



(12) 发明专利

(10) 授权公告号 CN 107222426 B

(45) 授权公告日 2021.07.20

(21) 申请号 201610161957.3

(22) 申请日 2016.03.21

(65) 同一申请的已公布的文献号  
申请公布号 CN 107222426 A

(43) 申请公布日 2017.09.29

(73) 专利权人 阿里巴巴集团控股有限公司  
地址 英属开曼群岛大开曼资本大厦一座四  
层847号邮箱

(72) 发明人 林锦成

(74) 专利代理机构 北京润泽恒知识产权代理有  
限公司 11319

代理人 赵娟

(51) Int. Cl.  
H04L 12/813 (2013.01)

(56) 对比文件

CN 103873282 A, 2014.06.18

CN 103369601 A, 2013.10.23

审查员 李红玲

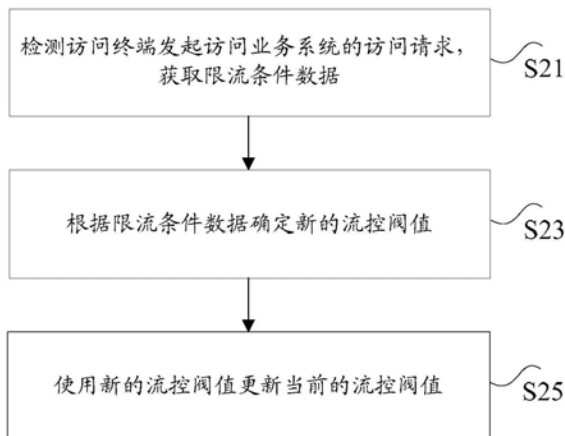
权利要求书3页 说明书18页 附图6页

(54) 发明名称

控流的方法、装置及系统

(57) 摘要

本发明公开了一种控流的方法、装置及系统。其中,该方法包括:检测访问终端发起访问业务系统的访问请求,获取限流条件数据;根据限流条件数据确定新的流控阈值;使用新的流控阈值更新当前的流控阈值。本发明解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。



1. 一种控流的方法,其特征在于,包括:

检测访问终端发起访问业务系统的访问请求,获取限流条件数据;所述限流条件数据包括当前的流控阈值和所述业务系统的负载数据;所述负载数据包括任务处理值或CPU占用率;

根据所述限流条件数据确定新的流控阈值;

使用所述新的流控阈值更新当前的流控阈值;

所述业务系统包括多个业务服务器,其中,所述根据所述限流条件数据确定新的流控阈值包括:

如果所述业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或所述CPU占用率大于等于第二预定值,则将所述流控阈值下降至第二预定流控阈值;

如果所述业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或所述CPU占用率小于等于第四预定值,则将所述流控阈值提高至第三预定流控阈值。

2. 根据权利要求1所述的方法,其特征在于,所述限流条件数据还包括:历史流控数据。

3. 根据权利要求2所述的方法,其特征在于,所述历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合。

4. 根据权利要求3所述的方法,其特征在于,所述流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,其中,根据所述限流条件数据确定新的流控阈值,还包括:

判断所述统计的限流次数是否超过预定次数,和/或所述限流频率是否超出预定频率值;

如果所述统计的限流次数超过所述预定次数,和/或所述限流频率超出所述预定频率值,则将所述流控阈值下降至第一预定流控阈值;

在根据所述历史流控数据确定将所述流控阈值调整至所述第一预定流控阈值,并根据所述业务系统的负载数据确定将所述流控阈值调整至所述第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为所述历史流控数据的优先级高于所述负载数据的优先级,则将所述第一预定流控阈值作为所述新的流控阈值;

如果所述限流控制优先级为所述负载数据的优先级高于所述历史流控数据的优先级,将所述第二预定流控阈值或所述第三预定流控阈值作为所述新的流控阈值。

5. 根据权利要求1所述的方法,其特征在于,在读取限流条件数据之前,所述方法还包括:

所述访问终端通过网关将所述访问请求发送给所述业务系统,其中,所述网关用于使用所述当前的流控阈值对所述访问请求进行限流控制;

所述业务系统将所述访问请求和进行所述限流控制时得到的流控数据写入日志系统,其中,所述日志系统用于记录历史流控数据;

其中,在日志系统写入所述访问请求的情况下,所述日志系统将所述历史流控数据发送至分析系统。

6. 根据权利要求5所述的方法,其特征在于,在使用所述新的流控阈值更新所述当前的流控阈值之后,所述方法还包括:

所述网关获取到更新后的流控阈值,或者将所述更新后的流控阈值发送给所述网关,

使得所述网关使用所述更新后的流控阈值对所述访问终端进行限流控制。

7. 根据权利要求5所述的方法,其特征在于,所述网关对所述访问请求进行如下任意一个或多个处理:协议转换处理、参数转换处理、访问控制处理和限流控制处理。

8. 一种控流的装置,其特征在于,包括:

分析模块,用于检测访问终端发起访问业务系统的访问请求,获取限流条件数据;所述限流条件数据包括当前的流控阈值和所述业务系统的负载数据;所述负载数据包括任务处理值或CPU占用率;

确定模块,用于根据所述限流条件数据确定新的流控阈值;

更新模块,用于使用所述新的流控阈值更新当前的流控阈值;

所述业务系统包括多个业务服务器,其中,所述确定模块包括:

第二下降模块,用于如果所述业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或所述CPU占用率大于等于第二预定值,则将所述流控阈值下降至第二预定流控阈值;

提高模块,用于如果所述业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或所述CPU占用率小于等于第四预定值,则将所述流控阈值提高至第三预定流控阈值。

9. 根据权利要求8所述的装置,其特征在于,所述限流条件数据还包括:历史流控数据。

10. 根据权利要求9所述的装置,其特征在于,所述历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合。

11. 根据权利要求10所述的装置,其特征在于,所述流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,其中,所述确定模块还包括:

判断模块,用于判断所述统计的限流次数是否超过预定次数,和/或所述限流频率是否超出预定频率值;

第一下降模块,用于如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值;

第一确认模块,用于在根据所述历史流控数据确定将所述流控阈值调整至所述第一预定流控阈值,并根据所述业务系统的负载数据确定将所述流控阈值调整至所述第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为所述历史流控数据的优先级高于所述负载数据的优先级,则将所述第一预定流控阈值作为所述新的流控阈值;

第二确认模块,用于如果所述限流控制优先级为所述负载数据的优先级高于所述历史流控数据的优先级,将所述第二预定流控阈值或所述第三预定流控阈值作为所述新的流控阈值。

12. 一种控流的系统,其特征在于,包括:

访问终端,向业务系统发起访问请求;

分析系统终端,与所述访问终端通讯,检测访问终端发起访问所述业务系统的访问请求,获取限流条件数据,根据所述限流条件数据确定新的流控阈值,并使用所述新的流控阈值更新当前的流控阈值;所述限流条件数据包括当前的流控阈值和所述业务系统的负载数据;所述负载数据包括任务处理值或CPU占用率;所述业务系统包括多个业务服务器,其中,

所述根据所述限流条件数据确定新的流控阈值包括：如果所述业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值，或所述CPU占用率大于等于第二预定值，则将所述流控阈值下降至第二预定流控阈值；如果所述业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值，或所述CPU占用率小于等于第四预定值，则将所述流控阈值提高至第三预定流控阈值。

13. 根据权利要求12所述的系统，其特征在于，所述限流条件数据还包括：历史流控数据。

14. 根据权利要求13所述的系统，其特征在于，所述历史流控数据包括：统计的限流次数和限流频率中的一个或二者的组合。

15. 根据权利要求14所述的系统，其特征在于，

所述分析系统终端还用于判断统计的限流次数是否超过预定次数，和/或限流频率是否超出预定频率值，如果所述统计的限流次数超过所述预定次数，和/或所述限流频率超出所述预定频率值，则将所述流控阈值下降至第一预定流控阈值；在根据所述历史流控数据确定将所述流控阈值调整至所述第一预定流控阈值，并根据所述业务系统的负载数据确定将所述流控阈值调整至所述第二预定流控阈值或第三预定流控阈值的情况下，如果限流控制优先级为所述历史流控数据的优先级高于所述负载数据的优先级，则将所述第一预定流控阈值作为所述新的流控阈值；如果所述限流控制优先级为所述负载数据的优先级高于所述历史流控数据的优先级，将所述第二预定流控阈值或所述第三预定流控阈值作为所述新的流控阈值。

16. 根据权利要求12所述的系统，其特征在于，所述系统还包括：

网关，分别与所述访问终端和所述分析系统终端通讯，用于使用所述当前的流控阈值对所述访问请求进行限流控制，并从所述分析系统终端获取到更新后的流控阈值，或者接收所述分析系统终端发送的所述更新后的流控阈值，使用所述更新后的流控阈值对所述访问终端进行限流控制。

## 控流的方法、装置及系统

### 技术领域

[0001] 本发明涉及网络应用技术领域,具体而言,涉及一种控流的方法、装置及系统。

### 背景技术

[0002] 在网络应用中,为了应对突发异常流量对于应用系统的影响,现有技术中常用的方法是采取流控或降级措施,即通过网关多维度地限制客户端访问者的请求,从而达到流控的目的,例如,限制每个客户端IP在每秒钟内的访问次数、限制某个API在每秒钟内的访问次数或限制某个用户对于某个API(Application Programming Interface,应用程序编程接口)的访问频率等。

[0003] 现有技术中通常采用的控流的方法为静态的限流策略,即,现有技术中的限流策略是通过配置系统或配置文件直接配置完成的,是固定不变的限流规则。图1是根据现有技术的一种控流系统的系统结构示意图,该系统结构可以提供如下控流方法对应的实现步骤:

[0004] A、客户端向网关发起访问请求;

[0005] B、网关获取预先配置好的流控规则;

[0006] C、网关使用预先配置好的流控规则对客户端进行限流控制;

[0007] D、网关将经过限流控制的访问请求发送至应用系统。

[0008] 然而在一些特殊场景下,例如某个区域的API出现了问题,或者被访问的应用程序的内部的压力比较大引起系统的响应时间较长,在上述情况下如果继续按照原有固定的限流策略进行限流,并有大量的访问请求的话,对被访问的应用系统来说,可能直接导致系统崩溃。

[0009] 针对现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题,目前尚未提出有效的解决方案。

### 发明内容

[0010] 本发明实施例提供了一种控流的方法、装置及系统,以至少解决现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。

[0011] 根据本发明实施例的一个方面,提供了一种控流的方法,包括:检测访问终端发起访问业务系统的访问请求,获取限流条件数据;根据限流条件数据确定新的流控阈值;使用新的流控阈值更新当前的流控阈值。

[0012] 根据本发明实施例的另一方面,还提供了一种控流的装置,包括:分析模块,用于检测访问终端发起访问业务系统的访问请求,获取限流条件数据;确定模块,用于根据限流条件数据确定新的流控阈值;更新模块,用于使用新的流控阈值更新当前的流控阈值。

[0013] 根据本发明实施例的另一方面,还提供了一种限流的系统,包括:访问终端,向业

务系统发起访问请求;分析系统终端,与所述访问终端通讯,检测所述访问终端发起所述访问业务系统的访问请求,则获取限流条件数据,根据所述限流条件数据确定新的流控阈值,并使用所述新的流控阈值更新所述当前的流控阈值。

[0014] 由于限流条件数据中至少包括了当前的流控阈值,因此根据限流条件获得的新的流控阈值是与当前的限流条件对应的流控阈值,可以认为,动态的新的流控阈值能够与当前的限流条件数据相对应。

[0015] 容易注意到的是,当业务系统的运行状态发生变动(例如:某个API出现故障、业务系统运行压力过大等)或终端对业务系统进行异常访问时,分析系统根据限流条件数据的变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端的访问请求进行限流;由于新的流控阈值与当前业务系统的运行状态(业务系统的运行状态体现于限流条件数据)相匹配,因此新的流控阈值也是与当前业务系统的运行状态相匹配的流控阈值,避免了业务系统的运行状态处于变动的状态但流控阈值为固定阈值导致的流控阈值与业务系统的运行状态不匹配的问题,从而避免了由于业务系统的运行状态和流控阈值不匹配导致的业务系统崩溃的问题。

[0016] 由此,本申请提供的上述实施例的方案解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。

## 附图说明

[0017] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

[0018] 图1是根据现有技术的一种控流系统的系统结构示意图;

[0019] 图2是根据本申请实施例1的一种控流方法的计算机终端的硬件结构框图;

[0020] 图3是根据本发明实施例1的控流的方法的流程图;

[0021] 图4是根据本发明实施例1的一种可选的控流的方法的示意图;

[0022] 图5是根据本申请实施例1的一种可选的控流的方法的信息交互图;

[0023] 图6是根据本申请实施例2的一种控流的装置的示意图;

[0024] 图7是根据本申请实施例2的一种可选的控流的装置的示意图;

[0025] 图8是根据本申请实施例2的一种可选的控流的装置的示意图;

[0026] 图9是根据本申请实施例2的一种可选的控流的装置的示意图;

[0027] 图10是根据本申请实施例2的一种可选的控流的装置的示意图;

[0028] 图11是根据本申请实施例3的一种可选的控流的系统的结构图;

[0029] 图12是根据本申请实施例的一种计算机终端的结构框图。

## 具体实施方式

[0030] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分的实施例,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,都应当属于本发明保护的范

围。

[0031] 需要说明的是,本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0032] 首先,在对本申请实施例进行描述的过程中出现的部分名词或术语适用于如下解释:

[0033] API:API(Application Programming Interface,应用程序编程接口)是一些预先定义的函数,目的是提供应用程序与开发人员基于某软件或硬件得以访问一组例程的能力,而又无需访问源码。

[0034] 日志系统:日志系统时用于记录硬盘、软件和系统问题的信息,同时还可以监视系统发生问题的原因。

[0035] 实施例1

[0036] 根据本发明实施例,提供了一种控流的方法实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0037] 本申请实施例一所提供的方法实施例可以在移动终端、计算机终端或者类似的运算装置中执行。以运行在计算机终端上为例,图2是根据本申请实施例1的一种控流的方法的计算机终端的硬件结构框图。如图2所示,计算机终端20可以包括一个或多个(图中仅示出一个)处理器202(处理器202可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器204、以及用于通信功能的传输装置206。本领域普通技术人员可以理解,图2所示的结构仅为示意,其并不对上述电子装置的结构造成限定。例如,计算机终端20还可包括比图2中所示更多或者更少的组件,或者具有与图2所示不同的配置。

[0038] 存储器204可用于存储应用软件的软件程序以及模块,如本发明实施例中的控流的方法对应的程序指令/模块,处理器202通过运行存储在存储器204内的软件程序以及模块,从而执行各种功能应用以及数据处理。存储器204可包括高速随机存储器,还可包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器204可进一步包括相对于处理器202远程设置的存储器,这些远程存储器可以通过网络连接至计算机终端20。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0039] 传输装置206用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机终端20的通信供应商提供的无线网络。在一个实例中,传输装置206包括一个网络适配器(Network Interface Controller, NIC),其可与其他网络设备相连从而可与互联网进行通讯。

[0040] 在上述运行环境下,本申请提供了如图3所示的控流的方法。图3是根据本发明实施例1的控流的方法的流程图,如图3所示的方法可以包括如下步骤:

[0041] 步骤S21,检测访问终端发起访问业务系统的访问请求,获取限流条件数据。

[0042] 在上述步骤中,访问终端可以是任意通过网络访问业务系统的终端,例如:智能手机、平板电脑、计算机等终端设备;业务系统可以是终端通过互联网访问的服务器提供的业务服务,例如:订单系统、物流系统等。

[0043] 结合图4所示的示例,在检测访问终端发起访问业务系统的访问请求,可以通过图4所示的业务系统来获取当前的限流条件数据,其中,本发明该实施例中的分析系统可以是一台独立的服务器/计算设备或耦合于其他服务器/计算设备中的应用软件,访问终端可以是一种具有应用功能的客户端,访问终端可以通过网关将访问请求透传至业务系统,然后通过业务系统或者进一步通过日志系统转发至分析系统,当前状态下,在客户端向网关发送访问请求时,此时网关可以从已有的流控规则中获取当前的流控阈值来对访问请求进行限流控制,在将访问请求成功透过网关发送至业务系统之后,此时分析系统就可以成功检测到访问终端发起的访问请求,从而触发获取限流条件数据。

[0044] 步骤S23,根据限流条件数据确定新的流控阈值。

[0045] 流控阈值用于表征网关所允许的最大的访问值,即在某一终端访问目标业务系统的访问值大于流控阈值的情况下,网关将禁止访问终端继续访问目标业务系统,其中,访问值可以是在预设时间内终端访问目标业务系统的次数、终端访问目标业务系统的频率、某一固定IP地址访问业务系统的次数或频率以及某一API访问业务系统的次数或频率。

[0046] 在上述步骤中,新的流控阈值为分析系统根据限流条件数据确定的,由于限流条件数据为动态数据,因此新的流控阈值也为动态数据。

[0047] 结合图4所示的示例,在一种可选的实施例中,可以通过分析系统来获得的新的流控阈值,其中,新的流控阈值可以是在当前流控阈值的基础上升高或者降低得到的流控阈值,分析系统将新的流控阈值会更新至当前的流控规则中。

[0048] 步骤S25,使用新的流控阈值更新当前的流控阈值,其中,可以通过图4所示的可选实施例中的分析系统来执行该步骤。

[0049] 在一种可选的实施例中,以业务系统为订单系统为例,在某一用户通过网关向目标订单系统发送访问请求的情况下,分析系统根据获取的限流条件数据,确定当前订单系统内部运行压力过大,需要降低流控阈值,以防过大的访问请求使订单系统崩溃,在分析系统得到新的流控阈值后,采用较低新的流控阈值来替代当前的流控阈值,网关设备则采用新的流控阈值来对用户的访问进行限流。

[0050] 此处需要说明的是,由于新的流控阈值为分析系统根据限流条件数据确定,并将当前的流控阈值替换为新的流控阈值,因此采用上述方案确定的流控阈值是根据限流条件数据变动的动态流控阈值。

[0051] 由于限流条件数据中可以包括当前的流控阈值,因此根据限流条件数据获得的新的流控阈值是与当前的限流条件数据对应的流控阈值,可以认为,动态的新的流控阈值能够与当前的限流条件数据相对应。

[0052] 容易注意到的是,当业务系统的运行状态发生变动(例如:某个API出现故障、业务系统运行压力过大等)或终端对业务系统进行异常访问时,分析系统根据限流条件数据的



变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端的访问请求进行限流;由于新的流控阈值与当前业务系统的运行状态(业务系统的运行状态体现于限流条件数据)相匹配,因此新的流控阈值也是与当前业务系统的运行状态相匹配的流控阈值,避免了业务系统的运行状态处于变动的状态但流控阈值为固定阈值导致的流控阈值与业务系统的运行状态不匹配的问题,从而避免了由于业务系统的运行状态和流控阈值不匹配导致的业务系统崩溃的问题。

[0053] 由此,本申请提供的上述实施例一的方案解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。

[0054] 在本申请上述实施例中,限流条件数据包括:当前的流控阈值,限流条件数据还包括:历史流控数据和业务系统的负载数据中的一个或二者的组合。

[0055] 容易注意到的是,上述实施例提供的方案可以单一的从历史流控数据或从业务系统的负载数据来得到新的流控阈值;也可以从历史流控数据和业务系统负载数据这两个维度的限流条件数据来得到新的流控阈值,兼顾业务系统的历史控流情况以及业务系统的负载情况,从而使得分析系统既能够以历史流控数据为基础得到新流控阈值,又能够获取与业务系统的负载情况相对应新的流控阈值。

[0056] 在本申请上述实施例中,历史流控数据可以包括:统计的限流次数和限流频率中的一个或二者的组合,其中,负载数据包括:任务处理值或CPU占用率。

[0057] 在一种可选的实施例中,结合图4所示的示例,分析系统可以从日志系统中进行日志采集,得到上述限流条件数据中的历史流控数据,同时还需访问当前应用系统负载来获取到各个业务系统的业务负载数据。

[0058] 在上述实施例中,历史流控数据所包括的限流次数和/或限流频率可以是某一终端被限流过的次数和/或在预设时间内某一终端被限流过的频率;业务服务集群可以是业务系统中运行的多个服务器构成的服务器集群,多个服务器用于同时处理业务系统的访问请求;任务处理值用于表示业务系统当前执行任务的任务值,CPU占用率用于表示当前业务系统所运行的程序占用的CPU的资源,当CPU占用率越高时,可以认为CPU已经被占用的资源越多,能够继续用来处理新的访问请求的资源越少。

[0059] 在本申请上述实施例中,流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,步骤S25,根据限流条件数据确定新的流控阈值,包括:

[0060] 步骤S251,判断统计的限流次数是否超过预定次数,和/或限流频率是否超出预定频率值。

[0061] 步骤S253,如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值。

[0062] 此处需要说明的是,访问用户的流控阈值可以是用户对某一固定业务系统的访问频率或访问次数;访问终端地址的流控阈值可以是某一访问终端地址对某一固定业务系统的访问频率或访问次数;API接口的流控阈值可以是API接口访问某一固定业务系统的访问频率或访问次数。

[0063] 在一种可选的实施例中,以预定次数为1000次,限流频率为20次/秒为例,在流控

阈值为访问用户的流控阈值的示例中,该访问用户访问某一业务系统超过流控阈值的次数为1243次,平均每秒超过控流阈值的次数为17.3次,因此该访问用户的流控阈值超过了预定次数,在上述情况下,将当前的流控阈值降低至第一预定流控阈值,以限制上述访问用户的访问请求。

[0064] 此处需要说明的是,在上述步骤中,限流次数和限流频率均为判断的条件,即限流次数和限流频率中任意一个值超过了对应的预设值,都将会引起流控阈值的下降。

[0065] 此处还需要说明的是,本实施例上述步骤采用判断限流次数的方式来确定是否降低流控阈值,目的在于在某一访问用户、访问终端或API接口被多次限流的情况下,该访问用户、访问终端或API接口为其访问的业务系统带了了较大的访问压力,并可以怀疑上述访问为恶意访问,为了防止上述访问用户、访问终端或API接口继续对业务系统进行高频率正常访问或恶意访问,分析系统在当前流控阈值的基础上降低新的流控阈值,以达到进一步限流的技术效果。

[0066] 在本申请上述实施例中,业务系统包括多个业务服务器,每个业务服务器的负载数据实时更新,步骤S25,根据限流条件数据确定新的流控阈值,包括:

[0067] 步骤S255,如果业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值,则将流控阈值下降至第二预定流控阈值。

[0068] 步骤S257,如果业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值,则将流控阈值提高至第三预定流控阈值。

[0069] 在上述步骤中,第一预定值可以是业务处理器在正常运行的基础上能够处理的较大任务处理值,第二预定值可以是CPU在正常运行的基础上能够使用的较大的CPU占用率,因此在业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值的情况下,如果继续使用当前的流控阈值,即使部分业务服务器或CPU超出负荷运行,则会出现部分业务服务器或CPU出现故障或崩溃的事件。

[0070] 第三预定值可以是业务服务器较为空闲时的任务处理直,第四预定值可以是CPU较为空闲时的CPU占用率,即在业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值的情况下,可以认为业务处理器集群或CPU为较为空闲的状态,仍然以当前的流控阈值进行限流会产生一定的资源浪费,业务服务器或CPU以较空闲的状态运行,而通过网关对业务系统进行访问的用户、地址或API均由于限流不能对业务系统进行访问,因此分析系统根据上述限流条件数据提高流控阈值,以使其他用户、地址或API能够尽快对业务系统进行访问。

[0071] 仍旧结合图4所示的示例,在一种可选的实施例中,上述方案可以实现分析系统获取动态的流控策略,其中,上述动态的流控策略中包含了根据当前可以获取到的历史流控数据和/或业务系统的负载数据得到新的流控数据的方法,或预先设定的与历史流控数据和/或业务系统负载数据对应的新的流控阈值。(例如:以4核4G的业务服务器为例,如果50%以上的业务服务器均出现 $load > 4$ ,或 $cpu > 100\%$ 的情况,则将新的流控阈值在当前流控阈值的基础上下降10%,如果50%以上的业务服务器均出现 $load < 0.5$ ,或 $cpu < 50\%$ 的情况,,则将新的流控阈值在当前阈值的基础上调10%。)

[0072] 由此可知,本申请上述步骤所实现的实施例可以实现,分析系统根据获取的历史流量数据和业务系统负责数据对流控规则中设置的流控策略进行动态调整。

[0073] 下面采用一种可选的实施例对上述示例进行详细说明,以第一预定值为20000,第二预定值为85%,第三预定值为8000,第四预定值为50%为例,在获取业务系统中的20个业务服务器有12个业务服务器的任务处理值大于200000,CPU的占用率为83%的情况下,分析系统将设置小于当前流控阈值的新的流控阈值;在获取业务系统中的20个业务服务器有3个业务服务器的任务处理值小于8000,CPU的占用率为46%的情况下,分析系统将设置大于当前流控阈值的新的流控阈值。

[0074] 此处需要说明的是,业务系统中包括多个业务服务器,在对业务服务器的任务处理值与第三预定值进行比较时,将业务系统中的每个业务服务器的任务处理值与第三预定值进行比较,其中,分析系统存储有预定数量,当任务处理值大于等于第一预定值的业务服务器的数量超过预定数量时,可以认为当前的流控阈值过大,业务处理器集群和CPU不能够处理当前的访问请求。

[0075] 此处还需要说明的是每个业务服务器的负载数据是实时更新的,分析系统在获取业务服务器的任务处理值时,获取的均为业务服务器最新的任务处理值。

[0076] 在本申请上述实施例中,限流条件数据至少还包括:预先设定的限流控制优先级,步骤S25,根据限流条件数据确定新的流控阈值,包括:

[0077] 步骤S259,在根据历史流控数据确定将流控阈值调整至第一预定流控阈值,并根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为历史流控数据的优先级高于负载数据的优先级,则将第一预定流控阈值作为新的流控阈值。

[0078] 在本申请上述步骤中,分析系统根据历史流控数据确定将流控阈值调整至第一预定流控阈值,根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值,而新的控流阈值只能是唯一控流阈值,因此需要对限流控制优先级做判断,选择限流控制优先级较高的数据对应的控流阈值作为新的控流阈值。

[0079] 在一种可选的实施例中,在业务系统内部服务器运行稳定的情况下,可以设置历史流控数据的优先级为较高优先级;在另一种可选的实施例中,还可以选择较大的流控阈值为新的流控阈值。

[0080] 在本申请上述实施例中,步骤S259,如果限流控制优先级为负载数据的优先级高于历史流控数据的优先级,将第二预定流控阈值或第三预定流控阈值作为新的流控阈值。

[0081] 此处需要说明的是,关于历史流控数据的优先级和负载数据的优先级,本申请不做具体限定,可以根据实际需求选择历史流控数据和负载数据的优先级。

[0082] 在一种可选的实施例中,在业务系统内部服务器压力较大的情况下,可以设置负载数据的优先级为较高优先级;在另一种可选的实施例中,还可以选择较大的流控阈值为新的流控阈值。

[0083] 在本申请上述实施例中,步骤S21,在分析系统读取限流条件数据之前,上述方法还包括:

[0084] 步骤S27,访问终端通过网关将访问请求发送给业务系统,其中,网关用于使用当前的流控阈值对访问请求进行限流控制。

[0085] 步骤S29,业务系统将访问请求和进行限流控制时得到的流控数据写入日志系统,其中,日志系统用于记录历史流控数据;其中,在日志系统写入访问请求的情况下,日志系

系统将历史流控数据发送至分析系统。

[0086] 在本申请上述实施例中,日志系统用于记录和存储历史流控数据,以使分析系统根据历史流控数据得到新的流控阈值。在一种可选的实施例中,日志系统以表格的方式呈现历史流控数据。

[0087] 结合图4所示的示例可知,上述步骤S27和步骤29可以实现访问终端(可以是客户端)通过网关将访问请求透传至业务系统,该业务系统可以是部署了应用软件的应用系统,在业务系统将访问请求和进行限流控制时得到的流控数据直接写入日志系统之后,或日至系统通过对应用系统进行日志采集得到访问请求和进行限流控制时得到的流控数据之后,可以将历史流控数据发送至分析系统,从而分析系统可以通过接收到的日志内容得到历史流控数据。

[0088] 此处需要说明的是,在本申请中的日志系统用于记录历史流控数据,但并不去限定本发明中的日志系统对其他数据的记录。

[0089] 在本申请上述实施例中,步骤S25,在分析系统使用新的流控阈值更新当前的流控阈值之后,方法还包括:

[0090] 步骤S211,网关从分析系统获取到更新后的流控阈值,或者分析系统将更新后的流控阈值发送给网关,使得网关使用更新后的流控阈值对访问终端进行限流控制。

[0091] 在一种可选的实施例中,分析系统根据限流条件数据获得新的流控阈值,网关从分析系统中获取新的流控阈值,并使用新的流控阈值对访问终端进行限流;在另一种可选的实施例中,分析系统将根据限流条件数据获得新的流控阈值发送至网关,以使网关对访问终端进行限流控制。

[0092] 在本申请上述实施例中,网关对访问请求进行如下任意一个或多个处理:协议转换处理、参数转换处理、访问控制处理和限流控制处理。

[0093] 在本申请上述实施例中,结合图4所示的实施例,上述实施例可以实现通过网关在经过协议转换、参数转换、访问控制、流量控制等处理后将访问请求发送至业务系统,其中,协议处理用于使协议不相同的网络进行互连,参数转换处理用于对不同的通讯格式、数据格式或语言的进行转换,访问控制可以是对访问终端进行鉴权,以确认该访问终端是否有权访问该业务系统,或该访问终端有权访问的信息项,限制流量处理用于限定用户终端在访问业务系统是速率,以防止瞬时访问高峰使业务系统和CPU崩溃,从而确保业务系统正常运行。

[0094] 综上可知,本申请提供的实施例一中的控流的方法可以实现当业务系统的运行状态发生变动或终端对业务系统进行异常访问时,分析系统根据限流条件数据的变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端发起的新的访问请求进行限流,从而避免了由于业务系统的运行状态和流控阈值不匹配导致的业务系统崩溃的问题。下面结合图5所示,就本申请的方案应用在一种浏览器客户端的访问进行控流的应用场景所实现的功能进行详细描述。

[0095] 图5是根据本申请实施例1的一种可选的控流的方法的信息交互图,结合图5所示,包括如下步骤:

[0096] 步骤S501,访问终端110向网关116发送访问请求。

[0097] 在上述步骤中,访问终端110可以是浏览器客户端的载体,在用户打开浏览器进行

访问时,可以实现访问终端110向网关116发送访问请求的功能。

[0098] 步骤S503,网关116经过协议转换、参数转换、访问控制和流量控制后将访问请求发送至业务系统终端114。

[0099] 步骤S505,业务系统终端114将访问请求发送至日志系统终端118。

[0100] 步骤S507,日志系统终端118将历史流控数据发送至分析系统终端112。

[0101] 步骤S509,业务服务器1110将负载数据发送至分析系统终端112。

[0102] 步骤S5011,分析系统终端112根据历史流控数据以及业务系统中每个业务服务器1110的负载数据确定新的流控阈值。

[0103] 在上述步骤中,在分析系统根据历史流控数据得到的新的流控阈值与根据业务服务器1110获得的流控阈值不相同的情况下,需要根据预先设定的优先级来确定新的流控阈值,在历史流控数据具有较高优先级的情况下,确认分析系统根据历史流控数据得到的新的流控阈值为新的流控阈值,否则,确认分析系统根据业务服务器1110获得的流控阈值为新的流控阈值。

[0104] 步骤S5013,网关116获取分析系统终端112确认的新的流控阈值,根据新的流控阈值对访问终端110进行限流控制。

[0105] 在上述步骤中,分析系统将新的流控阈值发送至网关116,使新的流控阈值替换当前流控阈值,网关116使用新的流控阈值对访问终端110的访问请求进行限流控制,使得控流的方法与当前限流条件相对应。

[0106] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本发明所必须的。

[0107] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0108] 实施例2

[0109] 根据本发明实施例,还提供了一种控流的装置,图6是根据本申请实施例2的一种控流的装置的示意图,如图6所示,该装置包括:分析模块60,确定模块62和更新模块64。

[0110] 分析模块60,用于检测访问终端发起访问业务系统的访问请求,获取限流条件数据;确定模块62,用于根据限流条件数据确定新的流控阈值;更新模块64,用于使用新的流控阈值更新当前的流控阈值。

[0111] 在上述步骤中,访问终端可以是任意通过网络访问业务系统的终端,例如:智能手机、平板电脑、计算机等终端设备;业务系统可以是终端通过互联网访问的服务器提供的业务服务,例如:订单系统、物流系统等。

[0112] 流控阈值用于表征网关所允许的最大的访问值,即在某一终端访问目标业务系统

的访问值大于流控阈值的情况下,网关将禁止访问终端继续访问目标业务系统,其中,访问值可以是在预设时间内终端访问目标业务系统的次数、终端访问目标业务系统的频率、某一固定IP地址访问业务系统的次数或频率以及某一API访问业务系统的次数或频率。

[0113] 在一种可选的实施例中,以业务系统为订单系统为例,在某一用户通过网关向目标订单系统发送访问请求的情况下,分析系统根据获取的限流条件数据,确定当前订单系统内部运行压力过大,需要降低流控阈值,以防过大的访问请求使订单系统崩溃,在分析系统得到新的流控阈值后,采用较低新的流控阈值来替代当前的流控阈值,网关设备则采用新的流控阈值来对用户的访问进行限流。

[0114] 此处需要说明的是,由于新的流控阈值为分析系统根据限流条件数据确定,并将当前的流控阈值替换为新的流控阈值,因此采用上述方案确定的流控阈值是根据限流条件数据变动的动态流控阈值。

[0115] 由于限流条件数据中至少包括了当前的流控阈值,因此根据限流条件获得的新的流控阈值是与当前的限流条件对应的流控阈值,可以认为,动态的新的流控阈值能够与当前的限流条件数据相对应。

[0116] 容易注意到的是,当业务系统的运行状态发生变动(例如:某个API出现故障、业务系统运行压力过大等)或终端对业务系统进行异常访问时,分析系统根据限流条件数据的变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端的访问请求进行限流;由于新的流控阈值与当前业务系统的运行状态(业务系统的运行状态体现于限流条件数据)相匹配,因此新的流控阈值也是与当前业务系统的运行状态相匹配的流控阈值,避免了业务系统的运行状态处于变动的状态但流控阈值为固定阈值导致的流控阈值与业务系统的运行状态不匹配的问题,从而避免了由于业务系统的运行状态和流控阈值不匹配导致的业务系统崩溃的问题。

[0117] 由此,本申请提供的上述实施例一的方案解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。

[0118] 在本申请上述实施例中,上述限流条件包括:当前的流控阈值,上述限流条件数据至少还包括:历史流控数据和业务系统的负载数据中的一个或二者的组合。

[0119] 容易注意到的是,上述实施例提供的方案可以单一的从历史流控数据或从业务系统的负载数据来得到新的流控阈值;也可以从历史流控数据和业务系统这两个维度的限流条件数据来得到新的流控阈值,兼顾业务系统的历史控流情况以及业务系统的负载情况,从而使得分析系统既能够以历史流控数据为基础得到新流控阈值,又能够获取与业务系统的负载情况相对应新的流控阈值。

[0120] 在本申请上述实施例中,历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合,负载数据包括:任务处理值或CPU占用率。

[0121] 在本申请上述实施例中,流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,其中,结合图7所示,上述确定模块62包括:

[0122] 判断模块70,用于判断统计的限流次数是否超过预定次数,和/或限流频率是否超出预定频率值;第一下降模块72,用于如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值。

[0123] 此处需要说明的是,访问用户的流控阈值可以是用户对某一固定业务系统的访问频率或访问次数;访问终端地址的流控阈值可以是某一访问终端地址对某一固定业务系统的访问频率或访问次数;API接口的流控阈值可以是API接口访问某一固定业务系统的访问频率或访问次数。

[0124] 在一种可选的实施例中,以预定次数为1000次,限流频率为20次/秒为例,在流控阈值为访问用户的流控阈值的示例中,该访问用户访问某一业务系统超过流控阈值的次数为1243次,平均每秒超过控流阈值的次数为17.3次,因此该访问用户的流控阈值超过了预定次数,在上述情况下,将当前的流控阈值降低至第一预定流控阈值,以限制上述访问用户的访问请求。

[0125] 此处需要说明的是,在上述步骤中,限流次数和限流频率均为判断的条件,即限流次数和限流频率中任意一个值超过了对应的预设值,都将会引起流控阈值的下降。

[0126] 此处还需要说明的是,本实施例上述步骤采用判断限流次数的方式来确定是否降低流控阈值,目的在于在某一访问用户、访问终端或API接口被多次限流的情况下,该访问用户、访问终端或API接口为其访问的业务系统带了了较大的访问压力,并可以怀疑上述访问为恶意访问,为了防止上述访问用户、访问终端或API接口继续对业务系统进行高频率正常访问或恶意访问,分析系统在当前流控阈值的基础上降低新的流控阈值,以达到进一步限流的技术效果。

[0127] 在本申请上述实施例中,业务系统包括多个业务服务器,每个业务服务器的负载数据实时更新,其中,结合图8所示,上述确定模块62包括:

[0128] 第二下降模块80,用于如果业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值,则将流控阈值下降至第二预定流控阈值;提高模块82,用于如果业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值,则将流控阈值提高至第三预定流控阈值。

[0129] 在本申请上述实施例中,限流条件数据至少还包括:预先设定的限流控制优先级,其中,结合图9所示,上述确定模块62还包括:

[0130] 第一确认模块90,用于在根据历史流控数据确定将流控阈值调整至第一预定流控阈值,并根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为历史流控数据的优先级高于负载数据的优先级,则将第一预定流控阈值作为新的流控阈值。

[0131] 本申请上述步骤中,分析系统根据历史流控数据确定将流控阈值调整至第一预定流控阈值,根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值,而新的控流阈值只能是唯一控流阈值,因此需要对限流控制优先级做判断,选择限流控制优先级较高的数据对应的控流阈值作为新的控流阈值。

[0132] 在一种可选的实施例中,在业务系统内部服务器运行稳定的情况下,可以设置历史流控数据的优先级为较高优先级;在另一种可选的实施例中,还可以选择较大的流控阈值为新的流控阈值。

[0133] 在本申请上述实施例中,结合图10所示,确定模块62还包括:

[0134] 第二确认模块100,用于如果限流控制优先级为负载数据的优先级高于历史流控

数据的优先级,将第二预定流控阈值或第三预定流控阈值作为新的流控阈值。

[0135] 此处需要说明的是,关于历史流控数据的优先级和负载数据的优先级,本申请不做具体限定,可以根据实际需求选择历史流控数据和负载数据的优先级。

[0136] 在一种可选的实施例中,在业务系统内部服务器压力较大的情况下,可以设置负载数据的优先级为较高优先级;在另一种可选的实施例中,还可以选择较大的流控阈值为新的流控阈值。

[0137] 第三预定值可以是业务服务器较为空闲时的任务处理直,第四预定值可以是CPU较为空闲时的CPU占用率,即在业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值的情况下,可以认为业务处理器集群或CPU为较为空闲的状态,仍然以当前的流控阈值进行限流会产生一定的资源浪费,业务服务器或CPU以较空闲的状态运行,而通过网关对业务系统进行访问的用户、地址或API均由于限流不能对业务系统进行访问,因此分析系统根据上述限流条件数据提高流控阈值,以使其他用户、地址或API能够尽快对业务系统进行访问。

[0138] 在一种可选的实施例中,以第一预定值为20000,第二预定值为85%,第三预定值为8000,第四预定值为50%为例,在获取业务系统中的20个业务服务器有12个业务服务器的任务处理值大于200000,CPU的占用率为83%的情况下,分析系统将设置小于当前流控阈值的新的流控阈值;在获取业务系统中的20个业务服务器有3个业务服务器的任务处理值小于8000,CPU的占用率为46%的情况下,分析系统将设置大于当前流控阈值的新的流控阈值。

[0139] 此处需要说明的是,业务系统中包括多个业务服务器,在对业务服务器的任务处理值与第三预定值进行比较时,将业务系统中的每个业务服务器的任务处理值与第三预定值进行比较,其中,分析系统存储有预定数量,当任务处理值大于等于第一预定值的业务服务器的数量超过预定数量时,可以认为当前的流控阈值过大,业务处理器集群和CPU不能够处理当前的访问请求。

[0140] 此处还需要说明的是每个业务服务器的负载数据是实时更新的,分析系统在获取业务服务器的任务处理值时,获取的均为业务服务器最新的任务处理值。

[0141] 实施例3

[0142] 根据本申请实施例,还提供了一种控流的系统,如图11所示,该系统可以包括:访问终端110和分析系统终端112。

[0143] 访问终端110,向业务系统发起访问请求。

[0144] 分析系统终端112,与访问终端通讯,检测访问终端发起访问业务系统终端的访问请求,获取限流条件数据,根据限流条件数据确定新的流控阈值,并使用新的流控阈值更新当前的流控阈值,其中,限流条件数据至少包括:当前的流控阈值。

[0145] 在一种可选的实施例中,以业务系统终端114为订单系统为例,在某一用户通过网关116向目标订单系统发送访问请求的情况下,分析系统终端根据获取的限流条件数据,确定当前订单系统内部运行压力过大,需要降低流控阈值,以防过大的访问请求使订单系统崩溃,在分析系统得到新的流控阈值后,采用较低新的流控阈值来替代当前的流控阈值,网关设备则采用新的流控阈值来对用户的访问进行限流。

[0146] 此处需要说明的是,由于新的流控阈值为分析系统根据限流条件数据确定,并将



当前的流控阈值替换为新的流控阈值,因此采用上述方案确定的流控阈值是根据限流条件数据变动的动态流控阈值。

[0147] 由于限流条件数据中至少包括了当前的流控阈值,因此根据限流条件获得的新的流控阈值是与当前的限流条件对应的流控阈值,可以认为,动态的新的流控阈值能够与当前的限流条件数据相对应。

[0148] 容易注意到的是,当业务系统终端的运行状态发生变动(例如:某个API出现故障、业务系统终端运行压力过大等)或终端对业务系统终端进行异常访问时,分析系统根据限流条件数据的变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端的访问请求进行限流;由于新的流控阈值与当前业务系统终端的运行状态(业务系统终端的运行状态体现于限流条件数据)相匹配,因此新的流控阈值也是与当前业务系统终端的运行状态相匹配的流控阈值,避免了业务系统终端的运行状态处于变动的状态但流控阈值为固定阈值导致的流控阈值与业务系统终端的运行状态不匹配的问题,从而避免了由于业务系统终端的运行状态和流控阈值不匹配导致的业务系统终端崩溃的问题。

[0149] 由此,本申请提供的上述实施例3的方案解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统终端的运行状态不符导致业务系统终端崩溃的技术问题。

[0150] 在本申请上述实施例中,上述限流条件包括:所述当前的流控阈值,所述限流条件数据还包括:历史流控数据和所述业务系统的负载数据中的一个或二者的组合。

[0151] 容易注意到的是,上述实施例提供的方案可以单一的从历史流控数据或从业务系统的负载数据来得到新的流控阈值;也可以从历史流控数据和业务系统负载数据这两个维度的限流条件数据来得到新的流控阈值,兼顾业务系统的历史控流情况以及业务系统的负载情况,从而使得分析系统既能够以历史流控数据为基础得到新流控阈值,又能够获取与业务系统的负载情况相对应新的流控阈值。

[0152] 在本申请上述实施例中,上述历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合,所述负载数据包括:任务处理值或CPU占用率。

[0153] 在本申请上述实施例中,分析系统终端还用于判断统计的限流次数是否超过预定次数,和/或限流频率是否超出预定频率值,如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值。

[0154] 在上述系统中,历史流控数据由日志系统终端118对访问请求日志进行采集得到,并将历史流控数据发送至分析系统终端。

[0155] 此处需要说明的是,访问用户的流控阈值可以是用户对某一固定业务系统终端的访问频率或访问次数;访问终端地址的流控阈值可以是某一访问终端地址对某一固定业务系统终端的访问频率或访问次数;API接口的流控阈值可以是API接口访问某一固定业务系统终端的访问频率或访问次数。

[0156] 在一种可选的实施例中,以预定次数为1000次,限流频率为20次/秒为例,在流控阈值为访问用户的流控阈值的示例中,该访问用户访问某一业务系统终端超过流控阈值的次数为1243次,平均每秒超过控流阈值的次数为17.3次,因此该访问用户的流控阈值超过了预定次数,在上述情况下,将当前的流控阈值降低至第一预定流控阈值,以限制上述访问用户的访问请求。

[0157] 此处需要说明的是,在上述系统中,限流次数和限流频率均为判断的条件,即限流次数和限流频率中任意一个值超过了对应的预设值,都将会引起流控阈值的下降。

[0158] 此处还需要说明的是,本实施例上述系统采用判断限流次数的方式来确定是否降低流控阈值,目的在于在某一访问用户、访问终端或API接口被多次限流的情况下,该访问用户、访问终端或API接口为其访问的业务系统终端带了了较大的访问压力,并可以怀疑上述访问为恶意访问,为了防止上述访问用户、访问终端或API接口继续对业务系统终端进行高频率正常访问或恶意访问,分析系统在当前流控阈值的基础上降低新的流控阈值,以达到进一步限流的技术效果。

[0159] 在本申请上述实施例中,分析系统终端还用于如果业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值,则将流控阈值下降至第二预定流控阈值,如果业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值,则将流控阈值提高至第三预定流控阈值。

[0160] 在上述系统中,业务服务器的负载数据由业务服务器1110发送至分析系统终端。

[0161] 在上述系统中,第一预定值可以是业务处理器在正常运行的基础上能够处理的较大任务处理值,第二预定值可以是CPU在正常运行的基础上能够使用的较大的CPU占用率,因此在业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值的情况下,如果继续使用当前的流控阈值,即使部分业务服务器或CPU超出负荷运行,则会出现部分业务服务器或CPU出现故障或崩溃的事件。

[0162] 第三预定值可以是业务服务器较为空闲时的任务处理直,第四预定值可以是CPU较为空闲时的CPU占用率,即在业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值的情况下,可以认为业务处理器集群或CPU为较为空闲的状态,仍然以当前的流控阈值进行限流会产生一定的资源浪费,业务服务器或CPU以较空闲的状态运行,而通过网关对业务系统终端进行访问的用户、地址或API均由于限流不能对业务系统终端进行访问,因此分析系统根据上述限流条件数据提高流控阈值,以使其他用户、地址或API能够尽快对业务系统终端进行访问。

[0163] 此处需要说明的是,业务系统中包括多个业务服务器,在对业务服务器的任务处理值与第三预定值进行比较时,将业务系统中的每个业务服务器的任务处理值与第三预定值进行比较,其中,分析系统存储有预定数量,当任务处理值大于等于第一预定值的业务服务器的数量超过预定数量时,可以认为当前的流控阈值过大,业务处理器集群和CPU不能够处理当前的访问请求。

[0164] 此处还需要说明的是每个业务服务器的负载数据是实时更新的,分析系统在获取业务服务器的任务处理值时,获取的均为业务服务器最新的任务处理值。

[0165] 在本申请上述实施例中,上述系统还包括:

[0166] 网关116,分别于与访问终端和分析系统终端通讯,用于使用当前的流控阈值对访问请求进行限流控制,并从分析系统终端获取到更新后的流控阈值,或者接收分析系统终端发送的更新后的流控阈值,使用更新后的流控阈值对访问终端进行限流控制。

[0167] 在一种可选的实施例中,分析系统根据限流条件数据获得新的流控阈值,网关从分析系统中获取新的流控阈值,并使用新的流控阈值对访问终端进行限流;在另一种可选

的实施例中,分析系统将根据限流条件数据获得新的流控阈值发送至网关,以使网关对访问终端进行限流控制。

[0168] 实施例4

[0169] 本发明的实施例可以提供一种计算机终端,该计算机终端可以是计算机终端群中的任意一个计算机终端设备。可选地,在本实施例中,上述计算机终端也可以替换为移动终端等终端设备。

[0170] 可选地,在本实施例中,上述计算机终端可以位于计算机网络的多个网络设备中的至少一个网络设备。

[0171] 在本实施例中,上述计算机终端可以执行应用程序的控流的方法中以下步骤的程序代码:检测访问终端发起访问业务系统的访问请求,获取限流条件数据;根据限流条件数据确定新的流控阈值;使用新的流控阈值更新当前的流控阈值。

[0172] 可选地,图12是根据本发明实施例的一种计算机终端的结构框图。如图12所示,该计算机终端1200可以包括:一个或多个(图中仅示出一个)处理器1202、存储器1204、传输装置1206、访问终端110和业务系统终端114。

[0173] 其中,存储器可用于存储软件程序以及模块,如本发明实施例中的控流的方法和装置对应的程序指令/模块,处理器通过运行存储在存储器内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的控流的方法。存储器可包括高速随机存储器,还可以包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器可进一步包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至终端A。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0174] 处理器可以通过传输装置调用存储器存储的信息及应用程序,以执行下述步骤:检测访问终端发起访问业务系统的访问请求,获取限流条件数据;根据限流条件数据确定新的流控阈值;使用新的流控阈值更新当前的流控阈值。

[0175] 可选的,上述处理器还可以执行如下步骤的程序代码:限流条件数据包括:当前的流控阈值,限流条件数据还包括:历史流控数据和业务系统的负载数据中的一个或二者的组合。

[0176] 可选的,上述处理器还可以执行如下步骤的程序代码:历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合,负载数据包括:任务处理值或CPU占用率。

[0177] 可选的,上述处理器还可以执行如下步骤的程序代码:流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,其中,根据限流条件数据确定新的流控阈值,包括:判断统计的限流次数是否超过预定次数,和/或限流频率是否超出预定频率值;如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值。

[0178] 可选的,上述处理器还可以执行如下步骤的程序代码:业务系统包括多个业务服务器,每个业务服务器的负载数据实时更新,其中,根据限流条件数据确定新的流控阈值,包括:

[0179] 如果业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值,则将流控阈值下降至第二预定流控阈值;如果业务系统

中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值,则将流控阈值提高至第三预定流控阈值。

[0180] 可选的,上述处理器还可以执行如下步骤的程序代码:限流条件数据至少还包括:预先设定的限流控制优先级,其中,根据限流条件数据确定新的流控阈值,包括:在根据历史流控数据确定将流控阈值调整至第一预定流控阈值,并根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为历史流控数据的优先级高于负载数据的优先级,则将第一预定流控阈值作为新的流控阈值。

[0181] 可选的,上述处理器还可以执行如下步骤的程序代码:如果限流控制优先级为负载数据的优先级高于历史流控数据的优先级,将第二预定流控阈值或第三预定流控阈值作为新的流控阈值。

[0182] 可选的,上述处理器还可以执行如下步骤的程序代码:访问终端通过网关将访问请求发送给业务系统,其中,网关用于使用当前的流控阈值对访问请求进行限流控制;业务系统将访问请求和进行限流控制时得到的流控数据写入日志系统,其中,日志系统用于记录历史流控数据;其中,在日志系统写入访问请求的情况下,日志系统将历史流控数据发送至分析系统。

[0183] 可选的,上述处理器还可以执行如下步骤的程序代码:网关从分析系统获取到更新后的流控阈值,或者分析系统将更新后的流控阈值发送给网关,使得网关使用更新后的流控阈值对访问终端进行限流控制。

[0184] 可选的,上述处理器还可以执行如下步骤的程序代码:网关对访问请求进行如下任意一个或多个处理:协议转换处理、参数转换处理、访问控制处理和限流控制处理。

[0185] 由于限流条件数据中至少包括了当前的流控阈值,因此根据限流条件获得的新的流控阈值是与当前的限流条件对应的流控阈值,可以认为,动态的新的流控阈值能够与当前的限流条件数据相对应。

[0186] 容易注意到的是,当业务系统的运行状态发生变动(例如:某个API出现故障、业务系统运行压力过大等)或终端对业务系统进行异常访问时,分析系统根据限流条件数据的变动发生变动,从而得到新的流控阈值,网关采用新的流控阈值对终端的访问请求进行限流;由于新的流控阈值与当前业务系统的运行状态(业务系统的运行状态体现于限流条件数据)相匹配,因此新的流控阈值也是与当前业务系统的运行状态相匹配的流控阈值,避免了业务系统的运行状态处于变动的状态但流控阈值为固定阈值导致的流控阈值与业务系统的运行状态不匹配的问题,从而避免了由于业务系统的运行状态和流控阈值不匹配导致的业务系统崩溃的问题。

[0187] 由此,本申请提供的上述实施例的方案解决了现有技术中由于限流策略为固定限流策略,使得在特殊情况下固定的限流策略与业务系统的运行状态不符导致业务系统崩溃的技术问题。本领域普通技术人员可以理解,图12所示的结构仅为示意,计算机终端也可以是智能手机(如Android手机、iOS手机等)、平板电脑、掌上电脑以及移动互联网设备(Mobile Internet Devices,MID)、PAD等终端设备。图12其并不对上述电子装置的结构造成限定。例如,计算机终端1200还可包括比图12中所示更多或者更少的组件(如网络接口、显示装置等),或者具有与图12所示不同的配置。

[0188] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令终端设备相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:闪存盘、只读存储器(Read-Only Memory,ROM)、随机存取器(Random Access Memory,RAM)、磁盘或光盘等。

[0189] 实施例5

[0190] 本发明的实施例还提供了一种存储介质。可选地,在本实施例中,上述存储介质可以用于保存上述实施例一所提供的控流的方法所执行的程序代码。

[0191] 可选地,在本实施例中,上述存储介质可以位于计算机网络中计算机终端群中的任意一个计算机终端中,或者位于移动终端群中的任意一个移动终端中。

[0192] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:检测访问终端发起访问业务系统的访问请求,获取限流条件数据;根据限流条件数据确定新的流控阈值;使用新的流控阈值更新当前的流控阈值。

[0193] 可选的,上述处理器还可以执行如下步骤的程序代码:限流条件数据包括:当前的流控阈值,限流条件数据还包括:历史流控数据和业务系统的负载数据中的一个或二者的组合。

[0194] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:历史流控数据包括:统计的限流次数和限流频率中的一个或二者的组合,负载数据包括:任务处理值或CPU占用率。

[0195] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:流控阈值为如下任意一个数据值:访问用户的流控阈值、访问终端地址的流控阈值、API接口的流控阈值,其中,根据限流条件数据确定新的流控阈值,包括:判断统计的限流次数是否超过预定次数,和/或限流频率是否超出预定频率值;如果统计的限流次数超过预定次数,和/或限流频率超出预定频率值,则将流控阈值下降至第一预定流控阈值。

[0196] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:业务系统包括多个业务服务器,每个业务服务器的负载数据实时更新,其中,根据限流条件数据确定新的流控阈值,包括:如果业务系统中超过预定数量的业务服务器的任务处理值大于等于第一预定值,或CPU占用率大于等于第二预定值,则将流控阈值下降至第二预定流控阈值;如果业务系统中超过预定数量的业务服务器的任务处理值小于等于第三预定值,或CPU占用率小于等于第四预定值,则将流控阈值提高至第三预定流控阈值。

[0197] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:限流条件数据至少还包括:预先设定的限流控制优先级,其中,根据限流条件数据确定新的流控阈值,包括:在根据历史流控数据确定将流控阈值调整至第一预定流控阈值,并根据业务系统的负载数据确定将流控阈值调整至第二预定流控阈值或第三预定流控阈值的情况下,如果限流控制优先级为历史流控数据的优先级高于负载数据的优先级,则将第一预定流控阈值作为新的流控阈值。

[0198] 可选的,上述处理器还可以执行如下步骤的程序代码:如果限流控制优先级为负载数据的优先级高于历史流控数据的优先级,将第二预定流控阈值或第三预定流控阈值作为新的流控阈值。

[0199] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:访

问终端通过网关将访问请求发送给业务系统,其中,网关用于使用当前的流控阈值对访问请求进行限流控制;业务系统将访问请求和进行限流控制时得到的流控数据写入日志系统,其中,日志系统用于记录历史流控数据;其中,在日志系统写入访问请求的情况下,日志系统将历史流控数据发送至分析系统。

[0200] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:网关从分析系统获取到更新后的流控阈值,或者分析系统将更新后的流控阈值发送给网关,使得网关使用更新后的流控阈值对访问终端进行限流控制。

[0201] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:网关对访问请求进行如下任意一个或多个处理:协议转换处理、参数转换处理、访问控制处理和限流控制处理。

[0202] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0203] 在本发明的上述实施例中,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0204] 在本申请所提供的几个实施例中,应该理解到,所揭露的技术内容,可通过其它的方式实现。其中,以上所描述的装置实施例仅仅是示意性的,例如所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,单元或模块的间接耦合或通信连接,可以是电性或其它的形式。

[0205] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0206] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0207] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0208] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

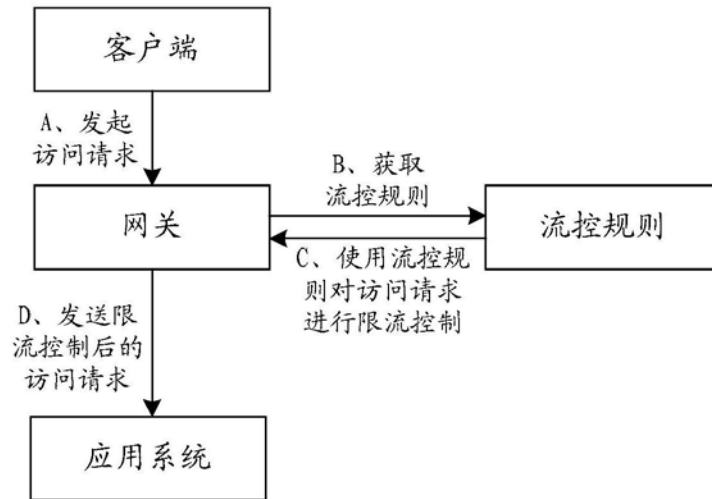


图1

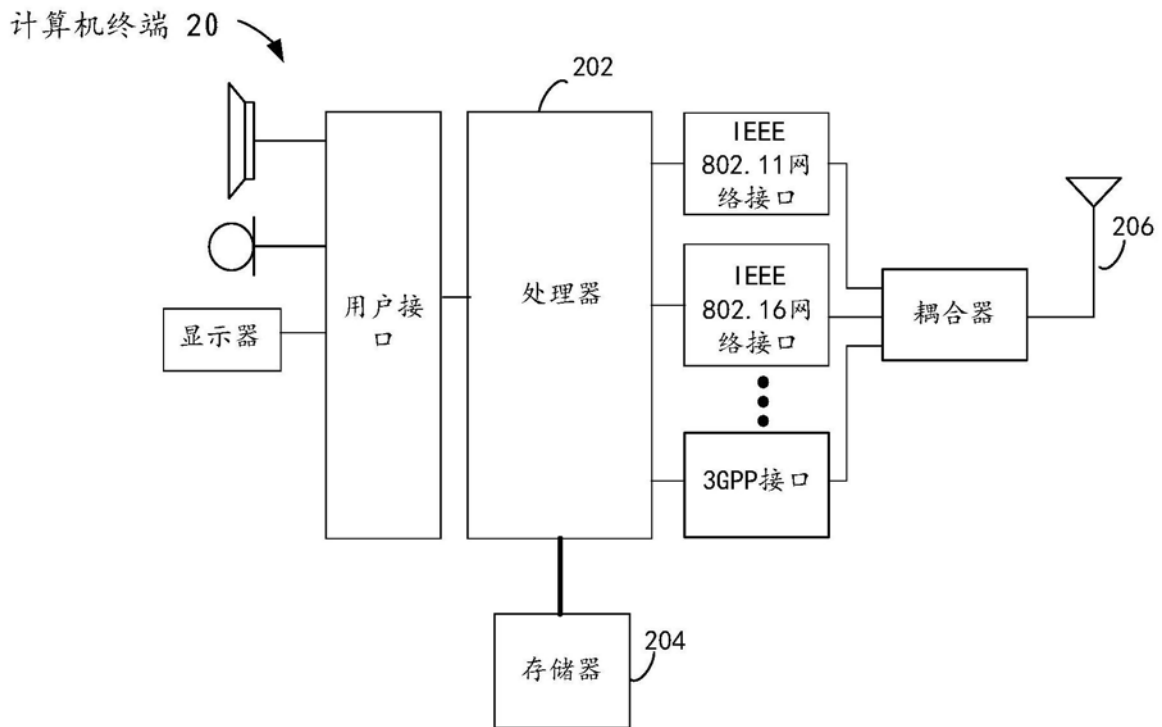


图2

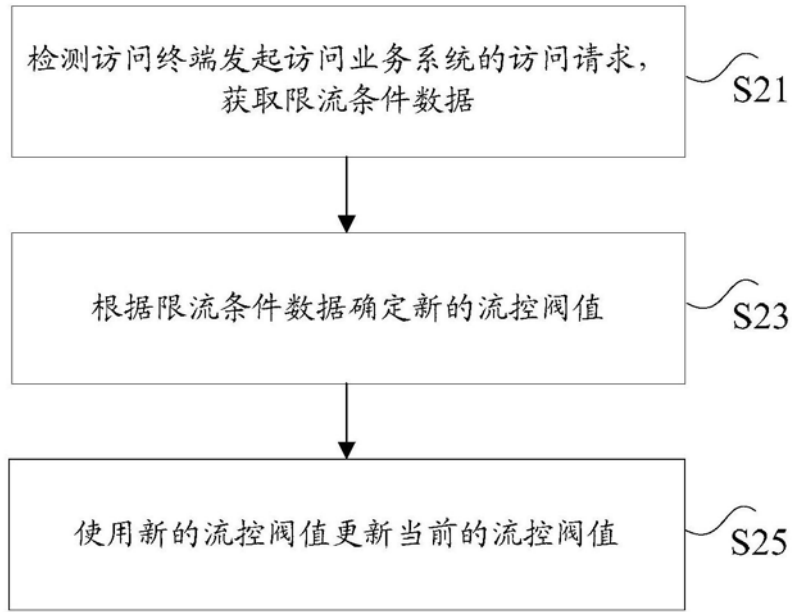


图3

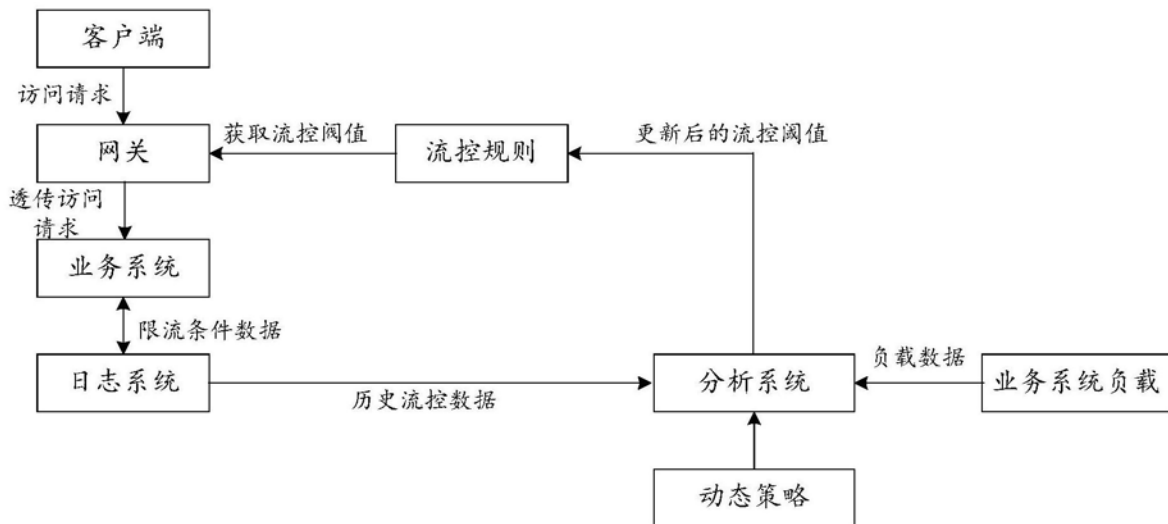


图4



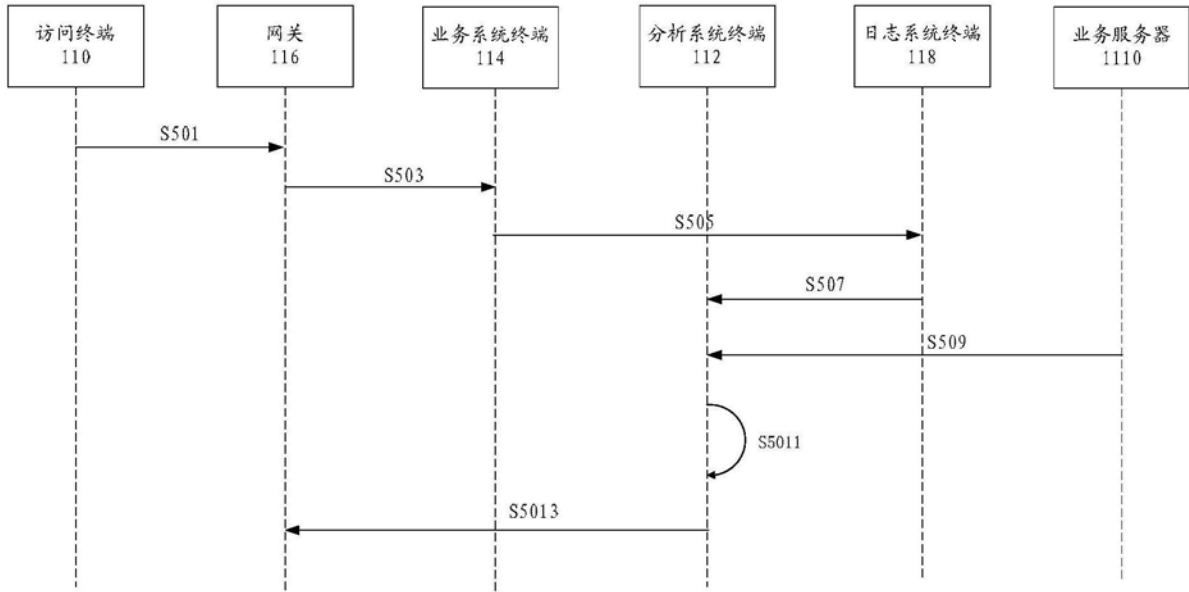


图5



图6



图7



图8

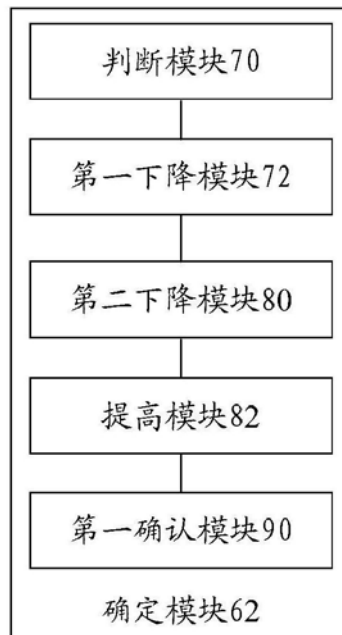


图9

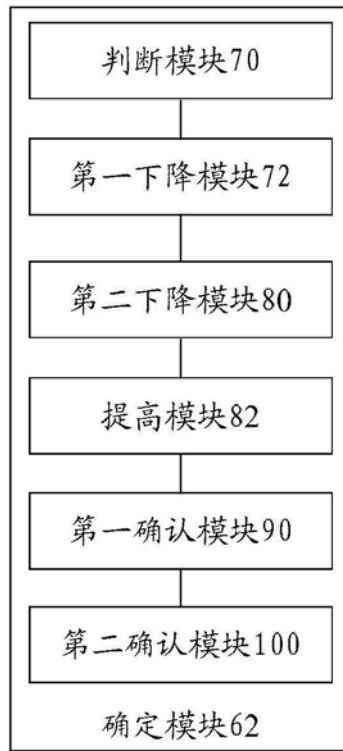


图10

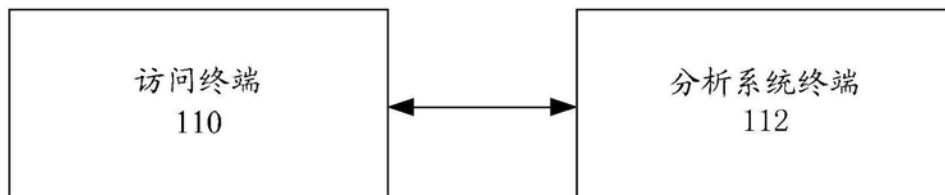


图11

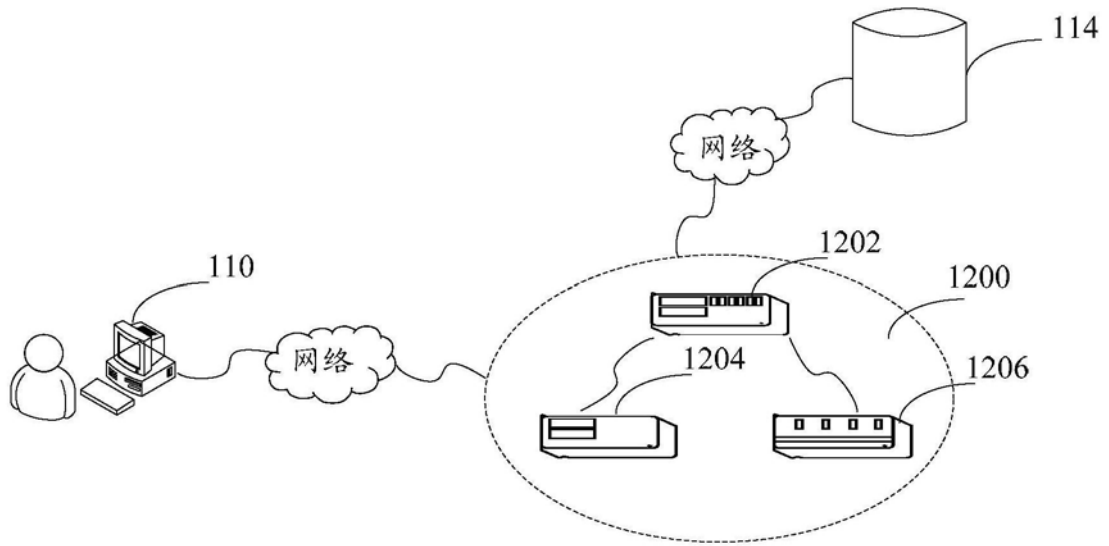


图12