



(12) 发明专利申请

(10) 申请公布号 CN 112948120 A

(43) 申请公布日 2021.06.11

(21) 申请号 202110291302.9

(22) 申请日 2021.03.18

(71) 申请人 广东好太太智能家居有限公司
地址 511434 广东省广州市番禺区化龙镇
石化路21号之一、之二

(72) 发明人 王妙玉 王伦 吴宁泉

(74) 专利代理机构 广州市越秀区哲力专利商标
事务所(普通合伙) 44288
代理人 成婵娟

(51) Int. Cl.
G06F 9/50 (2006.01)
H04L 29/08 (2006.01)

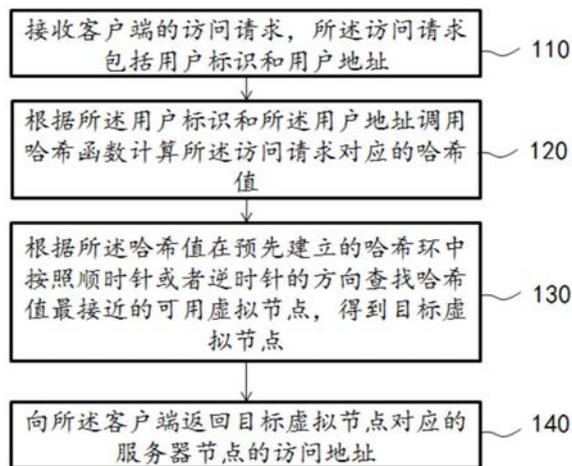
权利要求书2页 说明书6页 附图2页

(54) 发明名称

负载均衡方法、系统、装置和存储介质

(57) 摘要

本申请公开了一种负载均衡方法、系统、装置和存储介质,涉及互联网技术,方法包括以下步骤:根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。本方案可以实现负载均衡,减少雪崩效应的可能性。



1. 一种负载均衡方法,其特征在于,包括以下步骤:

接收客户端的访问请求,所述访问请求包括用户标识和用户地址;

根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;

根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;

向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

2. 根据权利要求1所述的负载均衡方法,其特征在于,所述根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点,包括:

根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向依次查找所述虚拟节点,查找到所述虚拟节点时判断当前的所述虚拟节点对应的所述服务器节点是否可用,若可用,则停止查找,并将当前的所述虚拟节点作为所述目标虚拟节点,若不可用,则继续查找下一个虚拟节点直到找到对应的服务器节点可用的所述虚拟节点。

3. 根据权利要求1所述的负载均衡方法,其特征在于,所述哈希环中相邻的两个虚拟节点所对应的服务器节点不相同。

4. 根据权利要求1所述的负载均衡方法,其特征在于,所述方法还包括以下步骤:

获取多个所述服务器节点的负载状态,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量。

5. 根据权利要求4所述的负载均衡方法,其特征在于,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量,包括:

确定多个所述服务器节点对应的负荷率中的最大值和最小值;

当所述最大值和最小值之差大于预设值时,向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,所述临时虚拟节点指向所述最小值对应的服务器节点。

6. 根据权利要求5所述的负载均衡方法,其特征在于,当按照顺时针方向查找所述虚拟节点时,所述向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,具体是:

向所述最大值对应的服务器节点所对应一部分所述虚拟节点的右侧插入临时虚拟节点。

7. 根据权利要求6所述的负载均衡方法,其特征在于,所述方法还包括以下步骤:

按照预设的周期重建哈希环,其中,重建哈希环时根据当前可用的服务器节点生成所述虚拟节点。

8. 一种负载均衡系统,其特征在于,包括:

接收模块,用于接收客户端的访问请求,所述访问请求包括用户标识和用户地址;

哈希值计算模块,用于根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;

节点确定单元,用于根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的

方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;

发送模块,用于向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

9. 一种负载均衡装置,其特征在于,包括:

程序;

存储器,用于存储所述程序;

处理器,用于加载所述程序以执行如权利要求1-7任一项所述的负载均衡方法。

10. 一种存储介质,其存储有程序,其特征在于,所述程序被处理器执行时实现如权利要求1-7任一项所述的负载均衡方法。

负载均衡方法、系统、装置和存储介质

技术领域

[0001] 本申请涉及互联网技术,特别是一种负载均衡方法、系统、装置和存储介质。

背景技术

[0002] 在互联网领域中,随着产品项目的发展,后台系统需要通过水平扩展来增加系统的吞吐量、数据处理能力,这时候就需要对流量做负载均衡,简单点说,负载均衡意思就是将流量分摊到具有相同功能部署在不同服务器上功能单元。通常可以用硬件负载,这类负载效果比较好。另外,可以通过引入负载均衡应用,典型的有nginx,它可以将访问同一地址的流量根据一定的负载算法转发到不同的服务器上。

[0003] 但是对于上述方案而言,成本相对较高。因此,亦有方案通过哈希环进行服务器节点查找,从而实现负载均衡。这一方式通过计算访问请求的哈希值,并在哈希环上寻找哈希值最接近的服务器节点作为访问对象。然而,当一个服务器崩溃时,其访问量会进入下一个节点的服务器,从而导致下一个节点的服务器因为负荷过大崩溃,接着两个崩溃的服务器的访问量一起进入下一个节点,引起连锁反应,导致所有的服务器崩溃。

发明内容

[0004] 本发明旨在至少解决现有技术中存在的技术问题之一。为此,本发明提出一种负载均衡方法、系统、装置和存储介质,以通过较低的成本实现负载均衡。

[0005] 一方面,本申请实施例提供了:

[0006] 一种负载均衡方法,包括以下步骤:

[0007] 接收客户端的访问请求,所述访问请求包括用户标识和用户地址;

[0008] 根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;

[0009] 根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;

[0010] 向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

[0011] 在一些实施例中,所述根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点,包括:

[0012] 根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向依次查找所述虚拟节点,查找到所述虚拟节点时判断当前的所述虚拟节点对应的所述服务器节点是否可用,若可用,则停止查找,并将当前的所述虚拟节点作为目标虚拟节点,若不可用,则继续查找下一个虚拟节点直到找到对应的服务器节点可用的所述虚拟节点。

[0013] 在一些实施例中,所述哈希环中相邻的两个虚拟节点所对应的服务器节点不相同。

[0014] 在一些实施例中,所述方法还包括以下步骤:

[0015] 获取多个所述服务器节点的负载状态,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量。

[0016] 在一些实施例中,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量,包括:

[0017] 确定多个所述服务器节点对应的负荷率中的最大值和最小值;

[0018] 当所述最大值和最小值之差大于预设值时,向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,所述临时虚拟节点指向所述最小值对应的服务器节点。

[0019] 在一些实施例中,当按照顺时针方向查找所述虚拟节点时,所述向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,具体是:

[0020] 向所述最大值对应的服务器节点所对应一部分所述虚拟节点的右侧插入临时虚拟节点。

[0021] 在一些实施例中,所述方法还包括以下步骤:

[0022] 按照预设的周期重建哈希环,其中,重建哈希环时根据当前可用的服务器节点生成所述虚拟节点。

[0023] 另一方面,本申请实施例提供了:

[0024] 一种负载均衡系统,包括:

[0025] 接收模块,用于接收客户端的访问请求,所述访问请求包括用户标识和用户地址;

[0026] 哈希值计算模块,用于根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;

[0027] 节点确定单元,用于根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;

[0028] 发送模块,用于向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

[0029] 另一方面,本申请实施例提供了:

[0030] 一种负载均衡装置,包括:

[0031] 程序;

[0032] 存储器,用于存储所述程序;

[0033] 处理器,用于加载所述程序以执行所述的负载均衡方法。

[0034] 另一方面,本申请实施例提供了:

[0035] 一种存储介质,其存储有程序,所述程序被处理器执行时实现所述的负载均衡方法。

[0036] 本申请实施例通过在哈希环中设置各服务器节点对应的多个虚拟节点,并通过将虚拟节点随机分布在上哈希环上来实现负载均衡,由于每个服务器节点对应多个虚拟节点,并且这些虚拟节点并不是顺序分配,因此即便某个服务器节点宕机,也不会导致该服务器节点的访问量全部涌入另一个服务器节点,而是相对平均地分摊到各服务器节点,可以降低雪崩状态的发生可能性。

附图说明

[0037] 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做一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,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。

[0038] 图1是本申请实施例提供的一种负载均衡方法的流程图;

[0039] 图2是本申请实施例提供的一种负载均衡方法所使用的哈希环的示意图;

[0040] 图3是本申请实施例提供的一种哈希环增加临时虚拟节点的示意图;

[0041] 图4是本申请实施例提供的一种现有的哈希环的示意图。

具体实施方式

[0042] 为使本发明的目的、技术方案和优点更加清楚,以下将参照本发明实施例中的附图,通过实施方式清楚、完整地描述本发明的技术方案,显然,所描述的实施例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他实施例,都属于本发明保护的范围。

[0043] 在本发明的描述中,若干的含义是一个以上,多个的含义是两个以上,大于、小于、超过等理解为不包括本数,以上、以下、以内等理解为包括本数。如果有描述到第一、第二只是用于区分技术特征为目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量或者隐含指明所指示的技术特征的先后关系。

[0044] 本发明的描述中,除非另有明确的限定,设置等词语应做广义理解,所属技术领域技术人员可以结合技术方案的具体内容合理确定上述词语在本发明中的具体含义。

[0045] 本发明的描述中,参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征或者特点可以在任何一个或多个实施例或示例中以合适的方式结合。

[0046] 参照图4,本实施例将对现有技术中存在的雪崩效应的原因进行分析,现有技术中,假设有ABCD四个服务器节点(每个节点有对应的哈希值),其分别对应于哈希环中的四个节点,在通过哈希值查找服务器节点时按照顺时针的方向。假设服务器节点A发生宕机失效,原本对应于服务器节点A的访问,将会依次进入到服务器节点B,假设每个服务器的负荷能力为M,若服务器节点A和服务器节点B的访问量均超过M/2时,服务器节点A崩溃时,涌入服务器节点B的访问量会让服务器节点B也崩溃,接着服务器节点A和服务器节点B的访问量一起涌入服务器节点C,节点C直接发生崩溃,同理,节点D也会被牵连而导致崩溃。

[0047] 参照图1,本实施例公开了一种负载均衡方法,可以应用在负载均衡服务器,本方法包括以下步骤:

[0048] 步骤110、接收客户端的访问请求,所述访问请求包括用户标识和用户地址。其中用户标识可以是用户账号,用户名称等等,用户地址通常是IP地址,在部分实施例中也可以是物理地址。

[0049] 步骤120、根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对

应的哈希值。在本步骤中,实际上是将用户标识和用户地址作为函数的输入值,然后利用设定的哈希函数计算访问请求对应的哈希值。

[0050] 步骤130、根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上。需要理解的是,本实施例所指的随机可以是完全随机,也可以是局部随机。例如,有ABCD四个服务器节点,每个服务器节点对应4个虚拟节点,那么可以将哈希环分成五段,每段中按照随机的顺序一共设置四个不同服务器节点对应的虚拟节点。可以理解的是,当每个服务器节点所对应的虚拟节点的数量足够多的时候,将这些虚拟节点随机分布,是可以达到均匀的效果的。其中,虚拟节点的数量可以设置为100个、500个或者更多。哈希环的数值范围可以是 $0\sim 2^{32}$ 。

[0051] 参照图2,图2以ABCD四个服务器节点,每个服务器节点对应4个虚拟节点为例,每个虚拟节点之间可以设置为等距离的。在部分例子中,按照顺时针的方式查找最接近的可用节点,例如,访问请求的哈希值处于虚拟节点C1和虚拟节点D4之间时,首先会将虚拟节点D4作为最接近的虚拟节点,而虚拟节点D4对应于服务器节点D,当服务器节点D宕机时是不能用的。因此,需要继续按照顺时针的方向查找下一个虚拟节点B4,因此,原本访问服务器节点D的部分访问量进入到服务器B,同理,原本对应到虚拟节点D2的访问量会访问服务器节点C,原来对应于虚拟节点D3的访问量会访问服务器节点A。可见,在服务器节点D宕机的情况下,服务器节点D的访问量会分配到其他不同的服务器节点中,而不是集中到一个服务器节点中。

[0052] 步骤140、向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

[0053] 本步骤向客户端返回服务器节点的访问地址,客户端可以基于该访问地址访问对应的服务器。

[0054] 在一些实施例中,所述根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点,包括:

[0055] 根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时针的方向依次查找所述虚拟节点,查找到所述虚拟节点时判断当前的所述虚拟节点对应的所述服务器节点是否可用,若可用,则停止查找,并将当前的所述虚拟节点作为目标虚拟节点,若不可用,则继续查找下一个虚拟节点直到找到对应的服务器节点可用的所述虚拟节点。

[0056] 通常地,如图2所示,当查找到一个虚拟节点对应的服务器节点不可用时,会自动寻找下一个虚拟节点,直到找到一个虚拟节点对应的服务器节点可用为止,例如,访问请求的哈希值落入虚拟节点C1和虚拟节点D4之间,在查找虚拟节点时,按照顺时针方向首先查找到虚拟节点D4,判断服务器节点D不可用后,继续查找到虚拟节点B4,由于服务器节点B可用,则停止查找其他虚拟节点,并将虚拟节点B4对应的服务器节点B的地址返回到终端。

[0057] 在一些实施例中,所述哈希环中相邻的两个虚拟节点所对应的服务器节点不相同。在一些实施例中,为了避免发生宕机时,出现相邻节点指向宕机服务器而导致查找虚拟节点时增加一次计算,可以在生成哈希环时将相邻的两个虚拟节点对应的服务器设置为不同。

[0058] 在一些实施例中,所述方法还包括以下步骤:

[0059] 获取多个所述服务器节点的负载状态,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量。

[0060] 事实上,虽然通过哈希环可以均匀访问的访问量,但是不同访问请求对应的业务实际上不同,对服务器节点造成的压力也不同,在极端情况下,即便通过哈希环进行负载均衡,也会有部分服务器节点明显高于另一些服务器节点,故当出现这些极端情况(服务器节点之间负荷率差异较大,且负荷率最大的服务器处于较高负荷率)时,需要进行调整,以避免有部分服务器节点宕机。例如,服务器A的负荷率(当前负荷量/最大负荷量之比)是80%,而服务器B的负荷率只有60%,其他服务器则处于70%左右。此时,可以通过增加服务器节点B对应的虚拟节点,减少服务器节点A对应的虚拟节点,从而增加服务器B的负荷量,降低服务器A的负荷量,这一过程只需要增减哈希环上的虚拟节点,而无需改变查找虚拟节点的程序。

[0061] 在一些实施例中,根据各所述服务器节点的负载状态调整所述哈希环中服务器节点所对应的所述虚拟节点的数量,包括:

[0062] 确定多个所述服务器节点对应的负荷率中的最大值和最小值;

[0063] 当所述最大值和最小值之差大于预设值时,向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,所述临时虚拟节点指向所述最小值对应的服务器节点。可以理解的是,虚拟节点的旁边是指该虚拟节点与相邻的虚拟节点之间的数值范围。

[0064] 在本实施例中,例如,负荷率最大的服务器节点A的负荷率是80%,负荷最小的服务器节点B的负荷率是60%,参照图3,通过在虚拟节点A3的右侧插入临时虚拟节点B6(指向服务器节点B),可以将原本哈希值落入虚拟节点D3和临时虚拟节点B6之间的访问量,从原本访问服务器节点A变为访问服务器节点B,从而实现负荷量的转移。

[0065] 在一些实施例中,当按照顺时针方向查找所述虚拟节点时,所述向所述最大值对应的服务器节点所对应一部分所述虚拟节点的旁边插入临时虚拟节点,具体是:

[0066] 向所述最大值对应的服务器节点所对应一部分所述虚拟节点的右侧插入临时虚拟节点。

[0067] 从上述例子可知,当查找是按照逆时针方向时,临时节点插入的位置位于需要分流的服务器节点的左侧。

[0068] 在一些实施例中,所述方法还包括以下步骤:

[0069] 按照预设的周期重建哈希环,其中,重建哈希环时根据当前可用的服务器节点生成所述虚拟节点。

[0070] 在该实施例中,当部分服务器节点永久退出或者在较长的时间内处于维修状态时,按照一定的周期进行哈希环的重建,有利于排除无用服务器节点,或者更新最近加入的服务器节点。

[0071] 本实施例公开了一种负载均衡系统,包括:

[0072] 接收模块,用于接收客户端的访问请求,所述访问请求包括用户标识和用户地址;

[0073] 哈希值计算模块,用于根据所述用户标识和所述用户地址调用哈希函数计算所述访问请求对应的哈希值;

[0074] 节点确定单元,用于根据所述哈希值在预先建立的哈希环中按照顺时针或者逆时

针的方向查找哈希值最接近的虚拟节点,得到目标虚拟节点;其中,所述哈希环中包括多个所述虚拟节点,每个所述虚拟节点对应一个服务器节点,所述服务器节点的数量有多个,每个服务器节点对应多个虚拟节点,多个所述虚拟节点随机分布在所述哈希环上;

[0075] 发送模块,用于向所述客户端返回目标虚拟节点对应的服务器节点的访问地址。

[0076] 本实施例公开了一种负载均衡装置,包括:

[0077] 程序;

[0078] 存储器,用于存储所述程序;

[0079] 处理器,用于加载所述程序以执行所述的负载均衡方法。

[0080] 本实施例公开了一种存储介质,其存储有程序,所述程序被处理器执行时实现所述的负载均衡方法。

[0081] 注意,上述仅为本申请的较佳实施例及所运用技术原理。本领域技术人员会理解,本申请不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本申请的保护范围。因此,虽然通过以上实施例对本申请进行了较为详细的说明,但是本申请不仅仅限于以上实施例,在不脱离本申请构思的情况下,还可以包括更多其他等效实施例,而本申请的范围由所附的权利要求范围决定。

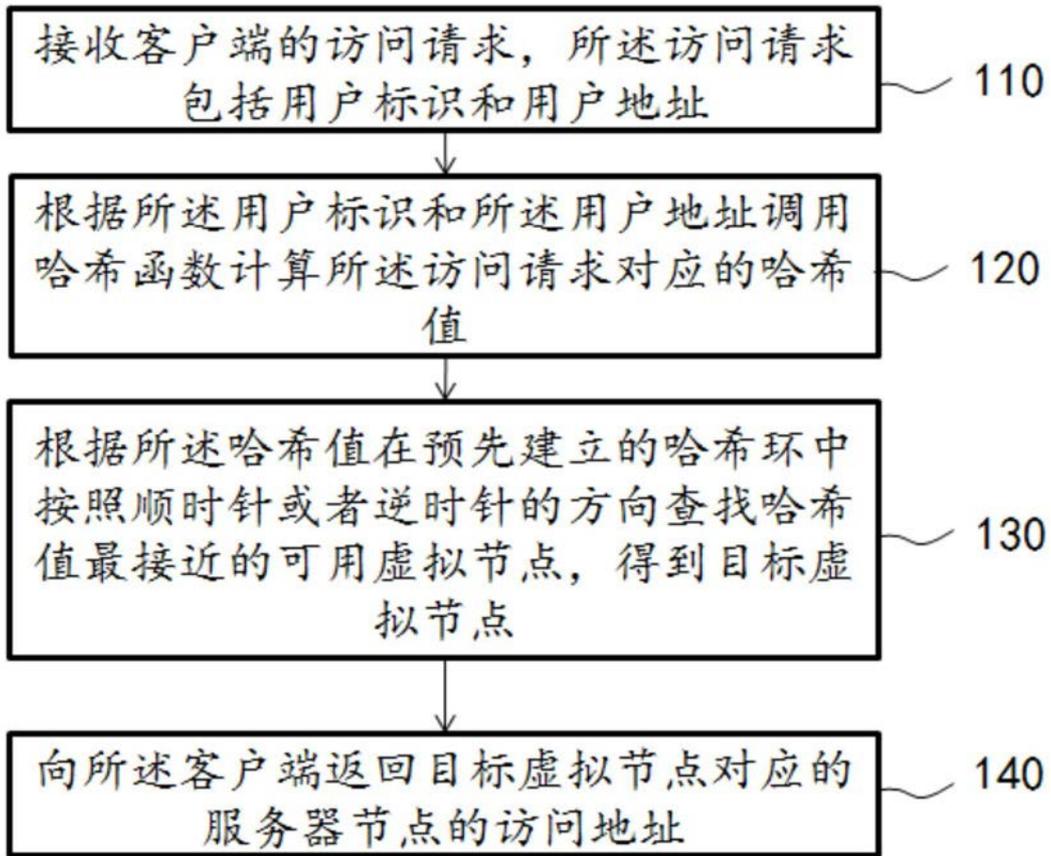


图1

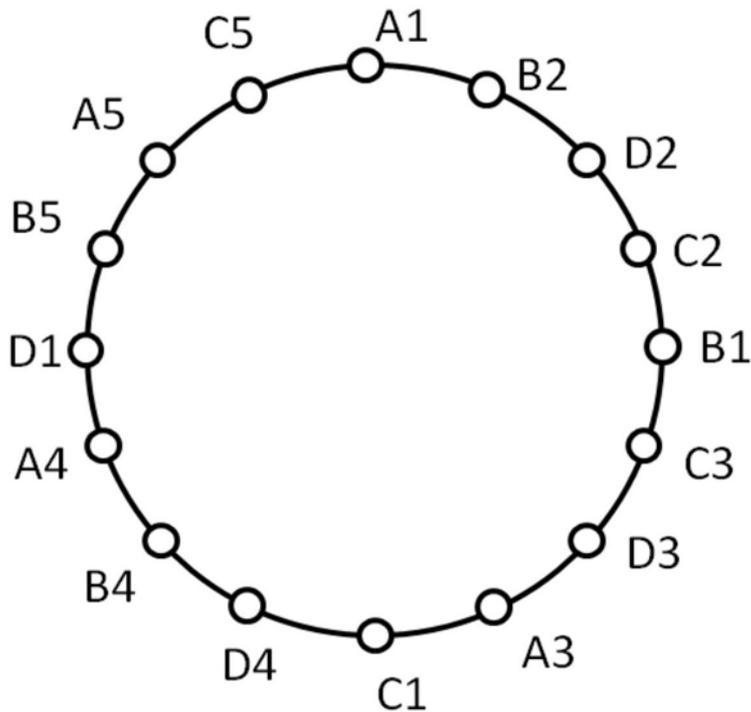


图2

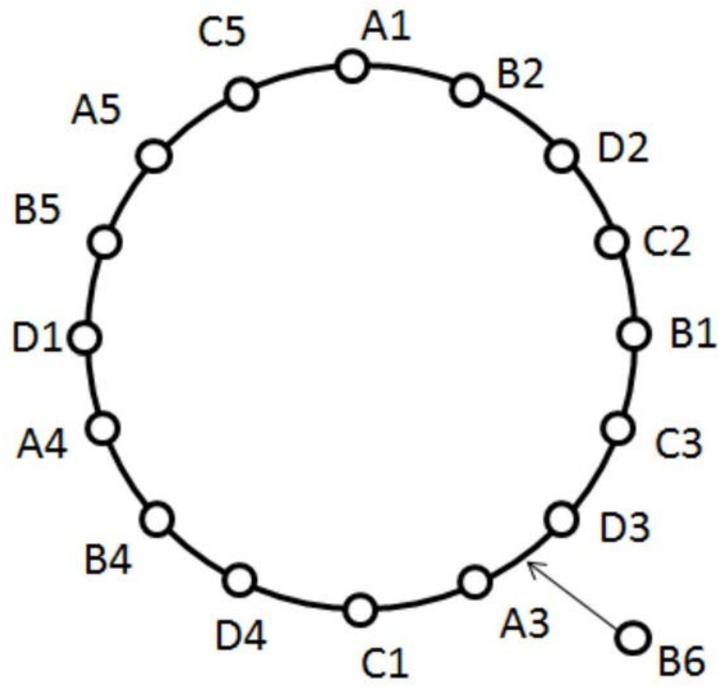


图3

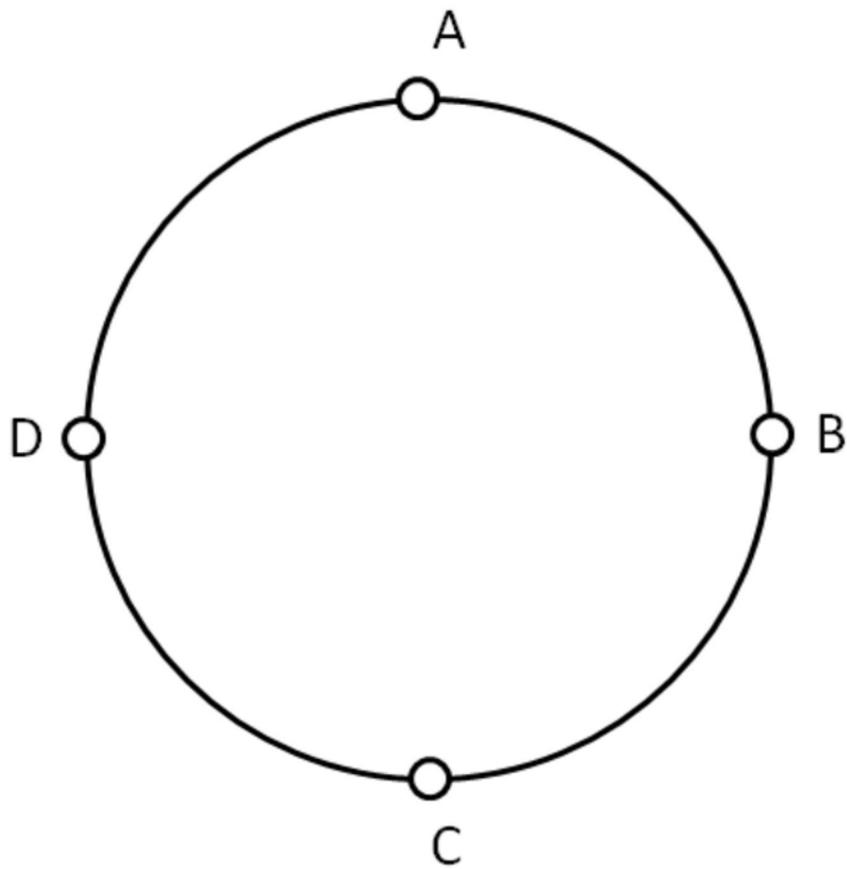


图4