◎ 汪中夏　刘　伟　编著
◎ 戴士剑　审校

数据恢复高级技术

信息产业部
电子行业技术培训专用教材

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

信息产业部电子行业技术培训专用教材

数据恢复高级技术

汪中夏　刘伟　编著
戴士剑　审校

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

数据恢复是指从被损坏的数据载体，如磁盘、磁带、光盘和半导体存储器等，以及被损坏或被删除的文件中获得有用的数据，对它的需求几乎是伴随着计算机的产生而产生的。本书是信息产业部电子行业技术培训专用教材，结合具体实例讲解了数据恢复基础、硬盘物理故障的判定及修复、PC-3000 在数据恢复中的应用、磁盘阵列的数据恢复、UNIX 系统的数据恢复、数据库修复技术、Office 文档修复和数据恢复在电子取证中的应用。

本书适合计算机系统管理人员、软件开发人员、计算机硬件维护维修人员、从事计算机取证的工作人员，以及信息安全及相关专业的大专院校的师生阅读。

图书在版编目（CIP）数据

数据恢复高级技术/汪中夏，刘伟编著. —北京：电子工业出版社，2006.11

ISBN 7-121-03310-0

Ⅰ. 数… Ⅱ. ①汪 ②刘… Ⅲ. 数据管理—安全技术 Ⅳ. TP309.3

中国版本图书馆 CIP 数据核字（2006）第 124119 号

责任编辑：刘海艳　特约编辑：杨晓红
印　　刷：北京天宇星印刷厂
装　　订：涿州市桃园装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×1 092　1/16　印张：21　字数：537.6 千字
印　　次：2006 年 11 月第 1 次印刷
印　　数：5 000 册　　定价：34.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系电话：（010）68279077；邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

序　言

很荣幸为本书做审校工作，并为之写序。自2003年《数据恢复技术》第1版出版以来，该学科建设一直没有值得一提的事。《数据恢复高级技术》的出版，应该是该学科建设的一件大事，值得我们庆贺。它的出版，对计算机科学的发展、国家信息化建设，以及电子取证和证据学建设，都会有独到的影响。

该书的作者，汪中夏老师，在我国计算机应用领域做过很多独创的工作，如汉字操作系统、软件的加解密、杀毒软件、国外软件的本地化等，数不胜数。了解我国计算机发展史的读者无不对这些技术了如指掌，而汪老师，正是这些技术发展背后的英雄。现在，汪老师又无私地将这些技术公开给读者，我们这些后生小将，能站在汪老师的肩膀上，为数据恢复事业做出一点点贡献，也不枉负汪老师的一片苦心。

初见刘伟老师，感觉是那种默默地踏踏实实做事的人。两年多来，果不其然，虽不爱多讲，却进步得非常快，为数据恢复事业做了大量的工作。我们期待着刘伟老师逐步将其研究成果都公布出来，促进数据恢复技术大踏步的发展。

本书在介绍部分基本的文件系统的基础上，重点介绍了硬盘物理故障的判定及修复技术、磁盘阵列的数据恢复技术、UNIX的数据恢复技术、数据库修复技术、Office文档修复技术。在全书的最后，介绍了电子取证的初步知识，全书构成了一个有机的体系。在介绍硬盘物理故障的章节中，详细介绍了硬盘物理故障的判定、检测和修复方法，以及专业修复工具的应用。通过这部分内容的学习，可以解决绝大部分的硬盘故障问题。磁盘阵列部分，不仅介绍了各个级别磁盘阵列的原理和恢复技术，还介绍了前沿的DAS、NAS和SAN技术。在UNIX的数据恢复技术介绍中，介绍了UNIX的系统体系、各系统之间的关系、fsck的应用技巧以及虚拟机的应用。通过这些技术，可以大大节省投资，提高研究效率。在数据库修复技术部分，重点介绍了已经搞清楚的SQL数据库的部分结构，这些结构信息，将是进一步研究其结构和修复技术的最好的敲门砖，有这个做基础，做更深入的研究，就不会再迷失在0、1的海洋里。在Office文档修复技术部分，介绍了Office文档的内部结构，与SQL数据库修复技术类似，通过介绍其基本的结构信息，提供进一步的研究基础，为手工修复和开发更好的修复软件提供了必要的入门知识。而电子取证的知识介绍，则将引导读者进入另一个世界。

尽管如此，数据恢复的技术发展仍远远不够，正如很多读者感觉的那样，本书很多的地方深度不够，其中的每一个章节都还有大量的技术宝矿需要深入开采，正因为如此，还希望更多的有志人士投身到这一领域，奉献自己的智慧和才华，为信息安全和保障做出自己的贡献。

总之，瑕不掩瑜，全书的知识体系在电子工业出版社出版的《数据恢复技术》的基

础上，有了全新的突破，是继2003年以来，给读者的最大奉献。所以，该书的出版，必将在数据恢复、电子取证事业发展史上留下自己的烙印。

戴士剑

前　言

作为一名计算机系统管理、软件开发或硬件维护人员，你是否看到自己管理的系统因意外崩溃而无法恢复？你是否面对着重要数据的丢失而无可奈何？你是否正在寻找一套行之有效的数据保护和恢复方案？如果你对上述问题的答案都是肯定的，那么本书对你而言实在是再合适不过的了。解决上述问题可以从加强数据备份工作和数据恢复两方面入手。从目前情况来说，经常地对数据进行备份是最直接有效的。但由于在某些情况下，数据备份也无法避免数据丢失，因此数据恢复成为挽救这些数据的最后机会。学习数据恢复技术已成为全面掌握数据安全知识的一个重要环节。本书将介绍许多有价值的数据恢复技术，以帮助读者更好地避免在数据恢复的过程中盲目地自行摸索，少走弯路，迅速提高读者自身数据恢复的技术水平。

关于本书

本书是由北京信息工程学院数据恢复实验室的（http://www.1huifu.com）众多高手共同打造而成的。

数据恢复的需求，几乎是伴随着计算机的产生而产生的。作为一种或然事件，最早由于使用计算机的用户不够多，能够提供此种技术服务的专业人员则更是少之又少。对于大多数普通计算机用户来说，“找寻失去的世界”是一件无法想像的事，甚至对于很多从事计算机技术研究的专业人员来说，数据恢复也是一种奢望和高深的“特异功能”。不仅在于过去，即使是到了近两年，从2004年夏天起，当我们作为信息产业部数据恢复职业技术培训的主创人员和主要师资，面向全国开展数据恢复专业技术培训项目的时候，发现各行各业，从政府机关到军方机构，从科研院所到企事业单位，对数据恢复技术的了解仍然处于空白或初级这种可怜的地步。

两年多来，我们配合国家信息产业部数据恢复培训项目的其他专家团队，成功地为全国各行各业的信息技术部门培训了数千名数据恢复专业技术人员。他们有的在本单位为信息系统的运行与数据存储的安全做出贡献，有的则走向社会，走向市场，利用自己所学成功地以数据恢复技术开创自己的事业，面向社会大众及各行各业的计算机用户提供数据恢复服务，取得了相当可观的经济和社会效益。

为满足此前已经初步获得信息产业部数据恢复培训证书的大批学员进一步学习更高水平的数据恢复专业技术技能的要求，根据近几年来我们对国际上数据恢复先进技术的进一步跟踪研究和探索，以及近两年来在数据恢复从业实践中所获得的一些实践经验和技巧，我们将这些前沿的新知识、新技术，以及切实有效的技术窍门与经验，总结成书，提供给有志于对更高级数据恢复技术进行研究和从事相关工作的同行们。需要说明

和强调的是，所有本书提供的这些内容，都是经过我们在日常的数据恢复业务实践中反复应用检验有效的，不是那种仅仅是翻译或者抄来的纸上谈兵之作，对于那些跟我们一样天天工作在数据恢复工作一线的技术工程师们来说，其价值远非一本书籍的定价所能涵盖的。

我们很高兴此书在付印前已经作为信息产业部数据恢复提高培训的试用教材，培训了一些学员，在经过学员的反馈，以及我们和其他专家的反复推敲之后，由电子工业出版社正式出版奉献给大家。相信今后作为数据恢复中级培训教材，本书在发挥更大作用的同时，也激励我们为大家奉献更多更好的内容。

本书的内容

要研究数据恢复技术必须对硬盘的物理结构和逻辑结构有一定的了解，掌握文件系统的底层知识。书中首先讲解了相关系统底层知识，如硬盘的内部结构及参数，FAT 和 NTFS 格式深入分析，RAID 原理，未公开的 Office 文件格式的深入研究等。在了解这些底层知识后，就可以将这些技术应用到自己的硬盘数据恢复方案中去，如各种文件和数据库修复技术的实现等。同时本书花了较大篇幅讲解了目前很热门的技术——PC-3000 的应用，以及如何恢复 UNIX 系统的数据等。

本书各章节内容安排如下：

第 1 章　数据恢复基础知识

任何技术都是以一些基本知识为基础的，数据恢复技术也不例外。理解硬盘的基本结构和文件系统的内部原理，有助于揭示众多数据丢失现象发生的真正原因，深刻理解各种数据恢复手段里面的内涵和奥秘。

第 2 章　硬盘物理故障的判定及修复

准确无误地判断硬盘物理故障的原因所在，掌握正确的故障检测方法，是硬盘物理数据恢复成败的关键。一些硬盘的专用工具软件，提供了相当强悍的硬盘镜像功能。利用它们可以修复硬盘“坏扇区”和部分固件，从而达到物理恢复数据的目的。

第 3 章　PC-3000 在数据恢复中的应用

PC-3000 的出现已绝非一日，但很多人可能还没有彻底了解 PC-3000 的工作机制。有关 PC-3000 的知识资料不是很多，详细资料就更少！仅有的一些资料大部分还是重复的，本章将使读者对 PC-3000 有一个全面的了解。PC-3000 不仅是一个强有力的硬盘维修工具，而且还广泛应用于数据恢复中。本章对 PC-3000 在数据恢复中的工作机制进行了深入探究，首先介绍 PC-3000 的原理，然后是 PC-3000 的用法详解，并介绍利用 PC-3000 恢复数据的实例。

第 4 章　磁盘阵列的数据恢复

研究 RAID 数据恢复技术必须先了解 RAID 的工作原理和存储机制。本章主要讲述 RAID 组成原理、RAID 故障分析、恢复方法和 RAID 数据恢复软件工具等内容。

第 5 章　UNIX 系统的数据恢复

与 Windows 系统不同，UNIX/Linux 支持文件系统种类繁多。如果要求一个技术人员熟知每一种文件系统的内部原理和底层知识，几乎是不可能的，也没有这个必要。其实，读者只要掌握了这些文件系统共同点和相互关联之处，就可以从容应对 UNIX 系统下棘手的数据恢复难题。本章介绍的 VMware 虚拟机软件，可以帮助读者在一台机器上建立多种 UNIX 平台，这为 UNIX 系统下硬盘文件的恢复提供了良好的环境和极大的便利。

第 6 章　数据库修复技术

MS SQL Server 数据库是目前应用最广泛的数据库，一旦出现数据库文件的毁坏，造成的损失常常无法估量。目前可利用的数据库修复手段和工具非常少，因此学习和了解数据库文件的内部存储格式将有助于读者解决各种难题，或者进一步自行开发出更实用方便的数据库修复软件。MSSQLRecovery 是一款比较成熟的 MS SQL Server 数据库修复软件，是本章介绍的重点。

第 7 章　Office 文档修复

现在 Office 文档修复工具众多，性能各有优劣。对专业的数据恢复人员而言，不但需要有功能强大的恢复工具，而且更希望了解这些工具的工作原理。本章的目的就是从原理入手，讲述 Office 文档的修复过程，以便读者能够针对 Office 文档的特点，手工对其进行修复，并研究出更好的恢复方法。

第 8 章　数据恢复在电子取证中的应用

电子取证技术是信息技术大体系下一门新兴的交叉学科，数据恢复在电子取证工作中的地位举足轻重。本章就电子取证的法律问题和取证工具做一个简要说明，并举例介绍数据恢复在电子取证中的应用。

对读者的要求

本书是针对具有一些数据恢复经验和一定编程基础的技术人员而写的。作为数据恢复技术人员，有必要对逻辑恢复和物理恢复两方面同时进行研究。也就是说，应该更多地从“软硬”结合的角度考虑，这样才可能比较合理地综合运用各种技术。

由于要与文件系统的底层打交道，不可避免地要使用编程语言，而且使用编程语言来阐述将是最清晰直接和易于理解的，因此需要读者有一定的编程基础。

本书适合对数据恢复技术感兴趣的技术人员、进行信息安全教学和学习的大专院校的教师和学生、对文件系统底层机制感兴趣的读者、从事计算机取证的工作人员、从事硬盘维修的技术人员阅读。

致谢

感谢北京信息工程学院的高级工程师张京生、张宇和宋杨老师对本书的大力支持！

感谢戴上剑老师对全书进行审校，并为本书撰写精彩的序言！并感谢数据恢复实验中心的唐雪梅、郭久武、王凤泰、张杰、李静、韦明等朋友为本书的出版所做的工作！

感谢房金信等众多朋友的支持和帮助！

感谢向本书提供参考资料的各位朋友的技术共享的精神，是他们的无私奉献使读者可以从本书中获得许多极有商业价值的技术。

本程序代码下载网址：http//yydz.phei.com.cn。

作　者

目　录

第1章

数据恢复基础知识

- 硬盘的基本技术介绍
- 数据恢复概述
- 硬盘的逻辑结构
- 常用文件系统介绍
- 磁盘编辑器的使用

1.1 硬盘的基本技术介绍

1956 年 IBM 公司推出了世界上第一个硬盘驱动器 RAMAC（Random Access Method for Accounting and Control）。其体积相当于两个并排放置的 200 立升的大冰箱，容量为 5 MB，价格超过 5 万美元。换句话说，就是每 1 GB 容量的价格为 10 000 万美元。现在，一个 3.5 英寸的硬盘的容量能达到 500 GB，而平均每 1 GB 容量的价格不足 1 美元。硬盘的改进速度大大超过了半导体存储器和电信数据的传输率。

1.1.1 硬盘的基本结构

硬盘是包含可移动部件的复杂的机电一体装置。虽然它通常能安全可靠地运行，但任何人造设备都是有寿命的。硬盘要完成其盘片数据的读/写操作，通常由几大部件组成，如图 1-1 所示，而其中任何一个部件发生故障，都可能导致数据丢失。

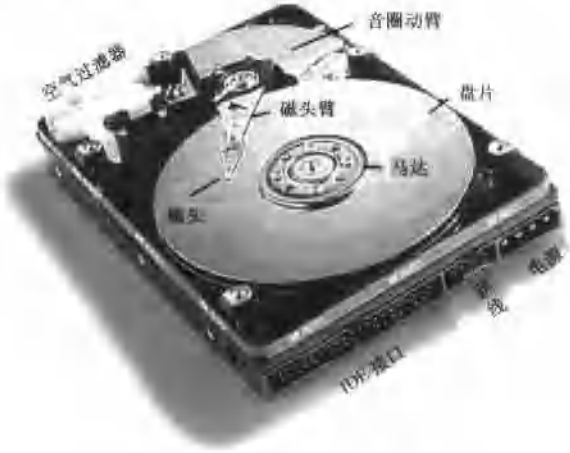

图 1-1 硬盘的基本结构

图 1-1 中，“马达”即“电动机”，为了便于行业内读者阅读，书中“马达”不再一一改为“电动机”。

1.1.2 盘片

一块硬盘可能包含一片或多片磁盘片，磁盘片是通过在铝或玻璃基片上再涂上一层微小的磁铁氧化物颗粒而制成的，磁盘片简称盘片。磁盘片连续转动，磁头沿着半径方向来回移动，这样就可以使磁头快速地到达磁盘片上的任何位置。一般台式计算机使用 3.5 in（英寸）硬盘，每块硬盘安装有 1－3 片磁盘片，可以存放大量的数据信息。每片盘片都被分隔成数万个同心圆，这些同心圆就是常说的磁道。因为每一个磁道上都要存储大量的信

息，为便于管理，进一步把磁道分隔成一个个同样大小的小区域，叫做扇区，每个扇区通常为512 B（字节）大小。磁盘片按照一定的规则安装在一个马达（spindle Motor）上。马达带着磁盘片高速旋转，IDE硬盘的转速一般是5400 rpm（转/分钟）和7200 rpm，SCSI硬盘的转速可高达15 000 rpm。盘片高速转动，与之相配合的磁头能够进行扇区定位操作。一般每片盘片都有一个磁头，并且所有磁头的移动必须完全同步，如图1-2所示。

图1-2 盘片

1.1.3 磁头

磁头由上、下多个读写头固定在一个磁头臂上，整体随磁头臂移动而移动，如图1-3所示。一般的读写原理是，例如，当写10 MB数据时，第1个磁头先写4096 B，第2个磁头再写4096 B，依此类推，呈垂直读/写，所以数据被分很多段存在各盘面上，读取时也是一样，这样，多磁头同时读/写可以大大提高磁盘工作效率。

磁头臂因需高速来回移动，不可抖动，并需精确地移至原地址，各磁头垂直度不可以超出允许误差，所以磁头和盘片都不可松动或偏移，否则将无法定位。目前技术水平还无法将误差的磁头或盘片调整回原位。

图1-3 磁头