

基于 Internet 的跨地域容灾系统*

廖竣锴, 李涛, 刘晓洁, 王健, 浦海挺

(四川大学 计算机系, 四川 成都 610065)

摘要: 实现的基于 Internet 的跨地域容灾系统, 不仅具备远程实时数据备份功能, 同时解决了数据传输安全问题, 并且提供服务切换功能, 是一种比较完善的容灾解决方案。

关键词: 容灾; 数据备份; Internet

中图分类号: TP393.4 文献标识码: A 文章编号: 1001-3695(2004)09-0202-02

A Tolerated Disaster System Based on Internet

LIAO Jun-kai, LI Tao, LIU Xiao-jie, WANG Jian, PU Hai-ting

(Dept. of Computer, Sichuan University, Chengdu Sichuan 610065, China)

Abstract: The implementation of a tolerated disaster system based on Internet is proposed. The system has the function of remote real-time data backup, resolved the issue of data transfer security, and provide with the function of service switch. It is a preferable scheme of tolerated disaster.

Key words: Tolerated Disaster; Data Backup; Internet

1 引言

现代企业的运作日益依赖于信息技术。信息已经成为公司拥有的最有价值的资产, 这些数据的丢失和损坏将对企业造成难以估量的损失。传统的数据备份技术和集群技术足以避免软、硬件故障和计算机病毒入侵所造成的破坏, 但对大规模的灾难性突发事件则无能为力。此时若想迅速恢复系统数据, 保持业务正常运行, 就必须建立异地的容灾系统。但是建立一套完整的容灾系统的代价非常昂贵, 因为按照目前的通行做法, 企业必须搭建一条专线用于本地和远程的数据传输, 其开销相当惊人。一般的企事业单位根本无法承受。

本文设计并实现了一种廉价的容灾系统基于 Internet 的跨地域容灾系统。它使用 Internet 这一廉价资源, 为用户建立一套廉价、安全并且稳定的跨地域容灾系统。该系统具有如下特点:

- (1) 通过 Internet 实现跨地域数据备份;
- (2) 数据加密传输, 确保数据在 Internet 上的传输安全;
- (3) 实时数据备份, 使变化的数据在极短的时间得到备份;
- (4) 服务自动切换, 保证业务的连续。

2 系统工作原理

基于 Internet 的跨地域容灾系统的构成如图 1 所示。

整个容灾系统从物理上可分为本地数据中心和远程数据中心两大部分。本地数据中心包括为企业提供正常业务的服务器群和本地容灾网关。所有服务器通过网关与 Internet 相连, 对外提供服务, 由网关监测本地服务器状态并控制 Internet

用户对它的访问。远程数据中心由备份服务器群和远程容灾网关组成。其拓扑结构基本上与本地数据中心相同, 但备份服务器数目可以小于本地服务器的数目。

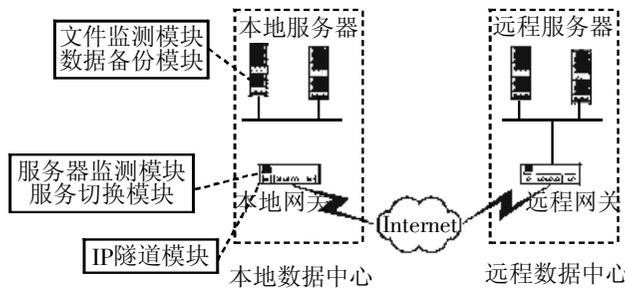


图 1 基于 Internet 的容灾系统体系结构

从功能上, 容灾系统包含三个子系统: 数据备份与恢复子系统、IP 隧道子系统和服务切换子系统。

(1) 数据备份与恢复子系统由两个模块组成: 文件监测模块和数据备份模块。其主要功能是实时备份数据到远程备份服务器, 并且可从远程备份服务器恢复数据。

(2) IP 隧道子系统为本地数据中心和远程数据中心在 Internet 之上的通信建立一条安全通信隧道, 其高强度的加密信道可确保传输数据万无一失。

(3) 服务切换子系统由服务器监测模块和服务切换模块构成。其目的是动态监测服务器状态, 并且能够自动切换服务到远程备份服务器。

系统模块体系结构如图 2 所示。

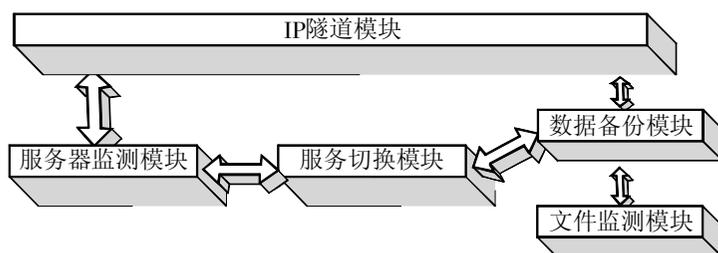


图 2 系统模块体系结构

正常情况下, 用户通过网关访问本地服务器, 文件监测模块监测服务器文件系统的状态, 一旦发现某文件状态发生变化, 立刻通知数据备份模块; 数据备份模块接到通知后, 采用差异备份的方式将该文件备份到远程数据中心的备份服务器上。备份数据在 Internet 的传输通过 IP 隧道进行。

当服务器出现故障时, 服务器监测模块检测到该故障后, 通知服务切换模块将用户请求转发到远程数据中心的备份服务器, 由备份服务器对外提供服务。当本地服务器从故障中恢复后, 服务器监测模块检测到服务器状态恢复正常后, 通知数据备份模块恢复丢失的数据。当数据恢复完成后, 通知服务切换模块将服务切换回本地服务器, 由本地服务器对外提供服务。

服务切换流程如图 3 所示。

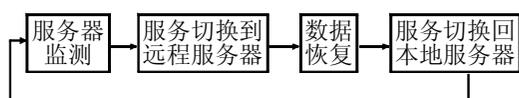


图 3 服务切换与恢复流程

3 关键技术

3.1 实时数据备份

实时数据备份要求当数据发生变化时, 立即将变化了的数据备份到远程数据中心。从而使得灾难发生时, 数据丢失量最小或不丢失数据。实时数据备份由两个模块实现: 文件监测模块和数据备份模块。文件监测模块负责监测文件的状态, 一旦发现文件发生变化则立刻通知数据备份模块, 数据备份模块接到消息后, 备份该文件。

3.1.1 文件监测模块

文件监测模块工作流程如图 4 所示。

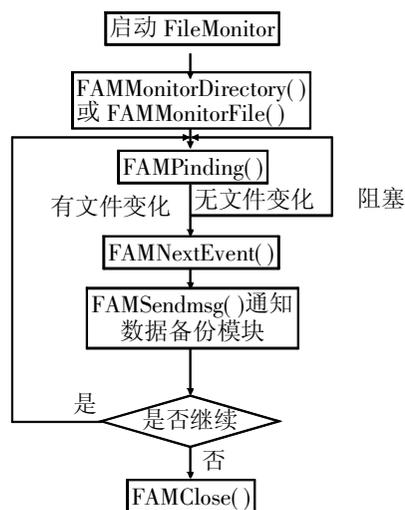


图 4 文件监测流程

文件监测模块在 Linux 平台上实现, 主要功能是监测文件系统的文件变化状况。在本系统中由数据备份模块调用文件监测模块, 以得到实时的文件变化信息。

系统管理员配置好需备份的目录或文件并启动数据备份, 调用文件监测模块并把需要监测的目录或文件作为参数传给它, 文件监测模块则开始监测该目录或文件。当文件状态发生变化时, 通知数据备份程序该文件发生了变化, 这种文件的变化包括文件创建、删除、修改、重命名、文件属性的改变, 其中文件属性的改变由包括文件大小、类型、用户、组、访问时间、修改时间、头节点、分配块数。

当应用程序暂时不需要监测该文件时, 可以挂起监测进程; 当需要再次监测该文件时, 恢复监测进程。当应用程序不

再需要监测该文件, 则可以关闭该监测进程。

3.1.2 数据备份模块

数据备份模块负责备份数据和恢复数据。为了提高数据备份的效率, 该模块采用差异备份方式备份数据。其最大优点在于传输数据量小, 差异备份比较两个文件的差异, 并且只备份两个文件中数据不同的部分, 而不是备份整个文件。因此大大提高了数据备份的效率, 并且减少了系统资源的开销。

差异备份的算法描述如下:

假设计算机 A 上有一文件, 计算机 B 上有一文件, 两个文件基本相似, 要将两个文件同步, 采用以下步骤:

(1) 将文件划分为一系列大小为 S 的数据块, 最后一块大小可以小于 S;

(2) 为每一数据块计算一个“弱 32 位滚动校验和”(该算法会在下面描述) 和一个“强 128 位 MD4 校验和”;

(3) 将校验和送到 A 上;

(4) 同样 A 搜索文件, 对每一个大小为 S 的数据块生成计算一个“弱 32 位滚动校验和”和一个“强 128 位 MD4 校验和”, 将它与 A 的校验和比较, 若匹配则说明数据块内的数据相同, 否则数据不同;

(5) 校验和检查完毕后, A 生成一个文件的拷贝并传给 B, 其中内容相同的数据块以一个引用表示, 内容不同的数据块包含数据内容。

结果, 文件与文件同步, 但只修改了与不同的部分。

为了提高检查文件的速度和效率, 差异备份算法中采用了滚动校验和算法, 它可以迅速地生成从文件任何位置开始的校验和, 并且采用滚动方式迅速完成校验和检查。所谓滚动方式是指每次检查完一个位置的校验和之后, 向前移动一个字节, 检查下一个位置的校验和, 直至文件尾。其算法如下:

(1) $s(k, l)$ 计算出从字节 $k \sim l$ 的校验和

$$a(k, l) = \left(\sum_{i=k}^l X_i \right) \bmod M$$

$$b(k, l) = \left(\sum_{i=k}^l (1-i) X_i \right) \bmod M$$

$$s(k, l) = a(k, l) + 2^{16} b(k, l)$$

(2) 利用(1)的结果计算从字节 $k+1 \sim l+1$ 的校验和

$$a(k+1, l+1) = (a(k, l) - X_k + X_{l+1}) \bmod M$$

$$b(k+1, l+1) = (b(k, l) - (1-k) X_k + a(k+1, l+1)) \bmod M$$

$$s(k+1, l+1) = a(k+1, l+1) + 2^{16} b(k+1, l+1)$$

3.2 IP 隧道加密传输

本地数据中心与远程数据中心的数据加密传输可采用基于 IPSec 的 IP 隧道技术来实现。

IP 隧道为数据传输提供了一条端到端的加密通道。将 IP 隧道模块部署在本地容灾网关和远程容灾网关上。当本地数据中心要与远程数据中心通信时, 由两个容灾网关相互协商建立一条安全通信隧道, 本地数据中心的数据首先在本地容灾网关上进行数据加密, 再通过 Internet 传送到远程容灾网关上, 在远程容灾网关上解密为明文并传送到备份服务器。所有数据在 Internet 上都以密文形式传输, 即使被黑客截获也无法得到明文信息。数据加密传输过程如图 5 所示。

接 OLE2.0 不同控件对象, 我们采用 PowerBuilder 的代码。



图 6 在质检程序中用 Word 编辑检验操作记录格式

(1) 将用户定制的格式文件读入到 OLE2.0 控件中

```
li_value = GetFileOpenName("Select File", ls_docname, ls_named, "DOC", "Text Files(*.TXT), *.TXT, Doc Files(*.DOC), *.DOC, all file + s(*.*)", *.*)
IF li_value = 1 THEN ole_1.insertfile(ls_docname)
```

(2) 将 OLE2.0 控件中的操作记录格式保存到数据库

```
lblob_data = ole_1.objectdata
UPDATEBLOB QM_SYNTAX1 SET record = :lblob_data
WHERE jpid = :is_jpid and jyzb = :is_jyzb
//筛选条件是指定检品的检验原子指标
USING sqlca;
```

(3) 控件对象的注册

```
string ls_controlname //控件对象文件名
string ls_inifile //ini 文件名(存放注册信息) ...
if ProfileString(ls_inifile, "REGISTER", ls_controlname, ) = then
//判断注册
run(regsvr32.exe ls_controlname)
setProfileString(ls_inifile, "REGISTER", ls_controlname, 1) ...
//写入注册文件
end if
```

3.2.4 优劣评价

使用 OLE2.0 控件可以调用外部程序处理不同格式文档, 增强了程序的通用性, 方便了用户, 同时减少了定制格式方面的编程工作; 在数据库表设计方面也很简单: 用户输入的所有数据只需要表中的一个 Blob 类型的字段来保存即可。

不足的是, OLE2.0 控件连接外部程序到打开文件需要等

待一些时间, 其执行的效率没有用动态数据窗口技术高。

4 结束语

本文分析了制药行业质检过程 GMP/GSP 管理的复杂性, 提出检验计划、检验操作记录用户统一定制的方案, 采用了 PowerBuilder 的动态窗口和 OLE2.0 技术, 很好地解决了问题, 为 GMP, GSP 的电子化管理探索出了一条新路。目前已在某一制药企业成功应用。

本文阐述了两种技术, 分析了在使用过程中的优劣, 实际上只需一种就能解决问题, 写在这里给大家提供参考, 笔者建议将两种配合使用效果更好。

对于本文讨论的格式定制的内容并不仅仅适用于制药行业质检部门的检验操作记录的 GMP, GSP 管理, 对于其他行业(如食品行业), 以及涉及到操作记录/报表比较多的其他部门(如销售或者财务部门)也有一定的参考价值。

参考文献:

[1] 陈鹏, 张洪伟. 企业资源计划系统(ERP)探讨[J]. 计算机应用研究, 2001, 18(增刊).

[2] 廖春华, 张洪伟. 制药行业 ERP 质量管理子系统的设计与实现[J]. 计算机应用, 2002, 22(5): 77 - 79.

[3] 张亚丰. 医药生产企业 GMP 与 GMP 认证实务全书[M]. 吉林: 吉林摄影出版社, 2002.

[4] 郑筱萸. 药品经营质量管理规范[EB/OL]. <http://www.chinamedicom.com/public/news/gspfaqui-2.jsp>, 2000-04-30.

[5] [美] Webby Boggs, Michael Boggs. UML 与 Rational Rose2002 从入门到精通[M]. 邱仲潘, 等. 北京: 电子工业出版社, 2002.

[6] [美] Simon J A Herbert. PowerBuilder 7.0 实用全书[M]. 张宝林, 韩平, 汪洋, 等. 北京: 电子工业出版社, 2000.

作者简介:

张红实(1977-), 男, 重庆人, 硕士研究生, 主要研究方向为数据库与信息系统及管理; 张洪伟(1955-), 男, 四川人, 教授, 博士后, 西德博士, 主要研究方向为数据库与计算机网络。

(上接第 203 页)

3.3 服务切换

服务切换包括状态监测技术和服务切换技术。能够在无人值守的情况下自动监测服务器的故障, 并且将服务切换到远程数据中心的备份服务器。另一方面, 当本地服务器从故障中恢复后, 再将服务切换回本地服务器。对用户而言, 服务始终保持正常, 没有任何影响。

3.3.1 服务器状态监测模块

服务器监测程序为一守护进程, 定期轮询服务器的状态, 有应答消息则说明服务器处于活动状态; 反之, 则认为服务器出现故障, 应该进行服务切换。其流程如图 6 所示。

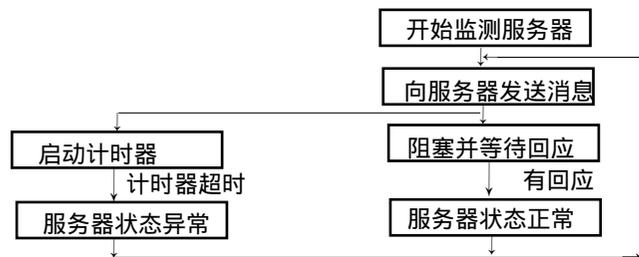


图 6 服务器状态监测流程图

3.3.2 服务切换模块

服务切换模块采用 Linux 的 IPTables 工具实现。IPTables 是 Linux 平台下的一个包过滤防火墙工具, 功能强大, 可实现状态包过滤、NAT(网络地址转换)、流量控制、负载均衡等功能。服务切换使用 IPTables 的 DNAT(Destination NAT), 当服

务器监测模块检测到本地服务器出现故障时, 服务切换模块动态地在 IPTables 防火墙 NAT 规则中添加一条 DNAT 规则, 将访问本地数据中心的请求转发到远程数据中心, 由远程数据中心的服务器提供服务; 当本地服务器从故障中恢复并且数据恢复完成后, 服务切换模块删除这条 DNAT 规则, 则访问本地数据中心的请求仍然发送到本地服务器。

4 结束语

本文提出了一种基于 Internet 的跨地域容灾系统方案。同时, 实现了在 Internet 上的远程数据备份与恢复, 远程数据传输加密以及服务切换等关键技术。整个系统的备份数据通过 Internet 加密传输, 而无须另外搭建专线, 是一种廉价、安全的容灾系统。

参考文献:

[1] tif Ghaffar. Real-time Data Mirroring under Linux[EB/OL]. <http://www.linuxfocus.org/English/March2001/article199>.

[2] usty Russell. Linux 2.4 NAT HOWTO[EB/OL]. <http://www.netfilter.org/documentation/HOWTO/NAT-HOWTO>.

[3] arlton R. Davis. IPsec VPN 的安全实施[M]. 北京: 清华大学出版社, 2002.

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

作者简介:

廖竣错(1976-), 男, 硕士研究生, 主研方向为网络信息安全; 李涛(1965-), 教授, 博士生导师, 研究方向为网络信息安全和人工智能; 刘晓洁, 副教授; 王健、浦海挺, 硕士研究生, 研究方向为网络信息安全。