



(12)发明专利

(10)授权公告号 CN 104915153 B

(45)授权公告日 2017.09.22

(21)申请号 201510310888.3

(56)对比文件

(22)申请日 2015.06.09

US 5459857 A, 1995.10.17,

(65)同一申请的已公布的文献号

US 2013/0117223 A1, 2013.05.09,

申请公布号 CN 104915153 A

CN 104378374 A, 2015.02.25,

(43)申请公布日 2015.09.16

CN 103391540 A, 2013.11.13,

(73)专利权人 山东超越数控电子有限公司

CN 101576837 A, 2009.11.11,

地址 250100 山东省济南市高新区孙村镇
科航路2877号

审查员 陈国耀

(72)发明人 张凡凡 吴登勇 李保来

(74)专利代理机构 济南信达专利事务所有限公司 37100

代理人 姜明

(51)Int.Cl.

G06F 3/06(2006.01)

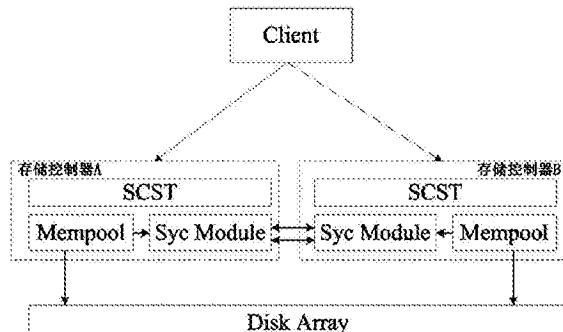
权利要求书2页 说明书5页 附图3页

(54)发明名称

一种基于SCST的双控缓存同步设计方法

(57)摘要

本发明提供一种基于SCST的双控缓存同步设计方法，其具体实现过程为：将客户机连接到两个相同的存储控制器A和B中，两存储控制器内均设置SCST缓存区、缓存池模块和同步模块，两存储控制器均连接磁盘阵列；当客户机数据输入一存储控制器时，该存储控制器内的数据备份到另一存储控制器，实现两存储控制器内的缓存同步，然后将存储控制器中的数据写入到磁盘阵列中。该基于SCST的双控缓存同步设计方法和现有技术相比，可以实现存储控制器数据的备份存储，缓存数据可以实现多物理传输介质的同步，满足海量阵列存储系统的吞吐量需求，实用性强，易于推广。



1. 一种基于SCST的双控缓存同步设计方法,其特征在于,该方法通过以下步骤实现:

一、将客户机连接到两个相同的存储控制器A和B中,两存储控制器内均设置SCST缓存区、缓存池模块和同步模块,两存储控制器均连接磁盘阵列;所述SCST缓存区以存储控制器主机的主存为存储器;缓存块大小按照内存页面的 4KB 大小设计,缓存块由一个内存页组成;

二、当客户机数据输入一存储控制器时,该存储控制器内的数据备份到另一存储控制器,实现两存储控制器内的缓存同步;

所述存储控制器内的数据缓存同步过程为:

数据传输到存储控制器A,首先放入其缓存池模块中,该缓存池模块同时保存未写入磁盘的数据和已写入磁盘的数据;

缓存池模块中的数据发送到同步模块与存储控制器B的同步模块中,最后传递到存储控制器B的缓存池模块中,使得每个存储控制器中的数据,既包含有本地存储控制器的缓存数据,也包含有另一存储控制器的缓存数据;

所述存储控制器A与存储控制器B进行数据缓存前,还包括进行握手通信的步骤,即两存储控制器建立连接的步骤;

所述握手通信的步骤为:

首先存储控制器A的缓存池模块发送一个随机数a至存储控制器A的同步模块;

存储控制器A的同步模块将随机数a作为种子参数生成随机数b,存储控制器A的同步模块存储随机数对(a,b),并将随机数b发送到存储控制器B的同步模块;

存储控制器B的同步模块将随机数b作为种子参数生成随机数c,存储控制器B的同步模块存储随机数对(b,c),并将随机数c发送到存储控制器B的缓存池模块;

存储控制器B 的缓存池模块将收到随机数c加1发送至存储控制器B的同步模块,存储控制器B的同步模块收到Ack确认信号c+1后根据随机数对(b,c)返回Ack确认信号b+1;

依次进行该过程,直到存储控制器A收到Ack确认信号;

三、最后将两存储控制器中的数据写入到磁盘阵列中。

2. 根据权利要求1所述的一种基于SCST的双控缓存同步设计方法,其特征在于,经过握手步骤后,数据开始传输缓存,数据传输过程两部分内容:一是数据流,另一个是确认流数据;

当数据从存储控制器A传输到存储控制器B中时,存储控制器A将不断进行数据流发送,存储控制器B将确认收到的数据进行确认,发送确认流;

当存储控制器A未收到存储控制器B的确认信号时,就不断重发,直到收到Ack确认信号或达到阈值时间后停止。

3. 根据权利要求2所述的一种基于SCST的双控缓存同步设计方法,其特征在于,每一个存储控制器的数据来源均为客户机或另一存储控制器,当数据源为客户机时,将数据缓存到缓存池模块中,且在缓存前先查询缓存池模块中是否存在该数据:若数据已经存在于缓存池模块中,那么将待缓存数据丢弃,并发送Ack确认信号确定已经执行成功;若数据不在缓存池模块中,将数据保存到缓存池模块中,并将数据传输到同步模块中;同步模块将数据同步到其他存储控制器并收到Ack确认信号以后,修改缓存池模块中数据;

当数据源为另一存储控制器时,将数据缓存到缓存池模块中,且在缓存前先查询缓存

池模块中是否存在该数据：若数据已经存在于缓存池模块中，那么将待缓存数据丢弃，并发送Ack确认信号确定已经执行成功；若数据不在缓存池模块中，将数据保存到缓存池模块中，发送Ack确认信号以后修改缓存池模块中数据。

4. 根据权利要求3所述的一种基于SCST的双控缓存同步设计方法，其特征在于，存储控制器A的同步模块的数据发送到存储控制器B的同步模块中采用滑动窗口算法。

5. 根据权利要求4所述的一种基于SCST的双控缓存同步设计方法，其特征在于，所述滑动窗口算法的具体过程为：

 定义存储控制器A的同步模块为发送方，存储控制器B的同步模块为接收方，初始态，发送方没有帧发出，发送窗口前后沿相重合；接收方0号窗口打开，等待接收0号帧；

 发送方打开0号窗口，表示已发出0帧但尚确认返回信息，此时接收窗口状态不变；

 发送方打开0、1号窗口，表示0、1号帧均在等待确认之列；至此，发送方打开的窗口数已达规定限度，在未收到新的确认返回帧之前，发送方将暂停发送新的数据帧，接收窗口此时状态仍未变；

 接收方已收到0号帧，0号窗口关闭，1号窗口打开，表示准备接收1号帧，此时发送窗口状态不变；

 发送方收到接收方发来的0号帧确认返回信息，关闭0号窗口，表示从重发表中删除0号帧，此时接收窗口状态仍不变；

 发送方继续发送2号帧，2号窗口打开，表示2号帧也纳入待确认之列，至此，发送方窗口又已达规定限度，在未收到新的确认返回帧之前，发送方将暂停发送新的数据帧，此时接收窗口状态仍不变；

 接收方已收到1号帧，1号窗口关闭，2号窗口打开，表示准备接收2号帧，时发送窗口状态不变；

 发送方收到接收方发来的1号帧的确认信息，关闭1号窗口，表示从重发表中删除1号帧，此时接收窗口状态仍不变。

一种基于SCST的双控缓存同步设计方法

技术领域

[0001] 本发明涉及计算机服务器存储技术领域,具体地说是一种基于SCST的双控缓存同步设计方法。

背景技术

[0002] 随着计算机技术、网络技术的快速发展,对于数据存储的可靠性也逐渐得到了重视。现在的存储方式多采用价格较为磁盘组合成为巨大容量的磁盘组,配合数据分散排列的设计,将数据分割为不同的区段,分别进行存储,从而构成了磁盘阵列。这样系统的可靠性瓶颈为磁盘阵列控制器,目前的方式可以采用多路磁盘阵列控制器共享磁盘阵列的方式,这样既可以增加阵列系统的可靠性,又可以通过多路控制器对外进行数据存储服务,提高了存储效率。

[0003] 目前来说,基于SCST的网络存储构架方案中,SCST已经为不同类型的目标端驱动程序提供统一的接口,屏蔽不同类型驱动程序的差异性,便于多种目标端驱动程序以统一的方式和底层各种存储设备的连接。SCST的核心模块处于 Linux 存储结构的块设备层之上。SCST支持多种I/O模式,常用的方式为Block I/O和File I/O两种模式。采用Block I/O模式,可以绕开系统cache,创建独立的缓存池结构,并设计不同的缓存刷写算法实现硬盘的刷写。但是单台存储控制器存在单点失效的问题,影响了数据的可靠存储。因此,实现在多台控制器之间的缓存同步是一个必要的技术。

[0004] 基于此,本发明提供一种基于SCST的双控缓存同步设计方法,该方法采用SCST来处理 I/O 请求,将数据缓存到系统缓存池中。通过设计同步机制和同步算法,解决了存储控制器的单点失效引起的数据丢失问题。

发明内容

[0005] 本发明的技术任务是针对在现有技术的不足,提供一种基于SCST的双控缓存同步设计方法。

[0006] 本发明的技术方案是按以下方式实现的,该一种基于SCST的双控缓存同步设计方法,该方法通过以下步骤实现:

[0007] 将客户机连接到两个相同的存储控制器A和B中,两存储控制器内均设置SCST缓存区、缓存池模块和同步模块,两存储控制器均连接磁盘阵列;

[0008] 当客户机数据输入一存储控制器时,该存储控制器内的数据备份到另一存储控制器,实现两存储控制器内的缓存同步;

[0009] 然后将两存储控制器中的数据写入到磁盘阵列中。

[0010] 所述SCST缓存区以存储控制器主机的主存为存储器;缓存块大小按照内存页面的4KB 大小设计,缓存块由一个内存页组成。

[0011] 所述存储控制器内的数据同步缓存过程为:

[0012] 数据传输到存储控制器A,首先放入其缓存池模块中,该缓存池模块同时保存未写

入磁盘的数据和已写入磁盘的数据；

[0013] 缓存池模块中的数据发送到同步模块A与存储控制器B的同步模块中，最后传递到存储控制器B的缓存池模块中，使得每个存储控制器中的数据，既包含有本地存储控制器的缓存数据，也包含有另一存储控制器的缓存数据。

[0014] 所述存储控制器A与B进行数据缓存前，需要进行握手通信的步骤，该握手通信的步骤为：

[0015] 首先是存储控制器A的缓存池模块发送一个随机数a至存储控制器A的同步模块；

[0016] 存储控制器A的同步模块将随机数a作为种子参数生成随机数b，存储控制器A的同步模块存储随机数对(a, b)，并将随机数b发送到存储控制器B的同步模块；

[0017] 存储控制器B的同步模块将随机数b作为种子参数生成随机数c，存储控制器B的同步模块存储随机数对(b, c)，并将随机数c发送到存储控制器B的缓存池模块；

[0018] 存储控制器B 的缓存池模块将收到随机数c加1发送至存储控制器B的同步模块，存储控制器B的同步模块收到Ack确认信号c+1后根据随机数对(b, c)返回Ack确认信号b+1；

[0019] 依次进行该过程，直到存储控制器A收到Ack确认信号。

[0020] 经过握手步骤后，数据开始传输缓存，数据传输过程两部分内容：一是数据流，另一个是确认流数据；

[0021] 当数据从存储控制器A传输到存储控制器B中时，存储控制器A将不断进行数据流发送，存储控制器B将确认收到的数据进行确认，发送确认流；

[0022] 当存储控制器A未收到存储控制器B的确认信号时，就不断重发，直到收到ACK确认信号或达到阈值时间后停止。

[0023] 所述每一个存储控制器的数据来源均为客户机或另一存储控制器，当数据源为客户机时，数据首先缓存到缓存池模块中查询数据是否存在；若数据已经存在于缓存池模块中，那么将数据丢弃，并发送Ack确认信号确定已经执行成功；若数据不在缓存池模块中，将数据保存到缓存池模块中，并将数据传输到同步模块中；同步模块将数据同步到其他控制器并收到Ack确认信号以后，修改缓存池模块中数据；

[0024] 当数据源为另一存储控制器时，数据首先缓存到缓存池模块中查询数据是否存在；若数据已经存在于缓存池模块中，那么将数据丢弃，并发送Ack确认信号确定已经执行成功；若数据不在缓存池模块中，将数据保存到缓存池模块中，发送Ack确认信号以后修改缓存池模块中数据。

[0025] 存储控制器A的同步模块的数据发送到存储控制器B的同步模块中采用滑动窗口算法，定义存储控制器A的同步模块为发送方，存储控制器B的同步模块为接收方，该算法的具体过程为：

[0026] 初始态，发送方没有帧发出，发送窗口前后沿相重合；接收方0号窗口打开，等待接收0号帧；

[0027] 发送方打开0号窗口，表示已发出0帧但尚确认返回信息，此时接收窗口状态不变；

[0028] 发送方打开0、1号窗口，表示0、1号帧均在等待确认之列；至此，发送方打开的窗口数已达规定限度，在未收到新的确认返回帧之前，发送方将暂停发送新的数据帧，接收窗口此时状态仍未变；

[0029] 接收方已收到0号帧，0号窗口关闭，1号窗口打开，表示准备接收1号帧，此时发送

窗口状态不变；

[0030] 发送方收到接收方发来的0号帧确认返回信息，关闭0号窗口，表示从重发表中删除0号帧，此时接收窗口状态仍不变；

[0031] 发送方继续发送2号帧，2号窗口打开，表示2号帧也纳入待确认之列，至此，发送方窗口又已达规定限度，在未收到新的确认返回帧之前，发送方将暂停发送新的数据帧，此时接收窗口状态仍不变；

[0032] 接收方已收到1号帧，1号窗口关闭，2号窗口打开，表示准备接收2号帧，时发送窗口状态不变；

[0033] 发送方收到接收方发来的1号帧的确认信息，关闭1号窗口，表示从重发表中删除1号帧，此时接收窗口状态仍不变。

[0034] 本发明与现有技术相比所产生的有益效果是：

[0035] 本发明的一种基于SCST的双控缓存同步设计方法可将控制器的数据备份到其他控制器中，可以实现存储控制器数据的备份存储，缓存数据可以实现多物理传输介质的同步，满足海量阵列存储系统的吞吐量需求；当某一控制器出现故障时，其它控制器可继续进行数据可靠存储，保证了数据的安全性和可靠性，实用性强，易于推广。

附图说明

[0036] 附图1是本发明的总体实现框图。

[0037] 附图2是本发明的缓存同步数据流示意图。

[0038] 附图3是本发明的缓存同步过程流程图。

[0039] 附图4是本发明的滑动窗口算法示意图。

具体实施方式

[0040] 下面结合附图对本发明所提供的一种基于SCST的双控缓存同步设计方法作以下详细说明。

[0041] 本发明提出一种基于SCST的双控缓存同步设计方法，该模块需要与SCST结合，实现数据的缓存同步。本发明将用于存储控制器中，方案需要设计同步机制和同步算法，保持两个控制器之间的数据一致性。该方案包括两个模块，分别为缓存池模块Mempool与同步模块Sync Module，主要设计传递数据结构，同步算法和同步策略等。

[0042] 如附图1所示，该方法通过以下步骤实现：

[0043] 将客户机连接到两个相同的存储控制器A和B中，两存储控制器内均设置SCST缓存区、缓存池模块和同步模块，两存储控制器均连接磁盘阵列；

[0044] 当客户机数据输入一存储控制器时，该存储控制器内的数据备份到另一存储控制器，实现两存储控制器内的缓存同步；

[0045] 然后将两存储控制器中的数据写入到磁盘阵列中。

[0046] SCST缓存区以控制器主机的主存为存储器。缓存块大小按照内存页面的 4KB 大小设计，缓存块由一个内存页组成。

[0047] 从客户机Client的数据传输到存储控制器A，首先放入Mempool中。Mempool需要同时保存未写入磁盘的数据和已写入磁盘的数据（加快文件读）。Mempool中的数据发送到Sync

Module与存储控制器B的Syc Module中,最后传递到存储控制器B的Mempool中。该过程中需要实时保证数据的一致性,即存储控制器A与B的Mempool是一致的。同理,对于存储控制器B的数据也是执行相同的操作。对于每个存储控制器中的数据,既包含有本地存储控制器的缓存数据,也包含有其他存储控制器的缓存数据,因此每个存储控制器都有全部数据的备份,保证在单点失效的情况下,其他存储控制器也可以将数据写入磁盘阵列Disk Array中。

[0048] 如附图2所示,所述存储控制器A与B进行数据缓存前,需要进行握手通信Handshaking的步骤,该握手通信的步骤为:

[0049] 首先是存储控制器A的缓存池模块发送一个随机数a至存储控制器A的同步模块;

[0050] 存储控制器A的同步模块将随机数a作为种子参数生成随机数b,存储控制器A的同步模块存储随机数对(a,b),并将随机数b发送到存储控制器B的同步模块;

[0051] 存储控制器B的同步模块将随机数b作为种子参数生成随机数c,存储控制器B的同步模块存储随机数对(b,c),并将随机数c发送到存储控制器B的缓存池模块;

[0052] 存储控制器B 的缓存池模块将收到随机数c加1发送至存储控制器B的同步模块,存储控制器B的同步模块收到Ack确认信号c+1后根据随机数对(b,c)返回Ack确认信号b+1;

[0053] 依次进行该过程,直到存储控制器A收到Ack确认信号。

[0054] 上述过程为模块通信前的Handshaking,经过握手以后,数据可以传输,数据传输过程两部分内容:一个是数据流Data Stream,另一个是确认流ACK确认信号 Stream。

[0055] 当数据从存储控制器A传输到存储控制器B中时,存储控制器A将不断进行数据流发送,存储控制器B将确认收到的数据进行确认,发送确认流;

[0056] 当存储控制器A未收到存储控制器B的确认信号时,就不断重发,直到收到ACK确认信号或达到阈值时间后停止。

[0057] 如附图3所示,所述每一个存储控制器的数据来源均为客户机或另一存储控制器,当数据源为客户机时,数据首先缓存到缓存池模块中查询数据是否存在;若数据已经存在于缓存池模块中,那么将数据丢弃,并发送Ack确认信号确定已经执行成功;若数据不在缓存池模块中,将数据保存到缓存池模块中,并将数据传输到同步模块中;同步模块将数据同步到其他控制器并收到Ack确认信号以后,修改缓存池模块中数据;

[0058] 当数据源为另一存储控制器时,数据首先缓存到缓存池模块中查询数据是否存在;若数据已经存在于缓存池模块中,那么将数据丢弃,并发送Ack确认信号确定已经执行成功;若数据不在缓存池模块中,将数据保存到缓存池模块中,发送ACK确认信号以后修改缓存池模块中数据。

[0059] 如附图4所示,存储控制器A的同步模块的数据发送到存储控制器B的同步模块中采用滑动窗口算法,定义Syc Module A为发送方,Syc Module B为接收方,该算法的具体过程为:

[0060] 初始态,发送方没有帧发出,发送窗口前后沿相重合;接收方0号窗口打开,等待接收0号帧;

[0061] 发送方打开0号窗口,表示已发出0帧但尚确认返回信息,此时接收窗口状态不变;

[0062] 发送方打开0、1号窗口,表示0、1号帧均在等待确认之列;至此,发送方打开的窗口数已达规定限度,在未收到新的确认返回帧之前,发送方将暂停发送新的数据帧,接收窗口此时状态仍未变;

[0063] 接收方已收到0号帧，0号窗口关闭，1号窗口打开，表示准备接收1号帧，此时发送窗口状态不变；

[0064] 发送方收到接收方发来的0号帧确认返回信息，关闭0号窗口，表示从重发表中删除0号帧，此时接收窗口状态仍不变；

[0065] 发送方继续发送2号帧，2号窗口打开，表示2号帧也纳入待确认之列，至此，发送方窗口又已达规定限度，在未收到新的确认返回帧之前，发送方将暂停发送新的数据帧，此时接收窗口状态仍不变；

[0066] 接收方已收到1号帧，1号窗口关闭，2号窗口打开，表示准备接收2号帧，时发送窗口状态不变；

[0067] 发送方收到接收方发来的1号帧的确认信息，关闭1号窗口，表示从重发表中删除1号帧，此时接收窗口状态仍不变。

[0068] 上述具体实施方式仅是本发明的具体个案，本发明的专利保护范围包括但不限于上述具体实施方式，任何符合本发明的一种基于SCST的双控缓存同步设计方法的权利要求书的且任何所述技术领域的普通技术人员对其所做的适当变化或替换，皆应落入本发明的专利保护范围。

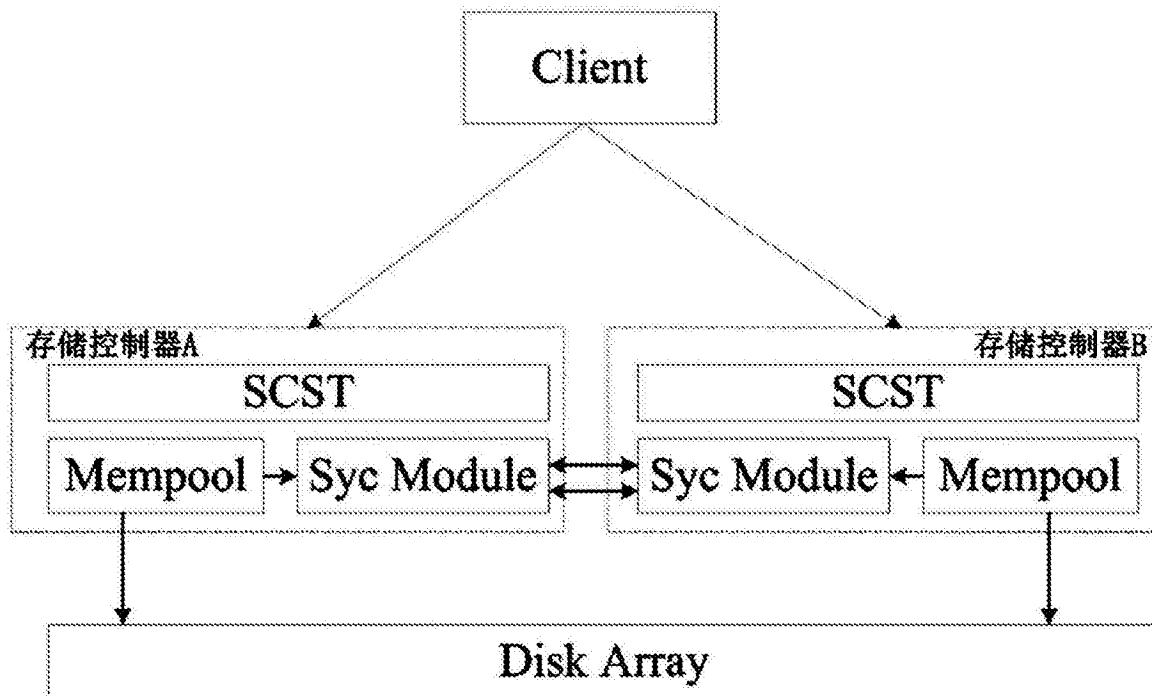


图1

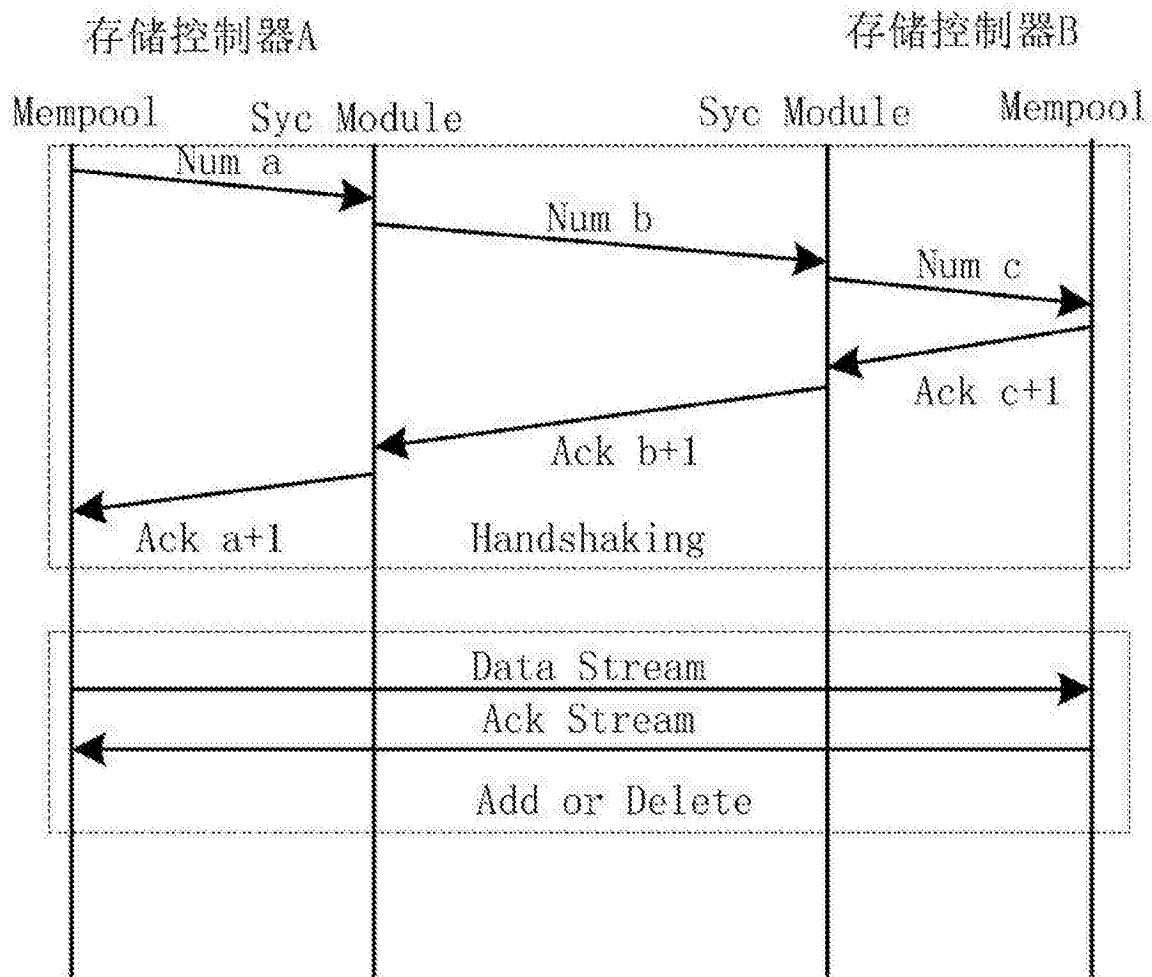


图2

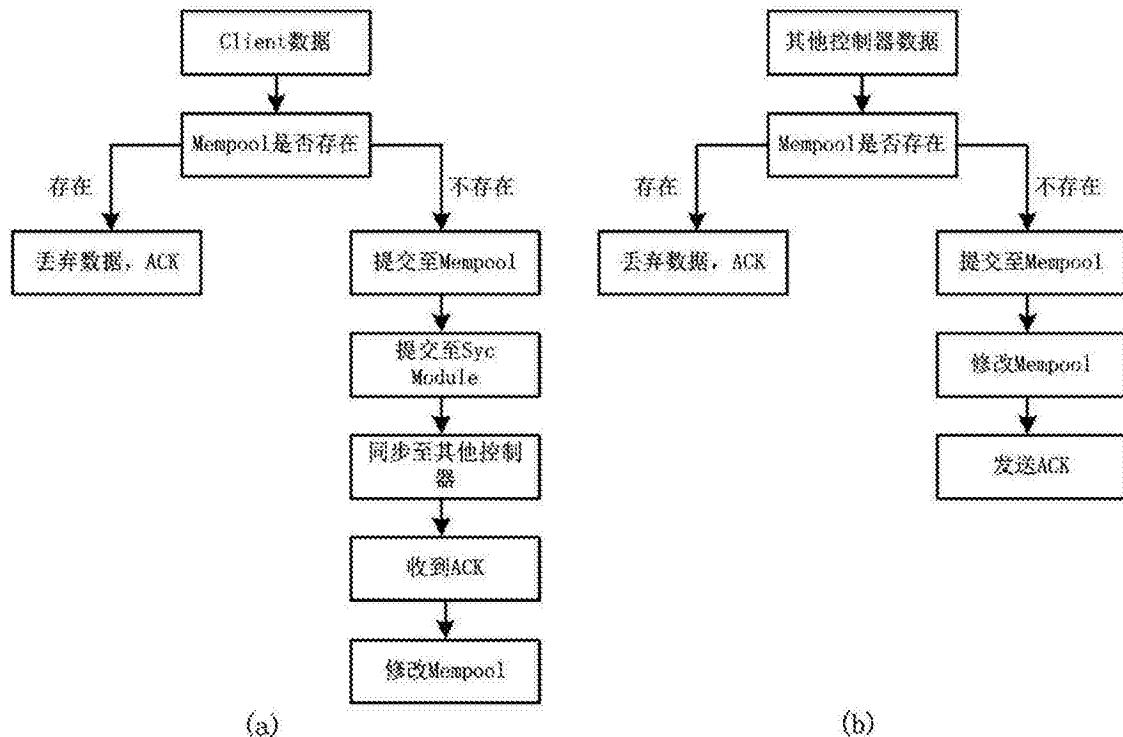


图3

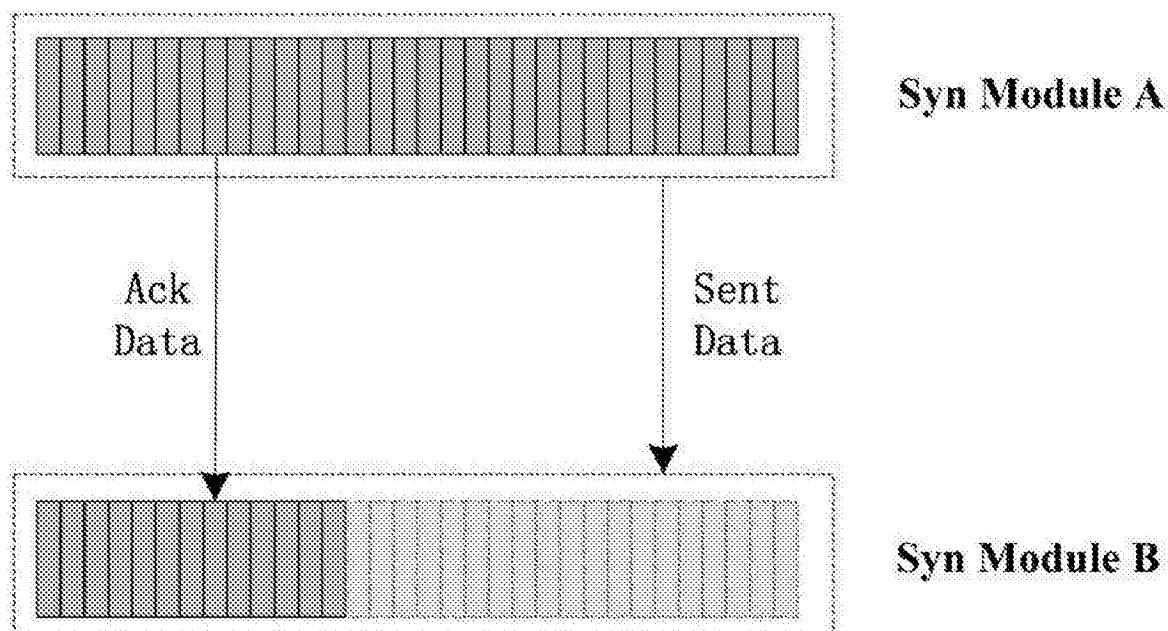


图4