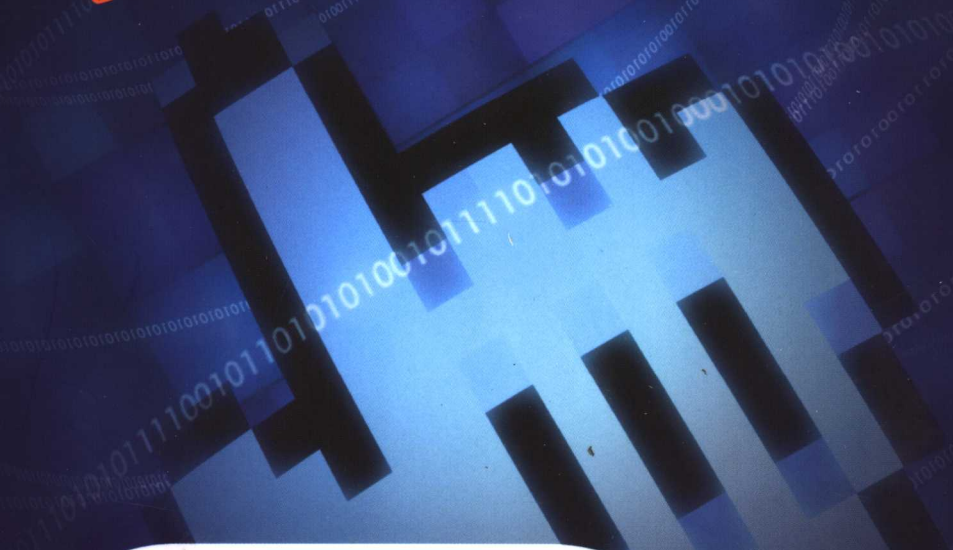
专家点拨 高手支招
解析全面 化解疑点

本书含光盘

# 数据安全与编程技术

涂彦晖 戴士剑 编著



深入剖析主流文件系统结构
防范黑客攻击与病毒破坏
全面揭示数据恢复技术
提供数据安全程序设计实例

清华大学出版社

# 数据安全与编程技术

涂彦晖　戴士剑　编著

清华大学出版社
北　京

## 内 容 简 介

随着各行业信息化程度的加深，数据安全越来越被人们所关注。计算机数据的保护、计算机数据的灾难恢复以及与此相关的程序设计，也逐渐成为一门新兴的技术。

本书由浅入深地对与数据安全有关的技术细节进行了深入的介绍。本书分为9章，内容包括了磁盘基础知识、硬盘的数据存储结构、FAT文件系统、NTFS文件系统、威胁数据安全的因素、数据恢复技术、数据安全程序设计基础、数据备份与恢复程序设计实例，以及与数据安全有关的一些程序设计实例。

本书可作为大专院校教材，也适合于IT系统客户服务人员、技术支持工程师、技术培训人员、数据恢复技术工程师和对数据安全程序设计有兴趣的读者。



**图书在版编目(CIP)数据**

数据安全与编程技术/涂彦晖，戴士剑编著. —北京：清华大学出版社，2005.8

ISBN 7-302-11080-8

Ⅰ. 数…　Ⅱ. ①涂…　②戴…　Ⅲ. 电子计算机-数据管理-安全技术　Ⅳ. TP309.2

中国版本图书馆CIP数据核字(2005)第050988号

**出 版 者**：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
**社 总 机**：010-62770175　　**客户服务**：010-62776969
**责任编辑**：魏江江
**印 装 者**：三河市春园印刷有限公司
**发 行 者**：新华书店总店北京发行所
**开　　本**：185×260　**印张**：20.75　**字数**：515千字
**版　　次**：2005年8月第1版　2005年8月第1次印刷
**书　　号**：ISBN 7-302-11080-8/TP·7337
**印　　数**：1～3000
**定　　价**：39.00元(含光盘)

# 前　言

随着信息化的发展，计算机和互联网络越来越成为了人们生活的一部分，人们的工作、学习、生活已经完全和计算机分不开了。计算机的操作逐渐趋向傻瓜化，即使一个没有太多计算机知识的人也能够不需培训而很快掌握计算机的操作和从网络中获取信息的方法。但是在计算机与互联网络发展的同时，病毒、蠕虫、黑客、误操作等也正在肆意破坏着人们宝贵的数据。因此如何防范对数据的破坏，如何恢复被病毒或人为破坏的数据，甚至如何编写与数据安全相关的程序越来越引起了人们的关注，这正是本书编写的目的。

## 读者对象

本书适合如下人员：

- 数据安全与磁盘存储技术的爱好者和从业人员
- 数据恢复爱好者和从业人员
- IT 系统客户服务人员
- 技术支持工程师
- 各大专院校的在校学生及教师
- 数据安全程序的设计爱好者和从业人员

## 主要内容

本书共分 3 篇。

第 1 篇是本书的基础知识，共包括 4 章，其中 NTFS 文件系统的分析在其他资料和图书中都鲜见提及。第 1 章简单介绍硬盘的物理结构及逻辑结构；第 2 章重点介绍数据在硬盘中存储的总体结构，介绍了两个对磁盘进行编辑的工具，并对主引导扇区的结构和硬盘分区表与扩展分区表的结构进行了详细的介绍；第 3 章重点分析 FAT 文件系统的结构，重点介绍了引导扇区的结构、磁盘文件分配表的结构以及目录项的结构，剖析了 FAT 文件系统文件的管理方法及 FAT 文件系统中树型目录的实现原理；第 4 章重点分析 NTFS 文件系统的结构，重点介绍了 NTFS 文件系统的引导扇区结构、文件属性结构、MFT 结构、NTFS 文件系统元数据文件结构以及 NTFS 文件系统的树型目录实现原理等。

第 2 篇具体阐述数据安全与恢复，共包括 2 章，尤其以数据恢复为重点。第 5 章分析了威胁数据安全的几个因素，如来自黑客与病毒的攻击，并分析了造成黑客入侵成功的几个原因及防范方法，介绍了几种常见病毒的分析与解除方法；第 6 章重点介绍了数据的恢复与备份技术，对一些常见的数据丢失原因及手工或工具恢复做了介绍，对一些重要的系统数据的备份方法做了概述。

第 3 篇牵涉到数据安全方面的具体程序设计，共包括 3 章。第 7 章是数据安全程序设

计基础知识，对在实模式下调用 BIOS 中断进行物理磁盘访问的方法及在保护模式下对磁盘扇区进行访问的方法，以及在 Windows NT 核心系统下对磁盘扇区的访问做了详细的介绍，并对微软公司的引导代码如主引导程序和引导记录进行了反汇编分析，这对于学习与数据存储相关的程序设计是很有帮助的；第 8 章用 3 个实例程序对数据备份与恢复做了介绍，其中 NTFS 文件系统中误删除文件的恢复程序充分体现了能够手工操作的就能够编程实现，读者可以在此程序的基础上自行开发和完善该程序的功能；第 9 章详细分析和介绍了数据安全的攻防两个方面的程序设计方法、多操作系统实现原理和程序设计方法，以及知识产权的保护程序设计等。

## 本书特点

本书具有以下特点：

- 读者对象广泛。该书从硬盘结构、文件系统等基础知识谈起，以 3 大篇幅探讨了与数据安全相关的知识，如数据恢复与备份、病毒与黑客的防范、数据安全相关程序设计等，该书既适合仅需了解磁盘结构及文件系统如 NTFS 文件系统的技术爱好者，又适合数据恢复技术的爱好者与从业人员，更适合关注数据安全的程序设计爱好者与程序员。
- 实用性与可操作性强。该书无论是对病毒与黑客的防范还是对数据恢复技术的讲解分析，以及本书中列举的大量原创程序，都非常具有实用性和可操作性。
- 知识点覆盖面广。本书探讨了几乎所有与数据安全相关的知识点，包括硬盘结构、文件系统、病毒与黑客防范、数据恢复与备份技术、实模式下对硬盘扇区的访问、保护模式下对硬盘扇区的访问、微软公司的部分代码的分析、数据备份程序的编写、数据恢复程序的编写、病毒破坏模块的原理分析、多操作系统引导程序的原理及编写实例、虚拟还原技术的分析与破解程序实例、硬盘加密程序的编写实例及共享软件注册模块的编写等。
- 讲解深入彻底。本书中所有对基础知识的讲解都非常深入，无论是对 FAT 文件系统还是对 NTFS 文件系统的讲解，或者是对磁盘逻辑组织结构的讲解，都深入到每一个有用的字节，对数据恢复的分析也不是停留在某个软件的使用下的。虽然为照顾一些基础比较弱的读者，书中也介绍了一些数据恢复软件的使用，但作者的本意是让所有的读者都能够了解数据恢复的原理，因此书中探讨的数据恢复技术是包含 3 个方面的，第一方面是工具的使用，第二方面是数据恢复原理的掌握，第三方面是手工数据恢复，这样读者才能够正确分析需恢复的数据，做到即使工具软件不能恢复的情况下也能正确恢复数据，并不造成二次的破坏。本书中所有的程序，作者都加入了非常详细的注释，有的程序如 NTFS 中误删除文件的恢复程序甚至是每一句都加了注释，虽然本书的程序都是用汇编语言编写的，但注释的详细加上程序原理的分析，足以让一个不懂汇编语言的程序爱好者看懂，从而可以轻易地将其改为其他语言编写的程序，更不用说一个略懂汇编语言的编程人员或者爱好者了。

本书由涂彦晖、戴士剑主编，由魏江江、涂彦广、胡艳芳、裘亦斌、黄小平、万仁甫、张宇、

黄志波等共同完成编写和审校工作。由于时间仓促，加之编者水平有限，书中难免会存在一些疏漏和不足之处，恳请广大读者和专家指正。

本书技术支持网站：http://www.itbook8.com。

编 者

# 目　录

## 第1篇　硬盘结构与文件系统

## 第 2 篇 数据安全与恢复



## 第 3 篇 数据安全程序设计

# 第1篇　硬盘结构与文件系统

本篇内容是基础知识，包括了文件系统的分析和介绍以及NTFS文件系统的分析，这些都是学习和掌握数据恢复技术、数据安全程序设计的基础。如果读者对这些知识已经掌握，可以跳过本篇，直接学习下一篇。

## 第1章　硬盘基础知识

硬盘是计算机中海量存储设备之一，本章将对硬盘的物理结构、逻辑结构进行简单的介绍。

### 1.1　闲话硬盘——从最大到最小

最近东芝公司推出了一款0.85英寸(1in=2.54cm)的硬盘，如图1-1所示。这款硬盘大小为32mm×24mm×(3～5)mm(长×宽×高)，质量只有10g，容量为2～4GB。这款硬盘只有1.8英寸硬盘的四分之一大小，可以作为数码播放器、数码相机、PDA等移动产品的储存器使用。这款硬盘被吉尼斯承认为一项“全球最小的HDD”世界记录。该硬盘产品在2003年年底发布，2004年夏天推出样机，到秋季正式实现批量生产。

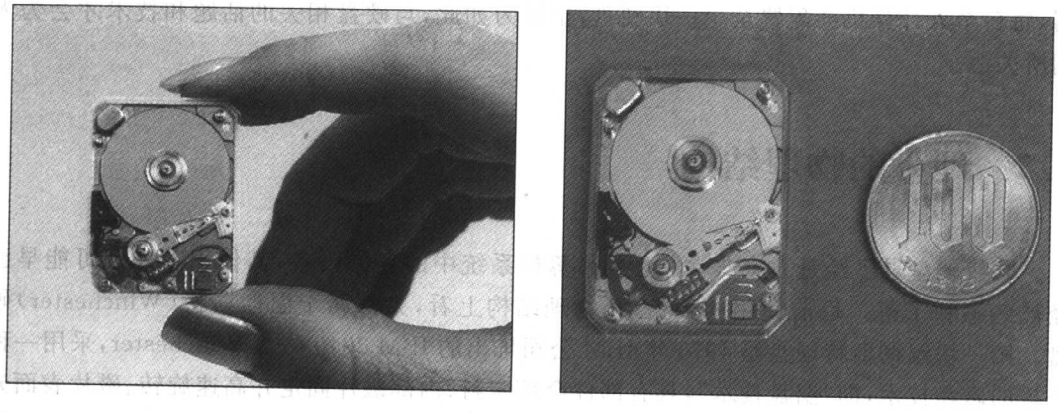


图1-1　世界上最小的硬盘

再看看1956年IBM公司推出的第一台硬盘驱动器IBM RAMAC 350，如图1-2所示，这块大硬盘使用的是24英寸的盘片，总共50片，具有5MB的容量，每张盘片容量0.1MB，转速1 200r/min，传输率为0.008 8MB/S，存储密度(单位面积存储的二进制位)为

2Kb/in²。这“块”(更应该说是台)硬盘有着巨大的体积,而且单位存储容量的价格高达 10 000 美元/MB。但它却是世界上的第一“块”硬盘驱动器。

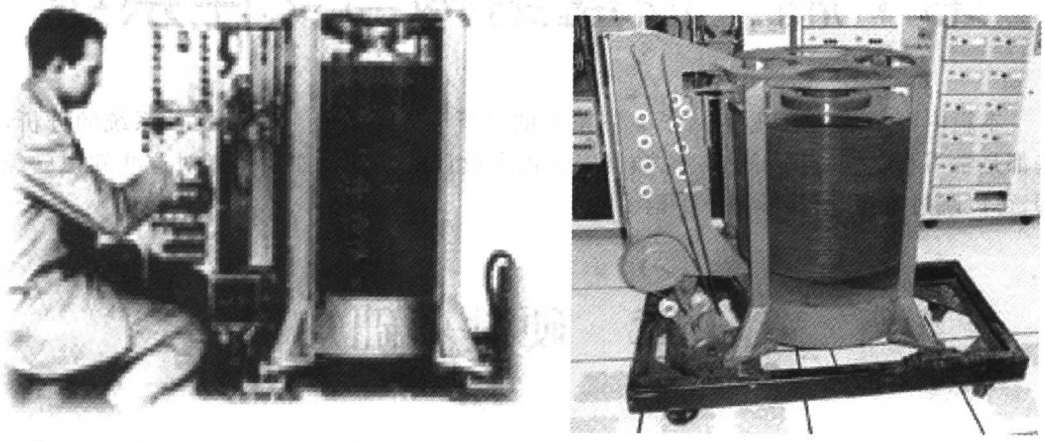


**图 1-2　第一台硬盘驱动器 IBM RAMAC 350**

日立公司在 2004 年 3 月发布了 Deskstar 7K400 硬盘,这款大容量硬盘采用了 S-ATA150 接口,8MB 的缓存,单碟容量为 80GB。

硬盘的发展之快,绝对不是仅仅体现在尺寸上或容量上,从 5MB 容量到今天的 400GB,从单碟 0.1MB 到单碟 80GB,从庞然大物到小巧玲珑,在这几十年里,硬盘驱动器在控制技术、接口标准、机械结构等方面都进行了一系列改进。正是这一系列技术上的研究与突破,使人们今天终于用上了容量更大、体积更小、速度更快、性能更可靠、价格更便宜的硬盘。

如今,其他存储设备的发展,虽然使得用户在大容量驱动器选择上大有挑选余地,但硬盘以其容量大、体积小、速度快、价格便宜等优点,依然成为个人计算机最主要的外部存储器,每台个人计算机会配置至少一块硬盘,正因为如此,与硬盘相关的话题和技术才会为人们所关注。

## 1.2　硬盘的物理结构

既然硬盘早已经是大家非常熟悉的计算机系统中的存储设备,其作用和外观可能早已经被熟悉和了解。目前使用的硬盘,从物理结构上看,大都属于温彻斯特(Winchester)硬盘。第一块温彻斯特硬盘是 1973 年 IBM 公司推出的 IBM 3340 磁盘 Winchester,采用一种先进的磁盘技术,即“温盘技术”,具有部件全部密封、内部磁片固定并高速旋转、磁片表面光滑、使用时磁头不与磁片直接接触等特点。打开硬盘外面的金属壳可以看到硬盘的内部物理结构,如图 1-3 所示。

硬盘的内部一般由磁片、磁头及定位系统、电动机以及电子线路构成。

- 磁片:磁片由表面镀有磁性物质的金属或玻璃等制成,多个磁片叠在一起共同构成磁盘体。

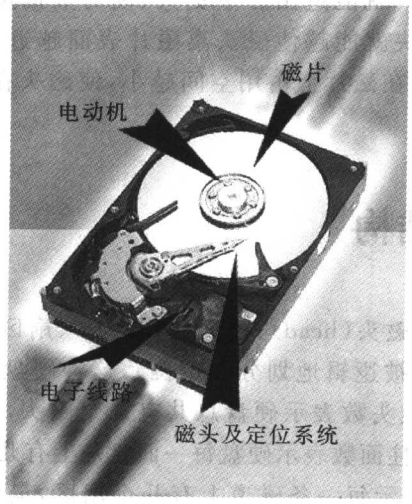



**图 1-3　硬盘的内部物理结构**

- 磁头及其定位系统：每张磁片的正反两面各有一个磁头，磁头是硬盘中最昂贵的部件，也是硬盘技术中最重要和最关键的部分。传统的磁头是读写合一的电磁感应式磁头，但是，硬盘的读、写却是两种截然不同的操作，为此，这种二合一磁头在设计时必须要同时兼顾到读、写两种特性，从而造成了硬盘设计上的局限。而 MR 磁头(Magnetoresistive Heads)，即磁阻磁头，采用的是分离式的磁头结构：写入磁头仍采用传统的磁感应磁头(MR 磁头不能进行写操作)，读取磁头则采用新型的 MR 磁头，即所谓的感应写、磁阻读。这样，在设计时就可以针对两者的不同特性分别进行优化，以得到最好的读、写性能。另外，MR 磁头是通过阻值变化而不是电流变化去感应信号幅度，因而对信号变化相当敏感，读取数据的准确性也相应提高。而且由于读取的信号幅度与磁道宽度无关，故磁道可以做得很窄，从而提高了磁片密度，达到 200MB/in$^2$，而使用传统的磁头只能达到 20MB/in$^2$，这也是 MR 磁头被广泛应用的最主要原因。目前，MR 磁头已得到广泛应用，而采用多层结构和磁阻效应更好的材料制作的 GMR 磁头(Giant Magnetoresistive Heads)也逐渐普及。
- 电动机：所有磁片都由主轴电动机带动旋转。
- 电子线路：是硬盘的控制集成电路板，其结构非常复杂，由硬盘 ROM(内有软件系统)、硬盘缓存(cache)和主控制芯片等构成。

磁片、磁头、电动机以及电子线路都密封在一个无尘的金属壳中。

硬盘存储数据是根据电、磁转换原理实现的。硬盘工作时，磁片高速旋转，设置在磁片表面的磁头则在电路控制下径向移动到指定位置，然后将数据存储或读取出来。当系统向硬盘写入数据时，磁头中“写数据”电流产生磁场使盘片表面磁性物质状态发生改变，并在“写数据”电流磁场消失后仍能保持，这样数据就存储下来了；当系统从硬盘中读数据时，磁头经过盘片指定区域，磁片表面磁场使磁头产生感应电流或线圈阻抗产生变化，经相关电路处理后还原成数据。因此只要能将磁片表面处理得更平滑、磁头设计得更精密

以及尽量提高盘片旋转速度，就能造出容量更大、读写数据速度更快的硬盘。这是因为磁片表面处理越平、转速越快就能越使磁头离磁片表面越近，提高读、写灵敏度和速度；磁头设计越精密就能使磁头在磁片上占用空间越小，使磁头在一张磁片上能建立更多的磁道以存储更多的数据。

## 1.3　硬盘的逻辑结构

早期的硬盘技术是通过磁头(head)、柱面(cylinder)、扇区(sector)对硬盘进行访问的(这个磁头、柱面和扇区一般被逻辑地划分和定位)，这种硬盘访问方式被称为 3D 寻址方式，如图 1-4 所示。其中，磁头数表示硬盘总共有几个磁头，也就是有几面盘片，最大为 255(用 8 个二进制位存储)；柱面数表示硬盘每一面盘片上有几条柱面，最大为 1 023 (用 10 个二进制位存储)；扇区数表示每一条磁道上有几个扇区，最大为 63(用 6 个二进制位存储)。每个扇区一般是 512B。在 BIOS 中断 13H 的入口参数中，寄存器 CH 是磁头号，其值为 0H～FEH(最多 255 个磁头)，寄存器 CL 中低 6 位为扇区号，其值为 1H～3FH(最多 63 个扇区)，寄存器 DH 为柱面号的低 8 位，寄存器 CL 中的高 2 位为柱面号的高 2 位，也就是说，柱面号最多由 10 位二进制数表示，$(1111111111)_2=(1023)_{10}$，即最多可以表示的柱面数为 0～1 023，共 1 024 个。由此可以看出基于这种访问方式最大能访问的磁盘容量为

$$255\times1024\times63\times512\ /\ 1048576\ =\ 8\,032.5\text{MB}$$

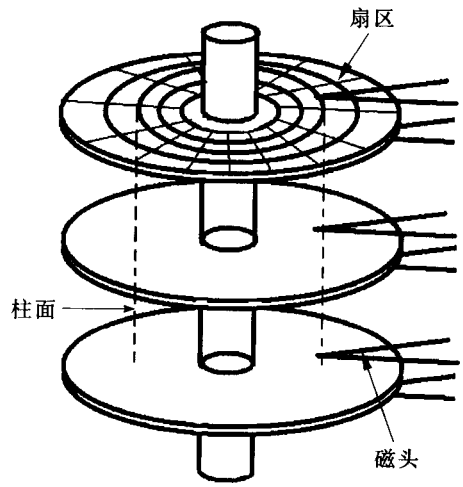



**图 1-4　硬盘的逻辑结构**

只有大约 8GB 的空间。这是因为早期磁盘容量还很小。就好像当年的计算机专家认为 1KB 内存已经很大了一样，让计算机用户很长一段时间都为配置 DOS 下的内存而烦恼。

硬盘技术的发展，总是会遇到一个一个的困难，硬盘大小受技术的限制也并不是第一次，早一点的硬盘用户一定不会忘记 528MB 的容量限制。

早期的硬盘容量都很小(比 8GB 还要小),笔者曾用的第一块硬盘容量是 200MB。当时设计 BIOS 时,INT 13H(磁盘中断)用来定位磁头的寄存器只用了 4 位。当时的磁盘柱面最大数为 1 024($2^{10}$),磁头的最大数是 16($2^4$),扇区的最大数是 63。因此能寻址的扇区数为 1024×16×63×512B/扇,这样 IDE 硬盘的最大容量大约为 528MB,528MB 的硬盘容量限制就出现了。

当时解决 528MB 限制的方法,是将 BIOS 中断 13H 的入口参数改为用寄存器 CH 存放磁头号,其值为 0H～FEH(最多 255 个磁道);用寄存器 CL 中低 6 位存放扇区号,其值为 1H～3FH(最多 63 个扇区);用寄存器 DH 存放磁头号的低 8 位;用寄存器 CL 中的高 2 位存放柱面号的高 2 位。也就是说,柱面号最多由 10 位二进制数表示,$(1111111111)_2=(1023)_{10}$,从 0 号柱面到 1 023 号柱面,即最多可以表示的柱面数为 1 024 个。不过由此又引发了新的问题,就是上面所谈到的硬盘 8GB 容量限制问题。其实当时为了解决 528MB 容量限制的问题,人们提出一些不同的办法,其中一个办法是将用于存放磁头的寄存器中没有用到的 4 位中的 2 位(确切地说是高 2 位)保留给柱面数的第 11、12 位使用。这样,最大的磁头数就是 64($2^6$)。但是这种方法最终并没有被采用,在当时只有极少数主板的 BIOS 采用了这种方法访问硬盘。

如今,大于 8GB 的硬盘是因为采用了一种更新的硬盘访问技术——扩展硬盘中断技术。该技术采用线性寻址方式存取硬盘,以扇区为单位进行存取,突破了 8 GB 的限制,并且加入了对可拆卸介质(如活动硬盘)的支持。

较早的硬盘是基于柱面、磁头、扇区寻址,其每个磁道的扇区数相等,所以外道的记录密度要远低于内道,因此会浪费很多磁盘空间。为了进一步提高硬盘容量,现在硬盘厂商都改用等密度结构生产硬盘。每个扇区在磁道上的长度相等,这样一来外道的扇区比内道多。采用这种结构后,明显能提高硬盘的单碟容量,硬盘不再具有物理上的 3D 参数,寻址方式也改为线性寻址,即以扇区为单位进行寻址。另外,老式的基于 3D 寻址方式的软件仍然可以使用,例如 Diskedit 软件还可以访问硬盘约 8GB 的内容,这是因为新的硬盘为了和使用 3D 寻址的软件兼容,厂商通常在硬盘控制器内部安装了一个地址翻译器,负责将老式 3D 参数翻译成新的线性参数。

当使用一些较新的磁盘编辑软件(比如 WinHex)时,仍然会用柱面、磁头、扇区来对磁盘位置进行描述,如图 1-5 所示,即使是柱面号已经超过了 1 023。一方面是因为用户习惯了用 3D 寻址方式来描述硬盘的位置,但实际上系统还是用线性地址寻址,也就是通过扇区号对硬盘位置进行描述和定位的。另一方面系统在对磁盘进行分区操作时,系统分区开始总是开始于柱面的起始扇区,分区结束总是在系统分区的最后一个扇区,即使是在柱面号大于 1 023 时,即分区开始的主引导区或扩展(虚拟)主引导区总是 X 柱面、0 磁头、1 扇区(其中 X 有可能大于 1 023),一个分区开始的起始扇区(引导扇区)总是从 X 柱面、1 磁头、1 扇区开始的(其中 X 有可能大于 1 023),分区结束的扇区总是 Y 柱面、254 磁头、63 扇区(其中 Y 有可能大于 1 023)。物理磁盘的最后一个扇区总是 Z 柱面、254 磁头、63 扇区(其中 Z 有可能大于 1 023)。


**图 1-5 WinHex 对大硬盘还是用柱面、磁头、扇区描述**