



(12) 发明专利申请

(10) 申请公布号 CN 114157666 A

(43) 申请公布日 2022.03.08

(21) 申请号 202111241808.5

(22) 申请日 2021.10.25

(71) 申请人 北京千方科技股份有限公司  
地址 100085 北京市海淀区东北旺西路8号  
中关村软件园一期27号楼B座501室

(72) 发明人 张竹林 宁志明 夏曙东

(74) 专利代理机构 北京辰权知识产权代理有限公司 11619

代理人 李小朋

(51) Int.Cl.

H04L 67/1004 (2022.01)

H04L 67/1008 (2022.01)

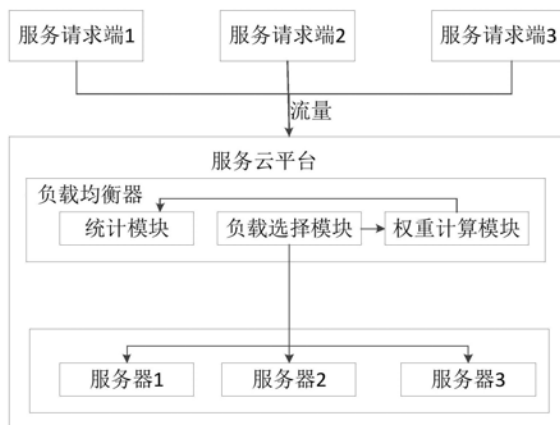
权利要求书2页 说明书10页 附图4页

(54) 发明名称

一种基于智能云平台的负载均衡系统、方法、介质及装置

(57) 摘要

本发明公开了一种基于智能云平台的负载均衡系统、方法、存储介质及负载均衡装置,系统包括服务请求端、服务云平台和服务端;当服务端设置于服务云平台内部时,服务云平台包括负载均衡器和多个服务端;负载均衡器包括统计模块、权重计算模块、负载选择模块;服务请求端生成服务请求发送至服务云平台;负载选择模块接收请求后从权重计算模块中获取各服务端的权重值,根据权重值确定服务端发起调用;统计模块记录并计算各服务端的异常请求比例来确定故障信息,将故障信息和各比例发送至权重计算模块计算各服务端的权重值,根据接收的故障信息调整异常服务端的权重。本申请能降低异常服务端的流量分配比例,提升系统的鲁棒性。



1. 一种基于智能云平台的负载均衡系统,其特征在于,所述系统包括服务请求端、服务云平台以及服务端;

当所述服务端设置于所述服务云平台的内部时,所述服务云平台包括负载均衡器和多个服务端;所述负载均衡器包括统计模块、权重计算模块、负载选择模块;

所述服务请求端,用于获取服务请求发送至所述服务云平台;

所述负载选择模块,用于当接收到服务请求时,从所述权重计算模块中获取每个服务端的权重值,并根据所述权重值确定出目标服务端,以及针对所述目标服务端发起调用;

所述统计模块,用于实时记录预设周期内所述每个服务端的请求正常次数和请求异常次数,并根据所述请求正常次数与请求异常次数计算每个服务端的异常请求比例,并根据所述每个服务端的异常请求比例确定异常服务端信息,并将所述异常服务端信息以及所述每个服务端的异常请求比例发送至所述权重计算模块;

所述权重计算模块,用于根据接收的所述每个服务端的异常请求比例计算所述每个服务端的当前权重值,并根据接收的所述异常服务端信息将所述异常服务端信息对应的异常服务端的当前权重调整为目标权重。

2. 根据权利要求1所述的系统,其特征在于,所述根据所述权重值确定出目标服务端,包括:

将所述每个服务端的权重值进行降序排列,得到排列后的权重值;

从所述排列后的权重值中选择最大权重值对应的服务端;

将所述最大权重值对应的服务端确定为目标服务端。

3. 根据权利要求1所述的系统,其特征在于,所述根据所述每个服务端的异常请求确定异常服务端信息,包括:

逐一判断所述每个服务端的异常请求比例是否大于预设阈值,生成多个判断结果;

从所述多个判断结果中挑选所述异常请求比例大于预设阈值的判断结果;

将所述异常请求比例大于预设阈值的判断结果对应的服务端确定为异常服务端。

4. 根据权利要求1所述的系统,其特征在于,所述根据接收的所述异常服务端信息将所述异常服务端信息对应的异常服务端的当前权重调整为目标权重,包括:

当接收到异常服务端信息时,基于所述异常服务端信息确定异常服务端;

获取所述异常服务端的当前权重;

获取预设权重表,并从所述预设权重表中确定出所述当前权重对应的目标权重;

将所述异常服务端的当前权重调整为所述目标权重。

5. 根据权利要求1所述的系统,其特征在于,

当所述服务端设置于所述服务云平台的外部时,所述服务云平台包括负载均衡器、服务注册器;所述服务端为多个边缘服务端;所述负载均衡器包括统计模块、权重计算模块、负载选择模块;

所述服务请求端,用于生成服务请求发送至所述服务云平台;

所述负载选择模块,用于当接收到服务请求时,根据所述服务注册器内注册的服务表单确定边缘服务端的标识信息,以及根据所述边缘服务端的标识信息从所述权重计算模块中获取每个边缘服务端的当前权重值,并根据所述权重值确定出目标边缘服务端,以及针对所述目标边缘服务端发起调用;

所述统计模块,用于实时记录预设周期内所述每个服务端的请求正常次数和请求异常次数,并根据所述请求正常次数与请求异常次数计算每个服务端的异常请求比例,并根据所述每个服务端的异常请求比例确定异常服务端信息,并将所述异常服务端信息以及所述每个服务端的异常请求比例发送至所述权重计算模块;

所述权重计算模块,用于根据接收的所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重值,并根据接收的所述异常边缘服务端信息将所述异常边缘服务端信息对应的异常边缘服务端的当前权重调整为目标权重。

6. 根据权利要求5所述的系统,其特征在于,所述根据接收的所述每个边缘服务端的异常请求比例计算所述每个边缘服务端的当前权重值之后,包括:

当至少存在两个权重值相同的可提供边缘服务端时,

计算所述可提供服务的边缘服务端中每个边缘服务端与所述服务请求端之间的距离,得到每个边缘服务端的距离值;

根据所述每个边缘服务端的距离值确定权重的参考系数,并结合所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重第一校正值,记为第一权重值。

7. 根据权利要求5所述的系统,其特征在于,所述根据接收的所述每个边缘服务端的异常请求比例计算所述每个边缘服务端的当前权重值后,还包括:

获取所述可提供服务的边缘服务端中每个边缘服务端的负载率;

根据所述每个边缘服务端的负载率确定权重的负载参考系数,并结合所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重第二校正值,记为第二权重值。

8. 一种负载均衡方法,用于服务云平台,其特征在于,所述方法包括:

统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数,并基于所述正常请求次数和异常请求次数计算异常请求比例,以及基于所述异常请求比例确定所述服务端是否发生故障;

当发生故障时,生成故障服务端信息发送给权重计算模块;

权重计算模块根据所述异常请求比例实时计算出每个服务端的权重值,并根据接收的所述故障服务端信息调整所述故障服务端权重;

负载选择模块根据所述权重计算模块中每个服务端的权重值确定出目标服务端,并针对所述目标服务端发起调用。

9. 一种计算机存储介质,其特征在于,所述计算机存储介质存储有多条指令,所述指令适于由处理器加载并执行如权利要求8的方法步骤。

10. 一种负载均衡装置,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序;

统计模块、权重计算模块以及负载选择模块;

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求8所述的方法。

## 一种基于智能云平台的负载均衡系统、方法、介质及装置

### 技术领域

[0001] 本发明涉及负载均衡技术领域,特别涉及一种基于智能云平台的负载均衡系统、方法、存储介质及负载均衡器。

### 背景技术

[0002] 对于智能交通云服务,时效性及准确性要求很高,尤其是在基于交通云平台为各交通主体(服务请求端)如车端,路侧,管理侧等提供云服务时,如何长期稳定保证各交通主体服务请求的响应时效,是智能交通服务提供商最为关注的问题之一。

[0003] 尤其是随着智能交通服务业务平台逐步运营,数据的积累逐步增加,业务平台系统的规模已经非常庞大,逐渐系统处理变慢等问题,会直接影响到服务请求的响应时效,严重影响用户体验和安全性。

[0004] 由此可见,如何获取有效的服务端,满足实际的服务需求,成为一个亟待解决的技术问题。

### 发明内容

[0005] 本申请实施例提供了一种基于智能云平台的负载均衡系统、方法、存储介质及负载均衡器。为了对披露的实施例的一些方面有一个基本的理解,下面给出了简单的概括。该概括部分不是泛泛评述,也不是要确定关键/重要组成元素或描绘这些实施例的保护范围。其唯一目的是用简单的形式呈现一些概念,以此作为后面的详细说明确定的序言。

[0006] 第一方面,本申请实施例提供了一种基于智能云平台的负载均衡系统,系统包括服务请求端、服务云平台以及服务端;

[0007] 当服务端设置于服务云平台的内部时,服务云平台包括负载均衡器和多个服务端;负载均衡器包括统计模块、权重计算模块、负载选择模块;

[0008] 服务请求端,用于生成服务请求发送至服务云平台;

[0009] 负载选择模块,用于当接收到服务请求时,从权重计算模块中获取每个服务端的权重值,并根据权重值确定出目标服务端,以及针对目标服务端发起调用;

[0010] 统计模块,用于实时记录预设周期内每个服务端的服务时效信息,并根据所述服务端的服务时效信息确定服务端的异常请求,并根据每个服务端的异常请求确定异常服务端信息,并将所述异常服务端信息以及所述每个服务端的异常请求发送至所述权重计算模块;

[0011] 权重计算模块,用于根据接收的所述每个服务端的异常请求计算所述每个服务端的当前权重值,并根据接收的所述异常服务端信息将所述异常服务端信息对应的异常服务端的当前权重调整为目标权重。

[0012] 可选的,根据权重值确定出目标服务端,包括:

[0013] 将每个服务端的权重值进行降序排列,得到排列后的权重值;

[0014] 从排列后的权重值中选择初始权重值对应的服务端;

- [0015] 将初始权重值对应的服务端确定为目标服务端。
- [0016] 可选的,根据每个服务端的异常请求比例确定异常服务端信息,包括:
- [0017] 逐一判断每个服务端的异常请求是否大于预设阈值,生成多个判断结果;
- [0018] 从所述多个判断结果中挑选所述异常请求大于预设阈值的判断结果;
- [0019] 将所述异常请求比例大于预设阈值的判断结果对应的服务端确定为异常服务端。
- [0020] 可选的,根据接收的异常服务端信息将异常服务端信息对应的异常服务端的当前权重调整为目标权重,包括:
- [0021] 当接收到异常服务端信息时,基于异常服务端信息确定异常服务端;
- [0022] 获取异常服务端的当前权重;
- [0023] 获取预设权重表,并从预设权重表中确定出当前权重对应的目标权重;
- [0024] 将异常服务端的当前权重调整为目标权重。
- [0025] 可选的,当服务端设置于服务云平台的外部时,服务云平台包括负载均衡器、服务注册器;服务端为多个边缘服务端;负载均衡器包括统计模块、权重计算模块、负载选择模块;
- [0026] 服务请求端,用于生成服务请求发送至服务云平台;
- [0027] 负载选择模块,用于当接收到服务请求时,根据服务注册器内注册的服务表单确定边缘服务端的标识信息,以及根据边缘服务端的标识信息从权重计算模块中获取每个边缘服务端的当前权重值,并根据权重值确定出目标边缘服务端,以及针对目标边缘服务端发起调用;
- [0028] 统计模块,用于实时记录预设周期内所述每个服务端的请求正常次数和请求异常次数,并根据所述请求正常次数与请求异常次数计算每个服务端的异常请求比例,,并根据所述每个服务端的异常请求比例确定异常服务端信息,并将所述异常服务端信息以及所述每个服务端的异常请求比例发送至所述权重计算模块;
- [0029] 权重计算模块,用于根据接收的每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重值,并根据接收的异常边缘服务端信息将异常边缘服务端信息对应的异常边缘服务端的当前权重调整为目标权重。
- [0030] 可选的,在根据接收的每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重值之后,还包括:
- [0031] 当至少存在两个权重值相同的可提供边缘服务端时,
- [0032] 计算所述可提供服务的边缘服务端中每个边缘服务端与所述服务请求端之间的距离,得到每个边缘服务端的距离值;
- [0033] 根据所述每个边缘服务端的距离值确定权重的参考系数,并结合所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重第一校正,记为第一权重值;
- [0034] 可选的,在根据接收的每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重值之后,还包括:
- [0035] 获取所述可提供服务的边缘服务端中每个边缘服务端的负载率;
- [0036] 根据所述每个边缘服务端的负载率确定权重的负载参考系数,并结合所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重第二校正,记为第二权重值。。
- [0037] 第二方面,本申请实施例提供了一种负载均衡方法,方法包括:

[0038] 统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数,并基于所述正常请求次数和异常请求次数计算异常请求比例,以及基于所述异常请求比例确定所述服务端是否发生故障;

[0039] 当发生故障时,生成故障服务端信息发送给权重计算模块;

[0040] 权重计算模块根据所述异常请求比例实时计算出每个服务端的权重值,并根据接收的所述故障服务端信息调整所述故障服务端权重;

[0041] 负载选择模块根据所述权重计算模块中每个服务端的权重值确定出目标服务端,并针对所述目标服务端发起调用。。

[0042] 第三方面,本申请实施例提供一种计算机存储介质,计算机存储介质存储有多条指令,指令适于由处理器加载并执行上述的方法步骤。

[0043] 第四方面,本申请实施例提供一种负载均衡装置,包括:

[0044] 一个或多个处理器;

[0045] 存储装置,用于存储一个或多个程序;

[0046] 统计模块、权重计算模块以及负载选择模块;

[0047] 当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现上述的负载均衡方法。

[0048] 本申请实施例提供的技术方案可以包括以下有益效果:

[0049] 在本申请实施例中,首先统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数,并基于正常请求次数和异常请求次数计算异常请求比例,以及基于异常请求比例确定服务端是否发生故障,当发生故障时,生成故障服务端信息发送给权重计算模块,然后权重计算模块根据异常请求比例实时计算出每个服务端的权重值,并根据接收的故障服务端信息调整故障服务端权重,最后负载选择模块根据权重计算模块中每个服务端的权重值确定出目标服务端,并针对目标服务端发起调用。由于本申请中通过调整服务端的权重来降低故障服务端的流量分配比例,从而提升系统的鲁棒性。

[0050] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本发明。

## 附图说明

[0051] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0052] 图1是本申请实施例提供的一种基于智能云平台的负载均衡系统示意图;

[0053] 图2是本申请实施例提供的另一种基于智能云平台的负载均衡系统示意图;

[0054] 图3是本申请实施例提供的一种负载均衡方法的流程示意图;

[0055] 图4是本申请实施例提供的一种故障感知方法示意图;

[0056] 图5是本申请实施例提供的一种故障转移方法示意图;

[0057] 图6是本申请实施例提供的一种基于云-边协同的智能交通云控平台;

[0058] 图7是本申请实施例提供的一种负载均衡装置的结构示意图;

[0059] 图8是本申请实施例提供的一种存储介质的示意图。

## 具体实施方式

[0060] 以下描述和附图充分地示出本发明的具体实施方案,以使本领域的技术人员能够实践它们。

[0061] 应当明确,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作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。

[0062] 下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0063] 在本发明的描述中,需要理解的是,术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性。对于本领域的普通技术人员而言,可以具体情况理解上述术语在本发明中的具体含义。此外,在本发明的描述中,除非另有说明,“多个”是指两个或两个以上。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0064] 本申请提供了一种基于智能云平台的负载均衡系统、方法、存储介质及负载均衡器,以解决上述相关技术问题中存在的问题。本申请提供的技术方案中,由于本申请中通过调整服务端的权重来降低异常服务端的流量分配比例,从而提升系统的鲁棒性,下面采用示例性的实施例进行详细说明。

[0065] 请参见图1,图1是本申请实施例提供的一种基于智能云平台的负载均衡系统示意图,该系统包括:服务请求端、服务云平台以及服务端;当服务端设置于服务云平台的内部时,服务云平台包括负载均衡器和多个服务端;负载均衡器包括统计模块、权重计算模块、负载选择模块;服务请求端,用于生成服务请求发送至服务云平台;负载选择模块,用于当接收到服务请求时,从权重计算模块中获取每个服务端的权重值,并根据权重值确定出目标服务端,以及针对目标服务端发起调用;统计模块,用于实时记录预设周期内所述每个服务端的请求正常次数和请求异常次数,并根据所述请求正常次数与请求异常次数计算每个服务端的异常请求比例,并根据所述每个服务端的异常请求比例确定异常服务端信息,并将所述异常服务端信息以及所述每个服务端的异常请求比例发送至所述权重计算模块;权重计算模块,用于根据接收的所述每个服务端的异常请求比例计算所述每个服务端的当前权重值,并根据接收的所述异常服务端信息将所述异常服务端信息对应的异常服务端的当前权重调整为目标权重。

[0066] 在本实施例中,在根据权重值确定出目标服务端时,首先将每个服务端的权重值进行降序排列,得到排列后的权重值,然后从排列后的权重值中选择初始权重值(权重最大的值)对应的服务端,最后将初始权重值对应的服务端确定为目标服务端。

[0067] 其中,降序排列是从将权重值从大到小进行排列,初始权重即排在第一个的最大权重值。

[0068] 在本实施例中,根据服务正常请求次数与异常请求次数计算每个服务端的异常请求比例,将正常请求次数与异常失败次数的比值确定为每个服务端的异常请求比例。

[0069] 具体地,服务请求异常包括服务响应时效异常,当某服务端的服务响应时间大于预设服务响应时间阈值时,即认为服务请求响应异常,记作一次异常请求。

[0070] 另外,服务异常还可以包括服务端不响应或响应报错等。

[0071] 在本实施例中,在根据每个服务端的异常请求比例确定异常服务端信息时,首先逐一判断每个服务端的异常请求比例是否大于预设阈值,生成多个判断结果,然后从多个判断结果中挑选异常请求比例大于预设阈值的判断结果,再将异常请求比例大于预设阈值的判断结果对应的服务端确定为异常服务端,最后根据异常服务端的参数信息生成异常服务端信息。

[0072] 在本实施例中,在根据接收的异常服务端信息将异常服务端信息对应的异常服务端的当前权重调整为目标权重时,首先当接收到异常服务端信息时,基于异常服务端信息确定异常服务端,然后获取异常服务端的当前权重,再获取预设权重表,并从预设权重表中确定出当前权重对应的目标权重,最后将异常服务端的当前权重调整为目标权重。

[0073] 具体的,预设权重表主要包含的字段包括服务端标识、异常服务请求持续时间、异常请求比例、每个异常请求比例对应当前权重及下一个阶段调整的目标权重。

[0074] 进一步地,请参见图2,图2是本申请实施例提供的另一种基于智能云平台的负载均衡系统示意图,该系统包括服务请求端、服务云平台以及服务端;

[0075] 当服务端设置于服务云平台的外部时,服务云平台包括负载均衡器、服务注册器;服务端为多个边缘服务端;负载均衡器包括统计模块、权重计算模块、负载选择模块;服务请求端,用于生成服务请求发送至服务云平台;负载选择模块,用于当接收到服务请求时,根据服务注册器内注册的服务表单确定边缘服务端的标识信息,以及根据边缘服务端的标识信息从权重计算模块中获取每个边缘服务端的当前权重值,并根据权重值确定出目标边缘服务端,以及针对目标边缘服务端发起调用;统计模块,用于实时记录预设周期内每个服务端的请求正常次数和请求异常次数,并根据请求正常次数与请求异常次数计算每个服务端的异常请求比例,并根据每个服务端的异常请求比例确定异常服务端信息,并将异常服务端信息以及所述每个服务端的异常请求比例发送至权重计算模块;权重计算模块,用于根据接收的每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重值,并根据接收的异常边缘服务端信息将异常边缘服务端信息对应的异常边缘服务端的当前权重调整为目标权重。

[0076] 在一实施方式中,根据接收的每个边缘服务端的异常请求比例计算得到每个边缘服务端的当前权重值,当至少存在两个权重值相同的可提供边缘服务端时,计算每个边缘服务端与服务请求端之间的距离,得到每个边缘服务端的距离值,根据每个边缘服务端的距离值与服务请求端的距离确定权重的参考系数,并结合所述每个边缘服务端的异常请求比例计算所述每个边缘服务端的当前权重第一校正系数,记为第一权重值。

[0077] 其中,根据每个边缘服务端距请求端的距离调整该边缘服务端对应的权重系数。具体的,在根据接收的每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重值时,首先计算可提供服务的边缘服务端中每个边缘服务端与服务请求端之间的距离,得到每个边缘服务端的距离值,再根据每个边缘服务端的距离值确定计算权重的第一参考系数,并结合每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重第一校正系数,记为第一权重值;



[0078] 在另一实施方式中,根据每个边缘服务端的负载率的差异调整该边缘服务端的权重,具体地,首先计算可提供服务的边缘服务端中每个边缘服务端的负载率,得到每个边缘服务端的负载率,然后根据每个边缘服务端的负载率确定为计算权重的第二参考系数,并结合每个边缘服务端的异常请求比例计算每个边缘服务端的当前权重第二校正,记为第二权重值。

[0079] 在本申请实施例中,首先统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数,并基于正常请求次数和异常请求次数计算异常请求比例,以及基于异常请求比例确定服务端是否发生故障,当发生故障时,生成异常服务端信息发送给权重计算模块,然后权重计算模块根据异常请求比例实时计算出每个服务端的权重值,并根据接收的异常服务端信息调整异常服务端权重,最后负载选择模块根据权重计算模块中每个服务端的权重值确定出目标服务端,并针对目标服务端发起调用。由于本申请中通过调整服务端的权重来降低异常服务端的流量分配比例,从而提升系统的鲁棒性。

[0080] 下面将结合附图3-附图6,对本申请实施例提供的负载均衡方法进行详细介绍。该方法可依赖于计算机程序实现,可运行于基于冯诺依曼体系的负载均衡装置上。该计算机程序可集成在应用中,也可作为独立的工具类应用运行。

[0081] 请参见图3,为本申请实施例提供了一种负载均衡方法的流程示意图。如图1所示,本申请实施例的方法可以包括以下步骤:

[0082] S101,统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数,并基于正常请求次数和异常请求次数计算异常请求比例,以及基于异常请求比例确定服务端是否发生故障;

[0083] 其中,统计模块、权重计算模块以及负载选择模块设置在负载均衡器中,负载均衡器设置在服务云平台。

[0084] 通常,当服务云平台中接收到服务请求端的服务请求时,先通过负载选择模块接收该请求,并获取权重计算模块中各服务端的权重值和预设算法确定出权重最大的服务端进行响应,当服务请求被正常响应时,统计模块记录一次正常响应,当服务请求被异常响应时,统计模块记录一次异常响应,正常响应及异常响应的记录包括服务请求端、异常服务端、正常/异常响应的时间等。

[0085] 进一步地,预设算法可以是轮询法、随机加权算法。

[0086] 在本申请实施例中,如图4所示,服务端每正常响应一次服务请求端的服务请求时,统计模块的正常请求次数即加1,每异常/超时一次,异常/超时次数+1,在一个统计周期内,根据统计的次数计算异常请求比例,当周期内异常请求比例超过预设阈值时,统计模块将异常的服务端信息发送至权重模块,权重模块即调整该异常服务端的权重值。

[0087] S102,当发生故障时,生成异常服务端信息发送给权重计算模块;

[0088] S103,权重计算模块根据异常请求比例实时计算出每个服务端的权重值,并根据接收的异常服务端信息调整异常服务端权重;

[0089] S104,负载选择模块根据权重计算模块中每个服务端的权重值确定出目标服务端,并针对目标服务端发起调用。

[0090] 在本申请实施例中,权重计算模块负责计算出每个服务端的权重值,假设未发生故障的时候权重是100,当服务端在当前的统计周期(15秒)内发生故障时,权重立刻调整成

1,即此时该服务端分配的流量是其他实例的1%,当流量相对较小的时候该服务端几乎不会分配到流量。如果在下一个统计周期(15秒),未发生故障,将调大该服务端的权重到4;如果在接下来的15秒内仍未发生故障,将该服务端的权重调整到64;在下一个15秒,权重调整到100,恢复正常。如果在任意一个时间窗口内,再次发生故障,流量权重调整成1。

[0091] 在一种可能的实现方式中,例如图5所示,负载选择模块在服务请求端向服务端发起请求的时候,负载选择模块通过获取权重模块根据统计模块发送的异常请求比例计算出服务端的权重,再利用权重分配算法,选择一个具体的服务端出来,发起调用。

[0092] 进一步地,例如图2所示,当服务端设置在服务云平台外部,即以集群的形式分布式部署在各边缘节点时,服务请求方通过服务请求端发起服务请求(服务请求可以为定位或导航服务请求),位于云服务平台的负载选择模块接收该服务请求后,根据云服务平台的服务注册器内注册的服务表单,确定边缘服务端所在的边缘服务段编号或容器编号,以及通过边缘服务段编号或容器编号从权重计算模块获取多个边缘服务端的权重值。

[0093] 当边缘服务端有多个时,初始化时,每个边缘服务端的流量分配比例默认为是一致的,均默认为100,负载选择模块根据轮询法,哈希算法,加权随机算法等选定服务端以响应该服务请求,当服务端对该服务请求正常响应时,在统计模块中增加一次正常响应记录,当服务端对该服务请求异常响应时,在统计模块中增加一次异常响应记录,统计每个周期的正常响应记录跟异常响应记录,当异常响应的比例超过预设阈值时,通过权重计算模块,调整该边缘服务端的流量比例(比如流量从100调整到1),以减少分配至该边缘服务端的服务请求。

[0094] 进一步地,当多个边缘服务端均可同一服务请求提供服务(即多个服务范围时),但该多个边缘服务端的服务覆盖范围不一致的(如多个服务端分布于路侧的边缘计算模块),在计算权重时,将边缘服务端与请求端的距离作为一个权重计算模块的一个参考系数,记作第一参考系数,并结合所述每个边缘服务端的异常请求计算所述每个边缘服务端的当前权重第一校正系数,记为第一权重值。其中,距离请求端越近的边缘服务端,服务响应的的时间越短,时效性越强,边缘服务端的权重值越大。

[0095] 根据距离调整后的权重值,即第一权重值计算公式如下:

$$[0096] \quad \alpha_1 = \alpha_0 + \left( \frac{\beta_1}{d} \right)^b + A$$

[0097] 其中, $\alpha_1$ 为根据距离调整后得到第一权重, $\alpha_0$ 为根据权重计算模块分配的权重值,即目标权重, $b \geq 1$ , $d$ 为边缘服务端与请求端的距离, $A$ 为距离影响常数, $\beta_1/d$ 为第一参考系数, $b$ 为第一参考系数的影响指数。

[0098] 因受负载率影响,受负载影响的权重影响因子 $\kappa_1 = \frac{\alpha_1}{\alpha_0}$ ;

[0099] 在另一个优选的实施方式中,在权重系数的计算过程中,将各边缘服务端的负载情况也作为整个权重分配的一个影响因素考虑,记作第二参考系数,具体地首先计算可提供服务的边缘服务端中每个边缘服务端的负载率,得到每个边缘服务端的负载率,然后根据每个边缘服务端的负载率确定为计算权重的第二参考系数。边缘服务端的负载越高,能够分配出来的计算资源越少,在负载均衡模块进行负载选择时,其分配的权重值适当调低,以降低该边缘服务端的负载压力。

[0100] 根据负载调整后的权重值为:

[0101]  $\alpha_2 = \alpha_0 - \beta * 1^x + B$ , 其中,  $\alpha_2$  为根据负载情况调整后的权重值, 记作第二权重值,  $\alpha_0$  为根据权重计算模块计算得到的初始权重值, 即基于当前权重通过预设权重表查找计算得到的目标权重值,  $x$  为负载的影响指数,  $\beta$  为边缘服务端的负载系数,  $1$  为该边缘服务端的当前负载,  $x \geq 1$ ,  $B$  为负载影响常数。

[0102] 因受负载率影响, 受负载影响的权重影响因子  $\kappa_2 = \frac{\alpha_2}{\alpha_0}$ ;

[0103] 在一个优选的实施方式中, 当一个服务端的分配权重既受负载率影响, 又受距离影响时, 其调整权重值记作第三权重值, 第三权重值  $\alpha_3 = \alpha_1 * \kappa_2$  或

[0104]  $\alpha_3 = \alpha_2 * \kappa_1$

[0105] 在本申请实施例中, 首先统计模块实时记录预设周期内每个服务端的正常请求次数和异常请求次数, 并基于正常请求次数和异常请求次数计算异常请求比例, 以及基于异常请求比例确定服务端是否发生故障, 当发生故障时, 生成异常服务端信息发送给权重计算模块, 然后权重计算模块根据异常请求比例实时计算出每个服务端的权重值, 并根据接收的异常服务端信息调整异常服务端权重, 最后负载选择模块根据权重计算模块中每个服务端的权重值确定出目标服务端, 并针对目标服务端发起调用。由于本申请中通过调整服务端的权重来降低异常服务端的流量分配比例, 从而提升系统的鲁棒性。

[0106] 请参见图6, 为本申请实施例提供了一种基于云-边协同的故障感知及负载均衡方法, 用于智能交通网联云平台。

[0107] 如附图6所示, 包括智能交通云平台及与智能交通云平台相连接的多个边缘端, 在智能交通云平台及多个边缘端均设置有负载均衡器。

[0108] 其中, 设置于智能交通云平台的负载均衡器的负载策略如下:

[0109] 接收到服务请求时, 先根据服务请求所在的位置跟区域, 查找可为服务请求提供设备的服务实例(根据智能终端所在位置及请求服务范围, 确定是边缘侧1, 2或是3中的一个为服务提供端), 当服务提供端(边缘端)中的服务实例有多个时, 位于边缘平台的负载均衡器根据轮询法, 随机法或加权随机法选定一个服务端进行响应, 负载均衡器记录响应结果, 当正常响应时, 在正常响应的记录上加1, 当异常相应时, 在异常响应的记录上加1, 按照预设周期统计每个周期的异常相应比例, 当异常响应比例超过预设阈值时, 调整服务端的分配权重及比例, 比如分配权重由100降低至1, 如果在下一个统计周期, 未发生故障, 将调大该服务实例的权重到4; 如果在接下来的一个统计周期内仍未发生故障, 将该实例的权重调整到64; 在下一个统计周期, 权重调整到100, 恢复正常。

[0110] 在另一实施方式中, 当智能云平台中的服务注册器根据服务请求, 查找出两个以上可提供服务响应的边缘计算端时, 负载均衡器分配该服务相应端时, 须考虑边缘计算端的服务响应时间, 其中, 服务响应时间是指以单位时间为基准, 各边缘计算端处理该服务请求所需的时间, 分别为  $t_1, t_2, t_3$ , 响应时间越短, 响应越快, 分配权重越高。

[0111] 在一种实施方式中, 在基于权重分配模块得到的目标权重的基础上, 可根据响应时效调整权重, 按照响应实现调整后的权重可采用如下方式进行计算,

[0112]  $\alpha_t = \alpha_0 + (\beta / t_n)^m + C$

[0113] 其中,  $\alpha_t$  为时间的调整系数,  $\alpha_0$  为初始的权重值, 即按照权重分配模块得到的目

标权重,  $t_n$  为第  $n$  个边缘计算端的响应时间,  $C$  为时间影响常数,  $m \geq 1$ ,  $m$  为时间影响指数。需要说明的是, 在本申请的实施方式中, 当可以确定时间调整权重时, 优先采用响应时间调整权重作为各边缘服务端的权重。采用加权随机分配法从多个边缘计算端选择唯一的边缘计算端。

[0114] (1) 当边缘计算端确定时, 如边缘计算端内仅有一个服务实例, 即确定该服务实例作为服务提供端,

[0115] 为服务请求提供响应, 服务响应完成后, 将响应结果 (正常响应或是异常响应) 发送至智能交通云控平台;

[0116] (2) 当边缘计算端有多个服务实例时, 根据位于边缘计算端的负载均衡器统计的各实施例的权重分配, 选择一个实施例进行响应, 并记录服务响应的结果, 根据统计周期内的某服务端异常响应的比例, 确定是否调整该服务端的分配权重。分配权重的调整策略如上述实施方式所示。

[0117] 在本实施方式中, 边缘计算端设置于路侧, 可接收路侧的智能感知数据, 根据服务请求计算与服务请求对应的服务响应结果, 并将服务响应结果发送至智能交通云控平台, 或者将边缘计算端的响应结果直接发送至请求端。但将响应正常与否的结果反馈至边缘计算端及智能交通平台的负载均衡模块。

[0118] 在一些优选的实施方式中, 边缘计算端内的服务以微服务的形式部署, 可部署于容器内, 也可部署于容器外。

[0119] 本申请实施方式还提供一种与前述实施方式所提供的负载均衡方法对应的负载均衡器, 以执行上述负载均衡方法。

[0120] 请参考图7, 其示出了本申请的一些实施方式所提供的一种负载均衡装置的示意图。如图7所示, 负载均衡器2包括: 处理器200, 存储器201, 总线202和通信接口203, 处理器200、通信接口203和存储器201通过总线202连接; 存储器201中存储有可在处理器200上运行的计算机程序, 处理器200运行计算机程序时执行本申请前述任一实施方式所提供的负载均衡方法。

[0121] 其中, 存储器201可能包含高速随机存取存储器 (RAM: Random Access Memory), 也可能还包括非不稳定的存储器 (non-volatile memory), 例如至少一个磁盘存储器。通过至少一个通信接口203 (可以是有线或者无线) 实现该系统网元与至少一个其他网元之间的通信连接, 可以使用互联网、广域网、本地网、城域网等。

[0122] 总线202可以是ISA总线、PCI总线或EISA总线等。总线可以分为地址总线、数据总线、控制总线等。其中, 存储器201用于存储程序, 处理器200在接收到执行指令后, 执行程序, 前述本申请实施例任一实施方式揭示的负载均衡方法可以应用于处理器200中, 或者由处理器200实现。

[0123] 处理器200可能是一种集成电路芯片, 具有信号的处理能力。在实现过程中, 上述方法的各步骤可以通过处理器200中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器200可以是通用处理器, 包括中央处理器 (Central Processing Unit, 简称CPU)、网络处理器 (Network Processor, 简称NP) 等; 还可以是数字信号处理器 (DSP)、专用集成电路 (ASIC)、现成可编程门阵列 (FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑

框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器201,处理器200读取存储器201中的信息,结合其硬件完成上述方法的步骤。

[0124] 本申请实施例提供的负载均衡器与本申请实施例提供的负载均衡方法出于相同的发明构思,具有与其采用、运行或实现的方法相同的有益效果。

[0125] 本申请实施方式还提供一种与前述实施方式所提供的负载均衡方法对应的计算机可读存储介质,请参考图8,其示出的计算机可读存储介质为光盘30,其上存储有计算机程序(即程序产品),计算机程序在被处理器运行时,会执行前述任意实施方式所提供的负载均衡方法。

[0126] 需要说明的是,计算机可读存储介质的例子还可以包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他光学、磁性存储介质,在此不再一一赘述。

[0127] 本申请的上述实施例提供的计算机可读存储介质与本申请实施例提供的负载均衡方法出于相同的发明构思,具有与其存储的应用程序所采用、运行或实现的方法相同的有益效果。

[0128] 以上,仅为本申请较佳的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

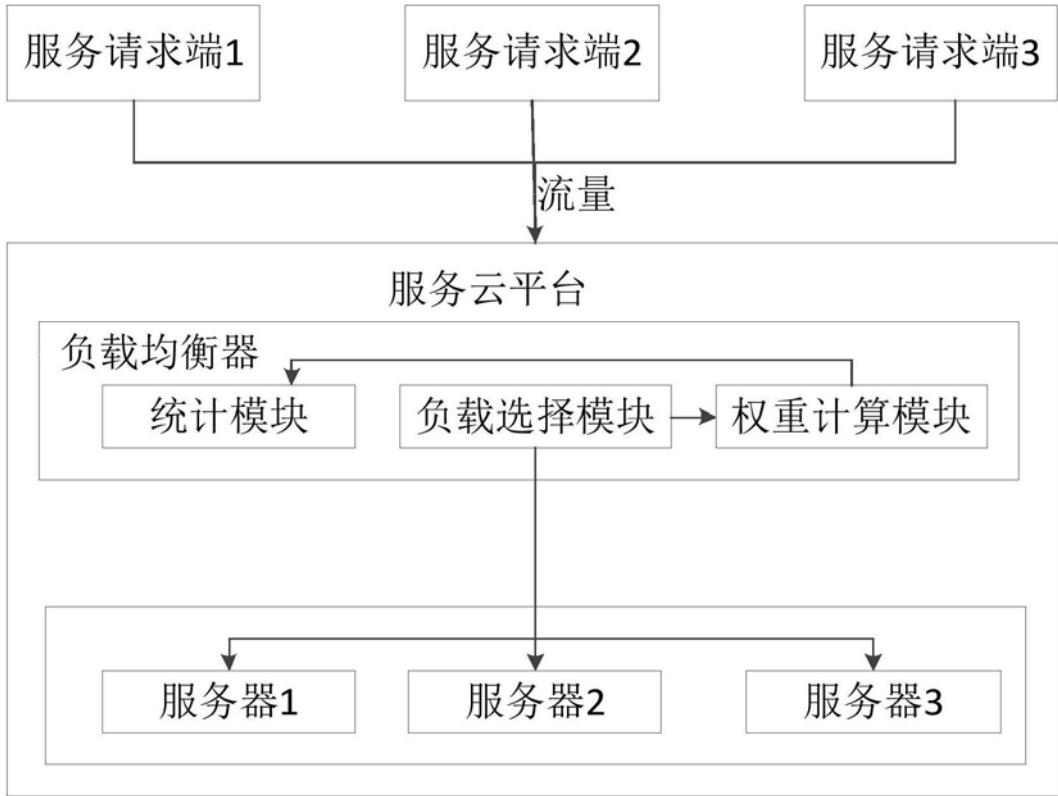


图1

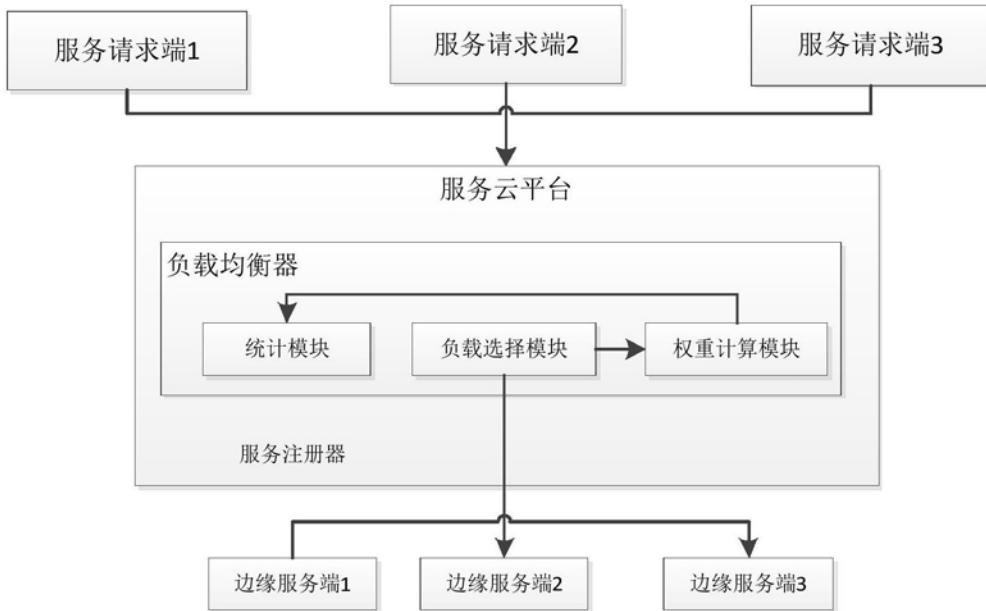


图2

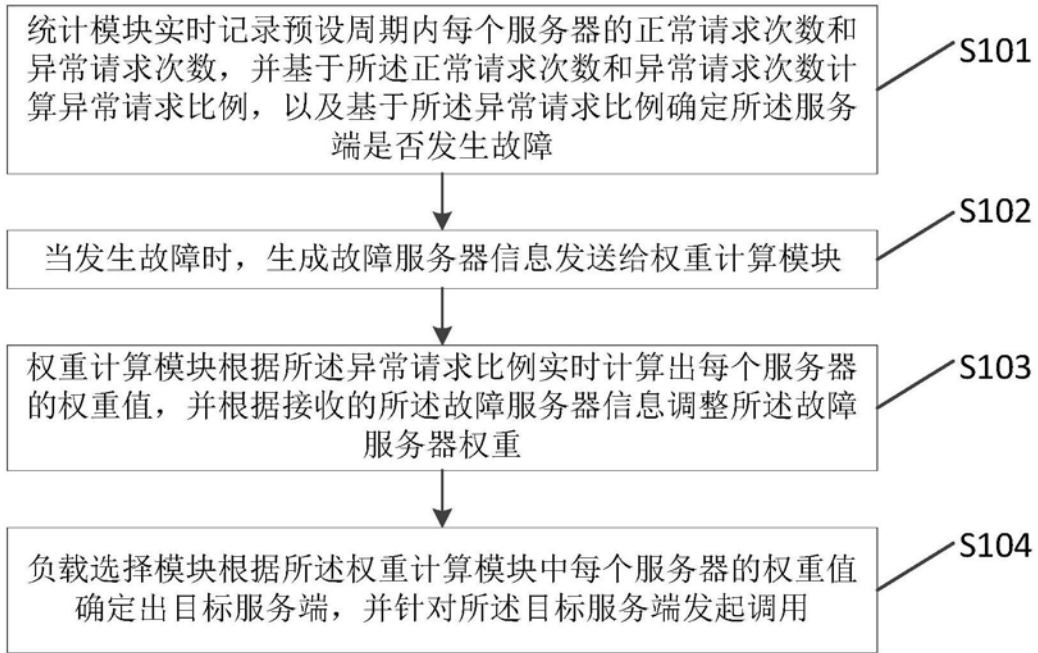


图3

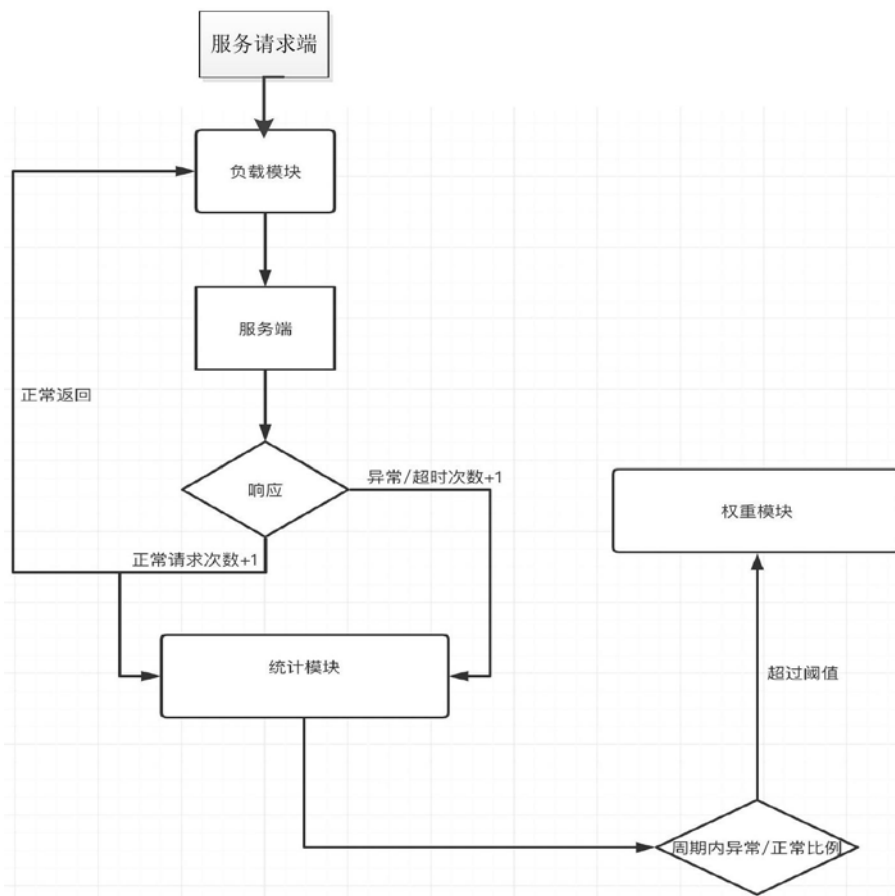


图4





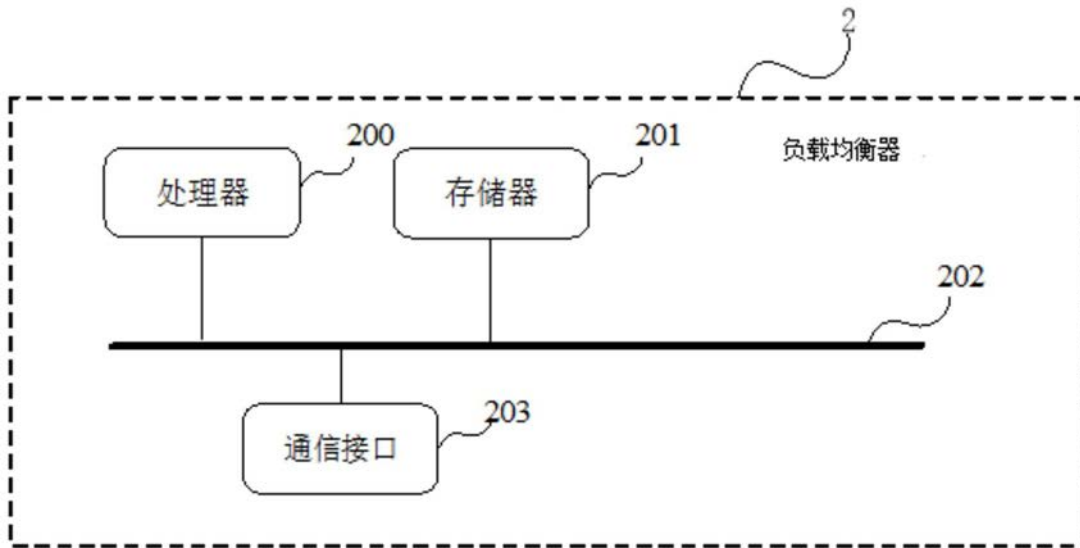


图7

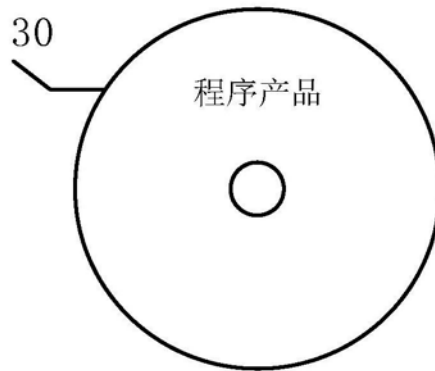


图8