高技术通讯 2012 年 第22 卷 第1 期:28 ~34

doi:10.3772/j.issn.1002-0470.2012.01.005

基于备份的 RAID5 在线重构框架①

徐 伟②* 朱旭东** 刘 浏***

(*国家安全生产监督管理总局通信信息中心 北京 100013)

(**浙江工商大学 杭州 310018)

(***中国科学院计算技术研究所 北京 100190)

摘 要 为解决国内外现有 RAID5 在线重构研究一直未解决的重负载持续访问下 RAID5 重构性能急剧恶化的问题,提出了利用外部存放的备份数据来加速 RAID5 在线重构的思想,构建了基于备份的 RAID5 在线重构框架。此框架利用备份系统所提供的恢复带宽将处于最近一次备份时间点的版本数据整合至 spare 盘,然后利用 RAID5 所提供的重构带宽将自最近一次备份时间点之后的已修改数据重构至 spare 盘。测试结果表明,该框架相对于现有重构方法显著改善了 RAID5 的重构性能和服务性能。

关键词 RAID5,在线重构,备份数据

0 引 言

磁盘阵列 RAID5 是一种兼顾数据存储性能、可靠性和存储成本的存储方案。应用负载所访问的磁盘阵列 RAID5 归于生产系统,由于生产系统直接对外提供服务,因此其可靠性和可用性非常重要,从而日常备份已经成为保证数据可靠性的一种常用手段。基于这种考虑,本文研究了基于备份的 RAID5 在线重构问题,提出了利用外部存放的备份数据来加速 RAID5 在线重构的思想,构建了基于备份的 RAID5 在线重构框架,并对其性能进行了分析。

1 相关知识

对 RAID5 的在线重构,国内外的研究主要是从预测失效、spare 布局、数据布局和重构策略四个方面来探索的。

预测失效主要指使用磁盘自动检测功能探测出即将不能正常运行的磁盘,在失效前将其复制到 spare 盘上,从而减少重构时间,提高磁盘阵列可靠性。大型 peer-to-peer 存储系统 Oceanstore 也采取了磁盘失效预测技术[1]。但这种方式不能完全准确地预测出磁盘失效。文献[2]指出,现有技术只能预告实际磁盘故障的 50%,其效果被夸大。

Spare 布局分为 Dedicated sparing[3]、Parity sparing[4]和 Distributed sparing[5-7]这三种方式。Dedicated sparing 模式专门用一块磁盘作为空闲盘,当磁盘阵列发生磁盘失效时,将失效磁盘上的数据完全重构到 spare 磁盘。Parity sparing 将 spare 磁盘作为第二块 parity 盘,减少 parity 组长度。当阵列中某块磁盘失效时,两个 parity 组融合生成一个更大的单一阵列,该阵列只有一个 parity 组。Distributed sparing 将 spare 空间分布在所有磁盘上,而不是专门用一个磁盘作为 spare 盘。当阵列中某块磁盘失效时,失效磁盘上的数据会被重构,并分布于所有磁盘的空闲空间。但当用户负载急剧增加时,磁盘阵列 RAID5 的重构性能显著降低,重构时间大幅增加。

数据布局研究主要采用 Decluster parity(cluster parity)[8,9]。Decluster parity 通过虚拟逻辑盘构成磁盘阵列,设物理磁盘个数为 C,虚拟逻辑盘个数为 G,将 $(G-1)/(C-1)$ 定义为 α,C 和 G 确定了 parity 数据消耗总磁盘空间,α 确定了系统的重构性能。α 值越小,重构所需时间则越少,parity 数据所占比例越大;α 值越大,重构所需时间则越大,parity 数据所占比例越小。可以调节 α,从而在重构和 parity

① 863 计划(2009AA01A403)资助项目。

② 男,1979 年生,博士;研究方向:网络存储;联系人,E-mail:xuw@chinasafety.gov.cn

(收稿日期:2010-03-17)

数据所占比例间取得平衡。但是,当 α 值很小时,用于 parity 数据的存储开销将非常大,而且当用户负载很大时,重构所需时间仍然相当大。

重构策略[10]主要由重构对象、重构顺序、重构与服务协作这三个方面组成。

重构对象主要分为面向条带重构、并行条带重构和面向磁盘重构。面向条带重构是按条带来进行重构的,并行条带重构采取多个并行的面向条带重构,面向磁盘重构采取与阵列中磁盘相同个数的进程,一个进程对应一个磁盘。但是,当用户负载较大时,重构所需时间仍然相当长,且对服务性能影响显著。文献[11]描述了基于磁道的重构算法,利用该算法可以重构磁道上丢失数据。

重构顺序主要分为 head-following、closet active stripe、multiple reconstruction points 和基于局部性的多线程重构。head-following 重构的主要原理是:重构磁盘总是从处于低地址的尚未重构单元进行顺序重构。Closest active stripe 重构的主要原理是:在完成用户请求时,总是从靠近磁头位置的尚未重构单元进行重构。Multiple reconstruction points 重构的主要原理是:将磁盘分为多个重构段,在完成用户请求时,从当前最近重构点开始重构。基于局部性的多线程重构的主要原理是:利用用户负载访问的局部性,优先重构频繁访问的区域[12]。

重构与服务协作方面主要分为用户请求操作、重构速率控制两方面。用户请求操作处理方式分为直读、回写和直写。直读主要原理是:如果用户对失效磁盘读请求所涉及数据已经被重构并已经在 spare 盘上,则直接从 spare 盘上读取该数据。回写的主要原理:用户对失效磁盘读请求重构出数据,该数据不仅被发送给用户,而且被写到 spare 盘上。直写的主要原理是:用户的写请求直接发送到 spare 盘上,写入 spare 盘的相应位置。重构速率控制主要在服务性能和重构速率之间寻找平衡点,从而使得服务性能处于用户可承受范围,尽可能提高重构速率[13,14]。文献[15,16]指出可以根据磁头移动轨迹来利用空闲带宽处理后台应用,这也有助于提高重构性能和服务性能。

尽管,国内外 RAID5 在线重构研究一直试图从预测失效、spare 分布、数据分布、重构策略这四个方面来提高磁盘阵列服务性能和重构性能,但是,在重负载持续访问的情况下 RAID5 的重构性能和服务性能急剧恶化这一问题始终没有得到解决。而企业存储网络系统多数由生产系统和备份系统构成,磁盘备份技术已被广泛用于保证数据可靠性,因此,本文基于"利用外部存放的备份数据来加速 RAID5 在线重构"的思想构建了基于备份的 RAID5 在线重构框架,以充分挖掘闲置磁盘备份的强大 IO 能力,显著提高繁忙生产系统内 RAID5 在线重构的速度。

2　设计思想

基于备份的 RAID5 在线重构框架的核心思想是:当生产系统磁盘阵列 RAID5 出现磁盘失效时,可以由备份系统虚拟出失效磁盘处于最近备份时间点的历史版本,通过将历史版本恢复至 spare 磁盘上,从而使得 spare 磁盘成为处于最近备份时间点的版本,最后,利用生产系统虚拟出的当前版本将失效磁盘上自最近一次备份时间点之后已修改数据重构至 spare 磁盘上。此框架的主要优点是:利用备份系统所提供的稳定恢复带宽,显著降低了应用负载对重构过程的影响,同时,显著减少了磁盘阵列 RAID5 参入重构,使得磁盘阵列 RAID5 优先满足用户服务。

3　设计实现

以蓝鲸虚拟存储系统(blue whale virtual storage device system, BW-VSDS)和蓝鲸备份系统(blue whale backup system, BW-BS)为基础平台,实现了基于备份的 RAID5 在线重构框架。

此框架的原型系统主要由面向数据的磁盘阵列架构、映射管理、恢复管理和重构管理四部分构成,通过这四部分协作,完成了磁盘阵列 RAID5 在线重构过程。重构过程分为两个阶段:版本恢复阶段和版本修复阶段。当生产系统出现磁盘失效时,激活 spare 磁盘,如图 1(a)所示。首先,重构过程进入版本恢复阶段, spare 硬盘从生产系统中换出,完全由备份系统将失效磁盘处于最近备份时间点的历史版本恢复至 spare 磁盘,从而将 spare 磁盘恢复成失效磁盘处于最近备份时间点的版本,如图 1(b)所示。然后,重构过程进入版本修复阶段,将 spare 磁盘重新加入到生产系统磁盘阵列中,利用虚拟当前版本将 spare 磁盘修复为当前版本,从而完成失效磁盘上自最近一次备份时间点之后已修改数据的重构,如图 1(c)所示。通过版本恢复阶段和版本修复阶段,完成 spare 磁盘的数据重构,从而将生产系统恢复到正常运行状态,如图 1(d)所示。

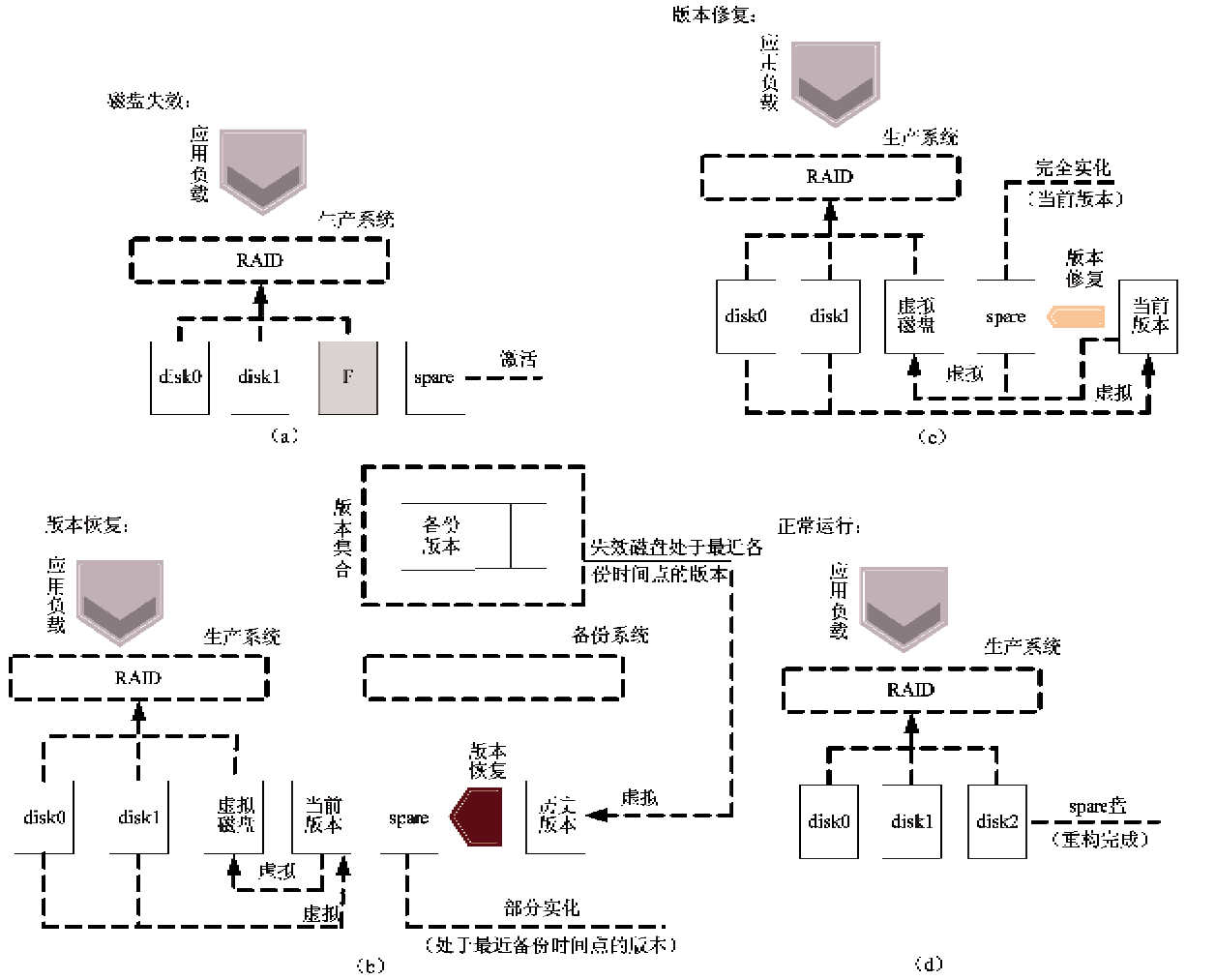

图 1 基于备份的 RAID5 在线重构框架的重构过程

3.1 面向数据的磁盘阵列架构

通过对传统磁盘阵列进行简单的改进,我们实现了面向数据的磁盘阵列架构。面向数据的磁盘阵列架构与传统磁盘阵列架构的根本差异是面向数据的磁盘阵列中未被使用的逻辑块上的数据一定为零。我们使用一个全局位图记录磁盘阵列中所有逻辑块(数据块和校验块)的使用情况,并通过在传统磁盘阵列转发读写请求路径上添加访问位图接口,从而实现了面向数据的磁盘阵列架构。面向数据磁盘阵列必须按照条带分配和释放逻辑单元,以避免引起附加的读写操作。

3.2 映射管理

映射管理主要负责以下三类映射关系的建立和维护:(1) 逻辑卷逻辑块与最新备份数据块的映射关系;(2) 失效磁盘逻辑块与最新备份数据块的映射关系;(3) 失效磁盘上自最近备份时间点之后修改数据块的位置标识。

图 2(a)描述了版本、逻辑卷、磁盘阵列和磁盘之间的映射关系。假设:磁盘阵列 RAID5 由 disk _ 0、disk _ 1、disk _ 2 三块磁盘构成;磁盘上逻辑块、RAID5 上 chunk、空间分配粒度和备份数据块都为 4KB;一个逻辑卷 lv1 已被创建;版本 version _ 1 附属于逻辑卷 lv1;位图 lv1-bmp 为自最近一次备份时间点之后的差别增量位图。

如图 2(a)所示,逻辑卷 lv1 的第 1 块映射到磁盘 disk _ 2 的第 1 块(d21),lv1 其余逻辑块与磁盘上逻辑块的映射关系不再赘述;逻辑卷 lv1 的 version _ 1 上数据块 v110,根据 version _ 1 的位图,v110 对应 lv1 的第 1 块数据(处于 version _ 1 时刻),其余版本所保存的数据块与逻辑卷 lv1 逻辑块的映射关系不再赘述;根据位图 lv1-bmp,逻辑卷 lv1 的第 1 个逻辑块自最近一次备份时间点之后被修改;磁盘逻辑块 d00、d10、p20 和 d01 未被使用,其上数据为零。

图 2(b)描述了生产系统上磁盘逻辑块与备份数据块之间映射关系。根据逻辑卷逻辑块与最新备

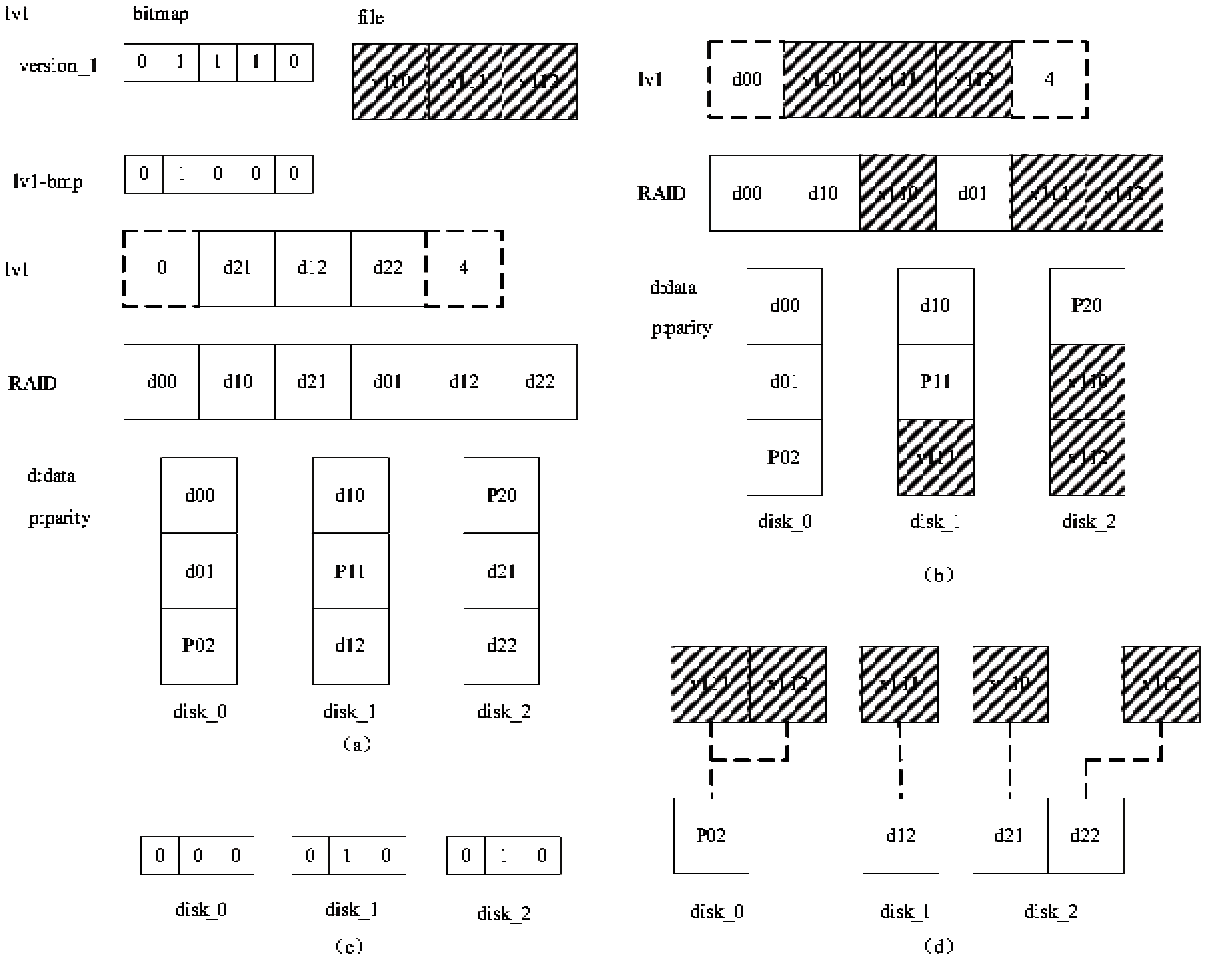

图 2　生产系统磁盘阵列 RAID5 上磁盘逻辑块与最新备份数据块的映射关系

份数据块的映射关系、逻辑卷与磁盘阵列逻辑块的映射关系以及磁盘阵列与磁盘逻辑块的映射关系,可以计算出任一磁盘逻辑块与最新备份数据块的映射关系。根据图 2(a)所描述的映射关系,计算出 disk _0,disk _1 和 disk _2 这三块磁盘的数据块与备份数据块的映射关系,如图 2(b)所示。在图 2(b)中,备份数据块 v110 映射在 disk _2 的第 1 块(d21)上,其余不再赘述。

图 2(c)描述了各磁盘上自最近一次备份时间点之后修改数据的位置标识(版本修复位图)。在图 2(b)中,根据位图 lv1-bmp,可知磁盘 disk _2 上逻辑块 d21 上数据自最近一次备份之后已被修改,因此,磁盘 disk _1 上逻辑块 p11 上数据自最近一次备份之后已被修改,因此,生成了磁盘 disk _0、disk _1 和 disk _2 上自最近一次备份时间点之后修改数据的位置标识。

图 2(d)中描述了各磁盘上逻辑块与最新备份数据块的映射关系。如图 2(b)所示,磁盘 disk _0 上逻辑块 p02 对应备份数据块 v111 和 v112;磁盘 disk _1 上逻辑块 d12 对应备份数据块 v111;磁盘 disk _2 上逻辑块 d21 对应备份数据块 v110,磁盘 disk _2 上逻辑块 d22 对应备份数据块 v112,因此,生成了磁盘 disk _0、disk _1 和 disk _2 上逻辑块与最新备份数据块的映射关系。

3.3　恢复管理

恢复管理构件由多个读线程和一个写线程构成,映射管理构件由一个线程构成。当恢复管理构件进行 spare 磁盘的恢复操作时,映射管理构件并行生成映射关系和版本修复位图。

恢复管理构件采取多个读线程并发读取备份数据。一个读线程从映射缓冲内顺序取出一个映射关系,对于奇偶校验码,读线程将所有相关数据(若干块备份数据)读出,并异或生成奇偶校验码,存放于数据缓冲相应位置;对于数据,则将相应数据(一块备份数据)读出,存放于数据缓冲相应位置。当某个读线程完成数据读取后,则顺序处理还未读取的数据。

恢复管理构件采用单个写线程将数据顺序写入

spare 磁盘。恢复管理构件有两个数据缓冲,当读线程读取数据并存放于一个缓冲时,写线程将已放满数据的另一个缓冲中数据写入 spare 盘中。

3.4 重构管理

在此原型系统中,重构管理构件仅负责版本修复阶段。通过版本修复位图和辅助位图两者合作,重构管理构件实现了版本修复功能。版本修复位图是自最近备份时间点之后修改数据的位置标识,辅助位图上所有位初始为 0,两个位图上每一位与磁盘上 4KB 逻辑块一一对应。通过版本修复位图和辅助位图之间合作,可以判断逻辑块上数据是否有效。spare 磁盘逻辑块上数据被定义了两种状态:

VALID——如果版本修复位图中某位为 0 或者辅助位图中某位为 1,则对应逻辑块上数据有效。

INVALID——如果版本修复位图中某位为 1,且辅助位图中相应位为 0,则对应逻辑块上数据无效,即逻辑块上数据自最近一次备份时间点之后已被修改。

在版本修复阶段,版本修复位图和辅助位图协作过程如下:

(1) 按照版本修复位图和辅助位图,顺序发出对于无效数据块的重构请求。当所计算出的数据被写入 spare 磁盘对应逻辑块之后,则将辅助位图上相应位设为 1,从而标识对应逻辑块上数据已经有效。

(2) 用户读请求访问 spare 磁盘时,被访问逻辑块上数据有效,则直接访问 spare 磁盘;否则,通过同一条带上其余数据块构建出被访问的数据块,并将其写入 spare 盘,同时,将辅助位图上相应位设为 1,从而标识被访问逻辑块上数据已经有效。

(3) 用户写请求访问 spare 磁盘时,将版本修复位图上相应位设为 1,并直接将写请求发送给 spare 磁盘;写请求完成之后,则将辅助位图上相应位设为 1,标识被访问逻辑块上数据已经有效。

通过面向数据的磁盘阵列架构、映射管理、恢复管理和重构管理四个构件,我们实现了基于备份的 RAID5 在线重构框架的原型系统。

4 性能评价

本文主要用面向磁盘重构(disk-oriented reconstruction,DOR)算法与基于备份的 RAID5 在线重构框架进行对比。因为 DOR 方法是现有重构方法中最有效的算法之一,而且已经被实现于许多软 RAID5 和硬 RAID5 产品中,并且得到最广泛的研究。

本节的测试配置如下:cello99(12-25、03-11、06-24、09-22 和 10-05)、F2. spc、F1. spc 和 tpcc94 应用模式;对于 cello99、F2. spc、F1. spc 和 tpcc94 这四种应用模式,生产系统 RAID5 单块磁盘容量分别设为 92GB、37GB、31GB 和 11GB;生产系统 RAID5 第 3 块磁盘失效;4KB、8KB 和 16KB 备份版本集合;每个逻辑卷对应备份版本集合包含了 184 个版本;备份系统 RAID5 由 6 块磁盘构成;生产系统 RAID5 由 9 块磁盘构成;生产系统运行时段分为工作时段(白天工作时段:8:00 - 20:00)和备份时段(晚上空闲时段:20:00 - 8:00);恢复管理模块里读线程数目设置为 30,写线程数目设置为 1。

4.1 重构性能

图 3 描述了基于备份的 RAID5 在线重构框架的重构性能相对于 DOR 提高的倍数。图中,4KB-Sl-W 表示备份版本集合的备份粒度为 4KB,生产系统运行时段为工作时段;4KB-Sl-B 表示备份版本集合的备份粒度为 4KB,生产系统运行时段为备份时段。8KB-Sl-W、8KB-Sl-B、16KB-Sl-W、16KB-Sl-B 含义与 4KB-Sl-W、4KB-Sl-B 可作类似解释。从图 3 可以得出以下结论:对于每天修改数据量较少且服务负载压力较大的应用模式,相对于 DOR,此框架将显著提高重构性能;对于每天修改数据量较多的应用模式,相对于 DOR,此框架对重构性能的提高幅度将显著降低。在图 3 中,对于 cello99-12-25 应用负载(每天修改数据量较小且负载压力较小),在工作时段,此框架的重构性能比 DOR 提高了 1.3 ~ 3 倍,即使在备份时段,此框架的重构性能也比 DOR 提高了 1 ~ 2.7 倍;而对于 cello99-10-05 应用负载(每天修改数据量较小且负载压力较大),在工作时段,此框架的重构性能比 DOR 提高了 11 ~ 14 倍,在备份时段,此框架的重构性能比 DOR 提高了 8 ~ 10 倍;对于 F2. spc 应用模式(每天修改数据量较小且负载压力较大),在工作时段,此框架的重构性能比 DOR 提高了 5 ~ 10 倍,在备份时段,此框架的重构性能比 DOR 提高了 4 ~ 9 倍;即使对于 F1. spc 和 tpcc94 应用模式(每天修改数据量非常大),在 F1. spc 应用负载下,在工作时段,此框架的重构性能也比 DOR 提高了 2.3 ~ 2.7 倍左右,在备份时段,此框架的重构性能也比 DOR 提高了 1.6 ~ 1.9 倍;在 tpcc94 应用负载下,在工作时段此框架的重构性能也比 DOR 提高了 1.7 倍左右,在备份时段,此框架的重构性能也比 DOR 提高了 1.1 倍左右。

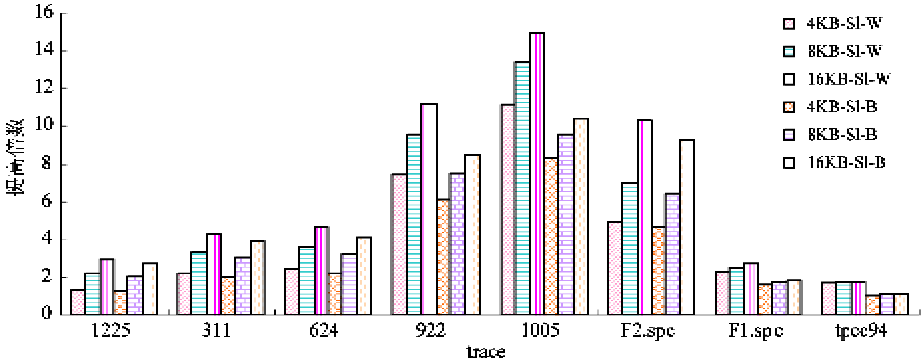

图 3　基于备份的 RAID5 在线重构框架的重构性能相对于 DOR 提高的倍数

当磁盘空间远大于实际分配空间时,相对于 DOR,此框架对重构性能有巨大改善。

4.2　服务性能

图 4 描述了基于备份的 RAID5 在线重构框架的总体平均响应时间相对于 DOR 降低的百分比。从图 4 可以得出以下结论:此框架相对于 DOR 显著改善了系统服务性能,提高了系统可用性。在图 4 中,对于 cello99 的五种应用模式,在工作时段,此框架的总体平均响应时间比 DOR 降低了 18% ~34%,在备份时段,此框架的总体平均响应时间比 DOR 降低了 17% ~33%;对于 F2. spc 应用模式,在工作时段,此框架的总体平均响应时间比 DOR 降低了 30%,在备份时段,此框架的总体平均响应时间也比 DOR 降低了 30% 左右;对于 F1. spc 应用模式,在工作时段,此框架的总体平均响应时间比 DOR 降低了 44% 左右,在备份时段,此框架的总体平均响应时间也比 DOR 降低了 35% 左右;在 tpcc94 应用负载下,在工作时段,此框架的总体平均响应时间比 DOR 降低了 20% 左右,在备份时段,此框架的总体平均响应时间也比 DOR 降低了 13% 左右。

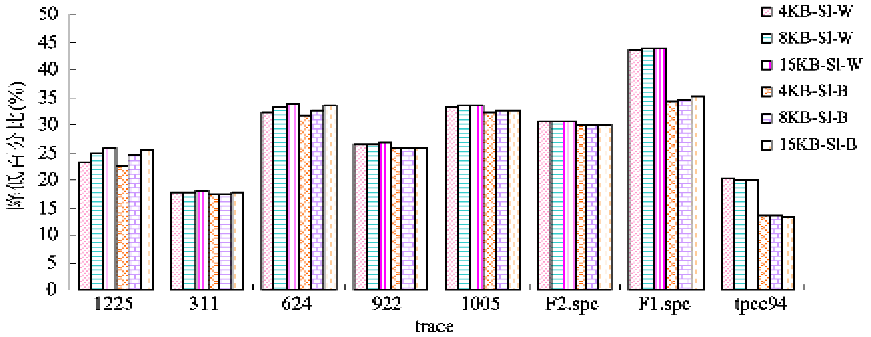

图 4　基于备份的 RAID5 在线重构框架的总体平均响应时间相对于 DOR 降低百分比

当磁盘空间远大于实际分配空间时,相对于 DOR,此框架对服务性能有巨大改善。

5　结　论

基于备份的 RAID5 在线重构框架利用备份系统所提供的稳定恢复带宽,显著降低了应用负载对重构过程的影响,同时,显著减少了磁盘阵列 RAID5 参入重构,使得磁盘阵列 RAID5 优先满足用户服务,显著改善了即时服务性能。相对于 DOR 算法(现在最常用且最有效重构算法之一),此框架将重构性能提高了 1.1 ~18 倍,平均响应时间(服务性能评价指标)改善了 3.5% ~44%。

下一步工作将根据应用负载的基本信息,考虑当前的系统服务性能和应用负载特征,确定磁盘阵列 RAID5 内重构带宽和备份系统上恢复带宽的利用比例,并能够随着应用负载变化进行自适应调整,以便在服务性能和重构性能之间取得平衡,完成数据重构。

参考文献

[1] Rhea S, Wells C, Eaton P, et al. Maintenance-free global data storage. *IEEE Internet Computing*, 2001, 5(5): 40 - 49

[2] Pinheiro E, Weber W D, Barroso L A. Failure trends in

a large disk drive population. In: Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07), San Jose, USA, 2007,4(3):33-48

[3] Menon J, Mattson D. Comparison of sparing alternative for disk arrays. In: Proceedings of the 19th International Symposium on Computer Architecture, Queensland, Australia,1992. 318-329

[4] Narasimha R A L, Chandy J, Banerjee P. Design and evaluation of gracefully degradable disk arrays. *Journal of Parallel and Distributed Computing*, 1993, 17: 28 - 40

[5] Thomasian A, Menon J. Performance analysis of RAID5 disk arrays with a vacationing server model for rebuild mode operation. In: Proceedings of the 10th International Conference on Data Engineering, Houston, USA, 1994. 111-119

[6] Xin Q, Miller E L, Schwarz T J E. Evaluation of distributed recovery in large-scale storage systems. In: Proceedings of the 13th International Symposium on High-performance Distributed Computing, Honolulu, USA, 2004, 2(1):172-181

[7] Alexander Thomasian. Comment on RAID5 performance with distributed sparing. *IEEE Transactions on Parallel and Distributed Systems*, 2006, 17(4):399-400

[8] Muntz R, Lui J. Performance analysis of disk arrays under failure. In: Proceedings of the 16th International Conference on Very Large Data Bases, Queensland, Australia, 1990. 162-173

[9] Watanabe A, Yokota H. Adaptive overlapped declustering: a highly available data-placement method balancing access load and space utilization. In: Proceedings of the 21th International Conference on Data Engineering, Tokyo, Japan, 2005. 828-839

[10] Fu G, Thomasian A, Han C Q, et al. Rebuild strategies for redundant disk arrays. In: Proceedings of the 12th Conference on Mass Storage Systems and Technologies, Greenbelt, USA, 2004, 2: 128-139

[11] Lee J Y B, Lui J C S. Automatic recovery from disk failure in continuous-media servers. *IEEE Transactions on Parallel and Distributed Systems*, 2002, 3: 499 - 515

[12] Tian L, Feng D, Jiang H, et al. PRO: a popularity-based multi-threaded reconstruction optimization for RAID-structured storage systems. In: Proceedings of the 5th USENIX Conference on File and Storage Technologies, San Jose, USA, 2007. 89-103

[13] Bachmat E, Schindler J. Analysis of methods for scheduling low priority disk drive tasks. In: Proceedings of the Special Interest Group on Measurement and Evaluation (SIGMETRICS), Marina Del Rey, USA, 2002. 103-111

[14] Tian L, Jiang H, Feng D, et al. Implementation and evaluation of a popularity-based reconstruction optimization algorithm in availability-oriented disk arrays. In: Proceedings of the 24th IEEE Conference on Mass Storage Systems and Technologies, San Diego, USA, 2007, 1: 233-238

[15] Lumb C R, Schindler J, Ganger G R, et al.. Towards higher disk head utilization: extracting free bandwidth from busy disk drives. In: Proceedings of the 4th Symposium on Operating System Design & Implementation, San Diego, USA, 2000. 7-7

[16] Thereska E, Schindler J, Bucy J, et al. A framework for building unobtrusive disk maintenance applications. In: Proceedings of the 3rd USENIX Conference on File and Storage Technologies, San Francisco, USA, 2004, 2: 213-226

The on-line reconstruction framework of RAID5 based on backup

Xu Wei*, Zhu Xudong**, Liu Liu***

(*Communication Information Center, State Administration of Work Safety, Beijing 100013)

(**Zhejiang Gongshang University, Hangzhou 310018)

(***Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

Abstract

To solve the unsolved problem in present research on RAID5 reconstruction that the reconstruction performance of RAID5 deteriorates rapidly under the continuous heavy workload, this thesis proposes the idea of accelerating the on-line reconstruction of RAID5 by use of the backup data in external storages, and completes the on-line reconstruction framework of RAID5 based on backup. The framework utilizes the restored bandwidth provided by the backup system for integrating the version data at the latest backup time point into the spare disk, and then utilizes the reconstruction bandwidth provided by RAID5 for reconstructing the modified data after the latest backup time point to the spare disk. The test result shows that compared with the existing reconstruction methods, the framework greatly improves the reconstruction performance and service performance of RAID5.

Key words: RAID5, on-line reconstruction, backup data