扇区读写技术与应用

李恒明
（空军航空大学，长春 130022）

〔摘　要〕　计算机中的硬盘只有在主引导记录、分区表、分区引导记录、文件分配表，文件目录表正常的情况下才能工作，然而硬盘在使用过程中，由于病毒、黑客误操作及其它方面的原因，常常使其中有一项不正常，从而使数据遭到破坏。有没有办法挽救硬盘，恢复数据，尽可能将损失减小呢？答案是肯定的。这就是本文要探讨一项技术——“扇区读写技术”以及在数据恢复等方面中的应用。

〔关键词〕　扇区；引导记录；分区表

〔Abstract〕　Only when the main boot record, partition table, partition boot record, file allocation table and file catalogue table are in the normal state, can the hard disk of the computer work. One of them, however, is often in an abnormal state, when the hard disk operates, for reasons of virus, hacking, misoperation and others, therefore making data destroyed. Is there any way to save the hard disk and recover data, minimizing the loss? The answer is positive. This paper discusses such a technology—“sector read/write” technology and its application in data recovery.

〔Key words〕　sector; boot record; partition table
〔中图分类号〕TP333.3+5　〔文献标识码〕B　〔文章编号〕1008－0821（2006）02－0223－03

1　硬盘的数据结构

硬盘只有建立起完整的数据结构，才能正常使用。通过读取扇区数据来恢复文件的方法，是完全不用考虑被操作硬盘安装的是什么操作系统，甚至都不用考虑硬盘上有没有操作系统，因为这种方法是通过调用 BIOS 磁盘服务程序来完成的。而 BIOS 对硬盘的管理级别高于所有的操作系统，但这种操作必须了解硬盘有关的数据结构和文件的存储方式，才能按照文件的存贮规律将它们恢复出来。因此有必要了解一下硬盘的数据结构。硬盘的数据结构是由 6 部分组成的，它们分别是主引导记录、主分区表和分区表链、分区引导记录、文件分配表、文件目录表以及数据区，分述如下：

1.1　主引导记录（MBR）

1 块硬盘上只有 1 个主引导记录，位于 0 柱面、0 磁头、1 扇区，共 512 个字节。分别由主引导记录、分区表和结束标志 55AA 组成。主引导记录是硬盘启动时最先加载的扇区数据。把自己读入内存后，就查找在分区表中是否有活动分区，找到活动分区后，就执行分区引导记录中的启动程序。将控制权交给操作系统，此外还得检查结束标志是否等于 AA55H。所以主引导记录不正常，后面所有的启动过程都不可能正常执行。主引导记录中的第二部分是分区表，位移从 1BEH 到 1FDH 共 64 个字节。它共有 4 个分区表项，一般只使用 2 个分区表项，另外 2 个表项全为 0。第一个分区表项记录着本分区的有关参数。它们是：本分区能否启动、本分区的起始磁头号、本分区的起始扇区值和柱面值、本分区操作系统的 ID 值、本分区结束磁头号、本分区结束扇区值和柱面值，本分区前扇区总数以及本分区内的扇区总数。第二个分区表项记录的是下一分区的有关参数。其大部分字段和第一分区记录的信息相同。但其中有一字段记录着下一个分区起始柱面地址和扇区地址。通常说的分区表指的是主分区表，另外在扩展分区的每一个逻辑驱动器中，都有一个分区链表，通过该字段就可以把所有分区表链连起来。通过分区表的解读和处理，我们就可以把硬盘上的所有逻辑驱动器的分区表扇区地址，以及分区引导记录扇区地址找出来，它是扇区读写技术中很重要的核心内容。主引导记录中的最后 2 个字节的结束标志 55AAH。

1.2　分区引导记录

硬盘主引导记录只有 1 个，而每一个逻辑驱动器都有 1 个分区引导记录，如你的硬盘分 C、D、E、F4 个逻辑驱动器，就应该有 4 个分区引导记录，分别存储在逻辑驱动器的第一个扇区。确定分区引导记录所在扇区的地址有两种方法：一种方法是对于 8.4GB 以下的硬盘，在使用前面已经找到的分区表所在扇区的 CHS 地址，“C”和“S”保护不变，“H”加 1 即可。如 D 盘的分区表扇区地址是 261 柱面 0 磁头 1 扇区，则该分区引导记录扇区地址是 261 柱面、1 磁头、1 扇区。另一种方法是对于 8.4G 以上的硬盘，CHS 寻址方式已不适用。必须使用线性寻址方式，在此不多介绍。分区引导记录主要由四部分组成：

1.2.1　BIOS 参数记录块 BPB；

1.2.2　磁盘标志记录表；

1.2.3　分区引导记录代码区；

收稿日期：2005－12－28
作者简介：李恒明（1959－），男，本科，空军航空大学信息中心副教授，从事计算机教学工作。

1.2.4　结束标志 55AA。

与数据恢复有关的只有 1、4 部分，下面简要对它进行一下讨论：

BPB 表所记录的参数，能帮助我们确定磁盘容量大小，文件分配表 FAT 的位置和大小，文件目录表 FDT 的位置和大小，结束标志 55AA 是对扇区进行搜索，用于查找分区引导记录所在扇区的地址的依据。BPB 表的结构与使用的分区格式有关。

1.3　文件分配表 FAT

操作系统或应用程序欲将数据写入一个磁盘文件时，必须在磁盘上找到可以使用的未用扇区，反过来要将数据以磁盘文件中读出时，也要在磁盘上找到已经储存了相应数据的有关扇区。要查找扇区必须知道扇区的地址，文件分配表 FAT 就是记录扇区地址的，因为硬盘的扇区非常多，如果将每个扇区的地址都记录在文件分配表里，势必造成文件分配表体积庞大，查找时效率将会很低，为解决这个问题，采用将扇区分组管理的方法。分组的过程称作扇区分簇，是由高级格式化程序在格式化磁盘时完成的。扇区分簇以后，将每个簇的地址记录到文件分配表 FAT 里去。这样文件分配表的体积就小多了，查找的速度就提高了。一个簇能包含多少扇区，是由分区格式和分区大小来决定的，如在 FAT16 文件系统中，1.2G 的分区它的簇包含着 64 扇区，而同样大小的分区，在 FAT32 文件系统中，簇所包含着扇区数却只有 8 个，可以看出 FAT32 的簇比 FAT16 的簇小得多，这就是 FAT32 能够节省磁盘空间的原因。当需要从磁盘读取文件时，首先从文件目录表中找到该文件的目录登记项，继而从目录登记项的有关字段，查到分配给文件的第一个簇号，根据第一簇号的内容可以计算出两组数据。其中一组数据指出了文件在数据区 DATA 里的第一簇扇区首地址。从第一簇扇区首地址开始数据是连续存放的，连续存放多少个扇区由分区格式和分区大小来决定，另一组数据指出了 FAT 表内簇登记项的地址，如果其值是结束标志 FFFFH（FAT16 格式）或 FFFFFF0FH（FAT32 格式），说明文件到此已经结束，否则该登记项的值为第二个簇号，据此又可以计算出两组数据，继而确定文件在数据区里第二簇扇区首地址和 FAT 表内第二个簇登记项的地址。继续重复上面的过程。就可以得到文件在 DATA 区里的全部数据。以及文件在 FAT 表里所有簇登记项的地址。当需要在硬盘上建立新文件时，其过程与此类似。FAT 表在磁盘文件系统中的地址是非常重要的。为了使磁盘文件操作安全可靠，文件系统的设计者们制定了两个内容相同的 FAT 表。这样当第一个 FAT 表被破坏时我们可以用第二个 FAT 表去修复它，因为第二个 FAT 表很少受到应用程序访问，这时利用“扇区读写编程技术”就有可能从硬盘上恢复出重要的数据来。查找 FAT 表首扇区地址的方法很简单，如果是 FAT16 分区格式的逻辑盘，只要将分区引导记录的 CHS 地址中的 S 加 1，其它不变，如果是 FAT32 分区格式的逻辑盘，则应将分区引导记录的扇区号加上 32。

1.4　文件目录表 FDT

操作系统为了管理磁盘上的目录和文件，在特定的扇区上建立了一个文件目录表 FDT，它是由高级格式化程序 FORMAT 在格式化磁盘时建立的。FAT16 分区格式的 FDT 表占用固定的 32 个扇区，扇区地址紧跟在第二个 FAT 表之后，FAT32 分区格式没有固定的 FDT 表，在第二个 FAT 表之后就是数据区 DATA。目录名和文件名也作为数据对待，存放在数据区内。两种分区格式都使用一个 32B 长的，“目录登记项”来说明目录或文件的有关特性，下表列出了目录登记项的各字段内容：

字节位移	字　节	内容说明
00H	8	文件名
08H	3	扩展名
0bH	1	属性
0cH	10	DOS 系统保留
16H	2	建立或最后修改时间
18H	2	建立或最后修改日期
1aH	2	起始簇号
1cH	4	文件长度

FDT 表的第一个扇区地址的计算可以沿用前面已经计算出的 FAT 表的地址继续推算，用前面计算出的第二个 FAT 表的首扇区地址，加上每个 FAT 表占用的扇区数 256，就得到 FDT 表的首扇区地址。

1.5　数据区 DATA

数据区 DATA 的所有扇区都划分成以簇为单位的逻辑结构，每一簇在 FAT 表里都有一个簇登记项与之对应。在对硬盘的逻辑故障进行修复时，一般只访问主引导记录，分区表和分区引导记录，极个别的情况下可能访问到 FAT 表和 FDT 表。而 DATA 区是访问不到的。但当硬盘的系统控制信息损坏很严重已经没有修复的可能时，为了抢救硬盘上的某些重要数据，我们可以利用“物理扇区读写技术”以簇为单位直接从 DATA 区将数据复制出来。

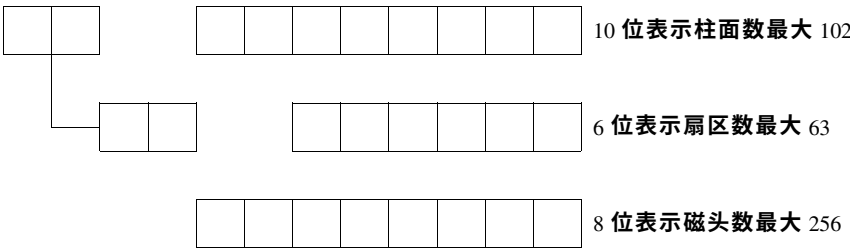
2　扇区读写技术有关的软件接口

硬盘正常工作时必须要与硬盘以外的某些电路或程序进行通信，这种通信是靠接口来进行的，硬盘有两种类型接口：一类是硬件接口，另一类是软件接口，即与某些程序进行通信的接口。应用扇区读写技术来恢复数据时要用到软件接口，这类接口有基本 INT13H，扩展 INT13H 和 ATA。操作系统或应用程序通过基本 INT13 中断或扩展 INT13 中断，调用主板 BIOS 中的磁盘服务程序。该服务程序再将 INT13H 中断请求转换为硬盘的 ATA 接口请求，从而实现对硬盘的数据存取。

2.1　ATA 接口

该接口是寄存器驱动型的并行接口。BIOS 磁盘服务程

序首先往 ATA 寄存器写入磁盘的有关数据，然后把指令按照约定的方式写入 CPU 的特定寄存器，继而实现对硬盘的读写操作。ATA 接口驱动程序有两种寻址方式：CHS 寻址方式和 LBA 寻址方式，CHS 寻址方式是将磁盘的柱面、磁头和扇区的值分别写入 4 个 8 位寄存器，其中两个 8 位寄存器连成 16 位存放柱面值，最大为 65 536。1 个 8 位寄存器存放扇区值，最大为 255。最后一个寄存器分成两个 4 位使用，分别存放驱动器号和磁头值，磁头值最大为 16。CHS 寻址方式所能寻址的最大扇区数是 $65536 \times 16 \times 255$。每扇区 512B、所以 ATA 接口驱动程序所能管理硬盘的容量是：$(65536 \times 16 \times 255 \times 512) \div 2^3 = 127.5G$，而 LBA 寻址方式是把原来分别表示 C、H、S 三个参数的 28 个位，看成是表示扇区地址的一个数，采用 LBA 寻址方式，ATA 接口驱动程序所能管理的硬盘容量是：$228 \times 512 \div 2^{30} = 128G$。



CHS 寻址方式能访问硬盘的最大容量是：

$$(1024 \times 256 \times 63 \times 512) \div 2^{30} = 7.875G$$

基本 INT13H 的 LBA 寻址方式，是将寄存器中原来表示 CHS 参数的 24 个位看成一个完整的 LBA 地址，这时能访问硬盘的最大容量是：$2^{24} \times 512 \div 2^{30} = 8G$。

2.3 扩展 INT13 接口

随着大容量硬盘的迅速普及，只能管理 8.4G 硬盘的寻址方式不能适应了，为了克服这个容量限制，微软公司和几个大硬盘生产商联合制定了一个新标准，改写了 BIOS 磁盘服务程序，重新定义了原来的 INT13H 中断功能，将新的接口称为扩展 INT13H 接口。新的磁盘服务程序不再使用寄存器传递硬盘的地址信息，而是通过主存储器传递一个磁盘地址数据包，该数据包是一个结构变量，长度为 16B，在结构内部定义了一个 64 位的结构成员，以线性方式来表示扇区地址。

在这种寻址方式中，扇区所能够表示的最大值是 2^{64}，扇区寻址范围极大地扩展了，新的磁盘服务程序同时兼容原来的 INT13H 接口规范，所以两种寻址方式可以并存。

3 扇区读写技术的具体应用

扇区读写技术应用范围很广泛，下面介绍几个典型的应用范例：

3.1 利用 INT13H 编制读扇区程序和写扇区程序

将主引导记录和主分区表所在的扇区数据备份到一个文件里而保存起来，以备后硬盘引导系统发生故障时，再将文件中的数据写回扇区中去，达到修复硬盘的目的。写扇区程序也可以按此方法编制。

3.2 修复主引导记录

修复主引导记录的数据非常容易，利用前面介绍的写扇区程序将事先备份的主引导记录写回到 0 号扇区中去，就可达到修复的目的。如果没有备份，找一个能正常启动的硬盘，将其 0 号扇区的数据读出来，作为源数据，然后使用复制数据块的方法把正常的主引导记录复制到损坏的扇区中去就可以了。

3.3 硬盘锁编制

所谓硬盘锁就是硬件或软件方法对硬盘加锁，以达到保护数据的目的，硬盘软件锁程序的编写方法大体有以下几种：

3.3.1 逻辑锁

修改硬盘主引导记录的内容，使引导过程陷入一个死循环，从而使硬盘不能启动；

3.3.2 密码锁

在硬盘主引导记录中加入识别代码，只有当操作者输入的密码被识别正确后，硬盘才能启动；

3.3.3 搬移式硬盘锁

将硬盘的主引导记录或分区表搬到另外的扇区，使系统在启动时找不到应该读取的数据，只有运行操作者的一段特殊程序后，才能正确启动。

这几种硬盘锁利用物理扇区读写技术编制硬盘锁程序。

3.4 清除潜伏病毒

有些病毒程序将其激活代码写入到 0 磁道的扇区中，等到病毒作者设置的条件成立时发作，我们如果利用扇区读写技术编制查看 0 磁道数据变化程序，就可以发现有没有新写入 0 磁道的可疑数据，如果有，及时将其删除，就能使这类病毒失去激活条件，防患于未然。

参 考 文 献

[1] 宋群生. 硬盘扇区读写技术 [M]. 北京：机械工业出版社，2004，(1).

[2] 陈渝生. 数据备份、恢复与急救完全手册 [M]. 重庆：重庆出版社，2003. 8，(1).