Research of File Recovery in Windows FAT32

Na Zhang [1, a] , Jia Wang[2, b]

[1] Department of Computer Science, Dalian Neusoft University of Information, Dalian, 116023, China

[2] Education Resources Development and Training Center, Dalian Neusoft University of Information, Dalian, 116023, China

[a]email: zhangna@neusoft.edu.cn , [b]email:wangjia@neusoft.edu.cn

Keywords: file system; file recovery; DOS boot record; file allocation table; file deleted.

Abstract. Files missing in file system often results in serious consequences. To solve this problem on Microsoft Windows, the file recovery solutions in FAT32 was proposed. File system structure in FAT32 was introduced. The specific processes of DBR recovery in FAT32 were illustrated. On the basis of file storage features in FAT32, an available recovery scheme about the deleted files was analyzed and proposed.

Introduction

Along with the rapid development of information technology, computers have acted a more and more important role in our works and lives and we have paid more and more attention to computer information security. A great deal of data information is stored in computer file system, so the damage or loss of data would result in irreparable consequences. How to recover the lost data correctly and quickly becomes vital.

At present, the market share of Microsoft Windows is more than 92% and FAT32 is the mainstream file system on Windows. So in this paper, the basic principle of file recovery in FAT32 was analyzed and a solution was proposed.

Techniques about File Recovery

Addressing Mode of Hard Disk Driver. There are two addressing modes of hard disk driver. One addressing mode is on the basis of cylinder/head/sector, hereinafter referred to as C/H/S. This three-dimensional mode is the first mode used in hard disk driver. At that time, the disk capacity was very small and each track in the disk has the same number of sectors, so the mode is reasonable. But with the increasing of disk capacity, the same number of sectors in each track will result in disk space wasted. Now the hard disk structure is the isodense structure, i.e., the number of sectors in the outer track is more than in the inner track. In this structure, the hard driver will no longer have the three-dimensional parameters, so the second addressing mode appeared. It is called LBA, short name of the Logic Block Address [1].

In the C/H/S addressing mode, the three-dimensional physical address of sector corresponds with the physical sector on the disk. In the LBA addressing mode, all physical sectors are numbered in ascending order from 0 in some way or rule. In order to ensure optimal performance of the disk, LBA of sector follows the following rule:

(1) No.0 cylinder / No.0 head / No.1 sector→No.0 sector (LBA)

(2) No.0 cylinder / No.0 head / No.2 sector→No.1 sector (LBA)

……

(62) No.0 cylinder / No.0 head / No.63 sector→No.62 sector (LBA)

(63) No.0 cylinder / No.1 head / No.1 sector→No.63 sector (LBA)

And so on.

This is the conversion equation between C/H/S and LBA:

LBA=C * 255 * 63 + H * 63 + (S - 1) (1)

C stands for the cylinder serial number, H stands for the head serial number, S stands for the sector serial number, LBA stands for the LBA of sector.

File System. For the convenience of access and storage, the information is stored in the form of file in the hard disk. File system is a method or a structure that is used to identify files on disks or partitions by operating system, that is a method used to organize files. File system is used to record the used and the free spaces on the storage, organize directories and files and record the physical address of a file, etc. Every operating system has a file system and every file system has its own logical organization.

Cluster. A cluster is the minimum unit of file storage and is also the basic unit of file access by operating system. A cluster is constituted of some sectors. In Windows, a cluster contains 2^n sectors and n is an integer. Before Windows 2000, the maximum number is 6 for n. After Windows 2000, the maximum number is 7. A file's space is one cluster even though its size is one byte.

File System Structure in FAT32

FAT32 file system can support a partition whose size is greater than 2G. The partition is constituted of DOS boot record sector (DBR, for short), file allocation table (FAT, for short) and data sectors. FAT32 treats a directory as a file, so no independent directory sectors for directory and all directory items stored in data sectors.

DBR. DBR is the abbreviation for DOS boot record. DBR is the first sector on logic driver and is constituted of jumping instruction which takes 3 bytes, manufacturer identification and operating system version which take 8 bytes, BIOS parameters block (BPB, for short) which takes 79 bytes, DOS booting program which takes 420 bytes and ending mark ("55AA") which takes 2 bytes[2].

Jumping instruction is constituted of jumping action ("EB"), booting program offset and null instruction ("90"). DOS booting program finishes positioning and loading IO.SYS and MSDOS.SYS files which are core system files in operating system. BPB is a description of partition including cluster size, FAT size, partition size and file system format etc.

FAT. FAT is the abbreviation for file allocation table. Cluster number of files is recorded in a chain structure in FAT. In order to ensure the safety, there are two FATs in operating system, one is the basic table and the other is the backup table. The length and content in two FATs are the same. The operating system version, partition size, cluster size and so on decide the sector number occupied by every FAT. FAT size is stored in BPB [3].

Every FAT entry (elaborated in the next part) occupies 4 bytes in FAT32 file system. FAT begins with "F8FFFF0F" and the first entry is a special entry that is "FFFFFFFF", the second entry records the used information of No.2 cluster in data area. Cluster is numbered from 2 in FAT32 file system. If the No.2 cluster was unused, the second entry is "00000000". If the NO.2 cluster was bad, the second entry is "FFFFFF7". If the NO.2 cluster was the last cluster of a file, the second entry is "0FFFFFFF". In other situations, the second entry is the next cluster number of a file. In a similar fashion, the used information of the No.n cluster is recorded in the No.n entry. So, the length of FAT is decided by the number of the clusters included in data area of partition.

File Directory Entry in FAT32. Every file or directory corresponds to at least one file directory entry. There are some information recorded in file directory entry, such as, the name of file or directory, extended name, file attributes, file size, the first cluster number of file or the next cluster number, creation time and modification time etc. Every file directory entry occupies 4 bytes in FAT32 file system, stored from the No.2 cluster.

The long file name is supported in FAT32, still stored in file directory entry[4]. For long file name, every 13 characters are in one group and every group is represented by one file directory entry. In long file name directory entry, the first cluster number of file, file size and date are not stored but directory sequence number, and 13 characters of file name etc. So, one file corresponds to several entries and these entries are in the reverse order in FAT. For the sake of distinguishing files name, long file name is coded in Unicode, occupied in 2 bytes.

In order to a long file name correctly read in operating system or program of the lower version, file system automatically creats a short file name for the long file name. Thus, the file data can be addressed in the long file name at the same time in the short file name. The short file name directory entry is stored after the long file name directory entry, including the short file name and file attributes and so on.

DBR Recovery in FAT32

DBR as the first sector of partition is very vital. DBR is responsable for managing file system structure and if the partition includes operating system, DBR is also responsable for booting operating system. So, if DBR was destoryed, partition would not be accessed.

There are two methods for DBR recovery. One is using the DBR backup. DBR backup is stored in the No.6 sector in FAT32 partition. Usually, if partition was formatted, DBR and its backup would be destroyed at the same time. In this case, the other method can be used.

By this means, first of all, a copy of DBR from other FAT32 system need be found and it is copied to the DBR sector of destroyed partition. After that, the emphasis is recovering several important parameters, including the sectors number of DBR, FAT size, the sectors number of a cluster and the partition size.

By searching the hexadecimal value “F8FFFF0FFFFFFFFF” in 0 offset of a sector, the basic FAT will be found and its LBA is assumed as FAT_1_LBA, the sectors number of DBR assumed as FAT_1_LBA. Then by the same searching, the backup FAT will be found and its LBA is assumed as FAT_2_LBA. The following is assumed.

Total_Size: the total sectors number;

Data_Size: the total sectors number in data area;

FAT_Size: FAT size;

S_per_C: the sectors number of a cluster.

Then there are the following formulas:

$$FAT\_Size=FAT_2\_LBA-FAT_1\_LBA \quad (2)$$

$$Data\_Size=Total\_Size-63-FAT_1\_LBA-2* FAT\_Size \quad (3)$$

$$S\_per\_C= Data\_Size/(FAT\_Size*512/4) \quad (4)$$

By these formulas, the results are computed and written into the DBR sector according offsets, so DBR is recovered successfully.

File Recovery after Deleted in FAT32

When a file is deleted from file system, the first byte is modified to “E5” in its file directory entry. If cluster number is great where this file is stored, the high bits in the cluster number will be reset to zero. At the same time, all FAT entries of this file in FAT will be also reset to zero. In this case, if this file is fragmental (that is, it occupies discontinuous clusters), or the data area of this file is covered, or the high bits of the file cluster number is reset to zero, this file will not be recovered. Byeond all that, the following method can be used to recover this file[5].

According to the directory of the deleted file, searching and reading file directory entries from the root directory. After file directory entry of the deleted file found, it is read to RAM and file name, the first cluster number and file size are read according to the offset. At last, data are read from the first cluster number up to enough file size in data area and written into a new file.

File Recovery after Formatted in FAT32

When partition is formatted as original file system format, basic FAT, backup FAT and root directory entry will be reset to zero, but other entry and file information will not be cleared. Thus, discontinuous files, files in root directory and covered files will not be recovered. Byeond all that, the above method can be used to recover these files.

Summary

In this paper, FAT32 file system structure was analyzed with emphasis on FAT structure and file storage principle in FAT32. After the recovery principle of DBR in FAT32 was elaborated, how to implement recovery of the deleted or formatted was proposed. With further research, the file recovery of NTFS file system will be concerned and file recovery tool will be developed.

References

[1] Na Zhang: Applied Mechanics and Materials, Vol.336-338 (2013), p.2221

[2] Zhongxia Wang, Wei Li: Advanced Data Recovery Technology, Publishing House of Electronics Industry, Beijing (2007). (In Chinese)

[3] Wenping Li: Application Research in Data Recovery Technology, Anhui Agriculture University, Anhui (2008). (In Chinese)

[4] Xiuyu Zhong: Computer Applications and Software, Vol. 25 (2008), p. 57. (In Chinese)

[5] Information on http://bbs.xlysoft.net/showforum-69.aspx