

一种网络备份系统的数据一致性检测方法*

易固武, 刘晓洁, 李 涛, 卢正添, 葛 亮, 周 煜
(四川大学 计算机学院, 成都 610065)

摘 要: 为了消除原始数据持续变化对数据一致性检测带来的影响, 设计并实现了一种基于 Windows 操作系统的网络备份系统数据一致性检测方法。本方法在一致性检测过程中使用实时的 Windows 卷影拷贝数据来替代原始的动态变化数据执行与备份数据的一致性检测。通过实验结果证明, 本方法的检测结果和使用原始数据与备份数据进行检测的结果一致。

关键词: 一致性检测; 数据备份; 卷影拷贝

中图分类号: TP393 文献标志码: A 文章编号: 1001-3695(2008)12-3700-02

Data consistency checking method in network backup system

YI Gu-wu, LIU Xiao-jie, LI Tao, LU Zheng-tian, GE Liang, ZHOU Yu
(School of Computer, Sichuan University, Chengdu, 610065, China)

Abstract: In order to eliminate the affection that the ever-changing original data bring to MS- the data consistency checking in the network backup system, this paper proposed a data consistency checking method based on MS-Windows OS. This method used a real-time volume shadow copy to check data consistency with the backup data instead of the ever-changing original data. The experiment shows that the result is the same as using original volume data.

Key words: data consistency checking; data backup; volume shadow copy

在网络数据备份中, 用户关注某一时刻的备份数据是否与对应时刻的原始数据一致。如果在同一时刻两边数据不一致, 就意味着备份数据不可用于恢复, 那么整个备份变得毫无意义。因此需要对原始数据和备份数据作一致性检测。但在网络数据备份中因为网络传输存在时间延迟, 即在 T_i 时刻原始数据发生的变化将在 $T'_i = T_i + \Delta T$ ($\Delta T > 0$) 时刻反映到备份数据中。从理论上讲在同一时刻两地的数据原本就是不一致的, 这导致在实践中数据一致性检测变得相当困难。为了验证两边数据是否一致, 目前常用的数据一致性检测方法均要求保证在检测时原始数据和备份数据不发生变化且在备份网络中没有数据的传输。在实际应用中, 这就要求在一致性检测开始前和检测进行过程中不仅要停止所有能对数据发生改变的业务, 而且还要停止对数据的备份。随着检测数据量的增大, 停止业务和备份时间会随之增长, 这对数据实时性和安全性要求高的企业是不能忍受的。为此, 本文设计了一种针对动态变化数据的一致性检测方法。

1 网络备份系统结构

网络备份系统在物理上分为本地数据生产中心和远程数据备份中心两个部分, 如图 1 所示。本地数据生产中心是数据备份任务的发起端; 远程数据备份中心是备份数据的接收端。备份任务就是将备份任务发起端指定的数据通过 Internet 传输到接收端存储。

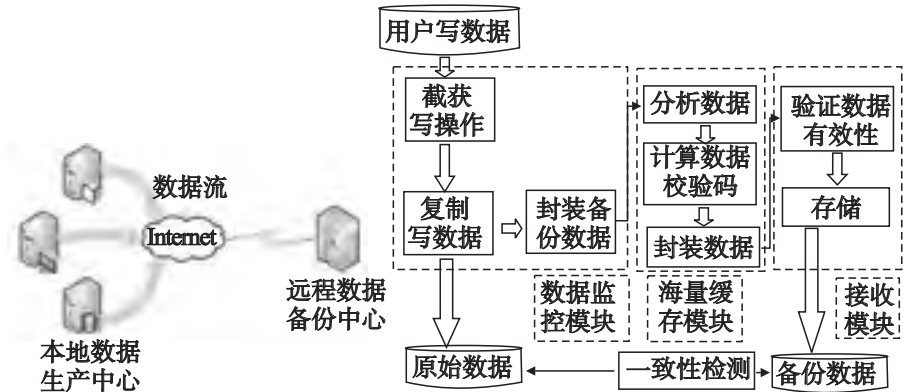
如图 2 所示, 本系统从功能上被划分为四个功能模块:

a) 数据生产中心监控模块(数据监控模块)。它负责截获监控卷的写操作 I/O 请求包 (IRP), 提取出 IRP 中的有效数据, 将有效数据按照固定格式封装写入海量缓存中。

b) 海量缓存模块。它负责分析写入海量缓存中的数据, 按照远程备份中心对数据的要求再次封装后发送给数据备份中心。

c) 远程备份中心数据接收模块(接收模块)。它负责接收来自数据生产中心的数据, 验证数据在传输过程中是否出现差错, 然后将其存储供以后使用。

d) 一致性检测模块。它用于验证备份中心的数据是否与数据生产中心的数据一致。



在备份任务启动前先将监控卷上的已有数据备份到数据备份中心, 使得在监控开始时备份中心数据与监控卷数据一致。

收稿日期: 2008-03-12; 修回日期: 2008-05-23 基金项目: 国家自然科学基金资助项目(60573130, 60502011); 国家“863”计划资助项目(2006AA01Z435); 国家教育部博士点基金资助项目(20070610032)

作者简介: 易固武(1983-), 男, 四川绵竹人, 硕士研究生, 主要研究方向为网络安全(zeaphyr@126.com); 刘晓洁(1965-), 女, 江苏南京人, 副教授, 主要研究方向为网络安全; 李涛(1965-), 男, 四川岳池人, 教授, 博导, 主要研究方向为网络安全及计算智能、信号处理; 卢正添(1981-), 男, 福建龙岩人, 讲师, 主要研究方向为网络安全; 葛亮(1983-), 男, 四川成都人, 硕士研究生, 主要研究方向为网络安全; 周煜(1985-), 男, 四川成都人, 硕士研究生, 主要研究方向为网络安全。

2 快照原理

快照是指将一个具体数据对象在一个时刻上的全部信息保留一个副本,使得在快照后对数据对象的修改不会影响到副本数据。在目前快照应用技术中,主要的实现方式大致分为三种,即中断镜像、按需复制和虚拟映像。其中,中断镜像和按需复制获得快照是原数据的物理拷贝,故其需要与原数据大小相同的空间来存放原数据的所有信息;而虚拟映像只保存了在快照建立后原数据中发生改变的部分,使得虚拟映像占用的存储空间一般只有原数据空间的 20% ~30%,但其快照依赖于原始数据而存在。

Windows 卷影拷贝服务 (VSS) 创建的快照是虚拟映像的一种。其快照生成采用写时拷贝 (copy-on-write) 技术,即在原始数据被修改时先将修改前的数据拷贝出来存储,再将修改后的数据写入。

3 动态数据一致性检测方法

对数据一致性检测定义如下: 设原始数据集 $P = \{ a, b, c \}$, 备份数据集 $P' = \{ a', b', c' \}$ 。 $P_i = (a_i, b_i, c_i)$ 为在 T_i 时刻原始数据集 P 的状态; $P'_i = (a'_i, b'_i, c'_i)$ 为备份数据集在 T'_i 时刻的状态。当 $\forall T_i = T'_i \rightarrow P_i = P'_i$ (式 (1)) 时,表示一致性检测通过。否则当 $\forall T_i = T'_i \rightarrow P_i \neq P'_i$ (式 (2)) 时,意味着备份数据和原始数据不一致,一致性检测未通过。

3.1 一致性检测算法描述

设当前监控卷数据集 D , 备份中心对应监控卷的备份数据集 D' ; 则备份数据集 D' 是 D 的像且 $D \rightarrow D'$ 是一个双射。集合 $D_1, D_2, \dots, D_n$ 和 $D'_1, D'_2, \dots, D'_n$ 分别为 D 和 D' 的子集且满足

$$\begin{aligned} \bigcup_{i=1}^n D_i &= D, \quad \bigcap_{i=1}^n D_i = \emptyset, \quad \bigcup_{i=1}^n D'_i = D', \quad \bigcap_{i=1}^n D'_i = \emptyset \\ \forall D &= D \Leftrightarrow D_1 = D'_1, \quad D_2 = D'_2, \dots, D_n = D'_n \end{aligned}$$

定义函数 $f(x)$ 为求集合 x 的摘要值, 则有 $\forall D = D' \Rightarrow f(D_1) = f(D'_1), \dots, f(D_n) = f(D'_n)$ (式 (3))。根据式 (1) ~ (3), 一致性检测表述为

- a) 当 $\forall T_i = T'_i, \quad \forall D_j \in D, D_j' \in D'$ 时, $\forall i = j \Rightarrow f(D_i) = f(D'_i) \Leftrightarrow D = D'$ 表示原始数据与备份数据相同, 一致性检测通过。
- b) 当 $\forall T_i = T'_i, \quad \exists D_i \in D, D_i' \in D'$ 时, $\exists i = j \Rightarrow f(D_i) \neq f(D'_i) \Leftrightarrow D \neq D'$ 表示原始数据与备份数据不相同, 一致性检测未通过。

对于网络备份系统来说, 因为网络传输, 在 T 时刻原始数据集发生的变化将在 $T'_i = T_i + \Delta T$ ($\Delta T > 0$) 时刻反映到备份数据中。针对网络备份系统的一致性检测表述为:

- a) 当 $\forall T_i = T'_j - \Delta T, \quad \forall D_i \in D, D_j' \in D'$ 时, $\forall i = j \Rightarrow f(D_i) = f(D'_j) \Leftrightarrow D = D'$ 表示原始数据和备份数据相同, 一致性检测通过。
- b) 当 $\forall T_i = T'_j - \Delta T, \quad \exists D_i \in D, D_j' \in D'$ 时, $\exists i = j \Rightarrow f(D_i) \neq f(D'_j) \Leftrightarrow D \neq D'$ 表示原始数据与备份数据不相同, 一致性检测未通过。

3.2 动态数据一致性检测难点及解决方法

由一致性检测算法描述可知, 在网络备份系统中, 面对动态数据, 需要重点解决的问题是: a) 怎样找到 ΔT 以确定原始数据发生变化后何时备份数据发生的相应改动; b) 怎样使得在一致性检测过程中原始数据和备份数据不发生变化。

问题 a) 使用海量缓存来解决。利用海量缓存模块, 可以容易控制发向备份中心的数据。结合备份中心在收到数据后给发送方的回复, 可以确认原始数据的改变已经反映在了备份数据集中。通过对接收方回复的确认, 间接创建了一个可控的备份数据集。原始数据集的每次数据改变通过接收方的回复确认被用户知道。借用这个方法, 虽然没有明确测出 ΔT 时间, 但可以完全控制原始数据的改变何时反映在备份数据中。

卷影拷贝解决了问题 b)。利用卷影拷贝, 可以在原始数据变化时创建某一时刻的快照。这个快照记录下了某一时刻卷的全部信息, 且快照时刻后原始数据的变化不会影响快照的数据。借助于快照, 获得了原始数据集某一时刻的副本; 再利用海量缓存将快照前的备份数据发送给备份中心, 获得了对应快照时间的备份数据集。使用这两个数据集作一致性检测, 其结果就反映了快照时刻两地数据是否一致。

3.3 动态数据一致性检测的实现

整个动态数据一致性检测由卷影拷贝服务、海量缓存模块和一致性检测模块协同完成, 其流程如下:

- a) 启动一致性检测模块。
- b) 要求海量缓存模块停止数据发送。
- c) 一致性检测模块确认数据已经停止发送后, 通知卷影拷贝服务创建快照。在快照生成时必须确认数据停止发送, 这样保证快照后产生的备份数据不会被海量缓存发送到备份中心。
- d) 海量缓存获取快照生成时间, 根据此时间发送海量缓存中快照前备份数据。
- e) 快照前备份数据发送完成后, 海量缓存停止发送备份数据, 然后通知一致性检测模块进行一致性检测。
- f) 一致性检测模块执行两地数据一致性检测。
- g) 一致性检测模块返回检测结果; 通知卷影拷贝服务销毁快照; 通知海量缓存模块继续发送数据。一致性检测完成。

动态数据一致性检测序列如图 3 所示。在过程 f) 中, 根据一致性检测算法描述, 结合实际应用, 作如下实现:

- a) 将原始数据集和备份数据集依照数据在卷上的偏移从小到大排列成有序数列。
- b) 将两个集合的数据按照在卷上的偏移一一对应从而构成双射。如果出现一个集合在另一个集合中找不到对应偏移的数据, 则在相应集合中以 0 与其对应。
- c) 将每个集合的有序数列以 512 Byte 划分, 构成与原始数据集和备份数据集的子集 $D_1, D_2, \dots, D_n$ 和 $D'_1, D'_2, \dots, D'_n$ 。
- d) 将原始数据集和备份数据集依照数据在卷上的偏移从小到大排列成有序数列。
- e) 将两个集合的数据按照在卷上的偏移一一对应从而构成双射。如果出现一个集合在另一个集合中找不到对应偏移的数据, 则在相应集合中以 0 与其对应。
- f) 将每个集合的有序数列以 512 Byte 划分, 构成与原始数据集和备份数据集的子集 $D_1, D_2, \dots, D_n$ 和 $D'_1, D'_2, \dots, D'_n$ 。
- g) 函数 $f(x)$ 取为计算每个集合的 MD5 摘要值。
- h) 比较对应摘要值得到一致性检测结果。
- i) 为了减少比较次数, 对算法作了一点优化。不按照算法直接比较每个子集的摘要值, 而是将计算所得的摘要值构成两个摘要值集合; 再将这两个集合作一致性检测, 从而提高性能。

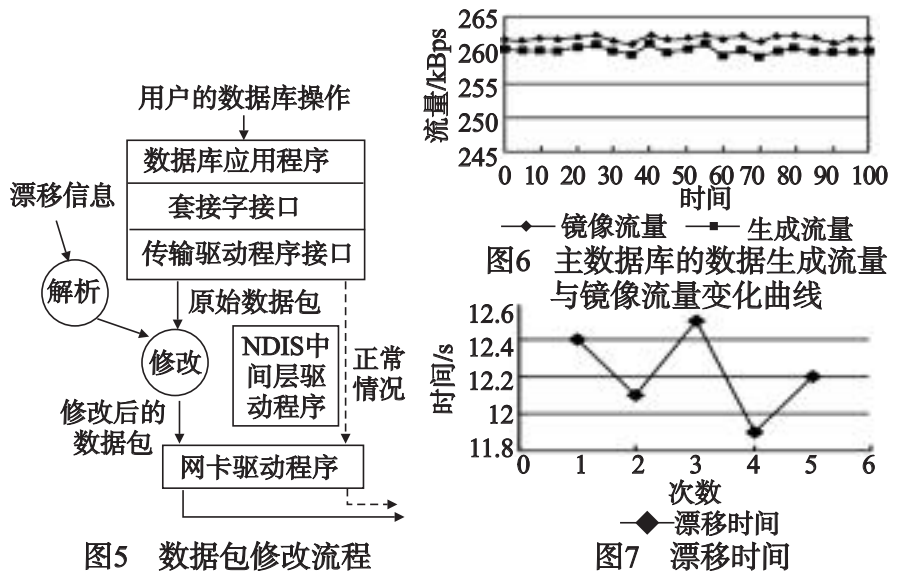
4 实验结果与结论

笔者搭建了一个网络备份环境来测试 (下转第 3709 页)

行插入、修改、删除等操作, 获得了主数据库的数据生成流量与镜像流量变化曲线, 如图 6 所示。镜像流量略大于生成流量, 表明主数据库的数据变化量被及时地镜像到了两个备用数据库上。若此时主数据库发生灾难, 其数据并不会丢失。该测试结果表明, 在 3 Mbps 慢速非稳定线路下, 系统仍能很好地完成数据库数据在异地的实时镜像, 可以实现数据零丢失。

3.2 服务漂移性能分析

关闭主数据库服务或者切断主数据库服务器的网络, 制造服务漂移发生的条件, 计算服务从主数据库漂移到备用数据库花费的时间, 即漂移时间。测试结果如图 7 所示, 漂移时间小于 13 s, 表明系统能以较短的时间完成数据库服务的漂移, 为数据库服务连续、安全、可靠地运行提供了强有力的保障。



4 结束语

本文设计并实现了一种基于慢速非稳定线路的数据库服务多点容灾系统, 通过数据监控、远程重放等方法实现了主数

据库数据到多个异地备用数据库的实时镜像, 保证了主备数据库的数据一致性, 同时对主数据库进行失效检测, 能够及时发现故障, 并以较短的时间(秒级)完成数据库服务的漂移, 有效保证了数据库服务的连续、安全、可靠运行, 为数据库服务容灾提供了一种廉价、高可用性、适应性强的解决方案。

参考文献:

[1] 沈昌祥, 张焕国, 冯登国, 等. 信息安全综述[J]. 中国科学 E 辑: 信息科学, 2007, 37(2): 129-150.

[2] 李涛, 刘晓洁, 曾金全, 等. 信息系统容灾抗毁原理与应用[M]. 北京: 人民邮电出版社, 2007.

[3] KEETON K, SANTOS C, BEYER D, et al. Designing for disasters [C] // Proc of the 3rd USENIX Conference on File and Storage Technologies. 2004: 59-72.

[4] Symantec (VERITAS). VERITAS global cluster manager: technical overview[R]. 2002.

[5] EMC. EMC SRDF family: powerful remote replication solutions improve organizational productivity, enhance online business continuity [R]. 2008.

[6] IBM. High availability clusters multi-processing XD (extended distance) for HAGEO technology: planning and administration guide [R]. 2003.

[7] WANG Yan-long, LI Zhan-hua, LIN Wei. RWAR: a resilient window-consistent asynchronous replication protocol[C] // Proc of the 2nd International Conference on Availability, Reliability and Security. 2007: 499-505.

[8] 杨朝红, 宫云战, 桑伟前, 等. 基于主从异步复制技术的容灾实时系统研究与实现[J]. 计算机研究与发展, 2003, 40(7): 1104-1109.

[9] 杨智军, 田地, 周斌. 基于 NDIS 中间层驱动程序的网络监测器 [J]. 吉林大学学报: 工学版, 2006, 36(2): 224-226.

(上接第 3701 页) 动态数据一致性检测的正确性并测试其检测耗时。其中本地服务器硬件配置为 Intel Pentium D 2.8 GHz CPU, 512 MB 内存, 120 GB 硬盘, 100 Mbps 网卡。远程服务器硬件配置为 Intel Celeron 2.4 GHz CPU, 内存 448 MB, 160 GB 硬盘, 100 Mbps 网卡。实验方法如下: 持续不断地向 2 GB 大小监控卷写入数据, 在写入数据的过程中执行一致性检测。测试结果如表 1 所示。实验结果如图 4 所示。

表 1 测试实验结果

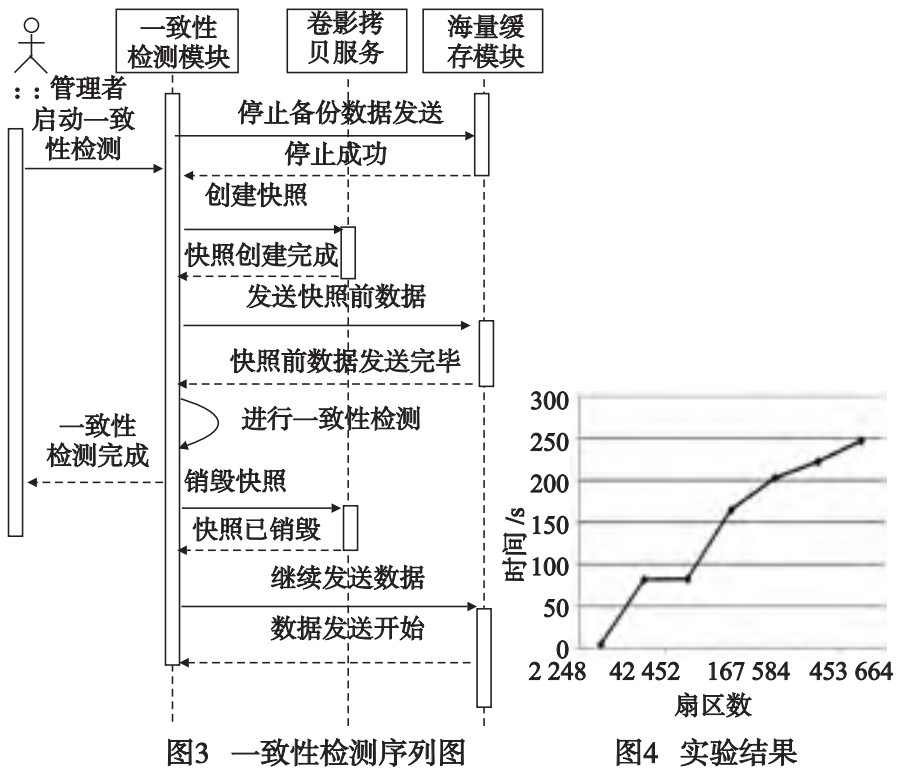
原始扇区数	快照扇区数	实测扇区数	检测耗时 / s	是否一致
2 248	2 248	2248	4.203	是
40 102	40 102	40 102	81.434	是
43 452	43 452	43 452	82.053	是
81 244	81 244	81 244	164.344	是
167 584	16 7584	167 584	202.328	是
183 032	183 032	183 032	221.975	是
453 664	453 664	453 664	247.047	是

实验结果表明: 动态数据一致性检测方法能在原始数据动态改变的同时执行一致性检测, 且检测结果能正确反映原始数据和备份数据的一致性。

5 结束语

本文设计并实现了一种网络备份系统的一致性检测方法。利用快照实时、快照数据不会改变的特性, 本方法采用一个实时的快照来代替持续变化的原始数据执行与备份数据的一致性检测。该方法解决了目前网络备份系统中作一致性检测必须要停止系统运行才能作数据一致性检测的缺点, 为目前网络

备份系统提供了一个安全、快捷、有效的数据一致性检测手段。



参考文献:

[1] 李涛. 网络安全概论[M]. 北京: 电子工业出版社, 2004.

[2] SOLOMON D, RUSSINOVICH M. Microsoft Windows internals[M]. 4th ed. Washington: Microsoft Press, 2005.

[3] ONEY W. Programming the Microsoft Windows driver model[M]. 2nd ed. Washington: Microsoft Press, 2003.

[4] 王志明. 容灾抗毁系统中数据一致性检测及快速恢复方法[D]. 成都: 四川大学, 2006.

[5] 王玮, 刘晓洁, 李涛, 等. 一种异地灾难恢复系统的设计与实现 [J]. 计算机应用研究, 2007, 24(9): 106-108.

[6] 周婧, 王意洁, 阮炜, 等. 面向海量数据的数据一致性研究[J]. 计算机科学, 2006, 33(4): 137-140.