式有很大的局限性：无法根据各台服务器的负载情况，将用户服务请求发送到不同的服务器上；在其中一台服务器出现问题无法工作的时候，系统仍然会将用户访问请求发送到出现故障的服务器上，造成一部分服务的中断；由于域名解释一般在各地的服务器上都会有 Cache 存在，因此会造成一个地区的用户访问请求将集中在同一台服务器上。因而实际上，采用域名轮转的方式来做系统负载分担，其效果并不明显。

因此，针对不同使用对象，负载分担技术大概分为四种模式：DNS、代理服务器、地址转换网关以及协议内部支持，高级服务器应用中代理服务器应用得最为广泛，因此这里重点分析代理服务器的现有技术背景。

使用代理服务器，可以将请求转发给内部的服务器，使用这种加速模式显然可以提升访问速度，代理服务器将请求均匀转发给多台服务器，从而达到负载均衡的目的。这种代理方式与普通的代理方式有所不同，标准代理方式是客户使用代理访问多个外部服务器，而这种代理方式是代理多个客户访问内部服务器，因此也被称为反向代理模式。虽然实现这个任务并不算是特别复杂，然而由于要求特别高的效率，实现起来并不简单。使用反向代理的好处是，可以将负载均衡和代理服务器的高速缓存技术结合在一起，提供有益的性能。这种代理服务器本身虽然可以达到很高效率，但是针对每一次代理，代理服务器就必须维护两个连接，一个对外的连接，一个对内的连接，因此特别高的连接请求，代理服务器的负载也就非常之大。反向代理方式下能应用优化的负载均衡策略，每次访问最空闲的内部服务器来提供服务。但是随着并发连接数量的增加，代理服务器本身的负载也变得非常大，最后反向代理服务器本身会成为服务的瓶颈。

而现在更为先进的代理服务器加入了許多算法，如流量负载分担方法。这种代理服务器会在每次发送数据包给服务器之前，预先计算出每条链路通道发送数据包所需的时间；然后选择出发送时间最短的通道，用该通道发送当前数据包。这种流量负载分担方法根据通道的繁忙程度动态选择链路，可应用于与数据通讯设备不同层次、不同应用的流量负载分担，但这种流量负载分担也有不足之处，如：客户端对服务器的请求数据不可能每次都相同，经过代理服务器找到的服务器每次都不一样，这样根本就保证不了客户端与服务器会话的一致性，而且，由于客户端每次请求服务器，代理服务器

都要为之进行流量计算，这样即浪费了很多的时间，又浪费了大量的代理服务器资源，从而造成代理服务器的负荷较大。

发明内容

考虑到上述问题而做出本发明，为此，本发明的主要目的在于，提供一种用于实现集群服务器内负载分担代理的方法。

根据本发明的用于实现集群服务器内负载分担代理的方法，包括以下步骤：客户端向代理服务器发送节点服务器选择请求；代理服务器根据集群服务器的负载情况，向客户端返回可选的节点服务器的列表；以及客户端通过与代理服务器进行负载分担算法协商，根据可选的节点服务器的列表，确定选择的节点服务器，并与选择的节点服务器直接建立连接，直到客户端与节点服务器间会话结束。

此外，在该方法中，可选的节点服务器的列表中的节点服务器分成与客户端相对应的一个最优节点服务器和一个或多个次优节点服务器。

此外，在该方法中，代理服务器根据客户端的请求和流量进行计算，以分析各个节点服务器的负载情况，找到最优节点服务器。

此外，在该方法中，客户端首先与最优节点服务器连接，如果与最优节点服务器的连接出现异常，则客户端与一个或多个次优节点服务器中的一个连接。

此外，在该方法中，客户端自主选择一个或多个次优节点服务器中的一个进行连接。而同时客户端根据客户端的情况决定代理服务器使用的负载分担算法。

通过本发明，由于采取了最优服务器节点和次优服务器节点均提供给客户端的负载分担代理方法，增强了集群服务器的对外稳定性，且代理服务器不转发客户端和服务器间的应用数据，代理服务器不容易出现网络瓶颈。客户端通过与代理服务器间进行负载分担算法协商机制，充分提高了选择的灵活性；最后客户端和服务器间一次会话的数据交换均同一个服务器节点间，保证了客户端与服务器间同一会话数据的一致性。

本发明的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

附图说明

附图用来提供对本发明的进一步理解，并且构成说明书的一部分，与本发明的实施例一起用于解释本发明，并不构成对本发明的限制。在附图中：

图 1 是示出了根据本发明实施例的用于实现集群服务器内负载分担代理的方法的流程图；

图 2 是示出了根据本发明实施例的客户端、代理服务器、集群服务器之间的关系图；

图 3 是示出了根据本发明实施例的负载分担代理的正常流程图；以及

图 4 是示出了根据本发明实施例的负载分担代理的异常流程图。

具体实施方式

以下结合附图对本发明的优选实施例进行说明，应当理解，此处所描述的优选实施例仅用于说明和解释本发明，并不用于限定本发明。

本发明是为了解决在使用代理服务器的情况下，提高客户端、代理服务器、集群服务器之间的可靠性、协商性、灵活性和会话一致性。

在集群服务环境内，每个节点服务器均具备相同的功能，客户端通过代理服务器与集群环境内的节点服务器进行连接。代理服务器根据客户端的请求和流量进行计算，分析各节点服务器的负载情况，找到最为合适的节点服务器，此时，节点服务器、代理服务器、客户端进行协商，只要客户端不主动断开与节点服务器的连接，该客户端就可以不通过代理服务器，直接与该节点服务器进行通讯。

图 1 是示出了根据本发明实施例的用于实现集群服务器内负载分担代理的方法的流程图。

如图 1 所示，包括以下步骤：步骤 S102，客户端向代理服务器发送节点服务器选择请求；步骤 S104，代理服务器根据集群服务器的负载情况，向客户端返回可选的节点服务器的列表；以及步骤 S106，客户端通过与代理服务器进行负载分担算法协商，根据可选的节点服务器的列表，确定选择的节点服务器，并与选择的节点服务器直接建立连接，直到客户端与节点服务器间会话结束。

此外，在该方法中，可选的节点服务器的列表中的节点服务器分成与客户端相对应的一个最优节点服务器和一个或多个次优节点服务器。

此外，在该方法中，代理服务器根据客户端的请求和流量进行计算，以分析各个节点服务器的负载情况，找到最优节点服务器。

此外，在该方法中，客户端首先与最优节点服务器连接，如果与最优节点服务器的连接出现异常，则客户端与一个或多个次优节点服务器中的一个连接。

此外，在该方法中，客户端自主选择一个或多个次优节点服务器中的一个进行连接。而同时客户端根据客户端的情况决定代理服务器使用的负载分担算法。

图2是示出了根据本发明实施例的客户端、代理服务器、集群服务器之间的关系图。

如图2所示，客户端通过代理服务器获取提供服务的服务器节点，并与节点服务器间直接进行连接。

图3是示出了根据本发明实施例的负载分担代理的正常流程图。以及图4是示出了根据本发明实施例的负载分担代理的异常流程图。

结合图3和图4进行详细描述，客户端发起连接请求，代理服务器分析集群内部节点服务器的负载情况，通知客户端可以连接的服务器节点，客户端获得连接许可，与节点服务器建立连接并进行通信。如果服务器节点不可用，或服务器节点中途出现异常，整个工作流程又重新从第三步开始。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。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。

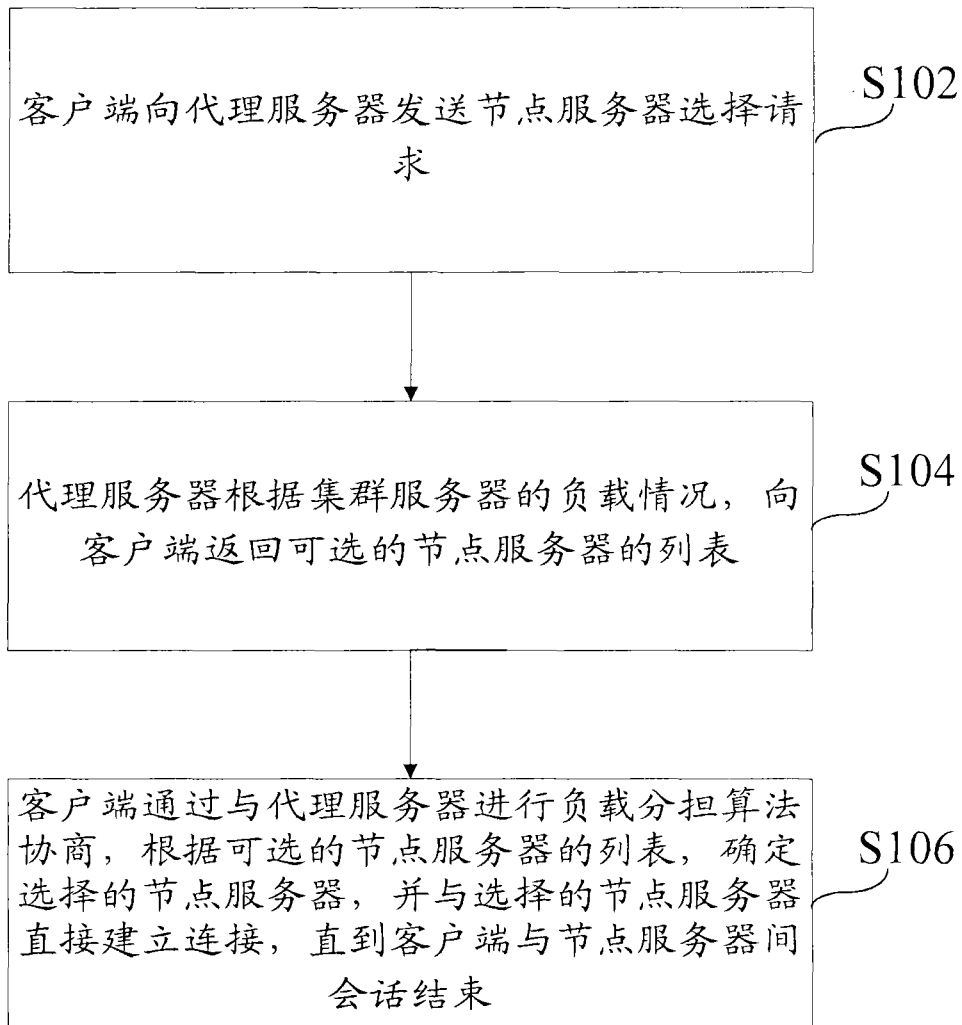


图 1

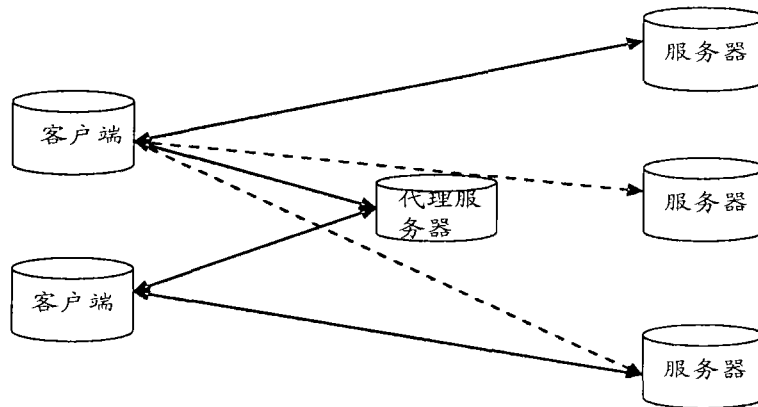


图 2

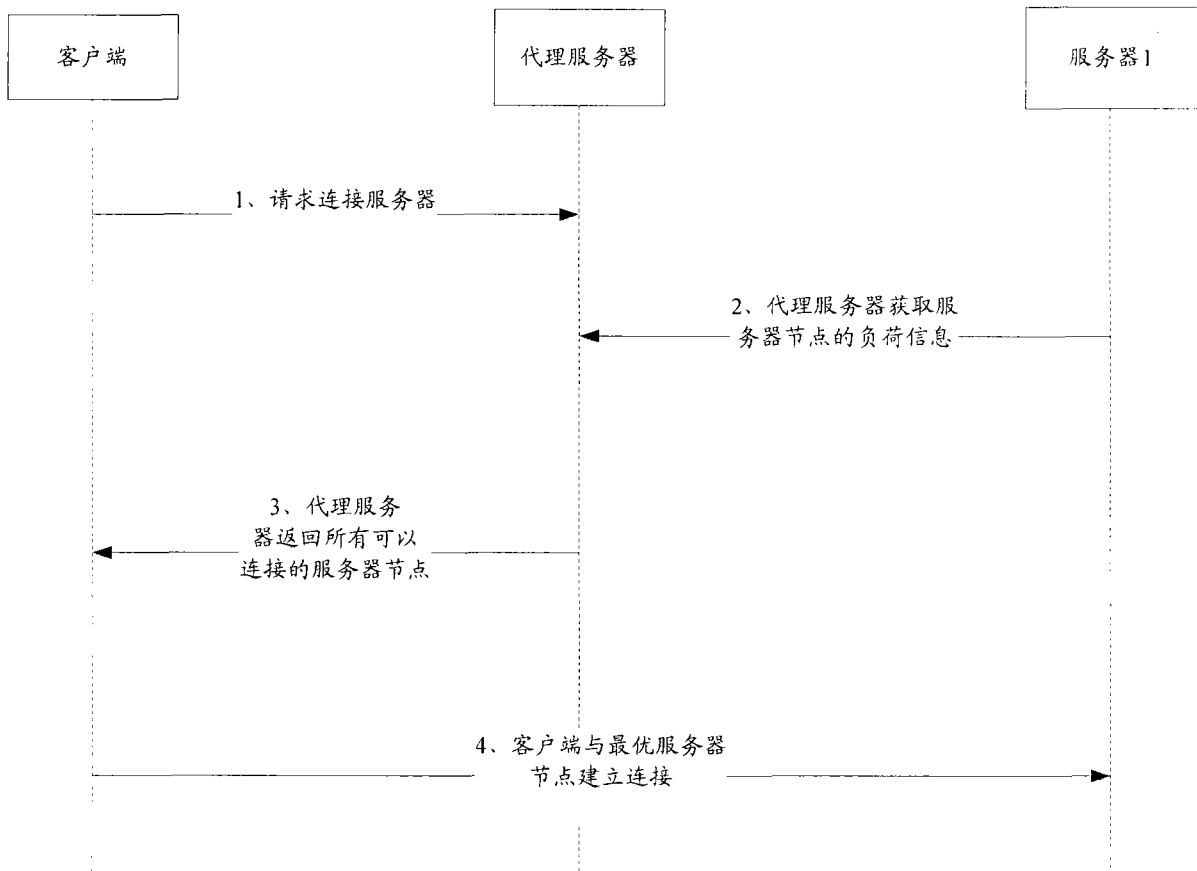


图 3

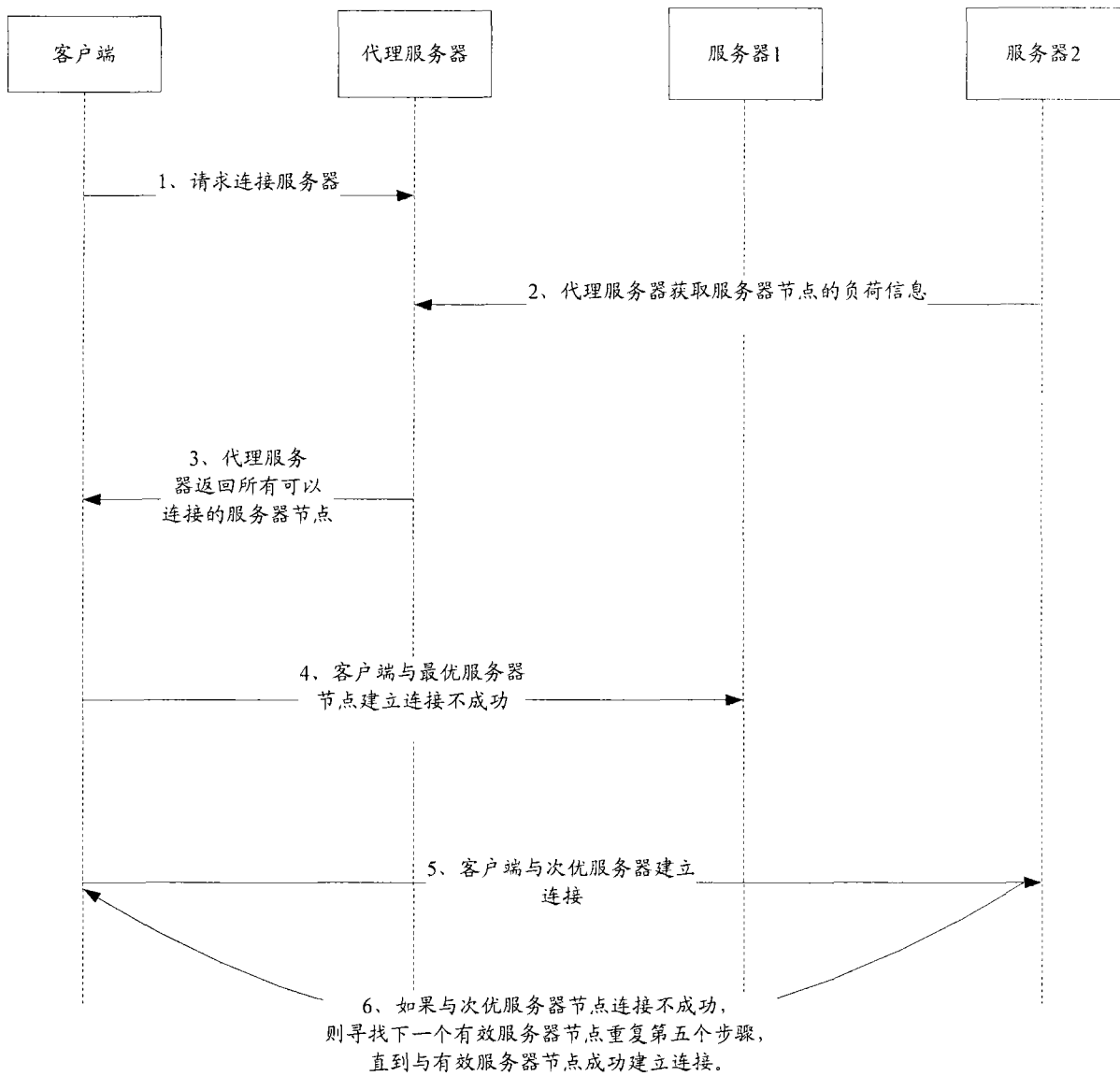


图 4