

基于数据容灾系统的服务漂移实现

向 坤, 刘晓洁, 赵 奎, 李 峰

(四川大学 计算机学院 网络安全实验室, 四川 成都 610065)

摘 要:传统的数据容灾技术只能提供数据的实时备份与恢复,无法保证服务的连续性。服务容灾技术可以在服务级别上保证业务连续性和数据一致性,但系统设计往往过于复杂。一种基于传统数据容灾系统的服务漂移方案可以有效地解决该问题。该方案在客户服务器出现故障后,使用另一台拥有相同服务配置的镜像服务器,通过 iSCSI 协议和存储服务器上的逻辑卷镜像建立链接,并基于此虚拟逻辑卷继续提供服务。待客户服务器故障修复后,再切换回本服务器提供服务。这种方案可以使传统的数据容灾系统实现服务级容灾,进一步提高企业关键业务的容灾能力。

关键词:数据容灾;服务容灾;服务漂移;服务连续性;数据一致性;iSCSI

中图分类号:TP309.2

文献标识码:A

文章编号:1673-629X(2010)07-0152-04

Implementation of Service Migration Based on Data - Level Disaster Recovery System

XIANG Kun, LIU Xiao-jie, ZHAO Kui, LI Feng

(Nisec Lab, College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: Traditional data - level disaster recovery technology can provide data - level real - time backup and restoration, while still have limitations in providing service continuity. The service - level disaster recovery technology can provide service continuity and data consistency in service - level, but often requires a complicated system design. An implementation of service migration based on the traditional remote disaster recovery system can solve all these problems. When the client server suffered from disasters and stopped providing services, another image server with the same software configuration use the iSCSI protocol to initiate a connection to the storage server and established an iSCSI based remote logic volume. With the virtual logic volume, the image server can keep providing services and ensure the data consistency. When the client server restored from the malfunction, the system can switch from the migrating state to the original state, and the client server can keep providing services. With this implementation the traditional remote disaster recovery system can provide service - level data consistency and service continuity, which will further improve the companies' disaster recovery capabilities of key business.

Key words: data - level disaster recovery; service - level disaster recovery; service migration; service continuity; data consistency; iSCSI

0 引 言

随着计算机系统和网络的不断普及,数据作为开展业务不可缺少的基础,其安全性成为了各公司、政府机构等需要考虑的问题^[1]。传统的数据容灾系统往往提供基于逻辑卷的备份能力,当客户服务器数据被破坏或者丢失后,可以通过远程存储服务器提供数据的

在线恢复,但是这种方式仍然无法避免服务中断给企业造成的损失^[2,3]。服务级容灾技术虽然可以提供服务级上的数据一致性和服务连续性^[4],但是系统实现往往过于复杂,从传统的数据容灾系统迁移到服务容灾系统的成本也比较大。

文中提出了一种基于远程数据容灾系统的服务漂移方案。当客户服务器出现故障或者数据被破坏时,另一个镜像服务器可以通过 iSCSI 协议发起对存储服务器上对应逻辑卷的连接请求,当虚拟逻辑建立后利用存储服务器上的数据继续提供服务。待客户服务器的故障修复或数据恢复后,再将服务切换回客户服务器。该方案拓展了传统数据容灾系统的功能,使其可以实现服务级别的容灾,也为企业从数据容灾系统迁移到服务容灾系统提供了一种可行的方案^[5]。

收稿日期:2009-11-24;修回日期:2010-02-21

基金项目:国家高技术产业示范工程(I2006J1780);信息产业部电子发展基金(I2007J292)

作者简介:向 坤(1985-),男,四川成都人,硕士,研究方向为计算机网络安全和数据存储;刘晓洁,副教授,硕士生导师,研究方向为网络安全;赵 奎,副教授,硕士生导师,研究方向为网络安全技术及应用。

1 服务漂移的原理与实现

1.1 iSCSI 介绍

iSCSI 是 Internet 小型计算机系统接口 (Internet Small Computer System Interface) 的简称。如图 1 所示, iSCSI 是一种基于 TCP/IP 的协议, 用来建立和管理 IP 存储设备、主机和客户机等之间的相互连接, 并创建存储区域网络 SAN。iSCSI 的主要功能是在 TCP/IP 网络上的主机系统 (启动器 Initiator) 和存储设备 (目标器 Target) 之间进行大量数据的封装和可靠传输过程^[6]。此外, iSCSI 提供了在 IP 网络封装 SCSI 命令, 且运行在 TCP 上。



图 1 iSCSI 结构图

iSCSI 的工作过程是: 当 iSCSI 主机应用程序发出数据读写请求后, 操作系统会生成一个相应的 SCSI 命令, 该 SCSI 命令在 iSCSI Initiator 层被封装成 iSCSI 消息包并通过 TCP/IP 传送到设备端, 设备端的 iSCSI target 层会解开 iSCSI 消息包, 得到 SCSI 命令的内容, 然后传送给 SCSI 设备执行; 设备执行 SCSI 命令后的响应, 在经过设备端 iSCSI Target 层时被封装成 iSCSI 响应包, 通过 TCP/IP 网络传送给主机的 iSCSI Initiator 层, iSCSI Initiator 会从 iSCSI 响应包里解析出 SCSI 响应并传送给操作系统, 操作系统再响应给应用程序^[7,8]。

在服务漂移系统中, 镜像服务器用于在客户服务器故障时代替其提供服务。镜像服务器上所有对于数据的访问请求都通过 iSCSI Initiator 封装成 TCP 数据包, 然后通过 IP 网路发送给存储服务器。存储服务器上 iSCSI Target 解封 TCP 数据包, 执行数据访问请求, 然后把结果封装成 TCP 数据包发送回去^[9,10]。通过这种方式在镜像服务器和存储服务器建立一种虚拟的逻辑卷链接, 所有数据的实际操作在存储服务器上执行, 而操作的发起却在镜像服务器上, 这样保证了存储服务器上的数据可以一直保持最新。

1.2 系统结构

服务漂移要求当客户服务器由于各种原因中断运行时, 可以由镜像服务器代替其提供同样的服务, 并且能够保证数据和服务访问地址的一致性。服务漂移的

实现应该包括为 3 部分: 请求的重定向、数据的同步以及服务的启用。

图 2 显示了服务漂移的系统架构。

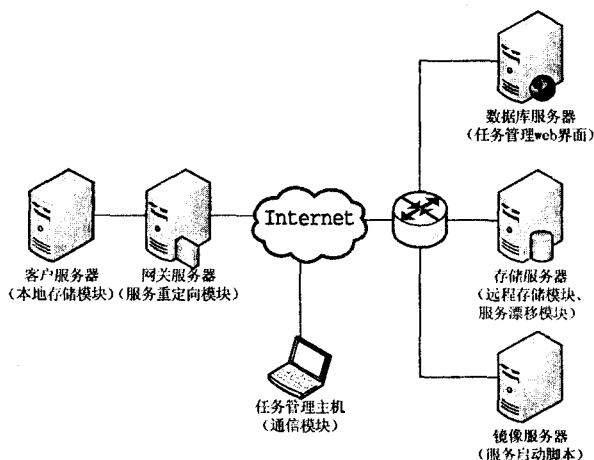


图 2 包含服务漂移功能的容灾系统架构图

1.2.1 客户服务器

客户服务是数据生产中心, 客户服务器上产生的数据通过 Internet 实时地备份到远程存储服务器中。

1.2.2 网关服务器

网关服务器处于客户服务器和外部网络之间, 将来自外部网络的服务请求定向到客户服务器, 同时保护客户不受外部攻击。当客户服务器服务中断时, 网关负责将服务请求定向到镜像服务器。

1.2.3 任务管理主机

任务管理主机可以是任何一台接入 Internet 的主机, 它通过 Web 界面和 ActiveX 控件分别与存储服务器、客户服务器进行通信, 实现对于备份任务的远程管理。

1.2.4 数据库服务器

数据服务器用于提供数据库服务和 Web 访问服务。关于任务的配置信息保存在数据库中, 其他模块在初始化时可以通过访问数据库获得任务的基本信息和状态。同时数据库服务也提供 Web 访问服务, 方便用户远程管理任务。

1.2.5 存储服务器

存储服务器用于保存客户服务器上的逻辑卷镜像。它不断接受来自客户服务器上的实时备份数据, 并写入到对应逻辑卷镜像。同时对于指定任务还可以执行数据同步、数据恢复等工作。

1.2.6 镜像服务器

镜像服务器用于在客户服务器出现故障无法提供服务时, 代替客户服务器提供服务。通过 iSCSI 协议将对逻辑卷的访问请求转发到存储服务器上对应的逻辑卷镜像。

1.3 模块实现

如图 3 所示,包含服务漂移功能的容灾系统包括三个子系统:数据存储子系统、配置管理子系统、服务漂移子系统。



图 3 容灾系统模块结构

1.3.1 数据存储子系统

数据存储子系统包括客户端和服务端端的存储模块,提供客户服务器和存储服务器之间的实时备份、数据同步、数据恢复、离线数据导出等功能。

1.3.2 配置管理子系统

配置管理主要为用户提供给予 Web 的远程管理界面,通过 ActiveX 和客户端消息处理模块实现控件和客户服务器的通信。

1.3.3 服务漂移子系统

服务漂移子系统主要负责当客户服务器中断服务时,启动镜像服务器上对应的服务,并将对于指向原客户服务的请求重新定向到镜像服务器。

1.4 服务漂移过程

容灾系统在正常工作情况下,存储服务器和客户服务器会保持定时的通信,以确定对方处于正常的工作状态。当客户服务器出现故障或灾难时,存储服务器会在检测到客户服务器长时间没有返回消息后判断客户服务器出现了故障,并且在界面上通知用户。

用户通过配置管理子系统启动服务漂移过程。如图 4 所示,服务漂移过程的启动主要经过三个阶段:客户服务器和存储服务器之间 iSCSI 连接的建立、镜像服务器上服务脚本的启动、网关服务器的请求重定向。

1.4.1 iSCSI 连接的建立

存储服务器在收到配置管理系统的消息后,首先会停止对于该客户服务器对应逻辑卷的更新,并将对应任务调整为停止状态。同时存储服务器使用 iSCSI Target 工具建立一个 iSCSI 目标,并将该任务对应逻辑卷加入到该 iSCSI 目标中。在存储服务器完成这些操作后,由镜像服务器上的 iSCSI Initiator 工具发起对存储服务器上 iSCSI 目标的连接。

1.4.2 服务脚本的启动

为了保证镜像服务器上能够提供和客户服务器上

相同的服务,并且保持数据的一致性,需要执行一系列的服务启动脚本。服务启动脚本的作用主要作用如下:发起 iSCSI 连接请求,建立和存储服务器的逻辑卷连接;然后调用系统 API 将连接的逻辑卷挂载到指定的盘符;启动与客户服务器上相同的服务进程(如 Web 服务器、FTP 服务器等等)。

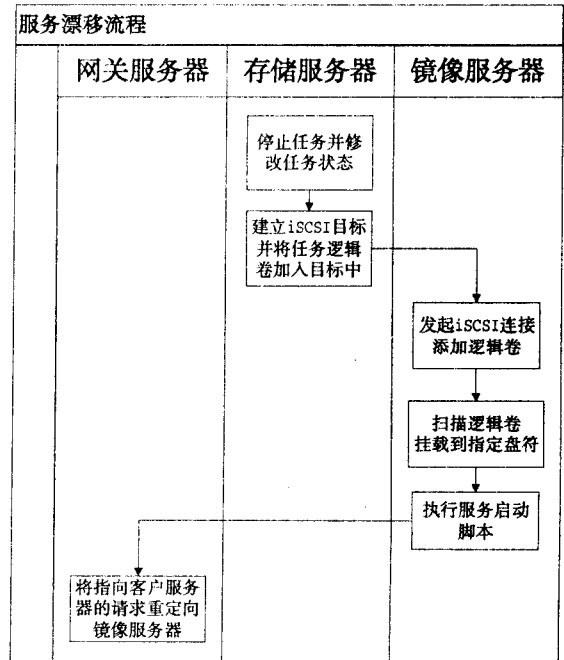


图 4 服务漂移的流程图

1.4.3 服务重定向

当完成以上两个过程后,镜像服务器已经能够替代客户服务器提供连续性的服务了。最后的服务重定向过程配置网关服务器将目前定位到客户服务器的请求重新定向到镜像服务器。位于网关服务器上的重定向模块通过运行 iptables 工具将所有指向原客户服务器的请求转发到镜像服务器。

2 性能测试

对于服务漂移系统的评估从两个方面考虑:一是服务漂移切换的时间,时间越短对于服务中断的影响越小,系统的性能也越好;二是服务漂移之后服务器的 IO 性能,因为漂移是通过 iSCSI 实现远程数据读写的,IO 性能可能不可能达到原来的磁盘直接读取速度,漂移后的 IO 性能越高说明系统的性能越好^[11,12]。

测试环境的存储服务器为 Intel Xeon 2.8Hz 双核 CPU,512MB 内存,希捷 ST3500630AV 7200 转硬盘,Intel 82541GI 网卡。其他服务均为普通 PC 机。试验环境下主要测试服务漂移速度和漂移后磁盘 IO 性能,数据如下:

在服务漂移的性能测试中,分别针对了 200m 到

3200m 数据进行了读写测试。结果如图 5 所示,漂移之前原始逻辑卷的读写平均速度为 28~29MB/s,漂移后通过 iSCSI 读写 IO 的平均速度为 10MB/s。造成这样的主要原因在于本系统使用的是百兆网卡,理论上的最高下行速度为 12.5MB/s,所以 iSCSI 传输速度实际上已经达到了网卡传输速度的上限。虽然目前漂移后 iSCSI 的性能比起原始磁盘的读取速度还有一定差距,但是在使用了千兆网卡后性能还能够获得进一步的提升。

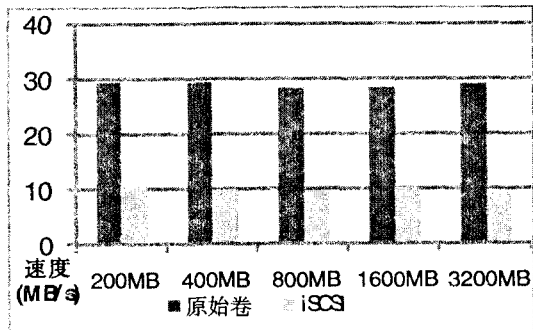


图 5 服务漂移的性能测试

在服务漂移切换时间的测试中,启动脚本主要执行两个操作,一是启动 Windows Server2003 上的 IIS 服务,二是启动 Microsoft SqlServer 服务。结果如图 6 所示,服务漂移的平均切换时间为 17.2s,这样的切换时间能够比较快速地完成服务漂移的切换,不会对服务中断造成太大的影响。

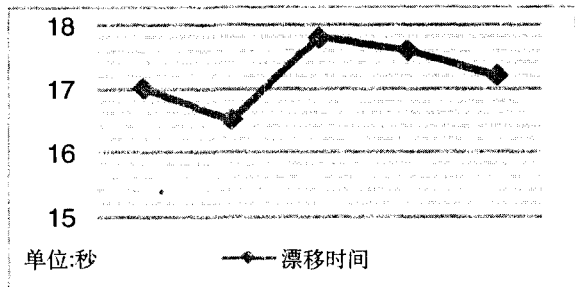


图 6 服务漂移的切换时间测试

3 结束语

提出并实现一种基于现有数据容灾系统的服务漂移方案,该服务漂移功能可以在客户服务器遭到灾难或故障时有效地实现服务连续性和数据一致性,并且具有较快的服务切换时间和,在 IO 性能上也具有较大的提升潜力。该方案是对现有数据容灾系统功能一个有力的扩充,适合作为中小型系统的服务容灾方案。

参考文献:

- [1] 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.
- [2] 李瑞浴,刘晓洁,李 涛,等.一种基于 ICE 通信的数据库容灾系统[J].信息安全与通信保密,2007(3):121-123.
- [3] 卢正添,李 涛,赵 奎,等.一种数据库服务多点容灾系统[J].计算机应用研究,2008,25(12):3707-3709.
- [4] 易固武,刘晓洁,李 涛,等.一种网络备份系统的数据一致性检测方法[J].计算机应用研究,2008,25(12):3700-3701.
- [5] 张 丁,刘晓洁,李 涛,等.数据库服务容灾系统的设计与实现[J].计算机工程与设计,2008,29(12):3247-3251.
- [6] Stevens W R. TCP/IP Illustrated Volume1: The Protocols [M]. [s.l.]:Addison Wesley, 1994.
- [7] 董欢庆,李战杯,林 伟,等. Linux 平台远程逻辑卷复制系统的设计[J].计算机工程与应用,2004(18):109-112.
- [8] 王彦龙,李战杯,董欢庆,等. Linux 平台远程逻辑卷复制系统的灾难恢复策略[J].微电子学与计算机,2005,22(3):68-72.
- [9] Symantec(VERITAS). VERITAS global cluster manager: technical over view[R]. [s.l.]:[s.n.],2002.
- [10] EMC. EMC SRDF family: powerful remote replication solutions improve organizational productivity, enhance online business continuity[R]. [s.l.]:[s.n.],2008.
- [11] Bryant R E, O'Hallaron D. Computer Systems: A Programmer's Perspective[M]. [s.l.]:Prentice Hall, 2003.
- [12] Stevens W R, Rago S A. Advanced Programming in the Unix Environment[M]. [s.l.]:Pearson Education, 2000.

(上接第 151 页)

- [3] Girgis M R, Mahmoud T M, Abd - El - Hafeez T. An Approach to Image Extraction and Accurate Skin Detection from Web Pages[J]. International Journal of Computer Science and Engineering, 2007, 12(2):88-96.
- [4] 许 强,江 早,赵 宏.基于图像内容过滤的智能防火墙系统研究与实现[J].计算机研究与发展,2000,37(4):458-464.
- [5] 甘利杰,丁明勇,杨永斌.基于 Winsock SPI 技术的包过滤研究[J].计算机科学,2007,34(8):122-124.
- [6] 吴 星,陈明锐.恶意网页防护系统的设计与实现[J].计

算机技术与发展,2009,19(5):154-157.

- [7] 李石君,李 洲,余 军,等.基于 URL 过滤与内容过滤的网络净化模型[J].计算机技术与发展,2009,19(1):5-8.
- [8] 周瀚宇.基于 URL 的网页内容过滤器的设计与实现[J].计算机工程,2006,32(7):81-83.
- [9] 王映辉,冯德民.大规模软件架构设计[M].北京:科学出版社,2003.
- [10] Drimbarean A F. Image processing techniques to detect and filter objectionable images based on skin tone and shape recognition[M]. [s.l.]:IEEE,2001.