



(12) 发明专利申请

(10) 申请公布号 CN 103049396 A

(43) 申请公布日 2013. 04. 17

(21) 申请号 201210529597. X

(22) 申请日 2012. 12. 10

(71) 申请人 浪潮(北京)电子信息产业有限公司
地址 100085 北京市海淀区上地信息路2号
2-1号C栋1层

(72) 发明人 王倩 张在理

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 栗若木 曲鹏

(51) Int. Cl.

G06F 12/08 (2006. 01)

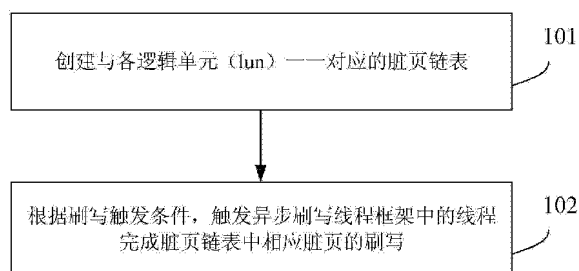
权利要求书 2 页 说明书 6 页 附图 3 页

(54) 发明名称

数据的刷写方法及装置

(57) 摘要

本发明提供了一种数据的刷写方法及装置,其中,该数据的刷写方法包括:根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。上述数据的刷写方法及装置,可以更快地将缓存中的数据写入到磁盘上,提高了缓存的使用率,从而提高了读写效率。



1. 一种数据的刷写方法,其特征在于,该方法包括:

根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。

2. 根据权利要求1所述的方法,其特征在于,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写之前,该方法还包括:

创建与各逻辑单元(lun)一一对应的脏页链表,其中,所述脏页链表为双向链表,所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应。

3. 根据权利要求1所述的方法,其特征在于,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,包括:

当写进程发现所述脏页的比例超过预定阈值时,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;

当读写处理进程需要申请空闲缓存页时,若空闲缓存页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,则触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;或者

当定时器定时检查到各lun的变脏时间超过指定阈值时,则触发异步刷写线程框架中的线程定时完成脏页链表中相应脏页的刷写。

4. 根据权利要求1-3任一权利要求所述的方法,其特征在于,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写之前,该方法还包括:

创建指定数目的空闲线程,并挂接到全局刷写线程链表中。

5. 根据权利要求4所述的方法,其特征在于,所述触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,包括:

通过函数从所述全局刷写线程链表中获得一个空闲线程的信息,设置调用的刷写函数及相关参数,并唤醒与该信息对应的线程,以便所述线程调用所述刷写函数刷写数据。

6. 根据权利要求5所述的方法,其特征在于,该方法还包括:

检查过去的预定时间内所述全局刷写线程链表是否为空,若为空且当前线程数小于所述指定数目,则创建一个新线程并加入到所述全局刷写线程链表中;

检查所述全局刷写线程链表中表头的线程在过去的预定时间内是否空闲,若空闲,则减少所述全局刷写线程链表中线程的数目。

7. 一种数据的刷写装置,其特征在于,该装置包括:

刷写模块,用于根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。

8. 根据权利要求7所述的装置,其特征在于,该装置还包括:

创建模块,用于创建与各逻辑单元(lun)一一对应的脏页链表,和/或创建指定数目的空闲线程,并挂接到全局刷写线程链表中;其中,所述脏页链表为双向链表,所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应。

9. 根据权利要求7或8所述的装置,其特征在于:

所述刷写模块,具体用于:

当写进程发现所述脏页的比例超过预定阈值时,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;

当读写处理进程需要申请空闲缓存页时,若空闲缓存页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,则触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;或者

当定时器定时检查到各 lun 的变脏时间超过指定阈值时,则触发异步刷写线程框架中的线程定时完成脏页链表中相应脏页的刷写。

10. 根据权利要求 8 权利要求所述的装置,其特征在于:

所述刷写模块,具体用于:通过函数从所述全局刷写线程链表中获得一个空闲线程的信息,设置调用的刷写函数及相关参数,并唤醒与该信息对应的线程,以便所述线程调用所述刷写函数刷写数据。

11. 根据权利要求 10 所述的装置,其特征在于,该装置还包括:

处理模块,用于检查过去的预定时间内所述全局刷写线程链表是否为空,若为空且当前线程数小于所述指定数目,则创建一个新线程并加入到所述全局刷写线程链表中;以及,检查所述全局刷写线程链表中表头的线程在过去的预定时间内是否空闲,若空闲,则减少所述全局刷写线程链表中线程的数目。

www.patviewer.com

数据的刷写方法及装置

技术领域

[0001] 本发明涉及数据存储技术,尤其涉及数据的刷写方法及装置。

背景技术

[0002] 随着计算机技术和存储技术的发展,用户对存储系统的可用性的要求越来越高。这里所谓的可用性包括存储系统的各个层面,不仅仅是存储本身,还包括存储数据的读写方法等。目前,存储数据的读写速度瓶颈主要在于对于磁盘本身进行读写时,磁头用于查找磁道的时间耗时较大。因此,存储系统中一般都增加了缓存,以提高读写速度。

[0003] 缓存的作用是在存储控制器的读写请求和磁盘读写之间增加了一层虚拟缓存,用于暂时保存数据。通过缓存提高读写速度是通过以下方式进行的:对于读请求,存储控制器先在虚拟缓存中查找是否存在请求的数据,如果存在,直接将数据返回,如果不存在,再通过磁盘读写获取数据。这样,对于数据在缓存中存在的情况,就省略了进行磁盘读写的动作,跳过了最为耗时的操作,从而提高了读效率。对于写请求,存储控制器先将要写到磁盘的数据写入缓存,然后返回写操作的结果。而将数据从缓存写入磁盘的操作,即刷写,则通过刷写方法在后台进行。由于在存储控制器的读请求流程中,没有读写磁盘的操作,所以写效率就得到了相应的提高。由于缓存大小是有限的,为了提高缓存的使用率,需要尽快将需要写入磁盘的数据写到磁盘上。

发明内容

[0004] 本发明提供了一种数据的刷写方法及装置,以解决在缓存大小是有限的情况下,如何尽快地将需要写入磁盘的数据写到磁盘上的问题。

[0005] 本发明提供了一种数据的刷写方法,该方法包括:

[0006] 根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。

[0007] 优选地,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写之前,该方法还包括:

[0008] 创建与各逻辑单元(lun)一一对应的脏页链表,其中,所述脏页链表为双向链表,所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应。

[0009] 优选地,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,包括:

[0010] 当写进程发现所述脏页的比例超过预定阈值时,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;

[0011] 当读写处理进程需要申请空闲缓存页时,若空闲缓存页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,则触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;或者

[0012] 当定时器定时检查到各lun的变脏时间超过指定阈值时,则触发异步刷写线程框

架中的线程定时完成脏页链表中相应脏页的刷写。

[0013] 优选地,所述根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写之前,该方法还包括:

[0014] 创建指定数目的空闲线程,并挂接到全局刷写线程链表中。

[0015] 优选地,所述触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,包括:

[0016] 通过函数从所述全局刷写线程链表中获得一个空闲线程的信息,设置调用的刷写函数及相关参数,并唤醒与该信息对应的线程,以便所述线程调用所述刷写函数刷写数据。

[0017] 优选地,该方法还包括:检查过去的预定时间内所述全局刷写线程链表是否为空,若为空且当前线程数小于所述指定数目,则创建一个新线程并加入到所述全局刷写线程链表中;检查所述全局刷写线程链表中表头的线程在过去的预定时间内是否空闲,若空闲,则减少所述全局刷写线程链表中线程的数目。

[0018] 本发明还提供了一种数据的刷写装置,该装置包括:

[0019] 刷写模块,用于根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。

[0020] 优选地,该装置还包括:创建模块,用于创建与各逻辑单元(lun)一一对应的脏页链表,和/或创建指定数目的空闲线程,并挂接到全局刷写线程链表中;其中,所述脏页链表为双向链表,所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应。

[0021] 优选地,所述刷写模块,具体用于:

[0022] 当写进程发现所述脏页的比例超过预定阈值时,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;

[0023] 当读写处理进程需要申请空闲缓存页时,若空闲缓存页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,则触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;或者

[0024] 当定时器定时检查到各lun的变脏时间超过指定阈值时,则触发异步刷写线程框架中的线程定时完成脏页链表中相应脏页的刷写。

[0025] 优选地,所述刷写模块,具体用于:通过函数从所述全局刷写线程链表中获得一个空闲线程的信息,设置调用的刷写函数及相关参数,并唤醒与该信息对应的线程,以便所述线程调用所述刷写函数刷写数据。

[0026] 优选地,该装置还包括:处理模块,用于检查过去的预定时间内所述全局刷写线程链表是否为空,若为空且当前线程数小于所述指定数目,则创建一个新线程并加入到所述全局刷写线程链表中;以及,检查所述全局刷写线程链表中表头的线程在过去的预定时间内是否空闲,若空闲,则减少所述全局刷写线程链表中线程的数目。

[0027] 上述数据的刷写方法及装置,可以更快地将缓存中的数据写入到磁盘上,提高了缓存的使用率,从而提高了读写效率。

附图说明

[0028] 图1是本发明数据的刷写方法实施例的过程示意图;

[0029] 图2是本发明中脏页链表的示意图;

[0030] 图 3 是本发明异步刷写线程框架中线程工作的流程图；

[0031] 图 4 是本发明数据的刷写装置实施例的结构示意图。

具体实施方式

[0032] 为使本发明的目的、技术方案和优点更加清楚明白，下文中将结合附图对本发明的实施例进行详细说明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互任意组合。

[0033] 如图 1 所示，是本发明数据的刷写方法实施例的过程示意图，该方法包括：

[0034] 步骤 101、创建与各逻辑单元 (lun) 一一对应的脏页链表；

[0035] 其中，所述脏页链表为双向链表，所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应；

[0036] 该步骤为可选步骤；

[0037] 步骤 102、根据刷写触发条件，触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写，其中，所述脏页为保存有数据的缓存页。

[0038] 其中，异步刷写线程框架中的线程为刷写线程，为描述方便，本实施例将刷写线程简称为线程。

[0039] 上述方法共涉及三方面内容：缓存的组织方式，刷写触发者和异步刷写线程框架：

[0040] 1) 缓存的组织方式

[0041] 本发明实施例采用的基本缓存大小是系统页大小，一般为 4KB，每一块基本缓存称为缓存页。每一个缓存页可以保存一个系统页大小的数据。

[0042] 逻辑单元 (lun) 是一个提供小型计算机系统接口 (scsi) 服务的单元，在块设备类型的 scsi 服务中通常由一个块设备如磁盘或者逻辑卷映射为一个 lun。在本实施例中，保存有效数据的页通过哈希表进行组织，查找页时可以根据给每个缓存页唯一指定的 lun 号和页号找到相应的缓存页。当在哈希表中找不到缓存页时，就申请一个新的缓存页，并将其索引到哈希表。由于缓存的大小是有限的，当申请不到缓存页时，系统就要适当地释放一部分可以释放的缓存页，以保证可以及时申请到缓存页。当存储控制器的写请求将数据暂时保存到缓存页中后，在这些缓存页中的数据没有被刷写到磁盘之前，这些缓存页都是不可以被释放的。本发明实施例通过保证缓存页数据更快地写入磁盘上，从而保证缓存页可以被及时释放，进而保证读写速度。

[0043] 存储控制器写请求会将即将写入磁盘的数据写入到缓存页中，将保存这些数据的缓存页称为脏页。在本发明实施例中，每个 lun 有一个脏页链表，如图 2 所示，链接这个 lun 中的所有脏页。这是一个双向链表，当一个缓存页变脏时，如果该页原本不是脏页，那么该页被直接插入到链表尾。如果该页原本就是脏页，那么就把该页从链表中本来的位置取出，然后插入到链表尾。当一个页频繁变脏时，它会被不断取出插入链表尾部，因此，最早变脏的页会逐渐被推到链表头。这就是该链表的特性。当对 lun 的脏页进行刷写时，直接从该 lun 对应的脏页链表中取出脏页刷写到存储设备。

[0044] 使用这种数据结构的优点是：

[0045] 1、刷写脏页时不需要扫描所有缓存页进行查找，省略了查找过程；

[0046] 2、脏页链表中,脏页的变脏顺序与脏页在链表中的位置相对应,最近变脏的页放在链表尾,最久变脏的页最先被刷写,避免了活动页的频繁刷写;

[0047] 3、在顺序写的情况下,相邻页在脏页链表中的位置也是相邻的,方便将多个相邻页合成为一个输入输出(I/O)请求写到磁盘上。

[0048] 2) 刷写触发者

[0049] 刷写触发者,是指触发异步刷写线程框架的刷写触发条件,是该框架的具体使用环境,脏页刷写触发者通过调用指定接口获取一个空闲的刷写线程并传递具体刷写函数指针来触发异步脏页刷写。

[0050] 在本存储系统中,脏页刷写的触发者有三个:

[0051] 1、写进程:当写进程发现脏页比例超过预定阈值时激活一个刷写线程进行刷写。

[0052] 2、I/O 缓存申请者:读写处理进程在发现读写的页不存在时需要申请空闲页,若空闲页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,将触发脏页刷写。

[0053] 3、定时器:模块定义一个定时器定时检查各个 lun 变脏时间超过指定阈值,超过则触发脏页刷写。刷写方式为定时刷写。

[0054] 为了提高刷写效率,每次刷写时,都会检查脏页链表中的相邻页在物理位置上是否也是相邻的,检查的依据是两个脏页的页索引是否相邻,如果相邻,则会把尽可能多的相邻页合并为一个 I/O 请求,写入到磁盘上。

[0055] 3) 异步刷写线程框架

[0056] 异步刷写线程框架是本发明的核心,用于根据系统实际的负载情况,不断调整刷写线程的数量。在 I/O 负载大时,及时增加刷写线程,在 I/O 负载小时,及时释放刷写线程。

[0057] 使用了异步刷写线程框架,可以支持多线程刷写以提高刷写效率。下面详细介绍异步刷写线程框架:

[0058] 1、初始时创建指定数目的空闲刷写线程,并挂接到全局刷写线程链表中。

[0059] 2、刷写触发者通过函数从全局刷写线程链表取得一个空闲的刷写线程信息,设置调用的刷写函数及其参数并唤醒该线程。

[0060] 3、每个刷写线程执行流程如下:

[0061] (1) 将本线程加入全局刷写链表并挂起;

[0062] (2) 线程被唤醒,调用刷写函数刷写数据。线程被唤醒的同时,刷写触发者会将该线程从全局刷写线程中删除;

[0063] (3) 检查过去的预定时间如 1 秒钟全局刷写链表是否为空,若是并且当前刷写线程数小于指定的最大刷写线程数,表示系统刷写忙,需要更多的线程执行动作,则创建一个刷写线程并加入到全局刷写线程链表中;

[0064] (4) 检查全局刷写线程链表最早的刷写线程在过去的预定时间如 1 秒钟是否空闲,若是,则表示系统当前不需要这么多数量的线程,本线程退出,否则返回(1)。

[0065] 如图 3 所示,该框架可以根据系统的 I/O 负载情况,适时调整线程个数,具体包括如下步骤:

[0066] 步骤 301、创建线程,增加线程个数的计数;

[0067] 步骤 302、判断线程停止标志是否置位,若置位,执行步骤 303,否则执行步骤 311;

[0068] 步骤 303、线程睡眠,等待唤醒;

[0069] 步骤 304、判断线程是否在刷写线程列表中,若是,转向步骤 302,否则执行步骤 305;

[0070] 步骤 305、执行刷写操作;

[0071] 步骤 306、检查过去的 1 秒钟全局刷写链表是否为空,若为空,执行步骤 307,否则,执行步骤 309;

[0072] 步骤 307、判断当前刷写线程总数是否小于最大刷写线程总数,若是,执行步骤 308;

[0073] 步骤 308、创建一个刷写线程并加入到全局刷写线程链表中;

[0074] 步骤 309、检查表头的刷写线程在过去的 1 秒钟是否空闲,若是,执行步骤 310,否则,转向步骤 302;

[0075] 步骤 310、判断当前刷写线程总数是否小于等于最小刷写线程总数,若是,转向步骤 302,否则,执行步骤 311;

[0076] 步骤 311、减少刷写线程计数。

[0077] 上述数据的刷写方法,可以更快地将缓存中的数据写入到磁盘上,提高了缓存的使用率,从而提高了读写效率。

[0078] 如图 4 所示,是本发明数据的刷写装置实施例的结构示意图,该装置包括:刷写模块 41,用于根据刷写触发条件,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写,其中,所述脏页为保存有数据的缓存页。

[0079] 另外,该装置还包括:创建模块 42,用于创建与各逻辑单元(lun)一一对应的脏页链表,和/或创建指定数目的空闲线程,并挂接到全局刷写线程链表中;其中,所述脏页链表为双向链表,所述脏页的变脏顺序与所述脏页在对应的脏页链表中的位置相对应。

[0080] 其中,所述刷写模块,具体用于:当写进程发现所述脏页的比例超过预定阈值时,触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;当读写处理进程需要申请空闲缓存页时,若空闲缓存页不够,则需要查找和释放部分缓存页,如果可释放的缓存页不足,则触发异步刷写线程框架中的线程完成脏页链表中相应脏页的刷写;或者,当定时器定时检查到各 lun 的变脏时间超过指定阈值时,则触发异步刷写线程框架中的线程定时完成脏页链表中相应脏页的刷写。

[0081] 另外,所述刷写模块,具体用于:通过函数从所述全局刷写线程链表中获得一个空闲线程的信息,设置调用的刷写函数及相关参数,并唤醒与该信息对应的线程,以便所述线程调用所述刷写函数刷写数据。

[0082] 进一步地,该装置还包括:处理模块 43,用于检查过去的预定时间内所述全局刷写线程链表是否为空,若为空且当前线程数小于所述指定数目,则创建一个新线程并加入到所述全局刷写线程链表中;以及,检查所述全局刷写线程链表中最早的线程在过去的预定时间内是否空闲,若空闲,则减少所述全局刷写线程链表中线程的数目。

[0083] 上述数据的刷写装置,可以更快地将缓存中的数据写入到磁盘上,提高了缓存的使用率,从而提高了读写效率。

[0084] 本领域普通技术人员可以理解上述方法中的全部或部分步骤可通过程序来指令相关硬件完成,上述程序可以存储于计算机可读存储介质中,如只读存储器、磁盘或光盘等。可选地,上述实施例的全部或部分步骤也可以使用一个或多个集成电路来实现。相应

地,上述实施例中的各模块/单元可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。本发明不限制于任何特定形式的硬件和软件的结合。

[0085] 以上实施例仅用以说明本发明的技术方案而非限制,仅仅参照较佳实施例对本发明进行了详细说明。本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的精神和范围,均应涵盖在本发明的权利要求范围当中。

www.patviewer.com

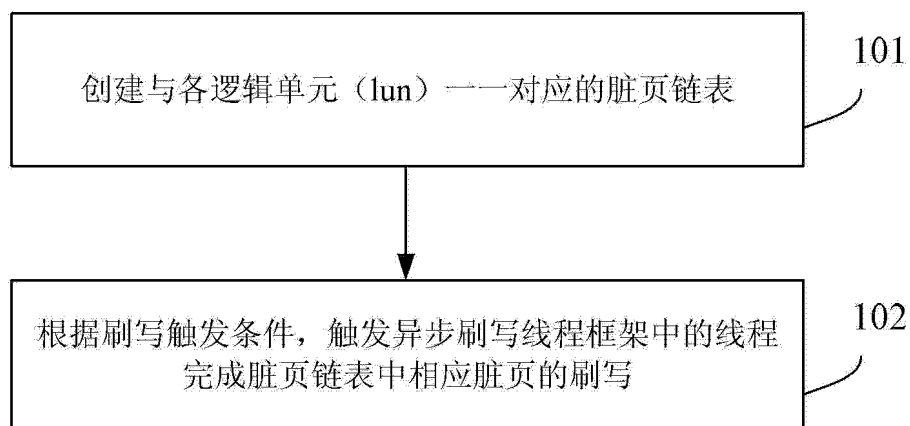


图 1

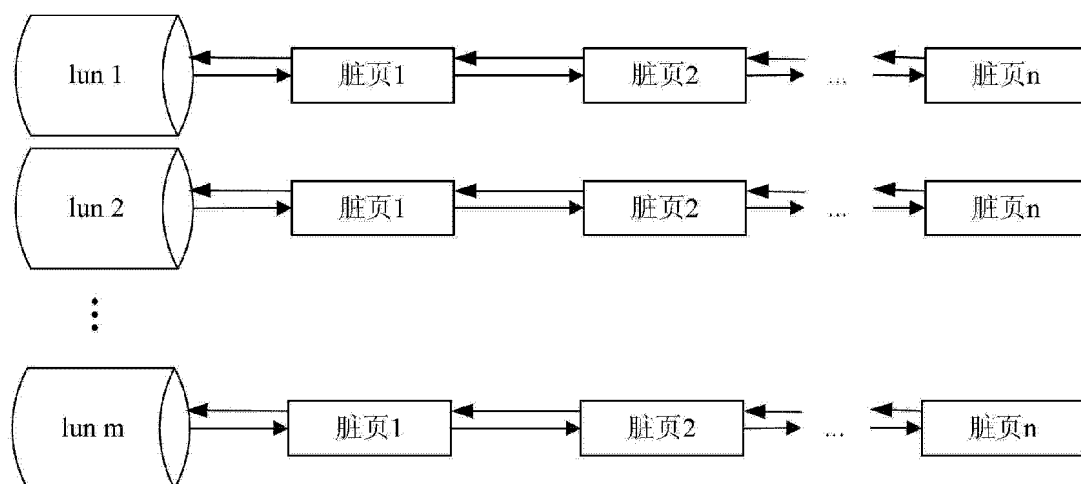


图 2

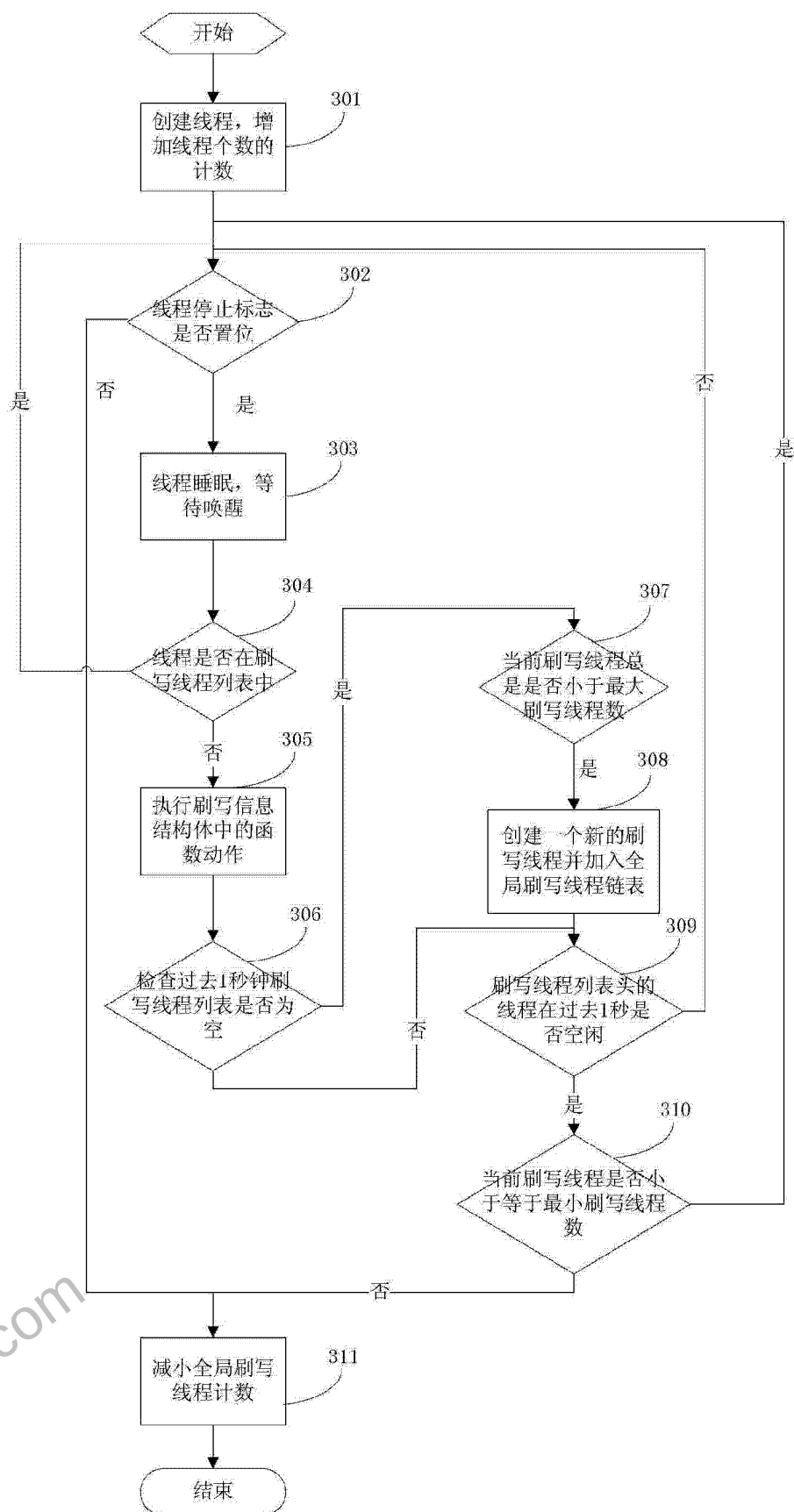


图 3

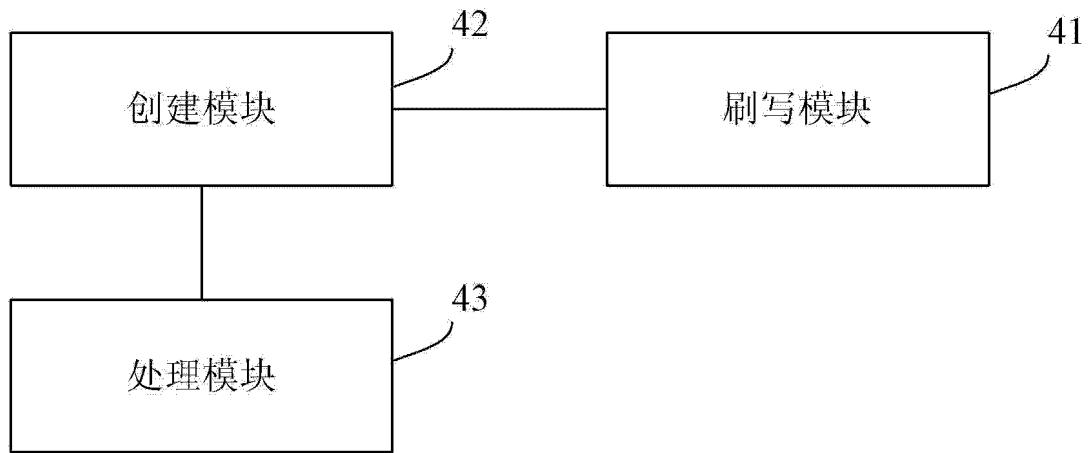


图 4