



(12)发明专利

(10)授权公告号 CN 103532873 B

(45)授权公告日 2018.07.24

(21)申请号 201310552008.4

[0002]、[0018]-[0027]段及附图.

(22)申请日 2013.11.08

CN 102170396 A,2011.08.31,说明书第

[0002]-[0019]、[0022]-[0050]段.

(65)同一申请的已公布的文献号

申请公布号 CN 103532873 A

审查员 申杨

(43)申请公布日 2014.01.22

(73)专利权人 北京邮电大学

地址 100876 北京市海淀区西土城路10号

(72)发明人 陆月明 尹珊

(51)Int.Cl.

H04L 12/813(2013.01)

H04L 29/08(2006.01)

(56)对比文件

CN 102170396 A,2011.08.31,说明书第
[0002]-[0019]、[0022]-[0050]段.

CN 102868638 A,2013.01.09,说明书第

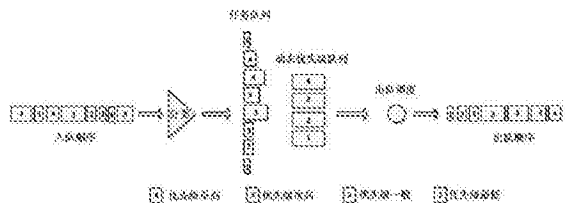
权利要求书1页 说明书6页 附图6页

(54)发明名称

应用于分布式文件系统的流量控制策略

(57)摘要

本发明涉及一种应用于分布式文件系统的流量控制策略,在应用层和网络层上进行流量控制,保证服务的可靠性及服务质量,属于计算机科学与技术领域。本发明应用于分布式文件系统中,可以实现流量分类和标记、排队与带宽分配、拥塞管理和突发流量控制功能,达到提高分布式文件系统的服务质量,完善系统功能等目标。



1. 一种基于分布式文件的系统的流量控制策略,包括管理服务器和数据服务器策略,其特征在于:

1) 管理服务器生成优先级标识位,动态标记数据流,同步该标识到相应的数据服务器;

2) 数据服务器根据各任务请求的优先级标识位分类排序,使用动态优先级判别和调整机制来处理任务请求;

其中,所述动态优先级判别和调整机制包括如下步骤:

1) 分别创建任务队列和动态优先级队列,对各服务请求进行分类操作,送入任务队列进行排队;

2) 根据任务队列的等级来填充优先级队列,由服务器系统维护;

3) 设定分配比例参数,设置计数器和空闲计时器;

4) 每完成一个任务,将其对应的计数器加一,当计数器累加到分配比例参数时,将计数器清零,并强制服务一个低优先级的任务;

5) 空闲计时器用于判断是否将任务从优先级队列去除。

2. 根据权利要求1所述的流量控制策略,其特征在于:所述优先级标识位特征如下:

1) 采用8bit优先级标识位,分别为1bit的上传和下载标识位,2bit的用户等级标识位,3bit的流量类型标识位和2bit的流量等级标识位;

2) 用户等级标识位根据付费方式分为备份数据、普通用户、高级用户和管理用户;

3) 流量类型标识位根据分布式文件系统中的常见流量类型进行分类;

4) 流量等级标识位为同一用户处理多个请求时自定义的流量等级。

3. 根据权利要求1所述的流量控制策略,其特征在于:所述排序操作主要为依次按照用户等级、流量类型、流量等级和任务到达时间来进行任务优先级排序。

4. 根据权利要求1所述的流量控制策略,其特征在于:所述空闲计时器算法如下:

1) 当剩余的最高优先级任务不足以使计数器达到设定的分配比例参数时,启动该优先级任务的空闲计时器;

2) 若指定时间内到达的任务优先级低于动态优先级队列中的最高优先级任务,或若指定时间内无任务到达,则空闲计时器超时,从优先级队列中删除当前最高优先级任务,下一优先级任务自动上升为最高优先级任务;

3) 若空闲计时器超时前有新任务到达,如果其优先级高于或等于动态优先级队列中的最高优先级任务,则动态优先队列中的最高优先级对应的空闲计时器清零,开始服务新任务并开启计时器。

应用于分布式文件系统的流量控制策略

技术领域

[0001] 本发明涉及一种应用于分布式文件系统的流量控制策略,该策略针对系统中的上传、下载及备份流量进行操作,完善现有系统功能,属于计算机科学与技术领域。

背景技术

[0002] 随着大数据时代的到来,分布式文件系统凭借其扩展能力和高可用性,受到越来越广泛的关注。分布式文件系统由管理服务器、数据服务器、客户端三大模块组成。管理服务器负责各个数据服务器的管理,文件读写调度,文件空间回收及恢复等;数据服务器负责连接管理服务器,听从管理服务器调度,提供存储空间,并为客户提供数据传输;客户端就是系统的使用者。

[0003] 在分布式文件系统中,所有的数据都被无区别地对待,系统尽最大的努力存储和下载所有用户的所有类型数据。但是由于共享服务器的局限性,系统不可能同时为所有的用户提供优良的服务,所以对于很多情况来说,这种无区别的服务会影响服务质量。例如,流媒体需要更多地考虑实时性。同时对于不同用户的不同需求,也需要有不同的服务效果。

[0004] 本发明提出一种应用于分布式文件系统的流量控制策略,对不同的数据流区别对待。本发明在应用层和网络层上进行流量控制,保证用户上传、下载流量的可靠性及服务质量。同时只要用户支付相应的费用,就可以在某些层面上享受更多的、体验更好的服务。

发明内容

[0005] 为了克服现有分布式文件系统对所有数据无区别等同对待的不足,本发明提供一种能够应用于分布式文件系统的流量控制解决方案。本发明由流量分类和标记、排队与带宽分配、拥塞管理和突发流量控制四大模块组成。

[0006] (1) 本发明组成模块

[0007] ●流量分类和标记模块:流量分类和标记是整个策略的核心部分。通过对流量的标识,并依据一定的匹配规则,就可以识别出对象,有区别地实施服务。

[0008] ●排队与带宽分配模块:该模块主要分为排队和带宽分配两个部分。第一部分是排队,将客户优先级信息放入队列中缓存,并采取调度算法安排流量的服务顺序。第二部分是带宽分配,结合IP QoS的功能,在带宽分配上区别对待不同等级流量。

[0009] ●拥塞管理模块:服务器的拥塞状态会影响服务器性能。当服务器发生拥塞时,通过降低正在服务中的任务带宽来减轻拥塞状态;当服务器过度拥塞时,通过主动放弃任务的操作来减轻过度拥塞。

[0010] ●突发流量控制模块:针对用户上传流量。如果系统中有突发的大量上传流量,则会造成服务器的过载状态,容易引起客户与服务器的连接中断现象。本发明采用令牌桶的方法来降低突发数据的速率,只有拥有令牌的流量才可以通过,这样就将数据写入速率控制在令牌产生速率之下。

[0011] (2) 本发明各模块的实现技术

[0012] 1、流量分类和标记

[0013] 采用长度为8bit的优先级标识位,动态标记数据流。当客户端向管理服务器发送请求时,管理服务器不仅要向客户返回数据服务器编号等信息,同时要通知数据服务器该客户即将采取操作,并将以下8bit优先级标识位发送给数据服务器。数据服务器暂存这些信息直到操作结束。

[0014] 表1优先级标识位

[0015]

上传或下载标识位	用户等级标识位	流量类型标识位	流量等级标识位
1bit	2bit	3bit	2bit

[0016] 下面将详细介绍各标识位的具体分类方法。

[0017] a. 上传或下载标识位

[0018] 上传为0,下载为1。

[0019] b. 用户等级标识位

[0020] 表2用户等级标识位

[0021]

标识	用户等级	用户类型
00	0	备份数据
01	1	普通用户
10	2	高级用户
11	3	管理用户

[0022] 默认排序:0优先级最低,3优先级最高。根据付费方式将用户划分为不同等级,管理用户可指定各用户等级。

[0023] c. 流量类型标识位

[0024] 表3流量类型标识位

[0025]

标识	流量级别	流量类型	备注
000	0	普通备份数据	优先级较低
001	1	普通数据业务	缺省使用的业务类型，无优先发送的要求，只要求“尽力而为”的服务质量。
010	2	时延敏感业务	流媒体等服务
011	3	紧急服务	用户支付相应费用，可以得到优先服务
100	4	保留	
101	5	保留	
110	6	保留	
111	7	保留	

[0026] 默认排序：0优先级最低7优先级最高。现在使用的有0-4等级，余下的等级作为保留。上传流量全部置为最高级。下载流量根据服务类型选择不同的优先级。

[0027] d. 流量等级标识位

[0028] 流量等级为0-3,0代表优先级最低,3代表优先级最高。所有流量等级缺省值均为3,即全力存取状态。若用户需要同时上传或下载多个数据,可通过自行设定流量等级来调整服务速率。

[0029] 分类的具体方法为,先读取所有流量用户等级,若用户等级高,则优先级最高,后面的不需比较;若用户的等级相同,则比较流量类型,若流量类型高,则优先级高,后面的不需比较;若用户等级和流量类型均相同,则根据流量等级来判断优先级。若经过一系列分类判断后,优先级相同,则根据任务到达时间排队。

[0030] 2、排队与带宽分配

[0031] 各数据服务器根据管理服务器提供的优先级标识位独立进行队列排序,形成任务队列及动态优先级队列。

[0032] 排队的具体过程包括队列的创建、报文的分类、队列调度等。如图1所示,首先将各服务请求进行分类操作,然后送入任务队列和动态优先级队列进行排队,此时出队调度机

制将按特定顺序进行服务。

[0033] 如图2所示为出队调度机制流程图,其中计数器 ≥ 2 :表示计数器到达分配比例参数;空闲计时器 ≥ 1 :表示空闲计时器超时。

[0034] 出队调度机制包括以下几个方面:

[0035] 1. 根据任务队列的等级来填充优先级队列并且由系统进行维护;

[0036] 2. 设定分配比例参数,并且设置计数器和空闲计时器,采用动态优先级判别和调整机制。分配比例参数和计数器用于防止高等级任务总是占用带宽,而低等级任务一直得不到服务的情况出现,实现方式为每服务一个相对高优先级的任务,将其对应等级的分配计数器值加一,当计数器累加到分配比例参数时,将计数器清零,并强制服务一个相对低优先级的任务。

[0037] 3. 空闲计时器用于判断是否将任务从优先级队列去除。当剩余最高优先级任务不足以使计数器达到设定值时,启动该优先级任务相应的空闲计时器。若指定时间内无任务到达,则空闲计时器超时,从优先级队列中删除当前最高优先级任务,下一优先级任务自动上升为最高优先级任务。若空闲计时器超时前有新任务到达,如果其优先级高于或等于动态优先级队列中的最高优先级任务,则动态优先级队列中的最高优先级对应的空闲计时器清零,开始服务新任务并开启计时器。如果新到达任务优先级低于动态优先级队列中的最高优先级任务,则仍服务当前任务。

[0038] 带宽分配机制包括以下几点:

[0039] 1. 按照优先级队列的先后顺序分配带宽,不同优先级任务享有不同的最大限制带宽,使得相应的服务速率也不同。

[0040] 2. 为各不同优先级任务设置最大限制带宽的两个门限值,即最大限制带宽的最高门限和最大限制带宽的最低门限。最高门限用于系统空闲的状态,给用户提供优秀的服务质量。最低门限用在系统繁忙时刻,给用户合格的服务质量。这两个门限的设置用于拥塞控制。

[0041] 3、拥塞管理

[0042] 正常情况下,给每个用户分配最大限制带宽的最高门限值。当拥塞发生时,即队列长度超过拥塞门限,则从最低等级流量开始,将分配带宽的门限降低到最大限制带宽的最低门限,以期使更多的用户得到服务,同时又保证正在进行中的服务的质量。当队列长度超过过度拥塞门限时,从最低等级流量开始抛弃一些请求,并发送重新请求信令。

[0043] 4、突发流量控制

[0044] 当令牌桶中没有令牌的时候,将不会给用户发送可以上传通知,数据也就不能被存储,只有等到桶中生成了新的令牌,数据才可以被存储,这就可以限制数据的流量只能小于等于令牌生成的速度,达到限制流量的目的。

[0045] (3) 本发明执行流程

[0046] 为了实现本发明“应用于分布式文件系统的流量控制策略”,要求对现有分布式文件系统进行修改。具体修改方法为在上传和下载操作之前对用户等级进行优先级排队。修改后的分布式文件系统上传和下载操作流程如图5和图6所示。

[0047] 如图5所示,上传操作具体的执行流程:当数据服务器收到管理服务器关于上传客户的优先级信息时,对这些信息进行排队,按队列顺序依次发放令牌,为得到令牌的用户分

配带宽,即使有新到达的高优先级客户也不能占用已分配带宽,并且立即向客户发送上传数据通知,客户在接到通知后再上传数据。

[0048] 如图6所示,下载操作具体的执行流程为,当数据服务器收到客户端发送的下载请求时,对这些客户对应的优先级信息进行排队,按照排队次序传送数据。

[0049] (4) 本发明特点

[0050] 本发明“应用于分布式文件系统的流量控制策略”的主要特点如下。

[0051] 1. 高效性:整个系统通过信令来完成信息交互,提高系统效率;

[0052] 2. 实用性:本发明在应用层和网络层上提供一种简单可行的流量控制方案;

[0053] 3. 普适性:本发明对大部分的分布式文件系统都适用;

[0054] 4. 可扩展性:本发明可根据应用场合及用户的具体需要进行扩展,且扩展十分简单,只需要改变优先级标识位的相关参数即可。

附图说明

[0055] 图1队列模型

[0056] 图2出队调度流程

[0057] 图3拥塞管理

[0058] 图4令牌桶

[0059] 图5分布式文件系统上传流程图

[0060] 图6分布式文件系统下载流程图

[0061] 图7排队实施例

[0062] 图8出队调度实施例

具体实施方式

[0063] 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,所描述的实施例也仅仅是本发明的一部分实施例,而不是全部实施例。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。

[0064] 为了说明本发明的排队机制,这里给出了一个排队的实施例,如图7所示,多个不同优先级的任务同时需要被服务,此时数据服务器将各任务按照优先级分类,生成任务队列和动态优先级队列,由出队调度机制管理服务顺序。

[0065] 如图8所示,出队调度过程如下:

[0066] 第一步,将待服务任务进行分类排序,并根据任务队列形成动态优先级队列;

[0067] 第二步,服务两个三级任务后,三级对应的计数器为2,达到分配比例参数;

[0068] 第三步,服务一个二级任务,三级对应的计数器清零,二级对应的计数器加一;

[0069] 第四步,服务一个三级任务,此时任务队列中已经没有三级任务,所以启动三级对应的空闲计时器;

[0070] 第五步,空闲计时器超时前没有三级或比三级任务优先级高的任务到达,所以在空闲计时器超时后,将三级从优先级队列中取出,二级自动升级为最高优先级,继续服务二级任务;

[0071] 第六步,二级计数器到达2时,服务一级任务并清零二级计数器。

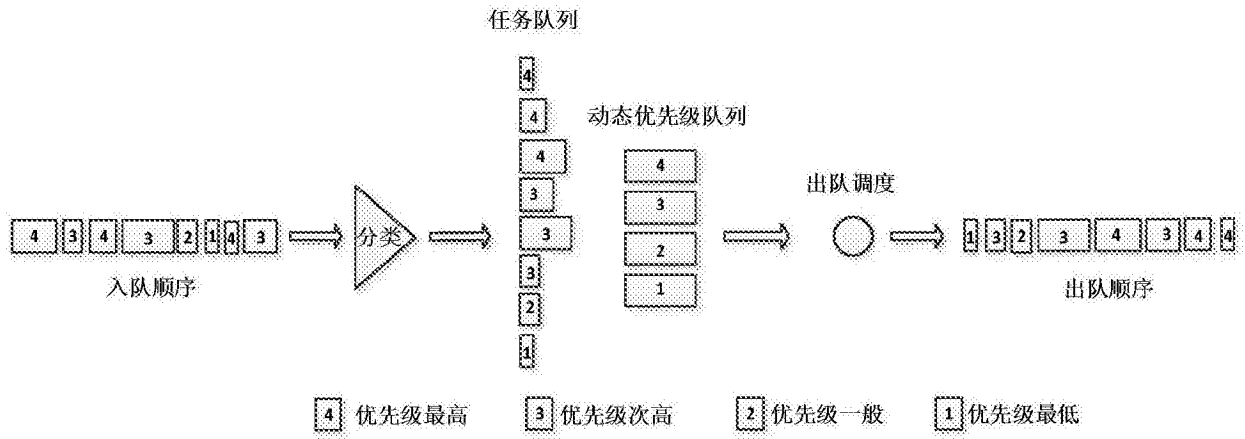


图1

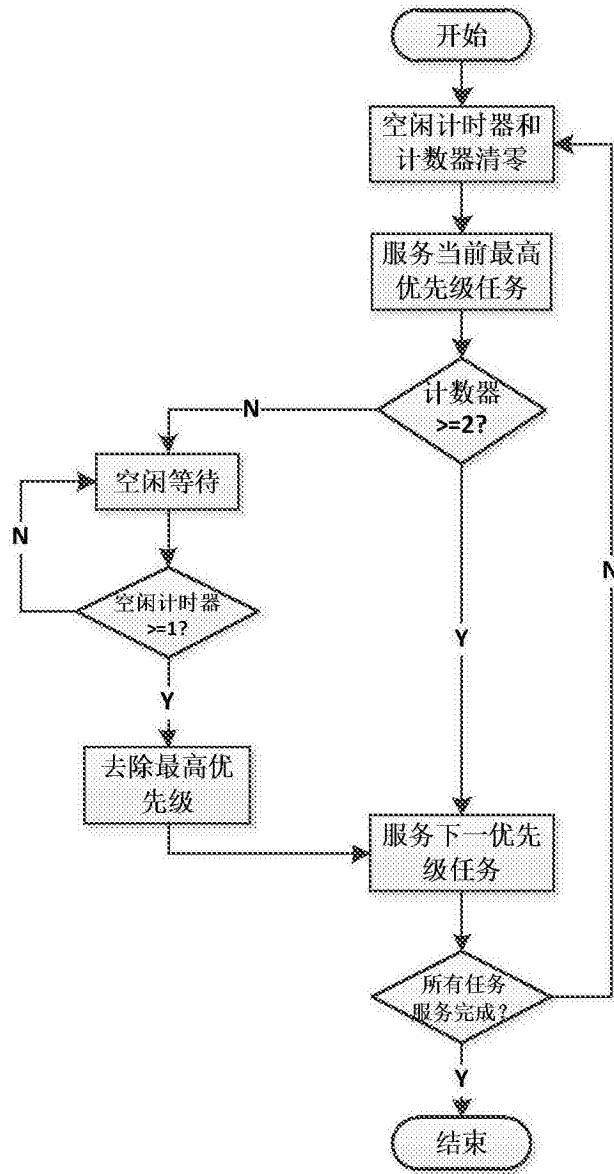


图2

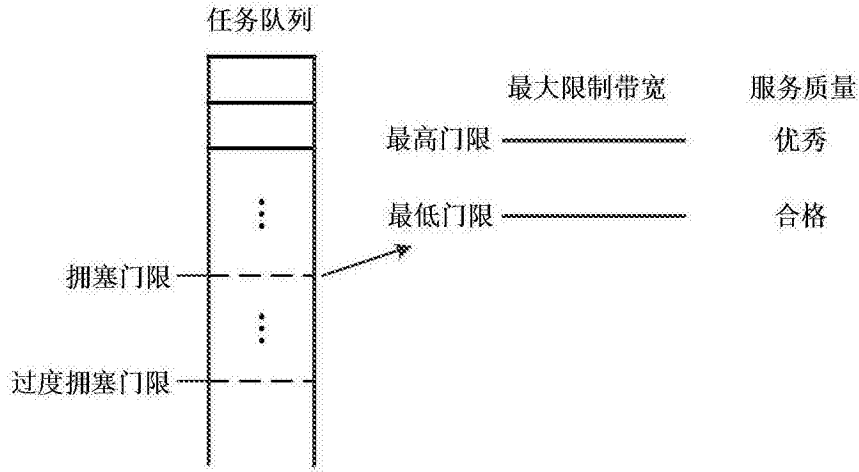


图3

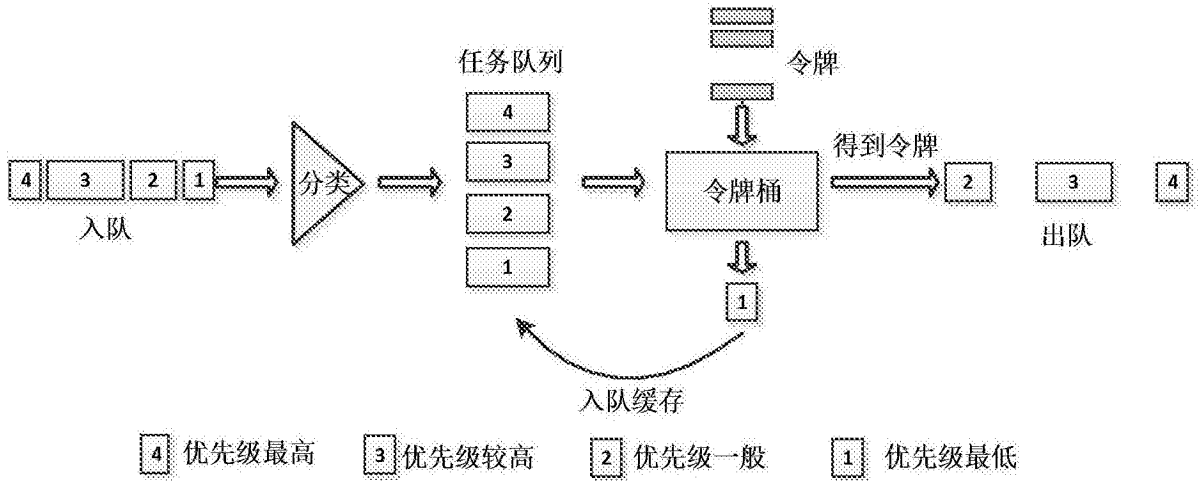


图4

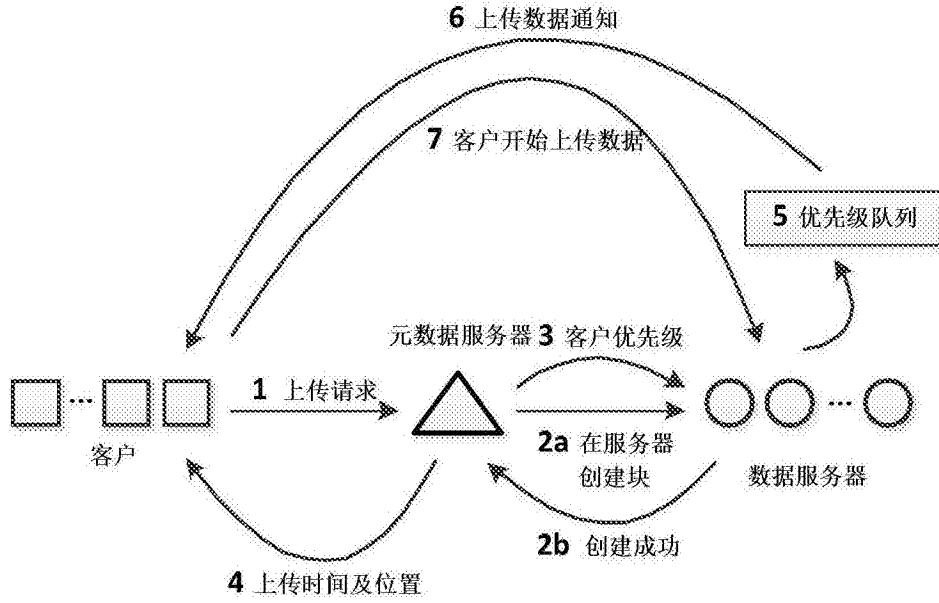


图5

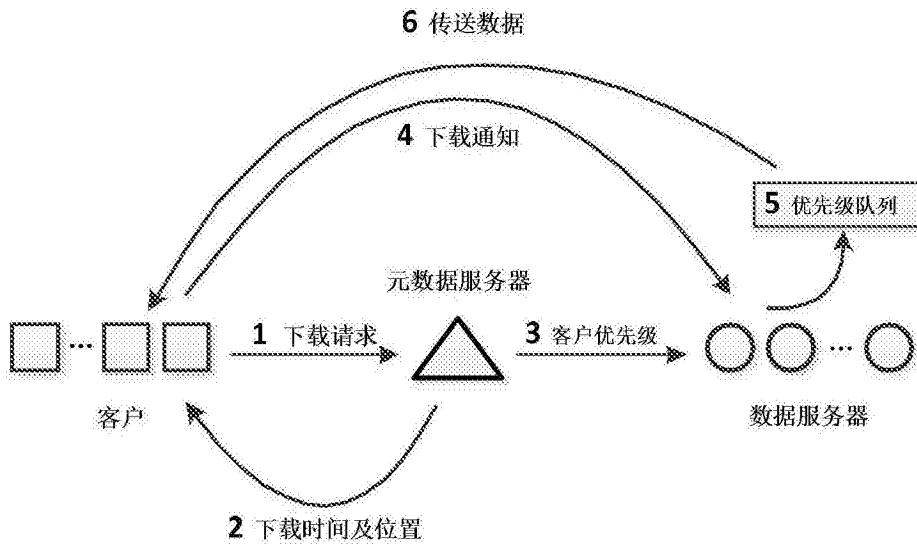


图6

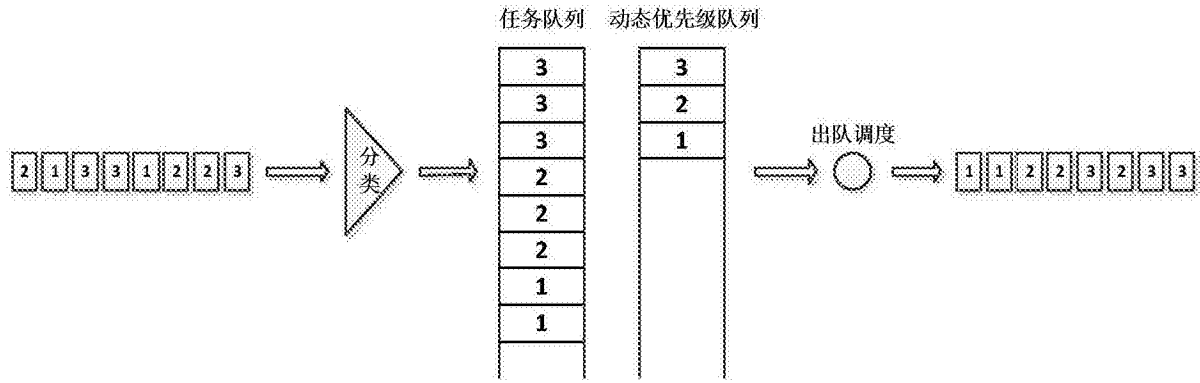


图7

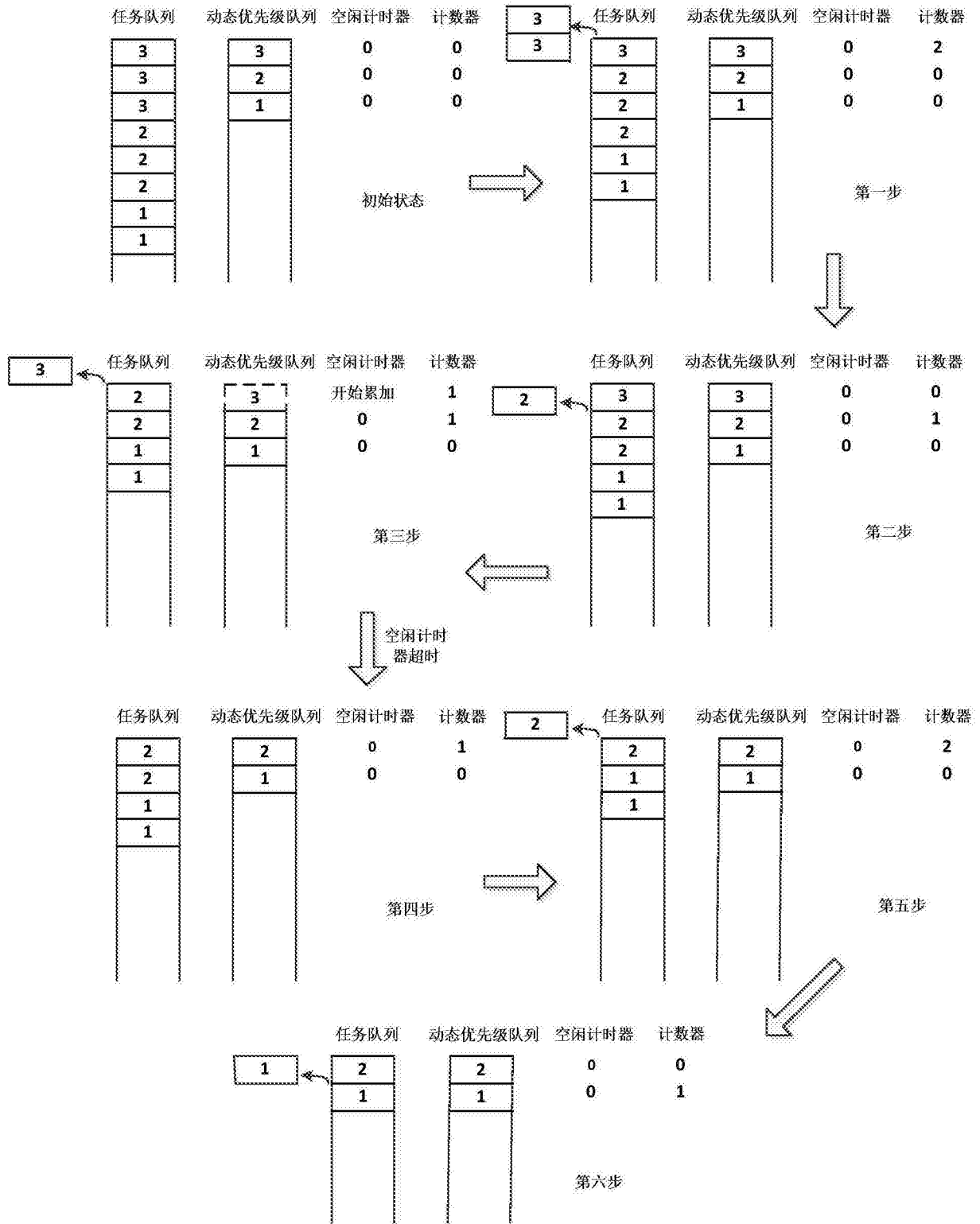


图8