一种基于 Internet 的异地容灾系统的设计和实现*

孙 洁, 刘晓洁, 李 涛, 刘 锦, 李 响, 吴 波, 皮璐琳
(四川大学 计算机学院, 四川 成都 610065)

摘 要: 基于 Internet 设计并实现了一种异地容灾系统。该系统集实时备份和监控、服务自动切换和快速恢复为一体, 支持多种操作系统和数据库, 并提供基于 Web 的远程管理。其结构简单、合理, 运行稳定, 具有较高的应用和推广价值。

关键词: 容灾; 数据备份; 灾难恢复
中图分类号: TP309.3 **文献标志码**: A **文章编号**: 1001-3695(2007)07-0171-02

Design and Implementation of Disaster Tolerant System Based on Internet

SUN Jie, LIU Xiao-jie, LI Tao, LIU Jin, LI Xiang, WU Bo, PI Lu-lin
(School of Computer, Sichuan University, Chengdu Sichuan 610065, China)

Abstract: A remote disaster tolerant system based on Internet has been presented. It has the ability of real-time mirroring, service switching automatically and fast recovering. And the system supports many operating systems and databases, providing remote management based on Web. The construction of the system is simple and rational, system working is good, and it has higher application and dissemination value.

Key words: disaster tolerant; data backup; disaster recovery

随着计算机技术的不断发展和信息化程度的不断提高, 信息已经成为企事业最具有价值的资产, 信息数据的丢失带来的灾难将是不可估量的损失。因此建立容灾系统, 保证数据完整性和业务连续、稳定在信息社会显得极为重要[1]。计算机系统的灾难备份和恢复建设受到高度重视并成为研究热点。与传统的容灾技术(如双机热备份、服务器集群技术等)相比, 异地远程容灾这种高性能的数据备份和灾难恢复技术, 更能充分保护系统中有价值的信息, 保证灾难发生时业务的连续性, 比其他容灾技术具有更多的优点[2,3]。但是容灾建设特别是高级别的异地容灾系统, 需要专线或光纤通信等特殊设备, 其投资巨大, 一些中小规模企业在资金不足的情形下, 无法顾及灾难备份, 所以灾难一旦发生, 后果将不堪设想。

基于上述因素, 针对中小型企事业单位提出的异地容灾系统, 利用 Internet 这一廉价资源, 实现了数据远程备份和快速恢复功能, 支持多种数据库和操作系统, 大大降低系统的开发和维护成本, 并确保用户业务不受影响或者将影响降低到最低。

1 系统结构

系统的拓扑结构如图 1 所示。

整个系统的拓扑结构以地域为界分为本地应用系统和远程备份系统两部分, 两者结构基本类似。本地应用系统由本地数据中心和本地网关组成, 其数据中心由一个或多个应用服务器组成, 它与本地网关距离近, 并由内部高速网络连接。同理, 远程数据中心也由多台备份服务器组成, 并与远程网关通过内部高速网络相连。

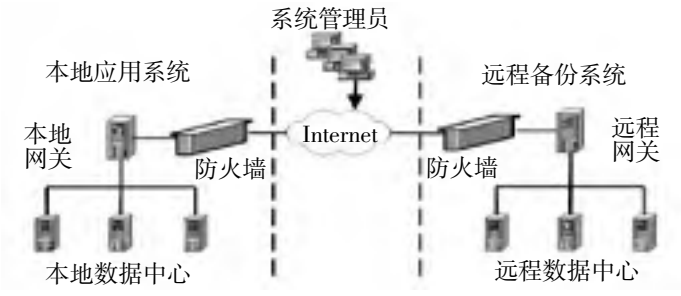

图 1 系统拓扑图

由于本地应用系统和远程备份系统结构上的相似性, 远程备份系统是与本地应用系统相当的备份应用系统, 系统的容灾层次为应用容灾[4], 即在灾难发生时, 高可用远程备份系统迅速接管本地应用系统的业务, 确保业务的连续性。

2 系统实现

2.1 容灾流程

容灾是项系统的工程, 必须理清容灾工作的流程, 制定详细的容灾计划。本系统整个流程共分五部分(图 2)。

系统管理员通过对本地应用系统可能遭受的灾难来源、受损程度、恢复要求等方面进行分析, 为备份和恢复打下基础。慎密地分析之后再根据生产系统的实际情况, 制定容灾计划(Disaster Tolerant Plan, DTP)。DTP 帮助系统管理员管理和控

收稿日期: 2006-05-02; 修返日期: 2006-07-28 基金项目: 国家自然科学基金资助项目(60373110, 60573130, 60502011); 教育部新世纪优秀人才计划资助项目(NCET-04-0870); 教育部博士点基金资助项目(20030610003); 四川省应用基础研究计划资助项目(05JY029-021-1)

作者简介: 孙洁(1981-), 女, 硕士研究生, 主要研究方向为网络安全技术及应用(sunjie334@163.com); 刘晓洁(1965-), 女, 副教授, 主要研究方向为网络安全技术及应用; 李涛(1965-), 男, 教授, 博导, 主要研究方向为网络安全及计算智能、信号处理; 刘锦(1982-), 男, 硕士研究生, 主要研究方向为网络安全; 李响(1982-), 男, 硕士研究生, 主要研究方向为网络安全; 吴波(1981-), 男, 硕士研究生, 主要研究方向为网络安全; 皮璐琳(1983-), 女, 硕士研究生, 主要研究方向为网络安全.

制容灾系统进行灾难恢复[5]。容灾计划包含容灾任务(Disaster Tolerant Task, DTT),并建立磁盘分区的对应关系进行数据镜像备份,同时实时监控各个 DTT 的运行。一旦灾难发生,按照预先拟定的容灾计划,进行数据恢复。恢复结束后,复查和评估该容灾计划,重新分析灾难事故,并及时更改容灾计划。

从系统的整个流程而言,容灾工作环环相扣,并相互促进、相互推动、灵活调整,具有极大的适应性。

2.2 容灾计划

容灾计划(DTP)由多个容灾任务(DTT)组成:

$$DTP = DTT_1 \cup DTT_2 \cup \cdots \cup DTT_n,\ DTT = \langle M, B, P\rangle,\ P = \langle S, R\rangle$$

其中,M 为本地生产中心的磁盘分区组,B 为远程备份中心的磁盘分区组,P 为该任务的容灾策略,包括备份镜像关系 S 和恢复策略 R,且有如下关系,$M \underset{R}{\overset{S}{\rightleftharpoons}} B$。其中本地数据中心的磁盘分区组 M = { LSP, LGDP, LGCP} ,LSP 是应用服务器提供服务和数据的磁盘分区,LGDP 是本地网关的磁盘分区,LGCP 是本地网关上缓冲写操作的磁盘分区。远程备份中心的磁盘分区组 N = { RGDP, RSP} ,RGDP 是远程网关的磁盘分区,RSP 是备份服务器的磁盘分区。

通过下述镜像关系,将本地应用服务器磁盘 LSP 上的写操作捕获并重放到远程备份服务器对应的磁盘分区 RSP 上,完成数据的备份。M、B 的对应关系如图 3 所示。

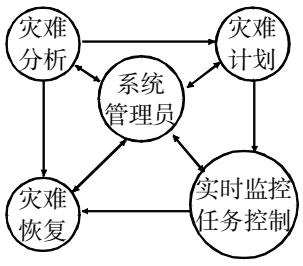

图 2 容灾流程

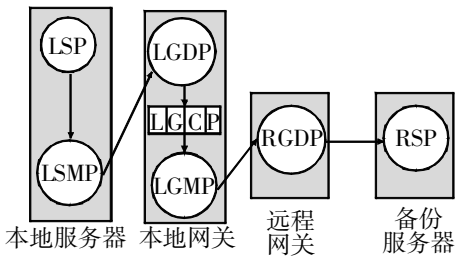

图 3 映像关系图

(1) 本地服务器到本地网关。首先通过磁盘镜像技术,本地网关设备 LGDP 映射到本地服务器上,记为 LSMP。对系统而言,LGDP 和 LSMP 视为同一逻辑单元,即当 LSMP 发生变化时,将写操作封装成 TCP/IP 数据包,通过内部高速网络传到 LGDP,取出该操作重新执行,故有 LGDP = LSMP。其次通过磁盘冗余技术,LSMP 和 LSP 同时发生数据更新,即 LSP = LSMP = LGDP,可推出 LSP = LGDP。

(2) 本地网关捕获写操作。考虑到本地网关与远程网关之间的外部网络性能不稳定性,所以系统监听到 LGDP 发生的写操作,截获写操作,放入缓冲分区 LGCP。

(3) 本地网关到远程网关。首先通过磁盘镜像技术,远程网关设备 RGDP 映射到本地网关的磁盘分区为 LGMP。LGMP 和 RGDP 被设为同一逻辑单元。写操作从缓冲分区 LGCP 中解封取出,重放到 LGMP。LGMP = RGDP,写操作也在 RGDP 上执行。

(4) 远程网关到备份服务器。因为 RGDP 和 RSP 互为镜像关系,利用内部的高速网络,对 RGDP 的写操作会同步执行到 RSP,从而 RGDP 与 RSP 的数据一致。

(5) 综上,应用服务器上的写操作通过 $LSP \Rightarrow LSMP \Rightarrow LGDP \Rightarrow LGCP \Rightarrow LGMP \Rightarrow RGDP \Rightarrow RSP$,完成数据的备份。

图 3 在上述映射关系的基础上,有多种恢复策略 P 供系统管理员选择:自动切换并恢复数据、手动恢复、定时恢复等。管理员可以根据实际情况,选择一种或多种恢复策略对任务进行快速恢复处理。

2.3 层次结构

系统的层次结构如图 4 所示,共分为三层,即用户层、中间层和核心层。通过分层,每一层功能明确、结构简单,便于系统的维护和扩展。

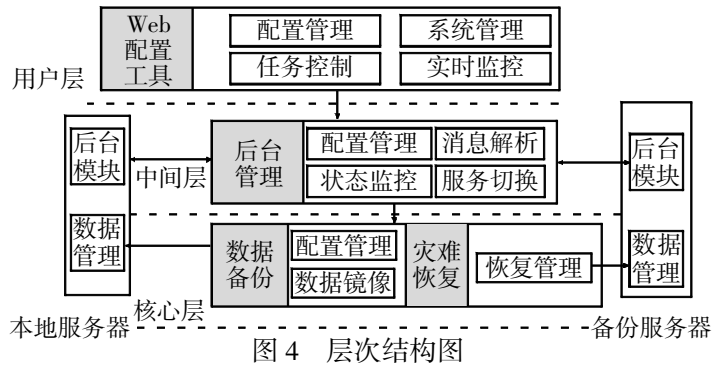

图 4 层次结构图

2.3.1 用户层

基于 Web 的远程管理,为用户提供友好的可视化图形界面。根据用户的要求完成备份任务组(DTT)的配置管理、灾难恢复、系统管理等操作。与下层之间以消息驱动的方式通信。

2.3.2 中间层

中间层即后台管理是用户层和核心层的中介,位于本地网关、本地服务器和备份服务器。本地服务器和备份服务器的后台管理主要功能是监控本机的运行状态并将结果反馈给本地网关。位于本地网关的后台管理含有四个模块,即消息解析、配置管理、状态监控和服务切换。

当系统加载并初始化之后,位于本地网关的后台管理通过特定的端口进行监听。当收到网页、本地服务器或者远程服务器发送的消息之后,通过消息解析模块,对消息进行分析和处理;如当收到配置任务组的消息时,后台将用户配置的任务组信息写入数据库,同时转发到本地服务器进行任务组配置。

状态监控有两项功能:①在用户设定的轮询时间内,查询各任务组的运行状态,并将查询结果反映到页面上,便于用户掌握系统的运行情况;②检测分散于各地的本地服务器、本地网关和远程网关的运行状态并将其组织起来,成为协同工作的一个整体。本地生产系统和远程备份系统可能相隔几千里,在检测时必须考虑网络迟延及其他因素,故本系统采用检查点技术,即主动检测技术。本系统的检测发起对象是本地网关,由它来负责感知整个系统的正常运行。当系统加载时,通过特定的端口,本地网关每隔一个轮询周期,就向检测对象本地服务器进行一次检测,如果在限定的时间内,检测对象返回了存活信息,即表明生产系统正常运行。同时,本地网关也每隔一个轮询周期,向远程备份系统发送自身存活信号。同理,在限定的时间内,远程备份系统收到了本地网关的存活信息后,即认为本地网关正常运行;若远程备份系统没有收到存活信号,即认为本地生产系统发生灾难。

当状态监控模块检测本地服务器发生了灾难,即自动进行服务切换到远程备份系统。整个过程对用户透明。

2.3.3 核心层

核心层由数据备份和灾难恢复两部分组成,是容灾系统的核心功能区。

(下转第 179 页)

的依赖关系。笔者在扩充的构件演化模型[17~19]的研究工作基础上,着眼于软件的维护阶段,针对的是多个演化的"遗留"框架实例,利用概念格技术从版本系统中挖掘蕴涵的框架变化点的实例化模式。

参考文献:

[1] 杨芙清.软件技术与软件产业[J].中国计算机学会通讯,2001(6).

[2] WILSON D A, WILSON S D. Writing frameworks-capturing your expertise about a problem domain, tutorial notes: proc. of the 8th Conference on Object-Oriented Programming Systems, Languages and Applications[C]. Washington: [s. n.], 1993.

[3] 杨芙清,梅宏,吴穹,等.基于异质构件复用的软件开发技术及其支持系统[J].中国科学,1997,27(3):275-281.

[4] HAKALA M, HAUTAMAKI J, KOSKIMIES K, *et al*. Generating application development environments for Java frameworks: proc. of the 3rd International Conference on Generative and Component-Based Software Engineering (GCSE'01) [C]. [S. l.]: [s. n.], 2001: 163-176.

[5] FONTOURA M, PREE W, RUMPE B. UML-F: a modeling language for object-oriented frameworks: proc. of the 14th European Conference on Object Oriented Programming (ECOOP 2000), Lecture Notes in Computer Science 1850[C]. Cannes, France: Springer, 2000: 63-82.

[6] Object Management Group Inc. OMG unified modeling language specification, version 1.3[R]. [S. l.]: [s. n.], 1999.

[7] Understand C++ [EB/OL]. http://www.scitools.com/.

[8] WILLS R. Restructuring lattice theory: an approach based on hierarchies of Concepts[M]. [S. l.]: Ordered Sets, 1982: 445-470.

[9] GODIN R, MISSAO U R, ALAOUI H. Incremental concept formation algorithms based on Galoisconcept lattices[J]. Computational Intelligence, 1995, 11(2): 246-267.

[10] ANDELFINGER U. Diskursive anforderung sanalyse. Ein Beitrag zum Reduktionsproblem bei Systementwicklungen in der Informatik[C]. Peter Lang, Frankfurt: [s. n.], 1997.

[11] LINDIG C, SNELTING G. Assessing modular structure of legacy code based on mathematical concept analysis: proc. of the International Conference on Software Engineering (ICSE'97) [C]. Boston: [s. n.], 1997: 349-359.

[12] TONELLA P. Concept analysis for module restructuring[J]. IEEE Trans on Software Engineering, 2001, 27(4): 351-363.

[13] SNELTING G. Software reengineering based on concept lattices: proc. of the 4th European Conference on Software Maintenance and Reengineering[C]. [S. l.]: IEEE, 2000: 3-12.

[14] RAY X. Views: understanding internals of classes gabriela arévalo, stéphane ducasse oscar nierstrasz: proc. of ASE 2003, IEEE Computer Society[C]. Montreal, Canada: [s. n.], 2003: 267-270.

[15] AR VALO G, DUCASSE S, NIERSTRA O. Discovering unanticipated dependency schemas in class hierarchies: proc. of CSMR 2005 of the 9th European Conference on Software Maintenance and Reengineering [C]. [S. l.]: IEEE Computer Society Press, 2005: 62-71.

[16] AR VALO G, BUCHLI F, NIERSTRASZ O. Detecting implicit collaboration patterns: proc. of WCRE 2004: the 11th Working Conference on Reverse Engineering[C]. [S. l.]: IEEE Computer Society Press, 2004: 122-131.

[17] 钟林辉,谢冰,邵维忠.扩充 CDL 支持基于构件的系统组装与演化[J].计算机研究与发展,2002,39(10):1361-1365.

[18] 钟林辉,谢冰,邵维忠.扩充 CDL 支持构件演化模型的方法研究[J].软件学报,2002,13.

[19] 钟林辉,陈宇,刘洋,等.软件配置管理系统 XML 数据模型及原型系统研究[J].计算机工程与应用,2001,37(19):82-84,120.

(上接第 172 页)数据备份以 Linux 内核模块方式运行于本地网关上,与网络存储技术相结合,通过图 3 所示的灾备任务磁盘分区之间的映射关系,将本地服务器的数据备份到远程备份服务器。且所有的实现均动作于本地网关,具有以下优点:①实现了本地服务器的逻辑隔离;②实现了本地服务器的平台无关性;③减小了对本地服务器的性能及其他方面的影响;④在实现远程镜像的同时也实现了数据在本地灾备网关上的完全镜像,增加了数据安全性。

灾难恢复采用差异恢复的方式,通过分析灾难源、故障点,恢复流程,选择一种或多种恢复策略。进行灾难恢复之前,任务必须满足图 3 所示的映射关系。在确定故障源之后,根据容灾计划中的恢复策略进行快速恢复。

3　与同类系统比较

在比较和评价各种异地容灾系统时,需要考虑的因素包括:数据的可用性、数据的可靠性、对应用程序处理效率的影响、系统的运行成本等。

基于以上因素,本系统与目前市场主流的容灾产品相比,优势在于:①利用 Internet 这一廉价的资源,不需要建设专网,对硬件没有硬性要求,降低了开发和维护成本;②通过数据镜像的差错控制机制,保证数据的一致性和可靠性;③不仅在异地实现了数据备份,在灾备网关上也同时实现了数据备份,提高了数据的冗余;④有多种恢复策略供用户选择,灵活多变;⑤对应用服务器上的应用程序透明,对系统性能影响较小。

4　结束语

本文设计并实现了一种基于 Internet 的异地容灾系统,实现了数据备份和灾难恢复,保证了灾难发生时服务自动切换及系统应用的不间断;支持多种平台,对硬件的要求低,具有广阔的发展前景。

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] KING R P, HALIM N, GARCIA-MOINA H, *et al*. Management of remote backup copy for disaster recovery[J]. ACM Trans on Database Systems, 1991, 16(2): 338-368.

[3] CHOY M H, LEONG H V, WONG M H. Disaster recovery techniques for Database System[J]. Communication of the ACM, 2000, 43(11): 272-280.

[4] 刘迎风,祁明. 容灾技术及应用[J].计算机应用研究,2002,19(6):7-10.

[5] WANG Kun, ZHOU Lihua, CAI Zhen, *et al*. A disaster recovery system model in an e-government system: proc of the 6th International Conference Parallel and Distributed Computing, Applications and Technologies[C]. [S. l.]: [s. n.], 2005: 247-250.