广西科学院学报
Journal of Guangxi Academy of Sciences
2006,22(4):370～374
Vol. 22,No. 4 November 2006

# 计算机取证系统核心技术分析
# Analysis on Key Techniques of Forensic Computing System

黄宁宁[1],苏红帆[2]
HUANG Ning-ning[1],SU Hong-fan[2]

(1. 南宁市公安局指挥中心,广西南宁 530012;2. 南宁市公安局交警支队,广西南宁 530028)
(1. Command Center, Nanning Bureau of Public Security, Nanning, Guangxi, 530012, China; 2. Traffic Police Detachment, Nanning Bureau of Public Security, Nanning, Guangxi, 530028, China)

**摘要**:介绍计算机取证的相关知识,分析实现计算机取证系统所涉及的硬盘的数据组织格式、文件系统解读、文件发现技术、删除数据的恢复技术等核心技术,为进一步实现计算机取证系统提供技术支撑。

**关键词**:计算机取证 数字证据 系统解读 数据恢复
**中图法分类号**:TP393 **文献标识码**:A **文章编号**:1002-7378(2006)04-0370-05

**Abstract**: Forensic computing involves processes of retrieving, preserving, analyzing, and presenting of data that have been stored in computer media. It is a hotspot in security field. The techniques about data format, unscrambling of system architecture, file finding, recovery of deleted files in the hard disk are discussed.

**Key words**: forensic computing, forensic duplicates, system architecture, recovery of data

各种基于安全的产品如:软、硬件防火墙、入侵监测系统(IDS)、虚拟专用网(VPN)、防病毒软件等从不同的系统层面,采用不同的策略和技术手段对来自网络内外的安全威胁进行防范,共同构成了一个网络安全防范体系,以防止信息安全事件的发生。

然而,大量计算机犯罪行为,包括信息的窃取、未授权的访问以及恶意传播恶意代码等事件依然层出不穷。据国际计算机应急响应组 CERT 统计显示,计算机攻击事件成上升趋势,仅 2006 年第一季度就高达 1597 起[1]。另据中国国家计算机网络应急技术处理协调中心数据统计,2006 年上半年非扫描类网络安全事件 6765 件,平均每月 1100 多件[2]。

计算机取证是一门涉及计算机领域与法学领域的交叉学科。通过对计算机存储设备中的数据进行采集、保存、分析以获取符合司法要求的数字证据,并通过适当的方式进行展示,以此来证明或推翻某种假设。计算机取证系统是获取计算机犯罪证据的工具。我国目前使用的都是国外的取证系统。建立拥有自主知识产权的计算机取证系统软件对于提高我国司法机关取证水准、确保软件与我国法律更好的结合具有重大的意义。本文就实现计算机取证系统的核心技术进行介绍分析,为进一步实现计算机取证系统提供技术支撑。

## 1 计算机取证

### 1.1 计算机取证的定义、目标与原则

最早关于计算机取证的经典定义是:计算机取证就是针对所关注的潜在的、合法的证据应用计算机进行调查和分析的技巧[3]。

计算机取证更为清晰、公认的定义是:计算机取证是涵盖计算机数据保存、识别、抽取、记录(Documentation)以及解释的一门涉及法学内容的学科,它有三个目标:将罪犯绳之以法;找到犯罪的动机和根源,避免事件再次发生;获得的证据最终得

收稿日期:2006-07-03
修回日期:2006-09-15
作者简介:黄宁宁(1974-),男,广西南宁人,硕士,主要从事计算机取证、计算机网络安全研究。

到法院的认可[4]。计算机取证的基本原则是:(1)在不对原有证物进行任何改动或损坏的前提下获取证物;(2)证明你所获得的证据与原始数据是相同的;(3)在不改动数据的前提下对其进行分析。

### 1.2 计算机取证步骤

计算机取证活动以获取数字证据为目的,其主要步骤或活动逻辑阶段有:保护和勘查、证据采集、证据分析、证据展示[5]。

#### 1.2.1 保护和勘查

一般称之为 Freezing Scene。内容包括:隔离目标设备、禁止未经授权人员靠近可疑设备、明确设备工作状况、搜索现场及设备所在区域并且记录现场的一切操作和程序,然后获取证据,并制作证据标签然后安全运送到指定地点。

#### 1.2.2 证据采集

对电子数据的原始载体比如硬盘、移动存储设备等,进行镜像得到司法鉴定复件。司法鉴定复件是对原始证据进行逐比特(bit)复制得到的与原始证据完全一致的证据副本[6]。

#### 1.2.3 证据分析

这是计算机取证的核心阶段。根据现场勘查获得的所有信息制定策略,进行证据查找分析。如:识别文件的真实格式、恢复被删除的文件、识别加密的文件和文件夹、识别隐藏文件等,找到有效证据后加以固定形成数字证据。

#### 1.2.4 证据展示

对分析结果进行全面的描述,并以符合司法诉讼的要求提供和展示获取的数字证据。

### 1.3 数字证据及其评估

#### 1.3.1 数字证据

计算机取证的目标就是获得证据以证明或推翻某种假设。关于证据的表述存在着很多形式,电子证据的定义在国际上也没有一个准确的定义和表述。从英文表述上有十几种之多[7],目前使用较为普遍的术语是电子证据(Electronic Evidence)、计算机证据(Computer Evidence)以及数字证据(Digital Evidence)。其中数字证据指以二进制存储或传输的能够被法庭认可的信息。

为方便讨论起见,本文以数字证据作为计算机取证的证据描述词汇。

#### 1.3.2 数字证据的评估

由于数字证据的不确定性、易篡改性使得电子证据的客观性、关联性及合法性与传统物理证据相比存在差距,使其可信性的评估非常困难。由于计算机系统的复杂性和多样性,在评估数字证据的可靠性方式上缺乏一致性。计算机犯罪调查专家 Eoghan Casey 提出一个数字证据确定性水平的分类等级,具有很高的借鉴价值,他把数字证据的可信度分为 7 个级别[8]:

C0:错误/不正确的,

C1:非常不确定,

C2:某种程度的不可信,

C3:有可能,

C4:很有可能,

C5:几乎确定,

C6:确定的。

## 2 计算机取证系统核心技术

计算机取证系统需要解读所获取的司法鉴定复件或合格司法鉴定复件(合格司法鉴定复件是一个大文件,除了涵盖原始证据的每一 bit 的信息之外,还可以加入一些校验信息),正确解读司法复件数据格式之后,识别文件系统,然后恢复被删除和隐藏的数据,在此基础上进行犯罪证据的查找。因此建立一套计算机取证系统要解决的几个主要问题是:硬盘数据组织的解读、各种常用文件系统的识别、隐藏数据的发现、被删除数据的恢复等问题。

### 2.1 磁盘数据组织

计算机取证系统是对通过 dd 镜像软件制作的特殊格式文件进行解读,在解读的基础上进行各项分析和调查工作。因此对磁盘的数据组织进行分析成为首要问题。

磁盘主要由盘片、定位伺服系统、磁头以及转动系统组成。硬盘由几个盘片组成,盘片两面都有磁层和一对磁头。磁头通过盘片转动产生的气流悬浮在盘片上移动,可以改变磁介质上的磁颗粒排列或感应其排列达到写和读的功能[8]。数据记录在盘片的同心圆上,称为磁道(track)。所有盘片中所有相同半径的磁道称为柱面(cylinders)。每一个磁道划分为更小的扇区(sector),每一扇区共有 512 字节(另有 45 字节用于低级数据解密,不能用于存储数据,因此通常称扇区大小是 512 字节。)。文件系统在读写数据时采用簇的概念,簇是一个文件系统进行存储时的一个逻辑概念,一个簇包含 4、8、16 个扇区,作为文件存储的最小单位。

硬盘在存储数据之前,一般需要经过低级格式化、分区和高级格式化三个步骤之后才能使用。最终在物理硬盘上建立一定的数据逻辑结构,一般分为

5 个区域：主引导记录区(MBR)、分区引导记录区、文件分配表区、文件目录标区和数据区实现对数据的存储与管理。5 个区域的详细位置如图 1 所示：

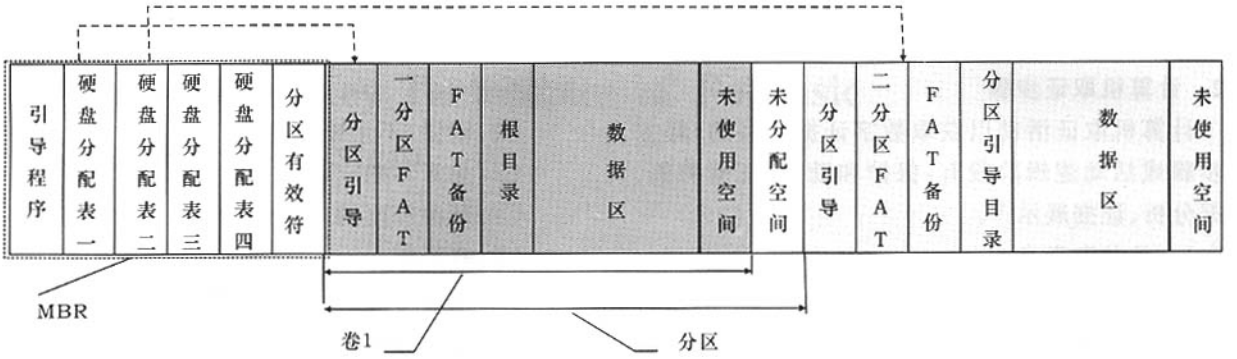


图 1 硬盘分区示意(FAT 文件系统)

## 2.2 FAT16/32 文件系统解读

文件系统是操作系统重要组成部分，它主要对以文件形式存放在外部存储器上的信息进行管理，主要包括如何以文件的形式组织信息，如何通过文件名对存储在介质上的文件进行操作，如何实现文件的共享、保护和保密等。

FAT 文件分配表定义了文件的其它存放簇号及结束标志。根据这两个表即可读取数据。下面给出一个简单的实现实例。

根据簇号获取磁头号、柱面号：

起始逻辑扇区＝隐含扇区数＋1＋2×每 FAT 扇区数＋FDT 扇区数＋(起始簇号－2)×每簇扇区数，其中起始簇号－2 是因为簇号从 2 开始。

柱面号＝起始逻辑扇区/每个柱面扇区数；磁头号＝余数/每磁道扇区数

获得柱面号、磁头号和扇区号后与本逻辑分区的起始柱面号、磁头号和扇区号即可以读取该文件。例如：可以根据文件 FILE1.DAT 的起始扇区号找到文件链的第一扇区，然后根据链读取下一个部分文件存储的扇区，直到读取文件的结束标志 EOF，整个文件即读取完毕。FAT 文件读取如图 2 所示。

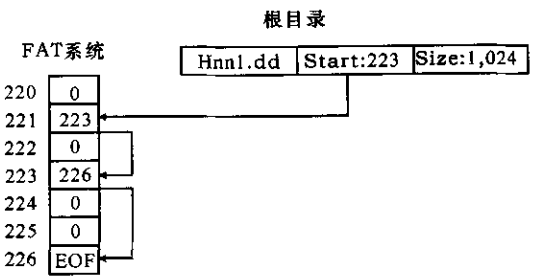


图 2 FAT 文件读取示意

## 2.3 NTFS 文件系统解读

文件系统 NTFS 是微软为操作系统 Windows NT、Windows 2000，Windows XP 及 Windows 2003 Server 设计的一种全新的文件系统。NTFS 将整个磁盘分区上每件事物都看作一个文件，而文件的相关事物又视为一个属性，比如数据属性、文件名属性等[9]。整个 NTFS 分区上每个扇区都被分配属于某些特殊文件，甚至描述文件系统本身的信息(元数据)也是一个文件。NTFS 把所有的文件索引放置在 MFT(Main File Table)中。MFT 则由文件记录(File Record)数组构成，最前端是 16 个基本元数据记录和所有文件的记录。MFT 中最开始的 16 条记录中存放了特殊的信息，从第 17 条记录开始，则全部用于记录磁盘分区上的文件和文件夹(同样被 NTFS 视作 1 个文件)。MFT 为每个文件夹都分配一个固定空间，文件的属性都写在这个固定的空间中。小文件和文件夹(≤1500 字节)可以完全被包含在 MFT 记录中，而大文件则使用“B-树”索引方式来指示扩展的 MFT 外部信息。NTFS 就这样依靠主文件表的详细记录来管理整个磁盘分区。

NTFS 系统使用簇来对文件数据进行定位。NTFS 使用逻辑簇号(LCN)和虚拟簇号(VCN)来对簇进行定位。VCN 是对属于特定文件的簇从头到尾进行编号，以便于引用文件中的数据。VCN 可以映射成 LCN，而不必要求在物理上连续。使用方式如图 3 所示：

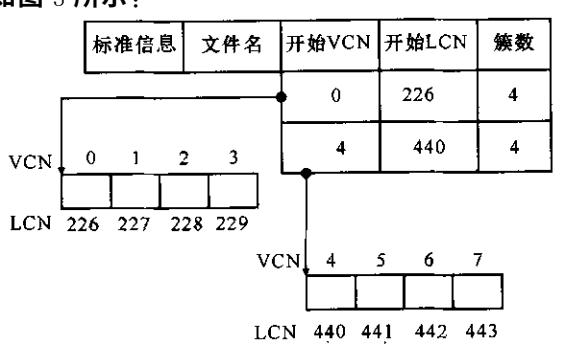


图 3 Data 区数据属性的 VCN-LCN 映射

## 2.4 EXT2/3 文件系统解读

Ext2/3 文件系统(second extended filesystem)

是 Linux 默认直接支持的文件系统,是 Linux 操作系统在计算机的硬盘上存储和检索数据的逻辑方法,包括本地驱动器、网络存储区域网络(SAN)上的导出共享[10]。EXT3 比 EXT2 增加了日志容量,以便文件系统在出现问题后的恢复和修复。EXT/2/3 分区的第一个磁盘块用于引导,其余部分被分成许多组,每一个组具有相同的结构布局:(1)文件系统超级块,描述文件系统的总体信息如 inode 总数、块总数等。(2)所有组的描述符,记录本组的描述信息。(3)块的位图,记录块分配位图的编号。(4)inode 位图,记录 inode 节点即信息节点位图的编号。(5)inode 表,信息节点起始块的块编号。(6)数据块,存储数据的最小单位,可以是 1024、2048 或 4096,安装文件系统时定制。EXT2 文件系统硬盘物理分布如图 4 所示。

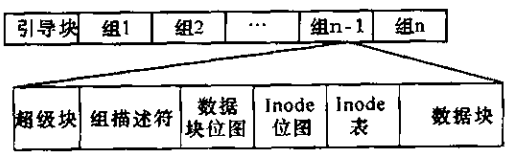


图 4 EXT2 文件系统硬盘物理分布

EXT2/3 文件系统采取索引结构管理文件,采取多级索引结构共有 15 项。0~10 登记项是一级指针,直接存放数据所在的磁盘块号。11 登记项是二级指针,指向的磁盘块是一系列的一级指针,12、13、14 登记项分别是三级、四级、五级指针,如图 5 所示。EXT2/3 的文件索引结构在支持大文件同时保证了对大量小文件访问效率[10]。

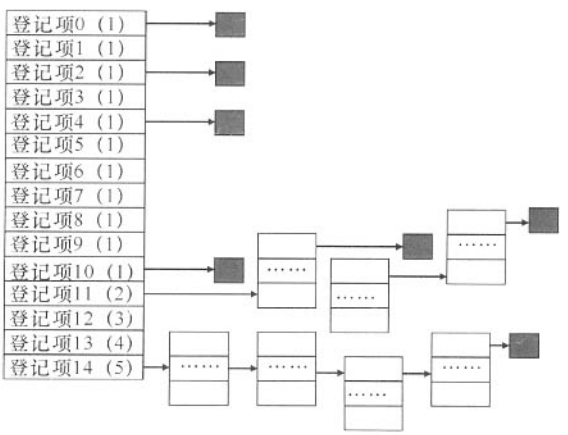


图 5 EXT2/3 文件索引结构

## 2.5 数据流隐藏数据的发现技术

数据流隐藏技术可以把任何的文件作为某个文件的另一个实例而附在文件中,而 Windows 资源管理器却无显示。该项技术源自 Macintosh 分级文件系统 HFS。尤其隐蔽的是,隐藏前后的文件的 MD5 值没有改变,这就给我们查找隐藏文件带来了困难。

目前,国外的一些取证软件已经实现了隐藏数据流的读取识别。由于 NTFS 系统将任何文件作为属性/属性值的集合来处理,数据作为属性值的存在而存储在 NTFS 的一个或多个运行中。由于文件的属性可以增加,因而可以增加多个属性值即数据并存储在运行中。在解读 NTFS 文件系统的证据镜像文件时,检查标准的文件属性类别即可识别出隐藏的文件。

## 2.6 闲散空间的数据发现技术

文件管理系统是按照"簇"的大小来读、写数据的。不论文件大小,每次读、写的大小都是指定的大小块。绝大多数的文件不会刚好等于簇或簇的整数倍大小。因而会有一些空间目前没有使用。对于删除的数据中,极有可能含有之前被删除文件的碎片。该区域就是 Slack Space 闲散空间。Slack space 包括 Slack RAM、Slack file 以及 Slack volume。文件末端到扇区末端的空间称为 Slack RAM,文件结束标志所在扇区到该簇末端的扇区称为 Slack file,在分区和卷之间的空间称之为 Slack Volume。Slack RAM 空闲块包含了一小部分内存中可执行文件的一个小片断,Slack File 文件空闲块则可能包含有删除数据以及硬盘厂商设置的原始数据。因此,在解读镜像文件的时候,按照簇链来读取文件的同时,可以记录或读取 Slack Space 中文件碎片的二进制数据留待进行证据分析时使用。Slack space 如图 6 所示:

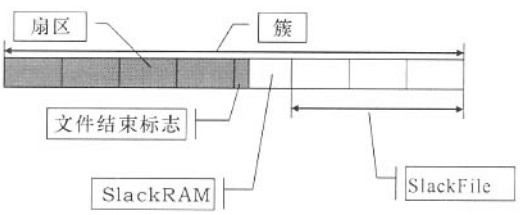


图 6 Slack Space 示意

## 2.7 删除数据的恢复技术

文件系统的数据删除,实质是把文件索引表清空,但实际的数据并未删除,直到存储该数据的物理扇区被覆盖之后实际数据才会被删除。下面以 FAT16/32 文件系统、EXT2/3 文件系统下的、未被覆盖的数据为例分析恢复技术。

### 2.7.1 FAT16/32 系统删除数据恢复

在 FAT 文件系统中,删除文件实质上是把文件移到回收站目录中并把文件目录项中的第一个字节更改为"E5H",其他字节没有变化[11]。对于长文件名是把描述长文件名的登记项的所有子节都改为"E5H",表示该文件已经删除。存储在实际簇中的文件片段依然存在,空间没有真正释放出来。如果此时

执行回收站的清空操作，就把FAT登记的目录项清0，实际的DATA区中的相应扇区没有变化。此时若没有新的数据写入，虽然FAT表链经过更改，但是实际数据还在，这就给找回数据提供了可能。文件分配表FAT在文件系统中是存在一个冗余备份的特点，把备份写回原扇区即可。

2.7.2 EXT2/3系统删除数据的恢复

对于EXT2/3文件系统中的文件被删除后，文件系统把文件的目录项隐藏，标记相关索引节点可以被其他文件占用，文件目录项、索引节点和数据依然保留在硬盘上，直至被覆盖。可以使用grep命令搜索关键字的方法来恢复文件，命令如下：

# grep -a B5 -A100 "computing forensic" /dev/sad4 >recover.txt。

命令在/dev/sda4上搜索文本"computing forensic"并且返回该字符串的前5行和后100行的文本，把结果重定向到recove.txt中。如果文件是碎片状态，则只能返回部分文件。前后行数设置不准确则会把前后二进制数据也存储进recover.txt中。

## 3 结束语

目前，在计算机取证领域中普遍使用的取证系统如EnCase、FTK等，都是国外的软件。开发拥有自主知识产权的取证系统，对于我国的取证工作的进一步规范化发展具有重大的现实意义。取证系统需要解决三个问题：数据的解读、取证分析及证据展示。本文对数据解读的文件系统的解读、删除数据的恢复等技术进行了详细的分析，为进一步实现计算机取证系统提供必要的基础。

参考文献：


[1] CERT/CC 计算机紧急响应组. CERT/CC Statistics 1988-2006：lnerabilites reported[EB/OL]. [2006-06-05]. http://www.cert.org/stats/cert stats.html.

[2] CNCERT/CC 国家计算机网络应急技术处理协调中心. 2006年上半年网络安全工作报告[M/DK]. [2006-08-2]. http://www.cert.org/stats/cert stats.html.

[3] ROBBINS JUDD. An Explanation of Computer Forensics [EB/OL]. [2006-07-08]. http://www.computerforemics.net/forensics.htm.

[4] WARREN G KRUSE Ⅱ, JAY G HEISER. 计算机取证：应急响应精要[M]. 段海新，刘武，赵乐南，译. 北京：人民邮电出版社，2003.

[5] 许榕生，吴海燕，刘宝旭. 计算机取证概述[J]. 计算机工程与应用，2001，37(21)：7-8.

[6] KEVIN MANDIA，CHRIS PROSISE. 应急响应：计算机犯罪调查[M]. 常晓波，译. 北京：清华大学出版社，2002.

[7] 何家弘. 电子证据法[M]. 北京：法律出版社，2002.

[8] EOGHAN CASEY. 数字证据与计算机犯罪[M]. 陈圣琳，汤代禄，韩建俊，等译. 第2版. 北京：电子工业出版社，2004.

[9] BRIAN CARRIER. File system forensic analysis[M]. Boston：Addison Wesley Professional，2005.

[10] MOSHE BAR. Linux 文件系统[M]. 天宏工作室译. 北京：清华大学出版社，2003.

[11] 戴士剑，陈永红. 数据恢复技术[M]. 北京：电子工业出版社，2003.


（责任编辑：韦廷宗）

## 柳州市城市应急联动指挥体系共享平台一期工程通过验收

广西信息产业局于2006年4月18日在柳州市组织"柳州市城市应急联动指挥体系共享平台一期工程"项目验收会。柳州市城市应急联动指挥体系共享平台一期工程项目构建了柳州市应急联动共享平台，该平台连接市应急联动指挥中心、市公安局110/122(含6个县)指挥中心、消防119(含二级终端)指挥中心、急救120(含二级终端)指挥中心，实现110、122、119、120四个部门的应急联动，建设了共享平台的基础数据库，建立了统一的应急联动接处警系统、GIS地理信息系统、GPS定位系统、大屏幕系统以及本地网系统。

项目总体方案采用在市政府统一指挥下的协同联动的管理模式，符合国际上现代化城市公共安全体系建设的发展趋势，系统采用物理分散、逻辑集中的总体结构，按照统一接警、分类处警的业务模式建设应急联动系统，较好地解决了应急联动与目前体制分散的矛盾，在应急联动系统模式—处警终端物理分散、逻辑总控集中方面有创新，达到国内同类系统的先进水平。

该系统于2005年9月全部投入使用，运行良好，提高了接处警的效率。

（陈友初）